GEOPHYSICS, VOL. 49, NO. 11 (NOVEMBER 1984); P. 1869–1880, 12 FIGS.

Signal/noise separation and velocity estimation

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

ABSTRACT

A signal/noise separation must recognize the lateral coherence of geologic events and their statistical predictability before extracting those components most useful for a particular process, such as velocity analysis. Events with recognizable coherence we call signal; the rest we term noise. Let us define "focusing" as increasing the statistical independence of samples with some invertible, linear transform L. By the central limit theorem, focused signal must become more non-Gaussian; the same L must defocus noise and make it more Gaussian. A measure F defined from cross entropy measures non-Gaussianity from local histograms of an array, and thereby measures focusing. Local histograms of the transformed data and of transformed, artificially incoherent data provide enough information to estimate the amplitude distributions of transformed signal and noise; errors only increase the estimate of noise. These distributions allow the recognition and extraction of samples containing the highest percentage of signal. Estimating signal and noise iteratively improves the extractions of each.

After the removal of bed reflections and noise, F will determine the best migration velocity for the remaining diffractions. Slant stacks map lines to points, greatly concentrating continuous reflections. We extract samples containing the highest concentration of this signal, invert, and subtract from the data, leaving diffractions and noise. Next, we migrate with many velocities, extract focused events, and invert. Then we find the least-squares sum of these events best resembling the diffractions in the original data. Migration of these diffractions maximizes F at the best velocity. We successfully extract diffractions and estimate velocities for a window of data containing a growth fault. A spatially variable least-squares superposition allows spatially variable velocity estimates.

Local slant stacks allow a laterally adaptable extraction of locally linear events. For a stacked section we successfully extract weak signal with highly variable coherency from behind strong Gaussian noise.

Unlike normal moveout (NMO), wave-equation migration of a few common midpoint (CMP) gathers can image the skewed hyperbolas of dipping reflectors correctly. Short local slant stacks along midpoint will extract reflections with different dips. A simple Stolt (1978) (f-k) type algorithm migrates these dipping events with appropriate dispersion relations. This migration may then be used to extract events containing velocity information over offset. Offset truncations become another removable form of noise.

One may remove non-Gaussian noise from shot gathers by first removing the most identifiable signal, then estimating the samples containing the highest percentage of noise. Those samples containing a significant percentage of signal may be zeroed; what remains represents the most identifiable noise and may be subtracted from the original data. With this procedure we successfully remove ground roll and other noise from a shot (field) gather.

INTRODUCTION

Interpreting noisy data requires recognition of the lateral coherence of geologic events and their statistical predictability. For example, reflections of continuous beds have roughly hyperbolic shapes in field and CMP gathers, and show lateral continuity in common-offset or stacked sections; faults and other abrupt interruptions appear as diffraction hyperbolas. Their lateral predictability means that such events are overspecified. For example, linear reflections could be specified by

Manuscript received by the Editor February 1, 1984; revised manuscript received May 7, 1984.
*Department of Geophysics, Stanford University, Stanford, CA 94305.
‡Dept. of Electronics, Politecnico, 20131 Milano, Italy.

slopes and vertical intercepts rather than depths every 50 m; hyperbolas could be specified by scattering locations and root-mean-square (rms) rock velocities. *Noise then must include components of the data whose description cannot be simplified, components showing no spatial coherence or predictability.* To qualify as signal, *a geologic component must have recognizable coherence.* If an algorithm fails to recognize the coherence, some geologic events may then function as noise. A given component of data (signal or noise) has a limited class of possible appearances; each possibility may be produced by a linear combination of more fundamental linearly independent "events." It will be our business to describe and recognize these events as efficiently as possible.

First we discuss the statistical motivation for signal extraction. Algorithms are outlined with as much generality as seems productive. We then emphasize the extraction of events containing velocity information because they illustrate well two possible complications.

(1) Before extracting the desired signal, remove both noise and the coherent events containing no velocity information.
(2) Estimate unknown velocities from the extracted signal; extract the signal by recognizing the coherence dependent upon these velocities. Yet, the extraction of signal must not bias the estimation of the velocities.

Later we discuss ways of making the description of signal sufficiently general and spatially adaptable. We explore improvements to prestack wave-equation methods and pure applications to separation of signal from noise.

The expense of specific applications will depend only on the chosen linear transformations (such as Stolt migration and the slant stack). Each application is repeated several times but they are easily vectorizable. Statistical calculations manipulate one-dimensional histograms and contribute negligibly to the cost.

FOCUSING AND EXTRACTION OF GEOLOGIC SIGNAL

Let us attempt to extract rather than merely to estimate signal. An estimate ordinarily finds the most probable signal when given the data. An extraction should accept only signal with a predetermined reliability and should eliminate those events significantly contaminated with noise. Alternately extracting signal and noise will place events in their most probable domain.

We establish the following results before describing specific algorithms. Using these results, one may identify some data component as signal (with predetermined reliability), and additionally identify the transform focusing this signal best.

(1) Define "focusing" as a linear transformation making the samples of an array statistically independent.
(2) A transform focusing signal must also defocus noise.
(3) Focused signal becomes more non-Gaussian; defocused noise becomes more Gaussian.
(4) A simple function will measure the local non-Gaussianity, and thereby the focusing of data.
(5) Estimate amplitude distributions of the transformed signal and noise from local histograms.
(6) With these distributions, identify and extract those samples of the transformed data containing most of the focused signal.
(7) To identify the best transform, extract the signal with several transforms and find a superposition best resembling the data. The focusing measure will identify the transform focusing this superposition best.

A measure of focusing

A linear transformation of an array may reduce the number of elements required to describe the signal. We call such a transformation focusing. These elements should be statistically independent, else the number could be reduced further. Most often one does not know the best focusing transform and must acknowledge the presence of some unknown parameters, such as rock velocities. Brute force would make each of these parameters a new dimension in the focusing transform (which must be invertible). A measure of focusing, however, could quickly identify the optimum values.

Describing signal by the smallest number of random variables (parameters to be estimated) allows the simplest statistical tools. Joint probability density functions (jpdf) allow the most arbitrary dependence between variables: the data never contain enough redundancy for their estimation. Marginal probability density functions (mdf) describe each sample independently. If a transformation has focused all variables, then jpdf's may be calculated from mdf's.

The data easily provide enough redundancy to estimate mdf's. An unbiased, robust statistical model should describe the possibilities to be found regionally in the data. Knowledge of one reliable event should increase the probability of finding a similar event nearby. Thus, one not only expects but desires that estimated mdfs change slowly over spatial dimensions and time. Because of this stationarity, a histogram prepared from a great many samples with identical mdfs will describe the possibilities open to them all.

The central-limit theorem requires that a linear transformation of *independent* random variables make the corresponding mdf's more Gaussian, that is, more like the Gaussian or "normal" distribution. By the contrapositive of this theorem, focusing an array with a linear transform must make the mdf's more non-Gaussian. Thus, a measure of the local non-Gaussianity of an array in turn measures the increase in focusing.

We derive a measure F of focusing in Appendix A using equation (A-8b). F measures the divergence of data histograms from the best-fitting Gaussians by the use of Kullback's (1959) cross-entropy. Calculation requires negligible computer time.

Extracting geologic signal

A linear transform focusing signal also defocuses noise. Noise, by our working definition, has no *recognizable* coherence spatially, though possibly over time. Noise includes all events that the chosen transform will not focus. Note that coherent noise such as multiples, sideswipe, ground roll, and cable noise may require wave and geologic models similar to those for signal. Because focused signal becomes more non-Gaussian, energy concentrates about narrow peaks, increasing the percentage of low values. A linear transform L must add coherence to previously incoherent noise. Because of the central-limit theorem, the noise mdf's become more Gaussian. Noise energy diffuses into overlapping coherent events. Thus, applying an invertible L to noise will always decrease F. After the removal of noise, the best L maximizes F.

With mdf's of transformed data and noise, one may estimate the signal present in each sample. Consider three components of the same transformed sample: $d' = s + n'$. Their distributions are related by

$$p_{d'}(x) = p_s(x) * p_{n'}(x). \tag{1}$$

The asterisk denotes convolution. Knowledge of two distributions determines the third. These three distributions determine the expected value of s given that of d':

$$E(s \mid d') = \int x p_{s|d'}(x \mid d')\, dx = \frac{\int x p_s(x) p_{n'}(d' - x)\, dx}{p_{d'}(d')}. \tag{2}$$

Use Bayes' rule twice (c.f., Van Trees, 1968).

An unbiased estimate of these distributions should derive directly from the data. Histograms of the transformed data will estimate data distribution. For a reliable signal extraction we prefer to overestimate the quantity of noise (see Appendix B). Briefly, one may artificially destroy coherence in the data before transformation (without harming marginal distributions) and then accept histograms after transformation as an overestimate of noise. Artificially incoherent signal behaves as noise by defocusing and becoming more Gaussian.

Components with Gaussian mdf's remain Gaussian after linear transformation, whether signal or noise. Because of the pessimistic estimation of the noise distribution, the estimator (2) must treat these indistinguishable components as noise. Extraction can accept only those components that have become more non-Gaussian after transformation, i.e., components that have been recognizably focused. The overestimate of the noise distribution will approach that of the Gaussian component as the most reliable signal and noise are iteratively subtracted.

Calculating the signal distribution requires deconvolution of equation (1), with appropriate constraints of positivity and unit area. To maximize the fit (the probability) of the data, one should minimize the divergence (as measured by cross-entropy) of the data mdf from the convolution of the signal and noise mdf's.

The estimate (2) must be modified somewhat for signal extraction. Zero all samples containing significant percentages of noise to avoid the conversion of any noise to coherent events. Analytic envelopes should be used in extractions (though not when preparing histograms); otherwise an extraction corrupts waveforms by deepening troughs between peaks. As another precaution, smooth an array of the $E(s \mid d')/d'$ values both spatially and temporally before multiplying $\bar{d}'$, thus avoiding sharp edges on events with residual coherence after transformation.

The highest amplitudes of the transformed data most commonly contain the highest concentration of signal. If transformation makes signal more non-Gaussian than noise, then $E(s \mid d')$ often has a characteristic shape: for amplitudes of d' above some abrupt cutoff, $s \approx d'$. Below this cutoff, noise begins to contribute quickly and significantly (c.f., Godfrey, 1979). A good estimate of this cutoff is obtained by examining a high quantile of transformed, artificially incoherent data. Yet such a priori assumptions about the shape of the estimation function are not necessary. Since the data will provide the necessary statistics, they should be used.

Finding the best extraction

Let us next treat a more general transform L, perhaps of higher dimensions, like migration. We extract signal over a range of a transform's parameter, linearly combine inverted events, and find the parameter value maximizing the focusing measure. This procedure both extracts the most reliable signal and determines the best transform.

Let the transform $\underset{\sim}{\mathbf{L}}$ depend upon v. We extract all signal focused at some v and then invert:

$$\bar{e}_v \equiv \underset{\sim}{\mathbf{L}}_v^{-1}\{\text{extract}\{\underset{\sim}{\mathbf{L}}_v\{\bar{d}\}\}\}, \tag{3}$$

where $\underset{\sim}{\mathbf{L}}_v^{-1}$, the least-squares inverse of $\underset{\sim}{\mathbf{L}}_v$, inverts all signal with the parameter v. Let $\underset{\sim}{\mathbf{L}}_v^{-1} = (\underset{\sim}{\mathbf{L}}^*\underset{\sim}{\mathbf{L}})^{-1}\underset{\sim}{\mathbf{L}}^*$, where an asterisk denotes the adjoint. With $\bar{e}_v$ for a physical range of v, find the least-squares sum best resembling the data

$$\hat{s} = \sum_v a_v e_v;$$
$$\min_{\bar{a}} |\hat{s} - \bar{d}|^2 \rightarrow \sum_w \langle \bar{e}_v, \bar{e}_w \rangle a_w = \langle \bar{e}_v, \bar{d} \rangle. \tag{4}$$

The brackets designate a simple scalar product. Set the scalar product equal to zero where the data are unknown to leave these values unconstrained. Solve (4) for all coefficients a_v and for $\hat{s}$ by inverting a symmetric matrix with an order equal to the number of velocities used. $\hat{s}$ contains the most reliable signal, without bias, for the chosen range of v. Transforming $\hat{s}$ over this same range will maximize F at the v focusing the signal best.

In this development we have assumed all signal in the data window corresponds to the same parameter v whereas this dependence may be spatially variable. We add such a dependence to equations (3) and (4) as the applications become clearer.

EXTRACTING DIFFRACTIONS FOR VELOCITY INFORMATION

Many velocity analyses extrapolate wave fields back in time, thereby concentrating signal and dispersing noise. For example, an NMO stack finds the image source of reflections from flat interfaces. Inverse scattering extrapolates wave fields back to residual-velocity scatterers and divides out the effect of extrapolated sources. Genuine seismic events always begin simply and produce increasingly diffuse wavefronts; the increase of thermodynamic entropy requires it. Because the wave equation is symmetric in time, seemingly isolated arrivals such as noise and missing data require more diffuse wavefronts at time zero. That extrapolation that concentrates the signal best while dispersing noise determines the best velocities.

Offset sections span greater distances on the surface and so detect deep velocity changes better than do shot and midpoint gathers. Diffraction events, such as reflections of bed truncations and point scatterers, contain all velocity information present in offset sections. Diffractions appear in the background of sections containing faults as seen in Figure 1a, a section stacked over offset. Migration with correct velocities focuses these diffractions; however, reflections of continuous beds do not focus, and noise defocuses. We must first remove bed reflections (Figure 1b) from the data, thereby exposing diffractions and noise (Figure 1c). We must next extract the most identifiable diffractions from the noise as shown in Figure 1d. Note that phase changes appear at peaks of the hyperbolas as predicted by theory. When these events are migrated, the focusing measure F is maximized at the best velocity (Figure 2). The remaining component of the data (Figure 1e), noise, is neither focused by slant stacks nor migration. Continuous beds, diffractions,

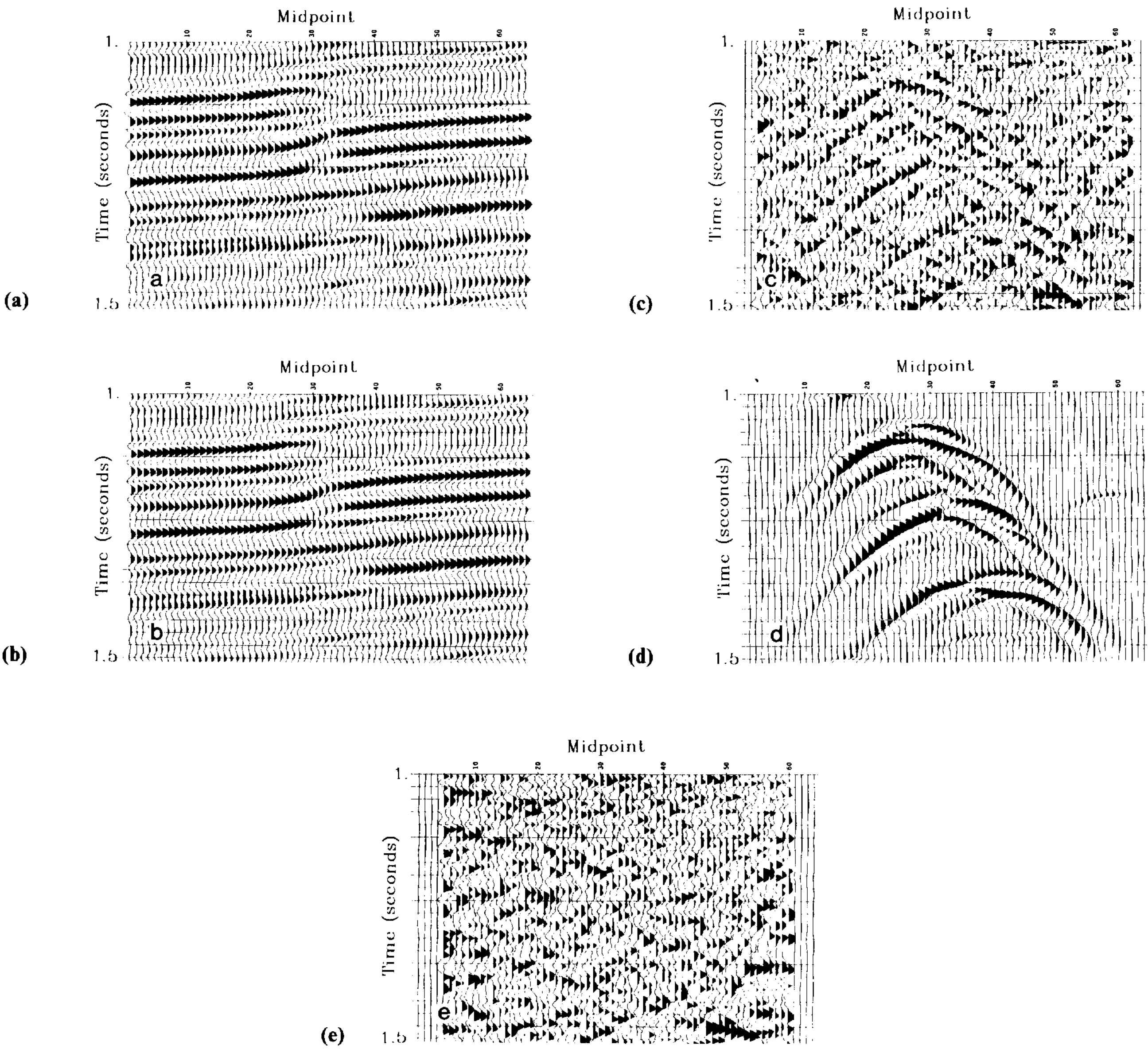

FIG. 1. (a) A window of stacked offshore Texas data contains a growth fault, with weak diffractions off truncated beds. Continuous reflections and noise obscure the velocity information in diffractions. (Data supplied by Western Geophysical Co.: extending 2 km at 33 m sampling.) We decompose this window into three components, (b), (d), and (e), which add up to the original data. (Panels are scaled differently for plotting.) (b) Continuous reflections are easily described as a superposition of lines. (c) Subtracting continuous reflections from data [(a) minus (b)] leaves diffractions and noise, neither easily described as a superposition of lines. (d) Diffractions are easily described as a superposition of diffraction hyperbolas. (e) Noise [(c) minus (d)] cannot be simplified by any invertible linear transformation.

and noise (Figures 1b, 1d, and 1e) add together to produce the original data (Figure 1a). (Diffractions and noise are amplified for plotting.)

Separating bed reflections from diffractions and noise

A window of stacked, offshore Texas data (provided by Western Geophysical Co.) contains a growth fault with weak diffractions off truncated beds (Figure 1a). Reflections of continuous beds may be separated from diffractions and noise after focusing with slant stacks. We define slant stacks by the inverse transform $d(x, t) = \int d'(p, \tau = t - px)$. (See Appendix A on slant stacks.) Slant stacks map lines of constant dip into points, thereby focusing bed reflections laterally. p specifies the slope of the dipping event, and τ denotes the vertical time intercept. Slant stacks may be said to remove the first-order predictability of gently curving events. Coherence remaining after a slant stack shows curvature in the original events. Events with rapidly changing dips, such as noise and diffractions, do not focus. Points map roughly to lines, thereby defocusing noise and diffusing its energy. More general transforms will allow better extractions of continuous reflections. We introduce localness or curvature into our slant stacks in a later section.

The results of previous sections apply directly to the estimation and removal of bed reflections.

(1) Slant stack artificially incoherent data (with randomly reversed traces) and estimate $\hat{p}_{n'}(x)$ locally from histograms.
(2) Slant stack data and estimate $p_{d'}(x)$ locally from histograms. Estimate $p_s(x)$ from step (1).
(3) Evaluate $E(s \mid d')$ for each sample of the transformed data. Smoothly zero samples containing significant noise.
(4) Invert the extracted signal and subtract from the original data (Figures 1b and 1c).
(5) Use this section to reestimate $\hat{p}_{n'}(x)$ and repeat steps (2)–(4).

In practice, a full slant stack is not necessary for step 1. Summing should be performed at enough dip values to provide the local statistics of noise.

Extracting diffractions from noise

To extract diffractions and estimate velocities implement equations (3) and (4) using migration and velocity as the appropriate transform and parameter. The appropriate migration is defined in Appendix D [equation (D-3)]. For the present, assume that all diffractions within the window express the same unknown velocity v. For interval velocities one could estimate velocities in upper windows, then downward continue lower events. Velocities can be estimated as locally as the density of diffractions permits. For particularly deep events the aperture of offsets is much smaller than the depth, so diffractions become the only source of velocity information.

We summarize the extraction of diffractions from noise. Previous methods are assumed for the estimation of the noise and signal distributions.

(1) Migrate the data, without continuous bed reflections, over a physical range of velocities.
(2) For each migrated section, smoothly zero those samples containing significant noise.
(3) Diffract (invert the migration of) each section at the extraction velocity.
(4) Find the least-squares superposition of these diffracted sections best resembling the data without bed reflections (Figure 1d).
(5) Migrate this superposition over the previous range of velocities.
(6) Determine the best migration velocity by evaluating the focusing measure (Figure 2).

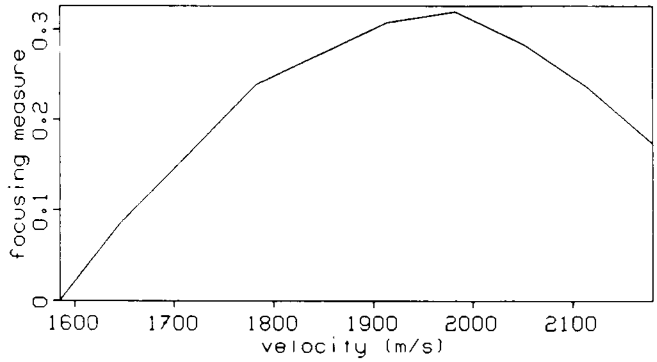

FIG. 2. Migration of the extracted diffractions in Figure 1c maximizes the focusing measure F at the best velocity.

To allow the rms velocities of extracted diffractions to vary vertically and laterally, use the modifications of Appendix E. De Vries and Berkhout (1982) examined the use of varimax norms for estimation of migration velocities. Harlan et al. (1983) emphasized the necessity of the extraction of diffractions.

FURTHER APPLICATIONS AND GENERALIZATIONS

Now let us exploit the generality of the previous sections. We use a local slant stack for a laterally adaptable extraction of locally linear events. Using both slant stacks and migration, velocity analyses of CMP sections can avoid noise and truncation problems and can image dipping beds. The most identifiable non-Gaussian noise may be subtracted from data without harming the signal underneath. Finally, a treatment of static shifts demonstrates how hidden signal coherence can often be recovered.

Extraction of continuous events with a local slant stack

A local slant stack specifies events as a sum of short, tapered line segments of all dips. Global slant stacks require the signal to be easily expressed as a sum of lines extending across the section; such an assumption produces corresponding artifacts when the data do not agree. The local slant stack maps a data set to a narrow cube with the transformation

$$d'(p, x_c, \tau) \equiv \left\{ \iint W\left(\frac{x - x_c}{X_w}\right)\right.$$

$$\cdot\, d[x, t = \tau + p(x - x_c)]\, dx \Big\} * \text{rho}(\tau). \tag{5}$$

$W(x)$ represents a windowing function, i.e., in the simplest case a rectangle function as in equation (A-7). x_c fixes the center of the window, and X_w the width. p is the slope of the short linear event, and τ the vertical time intercept at $x = x_c$. For an explanation of the convolution by a rho filter, see Appendix C on slant stacks. Choose the inverse transformation simply as

$$d(x, t) = \int d'(p, x_c = x, \tau = t)\, dp. \tag{6}$$

Equation (5) decomposes the array into planes containing narrow ranges of dip. Adding these planes together as in expression (6) reconstructs the original data. Frank Rieber (1936) first began applying similar local stacks, called sonograms, to prestack seismic data. We merely present a simple invertible expression for them.

The width X_w of the window specifies the number of traces over which an event should be linear. A natural uncertainty principle results: lateral resolution must be balanced with dip resolution. A larger X_w will improve the dip resolution of genuinely linear events. But wider windows may see greater curvature, increasing the number of p values needed to describe an event. At some critical X_w, analogous to a Fresnel zone, dip resolution will no longer improve.

Some simplifications make the local slant stack inexpensive. For a window of few traces, an $x - t$ implementation becomes less expensive than the Fourier implementation. The loop over p should be the outermost, with sums over various windows and intercepts inside. If the weighting function W is a rectangle, then forward transformation requires linear moveout of the entire section and horizontal sums over narrow windows of traces. The rho filter may be performed first. The sum for a window may be calculated from an adjacent one by the addition of a trace and the subtraction of another (cf., Robinson and Robbins, 1978). This transform becomes only twice as expensive as the equivalent global $x - t$ slant stack.

The great advantage of a local slant stack lies in its resistance to artifacts. Lateral adaptability prevents the oversimplification of focused events and prevents the straightening out or extension of their inverses after extraction. This transform will easily extract events with high curvature such as shallow diffractions if X_w is small, say 4–8 traces (50 m sampling).

A local slant stack successfully extracts weak signal from behind Gaussian noise in Figure 3. Non-Gaussian noise may be successfully removed without harming signal as we discuss in a later section. Gaussian noise, however, never focuses or defocuses after transformation. Only extracting the most reliable signal will ease the interpretation of this section. For the data window in Figure 3a, we extract those events showing significant coherence over at least four traces (Figure 3b). We take $X_w = \textit{four traces}$ and $W(x) = \exp(-\pi x^2)$ in equation (5). We subtract the most reliable signal from the data (Figure 3c). The chosen range of p values excludes that of highly dipping diffraction tails now appearing with the incoherent noise. In this way, extractions with different ranges of slopes will separate overlapping coherent events as well as noise and will even remove unwanted coherent events.

An even more general transformation could express the data as a sum of short second-order curves, defined as

$$d''(p, p', x_c', \tau') \equiv \Big\{ \iint W'\left(\frac{x - x_c'}{X_w'}\right) d'[p, x_c = x, \tau = \tau' + (p' + p)(x - x_c')]\, dx \Big\} * \text{rho}\,(\tau'). \tag{7}$$

This transformation, defined in terms of equation (5), detects linear spatial changes in the slope of events, i.e., the curvature. This more general transformation increases the focusing of curved events for a fixed window width. Alternatively, increas-

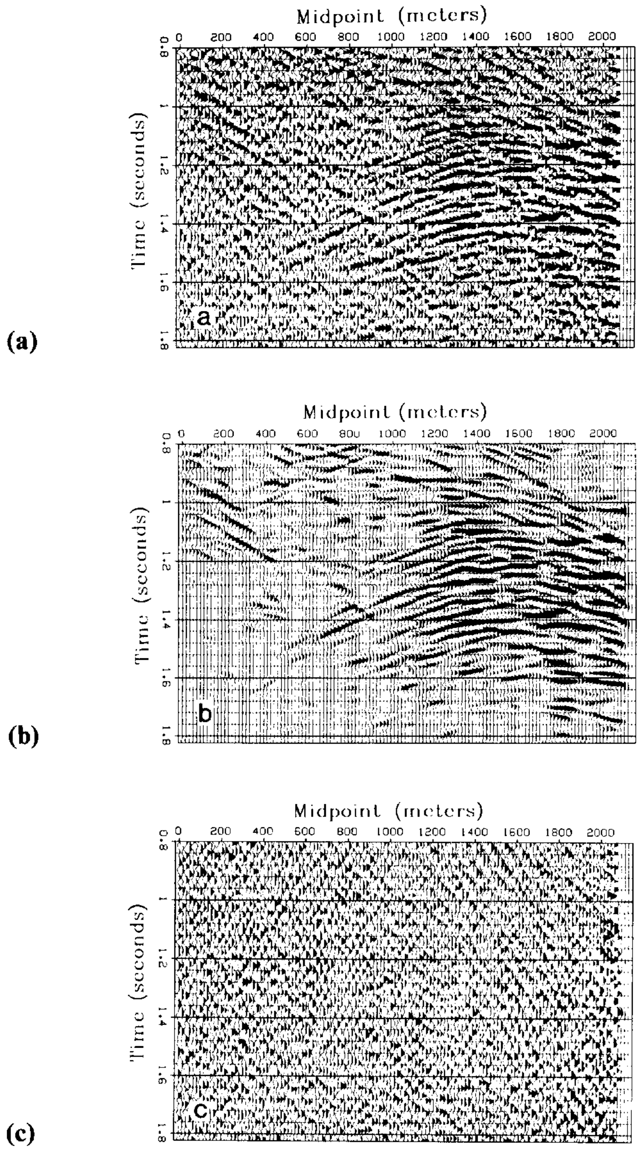

FIG. 3. (a) A window of stacked data contains weak, coherent reflections, much obscured by strong Gaussian noise. (Data courtesy of Trans-Pacific Geothermal.) (b) We extract that component of the data easily expressed as a sum of short, dipping line segments, that signal focused best by a local slant stack. (c) We subtract (b) from (a). The chosen range of p values excluded highly dipping diffraction tails, now appearing with the incoherent noise. (All panels are scaled the same.)

ing the width of windows increases the signal-to-noise (S/N) ratio in focused events. Summing planes over p and p' inverts the transformation.

General, local extractions of coherent events should considerably improve the appearance of stacked sections and sums of traces stretched to look alike. Stacking is generally expected to improve the S/N ratio; however, processors of noisy data often notice that reducing the number of traces in the stack improves the result. A simple guide for stacking results: *do not stack beyond the coherent width of an event*. Otherwise, samples containing only noise will add to the sum. As signal reduces to widths of a few traces, of course, it becomes indistinguishable from random alignments of noise. A sufficiently local extraction would not make signal wider or narrower than statistically reasonable. Stacking extracted signal is superfluous for reducing noise. Slices of the cube should be preferable to projections. Changes of waveforms with offset contain much stratigraphic information.

A closely related process, extrapolation, attempts to fill in missing data intervals in a way consistent with known data. One naturally prefers the simplest possible description of the missing events. Wider extrapolations require more global focusing transformations in order to span known data; the farther extrapolated events must become simpler. The window parameter X_w need only increase smoothly over such intervals.

Wave-equation velocity analyses

Wave-equation velocity analyses of CMP gathers have long been discouraging. Much non-Gaussian noise appears in gathers, including glitches, dead or overamplified traces, and static shifts. A wave-equation migration will image and focus signal, but defocused non-Gaussian noise overwhelms the eye with meaningless coherent events. The truncation of events at the maximum offset acts like severe non-Gaussian noise, containing the negative of the reasonable events one expects. Large amounts of data would seem necessary. Finally, without information orthogonal to the CMP gather, a migration must assume that reflections have no dip, just as will an NMO velocity analysis, the warhorse of conventional processing; estimated velocities suffer enormously.

We tackle each problem with tools already at hand. An signal/noise separation will avoid noise artifacts and allow the focusing measure to estimate velocities. A local slant stack will detect the dip of reflections along midpoint. Four to eight adjacent midpoint gathers suffice. A dip-dependent migration will recognize velocity-dependent coherence over offset. The algorithm follows

(1) Using the local slant stack equation (5), decompose a midpoint gather into events dipping differently over midpoint: $d'(h, t, p_y)$ where h is offset, p_y the dip over midpoint, and t the time at which the event intercepts the gather (equivalent to τ). See Appendix B.
(2) Remove noise by accepting only those samples containing a high percentage of signal.
(3) For each p_y migrate the corresponding gather using the dispersion relation [equation (D-4)] in Appendix D.
(4) Extract those events focused by migration at various velocities and dips.
(5) Find the least-squares superposition of these events best resembling the gather before extraction. Use the depth variable versions [equations (E-1) and (E-2)] in Appendix E.
(6) Estimate migration velocities locally with the focusing measure.

Noise disperses and avoids extraction; extracted events possess longer tails than the original events, extending into the truncated offsets. If the least-squares superposition does not constrain the missing tails of events, then only known offsets will affect the estimation of velocities. If too much signal is replaced by truncation noise, however, focusing will be weak, particularly for deep events with little curvature. Only a crude idea of the moveout of events makes the simplification of (E-3) equally effective.

Frank Rieber's controlled directional sensitivity (1936) first used many of the principles exploited here. Russian seismologists, particularly Dr. Rjabinkin and Boris Zavalishin (pers. comm.), have made Rieber's work the basis of their standard processing. Sonograms, slant stacks without rho filters, assign an angle to the energy arriving at a given time in a narrow window of the data cube. Ray tracing finds the corresponding images. Refinements certainly exist, but imaged reflections do not use waveforms. They successfully image dipping structures in data too noisy for Western methods.

Extracting non-Gaussian noise

Often, for interpretation, one would like to remove the most identifiable noise from a section without harming the signal underneath. We first remove the most identifiable signal from the section by one of the previous methods. An appropriate function $E(n \mid d)$ will then identify samples containing the highest percentage of noise. Those samples containing a significant percentage of signal may be zeroed; what remains represents the most reliable noise and may be subtracted from the original data. Gaussian noise, unfortunately, neither focuses nor diffuses after an invertible transformation and cannot be distinguished from signal. The removable non-Gaussian noise distracts the eye most, however.

The following algorithm seems effective.

(1) Remove the most recognizable coherent events from the data using a (not particularly) local slant stack.
(2) Zero samples containing significant amounts of remaining coherent events; the most reliable noise must remain.
(3) Subtract this noise from the original data.
(4) Repeat if necessary.

Once non-Gaussian noise has been removed, coherent events may be estimated more accurately. At least two iterations are recommended.

We apply this procedure to a shot gather with common noise problems. Figure 4a displays a common shot gather (provided by Western Geophysical Co.) corrupted by strong aliased groundroll, sharp static shifts, and overamplified traces. Figure 4b contains the extracted groundroll and other noise; no reflections are visible. Figure 4c shows the data with groundroll subtracted. Because groundroll possesses substantial coherence

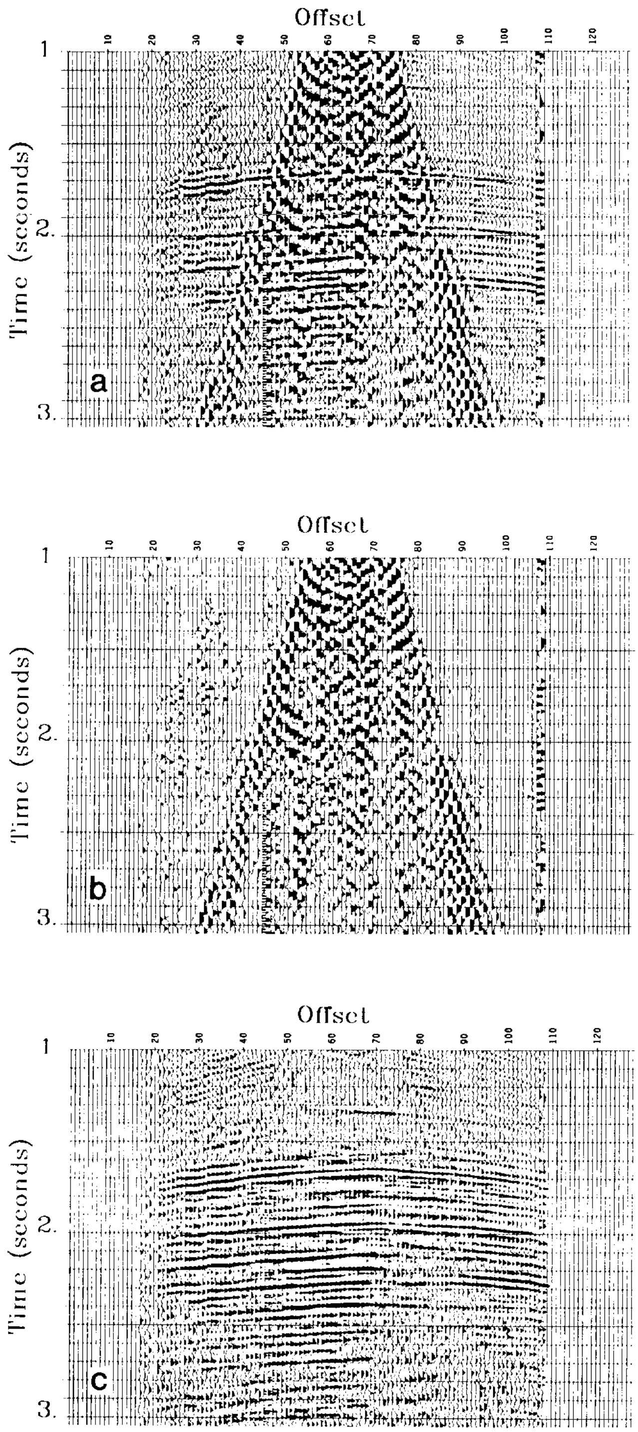

FIG. 4. (a) Strong, aliased ground roll obscures reflections in this window of a common-shot gather. (Data provided by Western Geophysical Co.: extending 4.8 km at 50 m sampling.) (b) By removing the most reliable signal from the data (employing slant stacks), we recognize samples containing a small percentage of signal and extract them as noise. (c) We subtract the most identifiable noise from the original data and uncover previously obscured events. After removing the extracted noise, we repeat the process with a better extraction of the signal. We used two iterations for these results. (Panels have slightly different scaling.)

of its own, a few residual events remain at dip aliases, but certainly most interfering waveforms are gone. Note that events not originally covered by groundroll or other noise are unaffected. A dip filter, by contrast, would have removed high dips from all events.

Land data often contain traces shifted by uneven topography and variations in surface velocities. Almost all information remains; only the exact zero time is uncertain. Yet if not corrected, this trace will interfere as pure noise. A static correction should be that shift of the entire trace adding the most coherency to the data.

Let us first extract the most identifiable signal in the data by a method from previous sections. High-frequency static shifts interfere as noise and do not contribute to the estimate. Let us crosscorrelate original traces with the extracted signal, shifting the original trace by varying amounts and taking the scalar product with the extracted trace. This correlation will appear largest at the static shift time, placing events in the most coherent position. Traces may then be corrected and the procedure repeated for second-order improvements.

CONCLUSIONS

By recognizing the lateral coherence of various geologic events, one may extract the components most useful for a velocity analysis or interpretation. Geophysical events possess distinctive coherence because of wave spreading and reflector continuity. Local statistics are easily estimated from a data array, but statistical dependencies between samples are underdetermined. Thus, signal is most easily estimated when the data samples are made statistically independent (focused) by an invertible transformation. Focusing signal necessarily defocuses noise. Because of the central-limit theorem, focusing makes signal more non-Gaussian and noise more Gaussian. Using cross entropy, one may define a measure of non-Gaussianity and thereby of focusing.

A Bayesian estimator exists to estimate focused signal locally. The estimator requires local probability density functions for signal and noise, easily estimated from local histograms. We extract signal by accepting those focused samples containing a low percentage of noise.

To estimate velocities from diffractions in stacked CMP sections, first remove continuous bed reflections containing no velocity information. Slant stacks focus continuous reflections well but not events with large curvature such as diffractions or noise. We successfully separate diffractions and noise. Some migration, dependent upon velocity, focuses diffractions best. We extract diffractions over a range of velocities and find the least-squares superposition best resembling the original data. The focusing measure shows which velocity focuses this superposition best. A spatially variable least-squares superposition allows local estimations of velocity.

A local slant stack allows extraction of signal most easily expressed as a sum of short dipping lines. We successfully extract weak events with laterally variable coherency from behind strong Gaussian noise. A more general transformation expresses events as a sum of short second-order curves.

The procedure used to extract diffractions may also be used for wave-equation stacks of field (shot) gathers and CMP gathers. Using local slant stacks in the orthogonal direction allows one to recognize various reflector dips and image them correctly in the stack. (NMO does not.)

Finally, one may wish merely to remove noise from data. Gaussian noise cannot be distinguished from signal, but non-Gaussian noise can. First we remove the most identifiable signal and extract those samples containing a high percentage of noise. Then we zero those samples containing a significant percentage of signal and subtract the remaining noise from the original data. We successfully remove groundroll, over-amplified traces, and sharp static shifts from a shot (field) gather.

ACKNOWLEDGMENTS

We thank the sponsors of the Stanford Exploration Project for their support of this research, particularly Western Geophysical Co., and Trans-Pacific Geothermal, who provided the data sets. We also thank Robert Burridge, Alphonso Gonzalez-Serrano, Dave Hale, Rick Ottolini, and Jeff Thorson for many valuable discussions.

REFERENCES

Claerbout, J. F., 1984, Imaging the earth's interior (in preparation).
De Vries, D., and Berkhout, A. J., 1982, Minimum entropy as a tool for velocity analysis: Proc. 52nd Annual International SEG Meeting, Dallas.
Gazdag, J., 1978, Wave equation migration with the phase shift method: Geophysics, **43**, 1342–1351.
Godfrey, R., 1979, A stochastic model for seismogram analysis: Ph.D. thesis, Stanford University, 88 p.
Harlan, W. S., Claerbout, J. F., and Rocca, F., 1983, Extracting velocities from diffractions: Presented at the 53rd Annual International SEG Meeting, Las Vegas.
Kullback, S., 1959, Information theory and statistics: John Wiley and Co.
Ottolini, R., 1982, Migration of reflection seismic data in angle-midpoint coordinates: Ph.D. thesis, Stanford University, 94 p.
Reiber, F., 1936, A new reflection system with controlled directional sensitivity: Geophysics, **1**, 97–106.
Rosenbaum, J. H., and Boudreaux, G. F., 1981, Rapid convergence of some seismic processing algorithms: Geophysics, **46**, 1667–1672.
Robinson, J., and Robbins, T., 1978, Dip-domain migration of two-dimensional seismic profiles: Geophysics, **43**, 77–93.
Shannon, C. E., and Weaver, W., 1963, The mathematical theory of communication: University of Illinois Press.
Stolt, R. H., 1978, Migration by Fourier transform: Geophysics, **43**, 23–48.
Thorson, J., 1978, Reconstruction of a wave field from slant stacks: **SEP-14**, 81–86; available from University Microfilms, Ann Arbor, Michigan.
Van Trees, H., 1968, Detection, estimation, and modulation theory, **1**: John Wiley and Sons.

APPENDIX A
DERIVING THE FOCUSING MEASURE

Cross entropy [defined first by Kullback (1954) as directed divergence] measures the unpredictability of a given $p_1(x)$ with respect to some reference $p_2(x)$:

$$I[p_1(x): p_2(x)] \equiv \int p_1(x) \log [p_1(x)/p_2(x)]\, dx. \qquad \text{(A-1)}$$

If p_2 is uniform then cross entropy becomes the negative of Shannon's statistical entropy. I approaches the minimum value of 0 when p_1 is most like p_2. Define a measure of non-Gaussianity F as the minimum cross entropy of the data mdf's with respect to Gaussian distributions of all variances σ^2. F increases with non-Gaussianity and thereby with the focusing of the data. (Assume zero-mean processes for simplicity in notation. A mean is easily estimated and subtracted.) For a single mdf, define

$$\begin{aligned} F[p(x)] &\equiv \min_\sigma I[p(x): \text{Gaussian } (\sigma, x)] \\ &= \min_\sigma \int p(x) \log \left[p(x) \Big/ \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} \right) \right] dx \\ &= \min_\sigma \left\{ \int p(x) \log p(x)\, dx \right. \\ &\quad \left. + \frac{1}{2\sigma^2} \int x^2 p(x)\, dx + \log \sigma + \log \sqrt{2\pi} \right\}. \qquad \text{(A-2)} \end{aligned}$$

Not surprisingly, F attains this minimum value when the Gaussian distribution possesses the same standard deviation as $p(x)$:

$$\begin{aligned} \frac{d}{d\sigma} I[p(x): \text{Gaussian } (\sigma, x)] &= -\frac{1}{\sigma^3} \int x^2 p(x)\, dx + \frac{1}{\sigma} = 0 \\ \rightarrow \sigma^2 &= \int x^2 p(x)\, dx. \end{aligned} \qquad \text{(A-3)}$$

Substitute this result into equation (A-2) to obtain

$$F[p(x)] \equiv \int p(x) \log p(x)\, dx + \frac{1}{2} \log \int x^2 p(x)\, dx + C, \qquad \text{(A-4)}$$

where $C = \log \sqrt{2\pi} + \frac{1}{2}$. Expression (A-4) provides a simpler working definition of F. Notice that expression (A-4) is scale-invariant: multiplying the random variable x by a constant α does not affect F:

$$F\left[\frac{1}{\alpha} p\left(\frac{x}{\alpha}\right)\right] = F[p(x)]. \qquad \text{(A-5)}$$

Finally, we may prove a posteriori that a Gaussian distribution minimizes equation (A-4). First replace $p(x)$ in equation (A-4) by a perturbed $(1 - \varepsilon)p(x) + \varepsilon\delta(x - x_0)$. $\delta(x)$ is the Dirac delta function with unit area. Then, setting the ε derivative equal to zero yields

$$\begin{aligned} -\int p(x) \log p(x)\, dx + \log p(x_0)\, dx + \frac{x_0^2}{2\sigma^2} - \frac{1}{2} = 0 \\ \rightarrow p(x_0) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x_0/2\sigma 2} \quad \text{and} \quad F[p(x)] = 0. \end{aligned} \qquad \text{(A-6)}$$

Expression (A-6) gives the equation of a Gaussian; (A-3) again defines the variance. The constraint of unit area requires that the measure attain a minimum value of 0.

In practice, evaluate equation (A-4) from discrete histograms functioning as mdf's. Represent the sampled versions as $\{p_i\}$ defining each sample (indexed with i) as an average of $p(x)$ over a short interval of x. Assume that the $\{p_i\}$ are sampled N times per standard deviation, and write

$$p_i = \frac{1}{(\sigma/N)} \int \prod \left(\frac{x - i\sigma/N}{\sigma/N} \right) p(x)\, dx, \qquad \text{(A-7)}$$

$$\prod(x) \equiv \begin{cases} 1 & -\frac{1}{2} \le x \le +\frac{1}{2} \\ 0 & \text{elsewhere} \end{cases}.$$

The sampling reduces $F[p(x)]$ by some ε made arbitrarily small by large N as follows:

$$F[p(x)] = \int p(x) \log p(x)\, dx + \log \sigma + C$$
$$= \sum_i \frac{\sigma}{N} p_i \log p_i + \log \sigma + C + \varepsilon$$
$$= \sum_i \left(\frac{\sigma}{N} p_i\right) \log \left(\frac{\sigma}{N} p_i\right) + \log N + C + \varepsilon. \quad \text{(A-8a)}$$

Let $\{q_i = \sigma p_i / N\}$ be the probabilities of the amplitude bins. If a histogram selects a fixed *N bins per standard deviation*, then the focusing measure equals Shannon and Weaver's (1963) statistical entropy plus constants, i.e.,

$$F[\{q_i\}] = \sum_i q_i \log q_i + \log N + C \quad \text{(A-8b)}$$

The focusing measure requires two inexpensive passes through the data: first to find the standard deviation from a sum of squares; and second, to calculate a coarse histogram scaled accordingly.

APPENDIX B
OVERESTIMATING TRANSFORMED NOISE

Let us extract the most reliable transformed signal from samples with low percentages of noise. Let $\bar{d} = L^{-1}\bar{s} + \bar{n}$, where $\bar{d}$ contains data, $\bar{s}$ focused signal, and $\bar{n}$ laterally incoherent events. L^{-1} is the right inverse of the linear transform L focusing signal. Write the linear transformation $\bar{d}' = L\bar{d}$ as

$$d'_i = \sum_j a_{ij} d_j. \quad \text{(B-1)}$$

Primes specify a component transformed out of its defined domain.

The transformed data and noise distributions are most easily estimated directly from the data. Local histograms of the transformed data will estimate $p_{d_i'}(x)$. Estimating a noise distribution is less straightforward. After transformation (B-1), the mdf's for transformed noise become multiple convolutions of the originals where these convolutions bring the distributions nearer Gaussianity, i.e.,

$$p_{n_i'}(x) = \prod_j * \left[\frac{1}{a_{ij}} p_{n_j}\left(\frac{x}{a_{ij}}\right)\right]. \quad \text{(B-2)}$$

Our notation specifies convolutions of all $p_{n_j}(x)$ with each other. If lateral coherence could be ignored or destroyed in the original data without harming marginal distributions, then the mdf's of transformed noise would not change. For example, randomly reversing the polarity of traces will artificially destroy lateral coherence without altering noise distributions. However, signal would defocus rather than focus after transformation, and resulting mdf's would become double convolutions of the originals, i.e.,

$$\tilde{p}_{s_i}(x) = \prod_{j,k} * \left[\frac{1}{a_{ik}^{-1} a_{kj}} p_{s_j}\left(\frac{x}{a_{ik}^{-1} a_{kj}}\right)\right]. \quad \text{(B-3)}$$

The a_{ik}^{-1} correspond to the inverse transform L^{-1}. Such adulterated signal approaches Gaussianity well. Transformation of artificially incoherent data produces the mdf

$$\hat{p}_{n_i'}(x) \equiv \tilde{p}_{s_i}(x) * p_{n_i'}(x). \quad \text{(B-4)}$$

The variances of the two right distributions are additive: $\hat{p}_{n_i'}(x)$ has a variance guaranteed larger than that of $p_{n_i'}(x)$. In the absence of signal, the signal mdf becomes a delta function and the estimate $\hat{p}_{n_i'}(x) = p_{n_i'}(x)$. $\hat{p}_{n_i'}(x)$ overestimates and prevents noise from appearing as signal. We can reestimate this noise distribution with greater accuracy after extracting the most identifiable signal, and then transform artificially incoherent data and calculate $\hat{p}_{n_i'}(x)$ from local histograms. Calculate $p_s(x)$ from the other two distributions in equation (1). Enough information then exists to find a lower bound for $E(s \mid d')$ in equation (2).

APPENDIX C
IMPLEMENTATION OF SLANT STACKS

We now derive formulations of global and local slant stacks. First we derive a global, frequency-domain version with sufficient dips for inversion. Then we discuss how to avoid possible artifacts from the frequency-domain interpolation.

Frequency-domain slant stacks

For sections of many traces a frequency-domain slant stack becomes less expensive than the space-time version. Lateral wraparound may be avoided with the proper interpolator. Define slant stacks by the inverse transformation

$$d(x, t) = L\bar{d}' = \int d'(p, \tau = t - px)\, dp. \quad \text{(C-1)}$$

Thus, a single point in the slant stack domain $d'(p, \tau) = \delta(p - p_0)\delta(\tau - \tau_0)$ will map to a line in the spatial domain. Two points map to two additively superimposed lines. Let the variables (x, t, p, τ) have the Fourier duals of (k, s, q, ν). Fourier transforming to $d(x, s)$ and $d'(q, \nu)$ yields the simplified relation

$$d(x, t) = \iiint e^{i2\pi\nu(t-px)} e^{i2\pi qp} d'(q, \nu)\, dp\, dq\, d\nu \quad \text{(C-2)}$$
$$= \iiint e^{i2\pi p(q-\nu x)} e^{i2\pi\nu t} d'(q, \nu)\, dp\, dq\, d\nu$$

$$= \int e^{i2\pi\nu t} d'(q = \nu x, \nu)\, d\nu$$

$$\rightarrow d(x, s) = d'(q = sx, \nu = s). \quad \text{(C-3)}$$

This transformation may be performed as a frequency-domain stretch. Unfortunately, in the forward direction a rectangle is mapped to a triangle. Dips not in the range of calculated p's must be aliased. The adjoint transformation, however, implicitly zeros those dips omitted from $\bar{d}'$, i.e.,

$$d'(p, \tau) = L^*\bar{d} = \int d(x, t = \tau + px)\, dx \quad \text{(C-4)}$$

and

$$\rightarrow d'(p, \nu) = d(k = -\nu p, s = \nu). \quad \text{(C-5)}$$

Use equation (C-3) by first performing expression (C-5), then following with an equivalent of the rho filter of Radon transforms (c.f., Thorson, 1978). This filter is the inverse $(L^*L)^{-1}$. The result is

$$\bar{d}'(q, \nu) = L^*L\bar{d}'$$

$$= \iint e^{i2\pi(\nu p)x} e^{-i2\pi qp} d'(q = \nu x, \nu)\, dx\, dp \quad \text{(C-6)}$$

$$= \int \delta(\nu x - q) d'(q = \nu x, \nu)\, dx$$

$$= \int \frac{1}{|\nu|} \delta\left(x - \frac{q}{\nu}\right) d'(q = \nu x, \nu)\, dx$$

$$= \frac{1}{|\nu|} d'(q, \nu)\, dx$$

$$\rightarrow d'(p, \nu) = L^{-1}\bar{d} = (L^*L)^{-1} L^*\bar{d}$$

$$= |\nu|\, d(k = -\nu p, s = \nu). \quad \text{(C-7)}$$

Forward transform with equation (C-7) to avoid aliasing dips and inverse transform with equation (C-3).

p values should be sampled well enough to avoid aliasing frequencies in traces at high x. Equivalently, the sampling in expression (C-7) should satisfy

$$\Delta p \cdot s_{\max} < \Delta k.$$

The discrete equivalent becomes

$$\Delta p < \frac{2\Delta t}{N_x \Delta x}, \quad \text{(C-8)}$$

where N_x is the number of x samples.

Frequency-domain interpolation

The discrete frequency-domain stretch over k requires an interpolation operator to zero implicitly space-domain replications responsible for "wraparound." Interpolation operators are equivalent to convolutions. Convolving by a function over k is equivalent to multiplying by its Fourier transform over x. $d(x, t)$ should be multiplied by a rectangle function that drops to zero before the first trace and after the last. Let us derive a general interpolator and allow arbitrary implicit windowing functions in the other domain. A local slant stack should use a Gaussian window. Stolt migration requires a similar stretch over temporal frequencies; a rectangular window should adjust to the location of zero time.

Define three parameters: x_z, the coordinate to be newly assigned the value of zero in the x domain; x_c, the center (in new coordinates) of the function to multiply the x domain; and X_w, the width of the function to multiply the x domain. Assume the function $W(x)$ to be symmetric about zero. Windowing $d(x)$ with $W(x)$ gives the transform

$$W\left(\frac{x - x_c}{X_w}\right) d(x) \rightarrow \int e^{-i2\pi x_c(k-k')} X_w W'[X_w(k - k')] \times e^{i2\pi x_z k'} d'(k')\, dk'. \quad \text{(C-9)}$$

(We suppress dimensions not being stretched.) $W'(s)$ is the Fourier transform of $W(x)$. Interpolation merely adapts equation (C-9) to the discrete case. Let n_z, n_c, and N_w be the important parameters in samples, and there results

$$d'_{n+\delta n} = \sum_{m=0}^{N_x-1} \frac{N_w}{N_x} W'\left(\frac{(n + \delta n - m)N_w}{N_x}\right) \times e^{-i2\pi n_c(n+\delta n-m)/N_x} e^{i2\pi n_z m/N_x} d'_m. \quad \text{(C-10)}$$

For a global slant stack and for Stolt migration with the surface at the first sample use $n_z = 0$, $n_c = N_x/2$, and $N_w = N_x$. Transform the rectangle function of equation (A-7) into the sine function. The following simplification occurs (cf., Rosenbaum and Boudreaux, 1981):

$$d_{n+\delta n} = \frac{1}{\pi} e^{-i\pi\delta n} \sin \pi\delta n \sum_{m=0}^{N_x-1} d_m/(n + \delta n - m). \quad \text{(C-11)}$$

Equation (C-11) may be tapered to a few terms.

When the windowing function $W(x)$ becomes a Gaussian, one may minimize the number of terms in the interpolation operator. Let $W(x) = \exp(-\pi x^2)$ so that X_w governs the distance between half-amplitude points $[W(0.5) \approx 0.5]$. Then $W'(s) = \exp(-\pi s^2)$. If we preserve all terms in expression (C-10) with coefficient magnitudes greater than 0.01, then

$$|n + \delta n - m| < 1.2 \frac{N_x}{N_w}.$$

The number of terms in the interpolator should be greater than $\sim 3N_x/N_w$.

Note that smaller N_w's allow larger Δp's. When $n_c = 0$, N_w may replace N_x in equation (C-8).

APPENDIX D
FREQUENCY-DOMAIN MIGRATION

For convenience we briefly derive the dispersion relations used for migration in this paper. Constant velocity formulations suffice because our least-squares superposition are spatially variable. Streamlined Stolt (1978) (f-k) or Gazdag (1978) (phase shift) algorithms are the most efficient for multiple constant-velocity migrations.

Let us begin in every case with the double square root (DSR) equation. Assume data are recorded as a function of (s, g, t)

which have the Fourier duals (k_s, k_g, ω). s is the horizontal coordinate of the shot, g of the geophone; t is the arrival time; z is the depth of an imaged reflector, and k_z its dual. Then we have

$$k_z = \frac{\omega}{v}(\sqrt{1 - S^2} + \sqrt{1 - G^2}), \quad \text{(D-1)}$$

where

$$S = \frac{vk_s}{\omega}, \qquad G = \frac{vk_g}{\omega}.$$

See Claerbout (1984) for a derivation and justification of this relation. In short, a single square root derives from the scalar wave equation $\omega^2 = k_z^2 + k_x^2$; reciprocity allows shots to be downward continued just as geophones. No one uses this relation directly. Nevertheless, in theory one could migrate by mapping the data from (k_s, k_g, ω) to (k_s, k_g, k_z) and by imaging at $(s = g, z)$. Ottolini (1982) provides some perceptive use of this relation in various coordinate systems.

Often, midpoint-offset coordinates are more convenient. For such a case $y = (g + s)/2$, $h = (g - s)/2$ and we have

$$k_z = \frac{\omega}{v}[\sqrt{1 - (Y + H)^2} + \sqrt{1 - (Y - H)^2}], \quad \text{(D-2)}$$

where

$$Y = \frac{vk_y}{2\omega}, \qquad H = \frac{vk_h}{2\omega}.$$

A stacked section is a sum of constant-offset sections stretched (by NMO and perhaps dip moveout) to resemble the zero offset. The data then are a function of (y, t) and supposedly invariant over h; thus $k_h \equiv 0$. Migration requires mapping from (k_y, ω) to (k_y, k_z) with

$$k_z = \frac{2\omega}{v}\sqrt{1 - Y^2}. \quad \text{(D-3)}$$

Stolt (1978) and Gazdag (1978) give two widely used algorithms for this mapping, the most common migration. Stolt's is fastest, but Gazdag's will treat depth variable velocities accurately. We use Stolt's method to estimate velocities from diffraction events. These algorithms also apply to the next dispersion relation.

Wave-equation stacks of CMP gathers should recognize dips in the orthogonal direction, along midpoint. We begin with a narrow cube of seismic data $d(y, h, t)$ containing four to eight adjacent midpoint gathers, decompose $d(y, h, t)$ over y and t with equation (5), and select the y_c of the central gather: $d'(y_c, h, t, p_y)$. Signal estimation should follow to discriminate against noise and artifacts from truncation of the data. A given CMP gather contains only events with a known dip along midpoint $p_y = k_y/\omega$. We next map (y_c, p_y, k_h, ω) to (y_c, p_y, k_h, k_z) with

$$k_z = \frac{\omega}{v}[\sqrt{1 - (H + vp_y/2)^2} + \sqrt{1 - (H - vp_y/2)^2}]. \quad \text{(D-4)}$$

Migrating with equation (D-4) as $\mathbf{L}_v$ in expressions (E-1) and (E-2) of Appendix E will provide a depth variable extraction of events containing velocity information. The focusing measure will then identify the best depth variable velocities.

APPENDIX E
EXTRACTING DIFFRACTIONS WITH SPATIALLY VARIABLE VELOCITIES

We prefer to extract diffractions over larger windows where velocities may vary vertically and laterally. The results of the previous section could be applied directly by partitioning the migrated sections and extracting diffractions independently in each. This approach, however, requires troublesome data organization and has particular problems at partition boundaries. Instead let us assume smooth polynomial variations in rock velocities. (Other smooth modulations, such as low-frequency sines and cosines, also apply.)

To allow polynomial variations of diffraction velocity with depth, rewrite equations (3) and (4) to read

$$\bar{e}_v^n \equiv \mathbf{L}_v^{-1}\{gain\ by\ z^n\{extract\{\mathbf{L}_v\{\bar{d}\}\}\}\}, \quad \text{(E-1)}$$

and

$$\hat{s} = \sum_n \sum_v a_v^n \bar{e}_v^n;$$
$$\min_{\hat{a}} |\hat{s} - \bar{d}|^2 \rightarrow \sum_m \sum_w \langle \bar{e}_v^n, \bar{e}_w^m \rangle a_w^m = \langle \bar{e}_v^n, \bar{d} \rangle. \quad \text{(E-2)}$$

An additional gain multiplies by powers of the depth coordinate z. Lateral changes could be added with lateral gains. The least-squares solution may now smoothly change the weighting of extracted events spatially. Solving (E-2) requires inverting a symmetric matrix of an order equal to the number of velocities times the number of gains. Unfortunately, each additional power in the polynomial variation should require another inverse transform $\mathbf{L}_v^{-1}$ for each velocity.

If diffractions with substantially different velocities do not overlap in the time section, however, then the gain and the inverse transform should commute well. The result is

$$\bar{e}_v^n \equiv gain\ by\ t^n\{\mathbf{L}_v^{-1}\{extract\{\mathbf{L}_v\{\bar{d}\}\}\}\} \quad \text{(E-3)}$$

Gains over the time coordinate t may be made implicitly in the scalar products. Thus, the additional inverse transforms may be avoided, and variable velocities may be added cheaply by calculation of additional scalar products.

Since the superposition allows spatially variable velocities, all migrations and diffractions may use constant velocities. To estimate the spatially variable velocities, migrate $\hat{s}$ over the proper range of velocities and evaluate the focusing measure in tapered windows.