Signal/noise separation with slant stacks and migration

William Harlan

Abstract

Geologic events strongly color a seismic section spatially. Removable noise is spatially white and non-gaussian (parsimonious). Transformations, such as slant stacks and migration, exist which make geology parsimonious. To estimate a parsimonious signal from a gaussian background, one should zero low amplitudes, which are not parsimonious. One should estimate geology after transformation and noise before--removing one from the data before estimating the other. Estimations with a slant stack separate real data into geology and noise very well. Similar existing algorithms (e.g. using slant stacks to prevent aliasing) do not make local estimations, thereby failing to remove noise or introducing artifacts.

Recognizing geologic color with slant stacks and migration

Geologic events on seismic sections show strong spatial color for two physical reasons: 1) geological time events may be represented as a superposition of hyperbolas because of the wave equation; and 2) earth reflectors are layered. Only geology contains color. Noise includes missing traces, improperly amplified traces, hardware noise, and cultural noise--any event unconfirmed by other measurements. We do not include multiples and sideswipe because their statistical sources are similar to that of desired primaries.

Because of the geologic color, a transformation exists which makes geology white and more non-gaussian (more parsimonious). Such a transformation makes noise colored and more gaussian. With any approximation to the best transformation (which depends on the rocks themselves), we may estimate all geology containing color recognized by the transformation. Ideally one prefers an adaptable transformation suited to the data at any interactive step. Migration and slant stacks suit most data sets sufficiently well.

Migration with the wave equation removes some color by inverting wave propagation. Geologic hyperbolas map to single points; noisy spikes become weak semicircles. Much color remains, however, because reflectors are very predictable. A slant stack of a seismic section, defined below, maps lines of all dips into points.

$$Q(p,t) \equiv \int_{-\infty}^{+\infty} P(y,t+py)dy$$

The integration is over the spatial dimension (midpoints for CMP stacks). Slant stacks remove the first order spatial color of events--faults and horizon curvature being higher order. Since curvature is predictable, geology will show some color over p values. Most geology is very limited in slopes and p values; noise occupies all p's. Slant stacks make geology much more parsimonious than do migrations and slightly less than both together. Migration alone removes an insufficient amount of color for a memory-less estimation of geology. Neither migration or slant stacks can be completely inverted, but adjoint transformations (diffraction and inverse slant stacks) can restore all geological information in the section. One need lose only high, non-physical dips.

Deriving an ideal separation

Let L represent a forward transformation chosen to make geological events more parsimonious. A decomposition of the given section, the array $\bar{d}$, becomes

$$\bar{d} = L^{-1}\bar{g} + \bar{n}$$

where L^{-1} is the adjoint of L, $\bar{g}$ the expected array of transformed geology, and $\bar{n}$ the expected array of untransformed noise. $\bar{d} - \bar{n} = L^{-1}\bar{g}$ is the desired section, without noise.

Before any decomposition one must estimate the statistics of $\bar{g}$ and $\bar{n}$ as probability density functions (p.d.f.'s) $p_n(\bar{n})$ and $p_g(\bar{g})$. With such information our best estimate of $\bar{g}$ (by subtraction implying a best $\bar{n}$) is a maximum a posteriori estimate (MAP estimate):

I. A MAP estimate finds the most probable $\bar{g}$ *given* $\bar{d}$, $p_n(\bar{n})$, *and* $p_g(\bar{g})$.

Another estimation method shall be called the generalized minimum entropy method:

II. The minimum entropy method solves for the $\bar{g}$ *from which one can estimate a most ordered (most predictable)* $\hat{p}_n(\bar{n})$ *and* $\hat{p}_g(\bar{g})$.

The entropy of a p.d.f. of an array $\bar{x}$ is defined as

$$-\int p(\bar{x}) \ln p(\bar{x}) d\bar{x}$$

Entropy measures the predictability of the array values. Method II is inherently iterative:

estimations of $\bar{g}$ imply better estimations of $\hat{p}_n(\bar{n})$ and $\hat{p}_g(\bar{g})$, implying better estimations of $\bar{g}$. $\bar{g}$ and $\bar{n}$ must possess redundant statistics in order to estimate the corresponding p.d.f.'s. By choosing L well, one assumes that the elements of $\bar{n}$ and $\bar{g}$, call them n_i and g_i, are not colored. Statistical independence implies

$$p_g(\bar{g}) \equiv \prod_i q_i^g(g_i) \; ; \quad p_n(\bar{n}) \equiv \prod_i q_i^n(n_i) \tag{1}$$

Events have no a priori most preferable distribution in $\bar{g}$ or $\bar{n}$, so all samples in one domain share the same p.d.f.'s. We further simplify

$$q_i^g(g_i) \equiv q_g(g_i) \; ; \quad q_i^n(n_i) \equiv q_n(n_i) \tag{2}$$

We may assume equations (1) and (2) successfully if L is well chosen. When array elements are independent, they must be as parsimonious as possible to satisfy II.

III. If $\bar{g}$ and $\bar{n}$ are uncolored, the minimum entropy estimate of $\bar{g}$ requires $\bar{g}$ and $\bar{n}$ to be as parsimonious as possible.

The elements of the best $\bar{g}$ (implying an $\bar{n}$) represent many samples from $q_g(g_i)$ and $q_n(n_i)$, so

IV. We can estimate $p_n(\bar{n})$ and $p_g(\bar{g})$ from a best $\bar{g}$.

Lastly, equations (1) and (2) allow us to say

V. A MAP and a minimum entropy estimate of $\bar{g}$ are equivalent.

The powerful generality of I may be achieved by the easily applied III. To prove V we begin with the explicit expression of a MAP estimate of $\bar{g}$.

$$\max_g \left\{ p(\bar{g} \mid \bar{d}) \right\} = \max_g \left\{ \frac{p(\bar{d} \mid \bar{g})\, p(\bar{g})}{p(\bar{d})} \right\}$$

$$\rightarrow \max_g \left\{ p_n(\bar{d} - L^{-1}\bar{g})\, p_g(\bar{g}) \right\} \rightarrow \min_g \left\{ -\ln p_n(\bar{d} - L^{-1}\bar{g}) - \ln p_g(\bar{g}) \right\}$$

We have used Bayes rule at the second step. Because of equations (1) and (2) the above is equivalent to

$$\rightarrow \min_g \left\{ \sum_i -\ln q_n[(\bar{d} - L^{-1}\bar{g})_i] - \sum_i \ln q_g(g_i) \right\} \tag{3}$$

where $n_i \equiv (\bar{d} - L^{-1}\bar{g})_i$. By statement IV we can calculate the estimated p.d.f.'s $\hat{q}_n$ and $\hat{q}_g$ from the $\bar{g}$ which minimizes (3). Quantity (3) is equal to

$$N_n \int_{-\infty}^{+\infty} -\hat{q}_n(n_i) \ln q_n(n_i) dn_i \;+\; N_g \int_{-\infty}^{+\infty} -\hat{q}_g(g_i) \ln q_g(g_i) dg_i$$

N_g and N_n are the number of samples in the arrays. We no longer sum over elements of the best $\bar{g}$ and $\bar{n}$. We integrate over all possible values of n_i and g_i. If our original choices of $q_g(g_i)$ and $q_n(n_i)$ were good, then $\hat{q}_g \approx q_g$ and $\hat{q}_n \approx q_n$ and the above minimized quantity equals the minimum entropy of the geology and noise p.d.f.'s.

Choosing a decomposition algorithm

After our chosen transformation the highest amplitude events are parsimonious. By selecting high amplitude events and zeroing the low, we obtain a parsimonious section, therefore a good estimate of our geology (by *III*). We shall see later how this high amplitude selection derives from the statistics of the geology and noise. Knowing a crude histogram of g_i and n_i (or the approximate number and magnitude of both events) determines a cutoff. We may always choose a cutoff conservatively high.

Ideally one prefers to eliminate colored noise overlying the extracted geology. One cannot use a prediction error filter because some color remains in the geology and because few adapt quickly enough to avoid biasing dips and frequencies. At the first iteration, however, one need only recognize the strongest geology. Noisy energy diffuses throughout the $p-t$ domain, so the amount of overlapping energy is small.

Next we inverse transform the strong geology and subtract it from the original data. Now strong, parsimonious noise dominates weak, diffuse geology. We estimate noise by selecting high amplitudes and subtract it from the original data. We may transform and iterate again.

Let us summarize an iteration of our decomposition:

1. *Forward transform the seismic section.*
2. *Estimate the strongest geology by zeroing small amplitudes.*
3. *Inverse transform this estimate and subtract from the original section.*
4. *Estimate the strongest noise by zeroing small amplitudes.*
5. *Subtract this estimate from the original section.*

One should subtract sections only in the original domain, otherwise noise will be lost by the inverse transformation. When estimating one should consider analytic envelope amplitudes, otherwise the troughs of high amplitude wavelets will deepen and alter the frequency content of events. We make our estimation somewhat dependent on neighboring time samples.

A sample decomposition

Figure 1 displays a CMP stack of a deep marine seismic survey over the Aleutian Arc region of the Pacific (courtesy of the USGS). The section ranges from 7 to 8 seconds, sampled at 4 ms and 100 m along midpoints. A hardware problem produced the short interrupting waveforms. Trace 40 was intentionally zeroed as missing data. The noise in this zeroed trace is the negative of the reasonable geology one expects. These features exemplify highly visible, spatially white noise.

We begin our decomposition by transforming the section to the slant stack domain (Figure 2). Color remains in the geology because of the original curvature. Barely visible linear events are noise, diffused through all p values. Since only physical p values appear here, the geology shows remarkable parsimony. Performing the slant stack in the frequency domain allowed the wrap-around to see dipping events in the corners of the section. Without an appropriate sinc-based interpolator, strong artifacts will appear.

Next we zero all samples with envelope amplitudes below 12% of the maximum. Figure 3 shows little change, though all non-overlapping noise is gone. The strongest 88% of geological amplitudes remain. We inverse slant stack this estimate (Figure 4), and subtract it from the original data for Figure 5. The noise events stand out from a weak background of colored geology. The 40th trace, as expected, contains the negative of the geology it hides.

To estimate noise we zero below 12 % of the maximum. The result, Figure 6, contains only spatially white events. We subtract the strongest noise from the original data (without removing any geological information). Where spikes or missing data stood, now stand the strongest geological events. To remove weaker noise, we use this section as the starting point of a new iteration. We lower cutoffs to 1% and arrive at improved estimates of the noise, in Figure 8, and of the geology, in Figure 9.

A calculation of the estimation function

We may derive a theoretical estimation function from the geology and noise statistics and compute it numerically for the sample decomposition. Let d'_i be a sample of transformed data, the sum of two components: $d'_i = g_i + n'_i$. Primes indicate that components are not in their original domains. We assume we know $q'_n(n'_i)$ and $q_g(g_i)$. The most probable value of g_i given d'_i is

$$E(g_i \mid d'_i) = \int_{-\infty}^{+\infty} g_i q(g_i \mid d'_i) dg_i = \frac{\int_{-\infty}^{+\infty} g_i q_g(g_i) q'_n(d'_i - g_i) dg_i}{q'_d(d'_i)} \qquad (4)$$

We have used Bayes rule in the last expression.

Let us calculate (4) for our sample data. We take the fully separated geology and noise in Figures 8 and 9 and compute from each an implied p.d.f. Figure 10 displays a histogram of the geology, and Figure 11 of the noise. The noise amplitudes cover a slightly greater range of values. The geologic components are slightly more gaussian. The differences are not enough to estimate noise first from the data.

After a slant stack the histograms are markedly different (Figures 12 and 13). The geology contains much higher amplitudes, and the noise is more gaussian. We compute a p.d.f. for the data (Figure 14) as a convolution of the geology and noise p.d.f.'s. Substituting into equation (4), we calculate $E(g_i \mid d'_i)/d'_i$ as a function of d'_i (Figure 15). The function reveals that values of d'_i above a low cutoff accurately estimate g_i. Lower amplitudes should be diminished because containing substantial noise. To keep only very probable geology and avoid any noise we zero these lowest samples. Performing an inverse slant stack and subtracting from our original data gave us Figure 5. Figure 11 still represents a histogram of the noise of this section because none has been removed. But the remaining geology is more gaussian and lower in amplitude (Figure 16). Using equation (4), we arrive at the estimation function in Figure 17. We can now estimate the noise well by discarding low amplitudes.

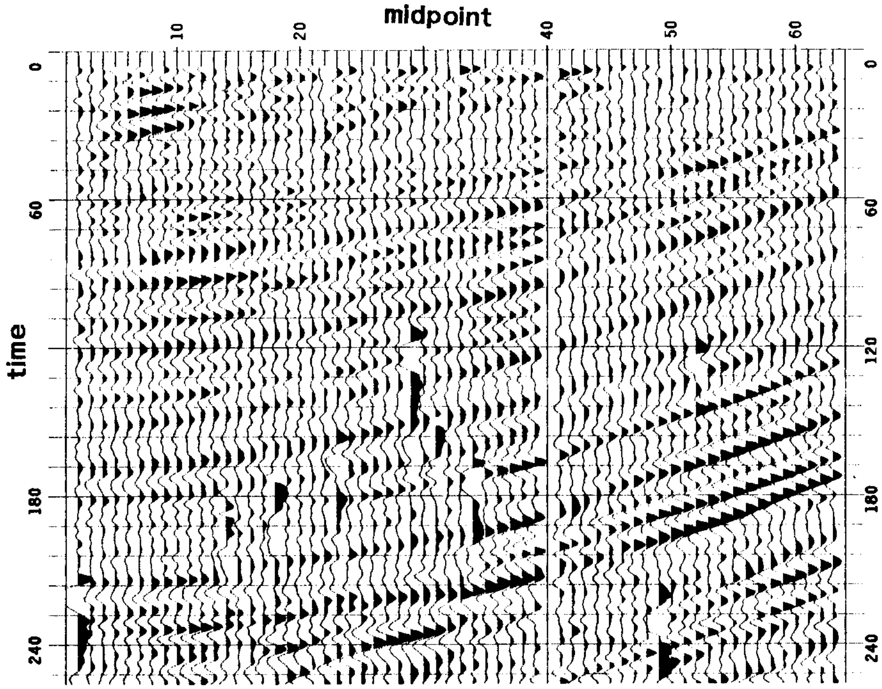
Other algorithms

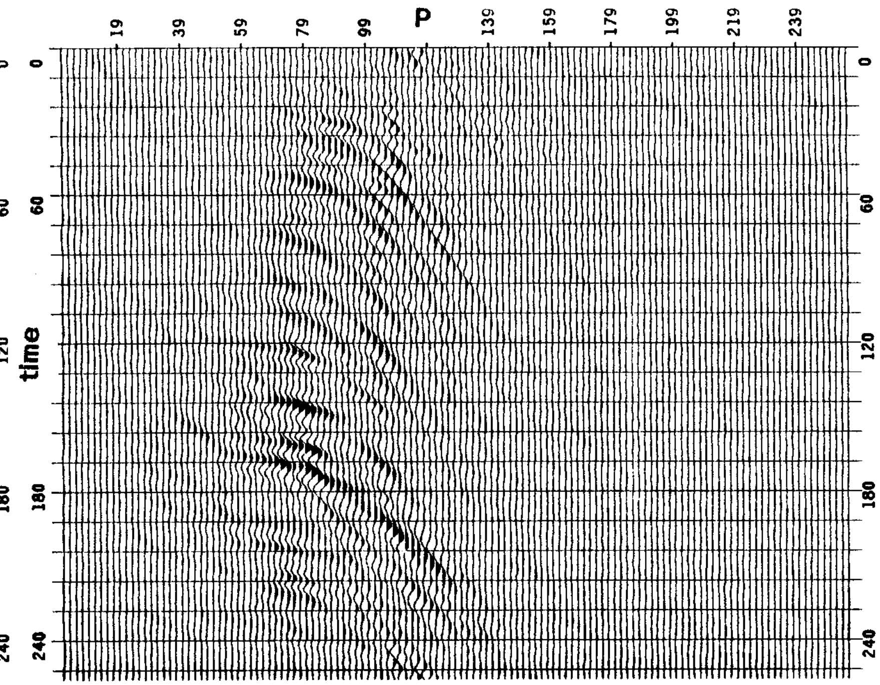
If one knows beforehand the location of missing traces, one can replace them after one iteration with the negative of extracted noise traces. Many attempting to fill in missing traces to prevent lateral aliasing will perform forward and inverse slant stacks, interpolating a greater number of traces on the return. Without a local operation in the $p-t$ domain one will see most of these missing traces again, with geology only weakly restored. Most who propose this cruder method restrict the range of p values narrowly around the highest amplitudes, thereby excluding most of the noise energy but introducing the artifacts of a sharply truncated dip filter. One needs a local operator in the $p-t$ domain.

REFERENCES

Godfrey, R., 1979, A stochastic model for seismogram analysis (dissertation): Stanford Exploration Project Report 17.

Van Trees, H., Detection, estimation, and modulation theory, v.1: New York, John Wiley & Sons.





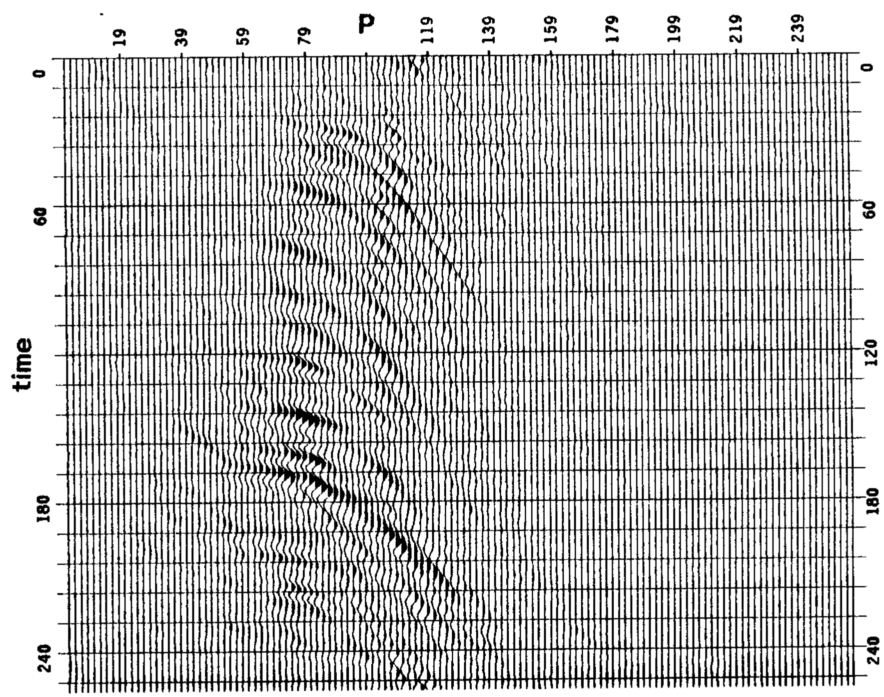
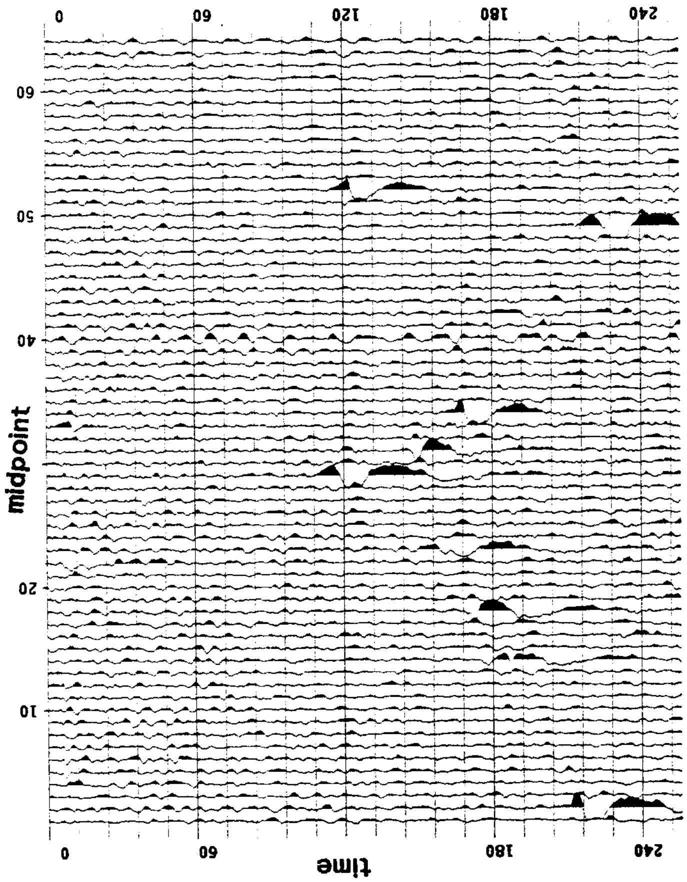
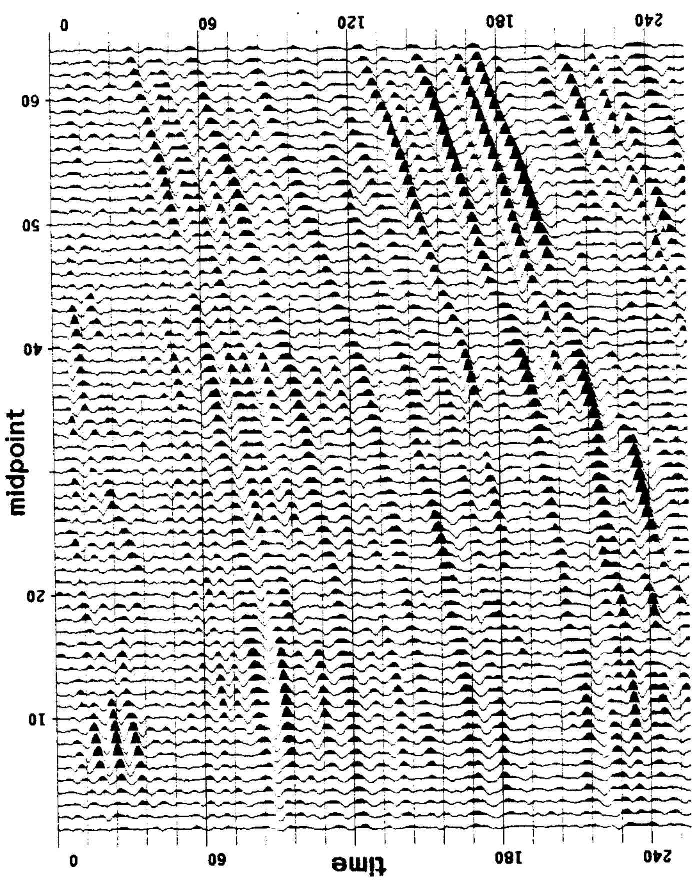


FIG. 1 (above left). The original stacked CMP data (recorded in deep water, 7-8 s, sampled at 4 ms and 100m) includes two types of noise: hardware noise (large waveforms) and a missing trace (40th). Geological events, by contrast, show strong lateral color.

FIG. 2 (above). A slant stack makes geological events whiter and more non-gaussian. Color remains because of second-order curvature of events. Noise (barely visible as straight lines) becomes strongly colored, more gaussian, and diffused through all p values. The figure displays only geological p's (p=0 at center trace).

FIG. 3 (left). To estimate geology we zero all events with analytic envelopes below a cutoff (12% of maximum). Though the change is barely distinguishable, only geology remains.





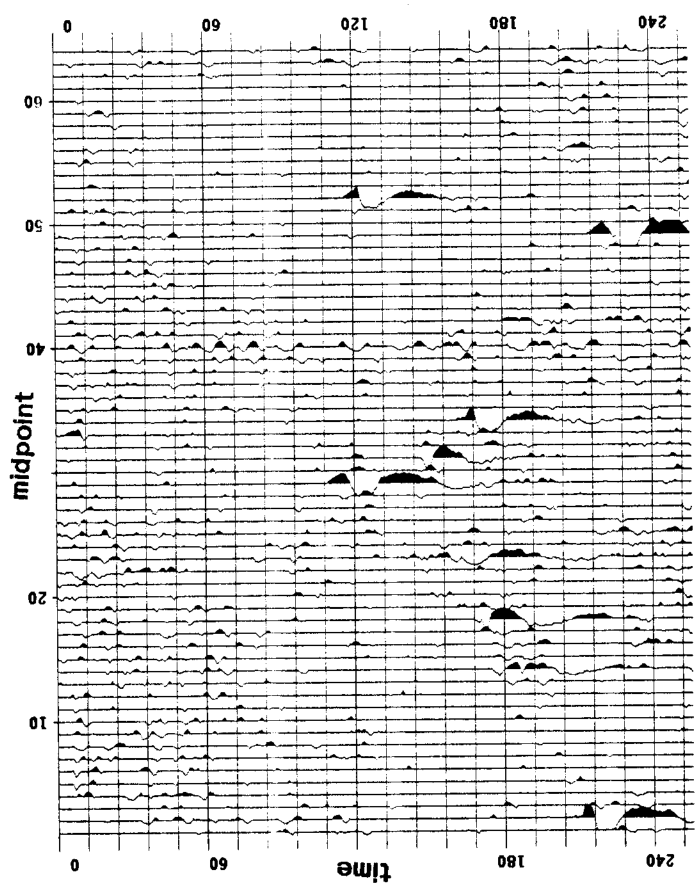
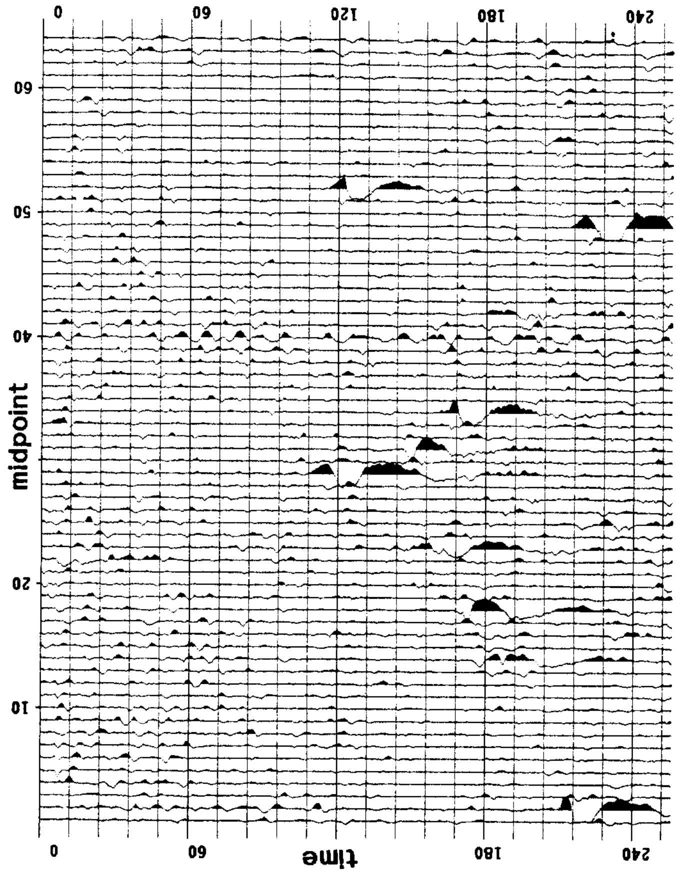
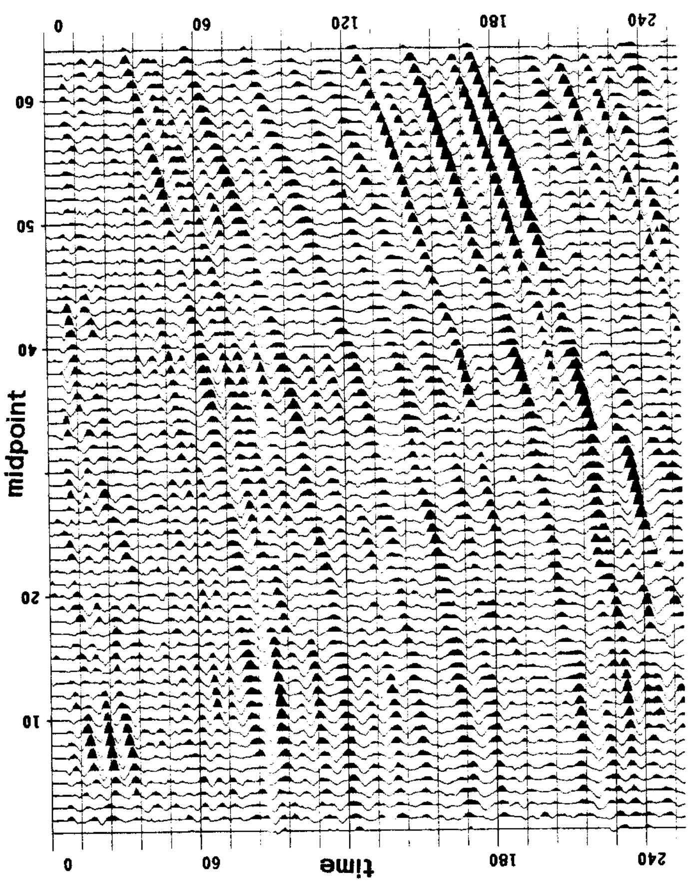


FIG. 4 (above left). An inverse slant stack confirms that our estimation chose the strongest and most colored geology. No white elements remain.

FIG. 5 (above). Removing the strongest geology from the original data leaves all noise with weak, colored, and fairly gaussian geology. The missing 40th trace appears as noise; a negative polarity exactly cancels the reasonable geology one expects in the original data.

FIG. 6 (left). A high amplitude selection (again 12%) estimates the strongest noise very well. No colored, geological events remain.





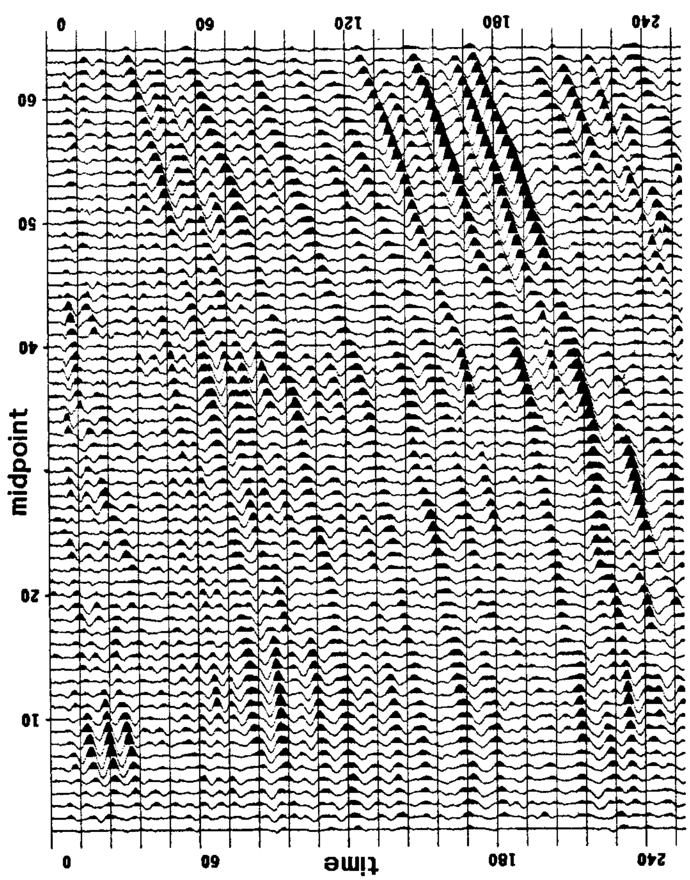


FIG. 7 (above left). Subtracting the strongest noise from the original data leaves *all* geological information, unobscured by the strongest noise. The previously missing trace now contains the strongest geological events. We slant stack this section for a new iteration.

FIG. 8 (above). Another full iteration, with cutoffs at 1%, extracts the strongest 99% of noise amplitudes.

FIG. 9 (left). A final estimation of the geology reveals clean, detailed events and largely restores the missing trace. Decomposition uncovers hidden geological information and does not harm any previously visible features.

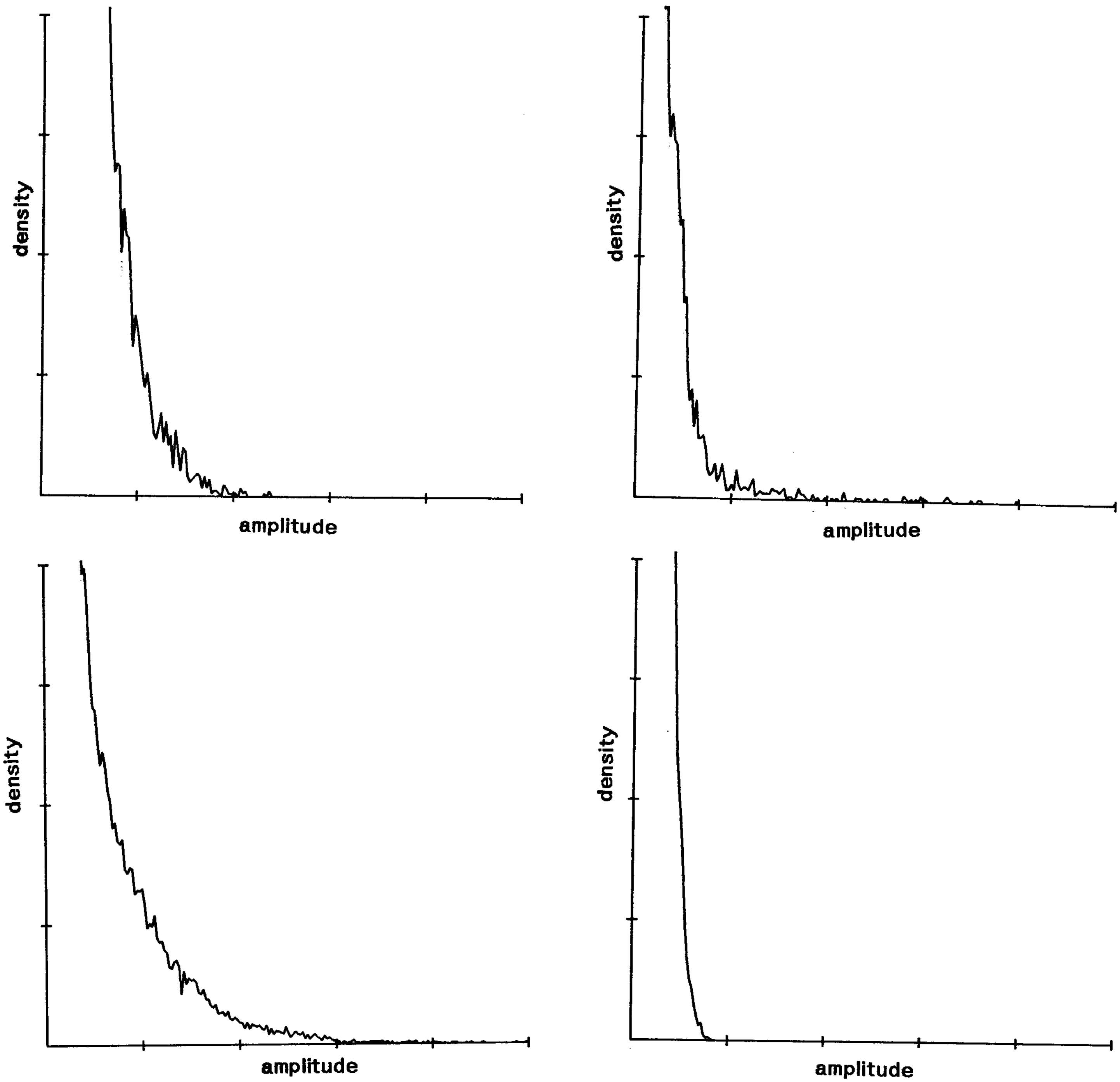
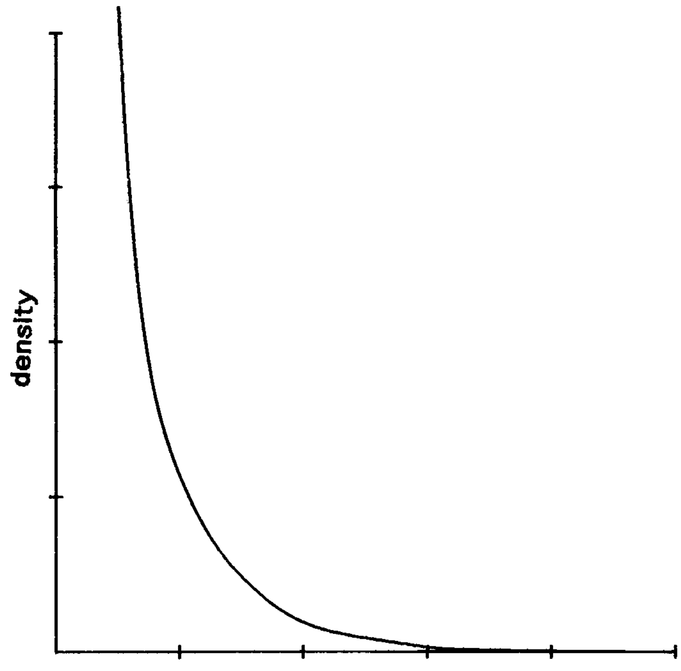


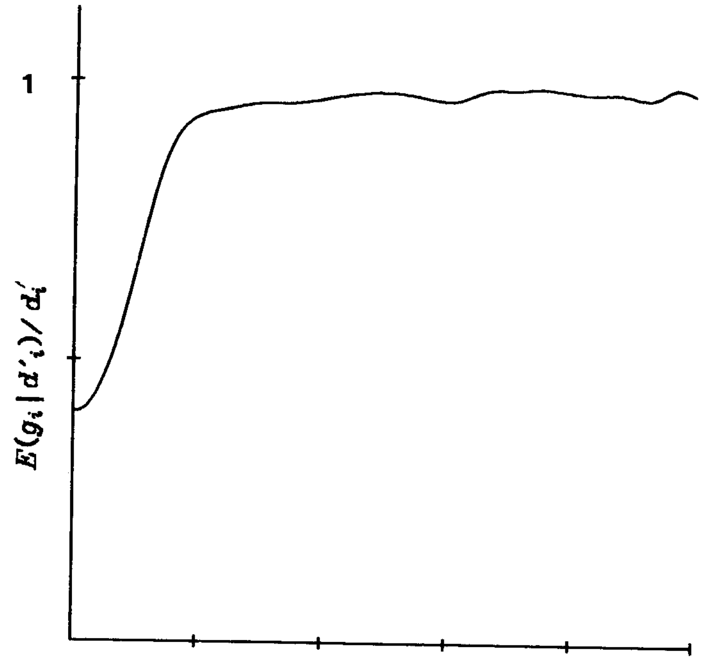
FIG. 10 (top left). The fully extracted geology (Figure 9) produces a histogram, (an approximate p.d.f, ignoring color) with a non-gaussian distribution of amplitudes.

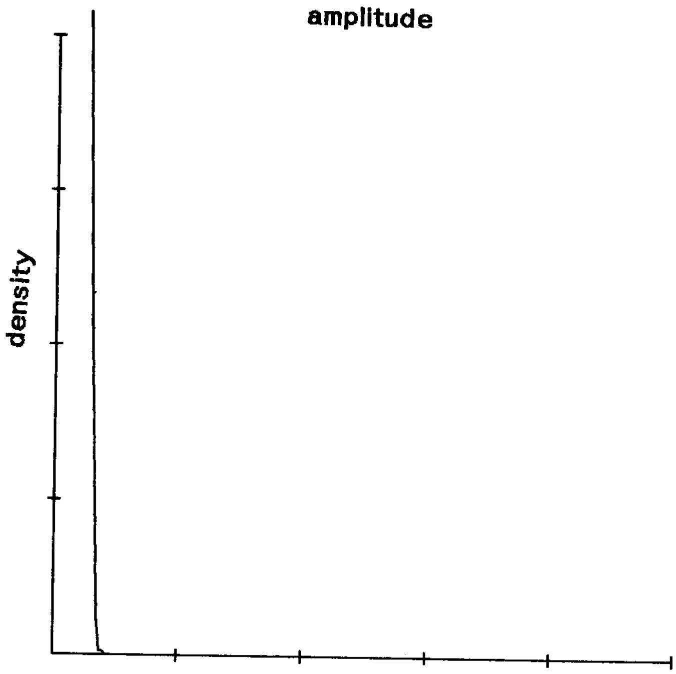
FIG. 11 (top right). The fully extracted noise (Figure 8) produces a more non-gaussian histogram than the geology but with a similar range of amplitudes. Estimating noise first from the original data would fail

FIG. 12 (bottom left). After a slant stack the samples of the extracted geology show a higher, more non-gaussian distribution of amplitudes.

FIG. 13 (bottom right). After a slant stack the fully extracted noise becomes much more gaussian. Amplitudes are very low.







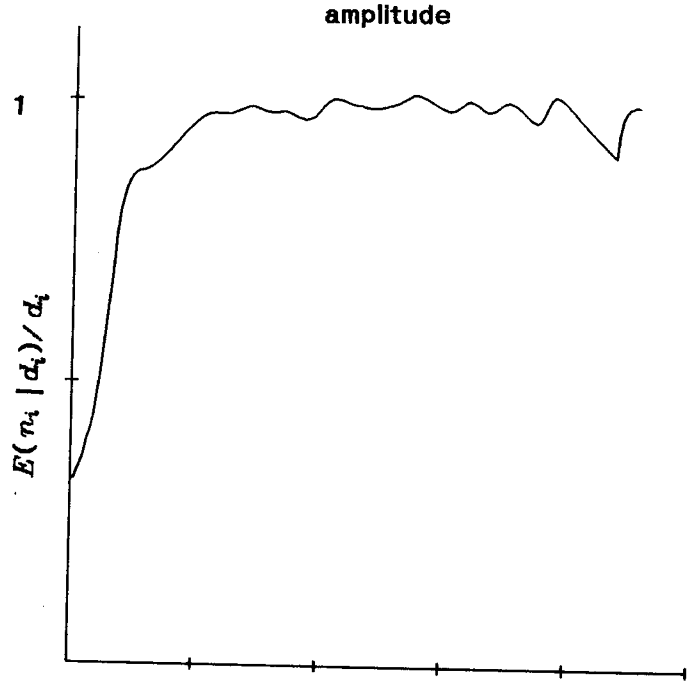


FIG. 14 (top left). A slant stack of the original data gives a distribution equal to a convolution of the geology and noise p.d.f.'s.

FIG. 15 (top right). Equation (4) gives $E(g_i\,|\,d'_i)/\,d'_i$ from the three previous p.d.f.'s. We estimate geology from a slant stack (Figure 2) by diminishing low amplitudes and keeping high ones at full strength. For a conservative estimate, we zero low amplitudes.

FIG. 16 (bottom left). With the high amplitude geology removed from the original data as in Figure 5, a histogram of the geology shows a lower, more gaussian distribution of amplitudes. The noise in Figure 5 remains as in Figure 11.

FIG. 17 (bottom right) With the strongest geology removed from the original data, the estimation function for noise also recommends diminishing small amplitudes.