Robust inversion of non-linear transformations with new notes on VSP's

William S. Harlan

MODEL SIMPLICITY

Let us begin with a simple illustration of how poorly seismic data may determine a geophysical model, even when redundancy is high. Figures 1, 2, 3, and 4 display two impedance functions and corresponding synthetic vertical seismic profiles (VSP's). Three other one-dimensional functions were used for the modeling -- a source waveform, geophone depths, and geophone amplification. See SEP-41, p. 283 for a fuller explanation of the forward transform. Figure 5 plots the subtraction of Figure 4 from 2, at the same scale. Differences in the modeled events are scarcely visible. The impedance functions, though showing some correlation, vary considerably. The difference between the two essentially belongs to the non-linear analog of a "null space": it does not greatly affect the modeled section. Subjectively, Figure 3 is easier to interpret than 1. What makes one impedance function appear simpler? It has fewer "events" -- in this case, fewer large non-zero samples in the first derivative.

Define an event as a physical feature statistically independent of other features, with a distinguishable expression in the data. A collection of events should easily model features of comparable importance in the data, excluding noise. We want iterations that converge on the most important events first since the reliability of later perturbations of the model depends on the reliability of the earlier.

Numerically, we may see the difference in the non-Gaussianity of the two impedance functions. Figures 6 and 7 show histograms corresponding to the first derivatives of these two functions. Also plotted are the best fitting generalized Gaussians and the best fitting Gaussians. See SEP 41, p.405 for the algorithm. Equation (5) shows the form of a generalized Gaussian. The "simpler" log of Figure 3 gives an exponential power (α) of 0.663, and Figure 1 of 2.15. Figure 1, which happens to be a least-squares inversion of a VSP, shows L2 or Gaussian statistics. Figure 3, from a more robust inversion to be derived in this paper, shows very non-Gaussian statistics, with a high-number of small values and sparse strong events.

THE MAP ESTIMATE

Let the data be a random process defined as a sum of noise and non-linearly transformed signal. Define signal $\overline{s}$ so that samples can be regarded as events; samples should be statistically independent.

$$d_i = f_i(\overline{s}) + n_i \quad (1)$$

$$or \ \overline{d} = \overline{f}(\overline{s}) + \overline{n}$$

We define $\overline{s}$ and $\overline{n}$ as focused, stationary random processes (random vectors).

We define geophysical noise as that untransformed component showing no spatial coherence (we allow some temporal coherence). If a component possesses significant coherence, then it should properly be defined after another transform, as a second variety of signal.

Let $p_{s_i}(\cdot)$ and $p_{n_i}(\cdot)$ be the corresponding marginal probability functions (mpf's). An mpf is defined by the following. The probability that random variable y will have an amplitude in the range $y - \Delta y \leq y \leq y + \Delta y$ is

$$p\left[y - \Delta y \leq y \leq y + \Delta y\right] = \int_{y=y-\Delta y}^{y=y+\Delta y} p_y(x)dx$$

x is a dummy variable.

The MAP inverse is the most probable $\overline{s}$ a given $\overline{d}$. (MAP abbreviates *maximum a posteriori,* so called because one assumes knowledge of the final transformed result.) We maximize the following conditional probability function

$$J_1(\overline{s}) = p_{\overline{s} \mid \overline{d}}(\overline{s} \mid \overline{d}) = \prod_i \frac{p_{s_i}(s_i)\, p_{n_i}[d_i - f_i(\overline{s})]}{p_{d_i}(d_i)} \quad (2)$$

The denominator merely normalizes, does not affect a maximization. Since the logarithm increases monotonically for positive functions, we may also maximize

$$J_2(\overline{s}) = \sum_i \ln p_{s_i}(s_i) + \sum_i \ln p_{n_i}[d_i - f_i(\overline{s})] + constants \quad (3)$$

Specific forms of $p_s(\cdot)$ and $p_n(\cdot)$ may simplify the form of the above functional. For example, if the signal and noise mpf's are generalized Gaussians (with zero means),

$$p_{s_i}(x) = \frac{1}{C_1} e^{-\left[\frac{|x|}{\beta_1}\right]^{\alpha_1}} , \quad p_{n_i}(x) = \frac{1}{C_2} e^{-\left[\frac{|x|}{\beta_2}\right]^{\alpha_2}} , \tag{4}$$

then equation (3) yields a classic inversion functional with L^p norms:

$$J_3 = \frac{1}{\beta_1} \sum_i | s_i |^{\alpha_1} + \frac{1}{\beta_2} \sum_i | d_i - f_i(\bar{s}) |^{\alpha_2} \tag{5}$$

Note that a least-squares norm results in the Gaussian case, when α is 2.

For a given estimate $\bar{s}_0$ of the signal, the gradient of J_2 is

$$\frac{\partial J_2}{\partial s_i} \Big|_{\bar{s}=\bar{s}_0} = \frac{p_{s_i}'(s_i^0)}{p_{s_i}(s_i^0)} - \sum_j F_{ij}^0 \frac{p_{n_i}'[d_j - f_j(\bar{s}_0)]}{p_{n_i}[d_j - f_j(\bar{s}_0)]} \tag{6}$$

$$where \ F_{ij}^0 \equiv \frac{\partial f_j(\bar{s}_0)}{\partial s_i}$$

This perturbation increases the probability of the model, but not necessarily the reliability of events. If a non-zero event is only marginally more probable than a zero, let us choose the zero and simplify the picture for the interpreter. One can estimate $p_s(\cdot)$ and $p_n(\cdot)$ from similar logs by using the cross-entropy methods described in SEP-41, p.405. One should not allow these distributions to depend on the data being inverted, otherwise the result will be non-unique and possibly unstable.

A ROBUST ESTIMATE

Again, we want iterations which converge on the most important events first. The most important are the most reliable. Let us measure the extent to which our perturbations are influenced by noise and the extent to which signal components can be distinguished from noise.

To make the following calculations possible, let us assume that the p_{n_i} are Gaussian. These distributions correspond to residual noise. We will iteratively extract the non-Gaussian noise, as we did for the VSP, so this assumption will become increasingly accurate. The important change is that the gradient in equation (6) is now a linear function of the residual events.

$$J_3 = \sum_i \ln p_{s_i}(s_i) + \sum_i \frac{1}{2\sigma_i^2}[d_i - f_i(\bar{s})]^2 \tag{7}$$

$$\frac{\partial J_3}{\partial s_i} \Big|_{\bar{s}=\bar{s}_0} = \frac{p_{s_i}'(s_i^0)}{p_{s_i}(s_i^0)} - \sum_j F_{ij}^0 \frac{1}{2\sigma_{n_i}^2} [d_j - f_j(\bar{s}_0)]$$

Define the following random variables as the residual signal, noise, and data:

$$\overline{s}\,' \equiv \overline{f}\,(\overline{s}\,) - \overline{f}\,(\overline{s}_0) \;\; ; \;\; \overline{n}\,' \equiv \overline{n} - \overline{n}_0 \;\; ; \;\; \overline{d}\,' \equiv \overline{s}\,' + \overline{n}\,' \tag{8}$$

$\overline{n}_0$ is previously extracted noise. Write the gradient $\overline{d}\,''$ as a single linear transform F of the residual data $\overline{d}\,'$.

$$d_i\,'' \equiv \sum_j F_{ij} \; d_i\,' + c_i \;\; ; \;\; \overline{d}\,'' \equiv F\overline{d}\,' \;\; ; \;\; \overline{s}\,'' \equiv F\overline{s}\,' \;\; ; \;\; \overline{n}\,'' \equiv F\overline{n}\,' \tag{9}$$

$$\overline{d}\,'' = \overline{s}\,'' + \overline{n}\,''$$

Each prime designates a transformation of a component away from the definition of equation (1).

Now let us estimate how much of a given sample of the gradient is only transformed signal. Let us postpone the estimation of the signal and noise mpf's. The Bayesian estimate of the signal, $s\,''$, in a sample of the linearly transformed data, $d\,''$, is

$$\hat{s}\,'' = E\,(s\,'' \mid d\,'') \tag{10}$$

$$= \int x \;\, p_{s''\,|\,d''}(x \mid d\,'')\,dx = \frac{\int x \;\, p_{s''}(x\,)\; p_{n''}(d\,''-x\,)\; dx}{p_{d''}(d\,'')}$$

We suppress sample subscripts.

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. 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. Accept a given perturbation $\hat{s}$ if

$$1 - e \le p\left[-cd\,'' \le s\,'' - \hat{s}\,'' \le cd\,'' \mid d\,''\right] \tag{11}$$

$$= \frac{\int_{-cd}^{cd} p_{s''}(\hat{s}\,'' - x\,)p_{n''}(x\,)dx}{\int_{-\infty}^{\infty} p_{s''}(\hat{s}\,'' - x\,)p_{n''}(x\,)dx}$$

where both c and e are small numbers.

SOME ADDITIONAL STABILITY

Some components of the MAP perturbation may be unreliable because the forward transform $\overline{f}$ largely destroys them. Another statistical simplification will allow us to suppress these sources of instability.

As we saw, the robust estimate requires only that the noise distribution be Gaussian. The signal distribution could remain arbitrary. Let us instead assume that residual signal is also Gaussian, and let us also temporarily replace the forward transform $\overline{f}$ by a linearization about $\overline{s}_0$. The functional changes accordingly.

$$f_i(\overline{s}_0 + \Delta\overline{s}) \approx f_i(\overline{s}_0) + \sum_j F_{ij}^0 \Delta s_j \tag{12}$$

$$J_4 = \sum_i \frac{1}{2\sigma_{s_i}^2} (s_i^0 + \Delta s_i)^2 + \sum_i \frac{1}{2\sigma_{n_i}^2} [d'_i - \sum_j F_{ij}^0 \Delta s_j]^2$$

Let $\overline{d} - \overline{f}(\overline{s}_0)$ be the residual data, $\overline{d}'$. Now repeated applications of the gradient (6) will remain a linear function of the noise. The cumulative perturbation $\overline{d}''$ would solve a least-squares inversion for the signal. Because of the linearity the robust estimation could still be applied.

This step becomes essential when unstable components completely obscure reliable events in the perturbations. This step will suppress poorly determined high frequencies that often plague the inversion of impedance functions.

ESTIMATING THE A PRIORI STATISTICS

We have already decided to make the *a priori* distributions Gaussian for the MAP estimate of equation (2). The standard deviation for the noise could be chosen pessimistically as equal to that of the data. The standard deviation for the signal is important for stability in the MAP perturbation. Let us first choose some physical upper bound, and if the result is unstable, let us reduce it.

Much more critical are the choices for the distributions used in the robust estimate of equations (10) and (11). We prefer to estimate these mpf's directly from the data by observing histograms of the residual data before and after linear transformation.

Stationarity

Let us first assume that the signal has a stationary dimension, so that enough redundancy exists for histograms to approximate mpf's. *A priori* statistics should reflect regional possibilities. Knowledge of one reliable event should increase the likelihood of finding such another event nearby. Thus, one not only expects but desires that estimated mpf's change slowly over spatial dimensions and time. Because of this stationarity, a histogram prepared from a great many samples with identical mpf's will describe the possibilities open to them all.

Likewise we can usually find some dimension over which we expect noise events to be equally likely and essentially similar. When we later drop subscripts from signal and noise distributions, we are assuming stationarity over some dimension.

Pessimistic estimates of distributions

We shall need the following relations between random variables, the z's and y's, and their corresponding mpf's. a is a constant, and x a dummy variable.

$$z = y + a \rightarrow p_z(x) = p_y(x - a) \; ; \tag{13}$$

$$z = a \cdot y \rightarrow p_z(x) = \frac{1}{a} p_y(x/a) \; ;$$

$$z = y_1 + y_2 \rightarrow p_z(x) = p_{y_1}(x) * p_{y_2}(x)$$

The star indicates convolution. Because we assumed $\overline{n}$ to be focused (samples statistically independent), equation (9) requires that

$$p_{n_i''}(x) = \prod_j * \left[\frac{1}{F_{ij}} p_{n'}\left(\frac{x - c_i}{F_{ij}}\right)\right] \tag{14}$$

The $\prod *$ indicates multiple convolutions. In many applications, including that of the VSP, the c_i's will change slowly enough to preserve local stationarity. We will then be able to suppress subscripts on mpf's.

Define an exaggerated estimate of $p_{n_i''}(x)$ by assuming all residual data are noise, equivalently by ignoring the coherence of any signal.

$$\hat{p}_{n_i''}(x) \equiv \prod_j * \left[\frac{1}{F_{ij}} p_{d'}\left(\frac{x - c_i}{F_{ij}}\right)\right] = p_{n_i''}(x) * \prod_j * \left[\frac{1}{F_{ij}} p_{s'}\left(\frac{x - c_i}{F_{ij}}\right)\right] \tag{15}$$

This mpf must overestimate the transformed noise and all positive moments. If the data contain no signal, then the estimate is perfect (the signal mpf becomes a delta function). Estimate (15) easily by generating a random, focused array with the same mpf's as the data, transforming with the same linear transformation, and taking local histograms. Because the signal and noise remain statistically independent and additive after transformation, choices of their mpf's determine that for the data:

$$p_{d''}(x) = p_{s''}(x) * p_{n''}(x) \tag{16}$$

With the assumption of local stationarity, estimate $p_{d''}(x)$ from local histograms of the transformed data. The divergence of the estimate from the *a priori* mpf in equation (16) should be minimal. Measure this divergence with the directed divergence (cross entropy) of Kullback. Minimizing $\int p_1(x) \ln[p_1(x)/p_2(x)]\,dx$ minimizes the "unpredictability" of

$p_1(x)$ assuming $p_2(x)$ to be the most predictable. Iteratively discover the best estimate of $p_{s''}(x)$, given $p_{d''}(x)$ and $p_{n''}(x)$, by minimizing the following (suppressing primes)

$$J_5[p_s(x)] = \int p_d(x) \ln[p_d(x) / \int p_s(x-y) p_n(y) dy] dx \tag{17}$$

$$+ \frac{\lambda_1}{2}[\int p_s(x) dx - 1]^2 + \frac{\lambda_2}{4}\int [p_s(x) - |p_s(x)|]^2 dx$$

Add two Lagrange multipliers for the constraints of unit area and of positivity. To calculate the gradient of J_5 with respect to each point of the function $p_s(x)$, perturb a previous estimate with a delta function: $p_s(x) + \epsilon\delta(x - x_0)$ and differentiate.

$$\frac{\partial}{\partial \epsilon} J_5[p_s(x) + \epsilon\delta(x - x_0)] \tag{18}$$

$$= -\int \frac{p_d(x)}{\int p_s(y) p_n(x-y) dy} p_n(x - x_0) dx$$

$$+ \lambda_1[\int p_s(x) - 1] + \lambda_2[p_s(x_0) - |p_s(x_0)|]$$

Iteratively perturb $p_s(x)$ with the negative of this gradient; an inexpensive line search finds the correct magnitude. The constraints easily determine the proper values of λ_1 and λ_2 for any magnitude of perturbation. The second term equally raises or lowers all points of $p_s(x)$ until the constraint of unit area is satisfied. The third term moves each point a sufficient positive distance to remove any negative excursions. The first term divides the estimate $p_d(x)$ by the *a priori* value and cross correlates with a shifted noise distribution, contributed by the perturbation of $p_s(x_0)$. The cross correlation thus identifies where the divergence is not uniform and compensates with appropriate perturbations.

EXTRACTING NON-GAUSSIAN NOISE FROM VSP'S

Figure 8 contains a portion of a VSP provided by L'Institut Français du Petrole. This section contains considerable Gaussian noise. A strong tube wave however acts as strong additive non-Gaussian noise that, because it violates the assumptions of equation (7), must distort the corresponding perturbations of the model. To circumvent this problem let us iteratively extract the most reliable signal and noise from the data.

I use the differential system of SEP-41, p.283 as a physical model to find an inversion for the most reliable signal, whose synthetic data appear in Figure 9. Subtracting Figure 9 from 8 leaves residuals containing noise and the least reliable signal (Figure 10). To find a pessimistic estimate of the mpf of the remaining signal, I find the synthetic VSP a least-squares inversion of the residuals and take a histogram. I then extract the samples of these residuals containing greater than 96% noise with greater than 50% probability (Figure 11). Figure 12 shows the remaining Gaussian noise (Figure 10 minus 11). Subtracting the most reliable noise from the original data (Figure 13) will insure that remaining noise is much more Gaussian.

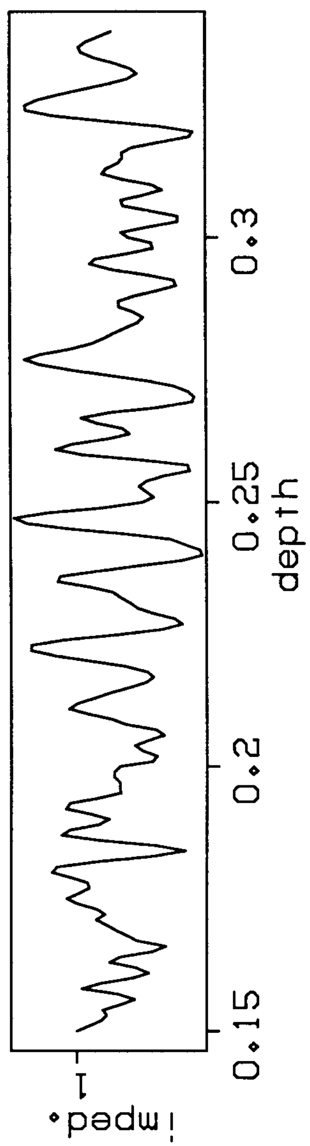

FIG. 1. This impedance function appears very Gaussian, with poorly recognizable events.

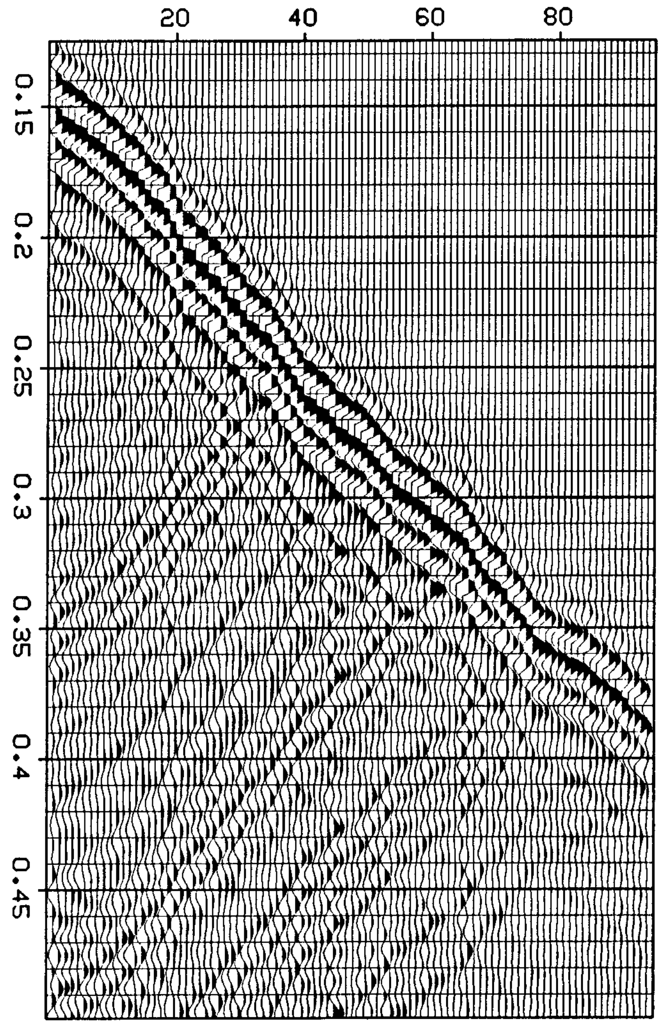

FIG. 2. A synthetic VSP using the impedance function of Figure 1.

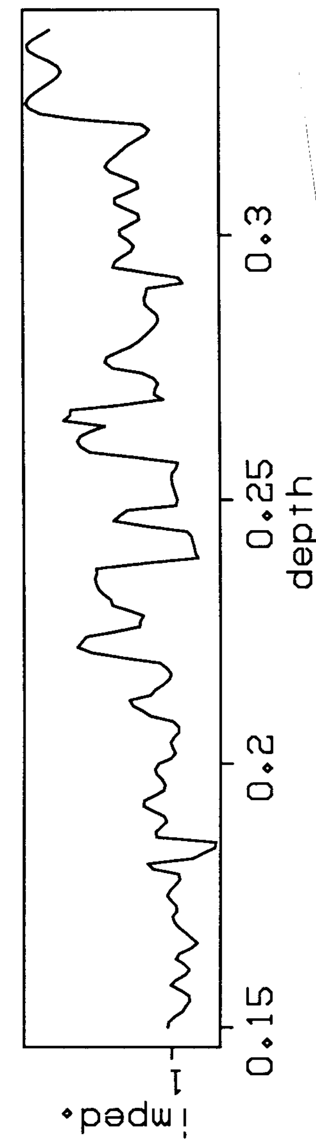

FIG. 3. This impedance function, though correlated with Figure 1, appears very different. Events are much more distinct and sparse.

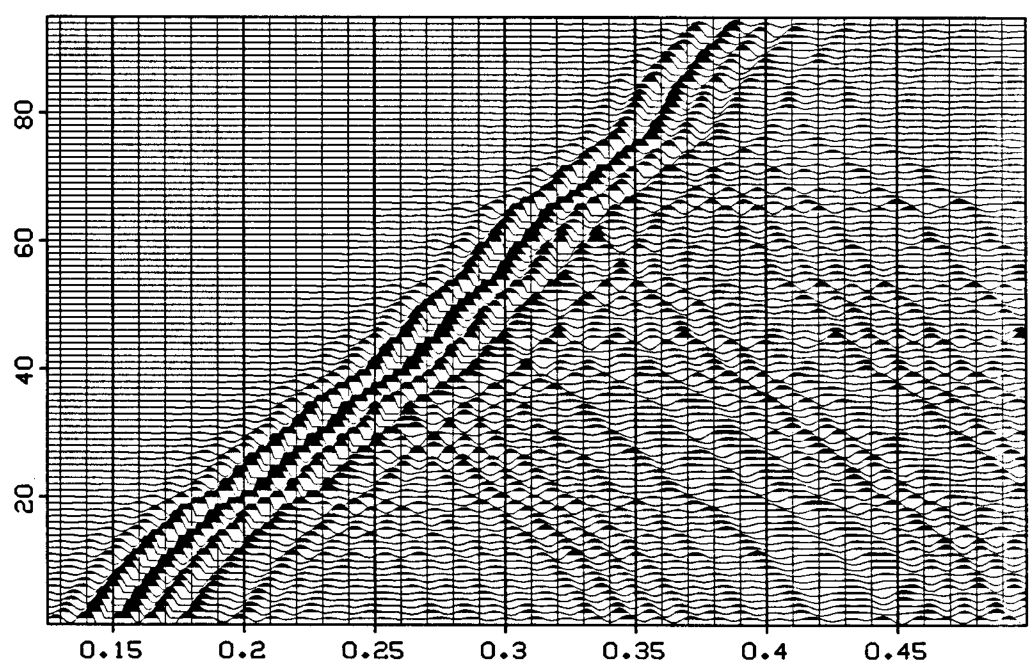

FIG. 4. A synthetic VSP using Figure 3 appears indistinguishable from Figure 2.

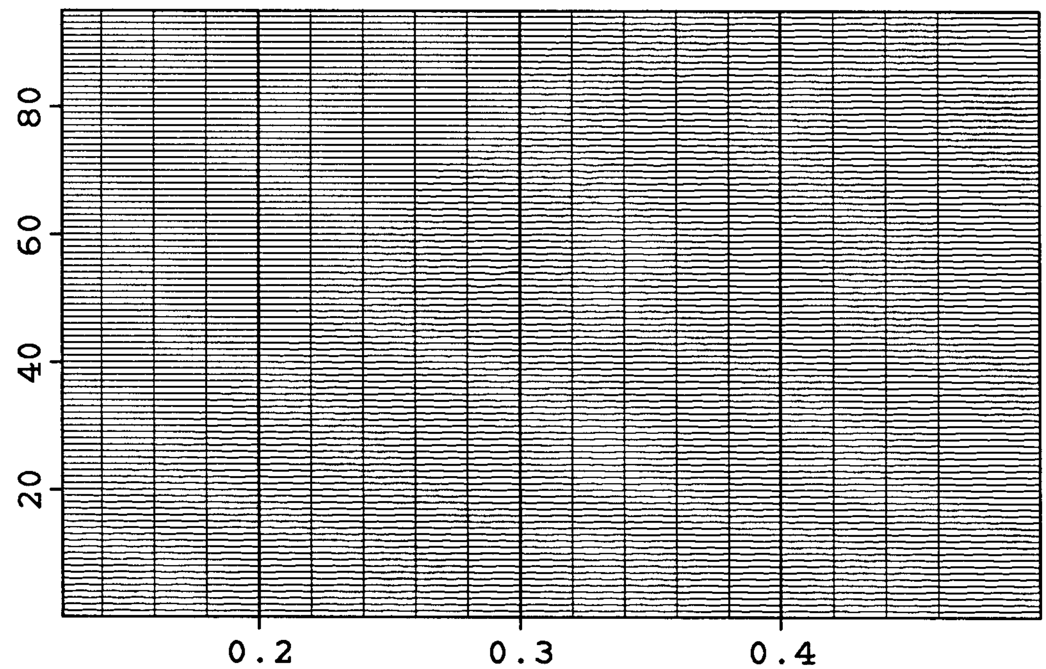

FIG. 5. Figure 2 minus 4 shows barely distinguishable differences when plotted at the same scale.

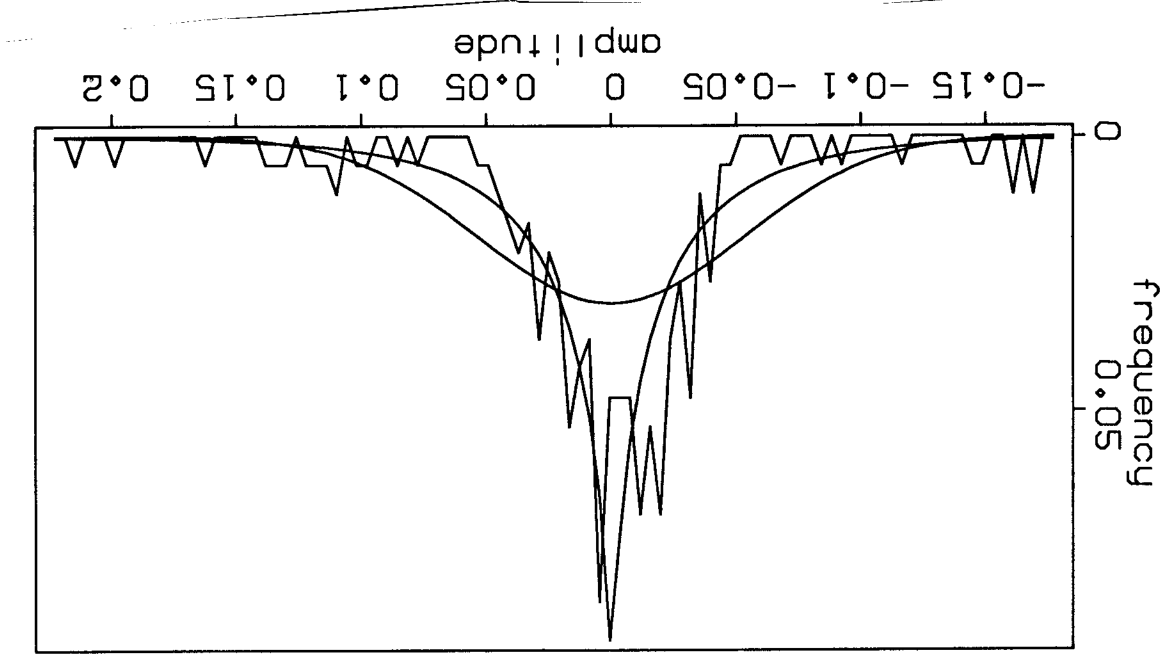

FIG. 7. A histogram of the first derivative of Figure 3. The best fitting generalized Gaussian (α=0.663) is quite far from the best fitting Gaussian.

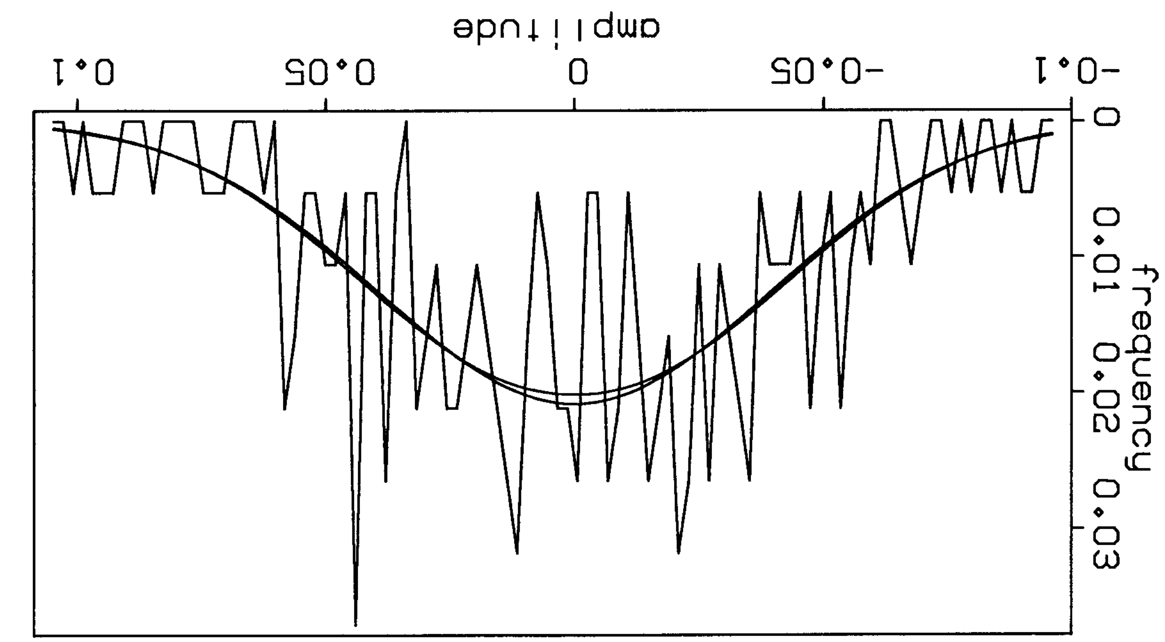

FIG. 6. A histogram of the first derivative of Figure 1. Also plotted are the best fitting generalized Gaussian (α=2.15) and the best fitting Gaussian (α=2).

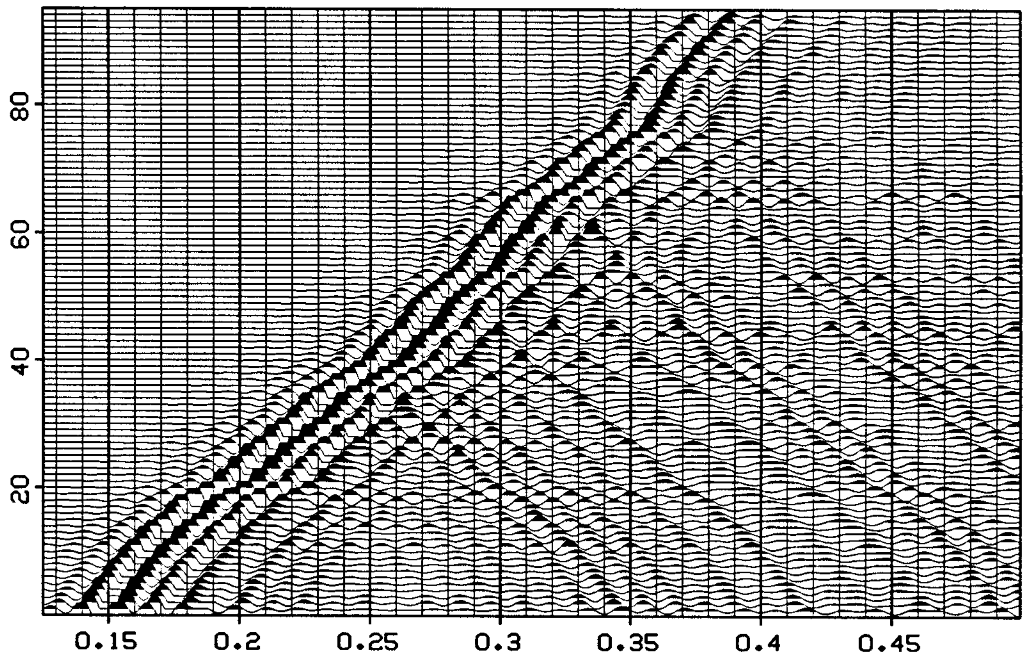

FIG. 9. A model of a robust inversion of the signal.

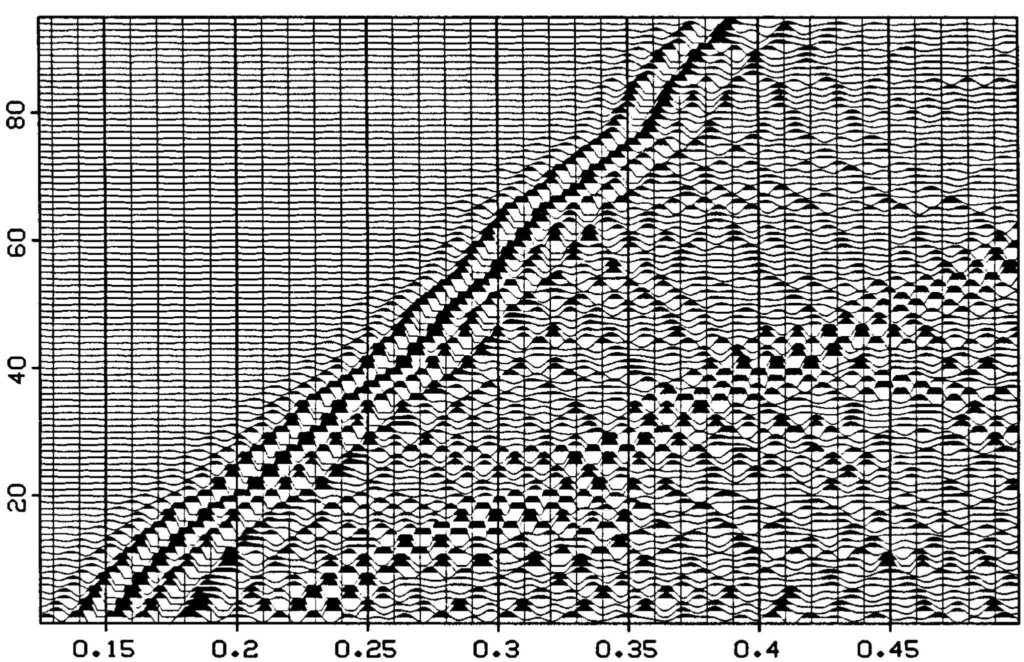

FIG. 8. A VSP provided by L'Institut Français du Petrole.

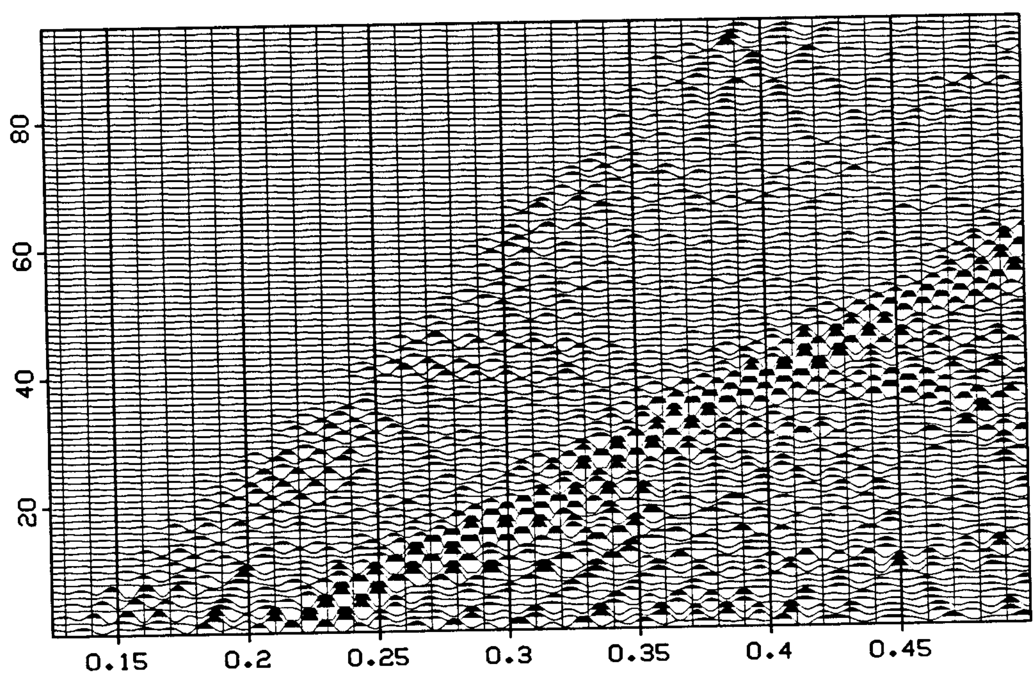

FIG. 10. Residual, uninverted events (Figure 8 minus 9).

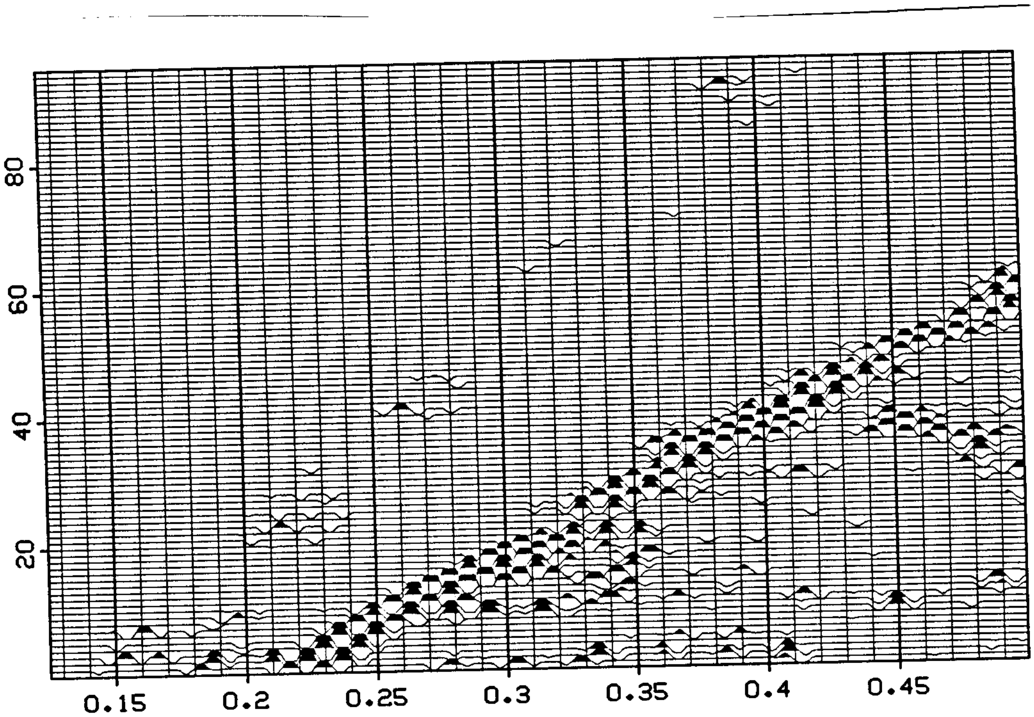

FIG. 11. The most reliable estimate of noise.

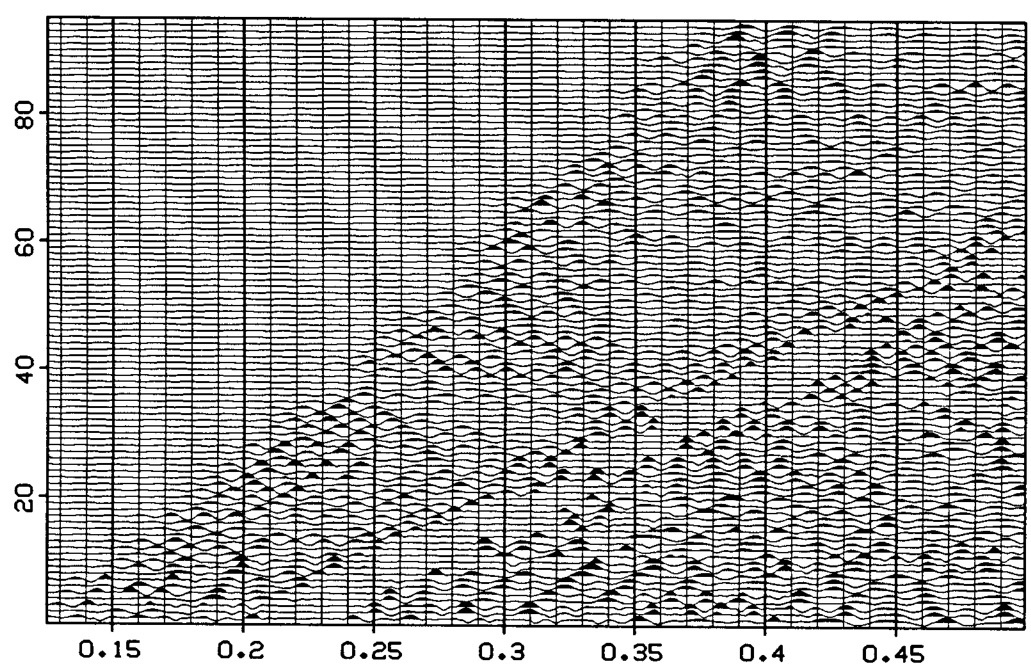

FIG. 12. The remaining Gaussian noise (Figure 10 minus 11).

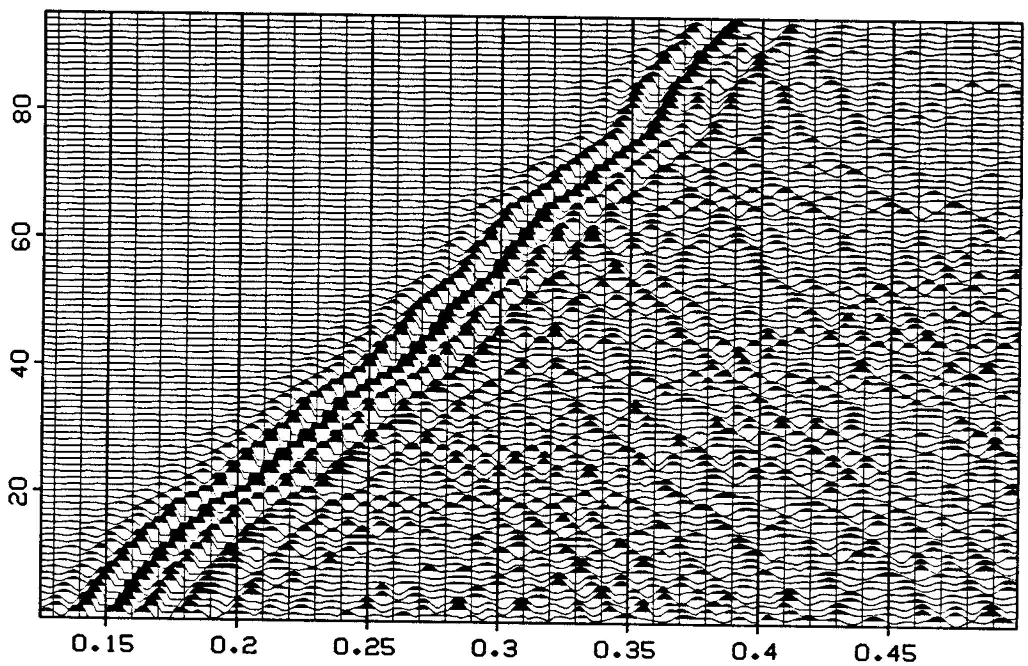

FIG. 13. The original data minus the most reliable noise (Figure 8 minus 11).