Geophysical Prospecting **39,** 613–624, 1991

INVERSION OF POST-STACK SEISMIC DATA USING SIMULATED ANNEALING[1]

PETER D. VESTERGAARD[2] and KLAUS MOSEGAARD[3]

ABSTRACT

VESTERGAARD, P.D. and MOSEGAARD, K. 1991. Inversion of post-stack seismic data using simulated annealing. *Geophysical Prospecting* **39,** 613–624.

Model-based inversion of seismic reflection data is a global optimization problem when prior information is sparse. We investigate the use of an efficient, global, stochastic optimization method, that of simulated annealing, for determining the two-way traveltimes and the reflection coefficients.

We exploit the advantage of an ensemble approach to the inversion of full-scale target zones on 2D seismic sections.

In our ensemble approach, several copies of the model-algorithm system are run in parallel. In this way, estimation of true ensemble statistics for the process is made possible, and improved annealing schedules can be produced.

It is shown that the method can produce reliable results efficiently in the 2D case, even when prior information is sparse.

INTRODUCTION

Automatic inversion schemes for the reconstruction of subsurface structures from seismic reflection data are used more and more frequently in the oil industry for detailed studies of oil and gasfields during the development phases. In cases of good well control such methods have often produced satisfactory predictions concerning the lithological columns seen in wells drilled at later stages.

However, it is frequently observed that surprisingly large errors in the prediction of reflector locations and acoustic impedance values occur in cases where the well

[1] Received November 1989, revision accepted December 1990, last material received January 1991.

[2] Imperial College, Department of Geology, Royal School of Mines, Prince Consort Road, London SW7 2BP, U.K. Formerly Ødegaard and Danneskiold-Samsøe ApS, Copenhagen.

[3] Geophysical Institute, University of Copenhagen, Haraldsgade 6, 2200 Copenhagen N, Denmark.

control is sparse or in cases where the correlation between seismic events and nearby wells is made difficult by fault zones, thinning of beds, local disappearance of impedance contrasts or by the presence of noise. Under such circumstances the prior information about the subsurface structure in the zone of interest is very limited and it is not possible to put strong constraints on the solution to the inverse problem.

An analysis of the principles behind the presently available inverse methods reveals that these techniques all belong to a category called 'local optimization methods'. A characteristic property of these algorithms is that they systematically adjust the subsurface model in such a way that the misfit function (measuring the misfit between synthetic data and actual data) decreases monotically. This property would have been desirable if the misfit function possessed only one minimum.

However, since seismic data is of a highly oscillating nature, the misfit function generally has a very large number of minima (Fig. 1). Moreover, secondary minima representing low values of the misfit function often correspond to subsurface models that are quite different from the true model. It is therefore imperative that a local model optimization method uses a starting model that is 'connected' to the optimal solution by a path along which the misfit function decreases monotonically. In practice, the only way to ensure that this is the case is to use data from a nearby well and provide a starting model that is very close to the optimal model.

These considerations lead to the conclusion that local model optimization algorithms are likely to fail in the previously mentioned cases of limited well control. In such cases, the probability that a starting model is sufficiently close to the optimal model is small and the corresponding probability that a local optimization method will be attracted by an irrelevant minimum for the misfit function is high.

The solution to the above-described problems is to employ a 'global' optimization method. Global optimization methods are capable of searching for the optimal subsurface model with only a small risk of being trapped by irrelevant minima for the misfit function. Global optimization methods are typically statistical techniques.

A prototype development and implementation of a global, full-scale, seismic model optimization program for inversion of seismic profiles is presented. This program is based on the global optimization method 'simulated annealing' and is aimed at inversion of selected parts of migrated, seismic profiles with the purpose of producing geological cross-sections showing the acoustic impedance and location of layer interfaces in the subsurface.

Classical simulated annealing is known to be a rather inefficient Monte Carlo technique, only applicable in cases where a very large number of iterations can be performed within the available computer resources. However, in the present implementation we employ a recent version of simulated annealing (Nulton and Salamon 1988; Andresen *et al.* 1988) in which we are able to extract important statistical information about the structure of the optimization problem during the computations. As a result, we have been able to speed up the algorithm significantly. The efficiency is improved by a factor of 7 to 100, and the fact that highly accurate results can be produced in a limited time has made the algorithm interesting from a practical point of view.

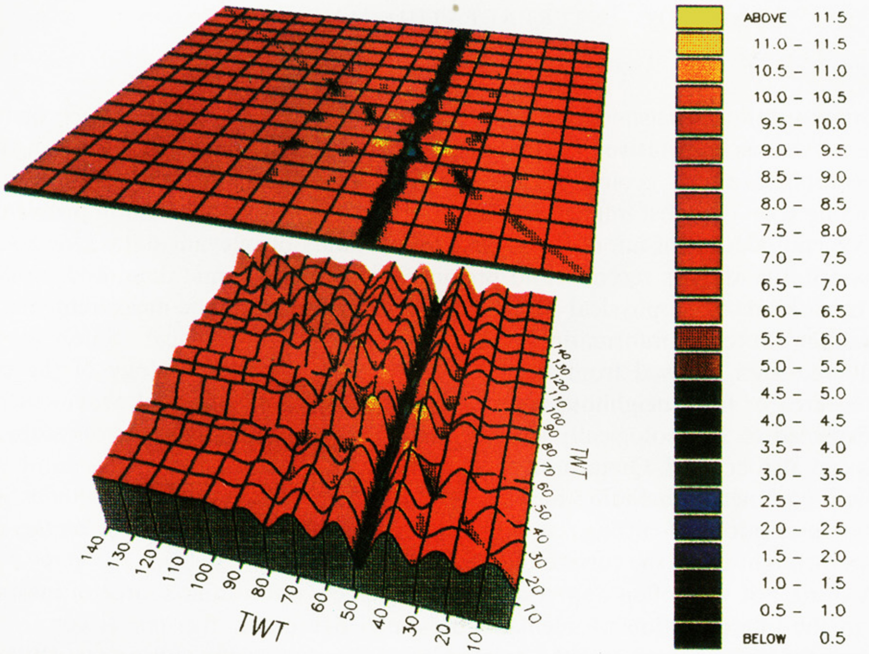

FIG. 1. Misfit function surface for a simple, single trace model optimization problem. The misfit surface is shown for a 2D cut through the parameter space. The independent parameters in the considered plane are two-way traveltimes of two reflectors.

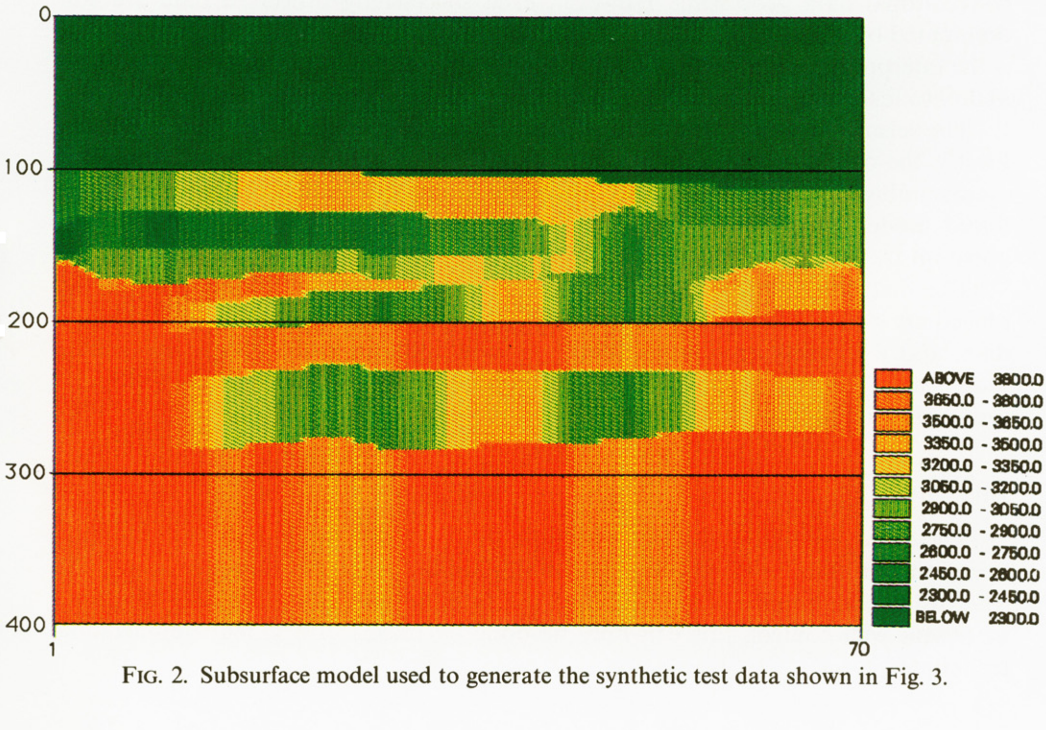

FIG. 2. Subsurface model used to generate the synthetic test data shown in Fig. 3.

Seismic Interpretation and Inversion

Interpretation

Interpretation of seismic data is an extremely complex process in which quantitative as well as qualitative information from several sources is compiled, weighted and combined into a geological model, which is displayed in such a way that it throws light on the most important aspects of the considered exploration problem.

The compilation of information includes all kinds of relevant data. One source of quantitative data is recently and previously recorded seismic data, and possibly also other kinds of geophysical data such as gravity and magnetic measurements.

Another source of information is qualitative, *a priori* geological models for the considered area, derived from the known, or partly known, geology of the considered area or from neighbouring areas, or from remote geological provinces that are expected to be geologically similar. Well data from the area or from adjacent areas can also be used. Quantitative well data include measurements of a number of physical parameters including mechanical ones. Qualitative well data includes geological descriptions of cuttings, small pieces of subsurface rock cut loose by the drill bit and brought up to the surface by the mud flow.

A third, not very often appreciated but extremely important, source of information in the interpretation problem is our knowledge of the theoretical connection between the seismic data and the mechanical properties of the subsurface. Without this information, an interpretation of the data would be impossible. Part of our theoretical knowledge of wave generation and propagation is applied during the conventional data processing. However, even successfully processed data are still dominated by at least one residual source or wave propagation effect: the wavelet. It is the interpreter's knowledge of the effect of this wavelet that initially determines his ability to resolve fine details in the subsurface.

The seismic wavelet gives rise to two important problems in data interpretation. Firstly, the oscillatory appearance of the wavelet makes traveltime determination of events ambiguous, and therefore serious reflector dislocations may occur in the produced model. Secondly, interference between events results in distortion of their apparent traveltimes and amplitudes.

The first dislocation effect can be removed in the vicinity of wells, where the reflectivity derived from the well data can be correlated directly with the seismic data, and a one-to-one correspondence between reflectors and reflections can be established. The second interference effect typically remains unsolved by the interpreter, due to the qualitative nature of the interpretation process.

Inversion

The above-mentioned interference effect can be solved by traditional, local optimization techniques such as steepest descent search, conjugate gradient search, etc. The dislocation problem, however, must be solved in advance. In case of sparse or absent well control, this can only be done by means of a global optimization technique.

In order to remove or avoid seismic reflector dislocations, it is natural to simulate a technique which is used for removal of crystallographic dislocations, namely chemical annealing. In this technique, the crystalline material is melted and subsequently cooled slowly through its melting point, allowing large, highly ordered crystals to grow. If annealing is performed sufficiently slowly, the crystalline material eventually consists of one crystal without dislocations. In this state the crystalline material has the lowest possible internal energy.

As pointed out by Mosegaard and Vestergaard (1991), the analogy between the seismic inversion process and chemical annealing can be reinforced in the following way: the subsurface models can be identified with the atomic configurations of the crystalline material. The misfit function used in the seismic model optimization as a measure of the difference between synthetic data, computed from a trial model, and the observed data, can be identified with the energy of the crystal. Furthermore, random changes in the sub-surface model during a stochastic search can be performed in a way that is analogous to the random movements of atoms in the melted material or in the crystal lattice. By this analogy, a gradual decrease in the average size of the 'thermal movements' of the model from large values down to zero is likely to result in a settling into a subsurface model possessing a low value of the misfit function. Such a model is exactly what we wish to find.

The technical details of this algorithm, which is known as 'simulated annealing' (Kirkpatrick, Gelatt and Vecchi 1983) are somewhat more involved than the above exposition suggests. The interested reader may consult the review by Mosegaard and Vestergaard (1991).

We have used a recently developed simulated annealing method (Nulton and Salamon 1988; Andresen *et al.* 1988). This method needs statistical information about the system to be optimized, in order to extract optimal annealing schedules. Reliable statistical information would be difficult to obtain from annealing with a single copy of the model-algorithm system, since the convergence property of the simulated annealing algorithm often results in sampling of a limited part of the model space, typically concentrated around the misfit minimum found by the algorithm. In order to reduce this problem, we run a number of copies of the annealing at the same time. These copies of the model-algorithm system share the same temperature schedule, but they use different random number sequences, and their initial states are distributed according the prior knowledge. Hence, their time evolution is different, and they sample widely different parts of the model space.

Calculation of the Misfit

In the present model optimization problem, the misfit function S is the error trace energy

$$S(\mathbf{r}, \boldsymbol{\tau}) = \sum_{n=0}^{N} (s(\mathbf{r}, \boldsymbol{\tau}, n) - d(n))^2, \tag{2}$$

where $d(n)$ is the nth data sample, $s(\mathbf{r}, \boldsymbol{\tau}, n)$ is the nth sample of the modelled trace, $\mathbf{r}$ and $\boldsymbol{\tau}$ are vectors of reflection coefficients and two-way traveltimes, respectively, and $N+1$ is the number of data samples. $s(\mathbf{r}, \boldsymbol{\tau}, n)$ is obtained from the convolutional model of the seismic trace:

$$s(\mathbf{r}, \boldsymbol{\tau}, n) = \sum_{k=1}^{K} r_k w_k(n - \tau_k), \tag{3}$$

where τ_k is the two-way traveltime of the kth reflector, w_k is the wavelet corresponding to the kth reflection, and K is the number of reflectors considered.

However, for the considered type of global model optimation problems, the time required to obtain a near-optimal solution by simulated annealing grows rapidly with the number of parameters to be determined. Therefore it is desirable to reduce the number of parameters to be optimized by means of simulated annealing.

In the present problem, it is observed that the assumed dependence of the modelled trace $s(\mathbf{r}, \boldsymbol{\tau}, n)$ on the reflection coefficients r_k is linear. Hence, it is possible to restrict the simulated annealing optimization to the two-way traveltime parameters τ_k only, and perform a simple, linear optimization of the reflection coefficients as part of the misfit calculations. We therefore redefine the misfit function as:

$$E(\boldsymbol{\tau}) = \min_{\forall r} S(\mathbf{r}, \boldsymbol{\tau}). \tag{4}$$

In the following, we assume that the wavelets $w_k(n)$ are non-zero only in the interval $0 \leq n \leq N_w$. For transient wavelets, this situation can always be obtained by introducing an appropriate time shift. We also assume that none of the individual reflection events are clipped at the end of the modelled trace. In other words, $0 \leq \tau_k \leq N - N_w$ for all k.

The partial derivatives of S with respect to the reflection coefficients r_k are given by

$$\frac{\partial S}{\partial r_k} = \sum_{n=0}^{N} 2(s(\mathbf{r}, \boldsymbol{\tau}, n) - d(n))w_k(n - \tau_k) = 2 \sum_{n=0}^{N} e(\mathbf{r}, \boldsymbol{\tau}, n)w_k(n - \tau_k), \tag{5}$$

where $e(\mathbf{r}, \boldsymbol{\tau}, n) = s(\mathbf{r}, \boldsymbol{\tau}, n) - d(n)$ is the error trace. The optimal values of r_k must satisfy

$$\frac{\partial S}{\partial r_k} = 0 \tag{6}$$

for all k. This leads to the following system of linear equations:

$$\sum_{n=0}^{N} \left[\left[\sum_{i=1}^{K} (r_i w_i(n - \tau_i)) - d(n) \right] w_k(n - \tau_k) \right] = 0 \tag{7}$$

for $k = 1, \ldots, K$. This system of equations is equivalent to the system

$$\sum_{i=1}^{K} r_i \left[\sum_{n=0}^{N} w_i(n - \tau_i) w_k(n - \tau_k) \right] = \sum_{n=0}^{N} d(n) w_k(n - \tau_k) \tag{8}$$

or

$$\sum_{i=1}^{K} r_i \left[\sum_{n=0}^{N_w} w_i(n + (\tau_k - \tau_i)) w_k(n) \right] = \sum_{n=0}^{N} d(n) w_k(n - \tau_k). \tag{9}$$

If we assume that the wavelet is the same, say $w(n)$, for all reflections, the equations reduce to

$$\sum_{i=1}^{K} r_i R_{\mathrm{ww}}(\tau_k - \tau_i) = R_{\mathrm{dw}}(-\tau_k), \tag{10}$$

where

$$R_{\mathrm{ww}}(\tau) = \sum_{n=0}^{N_w} w(n)w(n+\tau) \tag{11}$$

is the autocorrelation of the wavelet, and

$$R_{\mathrm{dw}}(\tau) = \sum_{n=0}^{N} d(n)w(n+\tau) \tag{12}$$

is the cross-correlation between the data and the wavelet. From (10), optimal reflection coefficients r_i can be found, if the two-way traveltimes τ_k are given. Equations of the form (10) are known in filter theory as the 'normal equations' and they can be solved very efficiently by the Wiener–Levinson algorithm. In each iteration, a new set of two-way traveltimes is selected to become candidate destinations for the next move in the two-way traveltime parameter space (having half the dimensions of the combined traveltime–reflection coefficient space). The misfit E is now calculated after having minimized S with respect to the reflection coefficients.

A necessary condition for the system of equations (10) to have a unique solution is that all the K two-way times τ_k are different (no reflectors coincide). This condition must be satisfied by each perturbation applied to the subsurface models. Moreover, reflectors are not allowed to be too close, since numerical instabillity will occur when the equations (10) are near singular. Geological situations such as layer pinchouts must therefore be treated separately.

Another problem to be mentioned is that a straight linear optimization of the reflection coefficients for given traveltimes may yield reflection coefficients that violate the constraints imposed by the prior information. This problem can be avoided by performing a constrained, linear optimization.

A Synthetic Example

The synthetic data test is based on the synthetic seismic response for a model of thin layers (Fig. 2). The model consists of a layered sequence between a low velocity half-space above and a high velocity half-space below. The acoustic impedance of the layers varies laterally, and two layers pinchout from left to right. The data set (Fig. 3) is generated by convolving the model reflectivity with a 40 ms long, band-pass filtered wavelet. The target zone is a 400 ms time window over 71 common depth points, which is a realistic size for many practical applications of target zone oriented inversion.

In this numerical example, the prior knowledge is sparse. The two-way traveltimes for the layer interfaces are limited to intervals that are slightly longer than one

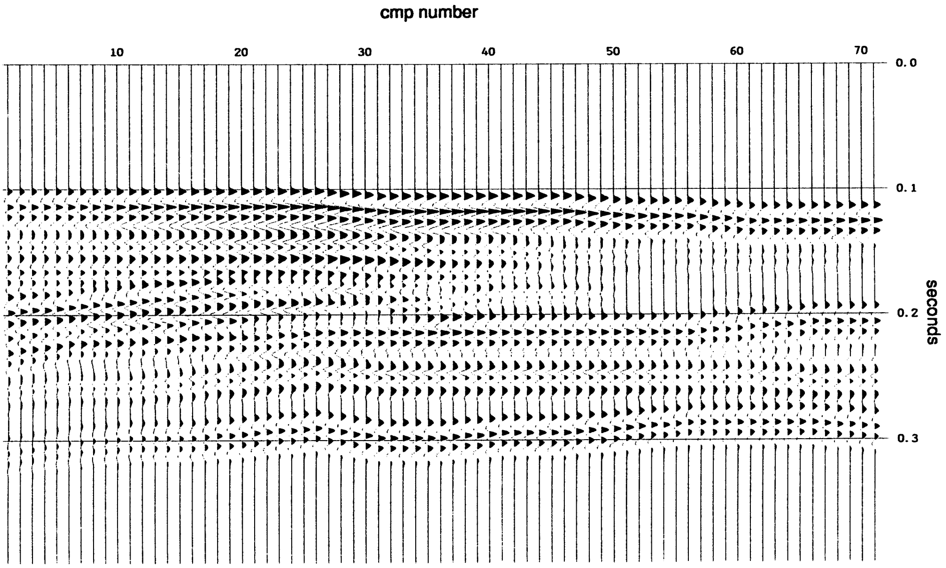

FIG. 3. Synthetic data set used in the test example.

wavelet-length, and the amplitudes of the reflections are constrained to be between -0.2 and 0.2. A weak, lateral smoothness constraint is applied to the reflectors during the optimization. Even when the wavelet is assumed to be known, these wide limits on the parameters impose a very large number of local minima on the misfit function for the model optimization problem.

Five independent copies of the model-algorithm system were allowed to perform 5000 iterations each. Two of these copies settled into a near-optimal solution. In order to illustrate how the solution to the model optimization problem was formed during the most successful of these annealing processes, a number of 'snapshots' of intermediate subsurface models are shown for increasing iteration numbers, corresponding to a temperature variation from a very high value (effectively infinity) down to zero. The main point to notice is how the models in the ensemble gradually change from being typical samples from the *a priori* distribution (model parameters uniformly distributed over the parameter intervals), to models that reflect the information contained in the seismic data, under the limitations imposed by the prior knowledge.

The first subsurface model in Fig. 4 is the result of a substantial number of iterations at a very high temperature. In this model, the seismic dislocations are very large, and the resulting model for the acoustic impedance is far from the optimal model. The illustrated model gives an impression of the weak model constraints used in this annealing run.

In the first part of the annealing, an initial ordering of the models takes place. Figure 5 shows a model after the first part of the annealing has taken place. It can be seen that a layered structure is growing in the upper part of the target zone, corresponding to the first ordering of a crystal structure in the physical analogy.

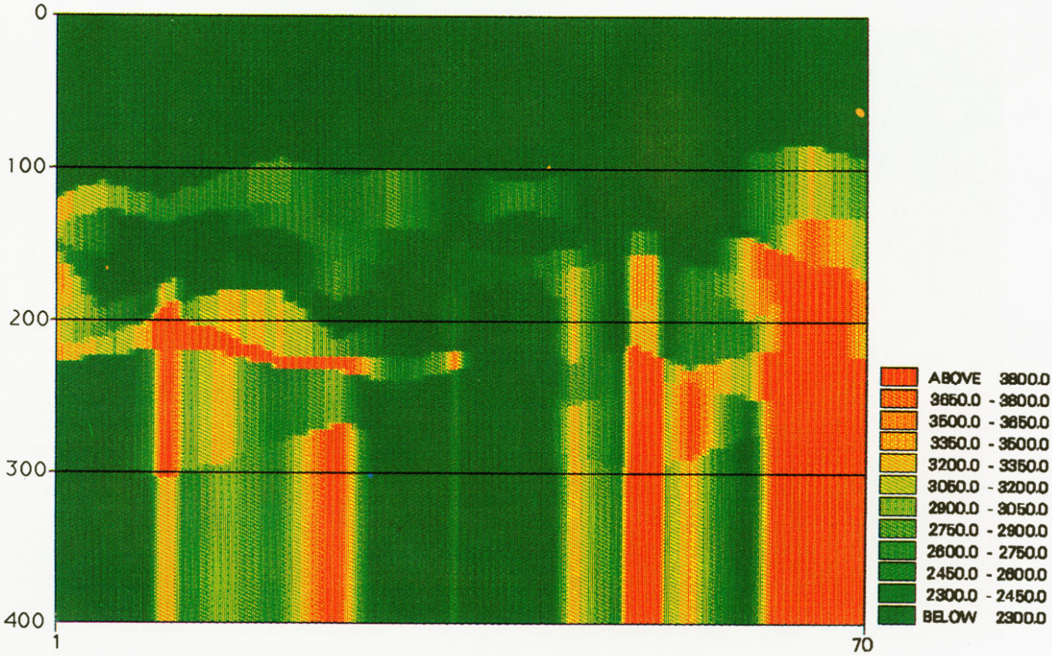

FIG. 4. A typical subsurface model obtained at infinite temperature. The plot shows the acoustic impedance as a function of two-way traveltimes.

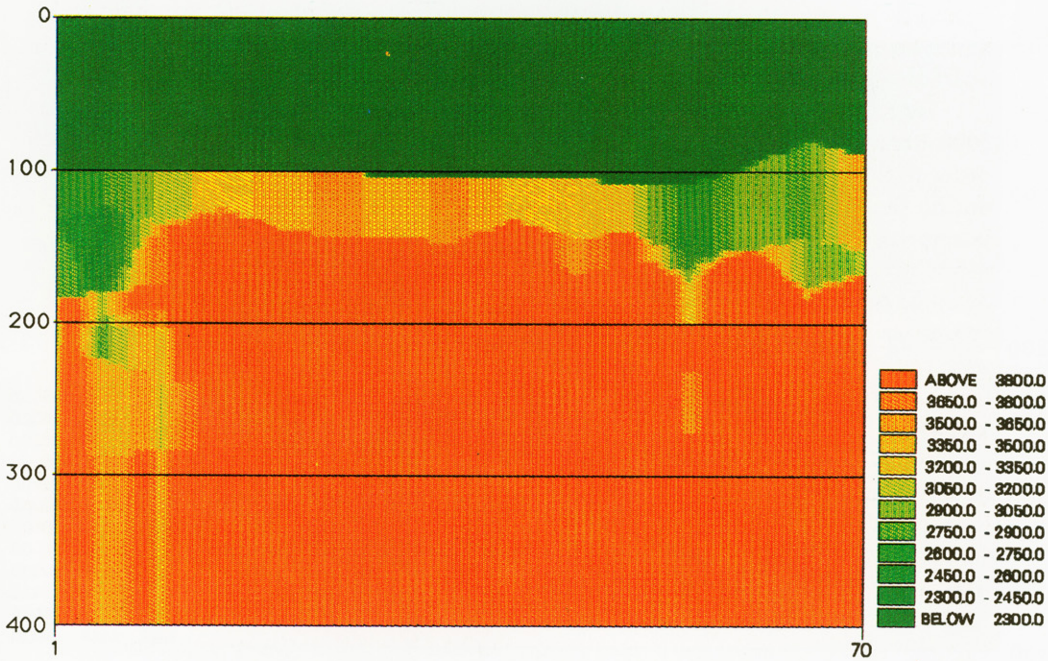

FIG. 5. A subsurface model from the ensemble, picked after the first part of the simulated annealing has taken place.

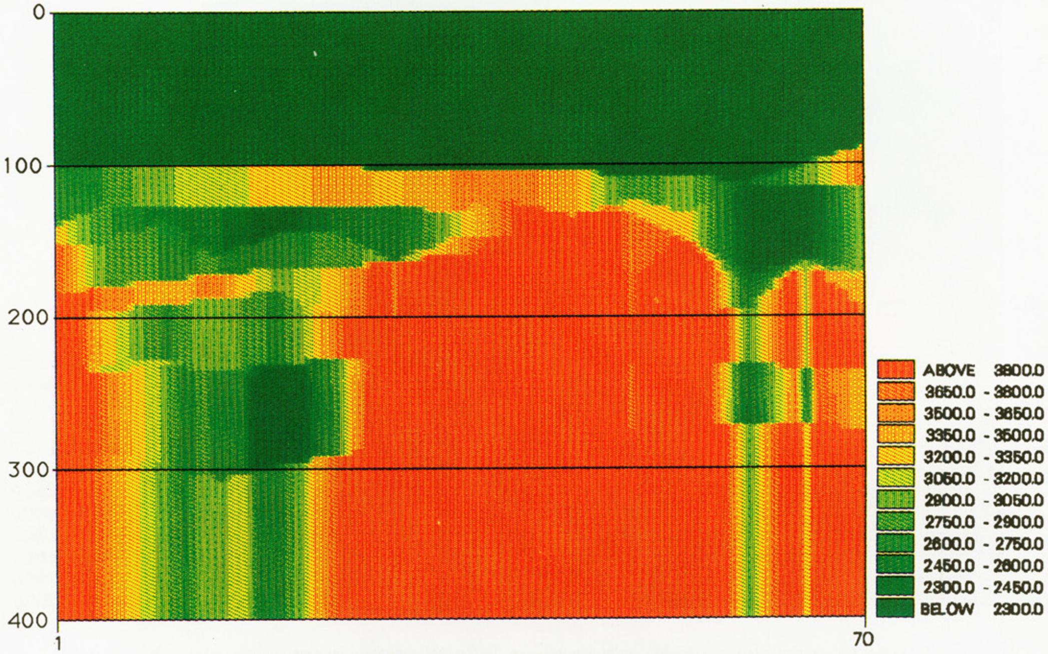

FIG. 6. An intermediate temperature subsurface model.

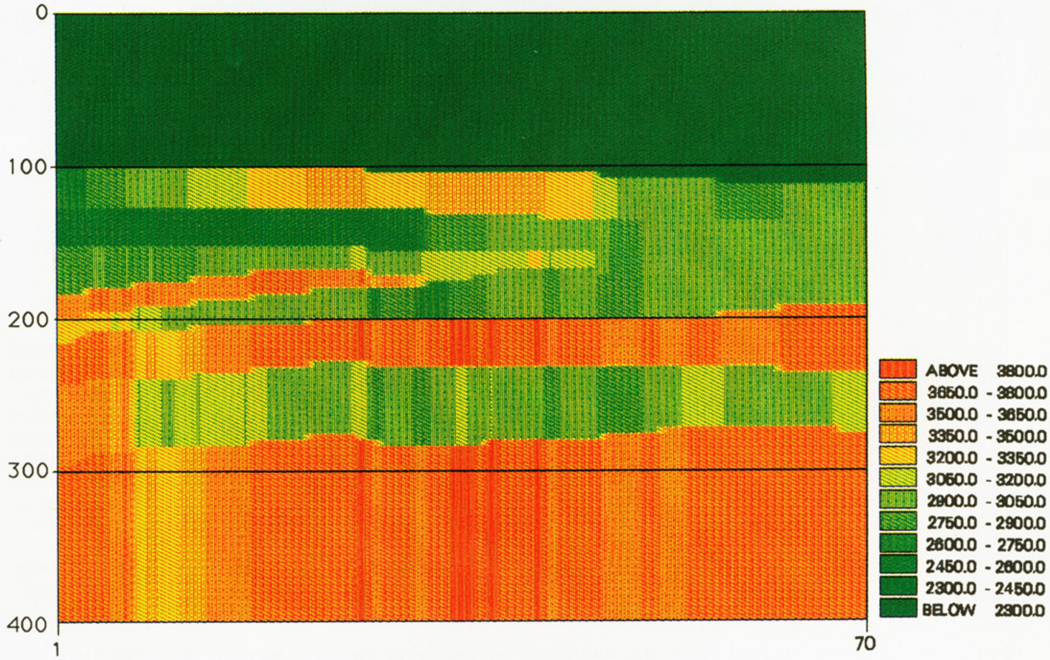

FIG. 7. The best subsurface model found close to zero temperature.

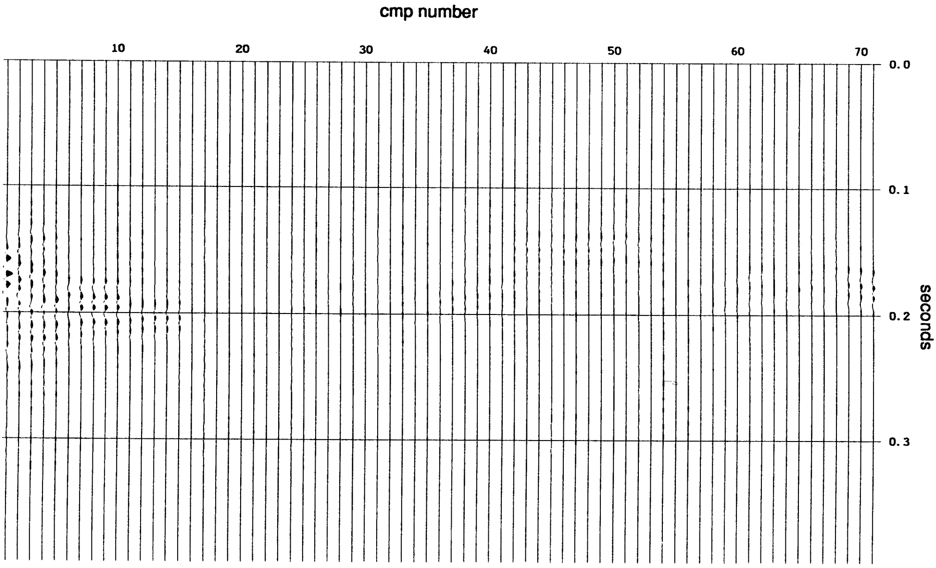

FIG. 8. Error traces for the model shown in Fig. 7.

After a further lowering of the temperature, models like that shown in Fig. 6 can be found in the ensemble. An increasing ordering of the model into a layered structure is seen. This illustrates how the influence of the data on the models increases as the annealing progresses. The major part of the target zone is not yet resolved at this intermediate temperature.

When the annealing temperature approaches zero, the models look like that shown in Fig. 7. These models are near optimal in the misfit sense, that is, the total error trace energy (Fig. 8) is small, compared to the total trace energy of the data. At a temperature close to zero, the influence of the data on the model is very strong, yielding a highly ordered subsurface model. The differences between the models in the ensemble reflect the limited resolution in the data, the non-uniqueness of the inverse problem, and possible imperfect convergence of the algorithm. It is seen that the actual target in this model, namely the pinch-out, located approx. 175 ms below the top of the target zone, developing from CDP 10 to CDP 40, is resolved at this point. However, due to end-effects, the error traces build up to the left of this zone.

If the annealing is terminated by a number of iterations at zero temperature (corresponding to a local optimization starting from the best model obtained from the annealing), the true model is reached, since the data used in this example is noise-free.

CONCLUSIONS

The process of seismic interpretation is made difficult by two main types of distortion, both caused by the seismic wavelet: the oscillatory appearance of the individual reflection events and interference between different events. The former effect results in a serious ambiguity in event identification, that is, in establishing a one-to-one correspondence between geological layer interfaces and features observed in the seismic data. The latter effect is responsible for minor errors in the estimated two-way traveltimes, and errors in reflection strengths observed in the seismic data.

Post-stack, sparse spike inversion methods provide quantitative methods for determining two-way times and reflection coefficients from carefully processed seismic data. If the event identification problem can be solved by comparing the seismic data with synthetic seismograms, calculated from well data, the remaining problem of removing interference effects can be solved by means of traditional, local optimization methods. However, if sufficient well data are not available, the event identification problem can only be solved quantitatively by means of a global optimization technique.

Global optimization methods are typically stochastic, and at present the most efficient is simulated annealing. In the present work, a recently developed, improved version of simulated annealing has been shown to produce near-optimal solutions to a seismic model optimization problem of a realistic size and complexity. An ensemble, consisting of several copies of the model-algorithm system, sharing the same temperature schedule but using different random number sequences, is used to collect statistical information about the model optimization problem, and efficient annealing temperature schedules are produced.

In order to reduce the computational workload considerably, the optimization of the reflection coefficients, which turns out to be a simple linear optimization, can be done separately as part of the misfit calculations. This reduces the dimensionality of the parameter space, in which the stochastic optimization is performed, to half the original dimensions. Consequently, the resulting number of accessible model configurations for the stochastic search decreases drastically, and the average time taken by the algorithm before a near-optimal model is located, is greatly reduced.

Acknowledgements

The authors are indebted to Jacob Mørch Pedersen, Peter Salamon, Bjarne Andresen and the simulated annealing group at the University of Copenhagen, whose comments and suggestions were invaluable. This work was sponsored by The Danish Ministry of Energy under contract No. 1313/87-12. Klaus Mosegaard was partly supported by the Danish Natural Science Research Council.

References

Andresen, B., Hoffman, K.H., Mosegaard, K., Nulton, J.D., Pedersen, J.M. and Salamon, P. 1988. On lumped models for thermodynamic properties of simulated annealing problems. *Journal de Physique* **49**, 1485–1492.

Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P. 1983. Optimization by simulated annealing. *Science* **220**, 671–680.

Mosegaard, K. and Vestergaard, P.D. 1991. A simulated annealing approach to seismic model optimization with sparse prior information. *Geophysical Prospecting* **39**, 599–611.

Nulton, J.D. and Salamon, P. 1988. Statistical mechanics of combinatorial optimization. *Physical Review* A **37**, 1351–1356.