Tomographic determination of velocity and depth in laterally varying media

T. N. Bishop*, K. P. Bube‡, R. T. Cutler§, R. T. Langan**,
P. L. Love**, J. R. Resnick**, R. T. Shuey**, D. A. Spindler**,
and H. W. Wyld‡‡

ABSTRACT

Estimation of reflector depth and seismic velocity from seismic reflection data can be formulated as a general inverse problem. The method used to solve this problem is similar to tomographic techniques in medical diagnosis and we refer to it as seismic reflection tomography.

Seismic tomography is formulated as an iterative Gauss-Newton algorithm that produces a velocity-depth model which minimizes the difference between traveltimes generated by tracing rays through the model and traveltimes measured from the data. The input to the process consists of traveltimes measured from selected events on unstacked seismic data and a first-guess velocity-depth model. Usually this first-guess model has velocities which are laterally constant and is usually based on nearby well information and/or an analysis of the stacked section. The final model generated by the tomographic method yields traveltimes from ray tracing which differ from the measured values in recorded data by approximately 5 ms root-mean-square.

The indeterminancy of the inversion and the associated nonuniqueness of the output model are both analyzed theoretically and tested numerically. It is found that certain aspects of the velocity field are poorly determined or undetermined.

This technique is applied to an example using real data where the presence of permafrost causes a near-surface lateral change in velocity. The permafrost is successfully imaged in the model output from tomography. In addition, depth estimates at the intersection of two lines differ by a significantly smaller amount than the corresponding estimates derived from conventional processing.

INTRODUCTION

Estimation of velocity and depth is often an important step in prospect evaluation in areas where lithology and structure undergo significant lateral change. Depth estimation is usually accomplished by converting zero-offset traveltimes, interpreted from a stacked section, to depth using a velocity field obtained from a normal-movement (NMO) analysis. In areas with complex lateral changes, a depth migration technique may be necessary to obtain the correct depth estimate (Larner et al., 1981). Both of these methods require an accurate representation of the root-mean-square (rms) velocity field. However, the stacking velocities used for such analyses can deviate significantly from rms velocities because analysis of stacking velocities assumes that the medium is laterally invariant and that traveltime trajectories for reflection events in CDP gathers are hyperbolic.

Media vary laterally due to either reflector dip or curvature, or due to lateral velocity variations, or both. A large portion of the effect of reflector dip or curvature on the stacking velocity can be removed approximately by first migrating common-offset panels with a first-guess velocity function, and then recalculating the stacking velocity in the migrated common-depth-point (CDP) gathers (Doherty and Claerbout, 1976). The influence of lateral variations in velocity on the stacking velocity cannot be corrected this way. For lateral variations in velocity whose wavelength is longer than a cable length (the maximum spread offset), the smoothed stacking velocity is usually a good representation of the rms velocity. Velocity variations on the scale of a cable length or shorter can produce large differences between the stacking velocity and the rms velocity. These differences are sometimes referred to as stacking-velocity anomalies (Pollet, 1974).

A conventional residual-statics approach can often correct for near-surface variations in velocity on a scale shorter than a cable length. However, the spatial resolution of statics analyses decreases rapidly beyond one cable length (Wiggins et al.,

1976). Another method was proposed in Lynn and Claerbout (1982) which inverts the observed stacking velocity and zero-offset traveltime, as a function of common midpoint, to obtain an approximate true-vertical rms velocity function. This method is limited to velocity variations longer and shallower than a cable length.

Here we present a method which uses traveltimes to selected events in unstacked seismic reflection data to image velocity anomalies above and between these events and to infer their depths. The depths are determined even in the presence of reflector dip or curvature and lateral changes in velocity. When thus determined, the depths are those that would be found from depth migration using the correct rms velocity field. With some limitations (to be discussed), the method is capable of imaging both short- and long-wavelength variations in velocity.

Conceptually, this new method is closely related to the iterative inversion method used in Hawley et al. (1981) for simultaneously determining hypocenters of earthquakes and three-dimensional (3-D) velocity variations. Their method is a generalization of the one-step general inverse method proposed in Aki and Lee (1976). The technique presented here differs from transmission problems (earthquake to receiver, see Anderson and Dziewonski, 1984) in that rays start at the surface, reflect off interfaces whose depths are to be determined, and return to the surface. Another major difference is the much larger number of model parameters in the reflection seismology problem due to the increased density of data. This difference makes this inversion method very computer-intensive. This method is also similar to work done by Kjartansson (1980). While his work also used reflected rays, it was limited to straight raypaths and fixed reflector depths.

All these methods are closely related to x-ray computerized tomography (CT) used in medical diagnosis. In CT the measured data can be modeled as line integrals through a discretized density field along a straight raypath. An efficient convolutional method exists for inverting these measurements to obtain the spatial distribution of the density. Recent geophysical applications of similar tomographic techniques have included imaging between boreholes for site characterization (Dines and Lytle, 1979) and mapping of large-scale changes in the density of the oceans (Cornuelle, 1982). The general inverse methods are computationally similar to CT, but they use traveltimes to image slowness as a discretized field. The tomography problem for seismic reflection data is complicated by the unknown depth to the reflectors, ray-bending effects which are nonlinear, and raypath coverage which is irregular and has limited view-angle coverage.

This paper discusses the following: the method of reflection tomography as a two-dimensional (2-D) general inverse problem; the inversion algorithm, illustrated using a synthetic example; the uniqueness of solutions to the inverse problem; and finally an example of velocity and depth determination for a seismic line where a large near-surface change in lateral velocity is caused by permafrost.

PROBLEM FORMULATION

First we establish notation, and then formulate the problem in a precise fashion. Let x be horizontal distance along the earth's surface and let z be depth. In the region of interest, there are n_r reflectors whose depths as a function of position on the

surface are denoted by $Z_1(x), Z_2(x), \ldots, Z_\rho(x), \ldots, Z_{n_r}(x)$. The slowness in the region is modeled by a function $w(x, z)$, which represents the reciprocal of seismic velocity at points (x, z) in the subsurface.

For computation, it is convenient to characterize the slowness and reflector depth functions by a finite set of parameters. Those functions are then restricted to lie in a certain finite-dimensional space of functions. The intent is to make the dimension of that space sufficiently high that functions which are not in the space, but may more accurately describe the real Earth, can be well approximated by functions which do lie in the space. However, the dimension must not be so high that the problem of inverting the available data is hopelessly indeterminate.

Consider first the slowness function $w(x, z)$. Divide the region of interest into a matrix of rectangular boxes having n_x columns and n_z rows (Figure 1). The horizontal dimension of a box is typically taken as four times the CDP spacing, with the vertical dimension roughly twice that. The value of $w(x, z)$ at the center of the box in the kth row and ℓth column is denoted $w_{k\ell}$. Elsewhere in the box the slowness is assumed to vary in such a fashion that the gradient of velocity within that box is constant. The value of slowness anywhere within the box can be obtained from an interpolation formula which depends on $w_{k\ell}$ and on $w_{k-1, \ell}, w_{k+1, \ell}, w_{k, \ell-1}, w_{k, \ell+1}$, that is, on the values of slowness at the centers of adjacent boxes. This procedure is motivated by the ray-tracing scheme. Rays are traced from source to receiver by a shooting method which uses the entrance angle at each box and the velocity gradient to compute the radius of curvature of an arc and the arclength in the box. This method is described in greater detail in Appendix A.

The reflector depth functions $Z_\rho(x)$ are parameterized by the depths at which the reflector intersects vertical boundaries of successive columns of boxes. For reflector ρ these depths are
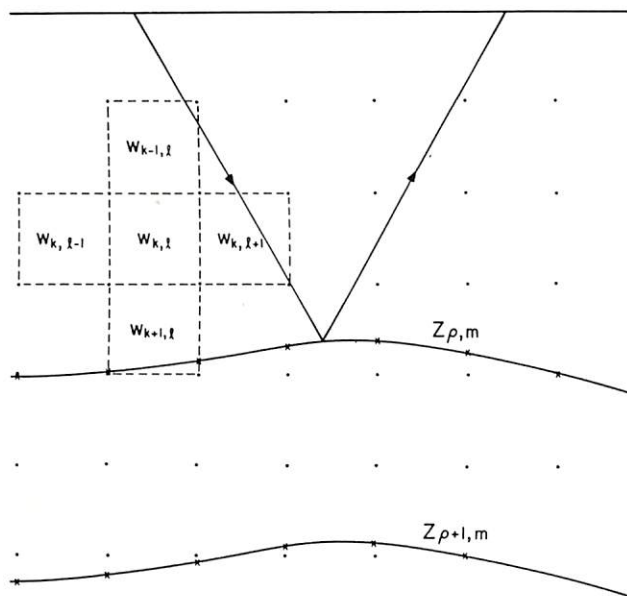

FIG. 1. Discretization of velocities and depths used in the model earth.

denoted $Z_{\rho m}$, $1 \le \rho \le n_r$, $0 \le m \le n_x$. Away from box boundaries, $Z_\rho(x)$ is regarded as a cubic spline function which interpolates the points specified by the $Z_{\rho m}$.

Let M be the total number of parameters which characterize this model of the earth. From the above discussion, it is clear that

$$M = n_x n_z + n_r(n_x + 1).$$

It will prove useful to define an M-dimensional vector $\mathbf{p}$ having the property that every parameter of the model appears as a component of $\mathbf{p}$. This can be done by reindexing the parameters: the first $n_x n_z$ components of $\mathbf{p}$ are the $w_{k\ell}$, and the last $n_r(n_x + 1)$ components are the $Z_{\rho m}$.

In addition to the vector $\mathbf{p}$, which is the primary descriptor of the model, some notation is needed to describe the data set of reflection tomography. The data consist of two-way traveltime measurements obtained from seismic recordings made on the surface. Let there be n_s shots with a maximum of n_g geophone station locations per shot. The observed traveltime for a ray which emerges from source μ, travels to reflector ρ, and returns to receiver ν is represented as $T_{\mu\rho\nu}$, $1 \le \mu \le n_s$, $1 \le \rho \le n_r$, $1 \le \nu \le n_g$. (In practice, every reflector will not necessarily yield reflections on every geophone trace of sufficient amplitude to facilitate the picking of event traveltimes; for such reflectors, the geophone index will remain below n_g in some shot gathers.) The total number of rays is denoted N. As was the case with the parameters, it is convenient to renumber the data so the traveltime for each ray appears as an entry in a single N-component vector. This vector of traveltime measurements is called $\mathbf{t_d}$.

For a particular choice of the parameter vector $\mathbf{p}$, rays can be traced through the model to yield a collection of traveltimes which correspond to the same shot-receiver pairs that are represented in the data. Indexing these traveltimes in a manner analogous to that of traveltimes from the seismic line yields $\mathbf{t(p)}$, an N-component vector function of $\mathbf{p}$.

The goal of reflection tomography is to find a model for which the model traveltimes closely match the real traveltimes. In more mathematical language, the problem can be posed as follows. Let the residual vector $\mathbf{r(p)}$ be defined by $\mathbf{r(p)} = \mathbf{t_d} - \mathbf{t(p)}$. Double bars ($\|\cdot\|$) denote the Euclidean norm of a vector and prime denotes the transpose. The objective of tomography is to choose $\mathbf{p}$ such that the quantity $\|\mathbf{r(p)}\|$ is minimized. The solution of this problem is identical to that of minimizing the mathematically more convenient function

$$\phi(\mathbf{p}) = \|\mathbf{r(p)}\|^2 = [\mathbf{t_d} - \mathbf{t(p)}]'[\mathbf{t_d} - \mathbf{t(p)}]. \quad (1)$$

The description of an algorithm to minimize $\phi(\mathbf{p})$ follows.

INVERSION ALGORITHM

To look for a parameter vector which minimizes $\phi(\mathbf{p})$, we have used an algorithm which has been used previously in geophysical applications (Hawley et al., 1981). The procedure is an iterative one which starts with an estimate $\mathbf{p}^{(k)}$ of the solution and then generates a new estimate $\mathbf{p}^{(k+1)}$ based on a linearized approximation to $\mathbf{t(p)}$ at $\mathbf{p}^{(k)}$. In the mathematics literature, this approach is known as the Gauss-Newton method (Marquardt, 1963; Gill et al., 1981).

A necessary condition for a local minimum of $\phi(\mathbf{p})$ is that at the minimum,

$$\nabla\phi(\mathbf{p}) = 0. \quad (2)$$

If $\underline{\mathbf{A}}(\mathbf{p})$ is the Jacobian matrix of $\mathbf{t(p)}$, that is, the matrix whose ijth element is $\partial t_i/\partial p_j$, then equation (2) implies

$$\underline{\mathbf{A}}'(\mathbf{p})[\mathbf{t(p)} - \mathbf{t_d}] = 0. \quad (3)$$

The algorithm actually attempts to find a solution of equation (3). To understand how the procedure works, suppose $\mathbf{p}^{(k)}$ is known and it is desired to find $\mathbf{p}^{(k+1)}$. Let $\Delta\mathbf{p} = \mathbf{p}^{(k+1)} - \mathbf{p}^{(k)}$, $\underline{\mathbf{A}}^{(k)} = \underline{\mathbf{A}}(\mathbf{p}^{(k)})$, and $\mathbf{r}^{(k)} = \mathbf{r}(\mathbf{p}^{(k)})$. Consider the quantity $\underline{\mathbf{A}}'^{(k)}\mathbf{t}(\mathbf{p}^{(k+1)})$. If $\mathbf{t}(\mathbf{p}^{(k+1)})$ is expanded in a Taylor series about $\mathbf{p}^{(k)}$, we have

$$\underline{\mathbf{A}}'^{(k)}[\mathbf{t}(\mathbf{p}^{(k+1)}) - \mathbf{t_d}] = \underline{\mathbf{A}}'^{(k)}\underline{\mathbf{A}}^{(k)}\Delta\mathbf{p} - \underline{\mathbf{A}}'^{(k)}\mathbf{r}^{(k)}$$
$$+ 0(\|\Delta\mathbf{p}\|^2). \quad (4)$$

The vector $\Delta\mathbf{p}$ is chosen so that, to first order, the right-hand side of equation (4) is zero. Since typically $\underline{\mathbf{A}}'^{(k)}\underline{\mathbf{A}}^{(k)}$ is either not invertible or is, at best, very poorly conditioned, there are many vectors $\Delta\mathbf{p}$ which can play such a role. The algorithm chooses $\Delta\mathbf{p}$ such that

$$[\underline{\mathbf{A}}'^{(k)}\underline{\mathbf{A}}^{(k)} + \alpha(\underline{\mathbf{W}}^{(k)})^2]\Delta\mathbf{p} = \underline{\mathbf{A}}'^{(k)}\mathbf{r}^{(k)}, \quad (5)$$

where α is a real constant and $\underline{\mathbf{W}}^{(k)}$ a weighting matrix which will be discussed below. With $\Delta\mathbf{p}$ chosen, $\mathbf{p}^{(k+1)}$ is then simply

$$\mathbf{p}^{(k+1)} = \mathbf{p}^{(k)} + \Delta\mathbf{p}. \quad (6)$$

Suppose the algorithm outlined in equations (5) and (6) converges. Call the vector to which $\mathbf{p}^{(k)}$ converges $\mathbf{p}^*$. Then $\mathbf{p}^*$ must be a solution of equation (3). This follows since $\mathbf{p}^{(k)} \to \mathbf{p}^*$ implies $\Delta\mathbf{p} \to 0$. So from equation (5),

$$\underline{\mathbf{A}}'(\mathbf{p}^*)\mathbf{r}(\mathbf{p}^*) = \lim_{k \to \infty} \underline{\mathbf{A}}'^{(k)}\mathbf{r}^{(k)} = 0.$$

The vector $\mathbf{p}^*$ is a solution regardless of how α or $\underline{\mathbf{W}}^{(k)}$ are chosen, although both α and $\underline{\mathbf{W}}^{(k)}$ can affect the rate of convergence. However, if many solutions of equation (3) exist, the values of α and $\underline{\mathbf{W}}^{(k)}$, as well as the choice of the initial model $\mathbf{p}^{(0)}$, determine which solution is obtained. Presumably all three terms may also influence whether or not convergence occurs. If the algorithm does converge, $\mathbf{p}^*$ satisfies the necessary condition [equation (3)] for a local minimum, but there is no a priori guarantee that it represents a global minimum.

In numerical tests it is observed that the residual decreases for several (between three and five) Gauss-Newton steps; after about five steps, changes in its value are sufficiently slow that further iteration seems pointless. A useful measure of the algorithm's effectiveness is the quantity

$$E = (\|r^{(k)}\|^2/N)^{1/2}, \quad (7)$$

which represents the rms difference between the traveltimes predicted by the model and those found in the data. Although E may be as large as 50 ms initially, by the third or fourth iteration it is often reduced to 5 ms, close to the sampling rate of standard seismic data.

In the scheme outlined above, the Jacobian plays a dominant role. A single row of the matrix $\underline{\mathbf{A}}^{(k)}$ consists of the derivatives of the traveltime for a particular ray with respect to each of the parameters in the model. The first $n_x n_z$ of these terms are slowness derivatives, and the rest are depth derivatives. Thus, the matrix $\underline{\mathbf{A}}^{(k)}$ can be partitioned into two matrices $\underline{\mathbf{A}}_w^{(k)}$ and $\underline{\mathbf{A}}_z^{(k)}$

denoted $Z_{\rho m}$, $1 \leq \rho \leq n_r$, $0 \leq m \leq n_x$. Away from box boundaries, $Z_\rho(x)$ is regarded as a cubic spline function which interpolates the points specified by the $Z_{\rho m}$.

Let M be the total number of parameters which characterize this model of the earth. From the above discussion, it is clear that

$$M = n_x n_z + n_r(n_x + 1).$$

It will prove useful to define an M-dimensional vector $\mathbf{p}$ having the property that every parameter of the model appears as a component of $\mathbf{p}$. This can be done by reindexing the parameters: the first $n_x n_z$ components of $\mathbf{p}$ are the $w_{k\ell}$, and the last $n_r(n_x + 1)$ components are the $Z_{\rho m}$.

In addition to the vector $\mathbf{p}$, which is the primary descriptor of the model, some notation is needed to describe the data set of reflection tomography. The data consist of two-way traveltime measurements obtained from seismic recordings made on the surface. Let there be n_s shots with a maximum of n_g geophone station locations per shot. The observed traveltime for a ray which emerges from source μ, travels to reflector ρ, and returns to receiver ν is represented as $T_{\mu\rho\nu}$, $1 \leq \mu \leq n_s$, $1 \leq \rho \leq n_r$, $1 \leq \nu \leq n_g$. (In practice, every reflector will not necessarily yield reflections on every geophone trace of sufficient amplitude to facilitate the picking of event traveltimes; for such reflectors, the geophone index will remain below n_g in some shot gathers.) The total number of rays is denoted N. As was the case with the parameters, it is convenient to renumber the data so the traveltime for each ray appears as an entry in a single N-component vector. This vector of traveltime measurements is called $\mathbf{t_d}$.

For a particular choice of the parameter vector $\mathbf{p}$, rays can be traced through the model to yield a collection of traveltimes which correspond to the same shot-receiver pairs that are represented in the data. Indexing these traveltimes in a manner analogous to that of traveltimes from the seismic line yields $\mathbf{t(p)}$, an N-component vector function of $\mathbf{p}$.

The goal of reflection tomography is to find a model for which the model traveltimes closely match the real traveltimes. In more mathematical language, the problem can be posed as follows. Let the residual vector $\mathbf{r(p)}$ be defined by $\mathbf{r(p)} = \mathbf{t_d} - \mathbf{t(p)}$. Double bars ($\|\cdot\|$) denote the Euclidean norm of a vector and prime denotes the transpose. The objective of tomography is to choose $\mathbf{p}$ such that the quantity $\|\mathbf{r(p)}\|$ is minimized. The solution of this problem is identical to that of minimizing the mathematically more convenient function

$$\phi(\mathbf{p}) = \|\mathbf{r(p)}\|^2 = [\mathbf{t_d} - \mathbf{t(p)}]'[\mathbf{t_d} - \mathbf{t(p)}]. \tag{1}$$

The description of an algorithm to minimize $\phi(\mathbf{p})$ follows.

INVERSION ALGORITHM

To look for a parameter vector which minimizes $\phi(\mathbf{p})$, we have used an algorithm which has been used previously in geophysical applications (Hawley et al., 1981). The procedure is an iterative one which starts with an estimate $\mathbf{p}^{(k)}$ of the solution and then generates a new estimate $\mathbf{p}^{(k+1)}$ based on a linearized approximation to $\mathbf{t(p)}$ at $\mathbf{p}^{(k)}$. In the mathematics literature, this approach is known as the Gauss-Newton method (Marquardt, 1963; Gill et al., 1981).

A necessary condition for a local minimum of $\phi(\mathbf{p})$ is that at the minimum,

$$\nabla\phi(\mathbf{p}) = 0. \tag{2}$$

If $\underline{\mathbf{A}}(\mathbf{p})$ is the Jacobian matrix of $\mathbf{t(p)}$, that is, the matrix whose ijth element is $\partial t_i / \partial p_j$, then equation (2) implies

$$\underline{\mathbf{A}}'(\mathbf{p})[\mathbf{t(p)} - \mathbf{t_d}] = 0. \tag{3}$$

The algorithm actually attempts to find a solution of equation (3). To understand how the procedure works, suppose $\mathbf{p}^{(k)}$ is known and it is desired to find $\mathbf{p}^{(k+1)}$. Let $\Delta\mathbf{p} = \mathbf{p}^{(k+1)} - \mathbf{p}^{(k)}$, $\underline{\mathbf{A}}^{(k)} = \underline{\mathbf{A}}(\mathbf{p}^{(k)})$, and $\mathbf{r}^{(k)} = \mathbf{r}(\mathbf{p}^{(k)})$. Consider the quantity $\underline{\mathbf{A}}'^{(k)}\mathbf{t}(\mathbf{p}^{(k+1)})$. If $\mathbf{t}(\mathbf{p}^{(k+1)})$ is expanded in a Taylor series about $\mathbf{p}^{(k)}$, we have

$$\underline{\mathbf{A}}'^{(k)}[\mathbf{t}(\mathbf{p}^{(k+1)}) - \mathbf{t_d}] = \underline{\mathbf{A}}'^{(k)}\underline{\mathbf{A}}^{(k)}\Delta\mathbf{p} - \underline{\mathbf{A}}'^{(k)}\mathbf{r}^{(k)}$$
$$+ O(\|\Delta\mathbf{p}\|^2). \tag{4}$$

The vector $\Delta\mathbf{p}$ is chosen so that, to first order, the right-hand side of equation (4) is zero. Since typically $\underline{\mathbf{A}}'^{(k)}\underline{\mathbf{A}}^{(k)}$ is either not invertible or is, at best, very poorly conditioned, there are many vectors $\Delta\mathbf{p}$ which can play such a role. The algorithm chooses $\Delta\mathbf{p}$ such that

$$[\underline{\mathbf{A}}'^{(k)}\underline{\mathbf{A}}^{(k)} + \alpha(\underline{\mathbf{W}}^{(k)})^2]\Delta\mathbf{p} = \underline{\mathbf{A}}'^{(k)}\mathbf{r}^{(k)}, \tag{5}$$

where α is a real constant and $\underline{\mathbf{W}}^{(k)}$ a weighting matrix which will be discussed below. With $\Delta\mathbf{p}$ chosen, $\mathbf{p}^{(k+1)}$ is then simply

$$\mathbf{p}^{(k+1)} = \mathbf{p}^{(k)} + \Delta\mathbf{p}. \tag{6}$$

Suppose the algorithm outlined in equations (5) and (6) converges. Call the vector to which $\mathbf{p}^{(k)}$ converges $\mathbf{p}^*$. Then $\mathbf{p}^*$ must be a solution of equation (3). This follows since $\mathbf{p}^{(k)} \rightarrow \mathbf{p}^*$ implies $\Delta\mathbf{p} \rightarrow 0$. So from equation (5),

$$\underline{\mathbf{A}}'(\mathbf{p}^*)\mathbf{r}(\mathbf{p}^*) = \lim_{k \rightarrow \infty} \underline{\mathbf{A}}'^{(k)}\mathbf{r}^{(k)} = 0.$$

The vector $\mathbf{p}^*$ is a solution regardless of how α or $\underline{\mathbf{W}}^{(k)}$ are chosen, although both α and $\underline{\mathbf{W}}^{(k)}$ can affect the rate of convergence. However, if many solutions of equation (3) exist, the values of α and $\underline{\mathbf{W}}^{(k)}$, as well as the choice of the initial model $\mathbf{p}^{(0)}$, determine which solution is obtained. Presumably all three terms may also influence whether or not convergence occurs. If the algorithm does converge, $\mathbf{p}^*$ satisfies the necessary condition [equation (3)] for a local minimum, but there is no a priori guarantee that it represents a global minimum.

In numerical tests it is observed that the residual decreases for several (between three and five) Gauss-Newton steps; after about five steps, changes in its value are sufficiently slow that further iteration seems pointless. A useful measure of the algorithm's effectiveness is the quantity

$$E = (\|r^{(k)}\|^2/N)^{1/2}, \tag{7}$$

which represents the rms difference between the traveltimes predicted by the model and those found in the data. Although E may be as large as 50 ms initially, by the third or fourth iteration it is often reduced to 5 ms, close to the sampling rate of standard seismic data.

In the scheme outlined above, the Jacobian plays a dominant role. A single row of the matrix $\underline{\mathbf{A}}^{(k)}$ consists of the derivatives of the traveltime for a particular ray with respect to each of the parameters in the model. The first $n_x n_z$ of these terms are slowness derivatives, and the rest are depth derivatives. Thus, the matrix $\underline{\mathbf{A}}^{(k)}$ can be partitioned into two matrices $\underline{\mathbf{A}}_w^{(k)}$ and $\underline{\mathbf{A}}_z^{(k)}$

$$\underline{A}^{(k)} = (\underline{A}_w^{(k)} \mid \underline{A}_z^{(k)}),$$

where $\underline{A}_w^{(k)}$ is $N \times n_x n_z$, and $\underline{A}_z^{(k)}$ is $N \times n_r(n_x + 1)$.

How are the entries in $\underline{A}^{(k)}$ calculated? First, since one ray passes through relatively few boxes, many of the entries in a row of $\underline{A}^{(k)}$ are zero, and consequently $\underline{A}^{(k)}$ is quite sparse. Moreover, an argument based on Fermat's principle of least time (see Appendix B) shows that the derivative with respect to slowness in a particular box is approximately the pathlength in that box. The depth derivative can also be obtained in closed form involving terms readily available from the ray tracing (see Appendix B).

The weighting matrix $\underline{W}^{(k)}$, which serves to stabilize the computation of Δp in equation (5), has been discussed by others in a variety of contexts (Franklin, 1970; Wiggins, 1972). In the present application it has the form

$$\underline{W}^{(k)} = \begin{bmatrix} W_w^{(k)} & & & & & \\ & \ddots & & 0 & & \\ & & W_w^{(k)} & & & \\ & & & W_z^{(k)} & & \\ 0 & & & & \ddots & \\ & & & & & W_z^{(k)} \end{bmatrix},$$

where $W_w^{(k)} = (\text{trace } \underline{A}_w'^{(k)} \underline{A}_w^{(k)} / N)^{1/2}$ and $W_z^{(k)} = (\text{trace } \underline{A}_z'^{(k)} \underline{A}_z^{(k)} / N)^{1/2}$. This choice serves to ensure that the unitless parameter α will have a value near 1.0 at the point at which it begins to affect significantly the solution of equation (5).

Despite the relative sparsity of $\underline{A}^{(k)}$, the large values of N and M typical of seismic data ($N \sim 4 \times 10^4$, $M \sim 2 \times 10^3$) make the determination of Δp in equation (5) a challenging task. The numerical technique employed for inferrring Δp consists of a Gauss-Seidel algorithm with successive overrelaxation (Varga, 1962). A considerable effort has been expended in optimizing

the code for use with an array processor attached to a UNIVAC 1100/84 mainframe.

To illustrate some of the ideas described, we present a computational example of reflection tomography. The example consists of a simulation designed to demonstrate the effects of a near-surface velocity anomaly. The simulated earth is shown in Figure 2. The simulated region contains flat reflectors and laterally homogeneous velocity, except in the uppermost layer, where the velocity changes laterally from 7 272 ft/s to 8 890 ft/s. The parameters of this simulated earth are denoted by p_{target}. Each box here is 880 ft long and 1 000 ft deep, with the total number of parameters $M = 2\,203$. (Also $n_x = 110$, $n_r = 3$, and $n_z = 17$.) The simulated data are obtained by tracing $N = 27\,300$ rays through this model. A single-shot gather of rays striking one of the reflectors is shown in Figure 3.

Five Gauss-Newton inversion steps are employed in this demonstration. The initial guess $p^{(0)}$, known as the "gray" model, is similar to p_{target} except that $p^{(0)}$ contains no lateral velocity anomaly. Figure 4 displays the difference between the target slowness structure and the gray-model slowness structure. Also shown are the reflector depths, which are identical for both p_{target} and $p^{(0)}$. The difference between the slowness distributions of $p^{(5)}$ and $p^{(0)}$ appears in Figure 5. Figure 6 exhibits the depth of the first reflector on an enlarged scale after one, three, and five Gauss-Newton steps, respectively.

If this procedure worked perfectly, Figures 4 and 5 would be identical, and the depth corresponding to $p^{(5)}$ in Figure 6 would simply be constant at 5 000 ft. It is clear that in the solution there is some coupling between the velocity anomaly and the depth. Nonetheless, it is significant that the depth has been correctly determined to within 40 ft. That the depth is determined relatively accurately is shown by analysis of what is "well-determined" and what is "poorly determined" in the solution of equation (5).
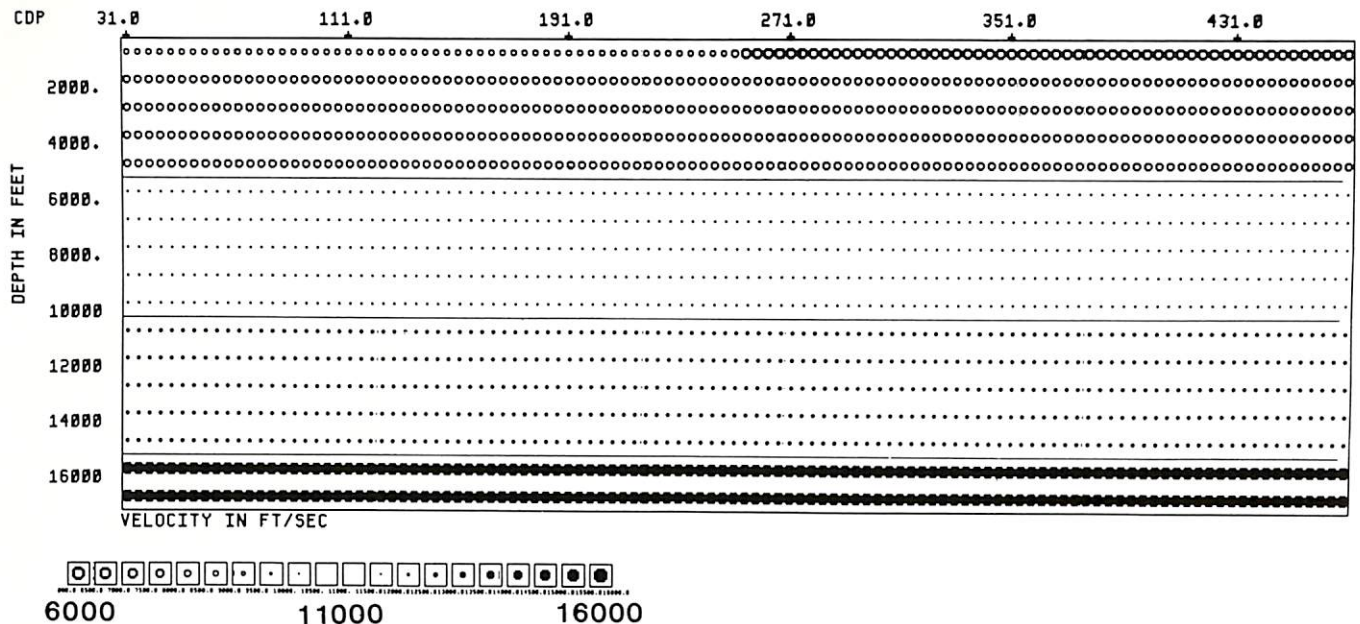

FIG. 2. Simulated earth model. Flat reflectors at 5 000, 10 000, and 15 000 ft. Velocities are laterally invariant except in the top layer.

THE EFFECTS OF LIMITED-ANGLE RAYS— UNDETERMINED AND POORLY DETERMINED MODEL FEATURES

There are limitations to the information about the subsurface that can be inferred from a particular set of traveltime measurements, even if such measurements are exact. These limitations are analyzed here for two problems, both of which are simpler than the actual tomographic inversion specified above. Based on experience from several numerical tests, it is found that the limitations described by these simpler analyses accurately characterize the results of the more general tomographic algorithm.

The vector $\Delta\mathbf{p}$ which is the solution of equation (5) is also the vector that minimizes the quantity

$$\| \underline{\mathbf{A}}^{(k)}\Delta\mathbf{p} - \mathbf{r}^{(k)} \|^2 + \alpha \| \underline{\mathbf{W}}^{(k)}\Delta\mathbf{p} \|^2. \tag{8}$$

The first term of expression (8) represents a linearized approximation to $\|\mathbf{r}^{(k+1)}\|^2$. Typically, there are many other choices of $\Delta\mathbf{p}$ for which this first term is of comparable magnitude. It is known that all of these vectors differ from each other by linear combinations of those eigenvectors of $\underline{\mathbf{A}}^{(k)'}\underline{\mathbf{A}}^{(k)}$ that correspond to small eigenvalues (Wiggins et al., 1976; Franklin, 1970). This includes all elements in the null space of $\underline{\mathbf{A}}^{(k)}$ for which the eigenvalues are zero, as well as a number of other components
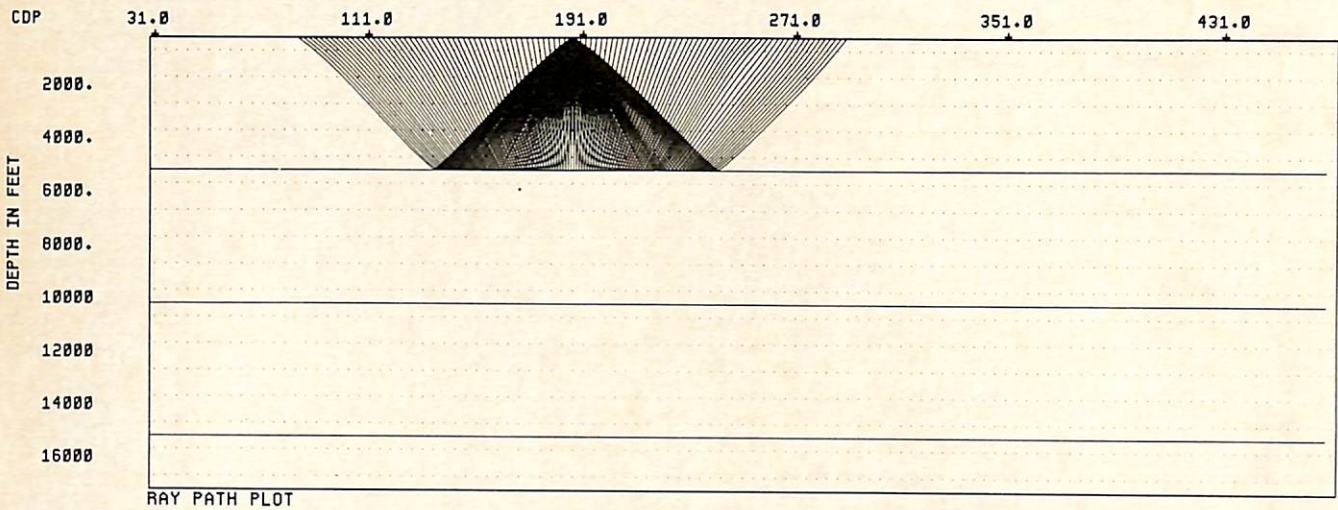

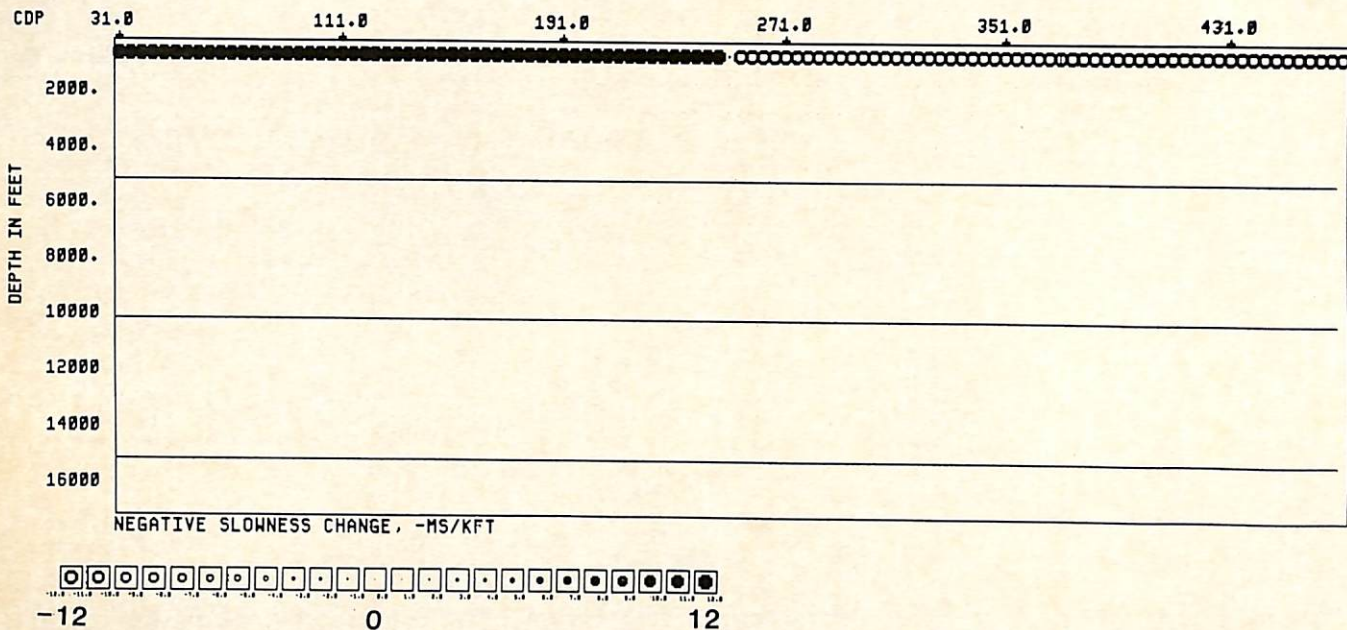FIG. 3. Raypaths from one shot to the top reflector traced through the model shown in Figure 2.

FIG. 4. Difference in slownesses between the simulated earth model shown in Figure 2 and "gray" model for tomography.

which in theory can be found by computing the "singular value decomposition" (SVD) of $\underline{A}^{(k)}$ (Golub and Reinsch, 1970; Forsythe, et al., 1977). Unfortunately, the SVD of a matrix as large as $\underline{A}^{(k)}$ is an extremely expensive computation. In what follows, although two variations on this general theme are pursued, SVD analysis of the large matrix $\underline{A}^{(k)}$ is avoided.

In the first approach, after certain simplifying assumptions are made, an exact specification is given of those parameters of the model earth which cannot be determined from the data.

This approach reveals the undetermined components in the solution of equation (3), the original, nonlinear problem. In the second approach, the most restrictive of the simplifying assumptions is relaxed, and a small model earth (151 parameters and 499 rays) is constructed for which an SVD analysis is feasible. The SVD analysis shows only those ambiguities that appear in the solution of the first Gauss-Newton step [equation (5)], but not necessarily in the full nonlinear problem. However, in addition to revealing those components which are impossible
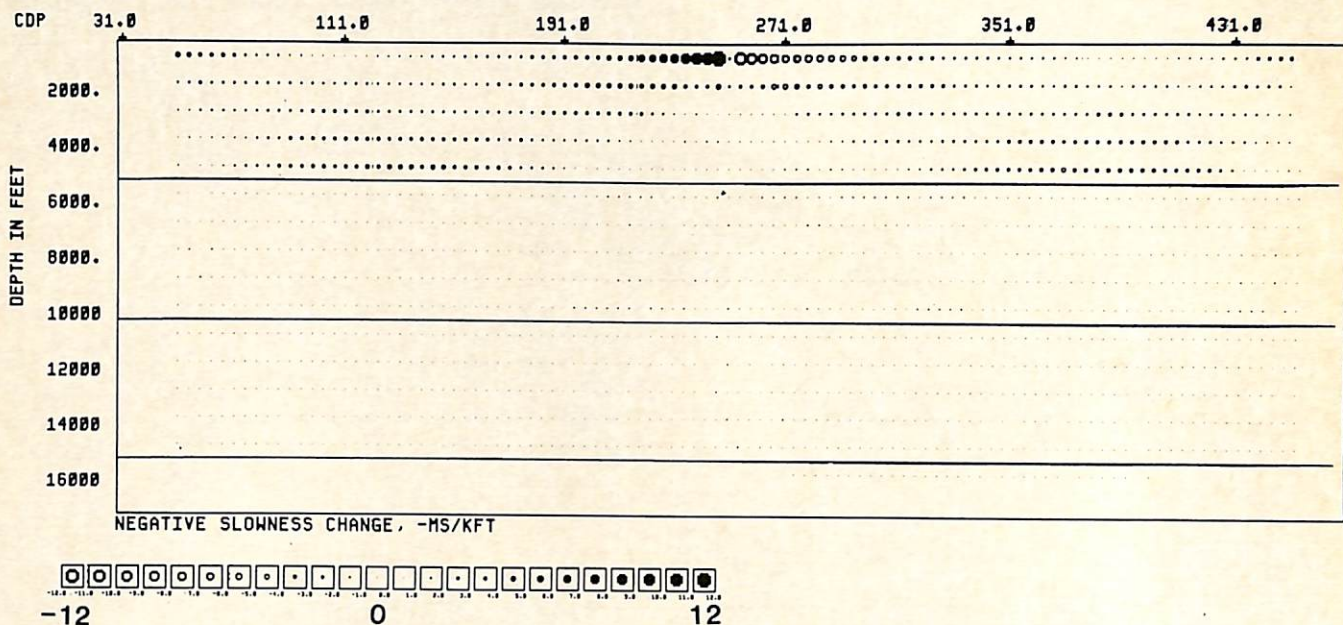

FIG. 5. Difference between the simulated earth model and output model after five Gauss-Newton steps.
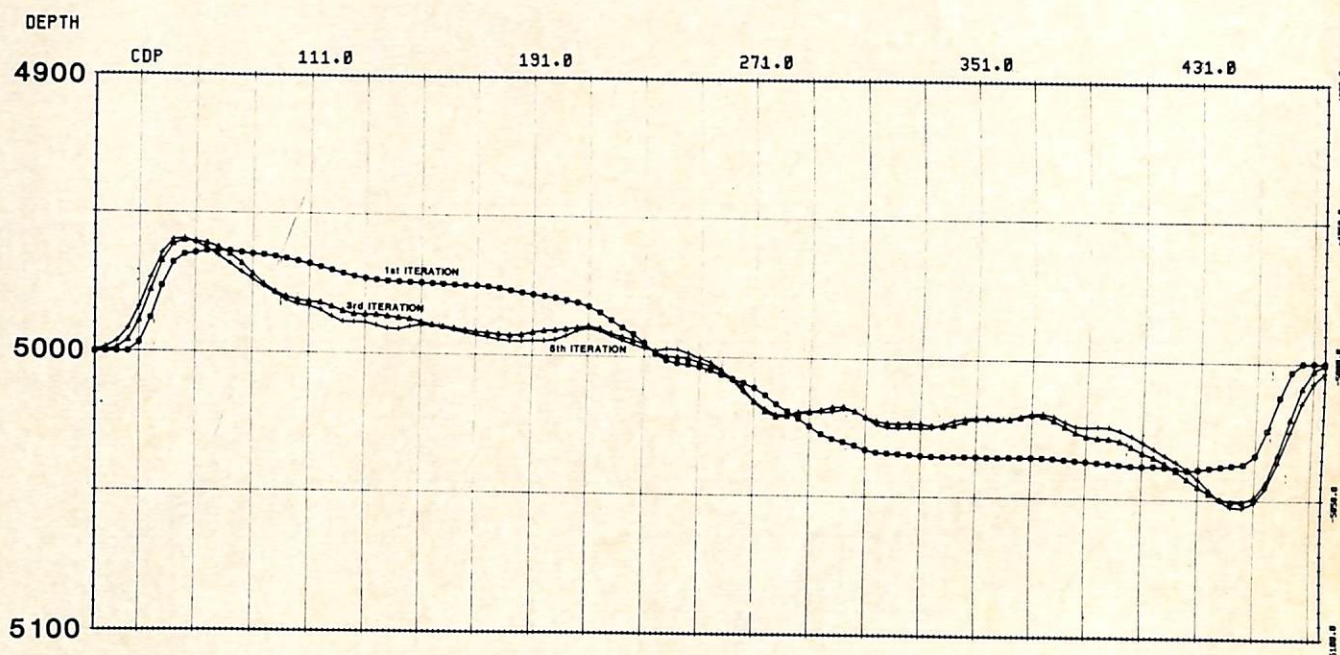

FIG. 6. Depth to the first reflector after one, three, and five Gauss-Newton steps. "Correct" answer would be flat lines at 5 000 ft.

to infer from the data, SVD also indicates which components are well-determined and which are poorly determined.

In both analyses, it is assumed that rays are straight, except for an equal-angle reflection at one of the reflectors. It is further assumed that the subsurface is discretized in boxes, with a constant velocity in each box. In the first analysis, it is also assumed that reflectors are flat and horizontal, so that only one parameter is needed to characterize each of n_r reflector depths. In addition, it is supposed that at least two distinct rays exist which are completely contained in one vertical column of boxes. Under these restrictions, the following results are proven in Appendix C.

(1) The depths of the flat and horizontal reflectors can be determined.
(2) Most of the velocity structure can be determined. Average slowness between reflectors can be determined. Lateral changes in slowness, including the vertical location of these changes, can be determined.
(3) Part of the velocity structure cannot be determined. Perturbations in slowness which are laterally constant and have zero-mean slowness between reflectors have no effect on the traveltime data and cannot be determined. Assuming that each box has at most one reflector in it, then the null space has dimension $n_z - n_r$. Recall that n_z is the number of horizontal layers of boxes.

Several of the assumptions made in this first analysis are rather restrictive. The SVD computations described next avoid some of these assumptions. Specifically, reflector depth is taken to be a continuous piece-wise linear function with bends at box boundaries, and a more realistic distribution of rays is considered. The points below summarize the results of a number of SVD calculations for a range of model choices.

There are singular engenfunctions with zero eigenvalues corresponding to

(1) the $n_z - n_r$ slowness perturbations described, and
(2) edge effects in slowness and reflector depths where there is an absence of rays or there are very few rays.

There are singular eigenfunctions with moderately small eigenvalues corresponding to

(3) more edge effects,
(4) linear variations in slowness in rows of boxes with zero vertical mean between reflectors, and
(5) sloping slowness compensating for sloping reflectors.

Small singular values corresponding to point (4) suggest that these variations in slowness are poorly determined. Similar but slightly larger eigenvalues associated with point (5) suggest a tradeoff between event dip across the whole model and a linear variation in slowness. This effect is seen in the example presented in the "Inversion algorithm" section. Figure 6 shows an inexact solution for depth which is characterized by a residual dip in the reflector of about 40 ft. Although each nonlinear pass of the inversion improves the depth estimate, the algorithm is converging to a depth function with a residual error in the slope of the reflector.

In the model examples considered, all the other singular values are an order of magnitude larger than the values detailed above, indicating that all other eigenfunctions of parameters are relatively well determined. This relationship has also been observed empirically by comparing target models with final inversion solutions.

In summary, we found that the Jacobian matrix has a null space corresponding to slowness variations in rows of boxes with zero vertical mean. The null space also contains some edge effects. Poorly determined quantities include linear slowness variations in rows with zero vertical mean and reflector tilt with a compensating linear slowness variation.

The above results are derived for a much simpler inversion problem than that encountered in the actual tomography algorithm. It is probable that the ray-bending and the discretization using a constant velocity gradient, which characterize the real algorithm, add information to the problem which is not taken into account in the simple analyses. This additional information may change an "undetermined" quantity into a "poorly determined" quantity, but it seems unlikely to alter the general picture which these studies establish of what can and what cannot be inferred from tomography. Indeed, numerical tests of the actual algorithm indicate that the indeterminancies described here do obtain in practice.

CONVENTIONAL AND TOMOGRAPHIC PROCESSING OF A SEISMIC LINE OVER A PERMAFROST BODY

We now compare conventional and tomographic processing of a seismic line. The line exhibits an abrupt lateral change in velocity due to permafrost. Proper depth interpretation was an important exploration problem on this line.

Figure 7 is a seismic cross-section (line A) from the Beaufort Sea. Shown in red is an interpretation of five major reflection events, labeled E1 through E5, which extend across most of the section. Near-surface reflection anomalies between CDPs 80 and 125 are due to permafrost. A time pull-up of events E1 and E2 is observed under the permafrost at the left end of the figure indicating a high-velocity, near-surface feature.

Further evidence for this interpretation comes from the stacking velocity analysis shown in Figure 8. This figure is a contour plot of stacking velocities after smoothing with a 22-CDP filter whose weights are given by a triangle function. There is a long-wavelength lateral gradient in velocity with the left end of the line faster than the right. The gradient is very strong at the surface and decreases to a rather small effect below 2.0 s. Superimposed on this gradient is a stacking-velocity anomaly from CDP 60 to CDP 190. The center of the anomaly is at CDP 125, corresponding to the edge of the shallow permafrost feature. The sense of the anomaly—high-velocity swing to the right and low-velocity swing to the left of the edge of the permafrost—corresponds to an abrupt lateral change to lower velocity from left to right.

The influence of an abrupt lateral velocity change on the traveltime and the stacking velocity for a reflection event below the permafrost is depicted schematically in Figure 9. In this figure the traveltime effect of the high-velocity layer is approximated by a static speed up. Note that the influence of the anomaly extends for a distance equal to a cable length centered on the edge of the layer.

To illustrate the effect of the stacking-velocity anomaly on depth conversion, the velocities shown in Figure 8 can be used
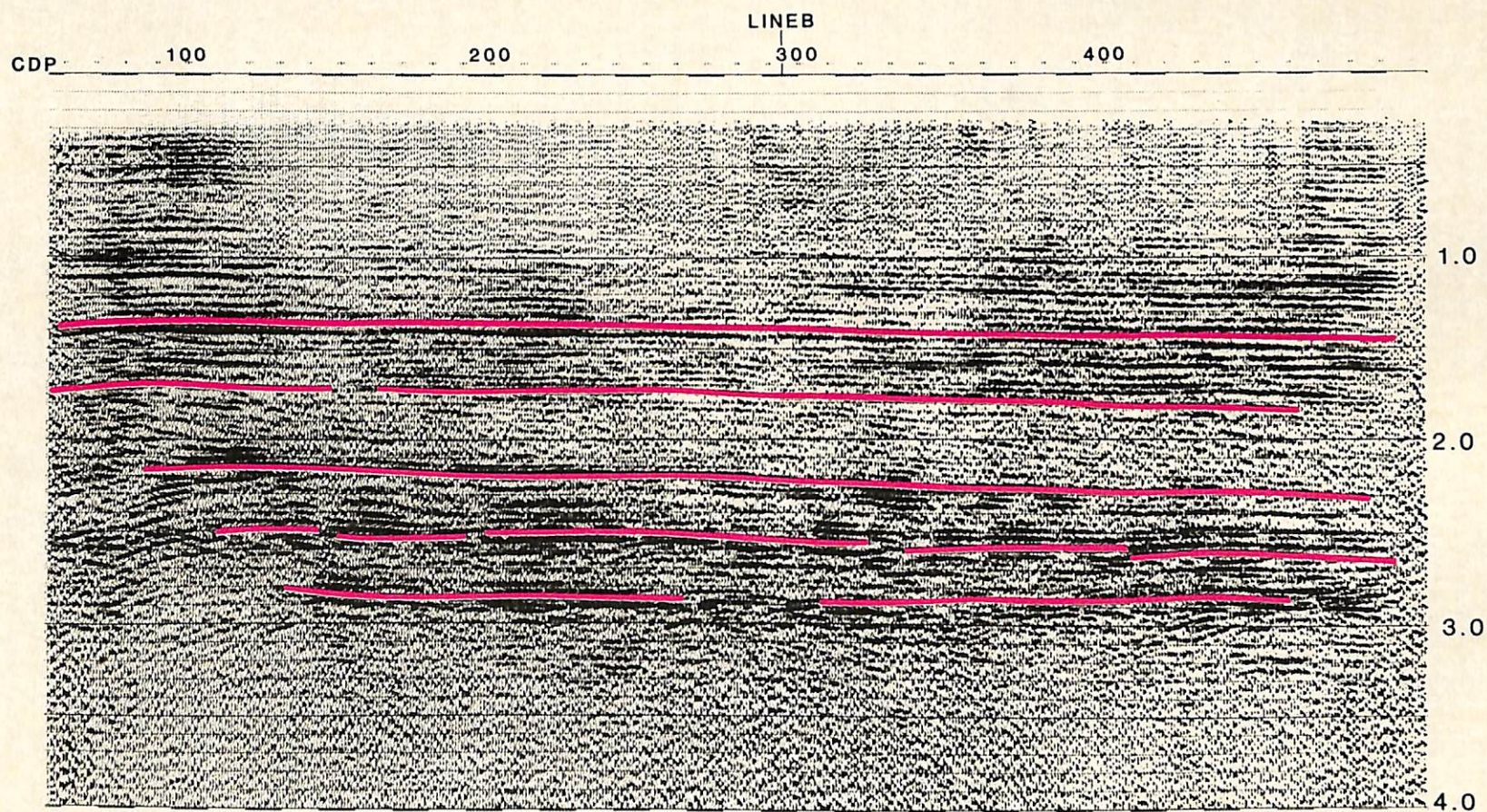
FIG. 7. Stacked section of Line A. The five events used in the tomographic analysis are indicated by red lines.

to convert the time section of Figure 7 to depth (Figure 10). The velocity anomaly is translated into a depth anticline centered around CDP 100 and a depth syncline centered around CDP 165 for all the events on the section. Deeper events are distorted more than shallow ones due to the increase in the size of the velocity anomaly with depth.

Typically, a longer smoothing function would be used to smooth the stacking velocities for depth conversion, and thus reduce the anomaly by averaging the negative and positive excursions. Further, having recognized the permafrost feature, most processors would attempt to correct for the time distortion it introduces by doing a statics analysis. The combination of statics corrections and long-wavelength velocity smoothing will reduce the influence of the permafrost on the inferred depth to the reflectors below it.

This conventional processing approach has two shortcomings. First, the permafrost does not produce purely static effects. That is, different events at the same offset have different delays (or speed-ups) due to raypath bending effects. Second, the lateral extent of the permafrost is a large fraction of a cable length. Solving for long-period delays is difficult with conventional residual-statics programs (Wiggins et al., 1976).

Figure 11 is the depth section after such a conventional processing sequence. A standard residual-statics program (Taner et al., 1974) was used to estimate static delays after moveout with a very slowly varying velocity function. These delays were then applied to the traces before moveout. A subsequent velocity analysis with a one-quarter cable-length triangular smoother was used to stack the data. The same velocity measurements, after the application of a one-cable-length smoother, were used to convert the stacked section to depth.

The red lines in Figure 11 show the depth positions of the events interpreted on the time section (Figure 7). Much of the depth distortion due to the velocity anomaly has now been removed. However, there are still indications of residual permafrost effects. A depth syncline persists for events E4 and E5 near CDP 165. There is also a considerable pull-up under the permafrost. This effect is seen most easily on event E1.

The exploration importance of the correct depth analysis is noted by examining event E4. An apparent anticline exists between CDPs 170 and 300 on E4 which could be interpreted as a favorable structure. The closure of that structure on the left end of the line is due to the synclinal feature around CDP 165. This raises the question of whether this closure is real or just a residual permafrost effect.

To answer the question, the five events discussed previously were used in a tomographic analysis. Traveltimes as a function of offset were digitized on shot gathers displayed on a television screen at an expanded time scale. The digitizing was checked by displaying these time picks on both CDP and offset gathers. Data quality was generally good and the five events were easy to follow on the shot gathers. In the few areas where the data were noisy, or the events ambiguous due to faulting, no time picks were made.

In this example, the first-guess model chosen consisted of velocities that did not vary laterally and depths which were constant except at faults. Jump discontinuities were input at CDP positions of faults which were interpreted on the time section (Figure 7). The theoretical times, obtained by ray tracing the first-guess model, differ from the data times by an rms average of 46 ms across the line.

The tomography algorithm attempts to minimize the difference between the theoretical times and the data times by modifying the velocities and depths of the model. Each iteration produces a new model with a corresponding set of theoretical times. Table 1 shows the misfit between the data times and the

Table 1. Root-mean-square misfit between data and theoretical times (ms), equation (7).

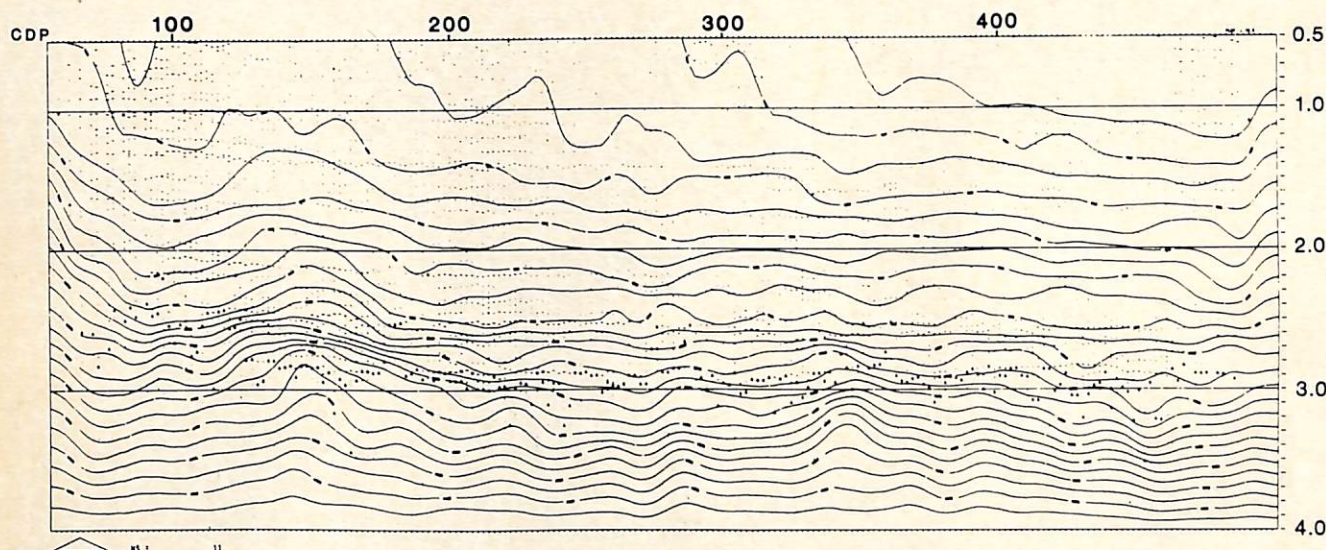|	
First-guess model	46.2
First iteration	6.3
Second iteration	5.4
Third iteration	5.1
Fourth iteration	5.1

FIG. 8. Contour plot of stacking velocities for line A. A stacking-velocity anomaly can be seen on the left side of the line.

A STACKING VELOCITY ANOMALY
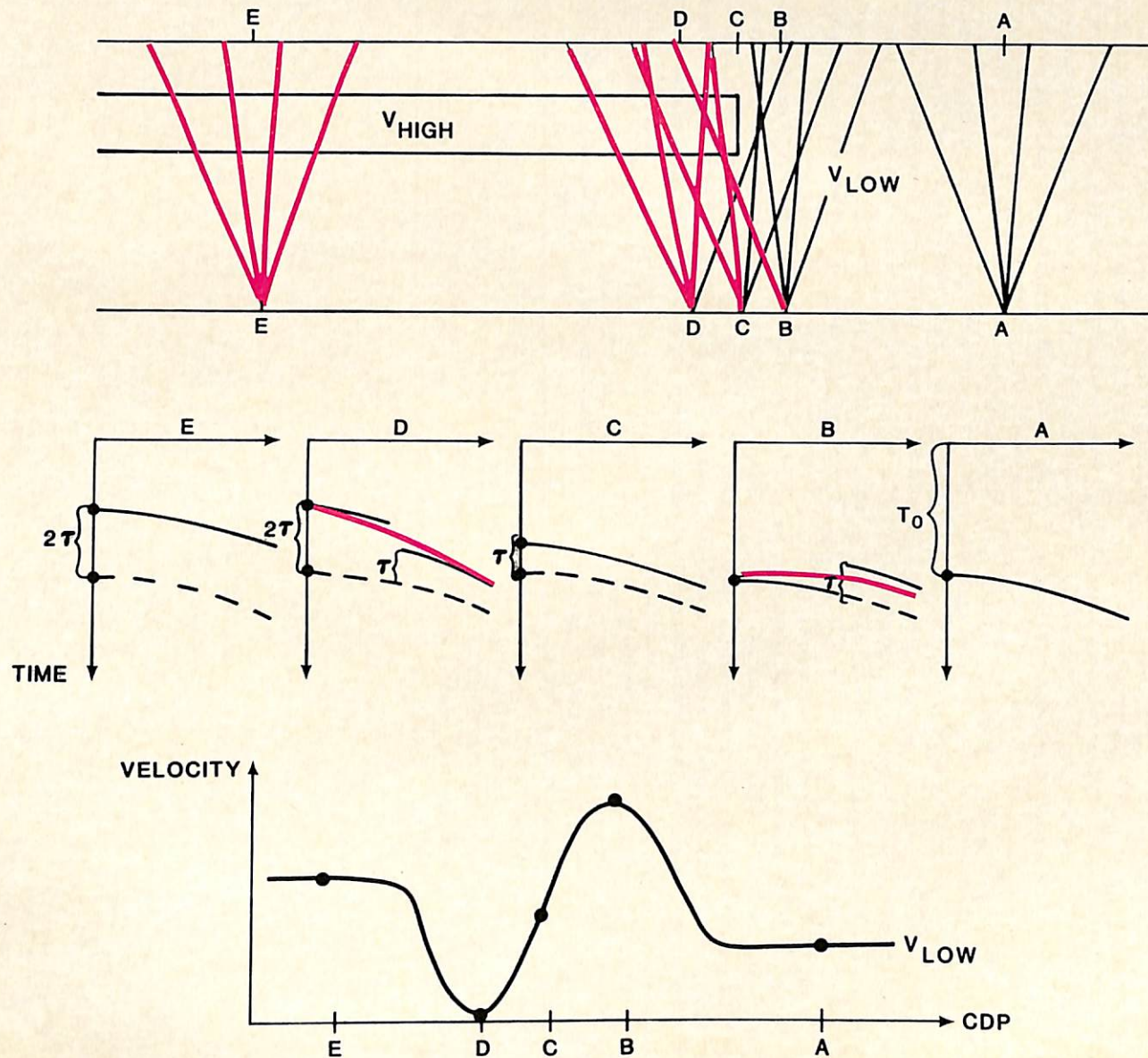# CAUSED BY A HIGH VELOCITY LAYER

FIG. 9. A stacking-velocity anomaly caused by a high-velocity layer. (a) Raypaths at various CDPs; red indicates rays traveling through the high-velocity layer. Layer is assumed to cause speedup τ relative to rays which miss the layer. (b) Moveout curves in CDP gathers. Red curves are best-fit hyperbolas at CDPs B and D. (c) Stacking velocity versus CDP showing velocity anomaly.
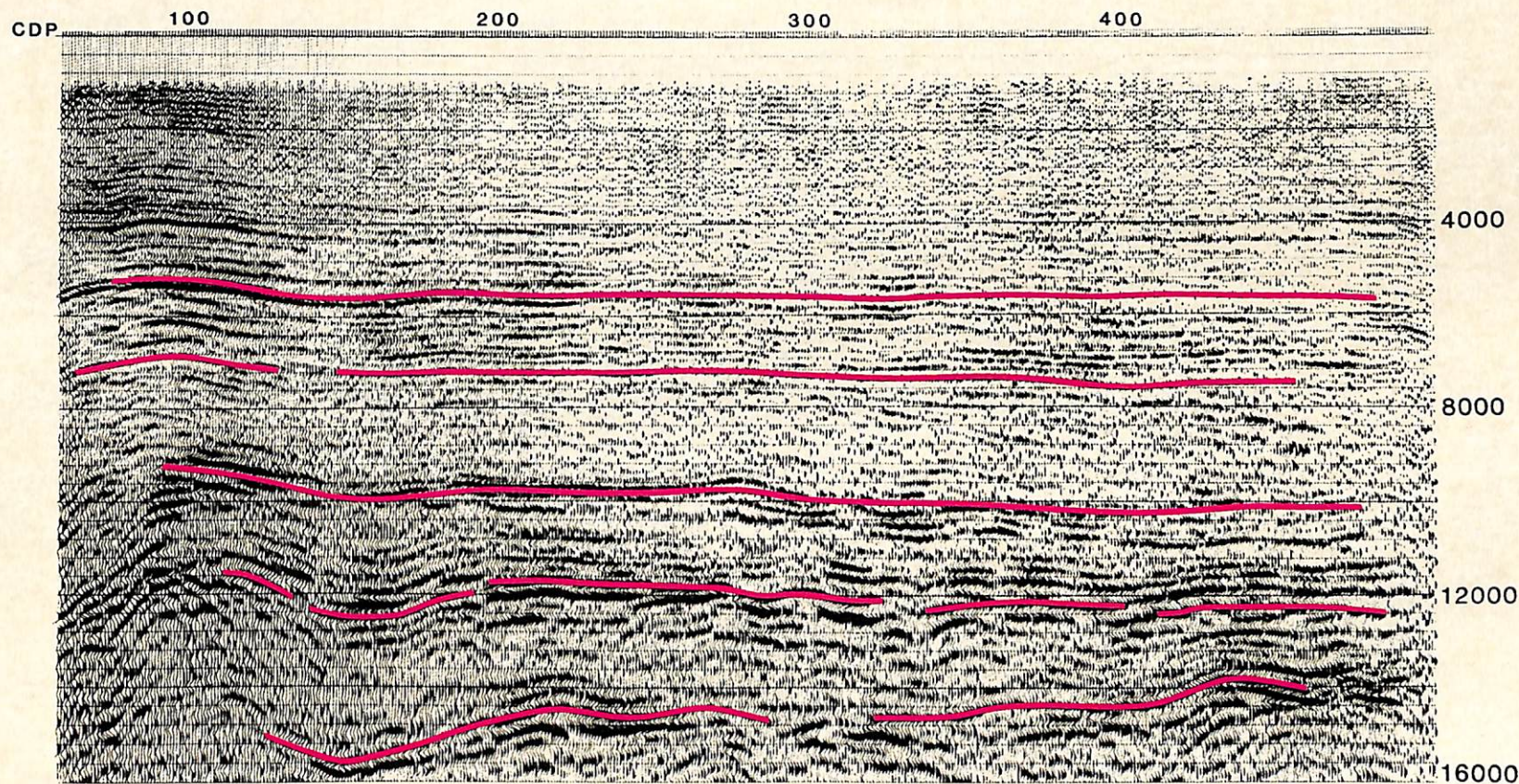
FIG. 10. Line A converted to depth using stacking velocities from Figure 8. Effect of velocity anomaly is obvious near the left edge of the line.
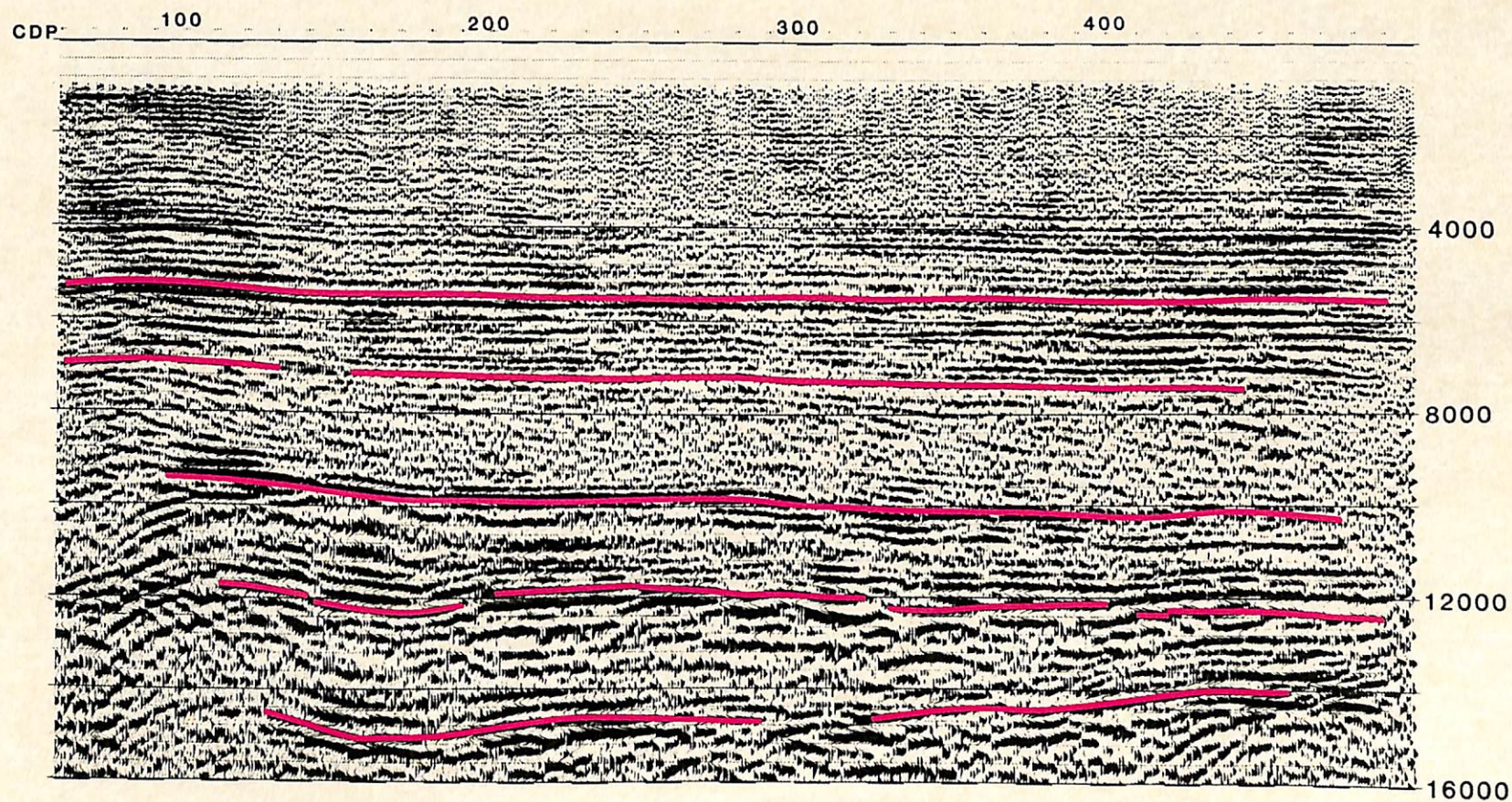
FIG. 11. Line A converted to depth using stacking velocities calculated after statics and smoothed with a long wavelength filter.

theoretical times for four iterations.

It is evident from Table 1 that the tomography algorithm does modify the model to reduce significantly the misfit and that most of the reduction occurs in the first iteration.

Figures 12 and 13 show the tomography model after the first and fourth iterations, respectively. The slownesses (plotted in color) are the differences between the original gray model and the result of the corresponding iteration. Red areas indicate where velocity has been increased and green where it has been decreased. The depths to events are represented by dotted black lines while the original gray-model depths are solid lines. Multiple dotted lines occur for each event where faults have been interpreted. The highlighted dots show the estimated event depth across the line.

Several interesting velocity features are seen in these models. There is a fast, shallow zone between CDPs 80 and 120 which corresponds to the range of CDPs where the near-surface seismic character indicates permafrost. The fast area is smeared down to the left and covers an area of the line greater than a focused image of the permafrost. This is due to both the limited vertical resolution of velocity between events and to the influence of the edge of the line. Velocities in the top layer tend to decrease to the right, corresponding to the lateral gradient observed on the plot of stacking velocities (Figure 8). The leftmost end of the line shows a region of velocity increase which corresponds to the edge of the data; this is probably an edge effect. The two deepest layers show velocity increases to the right. These increases may be due to lateral changes in lithology.

What distinguishes the results of the first iteration from the fourth? Most of the same velocity and depth features appear in both. The main differences are small, detailed changes in the parameters. One change is the improved resolution of the permafrost feature in the upper left end of the model.

To display the tomography results in a form comparable to the conventional processing, the tomographic velocity field is used to create the depth section shown in Figure 14. The depth of events in this display is obtained from the tomographic velocity field only. No direct use is made of the depth parameters. However, examination of the depth section indicates that the depths to the events are essentially the same as the estimates from tomography. This implies that the velocity-depth model found by tomography gives vertical-incidence traveltimes consistent with the stacked data. The red lines on this figure are then essentially the depth estimates from tomography.

The pull-up which lies under the permafrost feature, and which persists after conventional processing (Figure 11), is now corrected. There is also little or no evidence for structural closure on the left end of the line for the deeper events. We suggest that this apparent structure, present after conventional processing, is a residual permafrost effect which has now been corrected.

Similar analyses, both conventional and tomographic, were performed on a cross line (line B). The position of the line intersection is marked on the stacked section (Figure 7). Table 2 compares the depth for the first four events at the line intersection. The interpretation of the depth to the fifth event on Figure 11 is too ambiguous to allow a meaningful comparison.

Note that in both analyses, line A is consistently deeper than line B indicating that the depth mistie is not due to random error. It is likely that tomography has not completely removed systematic velocity-depth effects in the data. However, it has

done considerably better than the conventional processing as indicated by the improved depth ties shown in Table 2.

Some general conclusions can be drawn from this tomographic analysis. Traveltimes obtained by tracing rays through models constructed by this method fit the measured traveltime data to approximately 5 ms rms. This result is within the error in picking the recorded data. Significant shallow velocity variations were imaged in areas where they were predicted by other evidence (shallow waveform character, stacking-velocity anomaly). Depth estimates obtained indicated little or no structural closure on any of the events. The depths determined by tomography tied significantly better at the line intersection, indicating that although some indeterminancy remains, tomography has done a better job of distinguishing velocity effects from depth than the conventional processing scheme discussed earlier.

CONCLUSION

The technique presented uses a general inverse method to estimate velocities and depths from CDP seismic data. The technique attempts to produce a model of both the velocity field and depths to selected reflectors which minimizes the discrepancies between traveltimes derived from tracing rays through the model and those measured from the data. The minimization is achieved by using a Gauss-Newton method to modify a first-guess model and produce a series of model iterates. The first-guess model is constructed manually from well information or from the results of conventional seismic processing. Use of this technique on seismic data indicates that first-guess models with rms misfits in the range of 20 to 100 ms can be modified to produce models with misfits of approximately 5 ms in three to five Gauss-Newton steps.

An investigation of the uniqueness of the models reveals that most features of velocity and depth are determined. However, certain velocity features are undetermined while others are determined relatively poorly. The only degradation in depth resolution is a possible small overall tilt of reflectors.

The tomographic analysis was performed on a pair of intersecting CDP lines in an area with significant lateral-velocity variations due to the presence of permafrost. The technique produced models which fit the data measurements from the two lines to within 4.8 and 5.1 ms, respectively. The models included significant shallow velocity variations in areas where they were predicted by other evidence. The tomographic depth estimates at the intersection point agreed significantly better than the corresponding depth estimates derived from conventional processing.
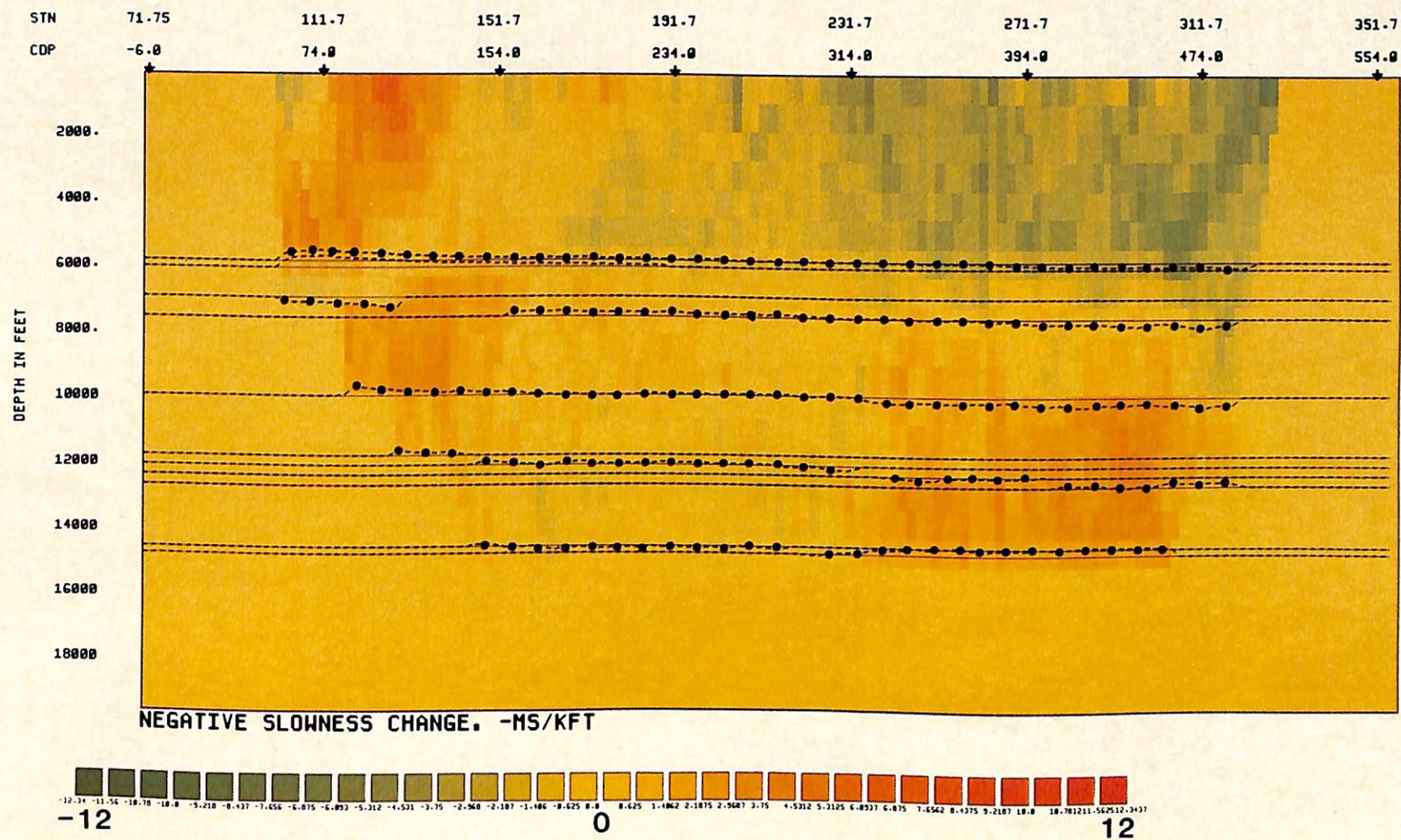
FIG. 12. Velocity-depth model for line A after one Gauss-Newton step.
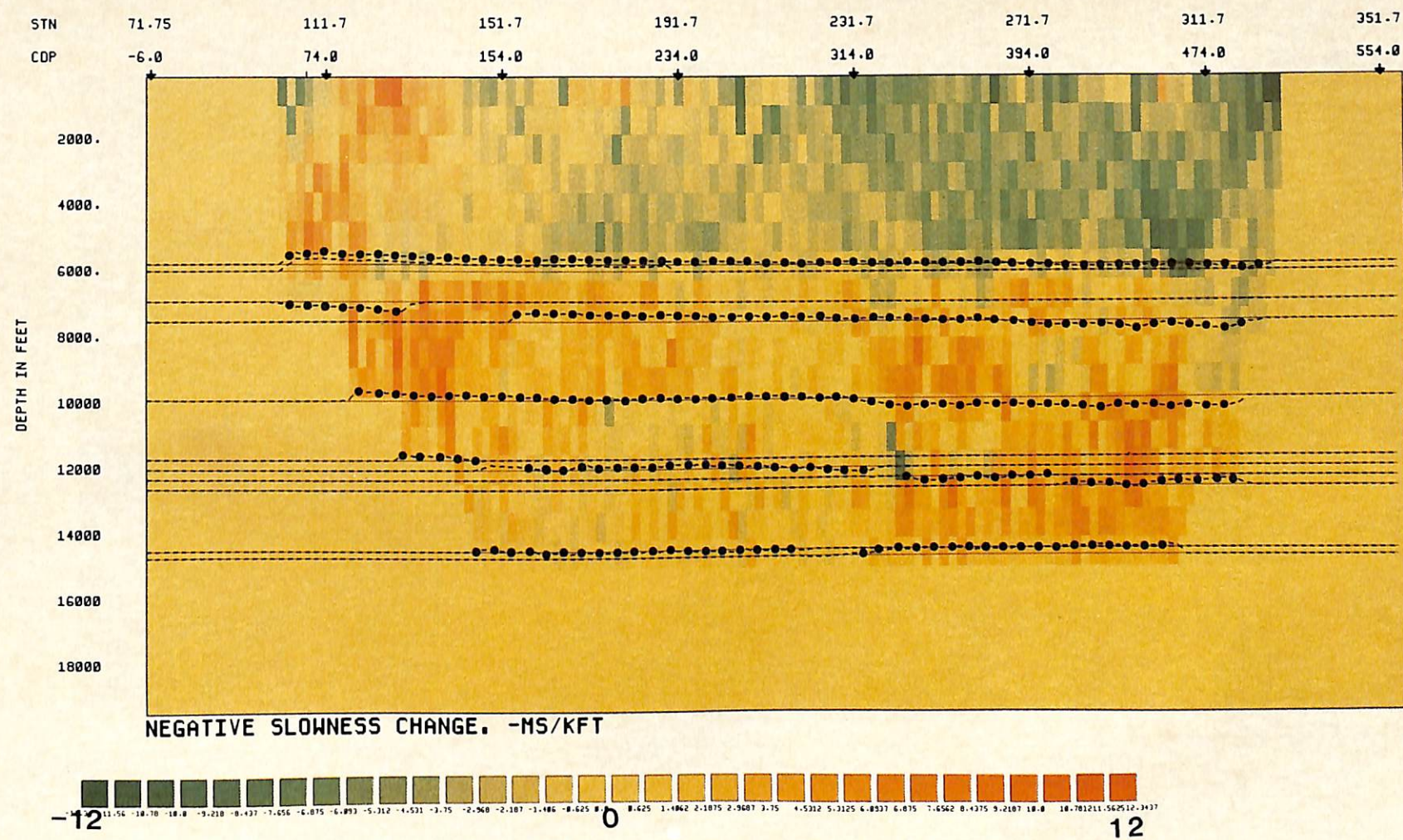
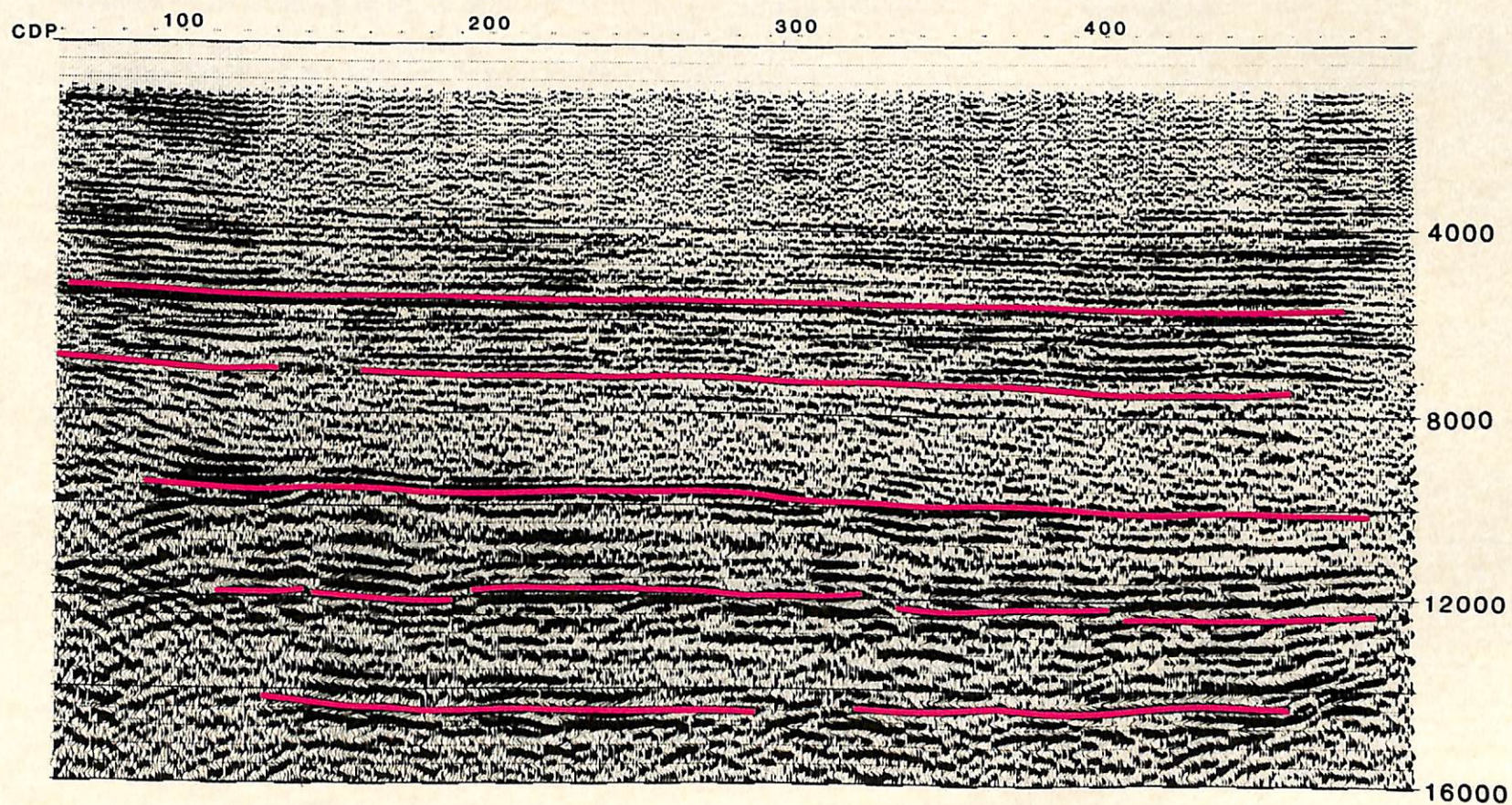FIG. 13. Velocity-depth model for line A after four Gauss-Newton steps.

FIG. 14. Line A converted to depth using velocities from the tomographic analysis.

Table 2. Depth comparison at the line intersection.

Event	Tomographic processing			Conventional processing		
	Line A	Line B	Difference	Line A	Line B	Difference
E1	5 626	5 587	+39	5 480	5 360	+120
E2	7 299	7 230	+69	7 280	7 080	+200
E3	9 757	9 713	+44	9 920	9 600	+320
E4	11 912	11 814	+98	11 880	11 620	+260
			Average = +62.5			Average = +225

REFERENCES

Aki, K., and Lee, W. H. K., 1976, Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes; 1. A homogeneous initial model: J. Geophys. Res., 81, 4381–4399.

Anderson, D. L., and Dziewonski, A. M., 1984, Seismic tomography: Scientific American, 251, 60–68.

Born, M., and Wolf, E., 1959, Principles of optics: Pergamon Press Inc.

Cornuelle, B. D., 1982, Acoustic tomography: Inst. of Elect. and Electron. Eng. on Geosci. and Remote Sensing, 6E-20, 326–332.

Dines, K. A., and Lytle, R. J., 1979, Computerized geophysical tomography: Proc. of the Inst. of Elect. and Electron. Eng., 67, 1065–1073.

Doherty, S. M., and Claerbout, J. F., 1976, Structure independent velocity estimation: Geophysics 41, 850–881.

Forsythe, G. E., Malcolm, M. A., and Moler, C. B., 1977, Computer methods for mathematical computations: Prentice-Hall, Inc.

Franklin, J. N., 1970, Well-posed stochastic extensions of ill-posed linear problems: J. Math Anal. Appl., 31, 682–716.

Gill, P. E., Murray, W. and Wright, M. H., 1981, Practical optimization: Academic Press, Inc.

Golub, G. H., and Reinsch, C., 1970, Singular value decomposition and least-squares solutions: Numer. Math, 14, 403–420.

Hawley, B. W., Zandt, G., and Smith, R. B., 1981, Simultaneous inversion for hypocenters and lateral velocity variations: an iterative solution with a layered model: J. Geophys. Res., 86, 7073–7086.

Kjartansson, E., 1980, Attenuation of seismic waves in rocks, Ph.D. thesis, Stanford Univ.

Langan, R. T., Lerche, I., and Cutler, R. T., 1985, Tracing of rays through an heterogeneous media: An accurate and efficient procedure: Geophysics, September.

Larner, K. L., Hatton, L., Gibson, B. S., and Hsu, I-Chi, 1981, Depth migration of imaged time sections: Geophysics, 46, 734–750.

Lynn, W. S., and Claerbout, J. F., 1982, Velocity estimation in laterally varying media: Geophysics, 47, 884–897.

Marquardt, D. W., 1963, An algorithm for least-squares estimation of non-linear parameters: Soc. of Industr. Appl. Math., 11, 431–441.

Pollet, A., 1974, Simple velocity modeling and the continuous velocity section: Presented at the 44th Annual International Soc. of Expl. Geophys., Meeting, Dallas.

Taner, M., Koehler, F., and Alhilali, K., 1974, Estimation in correction of near-surface time anomaly: Geophysics, 39, 441–463.

Telford, W. M., Geldart, L. P., Sheriff, R. E., and Keys, D. A., 1976, Applied geophysics: Cambridge Univ. Press.

Varga, R. S., 1962, Matrix iterative analysis: Prentice-Hall, Inc.

Wiggins, R. A., 1972, The general linear inverse problem: Implication of surface waves and free oscillations for earth structure: Rev. of Geophys. and Space Phys., 10, 251–285.

Wiggins, R. A., Larner, K. L., and Wisecup, R. D., 1976, Residual statics analysis as a linear inverse problem: Geophysics, 41, 922–938.

APPENDIX A
METHOD OF TRACING RAYS

During the development of seismic tomography we utilized two different ray-tracing methods. Our experience indicates that the difference in the results is principally one of efficiency and the ability to resolve features on the scale of one or two boxes. Our initial method was an iterative one based upon the analytic approach given in Telford et al. (1976). However, the ray-tracing method which we currently use for maximum efficiency and accuracy is presented in a companion paper (Langan et al., 1985). In this appendix we give the analytic approach and its iterative approximation.

The model of the earth used in tomography consists of rectangular boxes, with each box having a known slowness specified at the center and a constant gradient of velocity across the box. The components of the velocity gradient in each box are calculated using the central-difference approximation. For example, the x-component of the velocity gradient in the ith box λ_x is

$$\lambda_x = (1/w_{i+1} - 1/w_{i-1})/(2\Delta x), \qquad (A-1)$$

where w_i is the slowness in the ith box, Δx is the x-dimension of the box, and i increases in the x-direction.

Within a single box, it is known that the path of a ray is a circular arc. Thus, we have constructed a model where ray bending occurs within boxes and not at box boundaries. Therefore we can use simple geometry to express the orientation and traveltime of the ray in the box as a function of position, initial orientation, velocity gradient, and slowness at the center of the box. Telford et al. (1976) gave the expressions for the case where velocity increases linearly with depth only. We present similar equations for the more general case of an arbitrary direction for the gradient. In this case the velocity field is defined by

$$V(z) = V_0 + \lambda \cdot \mathbf{r}, \qquad (A-2)$$

where V_0 is the velocity at the origin, λ is the velocity gradient, and $\mathbf{r}$ is the position vector. For slowness, equation (A-2) is written

$$w(\mathbf{r}) = 1/(1/w_0 + \lambda \cdot \mathbf{r}), \qquad (A-3)$$

where w_0 is the slowness at the origin.

Suppose a ray enters a box at the point $(x = 0, z = 0)$ at an angle of entry ϕ_0 measured with respect to the vertical. Let the ray emerge from the box at (x, z) with angle ϕ. Then

$$\phi = \arccos\ [\cos\ \phi_0 - x/R], \qquad (A-4)$$

and

$$\phi = \arcsin\ [\sin\ \phi_0 + z/R], \qquad (A-5)$$

where R is the radius of curvature given by

$$R = 1/|w_0\lambda\ \sin\ (\phi_0 - \psi)|, \qquad (A-6)$$

and ψ is the angle the gradient makes with the vertical.

These equations can be used to find the position of the ray and its angle of exit from the box. For instance, if the ray exits from below, the depth of exit z is known, and ϕ and x then follow from equations (A-5) and (A-4), respectively. (The value of x obtained in this way must lie within the box; if it does not, the assumption that the ray exits from below is in error. In this case, it is really ϕ and z that must be computed with known x.) With the raypath determined, the path length s is given by

$$s = |R(\phi - \phi_0)|. \qquad (A-7)$$

The traveltime is then approximated as the product of path length and slowness at the center of the box.

Because this analytic approach is computationally slow, we performed our initial studies with a less precise, iterative method which approximates it. A tangent to the circular path is constructed at the entry point. The length of this tangent in the box is used as an initial estimate for arc length. An estimate of the change in ray orientation across the box is calculated from equation (A-7). An estimate of the position of the chord to the arc is constructed from this angle. The length of this chord is then used to calculate a new estimate of the arc length. If the chord and tangent exit different sides, this process is repeated.

APPENDIX B

CALCULATION OF SLOWNESS AND DEPTH DERIVATIVES

Entries in the Jacobian matrix $\mathbf{A}^{(k)}$ are of two kinds: derivatives of traveltime with respect to slowness, and derivatives of traveltime with respect to reflector depth. Both are computed immediately following the computation of the traveltimes $\mathbf{t}(\mathbf{p}^{(k)})$. Thus, the raypaths through the model are known when the derivatives are computed.

In the ray tracing used here, rays are continuous at box boundaries and refract inside boxes by moving along curvilinear paths determined by the slowness gradient. However, for simplicity the discussion will begin with an examination of the calculation of the slowness derivatives in a medium where the gradient is zero inside each box, and rays refract by bending at box boundaries. Also, reflectors will consist of piece-wise linear functions rather than cubic splines. Some discussion of the problem of computing the derivatives in the more general case will then follow.

From Fermat's principle, the raypath through the medium is that path which exhibits the minimum traveltime between the fixed source and receiver. The minimum is taken over all paths which lie in a certain "regular neighborhood" (Born and Wolf, 1959), and it is assumed that in that neighborhood there is only one such path. This guarantees that the traveltime is a true minimum rather than just an extremum.

In the finite-dimensional medium considered here, the path through a single box is simply a straight line. Moreover, any possible path through the entire medium can be specified by prescribing the depths and an additional finite set of independent path parameters. The latter will be denoted by the vector $\mathbf{q}$ with components q_n, $n = 1, 2, \ldots, \gamma$. (For example, one specific q_n might be the distance from a box corner to the point at which the ray enters or leaves the box, or the distance from a fixed point on a reflector to the point of reflection.)

Now define a real function $f_i(\mathbf{q}, \mathbf{p})$ to be the traveltime for the ith ray along a path through the medium specified by $\mathbf{q}$. From Fermat's principle, the true raypath traveltime $t_i(\mathbf{p})$ is given by

$$t_i(\mathbf{p}) = \min_{\mathbf{q}} f_i(\mathbf{q}, \mathbf{p}). \qquad (B-1)$$

The minimum in equation (B-1) is taken for $\mathbf{q}$ in a certain domain in $\mathscr{R}^\gamma$. Assume that the parameterization is such that the minimum does not occur on the boundary of the domain, and suppose that inside that domain the function f_i is smooth. It follows that at the minimum

$$\frac{\partial f_i}{\partial q_n} = 0; \qquad n = 1, 2, 3 \ldots \gamma. \qquad (B-2)$$

Let the vector of values of the components of $\mathbf{q}$ at the minimum be denoted by $\mathbf{q}^*$. The equations (B-2) provide an implicit relationship between $\mathbf{q}^*$ and $\mathbf{p}$. Assume it is possible to express this dependence as an explicit function of $\mathbf{p}$. This is equivalent to the assumption that the matrix with entries $\partial^2 f_i/\partial q_n\ \partial q_m$ is positive definite, and not just semidefinite, at the minimum. Thus at the minimum

$$\mathbf{q}^* = \mathbf{q}^*(\mathbf{p}), \qquad (B-3)$$

and equation (B-1) can be written

$$t_i(\mathbf{p}) = f_i(\mathbf{q}^*(\mathbf{p}), \mathbf{p}). \qquad (B-4)$$

Differentiating equation (B-4) with respect to p_j, evaluating all derivatives at the minimum, and making use of equations (B-2) yields

$$\frac{\partial t_i}{\partial p_j} = \sum_{n=1}^{\gamma} \left(\frac{\partial f_i}{\partial q_n}\bigg|_{\mathbf{q}=\mathbf{q}^*}\right)\frac{\partial q_n^*}{\partial p_j} + \frac{\partial f_i}{\partial p_j} = \frac{\partial f_i}{\partial p_j}. \qquad (B-5)$$

This shows that it is not necessary to know how the path parameters $\mathbf{q}^*$ depend upon the medium parameters in order to find $\partial t_i/\partial p_j$. To ascertain the effects of a small perturbation in slowness, only the original raypath (corresponding to the undisturbed parameters and depths) need be known. Similarly, for a perturbation in depth, the path required is the one specified by the unperturbed $\mathbf{q}^*$ and the new depth.

Obtaining closed-form expressions for the derivatives is now straightforward. The traveltime in a box is the product of the slowness $w_{k\ell}$ and the path length $s_{k\ell}$. The latter is a function of some of the q_n^* (and, if the box contains a reflector, of two depth

parameters), but it does not depend upon any of the slowness parameters. Therefore,

$$f_i = \sum s_{k\ell} w_{k\ell}, \qquad (B\text{-}6)$$

where the sum is taken over all boxes through which the ray passes. Using $w_{k\ell} = p_j$ in equation (B-5) then gives

$$\frac{\partial t_i}{\partial w_{k\ell}} = s_{k\ell}. \qquad (B\text{-}7)$$

For the depth derivative calculation, let the reflector pass through $(x_{m-1}, Z_{\rho, m-1})$ and $(x_m, Z_{\rho m})$, and let the point of reflection be (x_R, Z_R), with $x_{m-1} \le x_R \le x_m$, and $Z_{\rho, m-1} \le Z_R \le Z_{\rho m}$ (see Figure B-1). To compute $\partial t_i/\partial Z_{\rho m}$, it is necessary to evaluate the increase in path length Δs due to an increase in depth $\Delta Z_{\rho m}$ along a path characterized by the same $\mathbf{q}^*$ parameters as the unperturbed route. Consider the case where the $\mathbf{q}$ parameters represent the distance from a box corner to the point of entry into or exit from that box, or the distance from $(x_{m-1}, Z_{\rho, m-1})$ to (x_R, Z_R). Then the new path is identical to the original raypath except in the box containing the point of reflection. In that box, although the reflector is rotated clockwise, the point of entry of the new path into the box, the point of exit from the box, and the distance from $(x_{m-1}, Z_{\rho, m-1})$ to the reflection point must remain intact. Note that such a path *differs* from the actual path taken by a ray which reflects from the downward-rotated reflector. The proof demonstrates that the true raypath need not be known, because the difference between the traveltime along it and the traveltime along the path described is at most of second order.

Let θ be the angle between the incident ray and a line normal to the reflector passing through (x_R, Z_R). The reflector is inclined with respect to the horizontal by an angle $\beta = \tan^{-1}$ $(Z_{\rho m} - Z_{\rho, m-1})/(x_m - x_{m-1})$. Suppose a vertical line through (x_R, Z_R) intersects the perturbed reflector at $(x_R, Z_R + \Delta d)$. Referring to Figure B-1, it is clear that

$$\Delta s = 2\Delta d \cos \beta \cos \theta, \qquad (B\text{-}8)$$

and furthermore that

$$\Delta d = \Delta Z_{\rho m}(x_R - x_{m-1})/(x_m - x_{m-1}). \qquad (B\text{-}9)$$

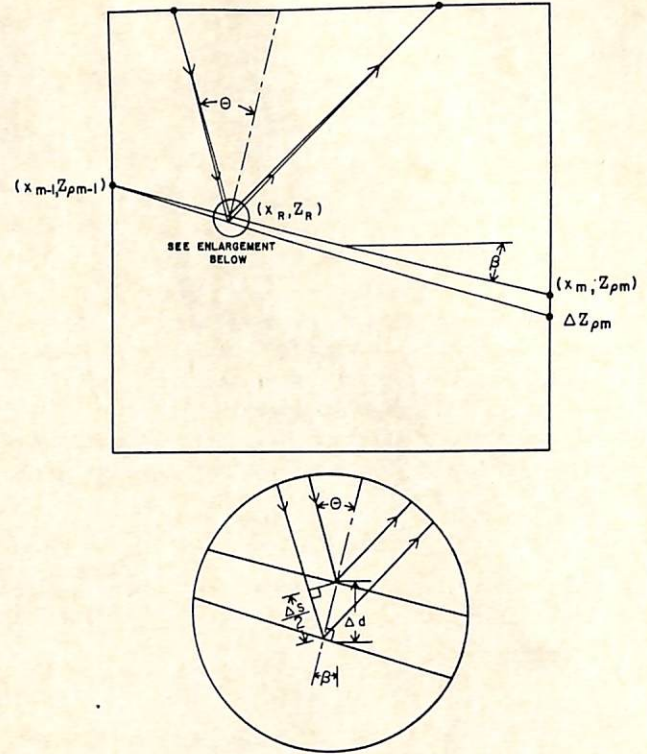Thus, if the box has slowness $w_{k\ell}$, the depth derivative is given by

FIG. B-1. Calculation of depth derivative.

$$\frac{\partial t_i}{\partial Z_{\rho m}} = 2 w_{k\ell} \cos \beta \cos \theta \left(\frac{x_R - x_{m-1}}{x_m - x_{m-1}} \right). \qquad (B\text{-}10)$$

As noted, these calculations pertain to a simplified version of ray tracing. In practice the ray tracing employed is that described in Appendix A. Nonetheless, these equations are used as shown here. The justification for this procedure ultimately lies in the success of the inversion algorithm; even with such simplifying assumptions, the residual is reduced to acceptable levels.

APPENDIX C
DETERMINED AND UNDETERMINED QUANTITIES

Appendix C gives a proof of statements made earlier regarding the determination of depth and slowness in a model earth in which ray paths are straight, reflectors are flat, and the gradient of slowness in each box is zero. (Slowness in adjacent boxes may differ; that is, this is not a restriction to laterally homogeneous models.) It is assumed that traveltime measurements for the model earth are available from sources and receivers placed at locations specified as necessary below. For simplicity, it is also assumed that reflectors lie on horizontal boundaries between boxes.

That the depth of each reflector is uniquely determined by the data can be proven as follows: consider a ray in a single column of boxes striking reflector ρ at depth Z_ρ (Figure C-1). Let the distance between shot and receiver be $2\Delta x$, let the round trip traveltime be t_1, and let θ_1 be the angle of incidence at which the ray encounters the reflector. Similarly, let $2\Delta x_2$, t_2, and θ_2 denote the analagous quantities for a second ray in the same vertical column of boxes. Then,

$$t_1 = \sum w_{k\ell} d_z \sec \theta_1 \qquad (C\text{-}1)$$

and

$$t_2 = \sum w_{k\ell} d_z \sec \theta_2, \qquad (C\text{-}2)$$

where the sum is over the slowness in the column of boxes which contains the rays. Hence,

$$\frac{t_1^2}{t_2^2} = \frac{\sec^2 \theta_1}{\sec^2 \theta_2} = \frac{Z_\rho^2 + \Delta x_1^2}{Z_\rho^2 + \Delta x_2^2}, \tag{C-3}$$

and, consequently,

$$Z_\rho^2 = \frac{(t_1^2 \Delta x_2^2 - t_2^2 \Delta x_1^2)}{t_2^2 - t_1^2}. \tag{C-4}$$

Thus the depth of each reflector is uniquely determined.

What can be said about the determination of the slowness parameters in this model? With the reflector depths inferred, they can be regarded as fixed, and in that case the traveltime function $t(p)$ can be considered to depend only on the $n_x n_z$ slowness parameters. The straight-ray approximation makes $t(p)$ a linear function of the slowness parameters p. Thus $t(p)$ can be written as

$$t(p) = \underline{A}p, \tag{C-5}$$

where the ijth entry in the $N \times M$ matrix $\underline{A}$ is the path length of ray i in box j. If the data vector for this system is t_d, then p must satisfy the equation

$$t_d = \underline{A}p. \tag{C-6}$$

Suppose one solution of equation (C-6) is $p = p^*$. If there is another solution $p = p^* + \Delta p$, then Δp is such that

$$\underline{A}\Delta p = 0, \tag{C-7}$$

that is, Δp is a member of $N(\underline{A})$, the null space of $\underline{A}$. It is easy to show that $N(\underline{A})$ is not empty. In particular, one element of this space consists of a slowness distribution Δp which is laterally homogeneous and has the property that the sum of the slowness components in a single vertical column of boxes above a reflector is zero. With n_R reflectors and n_z rows of boxes, the dimension of $N(\underline{A})$ is clearly at least $n_z - n_R$.

With certain distributions of rays, the dimension of the null space may be much larger than $n_z - n_R$. However, there is at least one finite set of rays for which no other elements in $N(\underline{A})$ exist. This is proven by exhibiting a set of rays for which the row space of $\underline{A}$ has dimension $M - (n_z - n_R)$. The null space of $\underline{A}$, which is the orthogonal complement of the row space, then has dimension at most $n_z - n_R$.

A collection of rays having this property is most easily described in connection with a specific example, such as that
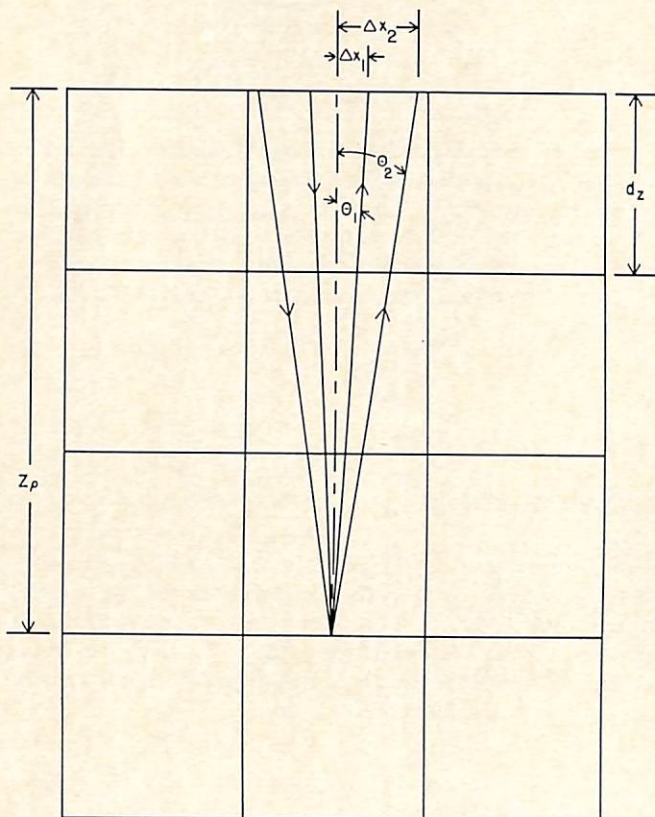

FIG. C-1. Determination of depth in the case of straight rays and a flat horizontal reflector.

ELEMENTS OF SPAN OF ROWS OF A

FIG. C-2. Elements of the row space of $\underline{A}$.

shown in Figure C-2. Here, $n_z = 3$, $n_R = 1$, $n_x = 4$, and $M = 12$. The numbers in each box of Figures C-2a and C-2b are proportional to the entries in that row of $\underline{A}$ which corresponds to the particular ray shown. The difference between these two rows, which is also in the row space, is shown in Figure C-2c. All the vectors seen in Figure C-3 except the last can be generated in this fashion. This yields a total of $n_z(n_x - 1)$ linearly independent vectors. In addition, using one ray in each column of boxes, elements in the row space of the type seen in the last diagram of Figure C-3 can be created. There is one such vector corresponding to each of the n_R reflectors. Thus the total

number of linearly independent vectors for this ray distribution is

$$n_z(n_x - 1) + n_R = M - (n_z - n_R). \tag{C-8}$$

Hence the null space has dimension $n_z - n_R$.

Note that elements in the null space affect neither the average slowness between reflectors nor the lateral change of slowness in a single row of boxes. This is the basis for the statements made in the discussion of effects of limited-angle rays in the text.

LINEARLY INDEPENDENT ELEMENTS OF SPAN OF ROWS OF A

1	-1	0	0
0	0	0	0
0	0	0	0

0	1	-1	0
0	0	0	0
0	0	0	0

0	0	1	-1
0	0	0	0
0	0	0	0

0	0	0	0
1	-1	0	0
0	0	0	0

0	0	0	0
0	1	-1	0
0	0	0	0

0	0	0	0
0	0	1	-1
0	0	0	0

0	0	0	0
0	0	0	0
1	-1	0	0

0	0	0	0
0	0	0	0
0	1	-1	0

0	0	0	0
0	0	0	0
0	0	1	-1

1	1	1	1
1	1	1	1
1	1	1	1

FIG. C-3. Linearly independent elements which span the row space of $\underline{A}$.