Inversion of seismoacoustic data using genetic algorithms and *a posteriori* probability distributions

Peter Gerstoft
SACLANT Undersea Research Centre, 19138 La Spezia, Italy

(Received 2 May 1993; accepted for publication 11 June 1993)

The goal of many underwater acoustic modeling problems is to find the physical parameters of the environment. With the increase in computer power and the development of advanced numerical models it is now feasible to carry out multiparameter inversion. The inversion is posed as an optimization problem, which is solved by a directed Monte Carlo search using genetic algorithms. The genetic algorithm presented in this paper is formulated by steady-state reproduction without duplicates. For the selection of "parents" the object function is scaled according to a Boltzmann distribution with a "temperature" equal to the fitness of one of the members in the population. The inversion would be incomplete if not followed by an analysis of the uncertainties of the result. When using genetic algorithms the response from many environmental parameter sets has to be computed in order to estimate the solution. The many samples of the models are used to estimate the *a posteriori* probabilities of the model parameters. Thus the uniqueness and uncertainty of the model parameters are assessed. Inversion methods are generally formulated independently of forward modeling routines. Here they are applied to the inversion of geoacoustic parameters (P and S velocities and layer thickness) in the bottom using a horizontally stratified environment. The examples show that for synthetic data it is feasible to carry out an inversion for bottom parameters using genetic algorithms.

PACS numbers: 43.30.Ma, 43.30.Pc, 43.60.Pt, 43.40.Ph

INTRODUCTION

The inversion of sound fields for determining the unknown environmental parameters can be separated into four parts: (1) discretization of the environment and discretization or transformation of the data; (2) efficient and accurate forward modeling; (3) efficient optimization procedures; (4) uncertainty analysis.

Item (1) is concerned with how to collect and discretize a wave field in order to have the necessary physical information available for the inversion, and also to determine which parameters it is feasible to invert for. Item (1) leads to a set of known environmental parameters and *a priori* bounds for the unknown parameters. Based on the above parameters a matched field can be computed by the forward acoustic model, item (2). Through an iterative scheme, item (3), the match between the observed and computed data is maximized by varying the environmental parameters. From the best models obtained, it is possible to provide estimates of the value of the parameters and their uncertainty and importance, item (4). The best solution is not very interesting without a proper statistical analysis of the result.

A complete inversion requires equal attention to all four items, but it is also clear that each item depends on its predecessor. Therefore it is natural that earlier research has focused on the first two items. The present paper is concerned with the development and application of items (3) and (4) to seismoacoustic problems.

This paper is organized as follows. In Sec. I an overview of optimization techniques is given with global optimization in mind, followed in Sec. II by an overview of uncertainty analysis. In Sec. III the theory and implementation of genetic algorithms is presented, while Sec. IV describes the statistical analysis method. Finally, in Sec. V the approach is applied to the inversion of geoacoustic parameters.

I. OPTIMIZATION METHODS

The nonlinear inverse problem can be stated as an optimization problem: Find the model vector $\mathbf{m}$ that minimizes the quadratic deviation

$$\phi(\mathbf{m}) = \| \, |\mathbf{d}_{\text{obs}}| - r|\mathbf{d}_{\text{cal}}(\mathbf{m})| \, \|^2 / \|\mathbf{d}_{\text{obs}}\|^2, \tag{1}$$

where

$$r = \frac{\|\mathbf{d}_{\text{obs}}\|}{\|\mathbf{d}_{\text{cal}}(\mathbf{m})\|}. \tag{2}$$

Here, $\|\cdot\|$ is the 2 norm of a vector, $|\cdot|$ is the absolute value of each observation in a vector, and ϕ is the object function. Minimizing this object function is similar to maximizing the ambiguity function in matched-field processing based on the correlation coefficient. Normally, in genetic algorithms the fitness function is maximized, but here we minimize the fitness function defined by Eq. (1). Here, $\mathbf{m}$ is the model vector consisting of the physical parameters and $\mathbf{d}_{\text{obs}}$ and $\mathbf{d}_{\text{cal}}$ are vectors containing the observed and calculated data, respectively, of n observations. These observations could be from n_1 ranges, n_2 depths, and n_3 frequencies, yielding a total of $n_1 n_2 n_3$ observations. The calculated data are obtained by calling the forward modeling routines with the model vector as input.

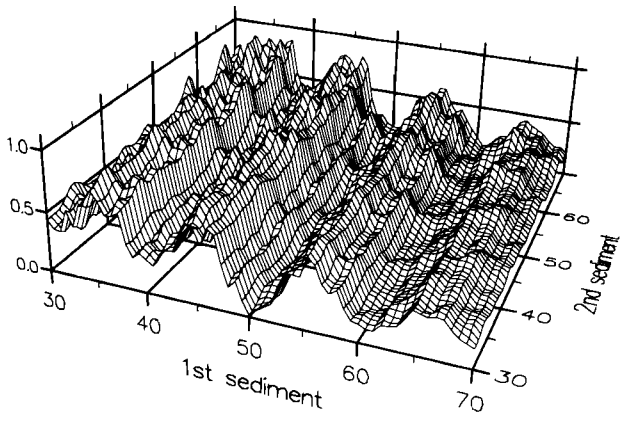

FIG. 1. Slice through the object function. All but the thickness of the first sediment layer and the thickness of the second sediment layer are kept to their correct values for the environment given in Table I.

If the phase information is available and reliable then it is better to use a modified version of the Bartlett processor as the object function,

$$\phi(\mathbf{m}) = \sqrt{1 - \frac{\|\mathbf{d}_{\text{obs}} * \mathbf{d}_{\text{cal}}(\mathbf{m})\|^2}{\|\mathbf{d}_{\text{obs}}\|^2 \|\mathbf{d}_{\text{cal}}\|^2}}. \tag{3}$$

This will give a less oscillating object function than Eq. (1).[1]

The object function ϕ has many names depending on the application and inversion method. In simulated annealing ϕ defines the energy function which is minimized. In genetic algorithms it is common to maximize the fitness, which is here defined as $1-\phi$, or it could be defined as the correlation coefficient. Here, as in simulated annealing, we are minimizing the object function.

Often the object function is highly oscillating, as indicated in Fig. 1, where a slice through a sample object function is shown by keeping all but two parameters constant, corresponding to the thickness of the first and second sediment layer. In order not to get trapped in a local minimum the starting point for both parameters must be within 5 m from the true minimum. For such a problem traditional local methods will have little chance of arriving at the correct solution, and other methods therefore have to be explored. Two approaches exist for resolving this problem: Reparametrization of the problem or use of global optimization methods.

The oscillations of the object function can often be reduced by another parametrization of the model. By transforming the data to another domain or by using a subset of the observed data to obtain a local problem, the object function can become more regular, and if there are only a few minima it can likely be inverted by a local method. Naturally, such a local method should in general be used since it is much faster. Moreover, the minimum will always be reached, provided it is a local problem. This method, however, requires a detailed knowledge of the object function and reparametrization may be impossible in multiparameter inversions. The local methods can further be stabilized by use of singular-value decomposition and regularization techniques.[2–4] Using the above approaches the compressional velocity of the ocean and bottom has been determined by inverting the modal eigenvalues in the water column,[3] and more recently, the bottom shear velocities have been determined by inverting the group velocity of the bottom interface wave.[5]

Global optimization methods accept that the object function is irregular and try to find the global minimum, without doing an exhaustive search. Advantages of global optimization are that it only requires the value of the object function at arbitrary points in space and the problem can then be solved without any further knowledge of the object function. It is thus expected that once the global inversion method has been tuned, any forward modeling method can be used without much change in the optimization parameters. Early solutions to the global problem were attempted using a simple Monte Carlo method, whereas modern methods use directional searches such as *genetic algorithms* (GA) or *simulated annealing* (SA).

An introduction to GA is given by Goldberg[6] and Davis.[7] The implementation used here is described in detail in Sec. III. GA are based on an analogy with biological evolution, one of the most efficient optimizing systems. GA have, to our knowledge, not yet been used in the underwater acoustic community, but these algorithms have provided promising results in the seismic community.[8–11]

The basic principle of GA is simple: From all possible model vectors, an initial population of q members is selected. The fitness of each member is computed based on the difference between the observed data and the computed data. Then through a set of evolutionary steps the initial population evolves in order to become more fit. An evolutionary step consists of selecting a parental distribution from the initial population based on the individual's fitness. The parents are then combined in pairs, and operators are applied to them to form a set of children. The operators are traditionally the crossover and mutation operators, see Sec. III. Finally the children replace part of the initial distribution to get a more fit population.

Current implementations of GA in the geophysical field are based on the total replacement algorithm, where the whole population is replaced for each generation. It has here been found that the steady-state replacement algorithm, where zz the least-fit fraction f of a population is replaced in each iteration, see e.g., Ref. 7, is both more robust and quicker than current implementations, see Sec. III and the examples in Sec. V.

Simulated annealing (SA)[12] has, since its rediscovery,[13] received much attention; first in seismic exploration applications and later in underwater acoustics.[1,14–16] The two main problems for SA are to find an adequate cooling schedule and to be able to control the "move class." The best annealing schedules are found by taking advantage of statistical information acquired during annealing, see, e.g., Refs. 17, 18. While control of the temperature is not a problem, little success has been obtained with controlling the "move class," i.e., the set of possible neighbors which the current model can move to. This is believed to be as important as using a good method for determining the temperature. For example, in the start of

the optimization we first determine some intervals for the most important parameters and later, while refining the former parameters, other parameters can then be determined. While SA cannot adaptively change the move class during the annealing, GA are very efficient in this respect. GA work simultaneously on several model vectors during the iterations. Thus when one model parameter is more important than another, the model vector containing the best fit for this parameter is more likely to be selected as a parent. When after some iterations a good fit is found for this parameter other parts of the model vector will become important.

Recently, several comparisons between SA and GA have been made for geophysical applications, generally favoring GA.[7–10] A meaningful comparison, however, is difficult to perform. Both SA and GA can be described by a Markov chain for meandering in search space. To obtain the next guess of the solution, GA can describe a wide range of operators, whereas SA just forms a random move from the current model. In this sense SA is more restricted and the developers of SA concentrate on how to regulate the temperature or find some essential temperatures where the transition of the optimization develops rapidly.

The difference between SA and GA can be further understood by considering the avenue of Monte Carlo methods. The simplest Monte Carlo method is based on one member meandering in the search space, randomly selecting the next step. This search can be significantly improved by applying some simple rules for meandering in the search space, as exactly both SA and GA do. SA is based on a single member meandering in the search space, while GA is based on a population which intercommunicates while meandering in the search space. The advantage of intercommunicating individuals is that the next guess of the solution can be based on operators combining information from all the individuals. Both SA and GA can be performed in parallel. This increases the possibility that the global minimum will be found and it also becomes possible to estimate the probability distribution for the solution.

II. UNCERTAINTY ANALYSIS

For an oscillating partially sampled object function, one can never be sure of having obtained the absolute minimum. In the case of data contaminated by noise, the global minimum might not be the global minimum in a noise-free environment, as the noise can offset this minimum. However, the minimum obtained in a noisy environment will still be the maximum likelihood estimate of the solution. In order to include this uncertainty it seems more appropriate to describe the solution to the inversion problem in terms of statistics, see, e.g., Refs. 18–20. This is done in a Bayesian framework. Guided by GA, samples are taken from the *a priori* distribution to obtain the *a posteriori* distribution. This function combines our knowledge of theory and all prior information about possible geoacoustic models. The *a posteriori* distribution can then be used as an input to the next step in the inversion, as described in the example in Sec. V C, or combined with other information to obtain a new probability distribution. This could, for instance, be the probability for a certain arrival time or the probability for finding oil.

The difficulty of solving the optimization problem is determined both by the number of local minima in the problem and the size of the search space. While the number of minima is not known *a priori*, the size of the search space is easily computed as the product of the possible values for each parameter. Thus we use the size of the search space as a measure of the complexity of the problem. The uncertainty and the size of the search space are closely related since a large problem is harder to solve and thus more uncertain than a smaller problem. In order to limit the search space, each parameter should only be discretized so finely that the difference between two neighboring values can be seen in the data. Naturally to know this *a priori* requires some expert knowledge. *A priori* knowledge could also exclude some parameter combinations or weight the search space according to where the minimum is most likely to be located.

A resolution and variance analysis[21] is an important part of the solution for locally linear methods. Here it will be shown that similar tools are available for global methods; both the marginal probability distribution for each parameter and the correlation coefficient will be computed, see Secs. IV and V.

III. GENETIC ALGORITHMS

The seismoacoustic environment will be discretized into M environmental parameters contained in a model vector $\mathbf{m}$. Each of these parameters, $j=1,...,M$, can take 2^{n_j} discrete values according to a rectangular probability distribution, where n_j is the number of bits in the jth parameter string. The *a priori* distribution of each parameter could easily be chosen as another distribution, such as a Gaussian probability distribution. The *a priori* rectangular distribution is given by the lower and upper bounds, $\theta_j^{\min}$ and $\theta_j^{\max}$. The parameter is discretized into 2^{n_j} values:

$$\theta_{j,i_j}=\theta_j^{\min}+\Delta\theta_j i_j, \quad i_j=0,...,2^{n_j}-1, \tag{4}$$

where

$$\Delta\theta_j=(\theta_j^{\max}-\theta_j^{\min})/(2^{n_j}-1). \tag{5}$$

Only when the forward modeling routines are called, is the model vector used. For the GA only the integer values i_j in Eq. (4) are used, but Eq. (4) defines a unique relationship between the two representations. For each parameter the integer values of the model vector are represented as a binary string of length n_j, where n_j is the number of bits in the jth parameter string. The reason for this binary coding is that the crossover and mutation operators, which are described later, work on the binary string. In accordance with natural terminology this bit string is called a gene. A given point in the data space can then be represented by a list of the M model parameters represented as a binary string, each of the binary strings having a length n_j. This gives a total search space of $\Omega = \Pi_{j=1}^{M} 2^{n_j}$.

To initialize the population, q members are selected randomly from the search space, and for each of these members $k=1,...,q$, the corresponding object function, Eq. (1), or in GA terminology, the fitness function $1-\phi(\mathbf{m}_k)$, is computed. The members in the generation are sorted according to their object function, $\phi(\mathbf{m}_1)\leqslant\phi(\mathbf{m}_k)\leqslant\phi(\mathbf{m}_q)$, where the most fit, i.e., these that have the best match to the data, are the ones with the lowest object function.

In order to establish a new population, also with q members, fq parents must be selected, where the fraction $0<f<1$. Each parent is chosen by assigning a probability to each of the q models based on their fitness. In the simplest selection method—stochastic sampling—the probability for selection of a model for reproduction is based on the ratio

$$p_k=\frac{1-\phi(\mathbf{m}_k)}{\Sigma_{l=1}^{q}[1-\phi(\mathbf{m}_l)]}, \quad k=1,...,q, \tag{6}$$

where $1-\phi(\mathbf{m}_k)$ is the fitness of each member in a population as given by the sum in Eq. (1). Thus the new candidates for reproduction are chosen in a probabilistic manner according to Eq. (6). In order to favor the best candidates, the object function is very often stretched before calculating the probability for selection, Eq. (6).[6,7] Stretching is the mapping of $\phi(\mathbf{m}_k)$ for all individual to another domain where the values can either be closer together or, normally, stretched further out. The purpose of this stretching is in the start of the evolution to assign a more equal probability to all parameters than in Eq. (6), and at later stages to put more emphasis on the fittest of the members.

By introducing the most successful control parameter in SA, the temperature, into the possibility of selection, very good results have been obtained for GA.[10] Thus an alternative probability of selection is

$$p_k=\frac{\exp[-\phi(\mathbf{m}_k)/T]}{\Sigma_{l=1}^{q}\exp[-\phi(\mathbf{m}_l)/T]}, \quad k=1,...,q, \tag{7}$$

where T is the temperature as defined in SA. From the behavior of the exponential function it can be seen that for high temperatures all model vectors are almost equally likely to be selected, whereas for low temperatures model vectors with lower values of the object function are more likely to be selected. Thus for low temperatures even small differences in the object function can be discriminated. An essential point is to choose a proper cooling schedule. In the SA literature advanced schedules have been found by studying the analogy with statistical physics,[18] or finite-time thermodynamics.[17] They are, however, not easy to implement in practice. A low temperature will drive the current population toward the minimum closest to the most fit member in the population. The disadvantage of a low temperature is that the variety in models will be lost by only selecting the best members. A good compromise between the two criteria is obtained when the temperature is of the same magnitude as the object function of the population, i.e., $T\approx\phi(\mathbf{m}_k)$. Experience has shown that choosing the temperature equal to the fittest in each generation gives good results. As the evolution continues the value of the object function will decrease and the temperature will also become lower.

After the selection of parents, several operators will be applied to the parents to form a new generation. These operators could work on real numbers and could be, for example, mean or random operators. Traditionally, this consists of the crossover operator and the mutation operator for the parameters coded in a binary format. In the present paper only traditional operators have been used.

The crossover is the first part of reproduction. For each set of parents, each consisting of a model vector, two children are constructed, and for each parameter in the model vector each child may either be a direct copy of one parent, with probability $1-p_x$, or it can be a bit crossover of the two parents with crossover probability p_x. The crossover is done by splicing together pieces of the binary string (genes) copied from the two parents. Let the binary strings for one parameter in the model vector for each of the two parents be given by

$$(\alpha_0,...,\alpha_{N-1}) \quad \text{and} \quad (\beta_0,...,\beta_{N-1}), \tag{8}$$

then a crossover point l is randomly selected from the interval $[1,N-1]$ and the two children are now given by

$$(\alpha_0,...,\alpha_{l-1},\beta_l,...,\beta_{N-1})$$

and (9)

$$(\beta_0,...,\beta_{l-1},\alpha_l,...,\alpha_{N-1}).$$

Since this is done for each parameter this method is called multiple-point crossover. There also exists the classical single-point crossover where the whole binary parameter vector is concatenated into one string and the children are determined by crossover from this string. This describes the classical crossover operator, but in general it could be any operator combining the information in the two parameters.

After the crossover operation, each bit of the parameter vector can be perturbed with low mutation probability p_m ($p_m\approx0.05$) in order to better explore the search space. This process is referred to as a mutation operator, which helps to insure that the process does not get stuck in a local minimum.

Before the children replace the least fit members of the intial population, a check is made that all the children are different and that none of them are present in the initial population. Thus in the steady-state algorithm, all q members of a generation will be different. This is in contrast to the total replacement method, where a large number, say 90%, of a generation are identical when the optimization has matured.

It is possible that one run of a GA will approach a local minimum. In order to increase the probability of finding the global minimum, several independent parallel populations M_{par} are started. This is similar to the ensemble approach in SA.[22,23] This is also advantageous for collecting statistical information to estimate the probability densities, as described below.

IV. *A POSTERIORI* STATISTICS

During the inversion process the samples of the search space are saved. From these observations the *a posteriori* probabilities can be estimated. For the problem of estimating N parameters this results in an N-dimensional space. To display part of this probability function two methods are followed. Tarantola *et al.*[24] recommend displaying the most likely model vectors and their relative probabilities. Frazer and Basu[18] use a graph-binning technique to sample the marginal probability density for each parameter. By this method the most likely model vectors may not be resolved, but useful statistics can be retrieved.

The marginal probability density functions can be estimated by sampling the model vectors as the evolution progresses. These samples are ordered according to their energy, value of object function, and when forming the probability distribution they are weighted according to a Boltzmann distribution, similarly to the weighting performed for the optimization, Eq. (7). Choosing the temperature equal to the energy of the fittest in the sample will favor the fittest part of the population, and choosing the temperature equal to the energy of the least fit will correspond to a more even weighting of the population. Experience has shown that a good temperature is the average of the best 50 samples. The probability for the kth model vector is then given by

$$\sigma(\mathbf{m}_k)=\frac{\exp[-\phi(\mathbf{m}_k)/T]}{\sum_{k=1}^{N_{\text{obs}}}\exp[-\phi(\mathbf{m}_k)/T]}. \tag{10}$$

For the ith parameter in the model vector the marginal probability distribution for obtaining the particular value κ can be found by summing Eq. (10):

$$\sigma^i(m^i=\kappa)=\frac{\sum_{k=1}^{N_{\text{obs}}}\exp[-\phi(\mathbf{m}_k)/T]\delta(m_k^i=\kappa)}{\sum_{k=1}^{N_{\text{obs}}}\exp[-\phi(\mathbf{m}_k)/T]}, \tag{11}$$

where N_{obs} is the number of observed model vectors and T is the temperature. It has been found that when several parallel runs are executed, it is sufficient to save the best part of the obtained model vectors in a population, $(1-f)q$ in each of the M_{par} populations. This is sufficient because in the steady-state algorithm the fittest part of a generation is kept for each generation.

The *a posteriori* mean value and covariance of the model parameters can also be estimated:

$$E(\mathbf{m})=\sum_{k=1}^{N_{\text{obs}}}\mathbf{m}_k\sigma(\mathbf{m}_k), \tag{12}$$

$$C(\mathbf{m})=E\{[\mathbf{m}-E(\mathbf{m})][\mathbf{m}-E(\mathbf{m})]^T\} \tag{13}$$

$$=\sum_{k=1}^{N_{\text{obs}}}\mathbf{m}_k(\mathbf{m}_k)^T\sigma(\mathbf{m}_k)-E(\mathbf{m})E(\mathbf{m})^T. \tag{14}$$

The diagonal of the covariance matrix is the variance of each parameter. It is useful to normalize the covariance matrix to obtain the correlation coefficient. For the correlation coefficient the off-diagonal terms show how different model parameters interact. A value of 1 means that, the parameters are fully correlated, 0 uncorrelated, and -1 negatively correlated.

TABLE I. Environmental model for inversion.

Medium	Lower interface m	Compression Speed m/s	Compression Atten. dB/λ	Shear Speed m/s	Shear Atten. dB/λ	Density kg/m^3
Water	50	1500	0	···	···	1000
Sediment 1	100	1600	0.1	500	0.2	1600
Sediment 2	150	1800	0.1	1000	0.2	2000
Basement	···	2800	0.1	1500	0.2	2200

V. EXAMPLES

Since GA are formulated independently of the forward modeling routine, it can easily be adapted to any acoustic propagation code. The versatility of GA is illustrated by applying either OASES,[25] Secs. IV A–C, or SNAP,[26] Sec. IV D, as the forward model.

The OASES program is an enhanced version of the SAFARI program;[27] it is a wave number integration code and assumes a horizontal stratification of the environment. The comparison between the observed and computed data here is carried out in the horizontal wave number domain. Thus the real observed data are recorded on a horizontal array. For real data, the estimation of the wave number spectrum can cause some problems since quite large apertures must be used in order to obtain reasonable estimates, see, e.g., Ref. 28. The advantage of using the wave number domain is that the output of the forward modeling is in this domain and that fewer wave numbers are required for the inversion than for reconstructing the real field. The physical parameters which could be inverted for each sediment layer are the thickness, P-wave velocity and attenuation, S-wave velocity and attenuation, and the density.

The examples presented in Secs. IV A–C are all based on the environment given in Table I. Thus the observed data will be computed using this environment. In order to obtain the maximum response from the bottom, both source and receivers are placed on the seabed. The source frequency is 100 Hz and the zz magnitude of the horizontal wave number spectrum is computed at 64 points in the phase velocity range from 1200 to 3000 m/s. First the convergence of the algorithm is assessed and compared to SA. Thereafter the algorithm is applied to obtain the compressional velocity profile and to estimate both the compressional and shear-velocity profiles and the thickness of the sediment layers. The sensitivity of the physical parameters to variation in the sampling parameters is explored.

The examples here are run by fixing some parameters and letting others vary. In a real inversion, all parameters would be allowed to vary, but some of the parameters would be more certain than others. The available computer time should be used with care. Thus the *a priori* importance and uncertainty of each parameter should be used to determine how much computer time should be used on each parameter. This can be expressed in terms of how

many discretizations, or bits, each parameter should have. For example, the P velocity would probably have more bits than the P attenuation, and P velocities close to the source and receivers will have more bits than those far away from the source and receivers.

A. Tuning the GA parameters and comparison to SA

The environment given in Table I was used to create the (synthetic) observed data. For the inversion it was assumed that the bottom consisted of 11 layers each 10 m thick. Thus the first five layers correspond to sediment 1, layers 6–10 to sediment 2, and layer 11 to the basement layer. In each of these, the P velocity could vary between 1500 and 3100 m/s with 256 discrete values, i.e., a discretization interval of 6.25 m/s. All other parameters in the layers are regarded as fixed. The problem size is thus $256^{11} \simeq 3 \times 10^{26}$. Such a large search space is prohibitive for an exhaustive search method.

In the present implementation there are relatively few GA parameters which have to be tuned for each application, and the precise value of each of these does not seem to be very important. Based on our limited experience, the following values are recommended.

(1) The population size q should be large enough that the model vectors can represent several minima, but also small enough that several iterations can be performed; $q=64$ seems to be a good compromise.

(2) The reproduction size f should be large enough that the fittest individuals stay in the population during the iterations; f should be less than 0.9, here $f=0.5$.

(3) The temperature T should be chosen so that it follows the best fit in each population, $T=\min\{\phi(\mathbf{m}_k)\}$.

(4) The crossover rate depends on how independent the parameters in the model vector are. A crossover rate p_x close to 1.0 seems to be a good choice for independent parameters; for dependent parameters a lower value of p_x is recommended, here $p_x=0.8$.

(5) It has been found that a high mutation rate gives the best result, here $p_m=0.05$.

(6) The number of forward computations for each population should be relatively low, 1000–5000, here 2000 is used.

(7) The number of parallel runs M_{par} depends on the application. To obtain a reasonable estimate of the inversion parameters, $M_{\text{par}}=1$ is sufficient. For computing the probability distribution it must be larger, e.g., $M_{\text{par}}=100$.

Having recast the inversion problem into an optimization problem, the best measure of the convergence is the energy of the fittest individual in the population as a function of computer time. Since most of the CPU time is spent in the forward modeling routines, the amount of computer time is proportional to the number of sample models. The fitness is the relative deviation of the observed and the calculated data as determined by Eq. (1) or (3).

Examples of convergence for different optimization parameters are shown in Fig. 2, where each curve is based on the average of 50 realizations. The energy is here the fittest individual in a population. The best result (solid line) is obtained for $q=64$, $p_m=0.05$, $p_x=0.8$, and $f=0.5$; the other curves are obtained by changing one parameter at a time. A low value of the mutation rate (dotted line) does not converge very well, as it is more likely that it will get trapped in a local minimum. Finally the dashed curve shows the performance for an optimal fixed temperature. The optimal temperature was found by trying several different temperatures. Other simple temperature schedules such as linear and power law schedules were also tried without being much better than when the temperature follows the energy of the population.

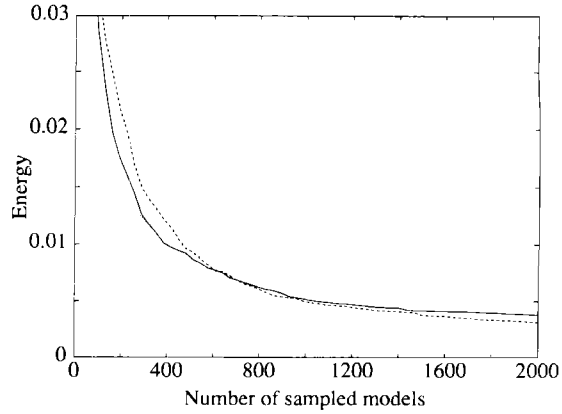

FIG. 2. Convergence of the fitness of the population versus number of models. The basic optimization parameters are $q=64$, $p_m=0.05$, $p_x=0.8$, and $f_g=0.5$. Solid line: only basic parameters. Dotted line: $p_m=0.01$. Dashed line: temperature is constant and equal to 0.5.

Using the same environmental setup, the GA algorithm was compared to the successful SA implementation of Collins and Kuperman,[14] Fig. 3. This SA implementation is based on the so-called fast SA method of Szu and Hartley.[29] Here it has been used with the starting temperature of 1. The SA was slower for this example. One reason is that the parameters are altered one at a time, a metropolis step. For a large class of inversion problems the inversion parameters can be ordered according to their sensitivity to the object function. The most sensitive parameters will usually be the most important parameters for the object function, and they can be ordered in a hierarchy according to their importance. Therefore, it is important to

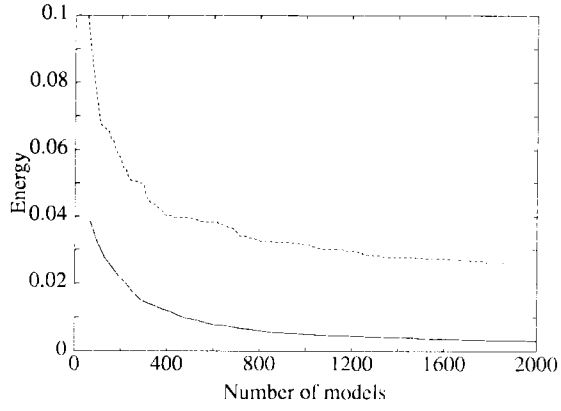

FIG. 3. Comparison between genetic algorithms (solid line) and simulated annealing (dashed line). The genetic algorithms use the same parameters as in Fig. 2.

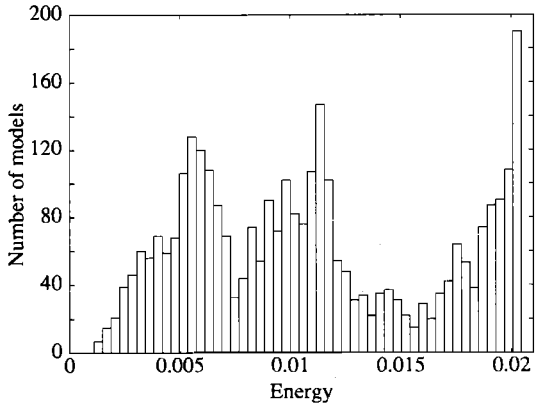

FIG. 4. Histogram of fitness of all observations for the inversion for the 11 P velocities.

TABLE II. The best inverted velocity profile and the difference from the true model.

Layer	P vel. m/s	Diff. m/s
1	1600	0
2	1600	0
3	1594	−6
4	1594	−6
5	1569	−31
6	1806	6
7	1538	−262
8	1800	0
9	1925	125
10	1918	118
11	2912	112

determine one parameter nearly correctly before attempting to find values for the other parameters. For example, in the above example the first velocity must be approximately determined before any of the other parameters can be estimated, so only every 11th model improves the estimated model vector. GA starts with $q=64$ different estimates of the first velocity, and during the optimization it works simultaneously on all parameters. For GA the move classes are changed as the iteration carries on. This is believed to be one of the major assets of GA.

B. Inversion for the velocity profile

The first example illustrates the use of global optimization to find the compressional velocity profile in the bottom. The same environment was used; also the parameters were fixed as when tuning the GA parameters. The inversion is done in $M_{\text{par}}=100$ parallel runs each sampling 2000 models. This gives a total of 2×10^5 modeling runs, which is not much compared to the 3×10^{26} possible models. The optimization part of the inversion was run in order to find the best models from the end of each population, in total $M_{\text{par}}qf=3200$ models. A postprocessor was then run to compute the *a posteriori* probability distribution based on these models.

From the histogram of the best energies, Fig. 4, it is seen that a few of the samples have a good (low) energy, the fittest individual has an energy of 0.001. Thus the mean relative error between the observed and computed data is $\sqrt{0.001}\simeq 3\%$ for the fittest individual. The model giving the best fit to the data, Table II, finds the velocity best in the upper layers, which are more important for the wave propagation in the ocean waveguide than the lower layers. The wave number spectrum based on the 64 sample points for this model and the true model is given in Fig. 5(a). A denser sampling of the wave number spectrum, again using the same best model and the true parameters, Fig. 5(b), shows that even though the inversion has not been based on the modal peaks these are determined quite well.

For all the best models the probability distribution for each parameter is estimated. This can either be done by equal sampling, Fig. 6(a), or by weighting the energy according to a Boltzmann distribution, Fig. 6(b), where the temperature is equal to the average energy for the 50 best models. Both graphs display the marginal probability distribution for the velocity in each of the 11 layers as a function of the velocity. The curves have been scaled so that the black area under each curve is the same. The true profile is the solid line. For equal sampling, Fig. 6(a), there are two peaks in the velocity probability distribution for the 2nd to 5th layer. They correspond to a velocity of about 1900 m/s, though with less probability than the true

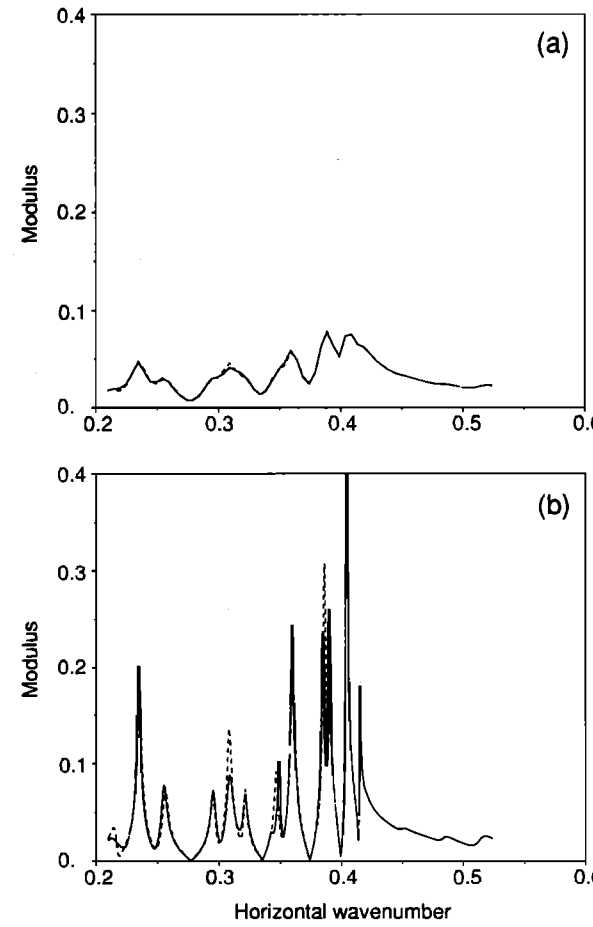

FIG. 5. Comparison of the best model (dashed line) and the true model (solid line) in the horizontal wave number domain: (a) Only 64 sample points, and (b) The densely sampled spectrum (1024 points).

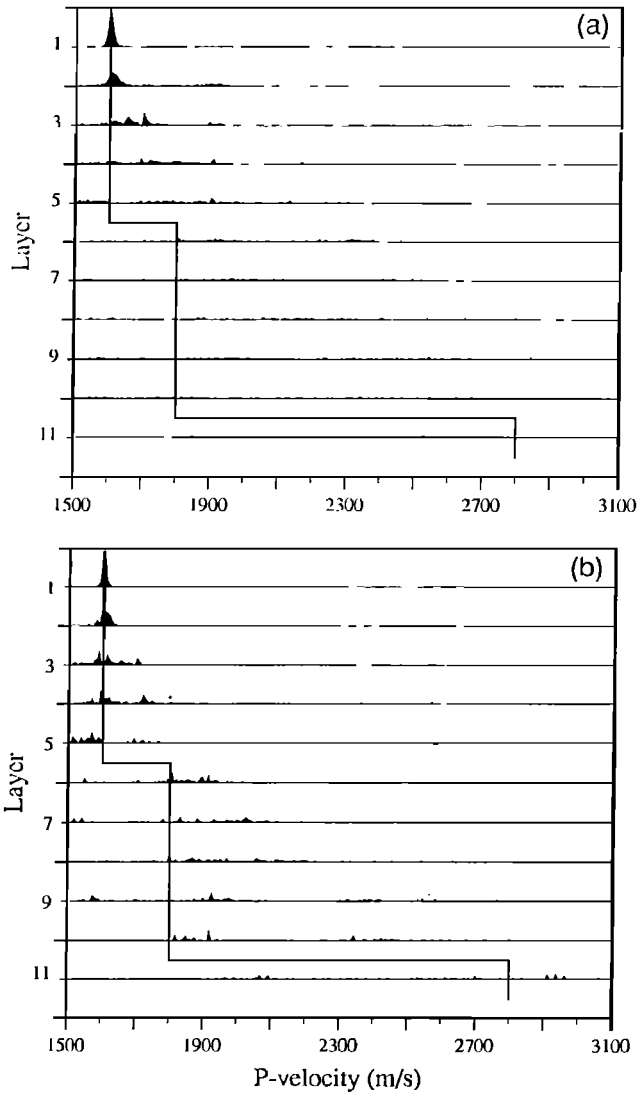

FIG. 6. Probability distribution for each of the 11 P velocities: (a) obtained by an equal representation of all the sampled models; (b) obtained by sampling from a Boltzmann distribution with a "temperature" corresponding to the average fitness of the 50 best models. (The solid line is the true model.)

velocity at 1600 m/s. This is in contrast to Fig. 6(b) where this ambiguity has been removed, and the values of the lower layers are also better resolved. Thus for equal sampling, the marginal probability distribution is more ambiguous than for the weighted probability distribution. In both graphs the velocities closest to the seabottom are best resolved as they receive the most energy. It is also seen that the velocities closest to an interface are less well resolved, i.e., the probability distribution is more spread out. A similar conclusion can be drawn by plotting the most likely model vectors, Fig. 7, for the most likely 20 models. Ideally this should be shown for one model at a time to obtain the variation between each parameter for a given model vector. By showing the 20 most likely models this display becomes similar to showing the marginal probability distribution.

The magnitude of the correlation coefficient between the inverted parameters, here the P velocity in each layer, is shown in Fig. 8. The first velocity in the bottom is determined nearly correctly in all 3200 models, and thus is so relatively dominating that it is only weakly correlated with all the other velocities. The velocities in the lowest layer are so insignificant that they are not coupled to the rest of the parameters

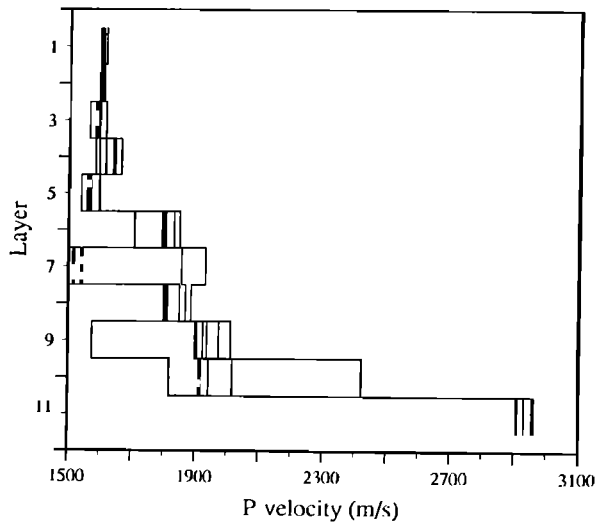

FIG. 7. The 20 most likely sets of velocities for the 11 layers.

C. Inversion for sediment layer properties

For the same environment it is now assumed that the number of sediment layers are known and we then invert for the thickness of the two sediment layers and the three compressional and shear velocities. For this example we again sample 2000 models per population, but with 200 parallel runs. The thickness of the sediments is assumed to be in the range of 20–84 m, the P velocity in the range of 1500–3100 m/s, and the S velocity in the range of 400–2000 m/s; all the parameters can assume 256 values, giving a total search space of $256^8 \simeq 2 \times 10^{19}$.

Again the histogram of the energy, Fig. 9, shows that some of the model vectors have obtained a good (low)

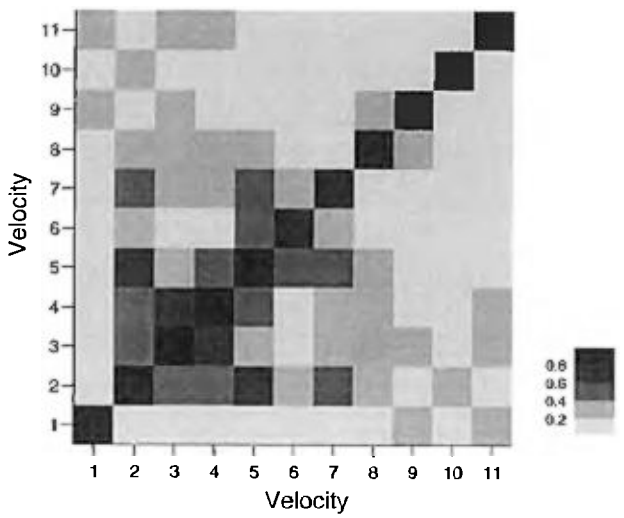

FIG. 8. Correlation coefficient between the 11 layer velocities.

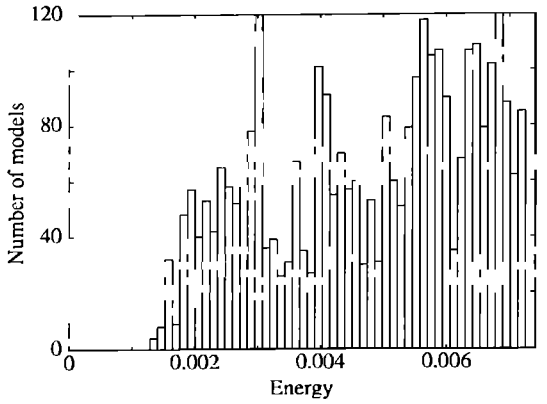

FIG. 9. Histogram of fitness for the inversion of the eight parameters.

energy. The marginal probability density, Fig. 10, is weighted according to a Boltzmann distribution, with a temperature equal to the average of the 50 fittest models. It is here seen that only the P velocity in the first layer is well resolved; for the other parameters there are considerable ambiguities. Some of these ambiguities could be removed by introducing more physical knowledge into the model. For example, the S velocity in layer 1 should not exceed the P velocity and thus the ambiguity at 1600 m/s could be removed.

The magnitude of the correlation coefficient, Fig. 11, between the inverted parameters, shows that the P and S velocities in the first layers are somewhat coupled. It can also be seen that the thickness of the second and the S velocity of the last layer are not coupled to the other parameters very much. This is because they are less important than the other parameters, i.e., they have a flatter distribution in Fig. 10.

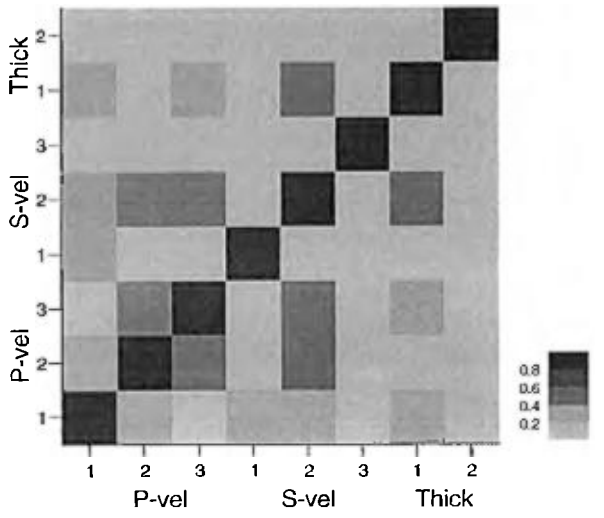

FIG. 11. Correlation coefficient between the eight parameters.

In order to investigate the sensitivity of the inverted parameters to the sampling of the wave field, the wave numbers were sampled in a larger phase velocity window, from 400–3000 m/s. From the physics of wave propagation it is clear that the range of the phase velocity interval

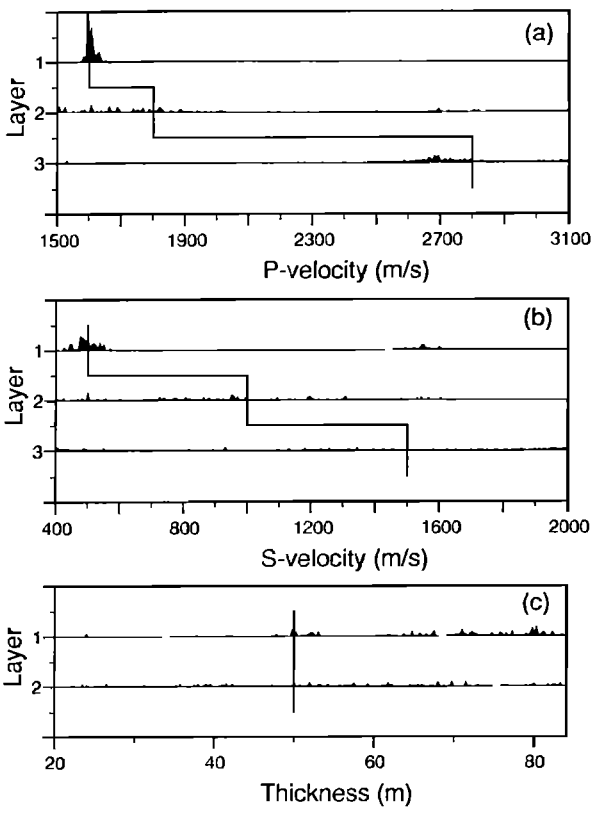

FIG. 10. Probability distribution for the eight parameters: (a) P velocity, (b) S velocity, and (c) thickness. (The solid line is the true model.)

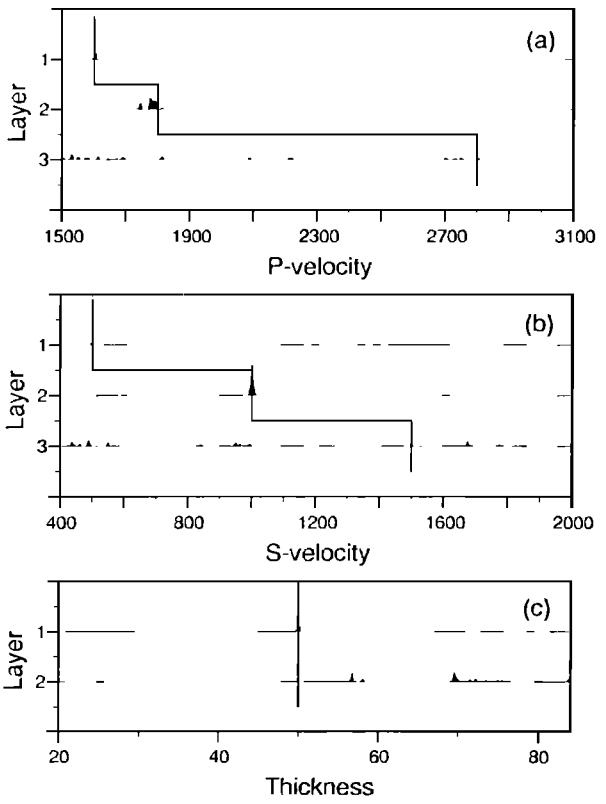

FIG. 12. Probability distribution for the eight parameters using a larger phase velocity interval from 400–3000 m/s, but still only 64 wave numbers: (a) P velocity, (b) S velocity, and (c) thickness. (The solid line is the true model.)

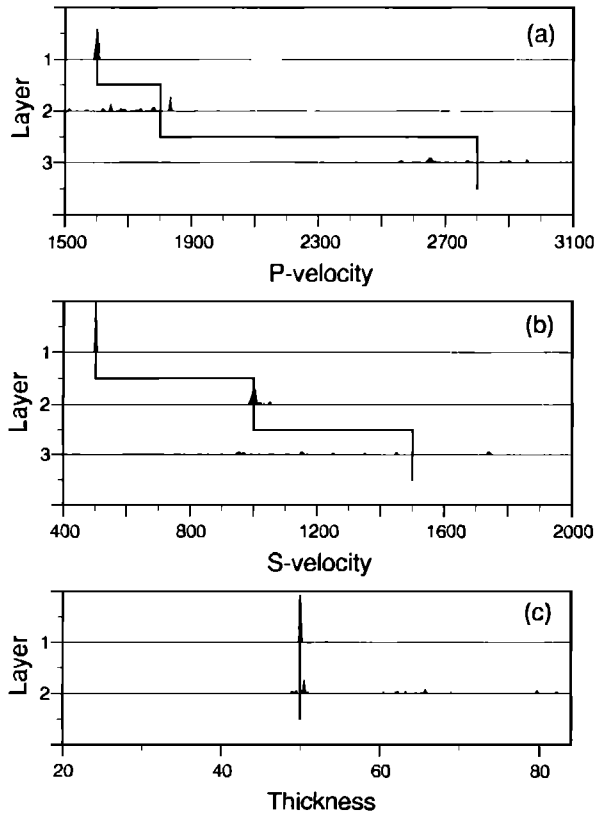

FIG. 13. Probability distribution for the eight parameters using a larger phase velocity interval from 400–3000 m/s and 256 wave numbers: (a) *P* velocity, (b) *S* velocity, and (c) thickness. (The solid line is the true model.)

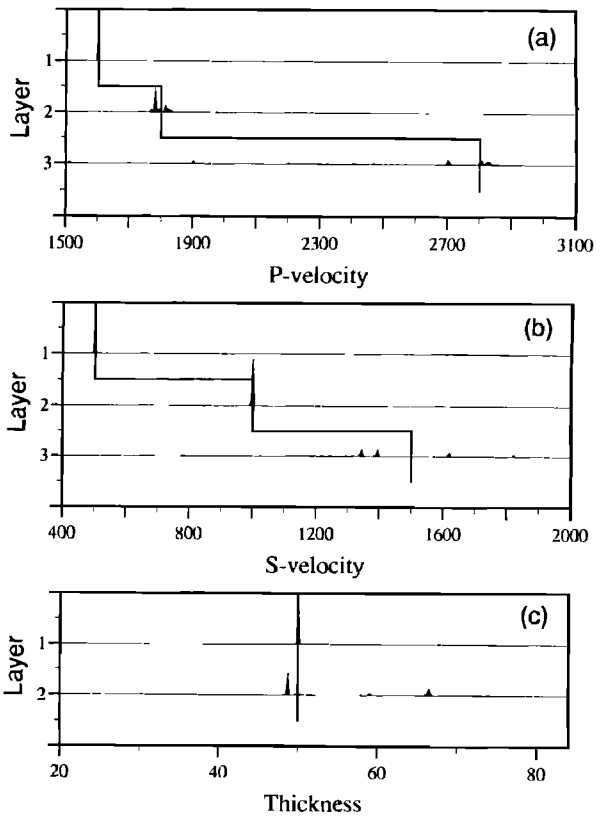

FIG. 14. Probability distribution for the eight parameters using also the phase of the wave number spectrum and using the modified Bartlett processor as the object function: (a) *P* velocity, (b) *S* velocity, and (c) thickness. (The solid line is the true model.)

should, in general, include the sound velocities to be estimated. By sampling at lower velocities it was observed, as expected, that the shear waves were better resolved and that the shallower layers are also better resolved, due to lower grazing angles, Fig. 12. By increasing the number of wave number samples to 256 in this interval, the overall resolution was improved a little, mainly for the lower layers, Fig. 13. In a noisy environment the more information that is available, the more reliable the solution. Thus for data contaminated with noise, more points should give a better estimate of the solution.

Instead of using only the magnitude of the wave number spectrum it might be possible also to extract the phase information. Then it is optimal to use the modified Bartlett processor, Eq. (3). For the larger phase velocity window, this gives, Fig. 14, a less ambiguous *a posteriori* distribution than just using the amplitude alone, Fig. 10.

D. Inversion for source location and environmental parameters

In a recent workshop on matched field processing[30] a blind test was given, Fig. 15, with uncertainties in both the source location and the environmental parameters. The uncertainties in the source depth and range were so large that they became the most important parameters, and it was thus more challenging to find the environmental parameters. Further, colored noise with a S/N=40 dB was applied to the data. The participants in the workshop received the above information as well as the pressure received on a vertical array spanning the water column with 20 receivers from 5 to 100 m.

The nine parameters, see Fig. 15, can all take 51 values, and thus the size of the search space is 2×10^{15}. Since the environment is range independent and it is a long-range propagation problem it is optimal to use normal modes as the forward model. Here a modified version of SNAP[26] is used. For the optimization 125 000 model vectors are sampled using 50 independent populations. Since this could be achieved in 5 CPU hours on an Alpha workstation, no attempt has been made to reduce the number of sampled model vectors.

The result of the inversion is shown in Fig. 16. It is

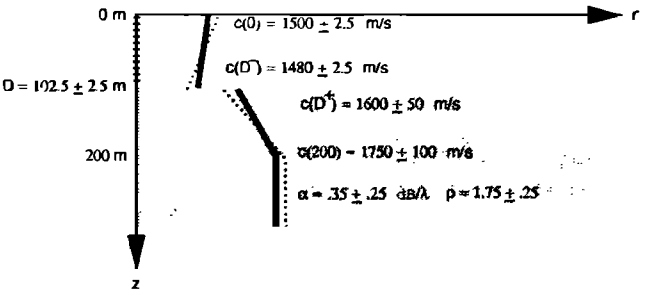

FIG. 15. The environment for the general mismatch case. The source frequency is 250 Hz and is located between 0–100 m in depth and 5–10 km in range.

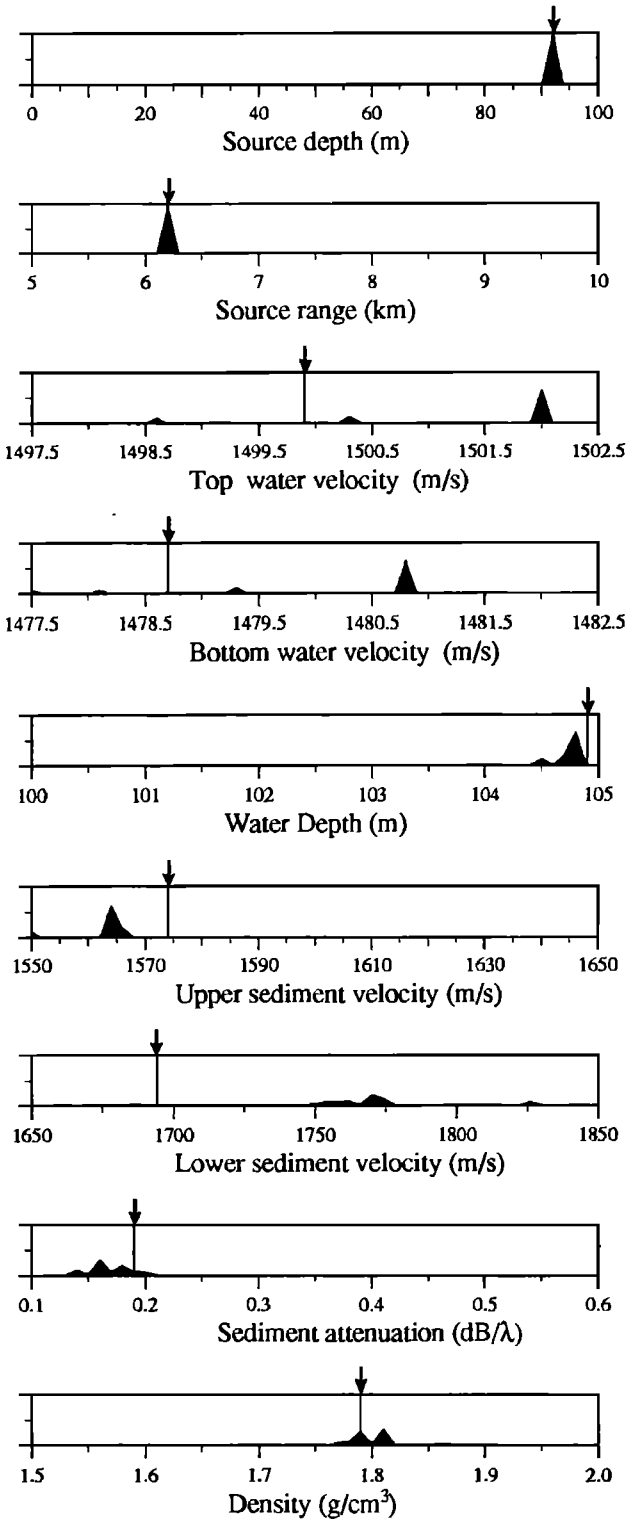

FIG. 16. Probability distribution for the general mismatch case based on the parameters in Fig. 15. The arrows indicate the correct values.

seen that the source range, source depth, and water depth are quite well determined. For the water velocities there are some ambiguities in the values. A plot of the two-dimensional marginal *a posteriori* probability distribution between the lower- and upper-water velocities reveal that these are correlated, Fig. 17, as there is a constant slope. It is easier to invert a set of uncorrelated parameters since a better fit can then be obtained by changing only one parameter as opposed to changing several parameters simultaneously. The water velocity profile is therefore better modeled by a constant shape function plus a sloping shape function; these could be thought of as the first two functions in a set of empirical orthogonal functions (EOF). The constant shape function is here taken as the velocity at the top and the slope representing the deviation from this at the bottom. The range of the two amplitudes are then 1497.5–1502.5 and 15–25 m/s.

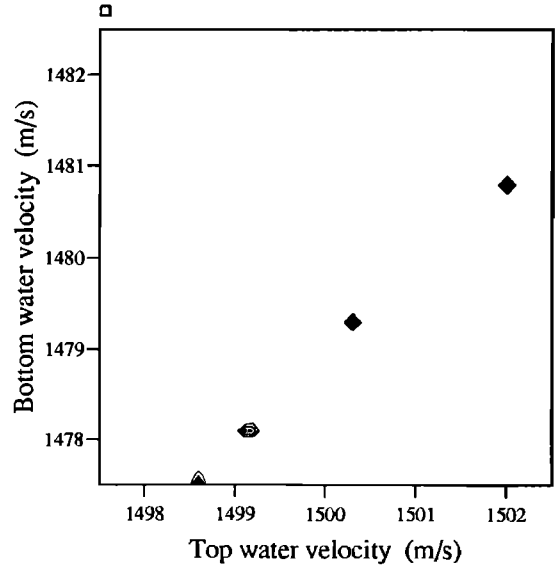

FIG. 17. 2-D marginal probability distribution between the upper- and lower-sound velocity in the water.

The result of using shape functions for the water velocity gives a high resolution of source range, source depth, water velocity slope, and water depth, Fig. 18. Since the water velocity profile now has been inverted better, the less important parameter in the sediment can also be better resolved, see Fig. 18 compared to Fig. 16. This shows that very different results can be obtained depending on how the environment is discretized.

VI. CONCLUSIONS

It has been demonstrated that the global optimization method genetic algorithms (GA) is quite robust as it requires little prior knowledge, and thus automatic inversion is not impossible. Specifically it was found that for selection of the parental distribution, a faster convergence is obtained if the energy is scaled with a temperature similar to that used in simulated annealing (SA). This temperature is conveniently selected as the value of the object function for the best individual in a population. For the replacement of a generation, only part of the population is replaced in each iteration (steady-state reproduction). This method seems faster than other implementations of GA and looks promising compared with SA.

The result of the optimization is displayed as an *a posteriori* probability distribution and is presented either as the 20 best models or as the marginal probability distribution for each parameter. Thus an indication of the impor-

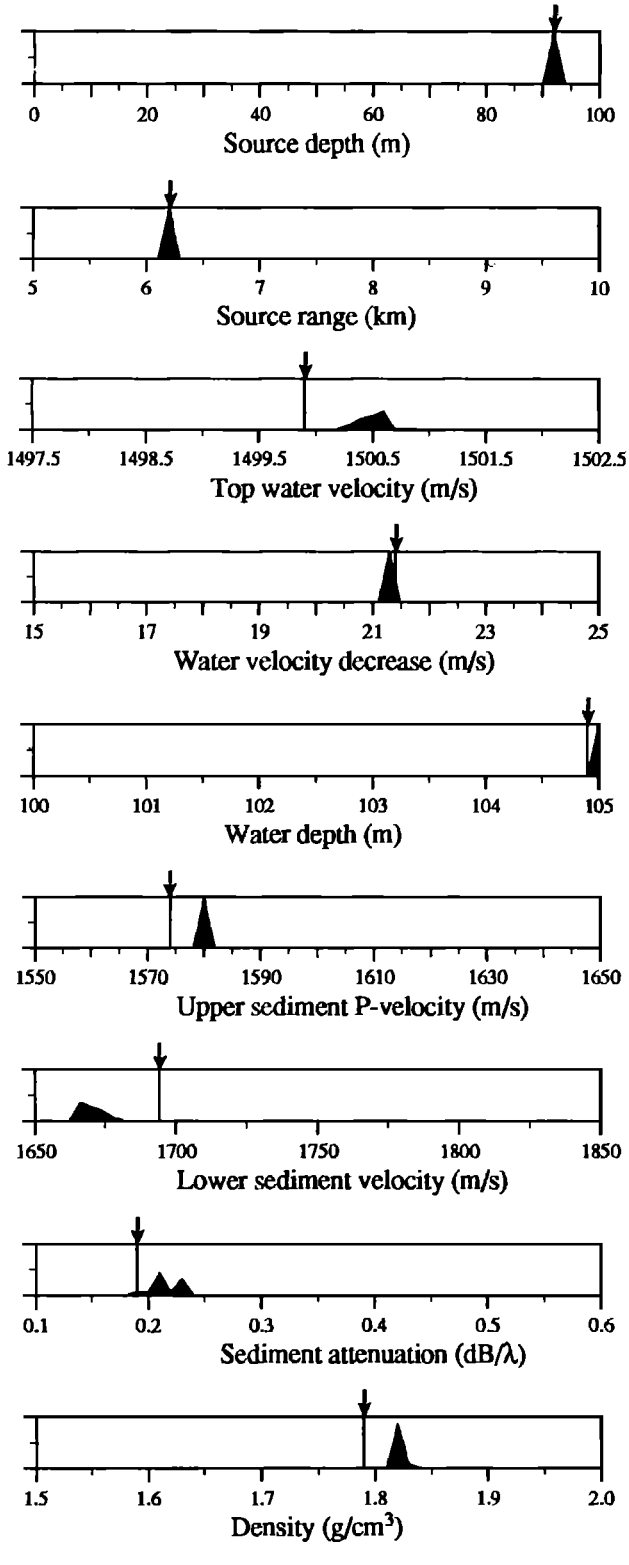

FIG. 18. Probability distribution for the general mismatch case based on the discretizion of the water velocity by shape functions. The arrows indicate the correct values.

tance and uniqueness of each parameter is obtained, and the uncertainty in the solution has been assessed.

The approach has been illustrated by several examples. These show that for noise-free synthetic data it is feasible to estimate the geoacoustic parameters and that the accuracy with which they are found is closely related to the physical importance of the individual parameters.

ACKNOWLEDGMENTS

The author wishes to thank Don Gingras, Finn Jensen, Hans Schneider, and Bjarne Stage for their invaluable comments which helped improve the quality of this paper.

[1] M. D. Collins, W. A. Kuperman, and H. Schmidt, "Nonlinear inversion for ocean-bottom properties," J. Acoust. Soc. Am. **92**, 2770–2783 (1992).

[2] L. R. Lines and S. Treitel, "A review of least squares inversion and its application to geophysical problems," Geophys. Prospect. **32**, 159–186 (1984).

[3] S. D. Rajan, J. F. Lynch, and G. V. Frisk, "Perturbative inversion methods for obtaining bottom geoacoustic parameters in shallow water," J. Acoust. Soc. Am. **82**, 998–1017 (1987).

[4] J. A. Scales, P. Docherty, and A. Gersztenkorn, "Regularisation of nonlinear inverse problems: imaging the near-surface weathering layer," Inverse Problems **6**, 115–131 (1990).

[5] A. Caiti, T. Akal, and R. D. Stoll. "Shear wave velocity in seafloor sediments by inversion of interface wave dispersion data," Rep. SR-205, (SACLANT Undersea Research Centre, Italy, 1993).

[6] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning* (Addison-Wesley, Reading, MA, 1987).

[7] L. Davis, ed., *Genetic Algorithms and Simulated Annealing* (Pithman, London, 1990).

[8] J. A. Scales, M. L. Smith, and T. L. Fisher, "Global optimization methods for highly nonlinear inverse problems," J. Comput. Phys. **103**, 258–268 (1992).

[9] M. Sambrigde and G. Drijkoningen, "Genetic algorithms in seismic waveform inversion," Geophys. J. Int. **109**, 323–343 (1992).

[10] P. L. Stoffa and M. K. Sen, "Multiparameter optimization using genetic algorithms: Inversion of plane wave seismograms," Geophysics **56**, 1794–1810 (1991).

[11] M. K. Sen and P. L. Stoffa, "Rapid sampling of model space using genetic algorithms: examples from seismic waveform inversion," Geophys. J. Int. **108**, 281–292 (1992).

[12] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of states done by fast computing machines," J. Chem. Phys. **1**, 1087–1092 (1953).

[13] S. Kirkpatrick, C. D. Gelatt, and M. Vecchi, "Optimization by simulated annealing," Science **220**, 671–680 (1983).

[14] M. D. Collins and W. A. Kuperman, "Focalization: Environmental focusing and source localization," J. Acoust. Soc. Am. **88**, 1011–1019 (1990).

[15] W. A. Kuperman, M. D. Collins, and H. Schmidt, "A fast simulated annealing algorithm for inversion of marine sediment seismo-acoustic parameters," in *Shear Waves in Marine Sediments*, edited by J. M. Hovem, M. D. Richardson, and R. D. Stoll (Kluwer, Dordrecht, 1991).

[16] S. E. Dosso, J. M. Ozard, and J. A. Fawcett, "Inversion of acoustic field data for bathymetry and bottom sound speed via simulated annealing," in *Acoustic Signal Processing for Ocean Exploration*, edited by J. M. F. Moura and I. M. G. Lourtie (Kluwer, Dordrecht, 1993).

[17] J. Nulton and P. Salamon, "Statistical mechanics of combinatorial optimization," Phys. Rev. A **37**, 1351 (1988).

[18] L. N. Frazer and A. Basu, "Inversion by statistical physics with an application to offset VSP," submitted to Geophys. J. Int. (1992).

[19] A. Tarantola, *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation* (Elsevier, Amsterdam, 1987).

[20] A. J. W. Duijndam, "Bayesian estimation in seismic inversion. Part I: Principles," Geophys. Prospect. **36**, 878–898 (1988).

[21] G. E. Backus and J. F. Gilbert, "The resolving power of gross Earth data," Geophys. J. R. Astron. Soc. **13**, 169–205 (1968).

[22] K. Mosegaard and P. D. Vestergaard, "A simulated annealing approach to seismic model optimization with sparse prior information," Geophys. Prospect. **39**, 599–611 (1991).

[23] P. D. Vestergaard and K. Mosegaard, "Inversion of post-stack seismic data using simulated annealing," Geophys. Prospect. **39**, 613–624 (1991).

[24] A. Tarantola, E. Crase, M. Jervis, Z. Koren, J. Lindgreen, K. Mosegaard, and M. Noble, "Nonlinear inversion of seismograms: State of the art," 60th Annual Mtg. Soc. Expl. Geophys., Expanded Abstr., 1193–1198 (1990).

[25] H. Schmidt, "OASES version 1.6. application and upgrade notes,"

Tech. Rep., Department of Ocean Engineering, Massachusetts Institute of Technology (1993).
[26]F. B. Jensen and M. C. Ferla, "SNAP: The saclantcen normal-mode acoustic propagation model," SM-121, SACLANT Undersea Research Centre, La Spezia, Italy (1979).
[27]H. Schmidt, "SAFARI: Seismo-acoustic fast field algorithm for range independent environments. User's guide," SR-113, SACLANT Undersea Research Centre, La Spezia, Italy (1987).
[28]L. Amundsen and B. Ursin, "Frequency-wavenumber inversion of acoustic data," Geophysics **56**, 1027–1039 (1991).
[29]H. Szu and R. Hartley, "Fast simulated annealing," Phys. Lett. **122**, 157–162 (1987).
[30]M. Porter and A. Tolstoy, *Workshop on Acoustic Model in Signal Processing*, 24–26 May 1993 (Naval Research Laboratory, Washington, DC, 1993).