Robust inversion of VSP's

Bill Harlan

INTRODUCTION

A previous paper (Harlan and Lailly, 1984) outlined a general method for the inversion of non-linear differential systems with an application to vertical seismic profiles (VSP's). That paper described all general strategies and statistical tools I shall require here. I turn to more mechanical difficulties now in hope of better illustrating the method's application to a given differential system.

Classical optimal control theory offers such powerful generality that I have given the first section of this paper to its elaboration. One is able to incorporate boundary conditions and stability criteria, yet treat a differential system in its continuous form.

Secondly I more explicitly describe my modifications of classical control as applied to VSP's -- where and how robustness should be added.

OPTIMAL CONTROL OF VSP'S

I review here a general method for the inversion of systems governed by partial differential equations. The mathematical techniques are due to Lyons (1968). The example of 1D inversion of seismic waves derives from Bamberger, Chavent, and Lailly (1982).

The differential system

The following differential system will determine the 1D propagation of acoustic waves.

$$\sigma \frac{\partial^2 y}{\partial t^2} - \frac{\partial}{\partial x}\left[\sigma \frac{\partial y}{\partial x}\right] = 0 \tag{1}$$

$$\sigma \frac{\partial y}{\partial x} \Big|_{x=0} = -g \tag{2}$$

$$y \mid_{t=0} = \frac{\partial y}{\partial t} \mid_{t=0} = 0 \tag{3}$$

$y(x,t)$ is the time derivative of the vertical displacement measured at time t and pseudo-depth x:

$$x = \int_0^z \frac{dz}{v_p(z)} . \tag{4}$$

x is the depth measured by the travel-time from the surface (Mace and Lailly, 1984). $\sigma(x) = v_p(x)\rho(x)$ is the impedance. $g(t)$ is the time derivative of the vertical traction (force per unit area) at the surface.

All functions will be invariant over the other two spatial dimensions. Assume data are recorded at a set of depth points $\{x_i\}$ for $0 \leq t \leq T$. These depths may be irregular and quite sparse. Positions of zero time and of zero depth are arbitrary for inversion. Assume that the first geophone is slightly below zero depth and invert for an appropriately shifted source function.

Let us invert for $g(t)$ over $0 \leq t \leq T$ and for $\sigma(x)$ over $0 \leq x \leq X$ where $X = x_{\max} + (T - x_{\max})/2$. X represents the maximum depth from which a reflection can be made in the given time span. One may safely assume that the wavefield is zero at deeper points without affecting the inversion of the earth above. Impedance may be varied at this depth without affecting the modeled data. Thus we obtain another boundary condition:

$$y \mid_{x=X} = 0 \tag{5}$$

The variational form

One may express equations (1), (2), and (4) as a single variational equation. Write (1) as

$$\int_0^T \int_0^X \left\{ \sigma \frac{\partial^2 y}{\partial t^2} q - \frac{\partial}{\partial x} [\sigma \frac{\partial y}{\partial x}] q \right\} dx \; dt = 0 \quad \textit{for all } q(x,t) \tag{6}$$

Integrate once by parts and incorporate boundary conditions (2) and (4).

$$\int_0^T \int_0^X \left\{ \sigma \frac{\partial^2 y}{\partial t^2} q + [\sigma \frac{\partial y}{\partial x}] \frac{\partial q}{\partial x} \right\} dx \; dt = \int_0^T gq \; dt \mid_{x=0} \tag{7}$$

Equation (7) contains all information in the original three equations.

Choosing a cost function

Now we wish to define the $\sigma(x)$ and $g(t)$ that best model the data. One should minimize the resulting error (noise) of the modeled data. For stability we want to minimize the norms of the inverted signal spaces.

$$J = \frac{1}{C_n^2} \sum_{x \in \{x_i\}} \int_0^T [d(x,t) - y(x,t)]^2 \, dt + \frac{1}{C_g^2} \sum_{t \in \{t_i\}} g^2(t) + \frac{1}{C_\sigma^2} \sum_{\chi \in \{\chi_i\}} [\frac{d\sigma}{dx}(\chi)]^2 \quad (8)$$

$\{t_i\}$ and $\{\chi_i\}$ represent the sets of points at which $g(t)$ and $\sigma(x)$ are to be inverted, and $\{x_i\}$ the depths at which $d(x,t)$ is measured. The C's represent variances. Even sampling of signal parameters is not necessary.

Minimization

A perturbation of y will create a corresponding perturbation of the cost function. From (8) we obtain

$$\delta J = \frac{-1}{C_n^2} \sum_{x \in \{x_i\}} \int_0^T [d - y] \delta y \, dt \quad (9)$$

$d - y$ is merely the error of the fit, or noise. Equation (9) allows us to find the relationship between perturbations in y and perturbations in σ and g.

$$\int_0^T \int_0^X [\frac{\partial^2 y}{\partial t^2} q + \frac{\partial y}{\partial x} \frac{\partial q}{\partial x}] \delta\sigma \, dx \, dt - \int_0^T q \, \delta g \, dt \mid_{x=0} \quad (10)$$

$$= -\int_0^T \int_0^X \left\{ \sigma q \frac{\partial^2}{\partial t^2} \delta y + [\sigma \frac{\partial}{\partial x} \delta y] \frac{\partial q}{\partial x} \right\} dx \, dt$$

$$= -\int_0^T \int_0^X \left\{ \sigma \frac{\partial^2 q}{\partial t^2} - \frac{\partial}{\partial x} [\sigma \frac{\partial q}{\partial x}] \right\} \delta y \, dx \, dt$$

Again we have integrated by parts. All unperturbed functions remain at reference values. Boundary condition (3) has destroyed two boundary terms. Suppress remaining boundary terms now and require appropriate boundary conditions in the following application.

Now for the cleverness. Choose a particular value of q such that the right-hand side of equation (10) equals the right-hand side of (9). Equations (9) and (10) will then yield the gradients $\delta J / \delta\sigma(x)$ and $\delta J / \delta g(t)$ for this chosen q and the reference field y.

$$\nabla_\sigma J = \int_0^T [\frac{\partial^2 y}{\partial t^2} q + \frac{\partial y}{\partial x} \frac{\partial q}{\partial x}] dt + \frac{1}{C_\sigma^2} \frac{d\sigma}{dx} \quad (11)$$

$$\nabla_g J = -q \mid_{x=0} + \frac{1}{C_g^2} g \tag{12}$$

To determine the proper q eliminate δy from the forced equality of (9) and (10).

$$\sigma \frac{\partial^2 q}{\partial t^2} - \frac{\partial}{\partial x}[\sigma \frac{\partial q}{\partial x}] = \frac{1}{C_n^2} \sum_{x_i} [d - y] \delta(x - x_i) \tag{13}$$

$$q \mid_{t=T} = \frac{\partial q}{\partial t} \mid_{t=T} = 0 \tag{14}$$

$$\frac{\partial q}{\partial x} \mid_{x=0} = \frac{\partial q}{\partial x} \mid_{x=X} = 0 \tag{15}$$

The above system is properly called the adjoint of our first differential system (1), (2), and (3). Boundary conditions (14) and (15) result from the suppressed boundary terms in equation (10). Condition (15) holds because of the assumption that we do not have a geophone at the surface nor at the depth too far to be reached by waves with $t < T$. Non-zero values may be introduced for such terms with a less convenient geometry. Our eliminations, however, do not restrict the application to VSP's.

Final steps

With the gradients (11) and (12) in our hands, the procedure for inverting as a classical control problem becomes quite comfortable. At any given iteration we will have estimates of σ and g and a corresponding estimated wavefield. Subtraction of this wavefield from the data provides a perturbation of the system δy, to be accounted for by appropriate perturbations $\delta\sigma$ and δg. A steepest-descent or Fletcher-Reeves algorithm would perform a line search in order to find the appropriate scale factor. The reference parameters should be perturbed, and a new δy calculated.

MODIFICATIONS OF OPTIMAL CONTROL FOR THE VSP

Here I discuss principle modifications of the classical method of optimal control described above. Again, I refer readers to the paper of Harlan and Lailly (1984) for explanation of the statistical tools used below. I shall include their functions and justifications, but omit algorithmic details involving manipulation of histograms.

Simplifying inverted impedances

In spite of its small dimensionality, the impedance function carries an important null space. One may completely obscure the inverted impedances with components that do not affect the modeled result. The difficulty lies in properly controlling the *a priori*

impedance statistics.

The first derivative with depth of an impedance log creates an array whose samples have negligible statistical dependence. That is, a single impedance transition tells nothing to an interpreter about the location or magnitude of other transitions. However, since the earth is made of largely homogeneous packages, transitions are few: a majority will be zero. Also, few interpreters would care to prejudice their inversions with *a priori* predictions about the overall distribution of transitions with depth. To summarize in statistical terms, I say that the samples of a differentiated impedance log should be an independent, identically distributed (IID), non-gaussian random process.

The least-squares (l.s.) constraint given in the functional (8) properly treats the IID property of $d\,\sigma(x)/dx$, but assumes that this function (when sampled) is gaussian. Real impedance logs are strongly non-gaussian. As a result, an inversion with this functional increases the *a priori* probability by exchanging sparse strong transitions for numerous small transitions. The result is predictably gaussian but difficult to interpret.

Figure 1 contains a noiseless synthetic VSP prepared from an impulsive source $g(t)$ and the impedance log $\sigma(x)$ in Figure 2. I began with a zeroed source and constant impedance. Three iterative perturbations of $g(t)$ and $\sigma(x)$ with the least-squares functional produced Figure 3 and the forward model and residuals of Figures 4 and 5. A low clip of amplitudes emphasizes the character of the uninverted events. The inverted log contains high-frequency events that largely obscure the important transitions. The loss of low frequency information destroys the blocky character.

Alternatively I proceeded as outlined by Harlan and Lailly (1984). Iterative estimation of the actual probability density function (pdf) for $d\,\sigma(x)/dx$, followed by Bayesian estimation and an extraction of the most reliable transitions yielded the impedance log of Figure 6. (A recapitulation of the algorithm will end this paper.) The fit with the data was no worse than that of the gaussian inversion, yet the first derivative of the log is zero most everywhere. The most important transitions have been captured, and the low frequency component is just.

The impossibility of linearizing the differential system

A temporary linearization of the system (1-3) with respect to background parameters would offer the attractive possibility of iteratively inverting a linearized least-squares system. However, a little testing shows that a non-zero impedance function $\sigma(x)$ as reference makes the linearized system unstable. Thus, the appropriate scale factor for perturbations must derive from a line search, or the expensive matrix inversion of Newton's method. I facilitate the line search by finding the appropriate scale factor for

FIG. 1. A noiseless synthetic VSP prepared from an impulsive source and the impedance function of Figure 2.

FIG. 2. A synthetic impedance function prepared by integration of a non-Gaussian random process.

FIG. 3. A l.s. inversion for impedance (minimizing functional [8]) obscures important transitions with unreliable high-frequency components. Low frequency structure is lost.

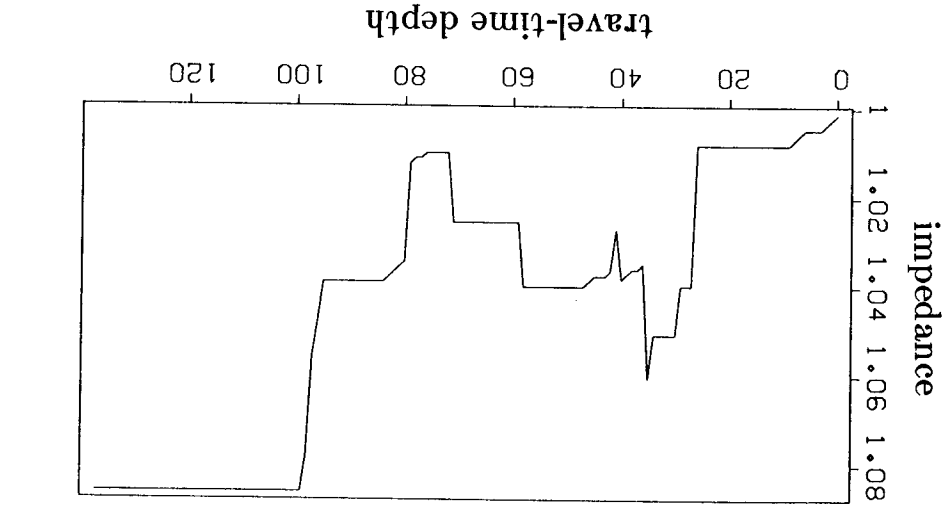

FIG. 6. A robust estimate of impedance captures the most important transitions. Modeled data is as good as for the l.s. solution.

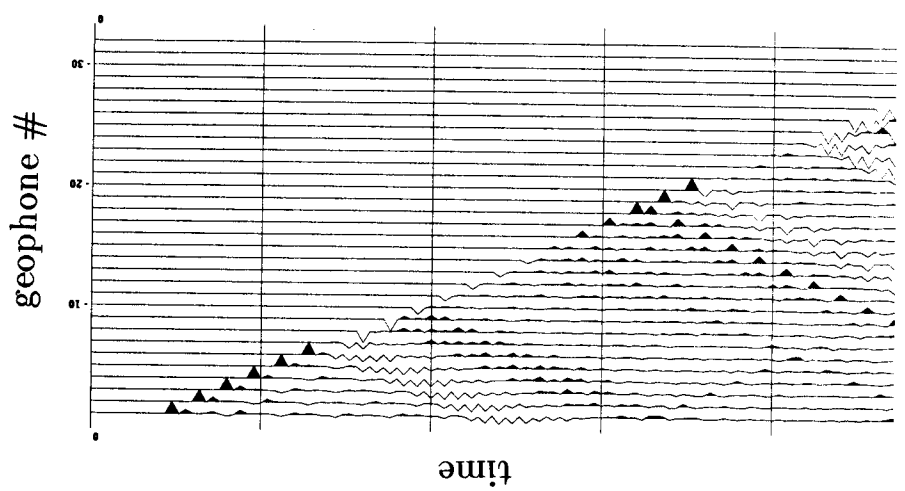

FIG. 5. Residual noise is evenly distributed. Gain is high and identical to Figures 3 and 4.

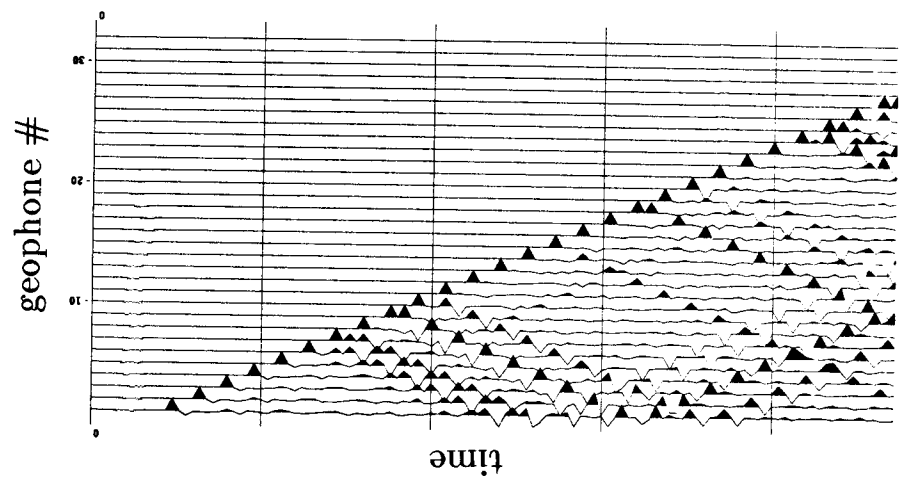

FIG. 4. A forward model using the inverted impedance of Figure 3. The amplitude clip is low to emphasize the character of errors.

a linearized system of *constant* reference impedance, then performing a line search for a value of comparable magnitude. For real data, the approximation was correct within a factor of two, but varied with the reliability given to the extracted events.

Adjusting for geophone amplification

Inversions of real data quickly showed that geophone coupling varied significantly and unpredictably. The data set of Figure 7 revealed that amplitudes could vary as much as 30% for equivalent events in neighboring traces. Inverted impedances were obliged first to accommodate these variations and only then could attempt to model the reflected energy. For this reason an additional physical parameter amplifying entire traces seemed desirable. After forward modeling $\sigma(x)$ and $g(t)$ with equations (1-3), I scaled each trace by independently determined constants $a(x)$.

$$y'(x,t) = a(x) \cdot y(x,t) \tag{16}$$

The same scaling must be applied before calculating the adjoint. I initialize $a(x) \equiv 1$.

I calculate $a(x)$ only after full perturbation of $\sigma(x)$ and $g(t)$. The first iteration attempts to model direct, down-going energy by perturbation of $g(t)$; $\sigma(x)$ remains constant until the reference wavefield is non-zero. The most important change of amplitudes in direct arrivals is due to irregular geophone amplifications rather than transmission losses. Subsequent inversions of reflected events overwhelmingly determines the perturbations of impedance; geophone amplification must adjust accordingly. At the end of each iteration I calculate the amplification $a(x)$ which makes the forward model resemble the data best in a least-squares sense. I assume a large variance of 10^2 for $a(x)$ and a variance for the noise equal to that of the data. The linear l.s. system requires only a few dot products for computation.

A ROBUST VSP INVERSION

Again, I refer readers to the paper of Harlan and Lailly (1984) for explanation of the statistical tools used below. I mention the function of each tool but omit algorithmic details involving the manipulation of histograms.

First I note the position of statistical calculations in the algorithm and then elaborate.

1. Calculate first-arrival times as the travel-time depths of the geophones.
2. Initialize $\sigma(x) \equiv 1$, $g(t) \equiv 0$, $a(x) \equiv 1$. Assume C_n^2 and C_g^2 equal the variance of the data samples, and $C_\sigma^2 \approx .1$

3. Calculate gradients (11) and (12) for $d\sigma(x)/dx$ and $g(t)$.
4. Estimate the signal and noise pdf's for the impedance gradient.
5. Apply the Bayesian estimator of signal to the impedance gradient.
6. Accept only those impedance perturbations containing a sufficiently low percentage of noise with a sufficiently high probability. Accept only the most reliable perturbations if too many satisfy this criterion. Stop if no innovations are acceptable.
7. Calculate the appropriate scale factor for the perturbations.
8. Update $g(t)$ and $\sigma(t)$.
9. Update $a(t)$
10. Go to step 3.

First-arrival times are signal parameters to be inverted. They are routinely estimated in VSP processing. Since only *relative* arrival times are important for indexing the traces, I recommend estimations from cross correlations. Lost of high frequencies with depth will not affect the peaks of these correlations. All events other than first arrivals ought to be suppressed. L'Institut Français du Petrole provided the VSP of Figure 7 and the first-arrival times of Figure 8.

Note that multiplying $\sigma(x)$ by a constant and dividing $g(x)$ by the same does not affect the modeled data. I choose unity as a convenient reference impedance.

I do not attempt an extraction of signal for the source-waveform perturbations. Experiments show these to be sufficiently Gaussian to satisfy the assumption of functional (8).

Estimating signal and noise pdf's are discussed at length by Harlan and Lailly (1984). In brief, the form of functional (8) insures that the gradient of $d\sigma(x)/dx$ and $g(t)$ are linear functions of the residual uninverted events. Thus, if noise is additive to signal in the data, noise will be additive in the gradient. Assuming signal and noise to be statistically independent requires the actual pdf of the gradient to be a convolution of pdf's for the signal and noise, $p_d(\cdot) = p_s(\cdot) * p_n(\cdot)$. A histogram of the gradient estimates the pdf $p_d(\cdot)$. One may obtain a pessimistic estimate of the pdf for noise (overestimating) by randomly rearranging the samples of the residuals before recalculating the gradient. Noise statistics (assumed stationary) should remain the same; signal, however, will now destructively interfere after linear transformation, and behave as additional noise. A histogram of this noise gradient pessimistically estimates the noise pdf, $p_n(\cdot)$. A minimum cross-entropy deconvolution of histograms gives a pessimistic estimate of the signal pdf, $p_s(\cdot)$. Iterative subtraction of the most reliable events from the data will

reduce the effect of signal converted into noise and increase the reliability of the pdf estimates.

Next take the signal and noise pdf's for the impedance gradient and calculate the most probable value for the signal present in each sample. Let d be the amplitude of a given sample. Then the best (Bayesian) estimate of the signal amplitude for this sample should be

$$\hat{s} = E(s \mid d) = \int s \; p_{s \mid d}(s \mid d) ds = \frac{\int s \; p_s(s) \; p_n(d-s) \; ds}{p_d(d)} \tag{17}$$

Though we may say that we now have the most *probable* values of the signal in our perturbations, we have not yet determined *how* probable these are. I define the reliability of a given estimate $\hat{s}$ as the probability that the actual value is no more than a fraction c of d away. I decide to accept a given perturbation $\hat{s}$ if

$$1 - e \leq p[-cd \leq s - \hat{s} \leq cd \mid d] \tag{18}$$
$$= \frac{\int_{-cd}^{cd} p_s(\hat{s} - s) p_n(s) ds}{\int_{-\infty}^{\infty} p_s(\hat{s} - s) p_n(s) ds}$$

where both c and e are small numbers. For the extraction of Figure 6, $c = e = .05$. Additionally, one should not accept too many innovations in a given iteration. If many samples meet condition (18) we accept only the most probable.

The remaining steps of the algorithm are as previously described. I use the golden mean algorithm for the line search of step 7.

Figures 9 through 13 show the results of applying this algorithm to the data of Figure 7. Three iterations were used at 20 minutes of VAX 11/780 computer time per iteration. Note that uneven sampling of geophones with depth did not prevent an accurate inversion of most events. The tube wave, having a velocity inconsistent with the first arrival, was effectively ignored. The most reliable noise could be extracted and subtracted from the original data for an improved second inversion. Here, $c = .05$ and $e = .5$ in equation (18). Convergence was reached at the third iteration, a bit too rapid for sufficient simplicity in the inverted impedance function. Reducing the number of perturbations allowed in an iteration would simplify the impedance function and slow convergence.

Impedance depths not spanned by the VSP direct arrival were not perturbed by the inversion. The impedance must be fixed for these depths before late reflections can

dominate the gradients.

ACKNOWLEDGMENTS

Thanks to Patrick Lailly of L'Institut Français du Petrole for his patient explanation of the application of control theory to the inversion of VSP's. Thanks also to Danielle Macé of L'I.F.P. for her detailed advice on the discretization and programming of such inversions. Thanks to L'IFP for accommodating me for six months and for providing the real VSP data.

REFERENCES

Bamberger A., Chavent G., and Hemon C., and Lailly P., 1982, Inversion of normal incidence seismograms: Geophysics, v. 47., p. 757-770.

Harlan W., and Lailly P., 1984, Robust inversion of non-linear transformations with an application to VSP's, SEP-38.

Lions J.L., 1968, Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles: Paris, Dunod-Gauthier Villars (Engl. trans., 1972, New York, Springer-Verlag).

Macé, D., and Lailly, P., 1984, A solution of an inverse problem with the 1D wave equation applied to the inversion of vertical seismic profiles, Sixth international conference on analysis and optimization of systems: Nice, Springer Verlag.

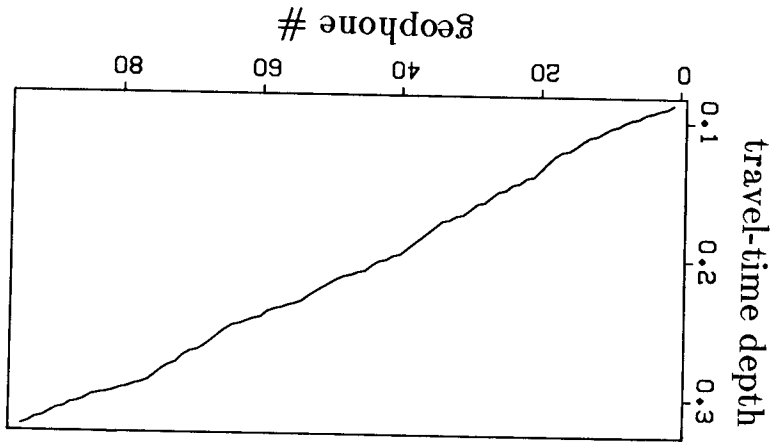

FIG. 8. First-arrival times indicate the travel-time depth of each geophone. (Also provided by L'I.F.P.)

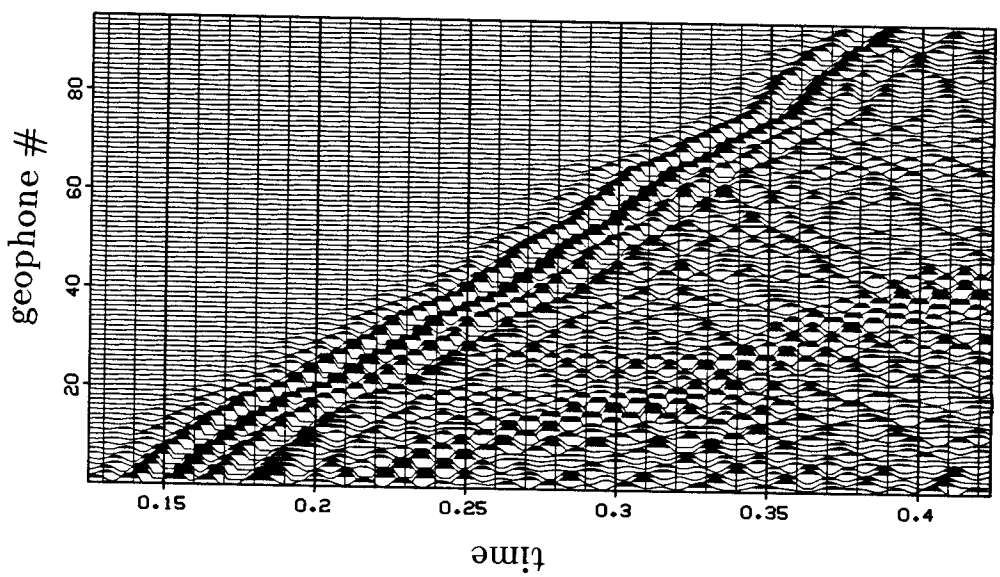

FIG. 7. A real VSP supplied by L'Institut Français du Petrole.

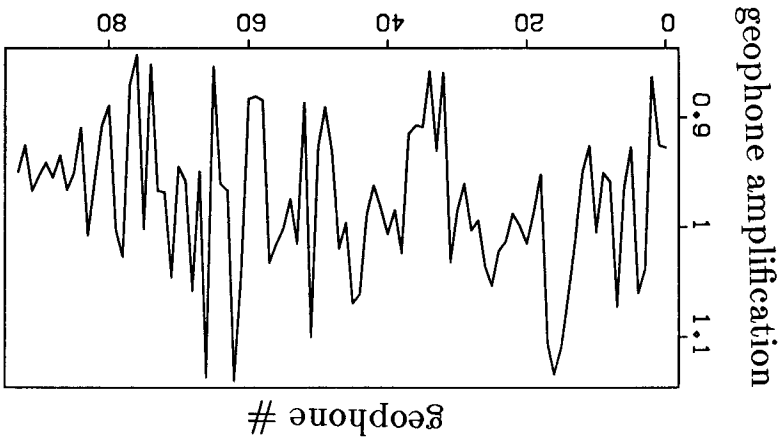

FIG. 11. The inverted geophone amplification after three iterations.

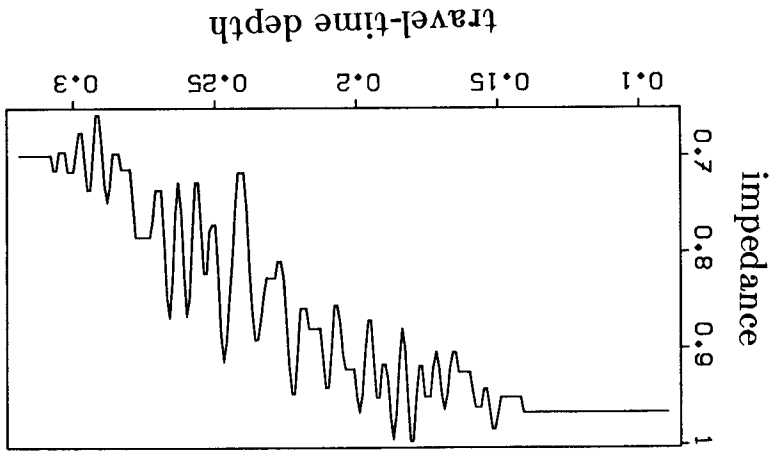

FIG. 10. The inverted impedance function after three iterations. Robust extractions were used to capture most important events.

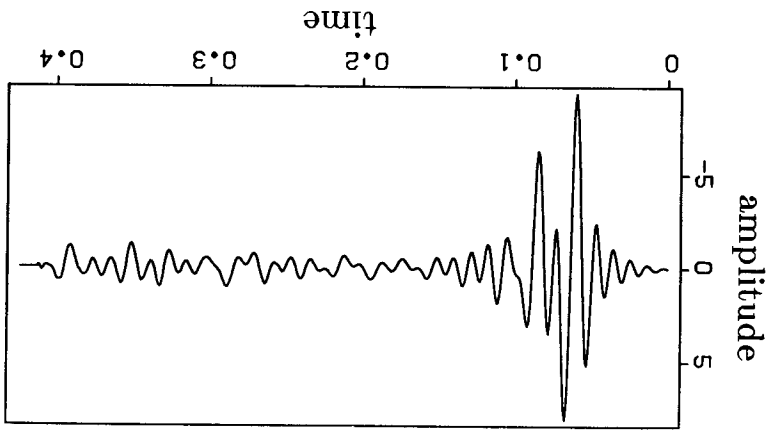

FIG. 9. The inverted source waveform after three iterations.

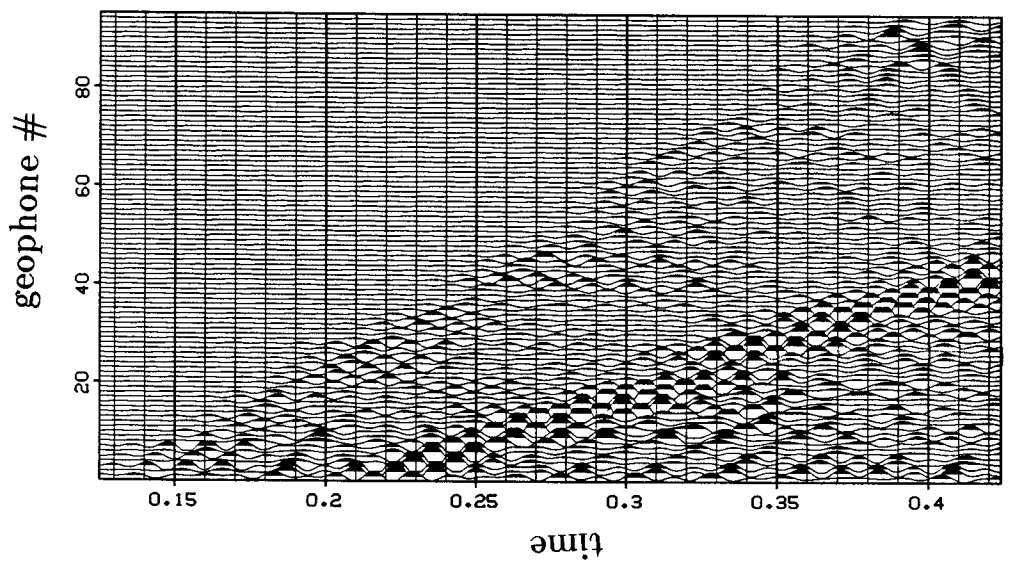

FIG. 13. Events remaining after subtraction of Figure 12 from 7. The tube wave was successfully ignored. Remaining reflections are weak, but could be successfully inverted after extracting and subtracting noise.

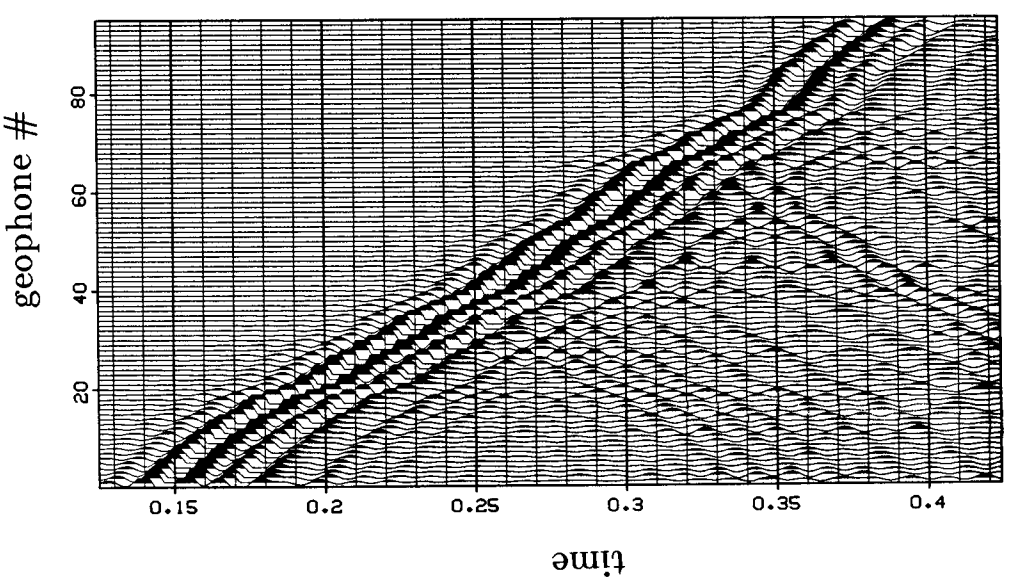

FIG. 12. Data modeled from information in Figures 8,9,10, and 11. Gain is identical to that of Figure 7.