Nonlinear inversion, simulated annealing, and residual statics estimation*

Daniel H. Rothman

ABSTRACT

Nonlinear inverse problems are usually solved with linearized techniques that depend strongly on the accuracy of initial estimates of the model parameters. With linearization, objective functions can be minimized efficiently, but the risk of local, rather than global, optimization can be severe.

This paper addresses the problem confronted in nonlinear inversion when no good initial guess of the model parameters can be made. The fully nonlinear approach presented here is rooted in statistical mechanics. Although a large nonlinear problem might appear computationally intractable without linearization, reformulation of the same problem into smaller, interdependent parts can lead to tractable computation, while preserving nonlinearities.

I formulate inversion as a problem of Bayesian estimation, in which the prior probability distribution is the Gibbs distribution of statistical mechanics. Solutions are obtained by maximizing the posterior probability of the model parameters. Optimization is performed with a Monte Carlo technique that was originally introduced to simulate the statistical physics of systems in equilibrium. The technique is applied to residual statics estimation when statics are unusually large and data are noise-contaminated. Poorly picked correlations ("cycle-skips" or "leg-jumps") appear as local minima of the objective function, but global optimization is successfully performed. Further applications to deconvolution and velocity estimation are proposed.

*This paper is a revision of my SEP-38 paper, "Nonlinear inversion by stochastic relaxation with applications to residual statics estimation."

INTRODUCTION

Many of the problems encountered in the analysis of geophysical data are essentially problems of data inversion: data are usually collected on the Earth's surface, and we try to infer a physical parameterization of the Earth's interior, given these surface observations. We do this by first constructing a mathematical model in which the physical parameters are unknowns. Inverse problems are then usually solved by minimizing a function (or functional) defined over the space of model parameters. Questions of uniqueness aside, these optimization problems can be broadly classified into two types: those that contain only one minimum (by the usual definitions from calculus) and those that contain more than one. The case with one minimum is called linear inversion because the solution may be obtained by (approximately) solving a set of linear equations. When there are many minima no such set of linear equations exists unless important additional assumptions are made. In the absence of these assumptions we must perform nonlinear inversion - ideally, the location of the global minimum in the presence of many local minima.

Conventional approaches to nonlinear inverse problems rely on obtaining good initial estimates of the model parameters so that the remaining perturbations of the parameters satisfy linear relations (Parker, 1977; Aki and Richards, 1980; Lines and Treitel, 1984). Global optimization without a good initial guess appears computationally intractable - the parameter space is simply too large for every possible choice of parameters to be evaluated.

This paper addresses the problem confronted in nonlinear inversion when no good initial guess of the model parameters can be made. To make the problem computationally tractable, a particular statistical representation of the model parameters is introduced. This statistical representation is adapted from stochastic models used in statistical physics to describe the interactions within systems containing many degrees of freedom. This model is appropriate for large problems that can be naturally subdivided into smaller, interdependent subproblems.

The use of a statistical model is formally expressed by defining a joint prior probability distribution for the model parameters and by casting the problem in the framework of Bayesian estimation. The resulting posterior probability distribution yields the most general form of the solution; locating the maximum of the posterior probability distribution determines the most probable set of physical parameters, given the observed data and the prior probability distribution.

For nonlinear problems the posterior probability distribution still contains many local extrema. Global optimization is possible, however, by exploiting the form of the statistical model (the Bayesian prior). Reasoning by analogy with statistical physics, I employ the method of *simulated annealing* (Kirkpatrick et al., 1983; Geman and Geman, 1984). Simulated annealing is a Monte Carlo optimization technique that mimics the physical process by which a crystal is grown from a melt. Crystallization is related to optimization, in that nonlinear inversion can be characterized as a transformation from disorder (ignorance) to order (the solution).

The need for accurate nonlinear inversion in reflection seismology is demonstrated most graphically in residual statics estimation. Broadly speaking, residual statics estimation is the estimation of near-surface velocity anomalies. The problem is notoriously difficult when the near-surface anomalies are severe, because recorded traces exhibit large static (constant) timing delays that severely distort the image of subsurface geology. Automatic estimation and removal of these timing delays depends on an algorithm's ability to accurately identify spatial correlations within the constraints imposed by the design of a seismic experiment (i.e., the solution must be surface-consistent). I present the estimation of these source- and receiver-consistent timing delays as a nonlinear inverse problem. Poorly picked correlations ("cycle-skips" or "leg-jumps") appear as local minima, but global optimization by simulated annealing can be performed without an initial estimate of the actual timing delays.

The paper begins with a review of the problems inherent in nonlinear inversion. It is emphasized that prior information plays an integral role in achieving any nonlinear solution; here, I advocate using a statistical mechanical model. The relevance of statistical mechanics to this problem is supported with a discussion of spatial Markov processes. Following Geman and Geman (1984), I incorporate these ideas into a general Bayesian framework that leads to an expression for the posterior probability of the model parameters. Solutions are then obtained by maximizing the posterior probability; this is performed by a Monte Carlo optimization procedure, which I review in detail. I then reformulate residual statics analysis as a nonlinear, rather than linear, inverse problem. A numerical example of residual statics estimation illustrates the features of a practical implementation of simulated annealing. Finally, I discuss directions for future research and indicate several further applications.

THE NONLINEAR INVERSE PROBLEM

Consider a physical system (for example, the Earth) that is characterized by a finite set of unknown *model parameters* $\mathbf{X} = \{X_1, \ldots, X_s\}$ in the s-dimensional *parameter space.* An experiment performed in this physical system produces a finite set of *data* $\mathbf{D} = \{D_1, \ldots, D_r\}$ in the r-dimensional *data space.* $\mathbf{X}$ and $\mathbf{D}$ are (random) variables that assume specific values $\mathbf{x} = \{x_1, \ldots, x_s\}$ or $\mathbf{d} = \{d_1, \ldots, d_r\}$.

Let a set of r (nonlinear) functions G_i be denoted by $\mathbf{G}$. $\mathbf{G}$ is a function of the model parameters $\mathbf{x}$ and describes the theoretical outcome of the experiment. The observed data $\mathbf{d}$ is contaminated by additive noise and is represented as

$$\mathbf{d} = \mathbf{G}(\mathbf{x}) + \mathbf{n} \ , \tag{1}$$

where $\mathbf{n} = \{n_1, \ldots, n_r\}$ is a realization of the random noise $\mathbf{N}$, which is assumed to be independent, identically distributed, and independent of $\mathbf{X}$. It is assumed that only discrete values are allowed when $\mathbf{X}$, $\mathbf{D}$, or $\mathbf{N}$ is the argument of a probability distribution. Extension to continuous random variables is straightforward.

Solving equation (1) for an $\mathbf{x}$ that represents the entire underlying set of model parameters is the most ambitious of geophysical inverse problems. In a more realistic approach, $\mathbf{x}$ represents a vector of parameters for a smaller problem in which many physical quantities are previously defined. Later, $\mathbf{x}$ will represent residual static timing delays, and it will be assumed that stacking velocities are known.

Inversion of (1) is generally performed by optimization. We solve for

$$\mathbf{x} = \min_{\mathbf{x}} f \, [\, \mathbf{d}, \mathbf{G}(\mathbf{x}) \,] \ . \tag{2}$$

In least squares formulations, f is typically the squared difference between $\mathbf{G}(\mathbf{x})$ and $\mathbf{d}$. Using $\mathbf{G}(\mathbf{x})$ implies that forward modeling is done; alternatively we can minimize a function that depends on $\mathbf{G}^{-1}$, the (approximate) inverse of $\mathbf{G}$. In this case

$$\mathbf{x} = \min_{\mathbf{x}} f \, [\, \mathbf{G}^{-1}(\mathbf{d}; \mathbf{x}) \,] \ . \tag{3}$$

$\mathbf{G}^{-1}$ operates on $\mathbf{d}$, but the precise form of $\mathbf{G}^{-1}$ may depend on $\mathbf{x}$.

Whether approach (2) or (3) is chosen, optimization remains a basic problem. In many geophysical inverse problems the function f , whatever its form, is rife with local minima. One might naively suggest that an exhaustive evaluation of all possible solutions be made. Usually, however, there is a large number of model parameters, each of which may assume some similarly large number (or a continuum) of values. Specifically, if s parameters can each assume one of q values, then there are q^s possible solutions. The utter enormity of this number definitely rules out the use of an exhaustive search.

Despite this assumed complexity, equations (2) and (3) do not necessarily pose intractable optimization problems. The key to their solution is prior information. For example, conventional approaches to nonlinear inversion usually incorporate an initial guess $\mathbf{x}^0$ for $\mathbf{x}$. The remaining perturbation $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}^0$ is then assumed to approximately satisfy the linear relation

$$\mathbf{G}(\mathbf{x}) \approx \mathbf{G}(\mathbf{x}^0) + \mathbf{F}\Delta\mathbf{x} \ ,$$

where $\mathbf{F}$ is a matrix of partial derivatives $\partial G_i \,/\, \partial x_j$ evaluated at $\mathbf{x}^0$. Then, letting $\mathbf{d}^0 = \mathbf{G}(\mathbf{x}^0) + \mathbf{n}$ and $\Delta d = \mathbf{d} - \mathbf{d}^0$, one solves for the Δx that satisfies

$$\mathbf{F}\Delta\mathbf{x} = \Delta\mathbf{d} \ .$$

This yields a possible solution $\mathbf{x}^1 = \mathbf{x}^0 + \Delta x$, which might or might not be satisfactory for minimizing (2) or (3). If it is not, the procedure is iterated by replacing $\mathbf{x}^0$ with $\mathbf{x}^1$, etc. Iterative techniques of this general form are widely used - reviews are contained in Parker (1977), Aki and Richards (1980), and Lines and Treitel (1984). The basic shortcoming of these techniques, however, is their reliance upon a good initial guess, without which they might fail severely.

What can be done, then, if there is no basis for an initial guess? Prior information, in the form of Bayesian inference, may still light the way. If we can formulate a prior probability distribution $\mathbf{P}(\mathbf{X}{=}\mathbf{x})$, we may be able to assign low probabilities to much of the parameter space, thereby effectively eliminating it. Later I will examine the problems encountered in residual statics estimation when statics are unusually large and no initial estimates of the timing delays are available. By relating probability to the power in a common midpoint stack, I will construct such an informative prior.

The prior distribution represents relative weights assigned for all possible $\mathbf{x}$ before any data are collected. After the data are observed, standard statistical analysis is used to combine the prior and the observed data to obtain the posterior probability distribution $\mathbf{P}(\mathbf{X}{=}\mathbf{x} \mid \mathbf{D}{=}\mathbf{d})$ via Bayes' theorem (Bard, 1974):

$$\mathbf{P}\,(\mathbf{X} = \mathbf{x} \mid \mathbf{D} = \mathbf{d}\,) = \frac{\mathbf{P}\,(\mathbf{D} = \mathbf{d} \mid \mathbf{X} = \mathbf{x}\,)\,\mathbf{P}\,(\mathbf{X} = \mathbf{x}\,)}{\mathbf{P}\,(\mathbf{D} = \mathbf{d}\,)} \ . \qquad (4)$$

Later we will want to find the $\mathbf{x}$ that maximizes the posterior probability - this step is often called maximum a posteriori (MAP) estimation, and presents yet another optimization problem.

One interesting generalized approach to the incorporation of prior information in nonlinear inversion is described in the paper by Tarantola and Valette (1982). Although their formalism is appealing, the generalized use of prior information alone will not necessarily make a problem computationally tractable. Nonlinear geophysical inversion will often require more - specifically, a method that employs available knowledge to obtain a *computable* solution.

GIBBS-MARKOV MODELS

Many large-scale problems in reflection seismology can be naturally subdivided to a much smaller, computationally convenient scale. The surface-consistent analysis of seismic data, for example, can be divided into related sub-units, the size of which depend on the seismic cablelength (typically much less than the length of a seismic survey). The analysis that follows attempts to unify estimates of parameters made in sub-units with estimates that are consistent with the entire dataset. The statistical model I will use can be described in either of two ways that have been shown to be formally equivalent. One model is derived from probability theory and is called a *Markov random field.* The other model is the *Gibbs distribution* of statistical physics. Our goal is to divide a large problem into smaller, more manageable parts while maintaining the relationships between the smaller parts as precisely as possible. Ultimately we will want to solve a large-scale nonlinear inversion problem. The Gibbs-Markov model forms the foundation of this objective.

Markov random fields (Kindermann and Snell, 1980) describe the structure common to the particular class of problems I address. A Markov random field is the spatial counterpart of the one-dimensional Markov chain. The simplest form of a Markov chain is a stationary sequence in which the conditional probability of an event at time t depends only on the value of the sequence at time $t-1$. Because the event at time t is independent of all times other than $t-1$, we may write, for a random sequence $\mathbf{X} = \{ X_0, X_1, \ldots, X_t \}$,

$$\mathbf{P}(X_t = x_t \mid X_{t-1} = x_{t-1}, \ldots, X_0 = x_0) = \mathbf{P}(X_t = x_t \mid X_{t-1} = x_{t-1}) \ .$$

A straightforward generalization leads to the consideration of a two-dimensionally indexed set of random variables $\mathbf{X} = \{X_{ij}\}$. The X_{ij} define a Markov random field if the value of each X_{ij} depends only on a *neighborhood* A_{ij} of (i, j). A_{ij} might contain only the nearest neighbors of x_{ij}:

$$A_{ij} = \{ (i+1, j), (i-1, j), (i, j+1), (i, j-1) \} \ .$$

Other, more complex neighborhood structures are possible (and are employed here). In general, neighborhoods contain only those model parameters that most immediately influence the values a given parameter may assume. A two-dimensional Markov random field with an arbitrary neighborhood structure A_{ij} is stated as

$$\mathbf{P}[X_{ij}=x_{ij} \mid X_{kl}=x_{kl}\,,(k\,,l\,)\neq(i\,,j\,)] = \mathbf{P}[X_{ij}=x_{ij} \mid X_{kl}=x_{kl}\,,(k\,,l\,)\in A_{ij}\,] \ . \tag{5a}$$

It is also required that all possible parameter vectors have positive probability:

$$\mathbf{P}(\mathbf{X}=\mathbf{x}\,) > 0 \ \text{for all} \ \mathbf{x} \ . \tag{5b}$$

Figure 1 shows an example of a Markov random field on a two-dimensional lattice.

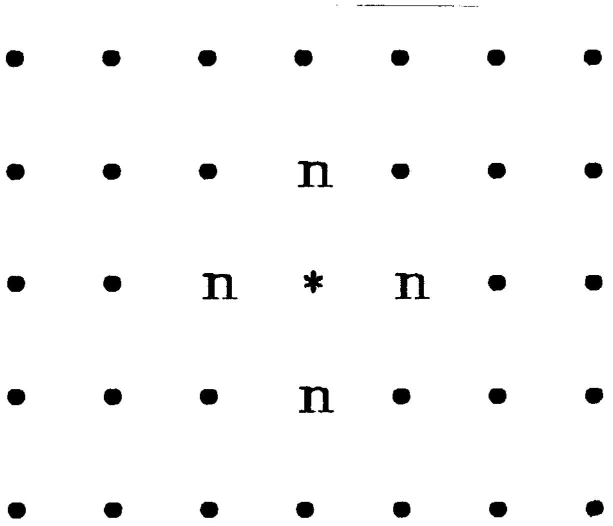

FIG. 1. A schematic representation of a Markov random field with nearest-neighbor interactions on a two-dimensional lattice. The probability distribution of the parameter located at position * depends only on its nearest neighbors, the n's; thus **P**(* | all else) = **P**(* | neighbors).

The notion of a Markov random field describes only the general form of local, conditional probabilities. While useful in concept, these local conditional probabilities provide only indirect information on the behavior of the complete set of parameters. More direct information should be in the form of a joint probability distribution **P**(**X**=**x**). This is fortunately available, because all Markov random fields exhibit a Gibbs probability distribution, and all Gibbs distributions define a Markov random field (Geman and Geman, 1984; Kindermann and Snell, 1980; Moussouris, 1974). Gibbs (or canonical) distributions arise in statistical physics in the study of systems in thermal equilibrium. **X** is Gibbs if

$$P(\mathbf{X}=\mathbf{x}) = \frac{1}{Z}\, e^{\frac{-E(\mathbf{x}\,)}{k_B\ T}} \ . \tag{6}$$

$E(\mathbf{x})$ is called the *energy,* and is the sum of local *potentials* $\phi_{A_{ij}}(\mathbf{x})$ such that

$$E(\mathbf{x}) = -\sum_{i,j} \phi_{A_{ij}}(\mathbf{x}) \ . \tag{7}$$

The $\phi_{A_{ij}}$ are evaluated over the same neighborhood structure A_{ij} used to specify the conditional probabilities (5a); i.e., $\phi_{A_{ij}}$ depends only on x_{kl}, $k,l \in A_{ij}$. T is *temperature* and k_B is *Boltzmann's constant.* Z is the normalizing constant

$$Z = \sum_{\mathbf{x}} e^{\frac{-E(\mathbf{x})}{k_B T}} \ , \tag{8}$$

called the *partition function.*

The Gibbs-Markov equivalence lends physical significance to the purely probabilistic concept of a Markov random field. If a problem can be divided into sub-units that satisfy equations (5), then physical insight gleaned from the statistical mechanics expressed by equations (6), (7), and (8) is applicable to that problem. The key to intuitively connecting the physics with the statistics lies in the energy function (7). Here it is evident why Gibbs distributions and Markov random fields are closely related: the same neighborhoods A_{ij} limit the spatial range of both the conditional probabilities (5a) and the x_{ij} needed to evaluate the local potentials $\phi_{A_{ij}}$. Only the general form of the energy function has been specified, however. Kirkpatrick et al. (1983) proposed an important additional bridge between the physics and the statistics by letting the energy represent an objective function in an optimization problem. In residual statics estimation, the objective function will be the negative power in a common midpoint stack. The Gibbs probability (6) then increases as stack power increases; T is expressed in the same units as the objective function and acts as a control parameter akin to the variance in a Gaussian distribution (but T has much physical significance).

The utility of the Gibbs representation will be evident shortly. Note now, however, not only that a large problem has been subdivided into smaller parts (Markov probabilities), but also that the joint Gibbs distribution describes the presumed interactions of these individual parts by supplying the prior probability of any given parameter *vector.* Moreover, given the prior, we can now derive the posterior distribution and complete the statistical statement of the problem.

THE GIBBS POSTERIOR

Assuming that $\mathbf{X}$ is a Markov random field, we may write the joint prior probability distribution for the model parameters as

$$\mathbf{P}(\mathbf{X}=\mathbf{x}) = \frac{1}{Z} e^{\frac{-E(\mathbf{x})}{T}} , \tag{9}$$

where for convenience we set $k_B = 1$. The choices of E and T are deferred until later discussions of applications. Following Geman and Geman (1984), I now show that the posterior probability $\mathbf{P}(\mathbf{X}=\mathbf{x} \mid \mathbf{D}=\mathbf{d})$ is also a Gibbs distribution.

Starting with Bayes' theorem [equation (4)], we substitute the Gibbs prior for $P(\mathbf{X}{=}\mathbf{x})$ and take $\mathbf{P}(\mathbf{D}=\mathbf{d})$ to be constant, to get

$$\mathbf{P}(\mathbf{X}=\mathbf{x} \mid \mathbf{D}=\mathbf{d}) = \frac{1}{Z} \mathbf{P}(\mathbf{D}=\mathbf{d} \mid \mathbf{X}=\mathbf{x})\, e^{\frac{-E(\mathbf{x})}{T}} \tag{10}$$

where Z is now a new constant. The noise $\mathbf{N} = \{N_1, \ldots, N_r\}$ is assumed to be independent, identically distributed, and independent of $\mathbf{X}$. For analytic convenience the probability distribution of the noise is assumed to be zero-mean with the form

$$\mathbf{P}(\mathbf{N}=\mathbf{n}) = c^{-1} e^{-\frac{1}{2}\left[\frac{||\mathbf{n}||_p}{\sigma}\right]^p} \tag{11}$$

where c and σ are constants and $|| \bullet ||_p$ is the L^p norm such that $(|| \mathbf{n} ||_p)^p = r^{-1}\sum_i^r n_i^p$. If $p = 2$ the noise is Gaussian, and if $p = 1$ the noise is exponential.

We now solve for the posterior. Equation (10) is rewritten as

$$\mathbf{P}(\mathbf{X}=\mathbf{x} \mid \mathbf{D}=\mathbf{d}) = \frac{1}{Z} \mathbf{P}[\,\mathbf{D}=\mathbf{G}(\mathbf{x})+\mathbf{n} \mid \mathbf{X}=\mathbf{x}\,]\, e^{\frac{-E(\mathbf{x})}{T}} \tag{12}$$

$$= \frac{1}{Z} \mathbf{P}[\,\mathbf{N}=\mathbf{d}-\mathbf{G}(\mathbf{x}) \mid \mathbf{X}=\mathbf{x}\,]\, e^{\frac{-E(\mathbf{x})}{T}} . \tag{13}$$

Because $\mathbf{N}$ is independent of $\mathbf{X}$,

$$\mathbf{P}(\mathbf{X}=\mathbf{x} \mid \mathbf{D}=\mathbf{d}) = \frac{1}{Z} \mathbf{P}[\,\mathbf{N}=\mathbf{d}-\mathbf{G}(\mathbf{x})\,]\, e^{\frac{-E(\mathbf{x})}{T}} \tag{14}$$

and by substituting from equation (11),

$$\mathbf{P}(\mathbf{X}=\mathbf{x} \mid \mathbf{D}=\mathbf{d}) = \frac{1}{Z} e^{\frac{-E(\mathbf{x})}{T} - \frac{1}{2}\left[\frac{||\mathbf{n}||_p}{\sigma}\right]^p} \tag{15}$$

where Z is again a new normalizing constant. By writing the exponent as

$$E'(\mathbf{x},\mathbf{d}) = \frac{E(\mathbf{x})}{T} + \frac{1}{2}\left(\frac{||\mathbf{d}-\mathbf{G}(\mathbf{x})||_p}{\sigma}\right)^p \tag{16}$$

we see that the posterior distribution is also Gibbs with energy function $E'(\mathbf{x},\mathbf{d})$:

$$\mathbf{P}(\mathbf{X}=\mathbf{x} \mid \mathbf{D}=\mathbf{d}) = \frac{1}{Z}\, e^{\frac{-E'(\mathbf{x},\mathbf{d})}{T}} . \tag{17}$$

Note that the form of the Gibbs posterior is unaffected by the value of p ; thus the noise need not be Gaussian.

Geman and Geman (1984) derive some additional results showing that the posterior neighborhood structure is slightly modified to include "second-order" neighbors (i.e., neighbors of neighbors). For computational purposes, however, I assume that the prior and posterior neighborhood structures can be taken equal.

The model parameters that best fit the data, from the viewpoint of Bayesian inference, are determined by maximizing the posterior probability (17). This is *maximum a posteriori* (MAP) estimation. Maximizing this posterior probability by conventional gradient techniques is often not possible for nonlinear problems because of the existence of many local extrema. We will next see that the method of simulated annealing can, however, perform global optimization because it is based on the statistical mechanics of equation (17).

SIMULATED ANNEALING

Kirkpatrick et al. (1983) introduced simulated annealing in an attempt to solve apparently intractable optimization problems. Simulated annealing is a Monte Carlo optimization procedure based on chemical annealing. Annealing is the way in which crystals are grown - a melt is cooled very slowly until a crystal is formed. The rate of cooling is important, because a non-crystalline, metastable glass can form if cooling is too rapid. Kirkpatrick et al. viewed the growth of a crystal as analogous to finding the global minimum in optimization, and the development of a glass as the analog to wrongly selecting a local minimum. Their primary application was directed at the combinatorial optimization problems that arise in the physical design of computers. Other researchers are now using simulated annealing in image restoration (Geman and Geman, 1984), artificial intelligence (Hinton and Sejnowski, 1983), and elsewhere.

The method proposed by Kirkpatrick et al. is a variant of a Monte Carlo integration procedure due to Metropolis et al. (1953). Metropolis et al. addressed the problem of random sampling from a Gibbs distribution at constant temperature, thereby simulating the average behavior of a physical system in thermal equilibrium. The Metropolis algorithm proceeds in the following way. For each model parameter (or magnetic spin, molecular position, etc.) $\mathbf{X}_{ij}$, a random perturbation is made, and the change in energy, ΔE, is computed. If $\Delta E \leq 0$ (i.e., if energy decreases), the perturbation is accepted. If ΔE is positive then the perturbation is accepted with probability

$$\mathbf{P}(\Delta E) = e^{\frac{-\Delta E}{T}} . \tag{18}$$

This conditional acceptance is easily implemented by choosing a random number α uniformly distributed between 0 and 1. If $\alpha \leq \mathbf{P}(\Delta E)$ then the perturbation is accepted; otherwise the existing value for the parameter is retained. Random perturbation according to these rules eventually causes the system to reach equilibrium, in which configurations $\mathbf{x}$ are realized with a Gibbs probability distribution. Because each step of the algorithm is dependent only on the present and not the past, the algorithm can be formally studied using Markov chain theory - see Fosdick (1963), Hammersley and Handscomb (1964), and Binder (1979, 1984) for reviews.

Kirkpatrick's optimization technique slowly lowers the temperature T during execution of the Metropolis algorithm. If the system is cooled sufficiently slowly and equilibrium conditions are maintained, the model parameters eventually converge to a (ground) state of minimum energy (or maximum probability). See Geman and Geman (1984) for a convergence proof. The essential characteristic of this optimization procedure is its ability to escape from local minima and locate the global minimum with a high degree of accuracy. Thus, in the Bayesian formulation here, perturbations that *lower* probability are accepted in accordance with equation (18), but the final result yields the model associated with *maximum* probability. When simulated annealing is employed to estimate residual static corrections, the algorithm will accept not only random guesses of residual statics that lead to increased stack power, but also some random guesses that decrease power. The final result will yield the stack with maximum power.

RESIDUAL STATICS ESTIMATION

When reflection seismic data are acquired on land, sources and receivers are generally placed on or near the surface. Eventual processing and interpretation of seismic data assumes that the data were collected on level terrain. If this is not the case, rough corrections for variations in elevation are made early in processing. These constant (static) time shifts are called *field statics* because the corrections are based on a surveyor's field measurements. Field static corrections are only approximate; the unconsolidated near-surface weathering layer can exhibit substantial variations in seismic velocity that also cause static timing differences. This latter problem can exist even when surface terrain is flat. These secondary corrections are derived from the data after field statics have been corrected and are called *residual statics.*

The conventional model for obtaining residual statics solutions (Wiggins et al., 1976; Taner et al., 1974) expresses the observed static time deviations t_{ij} of normal moveout corrected traces associated with the i th shot and j th receiver as

$$t_{ij} \;=\; s_i + r_j + g_k + m_k x_{ij}^2 \;. \tag{19}$$

The unknown surface-consistent time shifts due to near-surface velocity anomalies underneath the i th shot and j th receiver are denoted by s_i and r_j, respectively. The unknown subsurface-consistent part of the time shift due to variations in geologic structure at the k th midpoint is given by g_k. The last term represents the component that is due to residual normal moveout: m_k is the residual normal moveout coefficient, and x_{ij} is the distance between shot i and receiver j. The m_k are included in an attempt to account for the usually imperfect stacking velocities used prior to measuring t_{ij}.

The validity of the linear model (19) rests on two important assumptions: (1) waves travel through the near-surface with approximately vertical raypaths, and (2) the observed deviations t_{ij} are reasonably accurate. The first assumption is definitional; if near-surface raypaths were not vertical, "statics" would be "dynamics". The second assumption deserves further attention.

Each t_{ij} is usually measured by crosscorrelating an unstacked data trace against a "reference" or "pilot" trace, and then equating t_{ij} to the lag that yields the greatest value in the crosscorrelation function. This is an appropriate technique if the reference trace is reasonably similar to the data trace, but gross, irrecoverable errors can occur if the traces are sufficiently dissimilar (due either to excessive noise, large statics, or both). Equation (19) represents a large system of overdetermined and underconstrained linear equations (Wiggins et al., 1976) that are generally solved by least squares techniques. Least squares solutions are optimal if the errors between the observed t_{ij} and the "true" t_{ij} are Gaussian. If a small number of errors are non-Gaussian, robust estimation methods may still produce an effective least squares solution (Donoho, 1979). If the statics are sufficiently large and the data are sufficiently poor to cause a large number of gross errors in t_{ij}, the least squares method will fail - this is the "cycle-skipping" or "leg-jumping" problem. A nonlinear approach is then needed.

Let the data recorded at the j th receiver, after the i th shot has been fired at time $t = 0$, be denoted by $d_{ij}(t)$. Let $f_{ij}(t)$ be the same data but without noise $n_{ij}(t)$ and without static shifts. Then

$$d_{ij}(t) \;=\; G_{ij}(s_i, r_j)\, f_{ij}(t) + n_{ij}(t) \tag{20}$$

where

$$G_{ij}(s_i, r_j)\, f_{ij}(t) \;\equiv\; f_{ij}(t - s_i - r_j) \;. \tag{21}$$

We want to estimate the s_i and the r_j directly from the seismic data.

Although each G_{ij} is a simple linear operator, the recovery of the s_i and r_j can not be linear unless additional important assumptions are made. One such assumption is made in equation (19). The t_{ij} play the role of an initial guess: they are initial estimates of timing delays that are then decomposed into surface- and subsurface-consistent components. This decomposition is usually performed with a linear least squares technique. This linearization fails, however, when the t_{ij} contain gross errors due to large statics and noise-contaminated data.

I now cast the nonlinear statics problem of equation (20) in the form of a Gibbs-Markov model, in order to obtain a general solution that is valid for any severity of statics and noise. Seismic traces are first sorted to midpoint-offset $(y-h)$ coordinates and normal moveout corrected (with approximate velocities) to produce the new set of data $d_{yh}(t)$. It is assumed that each common midpoint gather contains traces that are identical except for surface-consistent time shifts and uncorrelated noise. The best estimate of shot and receiver statics will thus be those static shifts that maximize the total power in all common midpoint stacks. To express this formally, define the inverse of the shifting operator in midpoint-offset coordinates,

$$G_{yh}^{-1}\, d_{yh}(t) \;\equiv\; d_{yh}(t + s_{i(y,h)} + r_{j(y,h)}) \;, \tag{22}$$

where i and j are both functions of y and h. The objective I have selected is to minimize the negative of the total stack power as a function of the shot statics $\mathbf{s}$ and receiver statics $\mathbf{r}$; thus

$$[\mathbf{s},\mathbf{r}] \;=\; \min_{[\mathbf{s},\mathbf{r}]} \left\{ -\sum_y \sum_t \left[\sum_h G_{yh}^{-1}\, d_{yh}(t) \right]^2 \right\} \;, \tag{23}$$

where $\mathbf{s}$ and $\mathbf{r}$ determine the inverse operators G_{yh}^{-1} and the sums are taken over all y, t (within some time gate), and h. This is an optimization problem of the general form shown in equation (3). To make global optimization tractable when many local minima exist, the problem is divided into the interdependent parts of a Gibbs-Markov model.

Seismic cables are usually short relative to the length of a survey; consequently, each shot static s_i influences (in an immediate sense) the stack power of only a subset

Y_{s_i} of all midpoints y. Likewise each receiver static (immediately) affects only a subset Y_{r_j}. To measure the contribution to stack power due to s_i we compute

$$\phi_{s_i}(\mathbf{s},\mathbf{r}) \;=\; \sum_{y \in Y_{s_i}} \sum_t \left[\sum_h d_{yh}(\,t \,+ s_{i(y,h)} + r_{j(y,h)}\,) \right]^2 \;, \tag{24a}$$

where s_k, $k \neq i$, and all r_j are fixed. Similarly,

$$\phi_{r_j}(\mathbf{s},\mathbf{r}) \;=\; \sum_{y \in Y_{r_j}} \sum_t \left[\sum_h d_{yh}(\,t \,+ s_{i(y,h)} + r_{j(y,h)}\,) \right]^2 \;, \tag{24b}$$

where now r_k, $k \neq j$, and all s_i are fixed. The minimization problem in equation (23) is now replaced by

$$[\mathbf{s},\mathbf{r}] \;=\; \min_{[\mathbf{s},\mathbf{r}]} E(\mathbf{s},\mathbf{r}) \tag{25a}$$

where

$$E(\mathbf{s},\mathbf{r}) \;=\; -\sum_i \phi_{s_i}(\mathbf{s},\mathbf{r}) - \sum_j \phi_{r_j}(\mathbf{s},\mathbf{r}) \tag{25b}$$

In this formulation, the total stack power E plays the role of the energy in equation (17). Maximizing stack power is then equivalent to maximizing the posterior probability (17). Because the seismic cable is shorter than the survey line, E can be partitioned into the additive "local energies" ϕ_{s_i} and ϕ_{r_j}, each of which depends only on the s_i and r_j located within a cablelength. The cablelength determines the "nearest neighbors" in the sense of a Gibbs-Markov model; see Figure 2.

The subsurface-consistent terms g_k and m_k are not included in this approach. The g_k represent timing differences due to geologic variation from midpoint to midpoint, and are useful only with models like equation (19) whose solutions depend on measurements of trace-to-trace time deviations. However, the power computations in (24a,b) are performed within midpoint gathers, where the g_k are constant and therefore irrelevant. This is an important point. The statics solution presented here does *not* need to decompose structural and near-surface variations. Long wavelength statics, however, are still poorly resolved by the data and are generally suspect [see Wiggins et al. (1976) for more details]. Residual normal moveout, although ignored, is still an important parameter. Unfortunately, estimating the m_k with an objective function like (24) is cumbersome. It is of course possible, but my experience has shown that it would not merit the additional computational burden.

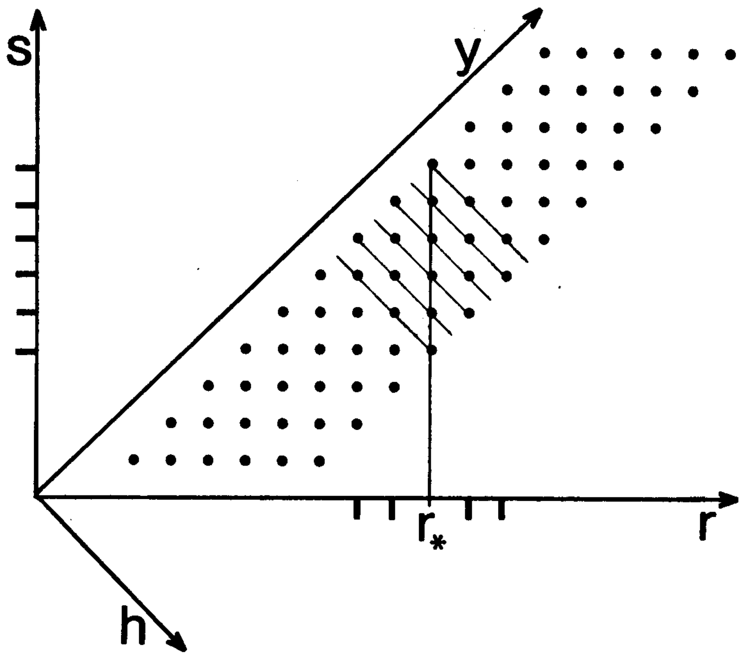

FIG. 2. Schematic representation of a seismic survey conducted with a 6-trace cable. s, y, r, and h denote the shot, midpoint, receiver, and offset axes, respectively. Each dot represents a single seismic trace (the time axis may be taken to go into the page). Each trace is uniquely specified by any 2 of the 4 coordinate axes. In this picture, all traces recorded at receiver location r_* are connected by a vertical line. Calculation of ϕ_{r_*} requires summing over the h-axis for all midpoint gathers containing a trace recorded at location r_*; this sum over h is depicted by the diagonal lines perpendicular to the y-axis. This calculation of stack power depends only on the shot statics and receiver statics located on the s- and r-axes at the positions marked by a dash. These are the "nearest neighbors" in the sense used in a Gibbs-Markov model.

IMPLEMENTATION

Simulated annealing is an iterative technique that continually creates samples from a Gibbs distribution while slowly decreasing the temperature parameter T (to exaggerate the peaks and troughs of the probability distribution). To help one choose the initial temperature T_0, it is often useful to compare the input stack power p_0 with the stack power p_r, which is computed after applying random shot and receiver statics. For a given β between 0 and 1, T_0 can then be chosen such that

$$\beta = e^{-\frac{p_0 - p_r}{T_0}} .$$

β determines the degree of "melting" prior to "annealing," and represents the probability of the algorithm accepting a decrease in power by the amount $p_0 - p_r$. Cooling can then proceed in any of a number of ways. My best results were attained with a logarithmic cooling function of the form $T_k = T_0/\log k$, where k is the number of iterations over each shot and receiver static. (One iteration includes one attempted perturbation of each shot and receiver static.) Geman and Geman (1984) proved that the algorithm converges when this logarithmic function is used. Practically, the most crucial requirement of any cooling function is that it be slow, especially near the "critical temperature" where convergence is rapid. Choosing a successful annealing schedule requires experience; ideally, the procedure would be interactive. I was usually able to produce a successful result after a few trials. I find it best to set β close to 1 for an initial exploratory run with $T_k = T_0\alpha^k$, with α equal to, say, .99. Then, after roughly determining the critical temperature, the more gentle logarithmic function may be used. My results thus far suggest that many iterations at a single temperature may suffice, if the temperature is chosen just below the critical temperature.

Once started, the next question to resolve is when to stop. In my tests I collected run statistics every 10 iterations. The algorithm simply stops if 10 iterations have passed in which few or no perturbations are accepted.

Hinton and Sejnowski (1983), Geman and Geman (1984), and others point out that simulated annealing can be implemented in parallel. Parallel computers are collections of processors that operate simultaneously while maintaining a continuous flow of information from each processor to its "neighbors." [See Hockney and Jesshope (1981) for a survey.] In theory, if there were N processors (say, one for each parameter), run time could be reduced by a factor of N. In certain cases, however, the Metropolis algorithm has only half this expected parallelism (Vichniac, 1984).

NUMERICAL EXAMPLE

Residual statics estimation by simulated annealing was tested on synthetic data that exhibit a severe surface-consistent statics problem. The data simulate the results of a survey conducted with a 12-trace cable, off-end shooting with a two receiver group gap, and evenly spaced shots and receiver groups. There are 100 6-fold common midpoint gathers. The sampling rate is 4 msec. and the data contain frequencies from 5-60 Hz. The data, prior to the introduction of static shifts, are shown in Figures 3a and 3b. Figure 3a shows four representative "moveout-corrected" common midpoint gathers, and Figure 3b is the common midpoint stack. The cablelength extends over 24 stacked traces. The signal-to-noise ratio (the total power of the signal divided by the total power of the noise) after stack is approximately 2.0. The entire dataset is scaled to an rms amplitude of 100. For all traces the signal is identical, except for the bulk time shift simulating a fault. (Real faults would not exhibit such a severe discontinuity before migration.) These data represent the desired solution for the test illustrated in the following figures.

Random shot and receiver statics are displayed in Figure 4. These statics vary between ±40 msec., in 4 msec. increments. Figure 5a shows the same common midpoint gathers of Figure 3a, but now with the traces shifted in accordance with the statics model in Figure 4. Figure 5b is the common midpoint stack after the model statics were applied. Because of the severity of the statics, almost no indications of reflection events can now be observed. The data in Figures 5a,b are the input to the statics estimation algorithm.

Figures 6a-e illustrate the results of applying the statics algorithm. Random guesses for shot and receiver statics were constrained to fall within ±40 msec., in 4 msec. increments. Three stages of the algorithm's execution are depicted: the stack after 2410 iterations (6a); after 3080 iterations (6b); and the final solution, after 4540 iterations (6c), which closely resembles the desired solution in Figure 3b. Figure 6d shows the four common midpoint gathers from Figure 5b; the statics solution has now been applied to them. For this example, $T_k = T_0 \log k_0 / \log(k_0 + 2k)$, with $T_0 = 4500$, $k_0 = 5000$, and k equal to the number of iterations. This is a mild annealing schedule: T changes by less than 11% from start to finish. Figure 6e is a graph of stack power versus iteration. Note that there is very little change in power until after approximately 2300 iterations. After 3000 iterations, the power sharply increases. This type of sudden change is analogous to rapid crystallization, and was observable in the results of most tests. By the time iteration 3080 was reached, the statics algorithm completed its most important work: solving for the shorter-wavelength statics, leaving only long-wavelength residuals. The longer wavelengths are the most poorly resolved components of the solution; this is as true for the linearized technique of Wiggins et al. (1976) as it is here. By iteration 4540 (the final solution), only a slight long-wavelength residual remains. Although we observe here, as elsewhere, the fundamental ambiguity of long wavelength statics and structure, it is important to note that the severe structural discontinuity implied by the artificial fault does not influence the solution.

The quality of the solution is measured by the objective function, stack power. For comparison with results, the power of the input stack in Figure 5b is normalized to 1. The final stack power for the solution in Figure 6c is 3.354. The known, desired solution has a stack power of 3.399, so the computed solution is in error by approximately 1.3%. The difference between the estimated statics and the true statics is graphed in Figure 7. Note that, for both shots and receivers, the basic error occurs as a slight kink about two-thirds along the line. The noise contamination for this test was strong enough so

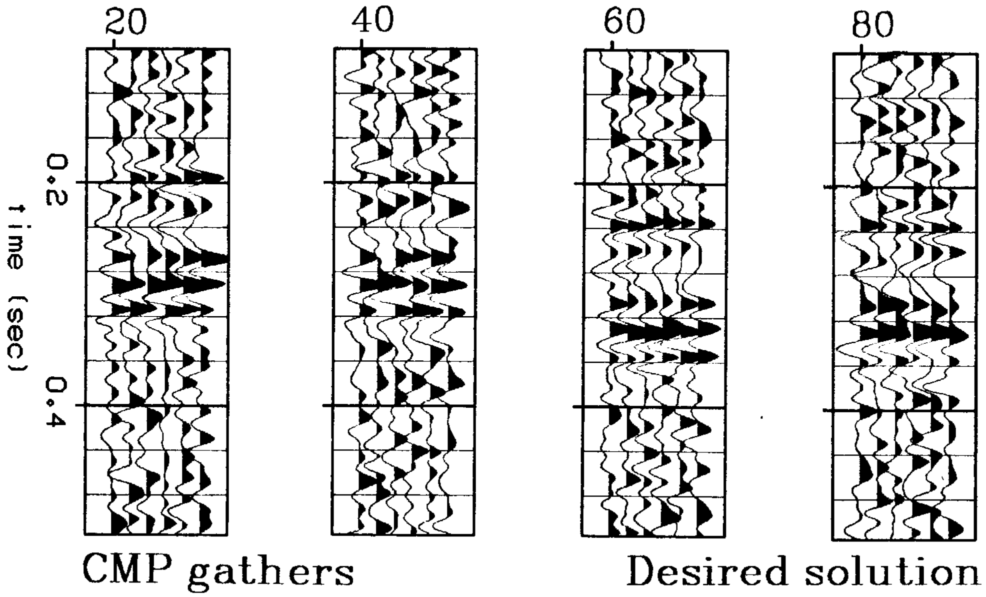

FIG. 3a. Four "moveout-corrected" common midpoint gathers. The gathers are shown without static shifts; there are 6 offsets in each gather. This correct alignment of traces is the desired solution for pre-stack data.

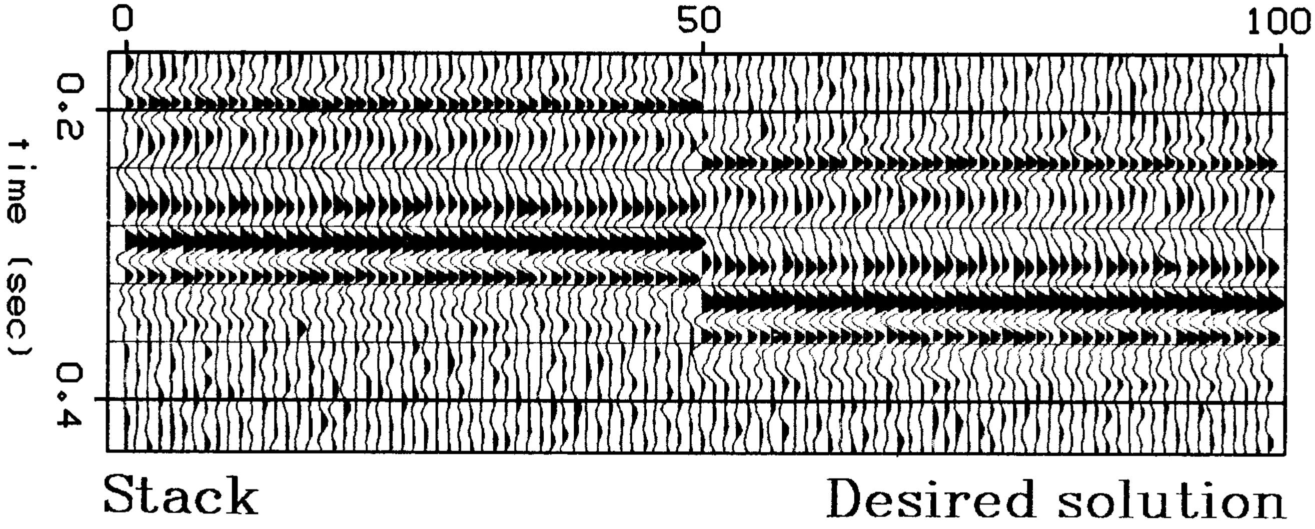

FIG. 3b. Common midpoint stack prior to the introduction of static shifts. The cablelength extends over 24 midpoints; there are 100 midpoints in total. The signal-to-noise ratio is approximately 2. This is the desired solution for stacked data.

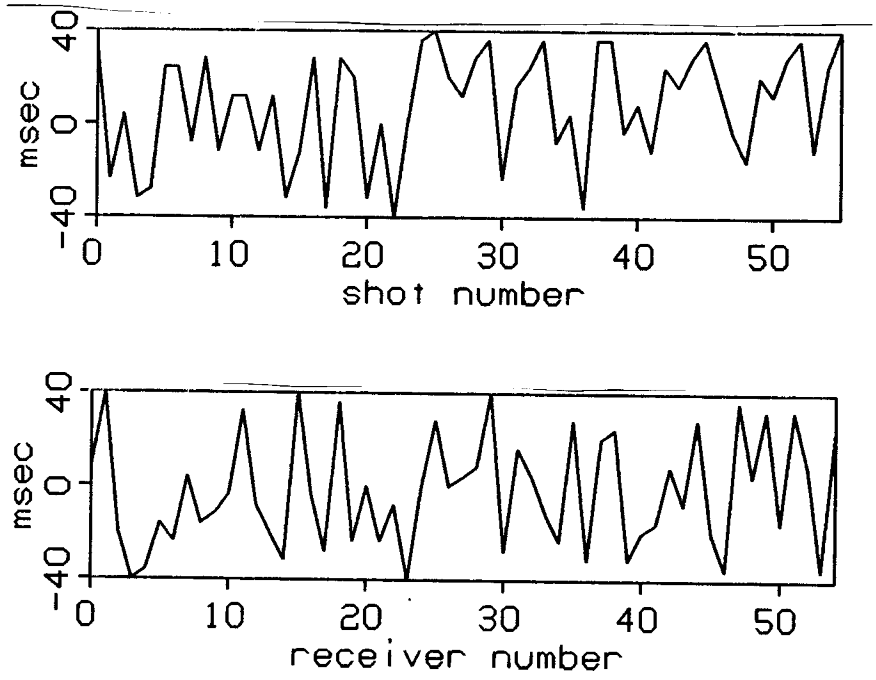

FIG. 4. Random shot statics (above) and receiver statics used to generate the test data in Figures 5a,b. Statics range between ±40 msec., in 4 msec. increments, for both shots and receivers.

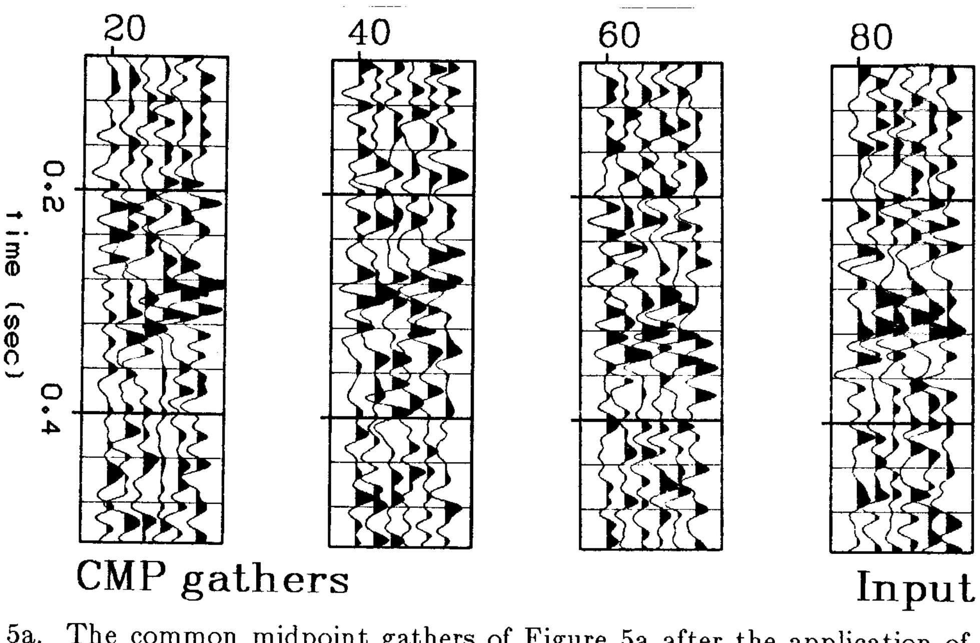

FIG. 5a. The common midpoint gathers of Figure 5a after the application of the static shifts in Figure 4. Note how the application of the statics has degraded the appearance of the data.

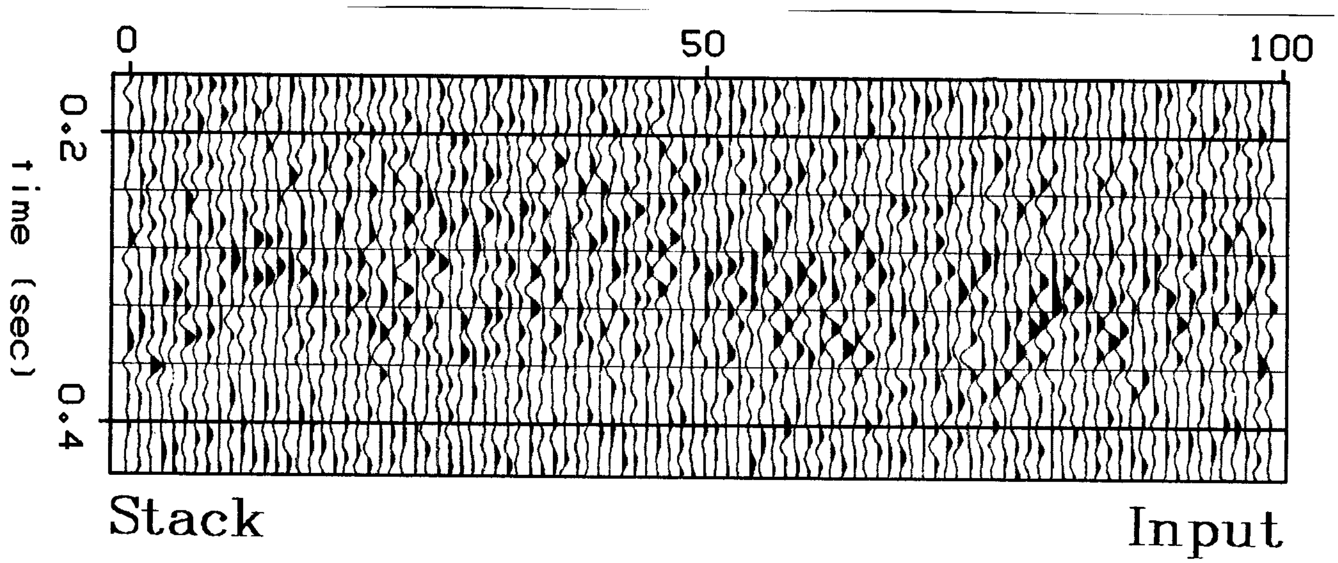

FIG. 5b. Common midpoint stack after application of the statics in Figure 4. Because the shifts are as much as 160 msec. apart, almost no indication of the reflection events in Figure 3b can now be observed. The data in Figure 5a and 5b are the input to the statics estimation algorithm.

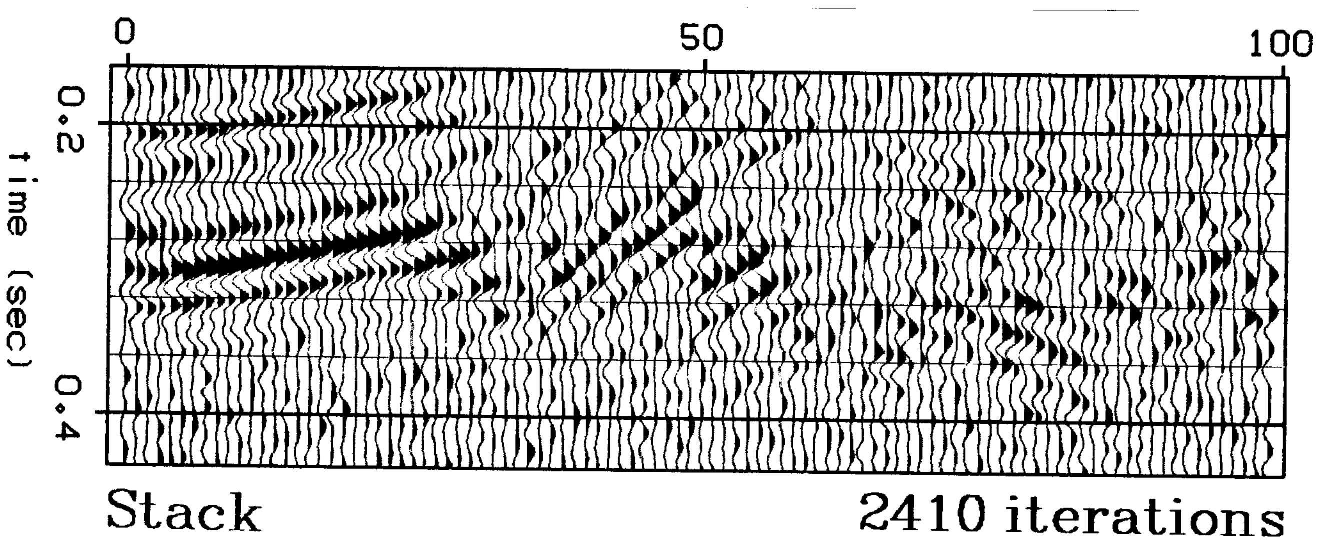

FIG. 6a. Common midpoint stack after 2410 iterations of the statics estimation algorithm. Good convergence already appears on the left, though the remainder of the section exhibits the effects of misaligned traces.

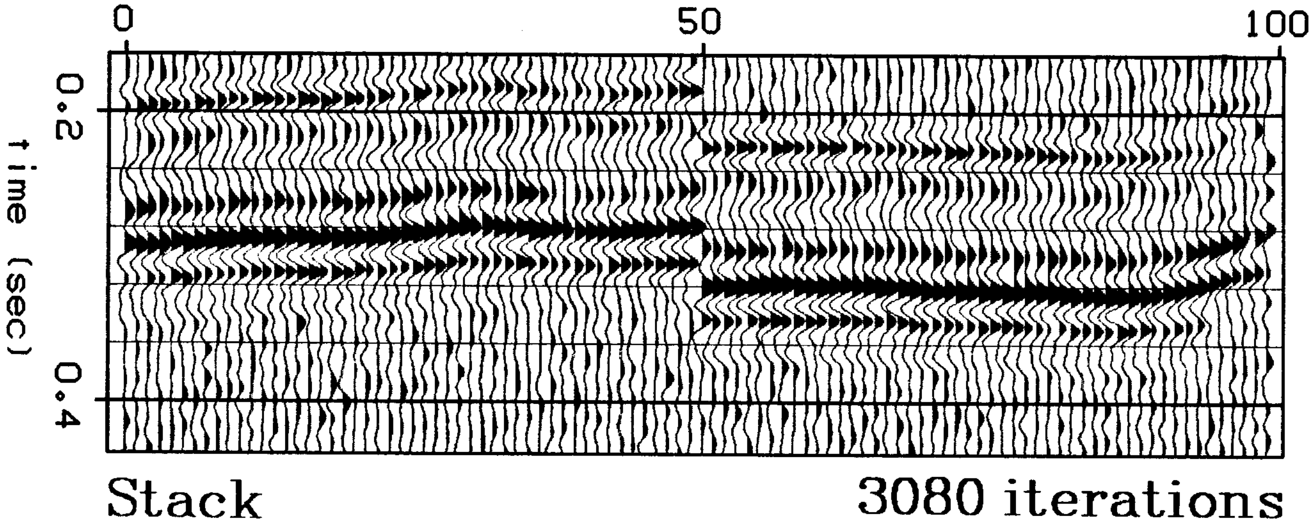

FIG. 6b. Common midpoint stack after 3080 iterations. Although long wavelength statics remain to be resolved, the bulk of the algorithm's work is completed. Note that, despite the ambiguity between structure and long wavelength statics, the artificial fault at trace 50 is properly resolved.

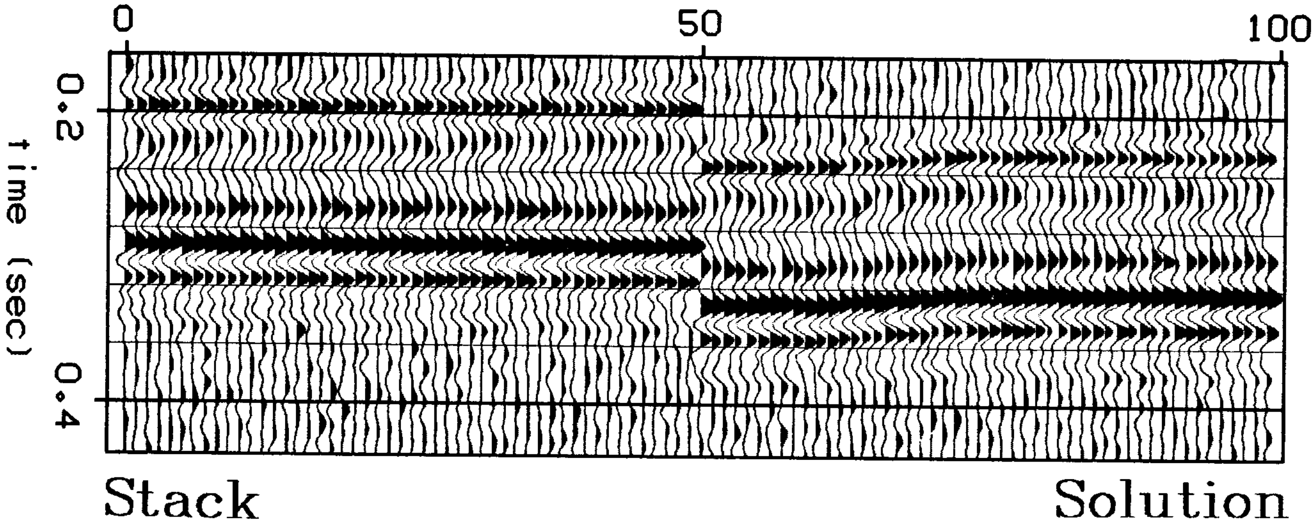

FIG. 6c. Common midpoint stack after 4540 iterations. This is the final solution, and should be compared with the input (Figure 5b) and the known, desired solution (Figure 3b). The 8 msec. rise on the right half of the section is a result of poorly resolved long-wavelength statics, due mostly to the noise contamination in the data.

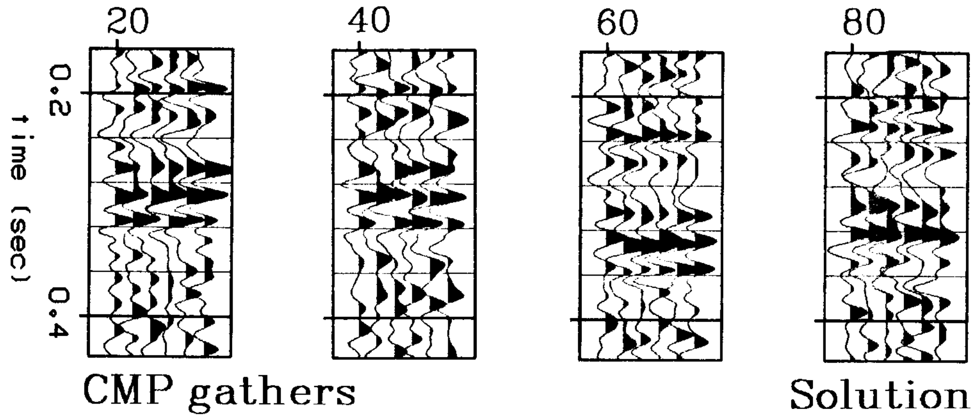

FIG. 6d. Common midpoint gathers after the statics solution has been applied. This should be compared to the input (Figure 5a) and the desired solution (Figure 3a). CMP 60 exhibits a slight error due to the poorly resolved long wavelength. The time axis is shorter now because the application of statics creates zeroes at early and late times.

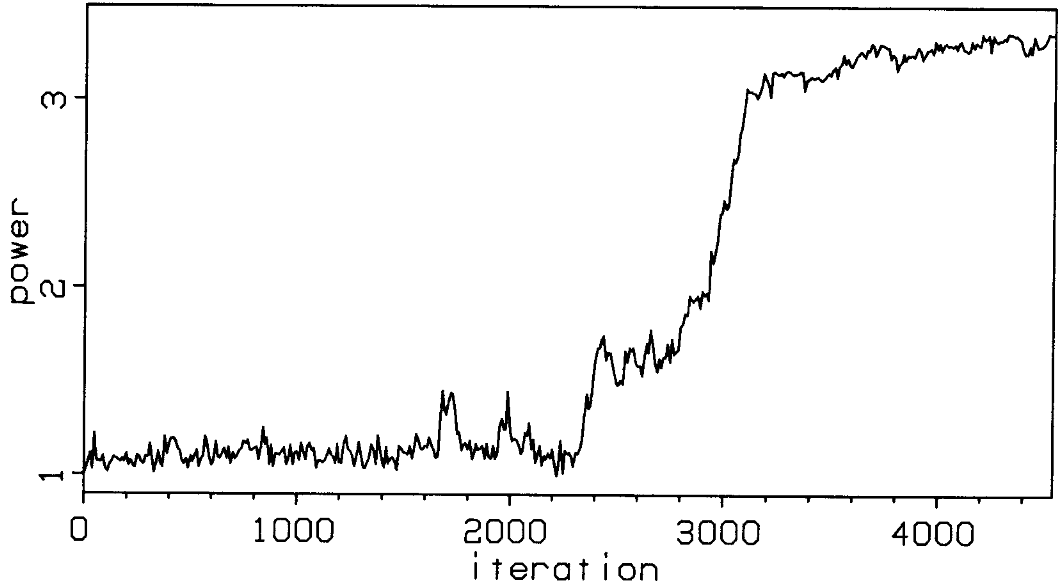

FIG. 6e. Stack power versus iteration number for the test leading to the result in Figure 6c. Stack power is computed every 10 iterations; the input stack power is normalized to 1. The final solution yields a stack power of 3.354, which is short of the true solution by 1.3%. Note the sudden increase in power after 3000 iterations. This abrupt change is analogous to rapid crystallization. Temperature decreases by less than 11% between the first and last iteration.

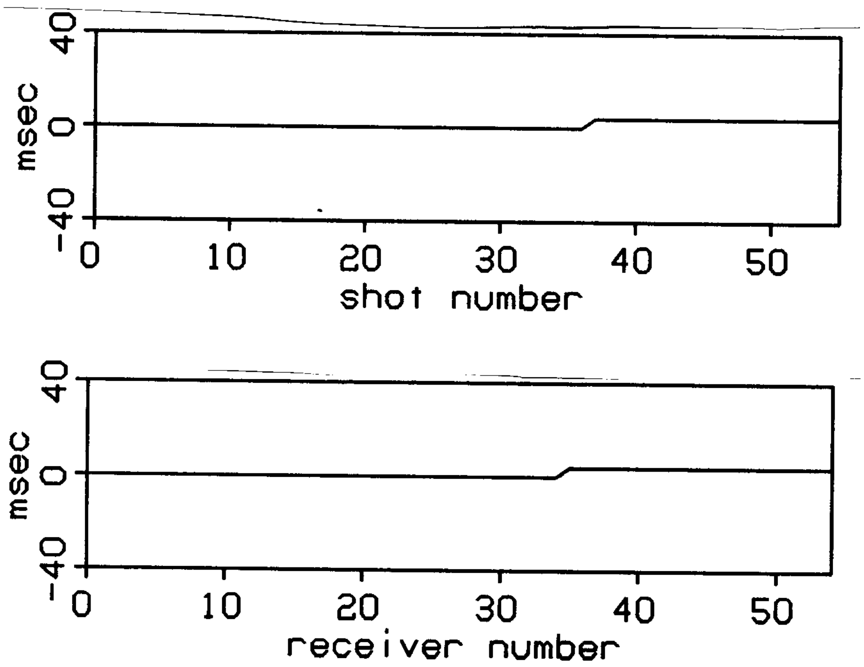

FIG. 7. Difference between the estimated statics and the true statics for the result in Figure 6c. The 8 msec. rise in the right half of Figure 6c is the result of the constant 4 msec. error for approximately the last 20 shot and receiver statics. The allowable values for statics fell within ±40 msec., in 4 msec. (1 sample) increments. The noise contamination for this test was too strong for the long-wavelength residual to be resolved.

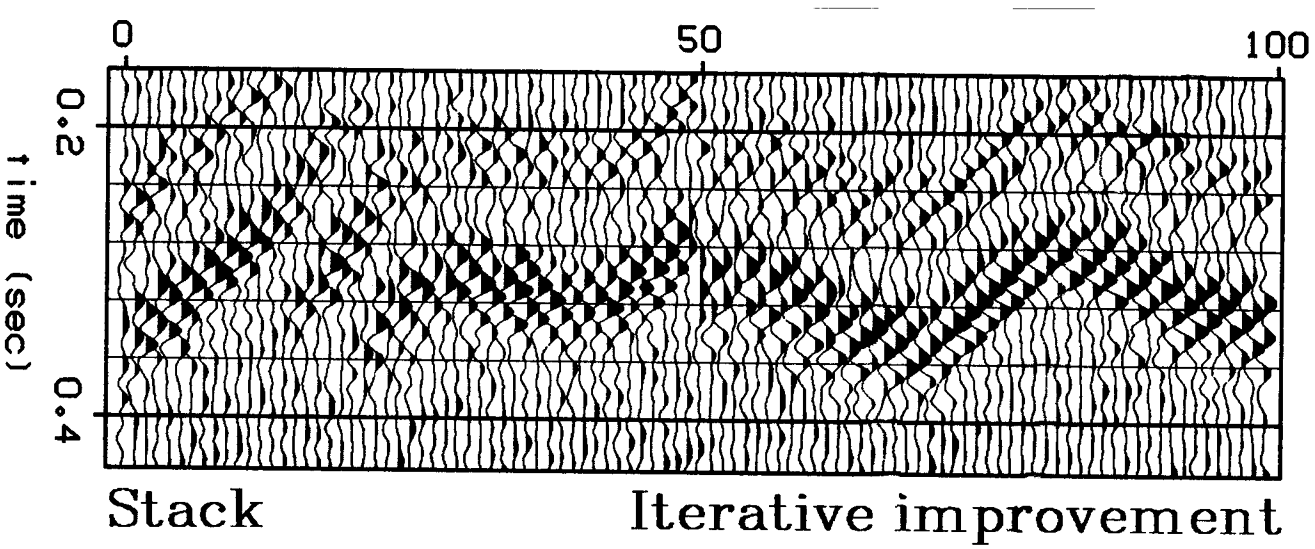

FIG. 8. The best result of 100 trials of residual statics estimation, made by iteratively choosing the best value for each shot and receiver static until a (local) power maximum is attained. Each trial was performed with a different random initialization of the shot and receiver statics. The stack power here is 29% short of the stack power of the annealing result in Figure 6c. The diagonal appearance of the stack is due to a severe cycle-skipping problem.

that this long wavelength residual could not be resolved; other tests (not shown) with higher signal-to-noise ratios more successfully resolved the long wavelengths.

The opposite of annealing is "quenching"; i.e., setting $T = 0$, so that only random perturbations that increase power are accepted. Efficient quenching can be accomplished by iteratively scanning all possible values for each shot and receiver static, and always choosing the shift that yields the greatest local stack power. This is optimization by *iterative improvement* (Lin, 1975; Kirkpatrick et al., 1983). Because iterative improvement always finds the nearest local minimum, it is customary to perform several runs with different starting positions (i.e., different initial values for $\mathbf{s}$ and $\mathbf{r}$) and to save the best result. I ran 100 such tests of iterative improvement on the statics test data; the best of these 100 results is shown in Figure 8. The stack power for this solution is 2.378, which is almost 30% less than the result obtained by simulating annealing. (The worst of the 100 trials was short by almost 50%.) The diagonal appearance of this result is due to a severe cycle-skipping problem, and is the visual manifestation of convergence to a local minimum. These 100 runs consumed 4-5 times the computing time needed to obtain the annealing result of Figure 6c. Although iterative improvement is not an appropriate method to use when a statics problem is as severe as the one illustrated here, it is nonetheless a powerful technique for performing statics estimation when statics are sufficiently small enough so that there are few local minima (Ronen, 1984).

FUTURE APPLICATIONS

Several further applications of simulated annealing in reflection seismology are conceptually straightforward. I mention just a few; none have yet been implemented.

The residual statics algorithm can be easily extended to the estimation of frequency-dependent statics. The statics model discussed thus far uses simple linear phase shifts. In a frequency-dependent model, however, phase shifts can be a more general function of frequency.

The problem of frequency-dependent statics is difficult to solve with models similar to (19), because phase shifts greater than $|\pi|$ are computationally ambiguous (Sword, 1983). Some form of phase unwrapping (Tribolet, 1979) is usually thought to be necessary. In principle, adaptation of the present statics algorithm is straightforward and requires only the application of two elementary theorems from Fourier transform theory (Bracewell, 1978). The Rayleigh-Parseval theorem states that power in the time domain equals power in the frequency domain. So for a function $f(t)$ and its Fourier transform $F(\omega)$,

$$\sum_t \mid f(t) \mid^2 = \sum_\omega \mid F(\omega) \mid^2 .$$

In addition, the shift theorem states that the Fourier analog of a time shift is multiplication by a complex exponential:

$$f(t - \tau) \supset e^{i\omega\tau} F(\omega) .$$

Then by letting the Fourier transform of $d_{yh}(t)$ be denoted by $D_{yh}(\omega)$, we may include frequency dependence in equation (24a) by writing

$$\phi_{s_i} [\mathbf{s}(\omega), \mathbf{r}(\omega)] = \sum_{y \in Y_{s_i}} \sum_\omega \left| \sum_h e^{i\omega[s_{i(y,h)}(\omega) + r_{j(y,h)}(\omega)]} D_{yh}(\omega) \right|^2 . \qquad (26)$$

Similar changes can be made to equation (24b). Note that the s_i and r_j are now functions of ω.

Residual statics algorithms find only relative time shifts; the longest spatial wavelength (the d.c. component) is fully unresolved by the data. This problem of resolution will be exacerbated in the frequency-dependent case if each ω-component is treated independently of the others. A physical model of frequency-dependent phase shifts should therefore require that the phase shifts be locally correlated with each other. This may be incorporated into equation (26) by smoothing $\mathbf{s}(\omega)$ and $\mathbf{r}(\omega)$ over ω. Representing these smoothed functions by $\overline{\mathbf{s}}(\omega)$ and $\overline{\mathbf{r}}(\omega)$, the energy function for frequency-dependent statics is

$$E[\overline{\mathbf{s}}(\omega), \overline{\mathbf{r}}(\omega)] = -\sum_i \phi_{s_i} [\overline{\mathbf{s}}(\omega), \overline{\mathbf{r}}(\omega)] - \sum_j \phi_{r_j} [\overline{\mathbf{s}}(\omega), \overline{\mathbf{r}}(\omega)] .$$

Velocity inversion can also be viewed as an extension of the residual statics algorithm. Statics are essentially the components of a one-dimensional velocity function. In velocity inversion, a two-dimensional grid would be parameterized by velocity, and we would seek the velocity distribution yielding the maximum stack power. Toldi (1984) and Loinger (1983) present techniques in which perturbations to a velocity model are linearly related to initial estimates of interval velocity. Optimization by simulated annealing would be valuable if this assumption of linearity were not valid (i.e., if the necessary perturbations were too large). However, this nonlinear approach might be computationally unacceptable, because each perturbation and power calculation might require far more effort than the simple shifts and sums needed in residual statics.

Deconvolution techniques that are designed to optimize an objective function such as "spikiness" or "simplicity" usually require some form of nonlinear optimization. Examples of these approaches to deconvolution include minimum entropy deconvolution

(Wiggins, 1978) and its various generalizations (Donoho, 1981). Usually, iterative descent from an initial guess is employed to estimate the coefficients of the deconvolution filter. However, if local minima are a problem, simulated annealing could be a valuable alternative. Here, the model parameters would be the filter coefficients and the "neighborhoods" would encompass the entire filter.

REMAINING ISSUES

Several important questions remain to be answered. The notion of a *critical temperature* is perhaps the most important, and the least understood. In physics, a critical temperature can be the temperature at which a liquid changes to a solid, or the temperature at which a ferromagnetic substance acquires permanent magnetization. These examples of the spontaneous ordering of matter are called *phase transitions* and have been the object of extensive study [see, for example, Stanley (1971)]. For the nonlinear inverse problems discussed here, the critical temperature may be broadly defined to be the largest value of T that leads to significant (non-local) correlations between parameters. Convergence is possible only below this temperature. The critical temperature is presently estimated empirically. An analytic approximation remains an open research problem.

In the original formulation of Metropolis et al. (1953), the Monte Carlo algorithm was used to estimate the ergodic averages

$$< f(\mathbf{x}) > \;=\; \sum_{\mathbf{x}} f(\mathbf{x}) P(\mathbf{X}=\mathbf{x}) \;=\; \frac{\sum_{\mathbf{x}} f(\mathbf{x})\, e^{\frac{-E(\mathbf{x})}{T_1}}}{\sum_{\mathbf{x}} e^{\frac{-E(\mathbf{x})}{T_1}}} \;, \qquad (27)$$

for a given $T = T_1$. These averages are valid only if the system has reached *equilibrium,* which means that the Monte Carlo algorithm has performed enough iterations so that the $\mathbf{x}$ are generated with a Gibbs probability. Ideally, T would be lowered in simulated annealing only if equilibrium has been attained. Equilibrium is theoretically defined in terms of the equilibrium distributions of Markov chains (Hammersley and Handscomb, 1964; Fosdick, 1963), but it is notoriously difficult to quantify in empirical studies (Binder, 1979; Binder, 1984). If one is certain that equilibrium has been attained, however, then the generation of the ergodic averages (27) can be useful for estimating means, variances, covariances, etc. One can also estimate the posterior probability distribution by constructing a histogram of the output of each iteration for constant temperature. Thus one can obtain not only a simple answer (the maximum a posteriori solution) but also estimates of resolution and accuracy. The posterior probability distribution is arguably the most fundamental information that can be provided by a solution to an inverse problem (Tarantola and Valette, 1982).

Ultimately, the question of computational efficiency must be addressed. My current implementation is slow (the example in Figures 6a-c required 4 hours of CPU time on a DEC VAX 11-780). The results are encouraging, however, because this particular statics problem appears to be unsolvable by existing techniques. The *heat bath method* (Rebbi, 1984; Creutz, 1980), an adaptation of the Metropolis algorithm, is currently being investigated for possible computational gains. The most significant gains in computational speed might await the eventual development of new parallel computing architecture.

CONCLUSIONS

Conventional techniques for solving nonlinear inverse problems rely on initial guesses of model parameters and subsequent linearization. These techniques perform well if initial estimates contain sufficiently small errors, but they can otherwise fail severely.

The Gibbs-Markov model provides guidelines for the reduction of a large nonlinear inverse problem into small, interdependent, and computationally manageable subproblems. This formulation does not depend on a good initial guess, and leads naturally to global optimization by simulated annealing.

Residual statics estimation should be formulated as a nonlinear inverse problem when statics are large and data are contaminated by noise. The benefit of a nonlinear formulation is substantial: no initial estimates of timing delays are needed. Although poorly picked spatial correlations ("cycle-skips") appear as local minima, global optimization can be successfully performed by simulated annealing. Large statics can considerably alter the appearance of seismic data. Although the estimation of statics by simulated annealing is computationally expensive, the benefits gained by its more accurate solution may far outweigh the additional cost.

Efforts toward an application to field data are currently underway and will be reported shortly. Both the broad applicability of the technique and its encouraging early results bode a promising future for this new approach to nonlinear inversion.

ACKNOWLEDGMENTS

Fabio Rocca helped inspire a reevaluation of residual statics analysis. Jon Claerbout suggested the power criterion for residual statics, and Shuki Ronen encouraged its use. Stew Levin critically reviewed the manuscript. I also had many helpful discussions with Paul Fowler, Peter Mora, Dale Morgan, and John Toldi. Visits I made to Exxon Engineering Research Co. and Schlumberger-Doll Research Co. were especially informative; I thank Peng Sheng of Exxon and Mike Oristaglio of Schlumberger for arranging those trips. A discussion with Ralphe Wiggins led to my comments on minimum entropy deconvolution. This work was supported by the Amoco Foundation and the sponsors of the Stanford Exploration Project.

REFERENCES

Aki, K. and Richards, P., 1980, Quantitative seismology: San Francisco, W. H. Freeman and Co., 932 p.

Bard, Y., 1974, Nonlinear parameter estimation: New York, Academic Press, 341 p.

Binder, K., ed., 1984, Applications of the Monte Carlo method in statistical physics: New York, Springer-Verlag, 311 p.

Binder, K., ed., 1979, Monte Carlo methods in statistical physics: New York, Springer-Verlag, 376 p.

Bracewell, R., 1978, The Fourier transform and its applications: New York, McGraw-Hill Book Co., 444 p.

Creutz, M., 1980, Monte Carlo study of quantized SU(2) gauge theory: Phys. Rev. D, v. 21, p. 2308-2315.

Donoho, D.L., 1979, Estimation of time delay at poor S/N: Paper presented at the 1979 EAEG, Hamburg.

Donoho, D.L., 1981, On minimum entropy deconvolution: in Applied time series analysis II, D. Findley, ed., p. 565-608.

Fosdick, L.D., 1963, Monte Carlo computations on the Ising lattice: Methods in Computational Physics, v. 1, p. 245-280.

Geman, S. and Geman, D., 1984, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images: IEEE Transactions on Pattern Analysis and Machine Intelligence (in press).

Hammersley, J.M. and Handscomb, D.C., 1964, Monte Carlo methods: London, Chapman and Hall, 178 p.

Hinton, G. and Sejnowski, T., 1983, Optimal perceptual inference: Proceedings IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition, p. 448-453.

Hockney, R.W., and Jesshope, C.R., 1981, Parallel computers: Bristol, Adam Hilger Ltd, 423 p.

Kindermann, R. and Snell, J.L., 1980, Markov random fields and their applications: Providence, American Mathematical Society, 142 p.

Kirkpatrick, S., Gelatt, C.D., Jr., and Vecchi, M.P., 1983, Optimization by simulated annealing: Science, v. 220, p. 671-680.

Lin, S., 1975, Heuristic programming as an aid to network design: Networks, v. 5, p. 33-43.

Lines, L.R. and Treitel, S., 1984, Tutorial: a review of least-squares inversion and its application to geophysical problems: Geophysical Prospecting, v. 32, p. 159-186.

Loinger, E., 1983, A linear model for velocity anomalies: Geophysical Prospecting, v. 31, p. 98-118.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E., 1953, Equation of state calculations by fast computing machines: Journal of Chemical Physics, v. 21, p. 1087-1092.

Moussouris, J., 1974, Gibbs and Markov random systems with constraints: Journal of Statistical Physics, v. 10, p. 11-33.

Parker, R.L., 1977, Understanding inverse theory: Ann. Rev. Earth Planet. Sci., v. 5, p. 35-64.

Rebbi, C., 1984, Monte Carlo calculations in lattice gauge theories: in Applications of the Monte Carlo method, K. Binder, ed., p. 277-298.

Ronen, S., 1984, Surface-consistent residual statics by stack optimization: Stanford Exploration Project Report 38, p. 27-37.

Stanley, H.E., 1971, Introduction to phase transitions and critical phenomena: New York, Oxford University Press, 308 p.

Sword, C., 1983, The generalized frequency-dependent surface-consistent problem: Stanford Exploration Project Report 35, p. 19-42..

Taner, M.T., Koehler, F., and Alhilali, K.A., 1974, Estimation and correction of near-surface time anomalies: Geophysics, v. 39, p. 441-463.

Tarantola, A. and Valette, B., 1982, Inverse problems = Quest for information: Journal of Geophysics, v. 50, p. 159-170.

Toldi, J., 1984, Laterally variable interval velocities from stacking velocities: Paper presented at the 1984 EAEG, London.

Tribolet, J., 1979, Seismic applications of homomorphic signal processing: Englewood Cliffs, N.J., Prentice Hall, Inc., 195 p.

Vichniac, G.Y., 1984, Simulating physics with cellular automata: in Cellular automata: proceedings of an interdisciplinary workshop, Los Alamos, New Mexico, D. Farmer, T. Toffoli, and S. Wolfram, eds.

Wiggins, R., Larner, K., and Wisecup, D., 1976, Residual statics analysis as a general linear inverse problem: Geophysics, v. 41, p. 922-938.

Wiggins, R., 1978, Minimum entropy deconvolution: Geoexploration, v. 16, p.21-35.

Are you ready to enter Moscow State University?

Moscow State University is the most prestigious university in the USSR. Entrance is (officially, at least) by competitive written and oral examinations.

Candidates wishing to study geophysics are expected to take the following examinations: mathematics (written and oral), physics (oral), and Russian language and literature (written). Each written examination lasts four hours.

Below is a sample mathematics written exam for applicants who would like to be admitted as undergraduates in geophysics. You have four hours. Ready? Go!

Geophysics exam

1. Find all real solutions of the equation

$$(x+1)\sqrt{16x+17} = (x+1)(8x-23).$$

2. In the triangle ABC, angle BAC is a right angle, and the lengths of the sides AB and BC are 1 and 2, respectively. The bisector of angle ABC intersects the side AC at point L, while point Q is point where the medians of triangle ABC intersect. Which is larger: the length of BL or the length of BQ?

3. Solve the inequality

$$\log_{\sin\frac{\pi}{3}}(x^2 - 3x + 2) \geq 2.$$

4. Find all pairs of real numbers x and y, which satisfy the system of equations

$$\begin{cases} 3\sin 3x + \cos y = -4, \\ x + y = \dfrac{3\pi}{2}. \end{cases}$$

5. A boat travels along a river whose current flows at 5 km/hr (kilometers per hour). According to its schedule, on a trip from A to D it covers a distance of 15 km in 1 hour. More specifically, leaving point A at 12:00, it is scheduled along the way to stop at points B and C, located at distances 11 km and 13 km respectively from point A, at 12:20 and 12:40. It is known that if the boat were to go from point A to point D without stopping, at a constant velocity v (relative to the water), then the sum of the absolute values of the deviation from the scheduled arrivals at points B, C, and D would not exceed the time, reduced by half an hour, that the boat would need to travel 5 km at velocity v in still water. Which of the points, A or D, is located upstream?

6. The base of pyramid $SABC$ is composed of isoceles triangle ABC, which has sides AB and AC of length 1, while the cosine of angle BAC equals $\frac{4}{5}$. Edge SA is perpendicular to edges AB and AC, and angle BAC is twice as large as angle BSC. Inside the pyramid is a right circular cylinder whose generatrix is parallel to BC. What is the maximum possible surface area of the sides (not including the bases) of such a cylinder?

Answers in the next SEP report