Extracting velocities from diffractions

William S. Harlan, Jon F. Claerbout, and Fabio Rocca

Abstract

By extracting diffraction events from stacked sections one may estimate seismic velocities with migration. Extrapolation of wavefronts back in time focuses signal, which originates from simple sources. Migration focuses diffractions, from bed truncations and point scatterers, but not reflections of continuous beds; migration disperses noise.

A measure, W, exists which is minimized and equal to the statistical entropy, H, when the data are statistically white. We calculate W from local histograms of the data. An invertible transformation, L, alters W but not H. Whitening the signal with L makes it more non-gaussian, focusing the energy; noise defocuses, becoming more gaussian. Thus, for extracted signal, the best L minimizes W.

If transformation makes signal more non-gaussian, then we may extract the highest amplitudes as signal. Only coherent summing accounts for amplitudes above some cutoff; noise sums destructively. Artificially whitened data, simulating noise, determine good cutoffs.

Slant stacks map lines to points, greatly concentrating bed reflections. We invert the strongest events and subtract from the data, leaving diffractions and noise. Next, we migrate with many velocities, extract focused events, and invert. We find a least-squares sum and migrate again: since only diffractions remain, W reaches a minimum at the best velocity. We estimate interval velocities by windowing and extrapolating downward.

We successfully extract diffractions and estimate velocities for a window of data containing a growth fault. Illustrations demonstrate each step of the extraction procedure.

The physical problem

Velocity analyses extrapolate wavefields back in time, thereby concentrating signal and dispersing noise. For example, an NMO stack finds the image source of plane reflections. Born inversion extrapolates wavefields back to residual-velocity scatterers and divides out the effect of extrapolated sources. Genuine seismic events begin simply and produce increasingly diffuse wavefronts. Since the wave equation is symmetric in time, seemingly isolated arrivals, such as noise and missing data, require more diffuse wavefronts at time zero. The extrapolation which best concentrates the signal, while dispersing the noise, determines the the best velocities.

Common-offset sections detect deep velocity changes better than do common-shot and midpoint gathers by spanning greater distances on the surface. Diffraction events, such as reflections of bed truncations and point scatterers, contain all velocity information in common-offset sections. Diffractions appear in the background of sections containing faults, such as Figure 1a. Migration with correct velocities focuses these diffractions; however, reflections of continuous beds do not focus, and noise defocuses. We shall first remove continuous reflections from the data, exposing diffractions and noise as in Figure 5c. We shall next extract the strongest diffractions from the noise as in Figure 10a. Note that phase changes appear at peaks of the hyperbolas, as predicted by theory. When these events are migrated, an appropriate focusing measure is minimized at the best velocity, as in Figure 12c.

Focusing and extraction of signal

We shall first define a whiteness measure for linearly transformed data. Let $\bar{d}$ be a data array, and let $p_{\bar{d}}(\bar{d})$ be the corresponding probability density function (p.d.f.). Shannon's statistical entropy, H, measures the predictability of, or the information contained in, the random process.

$$H \equiv -\int p_{\bar{d}}(\bar{x}) \ln(p_{\bar{d}}(\bar{x}))d\bar{x} \ . \tag{1}$$

H is large when $\bar{d}$ is most unpredictable. Shannon's fundamental theorem is

$$H \leq W \equiv -\sum_i \int p_{d_i}(x) \ln(p_{d_i}(x))dx \tag{2}$$

$$where \ p_{d_i}(x_i) = (\prod_{j \neq i} \int dx_j) p_{\bar{d}}(\bar{x}) \ .$$

The quantity defined as W might be called a whiteness measure. W is minimized and equal to H if and only if $\bar{d}$ is statistically white (the covariance matrix is diagonalized). If p_{d_i}

varies slowly as a function of d_i (local stationarity), then p_{d_i} may be estimated from local histograms of $\bar{d}$. An invertible linear transformation of the data can alter color but not information--can alter W but not H. Thus W is minimized by the linear transform which whitens the data most.

The linear transform which focuses the signal also defocuses the noise. Let $\bar{d}'=L^{-1}\bar{g} + \bar{n}'$, where the array $\bar{d}'$ contains data, $\bar{g}$ whitened geologic events, and $\bar{n}'$ white noise. L^{-1} is the right inverse of L. The transform $L\bar{d}'$ adds statistical color to the white noise events. Because of the central limit theorem, coloring white random variables makes them more gaussian. Similarly, whitening the signal makes it more non-gaussian. Thus, an ideal transform, L, is said to focus: geologic events become stronger and narrower; information is concentrated. We must remove noise, however, before the best L minimizes W.

If the data have been transformed so that signal is more non-gaussian than noise, then the highest amplitudes represent signal and may be extracted. Let d, g, and n be samples of $L\bar{d}'$, $\bar{g}$, and $L\bar{n}'$, so that $d = g + n$. Assume d to be large, relative to the width of p_d. Using Bayes rule, we can estimate the signal-to-noise ratio g/n for this large d:

$$E(g/n \mid d) \approx \frac{\int_{d-\Delta n}^{d+\Delta n} x p_g(x) p_n(d-x)\,dx}{\int_{d-\Delta g}^{d+\Delta g} x p_g(d-x) p_n(x)\,dx} \approx \frac{p_g(d)}{\Delta g} \frac{\Delta n}{p_n(d)}. \tag{3}$$

We reduce the integration limits about the largest contributions. Δg and Δn are widths, such as standard deviations, of the p.d.f.'s. The ratios at the far right of (3), though approximate, give the strength of the tail of a p.d.f. scaled by the width--a ratio measuring the divergence from a gaussian. When the signal is more non-gaussian than the noise, the signal-to-noise ratio is large for large d, and $g \approx d$. After an ideal transformation of the data, low amplitudes may be smoothly diminished to zero, the rest being the strongest signal. (Analytic envelopes measure the true amplitude of an event best; smoothing of the extraction function preserves the frequencies present.) Only highly non-gaussian noise will obscure the focused signal; such noise may be clipped before transforming. Let us now determine a reasonable signal cutoff.

One may artificially whiten data, transform, and select a signal cutoff near the highest resulting amplitude. Good cutoffs will appear at typical quantiles, a fixed percentage of the strongest samples, but one should avoid assumptions about data statistics. Let us artificially whiten the data set laterally but retain the same local distributions of amplitudes--perhaps randomly switching the polarity of traces. If the data contain only noise, then the original and altered transforms should both sum destructively and produce sections with

identical amplitude distributions. After transforming, we choose signal cutoffs in the original data equal to locally high quantiles of the altered data. Higher amplitudes must represent focusing of coherent signal.

Extracting diffractions: removing reflections of continuous beds

Migration of a section does not significantly change the amplitude distribution of continuous beds. A window of stacked offshore Texas data contains a growth fault with weak diffractions off truncated beds (Figure 1a). If we ignore the lateral statistical color of the beds, the amplitude distribution appears gaussian (Figure 1b). We prepare a cumulative-density function (c.d.f.) by integrating a histogram of the sample amplitudes. We compare this function to the gaussian c.d.f. and calculate a corresponding standard deviation at each amplitude. The standard deviation remains constant over amplitude, indicating gaussianity. The abrupt descent to zero marks the largest amplitude present. A constant-velocity Stolt migration at 1980 m/s focuses diffractions and sharpens the fault (Figure 1c), but sample amplitudes remain gaussian (Figure 1d).

One may remove reflections of continuous beds and retain diffractions by focusing with slant stacks. Slant stacks map lines of constant dip into points, with the transform $d(p,t) \equiv \int d(x,t-px)dx$, thereby whitening bed reflections laterally. Events with rapidly changing dips, such as noise and diffractions, diffuse instead. Figure 2a shows a slant stack over the slopes of ± 0.12 ms/m. The sample amplitudes have become strikingly non-gaussian (Figure 2b).

We estimate a signal cutoff first by artificially whitening the data (Figure 3a). Amplitudes retain the same distribution (Figure 3b). A slant stack disperses events throughout with strong statistical color (Figure 3c). The transformed distribution remains very gaussian (Figure 3d). The highest amplitude, at $\approx$7500, represents the highest possible value for noise. After transforming, we extract the strongest focused events, (Figure 4a), then invert (Figure 4b) and subtract from the original data (Figure 4c). By removing beds we uncover diffractions, though noise still remains. We improve our extraction by artificially whitening this result, transforming, and finding the maximum values as before. Figure 5b shows the new estimate of the strongest beds. Diffractions appear more clearly after subtraction of the beds (Figure 5c). This estimation procedure does not remove low dips from diffractions as would a dip filter.

Extracting diffractions from noise

Now we treat a more general transform L, perhaps of higher dimensions, like migration. We may extract signal over a range of a transform's parameter, linearly combine inverted events, and find the parameter which minimizes the whiteness measure. Let our transform L_v depend on v. We extract all signal focused at some v and then invert:

$$\bar{s}_v \equiv L_v^I\{Extract\{L_v\{d'\}\}\} \tag{4}$$

where L_v^I, the pseudo-inverse of L_v, inverts all signal with parameter v. Having $\bar{s}_v$ for a physical range of v, we find the least-squares sum which best reproduces the data:

$$\hat{g} = \sum_v a_v \bar{s}_v \; ; \quad \min_a (\hat{g} - \bar{d}')^2 \rightarrow \sum_w <\bar{s}_v, \bar{s}_w> a_w = <\bar{s}_v, \bar{d}'>. \tag{5}$$

The brackets indicate a simple dot product. We may solve (5) for all a_v and for $\hat{g}$. $\hat{g}$ contains the strongest signal, without bias, for the chosen range of v. Transforming $\hat{g}$ over this same range will minimize W at the v which focuses the signal best.

To extract diffractions and estimate velocities, we use equations (4) and (5), with migration and velocity as the appropriate transform and parameter. First we migrate the data without beds over a suite of velocities (Figure 6). Migration adds a small non-gaussianity at the highest amplitudes (Figure 7). If the noise is very gaussian, as here, one may avoid the artificial-whitening procedure and use the maximum amplitude of the data before migration as a cutoff. We extract the focused events (Figure 8) and diffract (inverse of migration) at the same velocities (Figure 9). Figure 10a contains the least-squares estimate, $\hat{g}$, of the diffractions. Subtracting these diffractions from the data without beds reveals only gaussian noise (Figure 10b).

We migrate the extracted diffractions over the same range of velocities (Figure 11) and find the non-gaussianity increases (Figures 12a and 12b). The whiteness measure, W, finds a distinct minimum at the best velocity (Figure 12c). For interval velocities one should extract upper events first then extrapolate lower events through the resulting velocities. Velocities may be measured as locally as the density of diffractions permits. Figure 13 shows the extraction of diffractions and linear beds from particularly deep marine data. The window extends 9.6-11.6s and 6km, at 4ms and 50m sampling. The aperture of offsets is very small relative to the depth, so such diffractions are the only source of velocity information.

Other algorithms

An automatic velocity analysis could use common-shot or midpoint gathers. For midpoint gathers one could let L_v be the normal-moveout stack. We should prefer imaging shot gathers to stacking midpoint gathers. Let L_v be a migration using the real two-way travel times. Such a migration finds image sources of reflectors. Truncation of gathers qualifies as noise to be estimated and removed. One begins by extracting shallow reflections, just as we have extracted diffractions, and using the corresponding velocities to image lower events. With only low-frequency velocity information extracted one could not hope to image the source in its proper location at the surface, unless one were willing to manipulate discrete velocity boundaries. Since interval velocities are to be determined, spurious events such as multiples and sideswipe will often be impossible to extract and thereby be eliminated from the velocity estimations. Retaining only the near offsets of the image (an assumption of low dip) reduces the cost. Arbitrary accuracy may be obtained at a given depth by using an iterative algorithm. The best of five widely spaced velocities could be determined, then of five velocities closely spaced about this best one, and so on.

One may remove highly non-gaussian noise by first estimating the strongest signal with migration and slant stacks and removing the signal from the data. Highest amplitudes then represent noise and may be extracted. Figure 14 demonstrates the effectiveness of this procedure for removal of hardware glitches and missing traces. Missing traces appear as the negative of the reasonable geology one expects in its place. Subtracting extracted noise from the original data restores previously missing portions of coherent events.

REFERENCES

Berkhout, A.J., 1980, Seismic migration--Imaging of acoustic energy by wave field extrapolation: Amsterdam/New York, Elsevier/North Holland Publ. Co.

Claerbout, Jon F., 1983, Imaging the earth's interior (in press).

Godfrey, R., 1979, A stochastic model for seismogram analysis: Ph.D. thesis, Stanford University, 88 p.

Harlan, W., 1982, Signal/noise separation with slant stacks and migration: SEP-32, p.25-35.

Shannon, Claude E. and Warren Weaver, 1963, The mathematical theory of communication: Chicago, University of Illinois Press.

Van Trees, H., 1968, Detection, estimation, and modulation theory, v.1: New York, John Wiley.

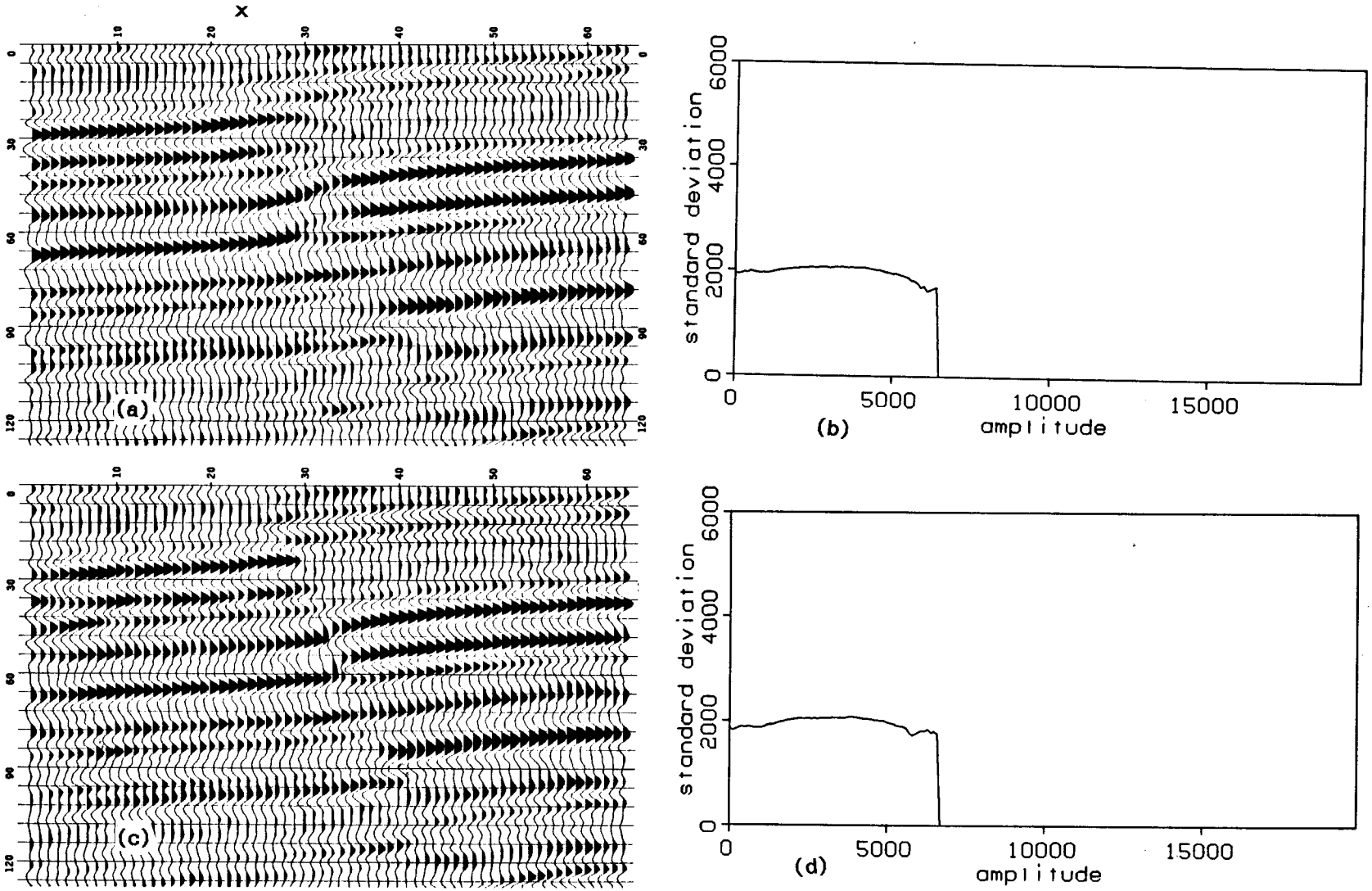


FIG. 1. (a) A window of stacked offshore Texas data contains a growth fault, with weak diffractions off truncated beds. The events from continuous beds and noise obscure the velocity information contained in the diffractions. The data were supplied by Western Geophysical (window extends 1-1.5 s and 2 km, at 4 ms and 33 m sampling).

(b) If the lateral color of events is ignored the data of Figure 1(a) appear gaussian. We prepare a cumulative-density function (c.d.f.) from a histogram of sample amplitudes, then calculate an appropriate standard deviation for each amplitude from the gaussian c.d.f. Little change in standard deviation indicates a gaussian distribution. The abrupt descent to zero marks the largest amplitude present.

(c) A constant-velocity Stolt migration at 1980 m/s focuses diffractions and sharpens the fault. But the strong lateral statistical color of the beds remains, and noise diffuses.

(d) After migration sample amplitudes remain as gaussian as before. No focusing measure could detect an improvement before we remove beds and noise.

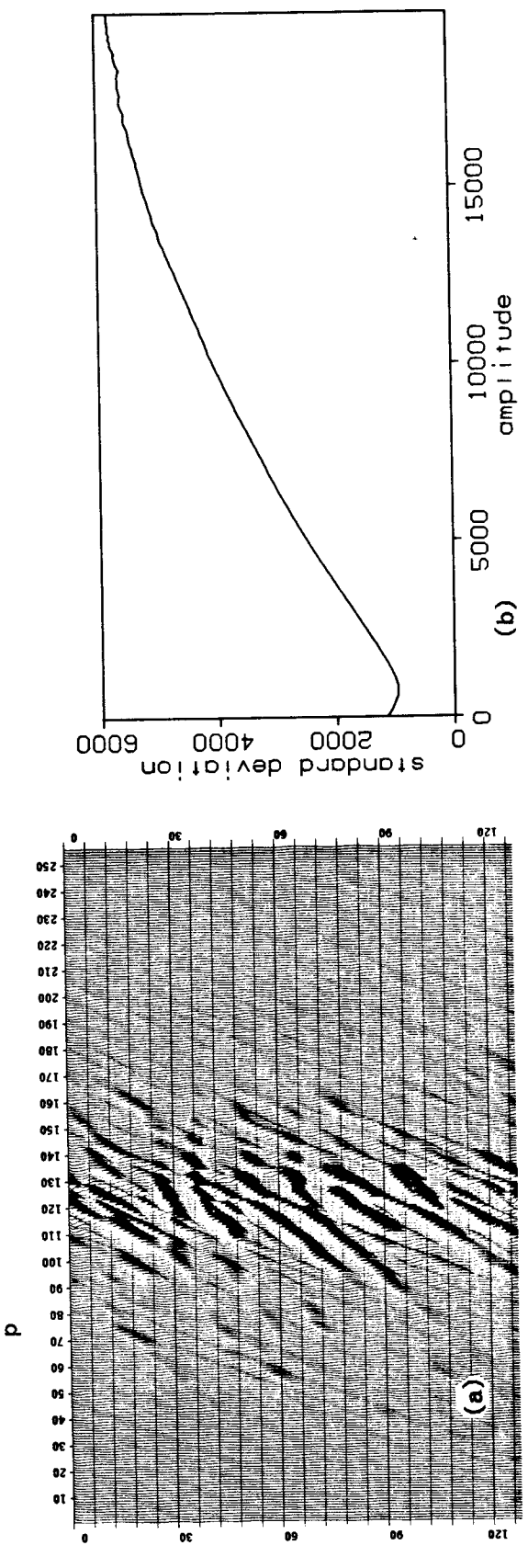
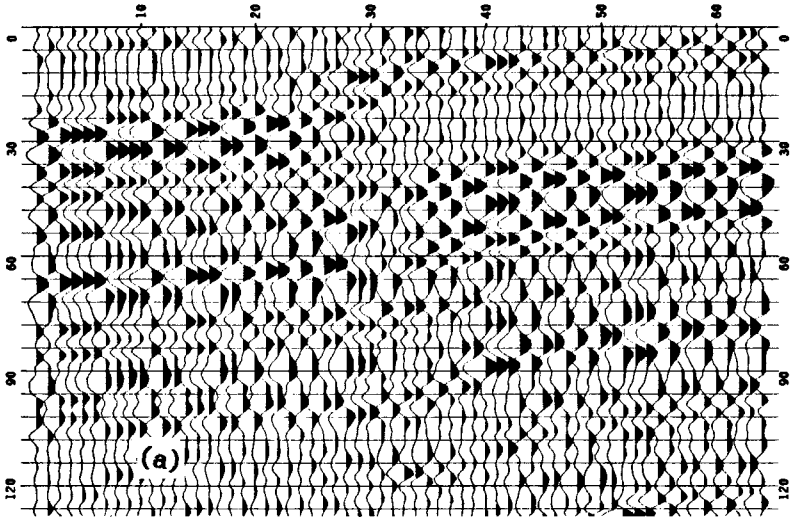
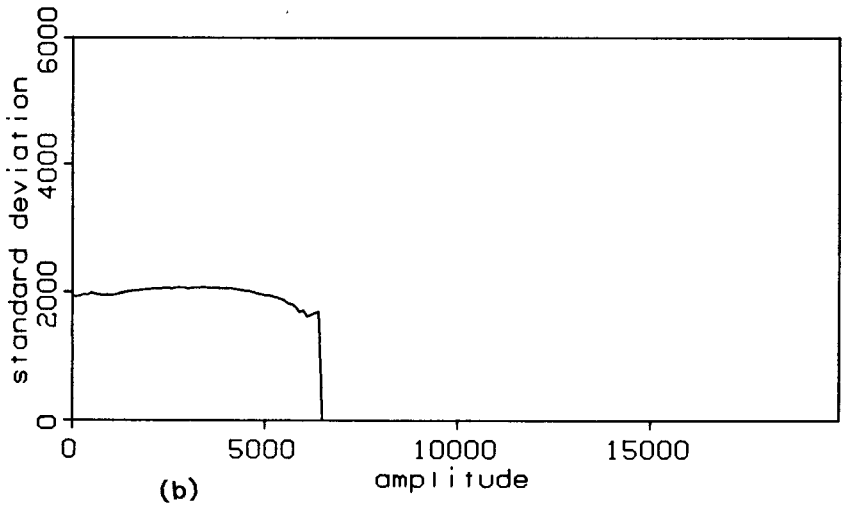
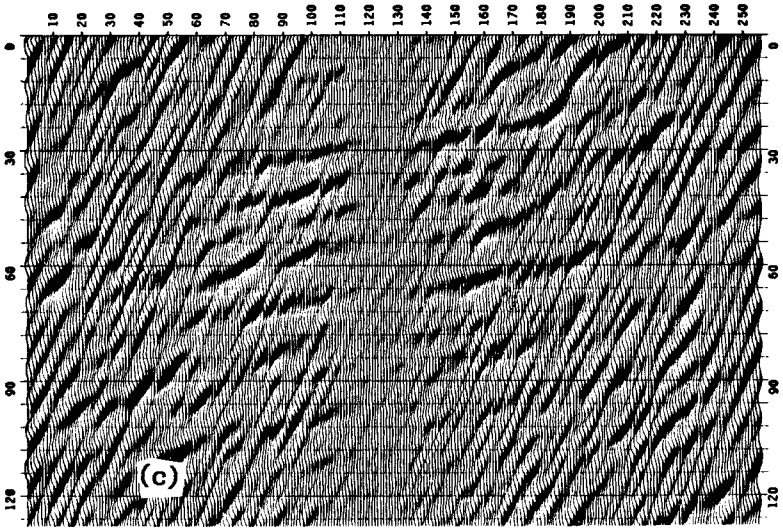


FIG. 2. (a) A slant-stack transformation of the data over the range ±0.12 ms/m concentrates the energy of the continuous beds, reducing their statistical color. Events containing too many dips, diffractions and noise, are dispersed.
(b) After the slant stack (Figure 2(a)), sample amplitudes are strikingly non-gaussian. Events with higher amplitudes suggest higher standard deviations. Moreover, the magnitude of the largest samples increases. Samples above some cutoff, determined in Figure 3, must represent the focusing of continuous beds.







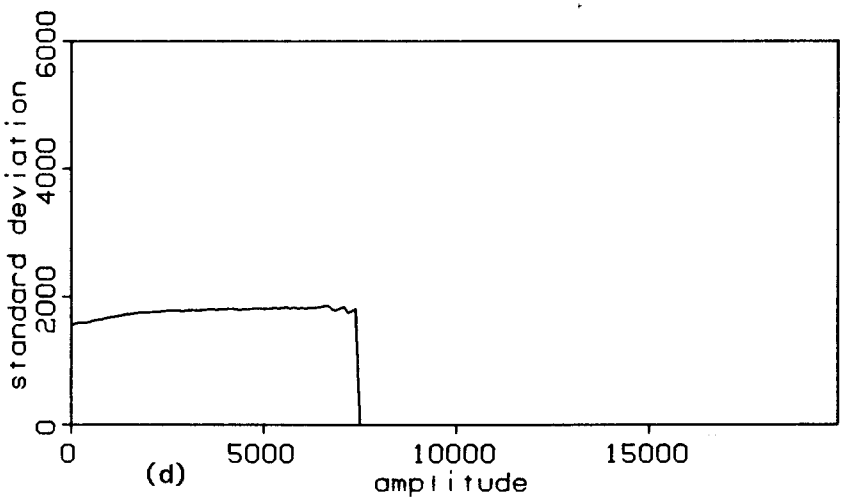
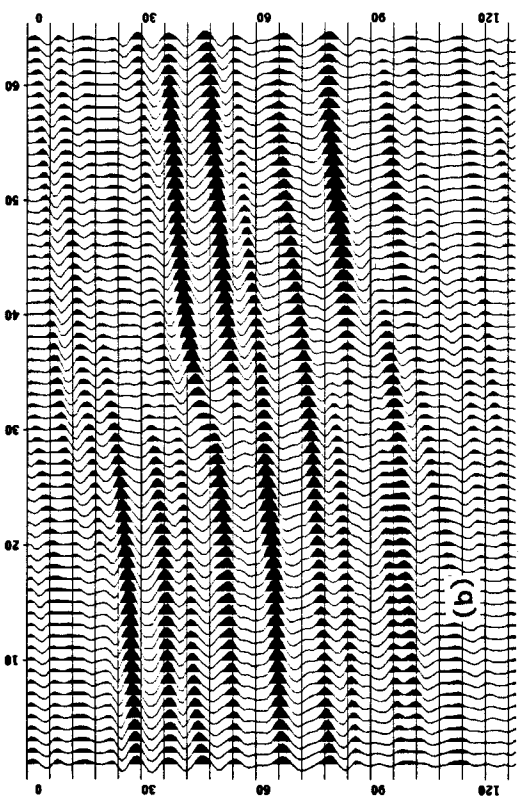
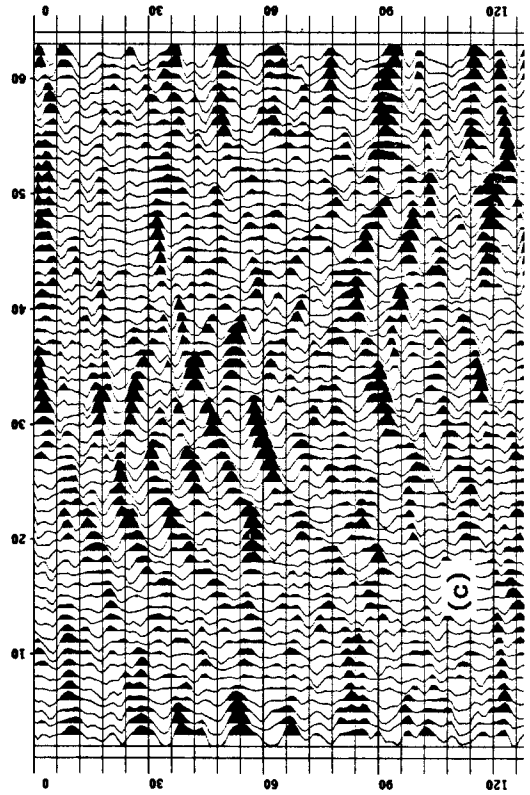


FIG. 3. (a) We artificially remove all lateral color from the data in Figure 1(a) by randomly reversing the polarity of the traces; we retain the same distribution of amplitudes. If the data were entirely of noise, then the statistics of the slant stack should be the same.
(b) Artificially whitening the data does not change the amplitude distribution.
(c) The slant stack no longer focuses as in Figure 2(a). Events distribute their energy throughout the section with strong statistical color.
(d) A slant stack of noise does not become more non-gaussian. The highest amplitude, at ~7500, represents the highest possible value for noise. In Figure 4 we extract higher amplitudes, belonging to focused continuous beds.





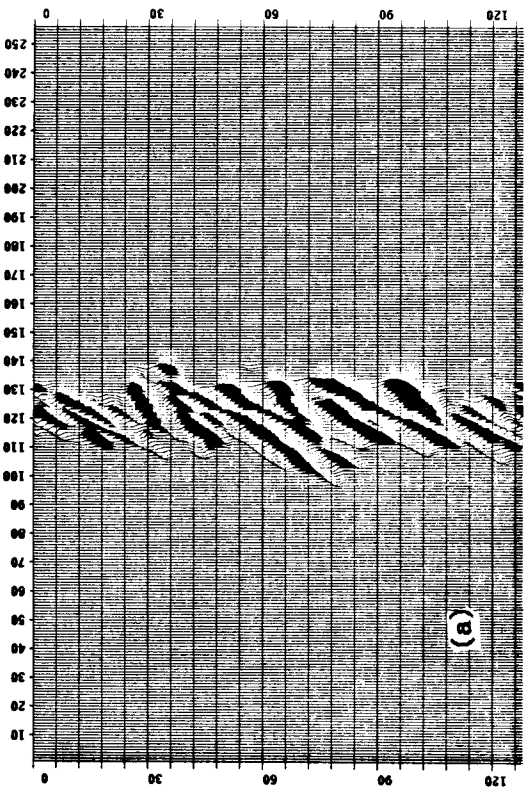
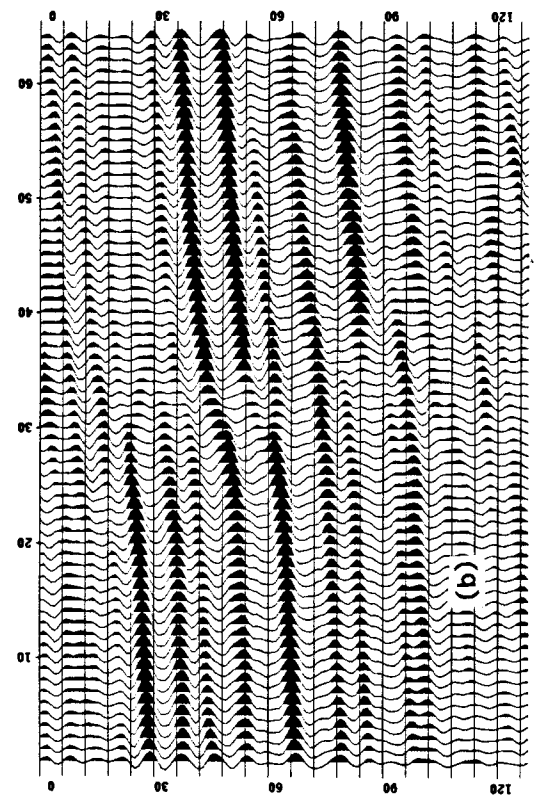
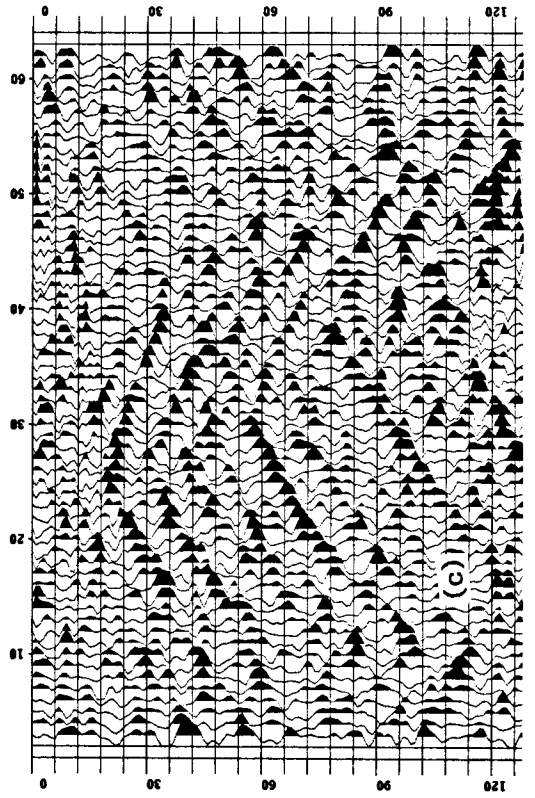


FIG. 4. (a) We extract those samples of the transformed data (Figure 2(a)) whose amplitudes are greater than the corresponding values of Figure 3(c). The amplitude cutoffs should be smoothly variable over the section and be measured from the analytic envelope of the data.
(b) An inverse slant stack reveals the strongest continuous beds.
(c) Subtracting the strongest beds from the original data (Figure 1(a)) reveals noise, diffractions, and weaker beds. (Gain has been increased to make these visible.) Since the noise is considerably weaker than the beds, we obtain a better signal cutoff by artificially whitening this section, forward transforming, and finding the maximum values--just as before. In Figure 5 we extract with the improved cutoff.





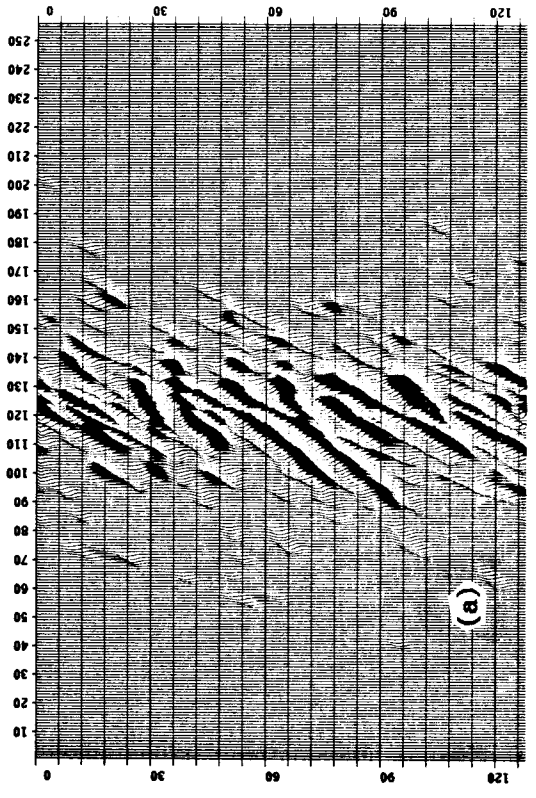
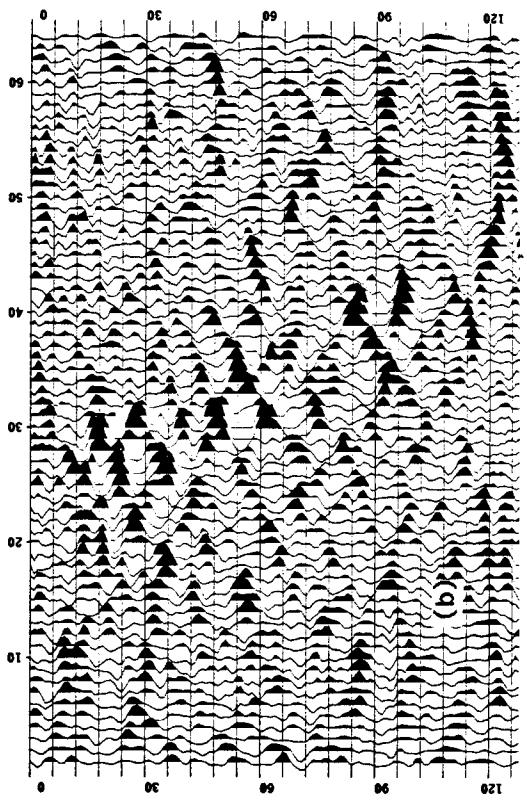
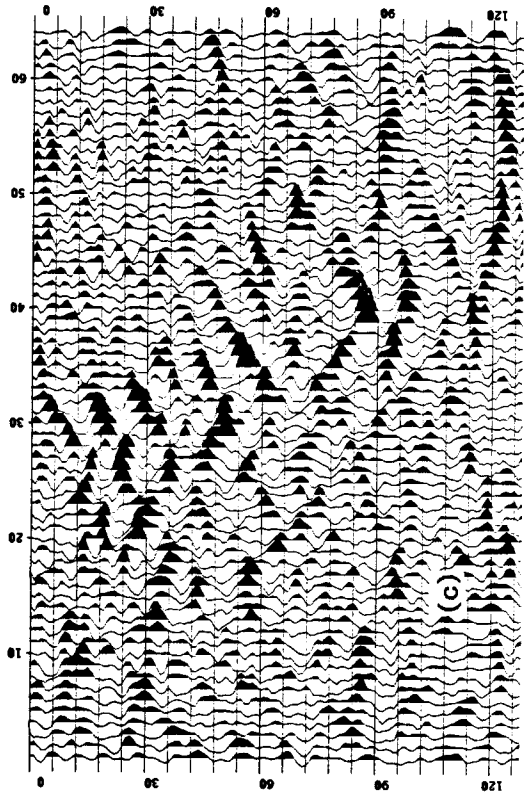


FIG. 5. (a) We extract the strongest events of the transformed data at the new cutoff.
(b) The inverse slant stack shows an improved estimate of the strongest beds.
(c) Subtracting the strongest beds from the original data reveals noise and diffractions, with no visible beds remaining.





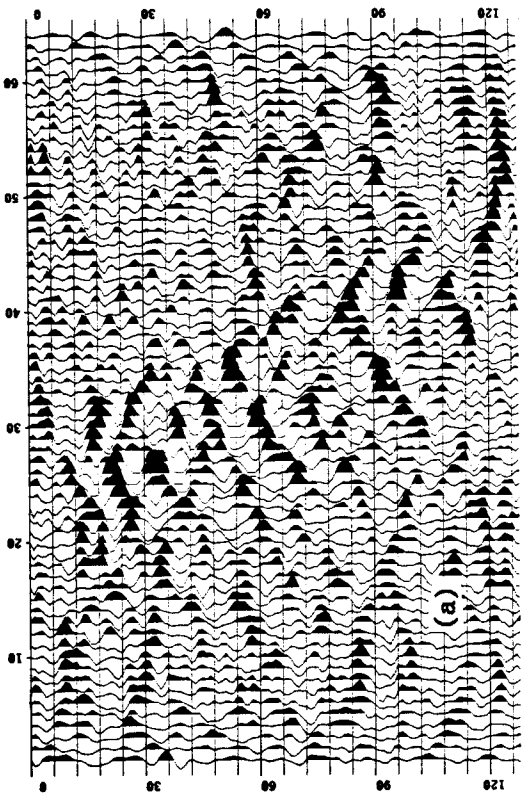
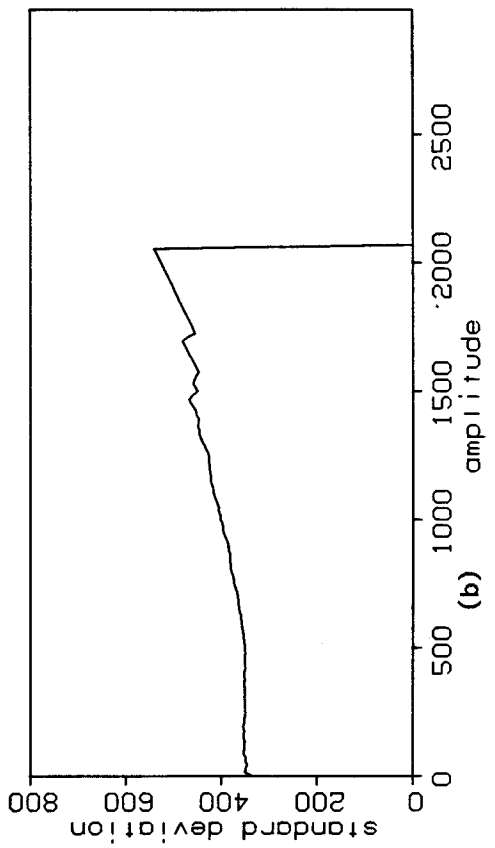


FIG. 6. Migration may now be used to distinguish diffractions from noise. First we migrate Figure 5(c) over a range of physical velocities. We display three: (a) an under-migration at 1650 m/s; (b) a good migration at 1980 m/s; and (c) an over-migration at 2310 m/s. Noise remains diffuse.



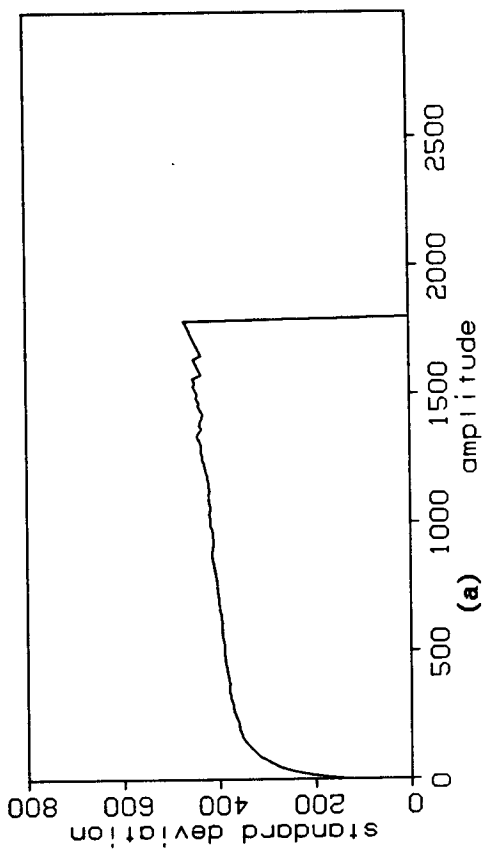
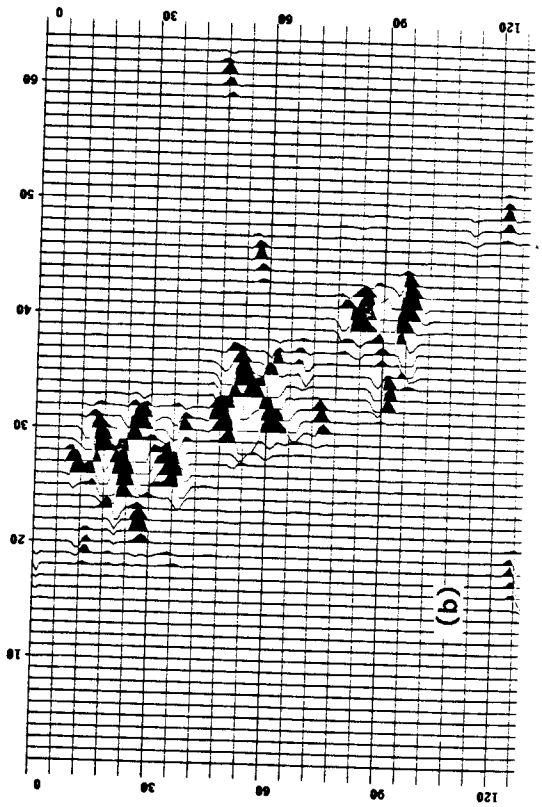
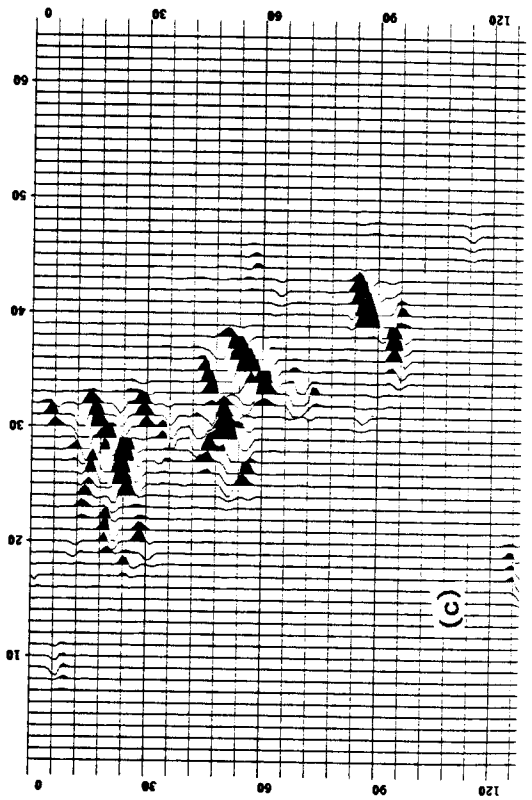


FIG. 7. (a) Because of noise, the amplitude distribution of Figure 5(c) remains very gaussian. (b) Migration at 1980 m/s (Figure 6(b)) adds a small non-gaussianity at the highest amplitudes. Since migration can only diffuse noise, these high amplitudes must represent focussed diffractions.





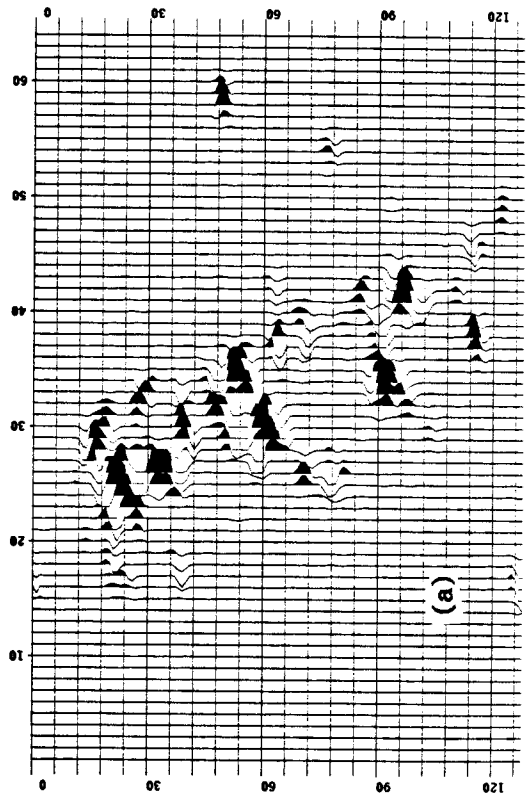
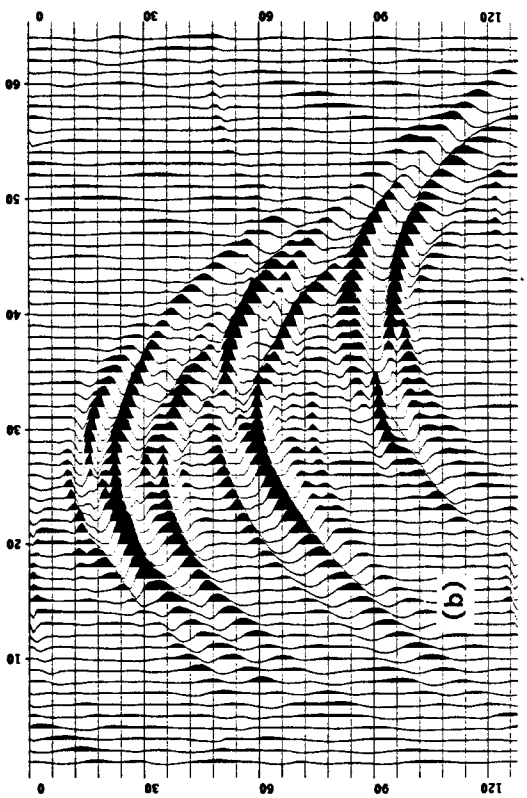
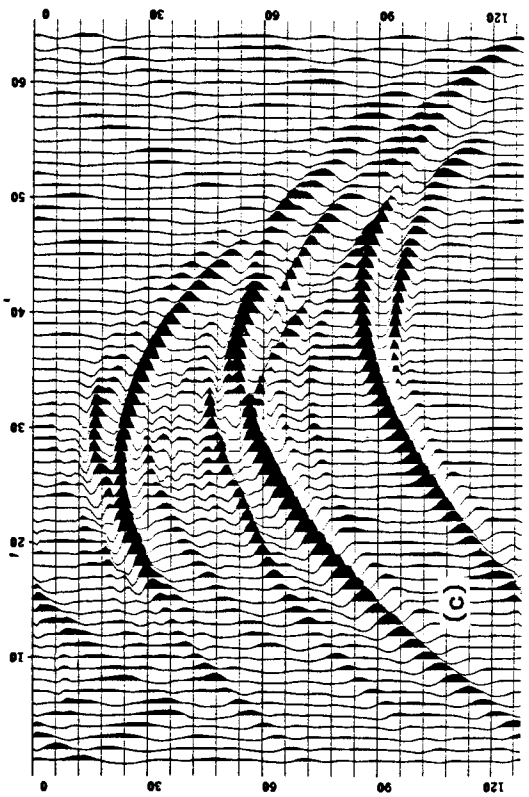


FIG. 8. We extract samples with amplitudes above some cutoff--amplitudes resulting from the focusing of diffractions. If the noise is very gaussian, one may choose the maximum value of the unmigrated section (Figure 5(c)) as the cutoff. Otherwise, one should proceed as in Figure 3 with artificial whitening. Cutoffs should be smoothly variable over midpoint and time. Three extractions are shown: (a) 1650 m/s, (b) 1980 m/s, and (c) 2310 m/s.





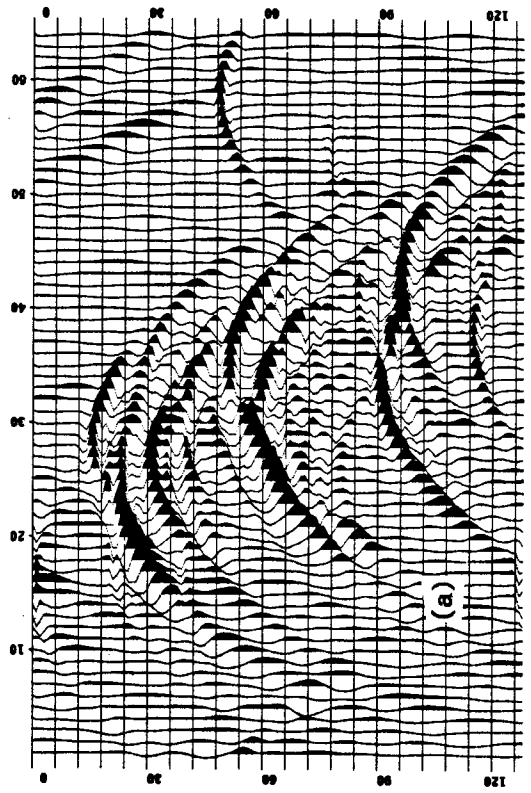


FIG. 9. We diffract (forward model) the events of Figure 8 at the velocities at which they were extracted. Each result biases events toward its corresponding extraction velocity because of the loss of residual tails. Three diffractions are shown: (a) 1650 m/s, (b) 1980 m/s, and (c) 2310 m/s.

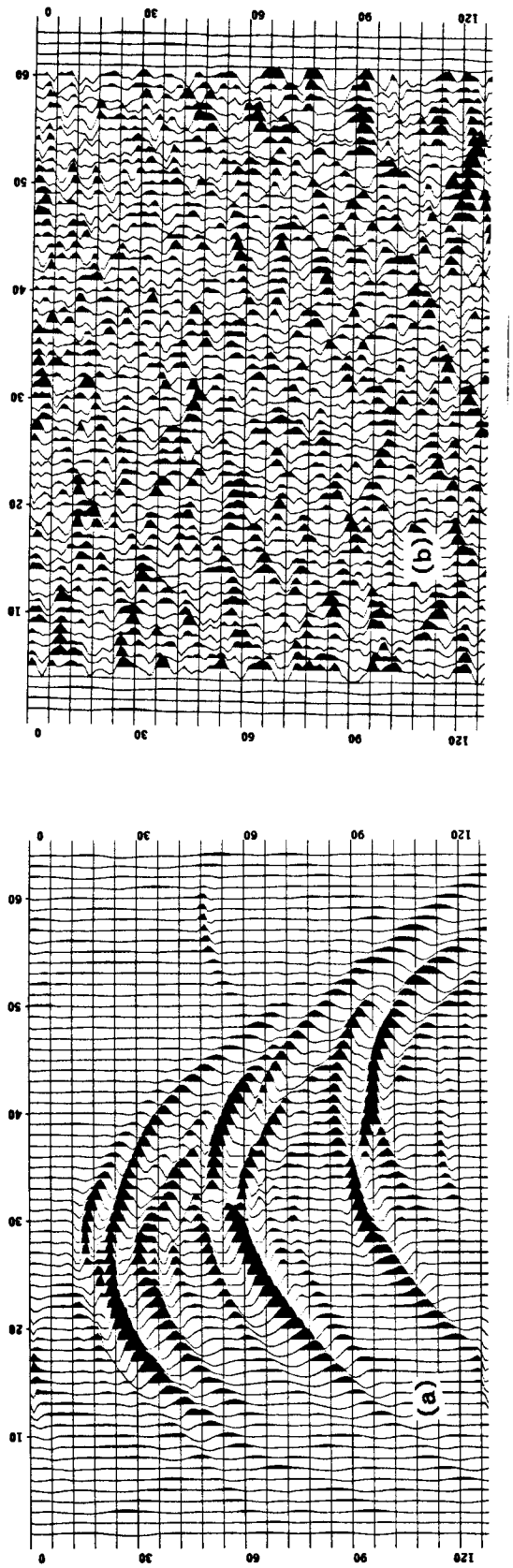
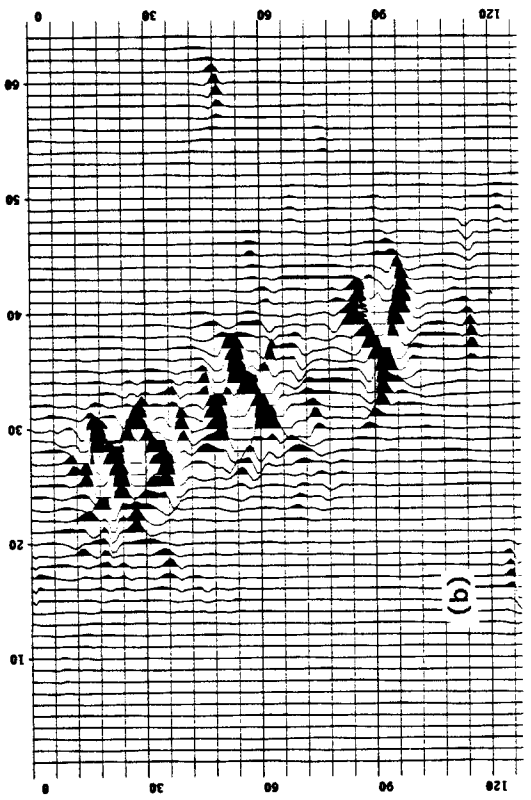
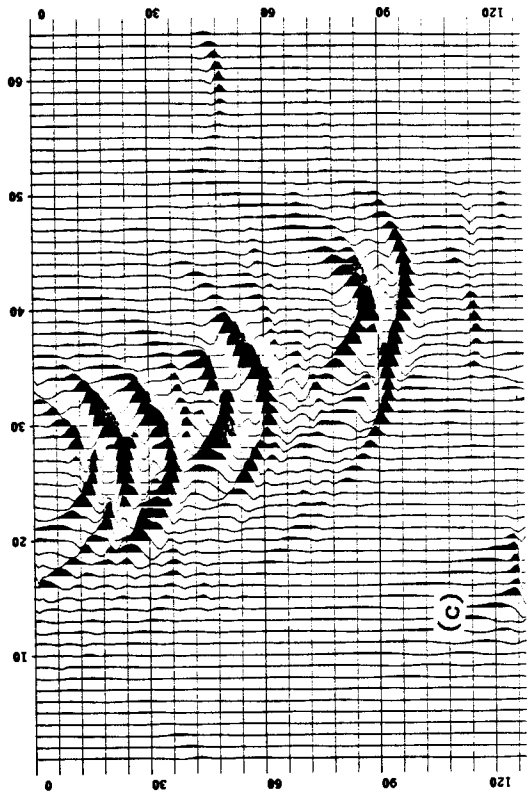


FIG. 10. (a) We find the least-squares sum of the extracted diffractions (Figure 9) which best resembles the section without beds (Figure 5(c)). Only those events which resemble the true diffractions will contribute to this least-squares sum. We then have the strongest diffractions within a range of velocities.
(b) Subtracting the extracted diffractions from the section without beds reveals only gaussian noise.





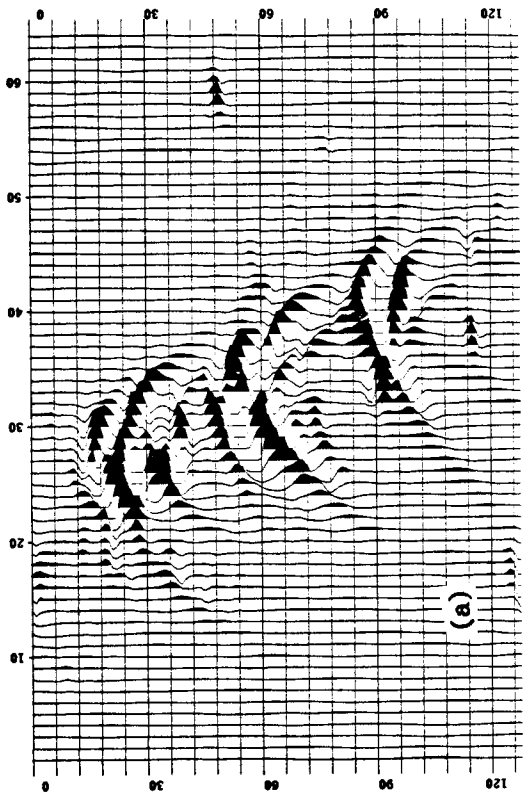


FIG. 11. We migrate the extracted diffractions of Figure 10(a) over the same range of velocities used to extract them. Three migrations are shown: (a) 1650 m/s, (b) 1980 m/s, and (c) 2310 m/s.

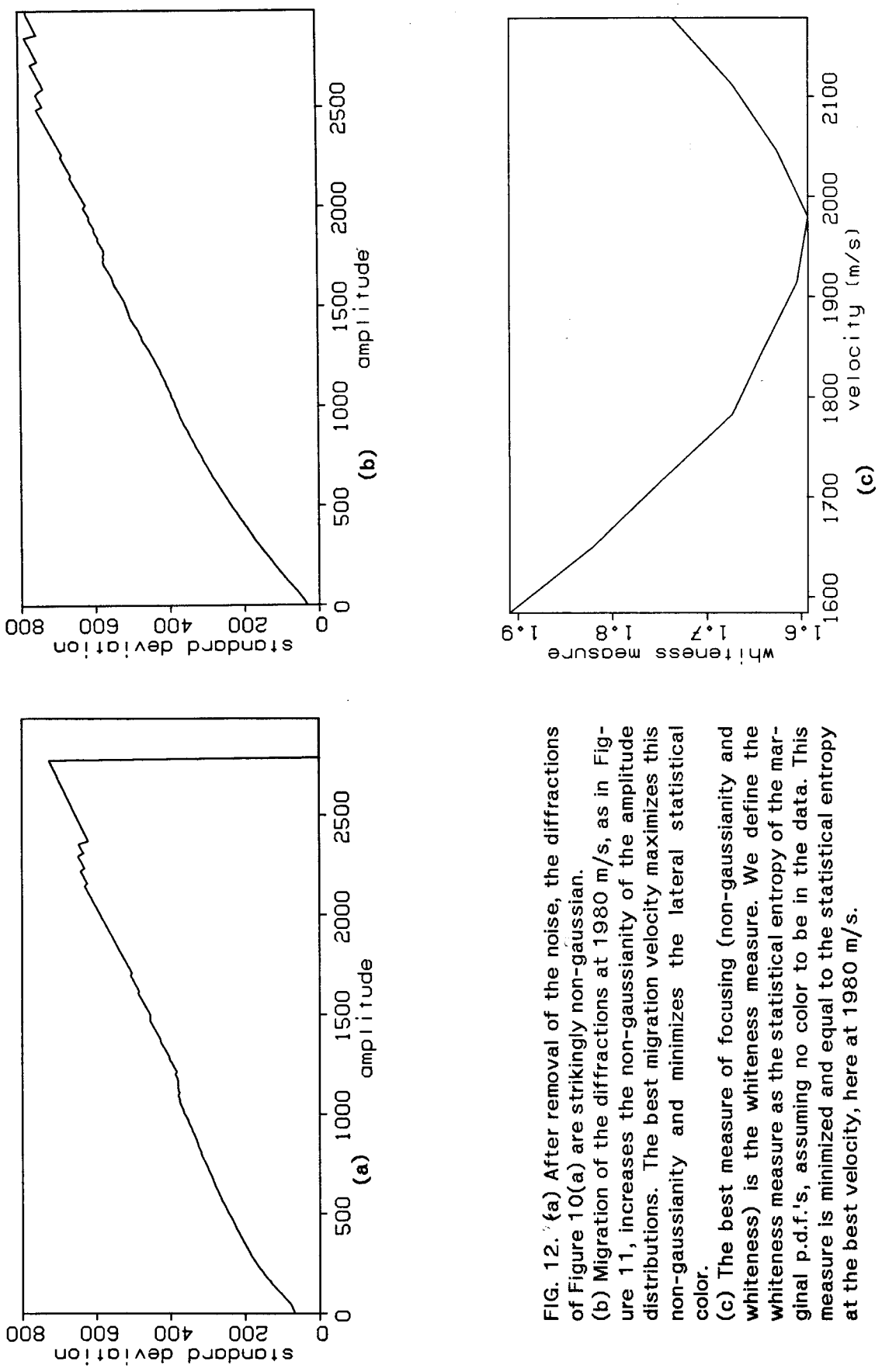


FIG. 12. (a) After removal of the noise, the diffractions of Figure 10(a) are strikingly non-gaussian.
(b) Migration of the diffractions at 1980 m/s, as in Figure 11, increases the non-gaussianity of the amplitude distributions. The best migration velocity maximizes this non-gaussianity and minimizes the lateral statistical color.
(c) The best measure of focusing (non-gaussianity and whiteness) is the whiteness measure. We define the whiteness measure as the statistical entropy of the marginal p.d.f.'s, assuming no color to be in the data. This measure is minimized and equal to the statistical entropy at the best velocity, here at 1980 m/s.

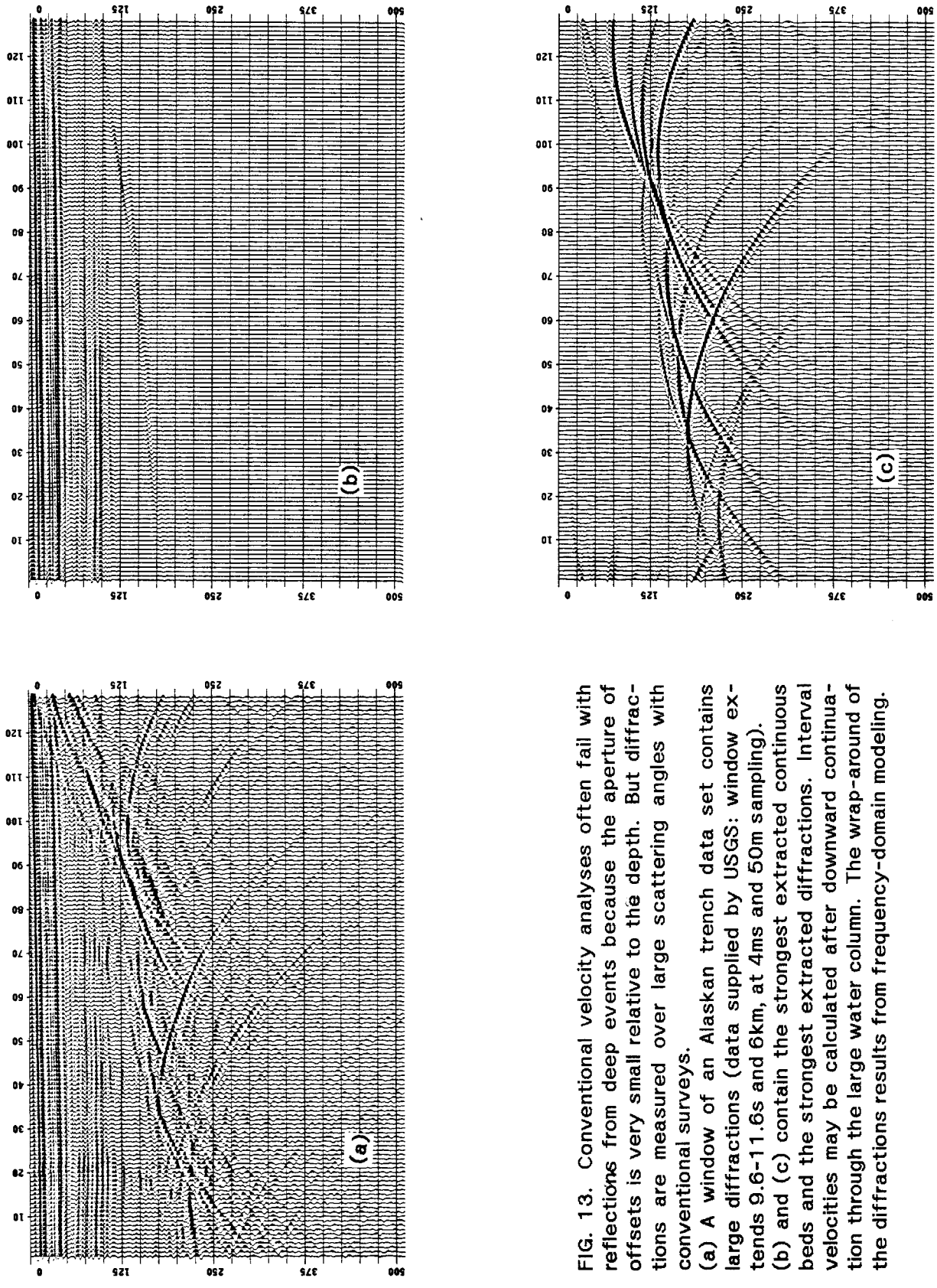


FIG. 13. Conventional velocity analyses often fail with reflections from deep events because the aperture of offsets is very small relative to the depth. But diffractions are measured over large scattering angles with conventional surveys.
(a) A window of an Alaskan trench data set contains large diffractions (data supplied by USGS: window extends 9.6-11.6s and 6km, at 4ms and 50m sampling).
(b) and (c) contain the strongest extracted continuous beds and the strongest extracted diffractions. Interval velocities may be calculated after downward continuation through the large water column. The wrap-around of the diffractions results from frequency-domain modeling.

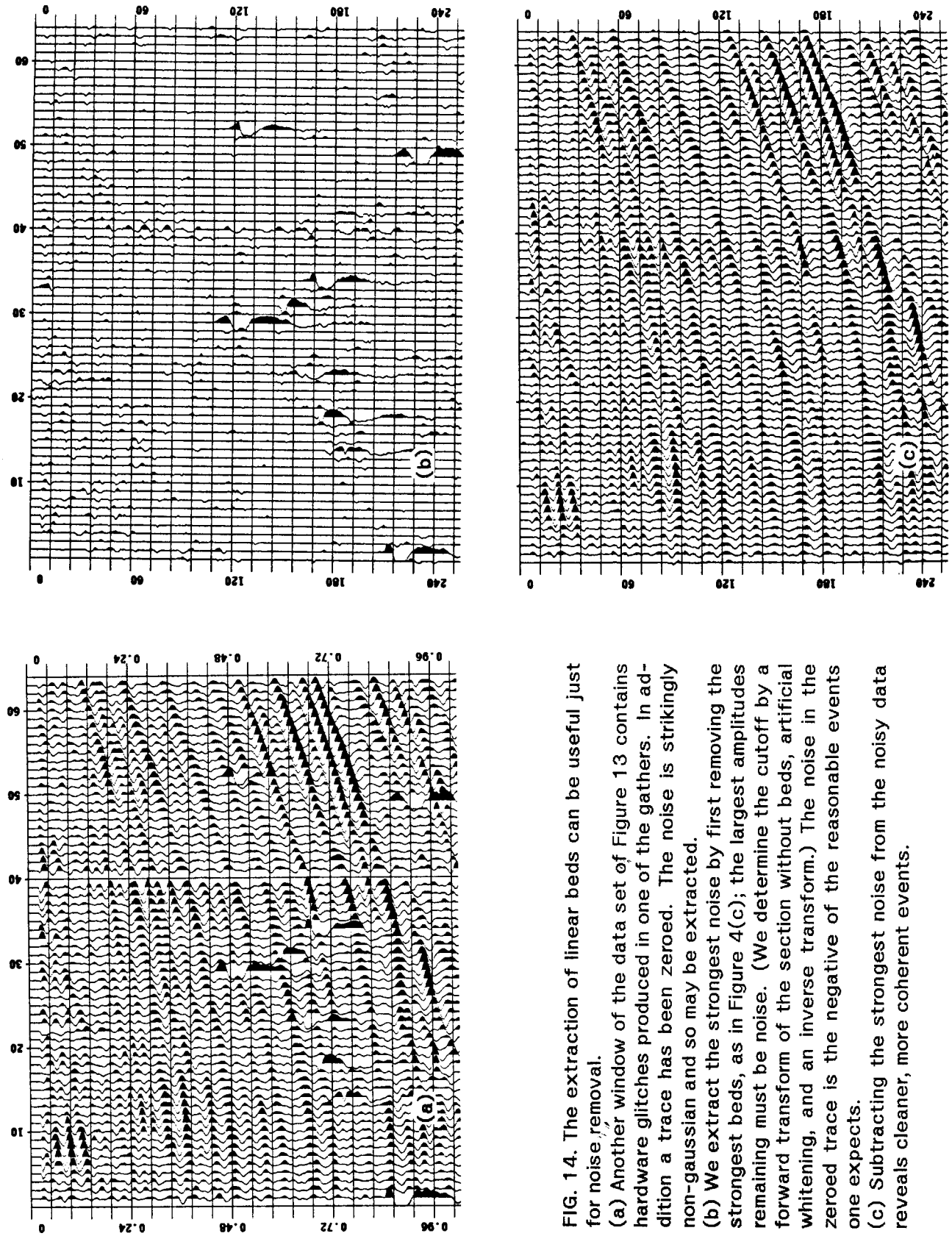


FIG. 14. The extraction of linear beds can be useful just for noise removal.
(a) Another window of the data set of Figure 13 contains hardware glitches produced in one of the gathers. In addition a trace has been zeroed. The noise is strikingly non-gaussian and so may be extracted.
(b) We extract the strongest noise by first removing the strongest beds, as in Figure 4(c); the largest amplitudes remaining must be noise. (We determine the cutoff by a forward transform of the section without beds, artificial whitening, and an inverse transform.) The noise in the zeroed trace is the negative of the reasonable events one expects.
(c) Subtracting the strongest noise from the noisy data reveals cleaner, more coherent events.