Robust velocity stacks

William S. Harlan

DEFINING AN INVERSION

Non-uniqueness makes inversion an act of creation or of selection. If the result is to be examined and interpreted by a man, not another program, then the inversion should not only determine what features are most probable but which are most reliable. An interpretation of a non-unique inversion should begin with the most reliable events.

Define a data set as a superposition of good and bad events. Define "events" as statistically independent physical parameters that model distinguishable features of the data. Let the "signal" be those events one wishes to model. Uninteresting events ("noise") and their possible resemblance to signal must be anticipated.

Define inversion as a choice among events. "Reliable" signal events should not be easily described by a chance combination of noise events. An "inversion" should efficiently describe the data with reliable events. An iterative inversion should converge on the most reliable events first.

We shall now examine the common procedure of velocity (normal moveout) stacking and see how the procedure may be refined as an inversion.

STACKING AS AN INCOMPLETE INVERSION

We begin by replacing the conventional velocity stack with a linear inversion of a simple hyperbolic model. We then examine the non-linear statistical tools that allow us to distinguish signal from noise and to estimate its reliability.

An alternative least-squares inverse

Summing common-midpoint (CMP data along hyperbolic paths does not discriminate hyperbolic events optimally. The conventional normal-moveout stack may be defined as

$$model(\tau,v) \equiv \sum_x data(t = \sqrt{\tau^2 + x^2/v^2}, x) \tag{1}$$

Where x is offset, t time, τ zero-offset time, and v stacking velocity. We assume that geometric changes in amplitude with time have been backed out by a multiplication with t (t^2 often corrects for other dissipations of energy). One desires that each point in the model should describe a single hyperbolic path in the data. Two difficulties prevent this property from being true.

- Hyperbolic events overlap at low offsets. Events are dispersed in the velocity-space model.
- Stacks assume data are measured evenly, without truncations or missing traces. Zeros are implicitly interpolated.

The result is a description of the data with "artifacts" -- hyperbolas not found in the data are created in the model.

Instead, first define a simple forward transform from the model to the data (e.g.):

$$data(t,x) = \sum_v model(\tau = \sqrt{t^2 - x^2/v^2}, v) \tag{2}$$

This equation maps single points to hyperbolas; equation (1) is most definitely not the inverse. Next, define a objective function to optimize the data fit and the model simplicity (e.g.):

$$\min_{model(\tau,v)} \sum_{t,x} \{data(t,x) - \sum_v model[\tau(t,x),v]\}^2 \tag{3}$$

$$+ \alpha \sum_{\tau,v} [model(\tau,v)]^2$$

This functional assumes provides a maximum-likelihood estimate of the model assuming samples of the model and noise are Gaussian and IID (independent and identically distributed). α equals the ratio of the variance of the noise to the variance of the signal (model). The LS inverse makes two improvements over stacking:

- We can fit irregularly sampled data, without assumptions elsewhere, by limiting sums in the objective function.
- We can constrain the model for stability, to suppress those components of the model destroyed by forward transformation.

A synthetic example shows how artifacts can be diminished. Figure 1 contains a synthetic midpoint gather with a single spike of unit amplitude moved out along a hyperbolic path. The correct moveout velocity is one. Sinc interpolations were used. Figure 2 shows the result of applying conventional stacking (equation [1]) and of finding the least-squares (LS) solution (equation [3]) with a conjugate-gradient algorithm. Three velocities were used in each inversion: the correct, 25% high, and 25% low. Note that the conventional stack has much more energy at the incorrect velocities.

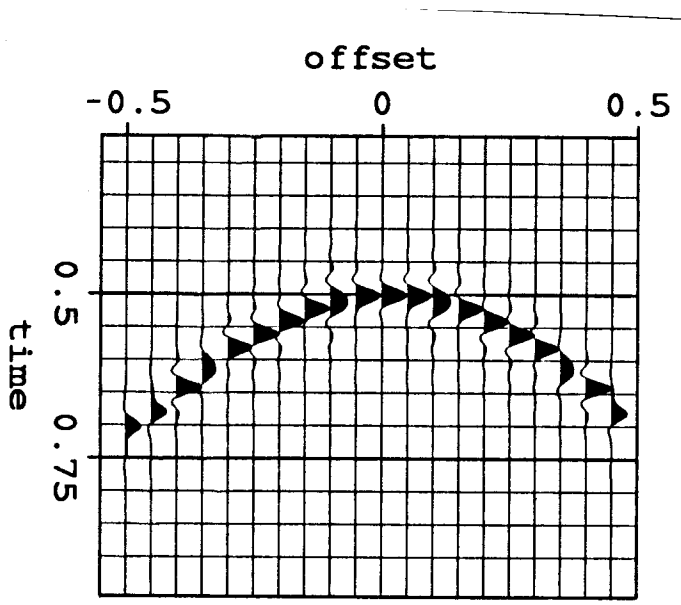


FIG. 1. A synthetic midpoint gather contains a single spike of unit amplitude moved out along a hyperbolic path of velocity 1.

Figure 3 shows the result of recreating the data from the conventional stack with equation 2. Artifacts are now visible as two hyperbolas flanking the correct one. Subtracting this model from the original data leaves strong residuals. Figure 4 recreates the data from the LS model. By contrast, artifacts are scarcely visible; residuals are very small.

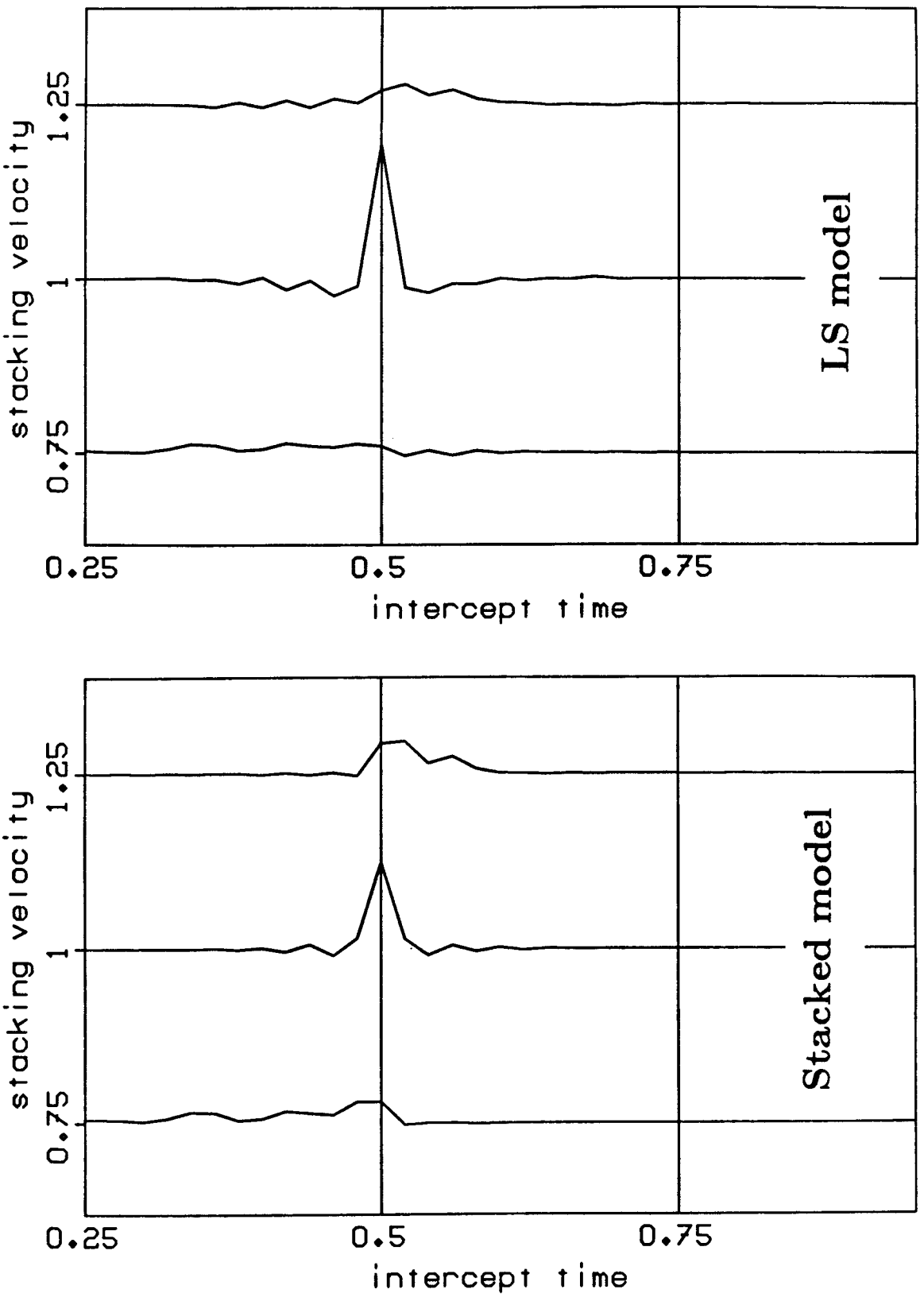


FIG. 2. The synthetic data is inverted for three velocities: the correct, 25% high, and 25% low. The least-squares (LS) inversion of equations (2) and (3) gives better resolution than the stacking of equation (1).

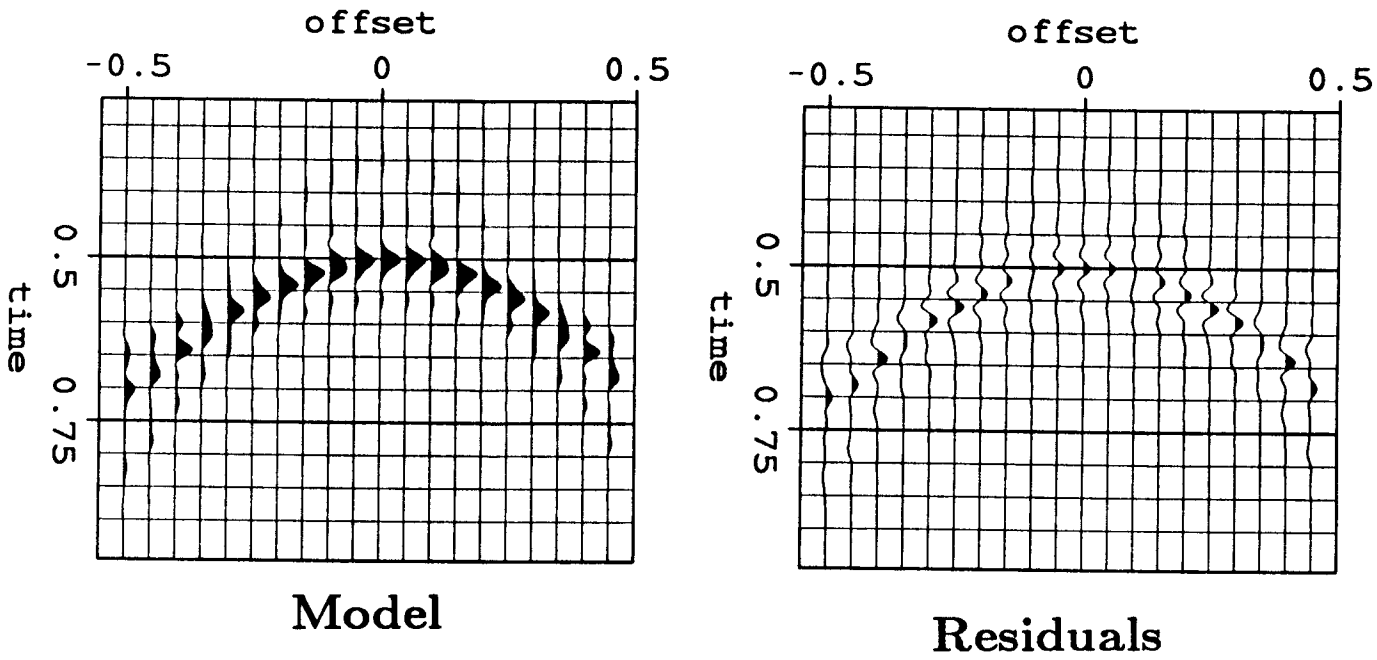


FIG. 3. The stacked model creates data with strong artifacts. Two false hyperbolas flank the correct one.

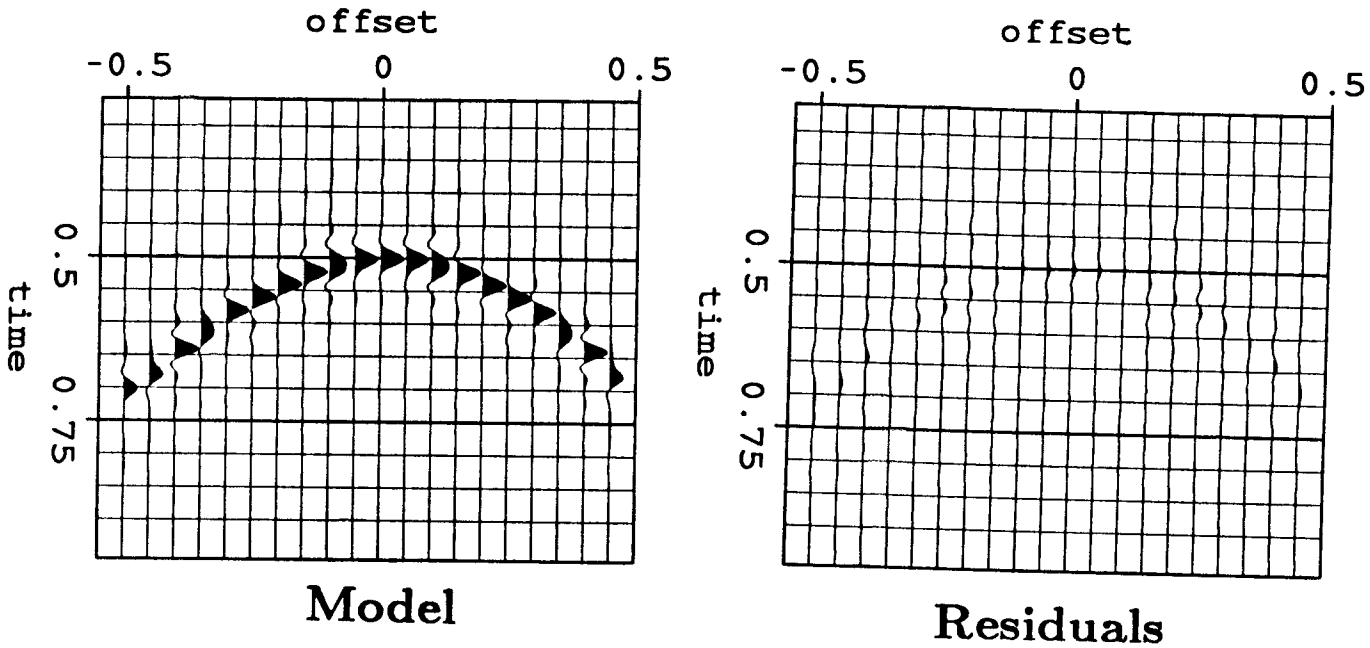


FIG. 4. The LS model creates data with few artifacts.

The LS inversion fits the data as well as need be. However, the stacked traces are far from unique and can be made simpler and more interpretable. The question remains: how reliable are the inverted events? The LS inversion finds the most probable Gaussian signal and noise. True events will most often not be Gaussian, and the Gaussian assumption will distort them. Also an event may be only slightly more probable as signal than as noise. Whether or not they improve the data fit, unreliable events should not be allowed to obscure others

Choosing reliable events

Figure 5 shows a midpoint gather of incoherent noise and the events fitted by LS inversion. The noisy data were created by randomly reordering the traces of the original data in Figure 1. Note that fortuitous alignments have allowed some hyperbolas to be fit. Figure 6 compares the LS inversion of the original data to that of the pure noise. Note that false events in the noise model are equal in strength to the artifacts in the LS model. Thus the artifacts in the original LS inversion (Figure 2) could be simply dismissed as random alignments of incoherent noise.

Scrambling the original data did not affect local statistics of the data (the marginal distributions) before inversion. But, because of the lack of coherence in the data, samples added destructively rather than constructively in the LS inversion. As predicted by the Central Limit Theorem, sums of the incoherent data were more Gaussian than sums of the original data. If the original data had been entirely noise, then scrambling the data would not have affected the statistics of the LS model.

To estimate statistics for the signal and noise, we note two properties:

- If signal and noise are additive in the data, they remain additive in the least-squares inversion, a linear transformation.
- Adding independent random variables convolves their probability distribution functions (pdf's)

A sample of data (d, transformed or not) equals the sum of signal and noise random variables (s and n).

$$d = s + n \;;\; p_d(x) = p_s(x) * p_n(x) \tag{4}$$

where the corresponding probability distribution functions (pdf's) are defined by

$$probability\ [d_0 - \Delta d < d < d_0 + \Delta d] = \int_{d_0 - \Delta d}^{d_0 + \Delta d} p_d(x)\,dx \tag{5}$$

The probability of a random variable falling in an interval is equal to the integration of

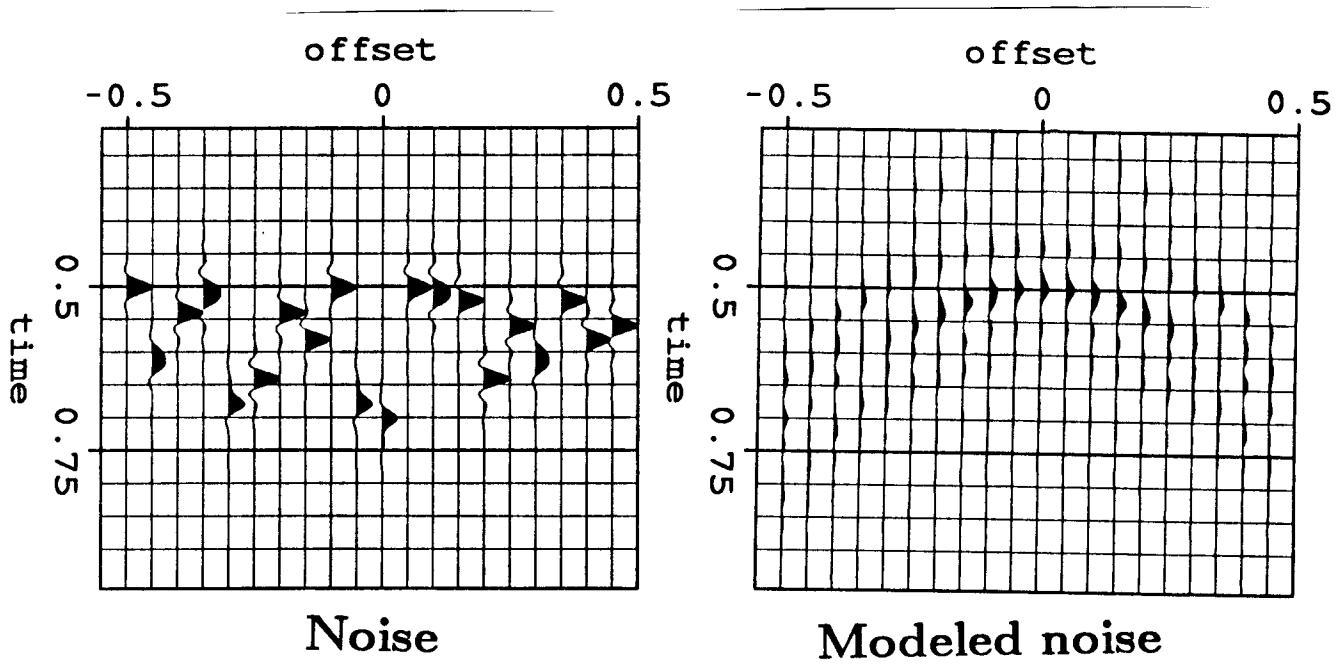


FIG. 5. Invert noise with local statistics equal to that of the data. (The noise randomly reorders the original data traces.) Random alignments have allowed some hyperbolas to be fitted.

the pdf over that interval.

Figure 7 contains histograms of the inverted data and noise. These represent estimates of $p_d(\cdot)$ and $p_n(\cdot)$. The inner peak of the noise distribution is somewhat narrower than that of the data. The data distribution also contains a second peak at high amplitude (0.8), corresponding to the one true hyperbola.

Figure 8 contains an estimate of $p_s(\cdot)$, calculated from equation (4) by deconvolving $p_d(\cdot)$ with $p_n(\cdot)$. Assuming $p_n(\cdot)$ to be exact, we find the $p_s(\cdot)$ that makes the data distribution most probable. This maximum likelihood estimate equivalently minimizes the cross-entropy of $p_d(\cdot)$ with respect to $p_s(x) * p_n(x)$. See the appendix for equations. The resulting convolution $p_s(x) * p_n(x)$ (Figure 8) does not and cannot fit the data distribution $p_d(\cdot)$ perfectly because the noise distribution is broader than the secondary peak (at 0.8) in $p_d(\cdot)$.

With the necessary distributions $p_s(\cdot)$, $p_n(\cdot)$, and $p_d(\cdot)$, we can now estimate the most probable values of transformed signal with known reliability. Figure 9 displays the

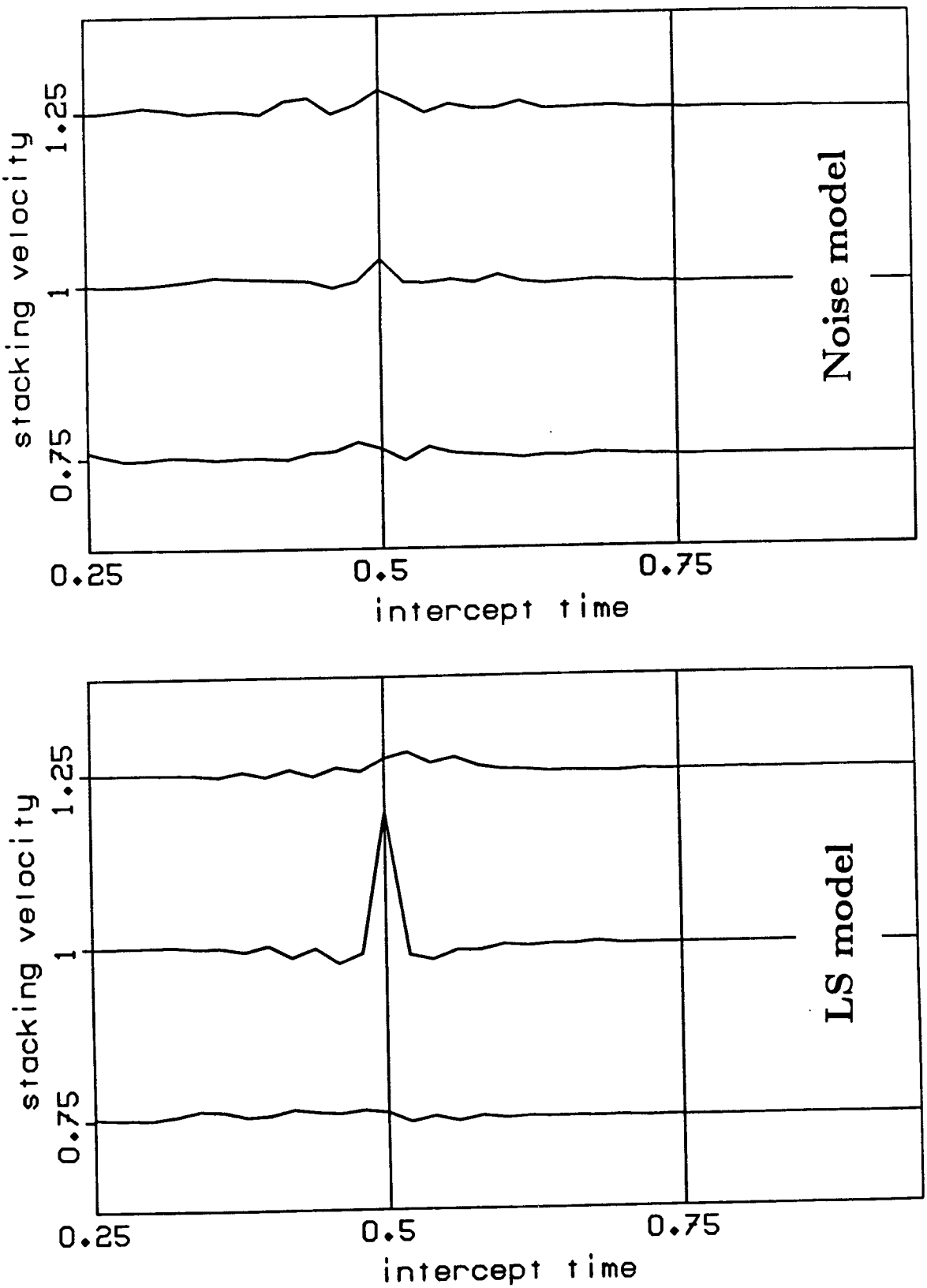


FIG. 6. The sidelobes of the LS model are as strong as those of the inverted noise.

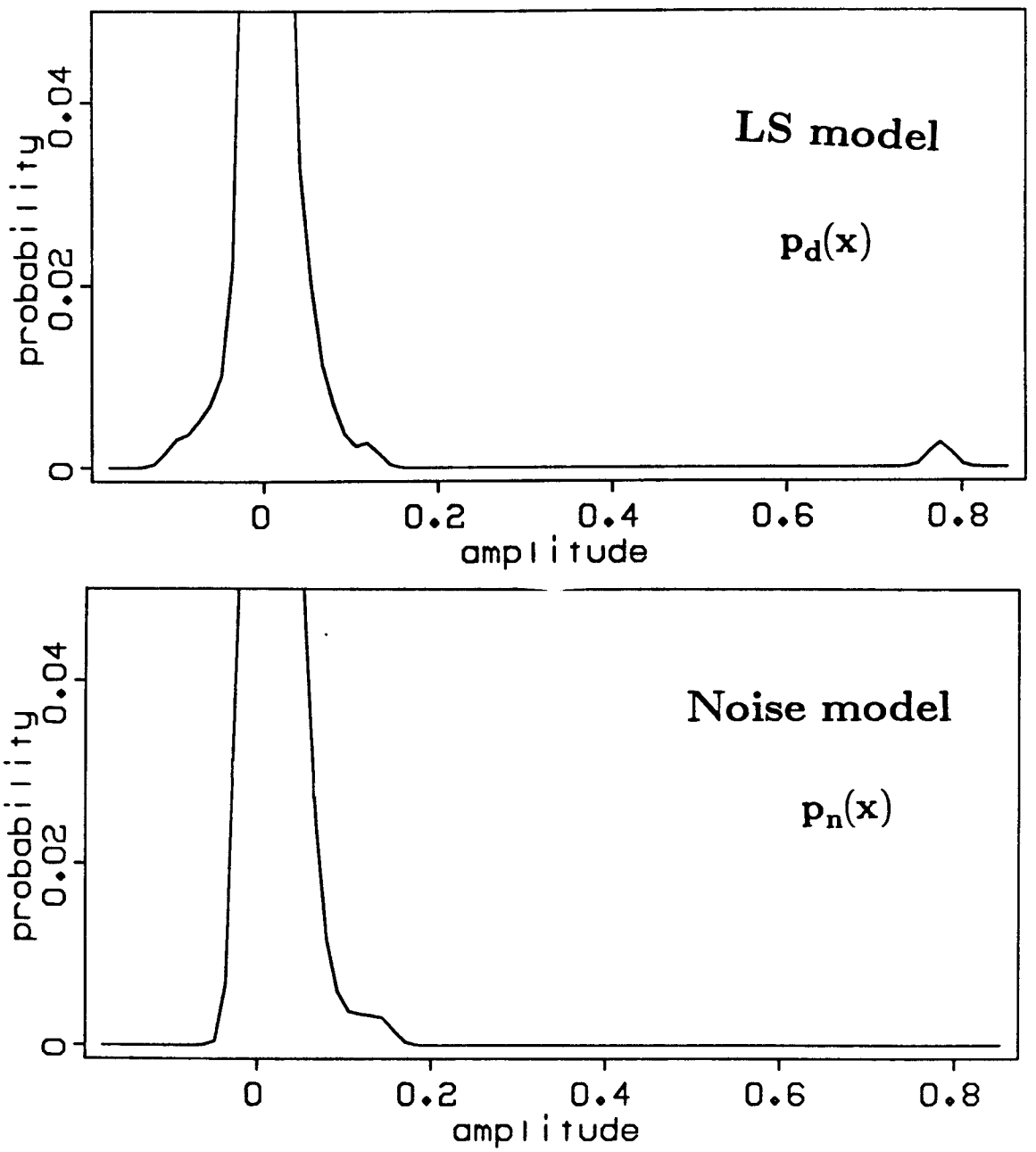


FIG. 7. Histograms of the LS and noise models: the noise histogram overestimates the noise in the LS model. These histograms will be used as (pessimistic) estimates of probability distribution functions (pdf's) for the transformed data and noise. The transformed data shows a sample at high amplitude (near 0.8) corresponding to the one true hyperbola, clearly improbable as transformed noise.

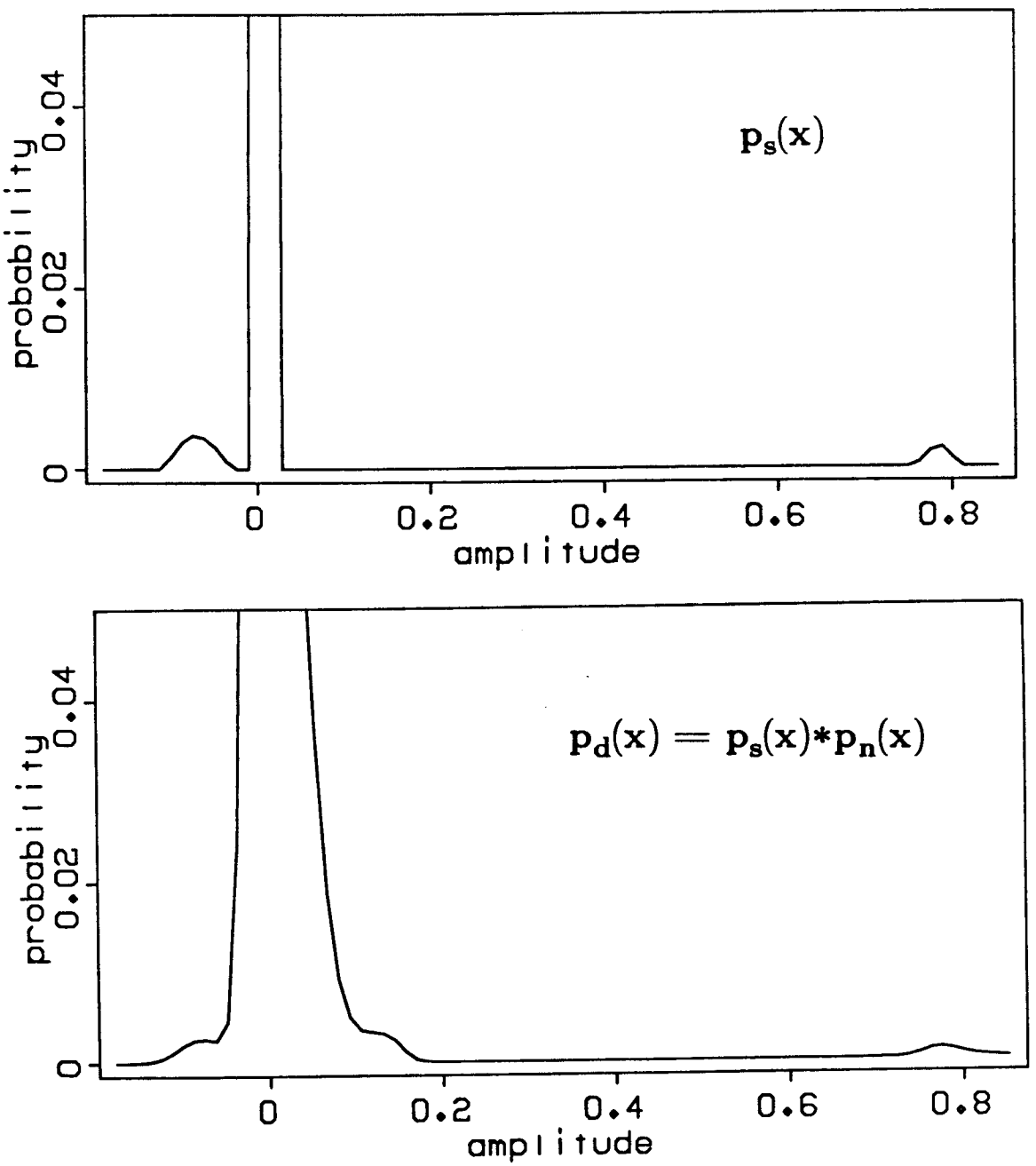


FIG. 8. A maximum-likelihood estimate of the pdf for transformed signal implies a distribution for the sum of signal and noise. Because signal and noise remain additive and independent in the LS inversion, their pdf's will convolve. The implied distribution cannot match the histogram for the transformed data perfectly.

Bayesian estimate of the signal present in a sample of transformed data, containing signal plus noise (equation in appendix). For most amplitudes of the transformed data, the most probable amplitude of the signal is zero. The most striking exception appears at the high amplitude peak (near 0.8) corresponding to the true hyperbola. This peak is determined to be all transformed signal. Applying the Bayesian estimate to the data (Figure 10) suppresses much of the energy of sidelobes. Some spurious events remain.

Figure 11 displays the "reliability" of the Bayesian estimate: the probability that the percentage error is less than 5% (equation in appendix). The reliability of the peak with amplitude 0.8 is seen to be almost unity. Other events are much less reliable. We extract those signal events with greater than 95% reliability (Figure 12). Only the central correct peak survives, without sidelobes. Recreating the data from this model still fits the data well (Figure 13), as did the LS inversion in Figure 4. Yet the LS model of Figure 2 minimized the LS functional of equation (3) more effectively than Figure 12. Because the LS functional incorrectly assumed a Gaussian model, resolution was lost as energy was dispersed into greater Gaussianity.

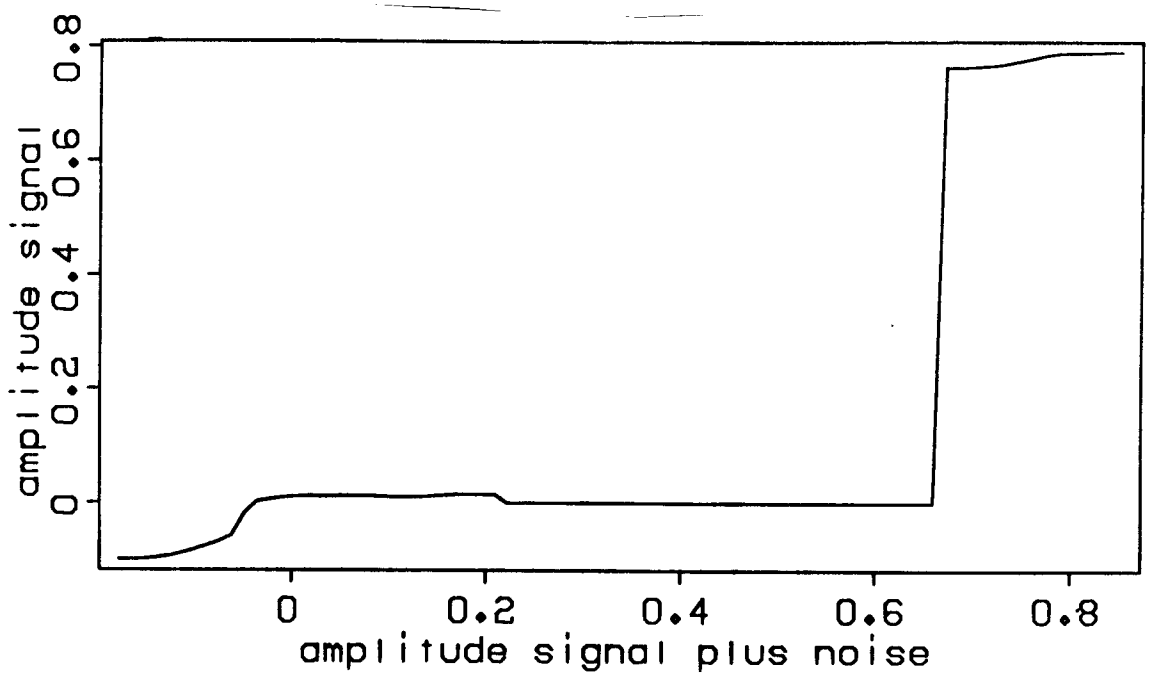
MORE ELABORATE VELOCITY STACKS

In this section we define two alternative NMO stacking operators, each of which uses the algorithm of the previous section (and appendix). The first, a radial stack, attempts foremost to represent amplitude changes with offset clearly in the inverted model. The second, an offset-local stack, will stress fitting the local non-stationarity in the data to distinguish the signal and noise best.

A radial stack

Our first alternative is designed with an eye to improving the quality and interpretability of the stacks. Fitting the data well will be of secondary importance.

If changes in amplitude with offset are of use to interpreters then stacked sections should preserve this information. Hyperbolic events contain amplitude changes with offset because of (1) geometric spreading, (2) dispersion of waves from lateral velocity anomalies, and (3) changes in reflection coefficients with offset. We have assumed that multiplying traces by a power of the arrival time has corrected for geometric spreading. The other two causes, however, cannot be so easily parametrized. These anomalies, particularly the third, are most evenly distributed in arrival angles. In a constant-velocity, stratified medium, constant arrival angles are equivalent to constant radial angles, defined by $r = x/t$, where x is arrival offset and t arrival time. We imagine first a small number of stacked sections representing known ranges of r. As in the previous



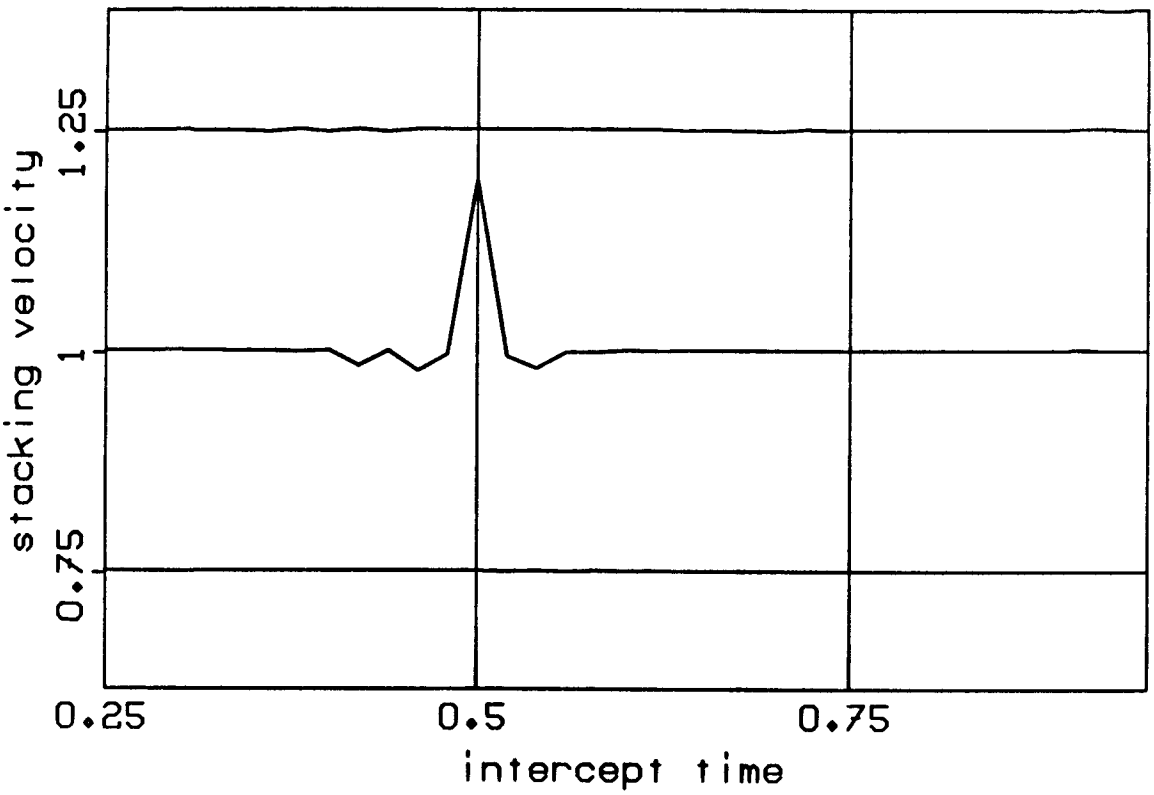
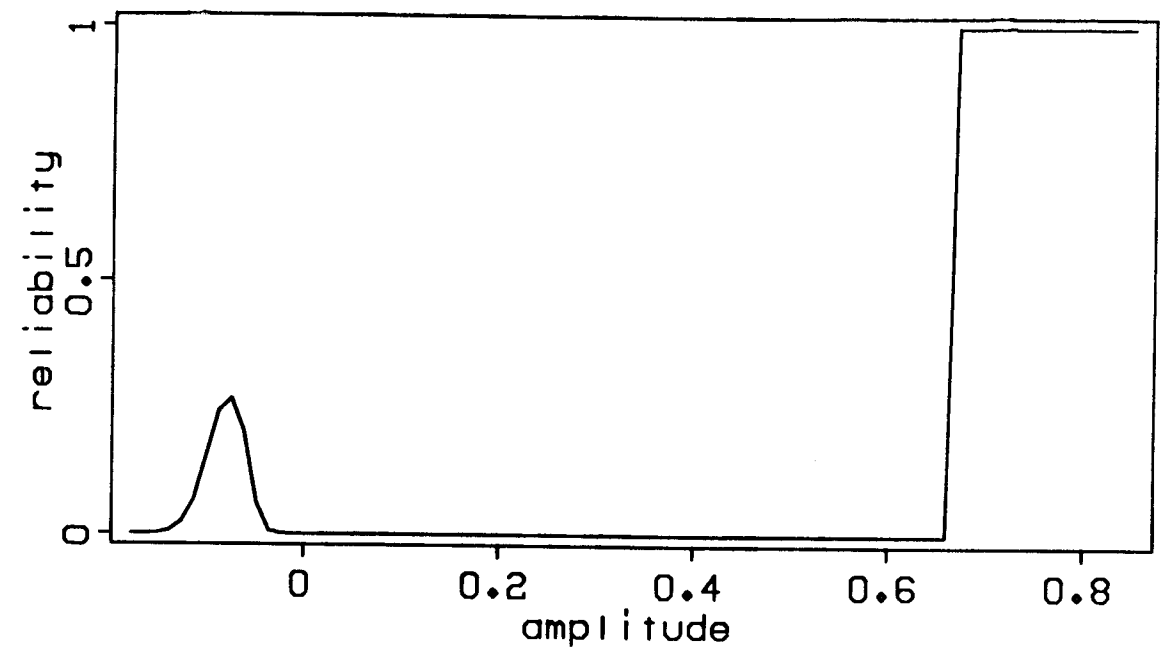


FIG. 9. A Bayesian estimate gives the expected value of signal when the signal plus noise is known. Most transformed data amplitudes are expected to have nearly zero signal, excepting the high amplitude (near 0.8).

FIG. 10. The Bayesian estimate suppresses much of the energy of sidelobes. Some spurious events remain.



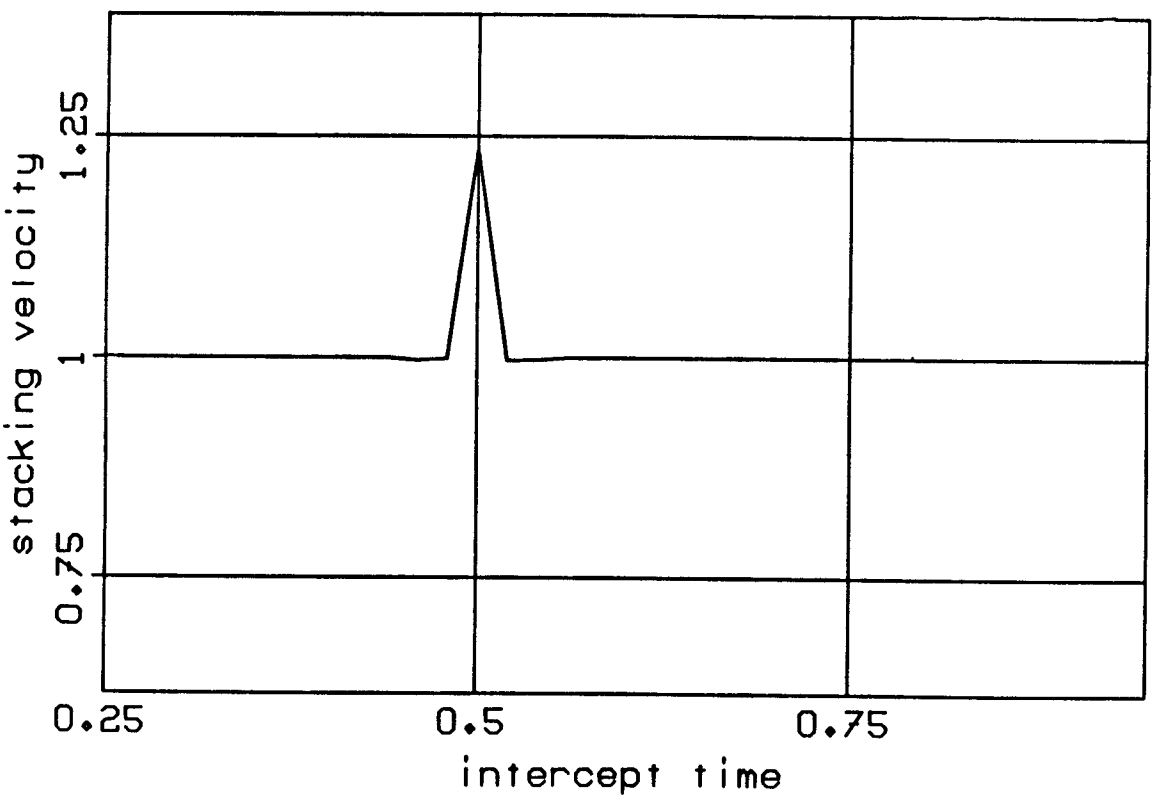


FIG. 11. The reliability of the Bayesian estimate is the probability that noise is less than 5% the estimated signal. The peak with an amplitude near 0.8 has a reliability of almost one; other events are much less reliable.

FIG. 12. Extracting those events with greater than 95% reliability eliminates all but the central correct peak, without sidelobes.

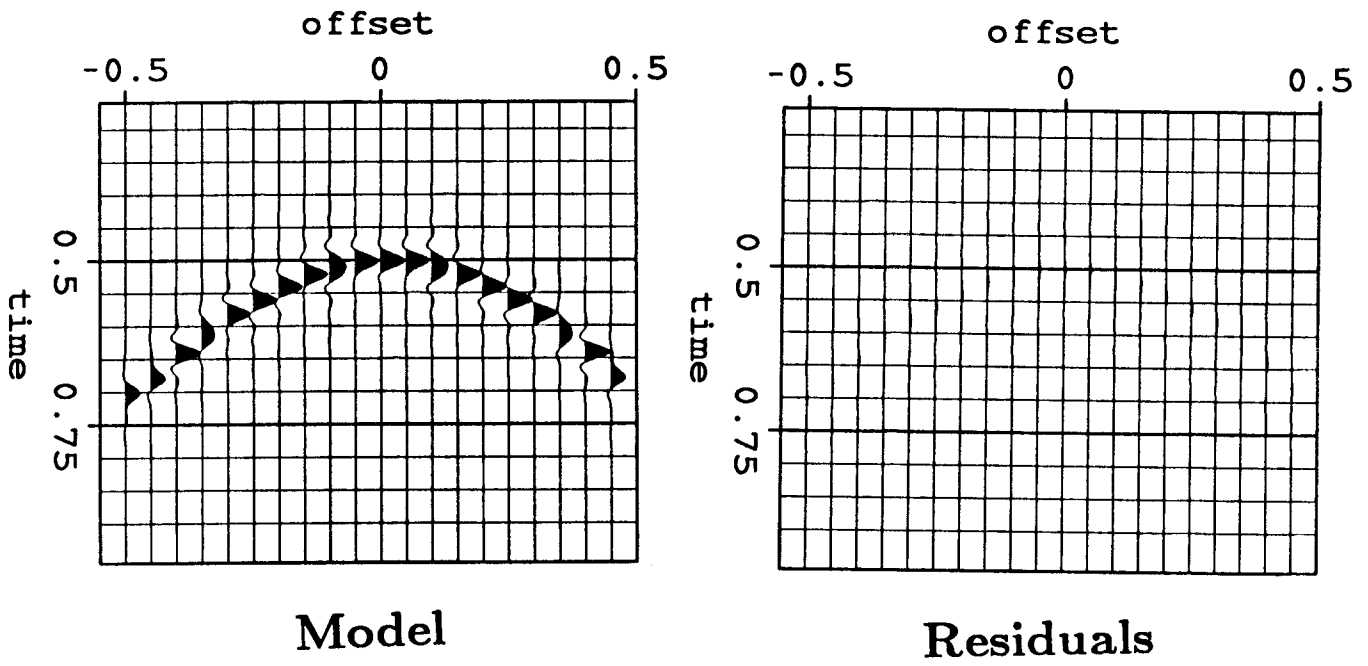


FIG. 13. The reliable events fit the data very well.

section, we will define the stack as the inverse of a simple modeling equation:

$$data_{x,t} = \sum_{r} weight_{r,x,t} \sum_{v,\tau} moveout_{v,\tau,x,t}\, model_{v,\tau,r} \tag{6}$$

where

$$moveout_{v,\tau,x,t} \equiv \delta(t - \sqrt{\tau^2 + x^2/v^2}) \tag{7}$$

$$weight_{r,x,t} \equiv W(\frac{r - x/t}{\Delta r}) \tag{8}$$

Let $W(\cdot)$ be a symmetric windowing function of unit area, such as the triangle function:

$$\Lambda(x) \equiv \begin{cases} 1 - |x| & |x| \leq 1 \\ 0 & |x| > 1 \end{cases} \tag{9}$$

Figure 14 windows a common-shot gather supplied by Western Geophysical. The reflecting beds are not perfectly flat . Near surface anomalies distort the hyperbolas locally and make the event amplitudes variable with offset. The groundroll is present as

strong, non-Gaussian additive noise.

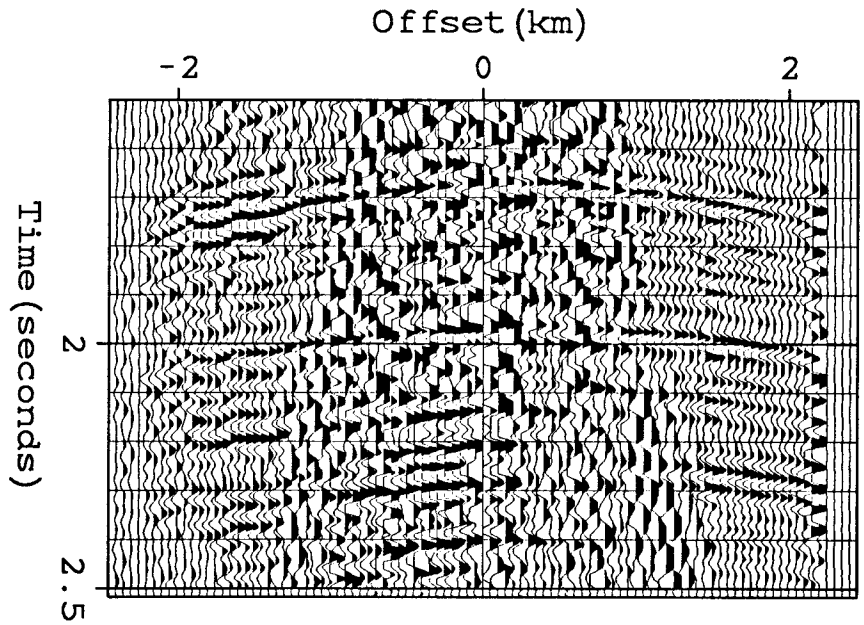


FIG. 14. A common-shot gather supplied by Western Geophysical. Reflecting beds are not perfectly vertical. Near surface anomalies distort hyperbolas locally and make amplitudes variable with offset. Groundroll is strong additive noise.

Figure 15 displays a decomposition of the data into three ranges of r, without the summing over r in equation (6):

$$decompostion_{r,x,t} = weight_{r,x,t} \sum_{v,\tau} moveout_{v,\tau,x,t}\, model_{v,\tau,r} \tag{10}$$

$\Delta r = 1$. The velocities were 2.5, 3.0, and 3.5 km/s. We shall not consider the quality of the stack here, but rather the fit with the data. When the panels of Figure 15 are summed together as in equation (6), the defects become very visible (Figure 16). The three ranges of r do not merge well. Subtracting the sum from the original data leaves a considerable number of hyperbolic events behind (Figure 17). These events will interfere with any subsequent extraction of the groundroll. The radial stack has the advantage of keeping the inverted model to a size that is easily managed by an interpreter. To account for most of the non-stationarity of the hyperbolic events, however, requires

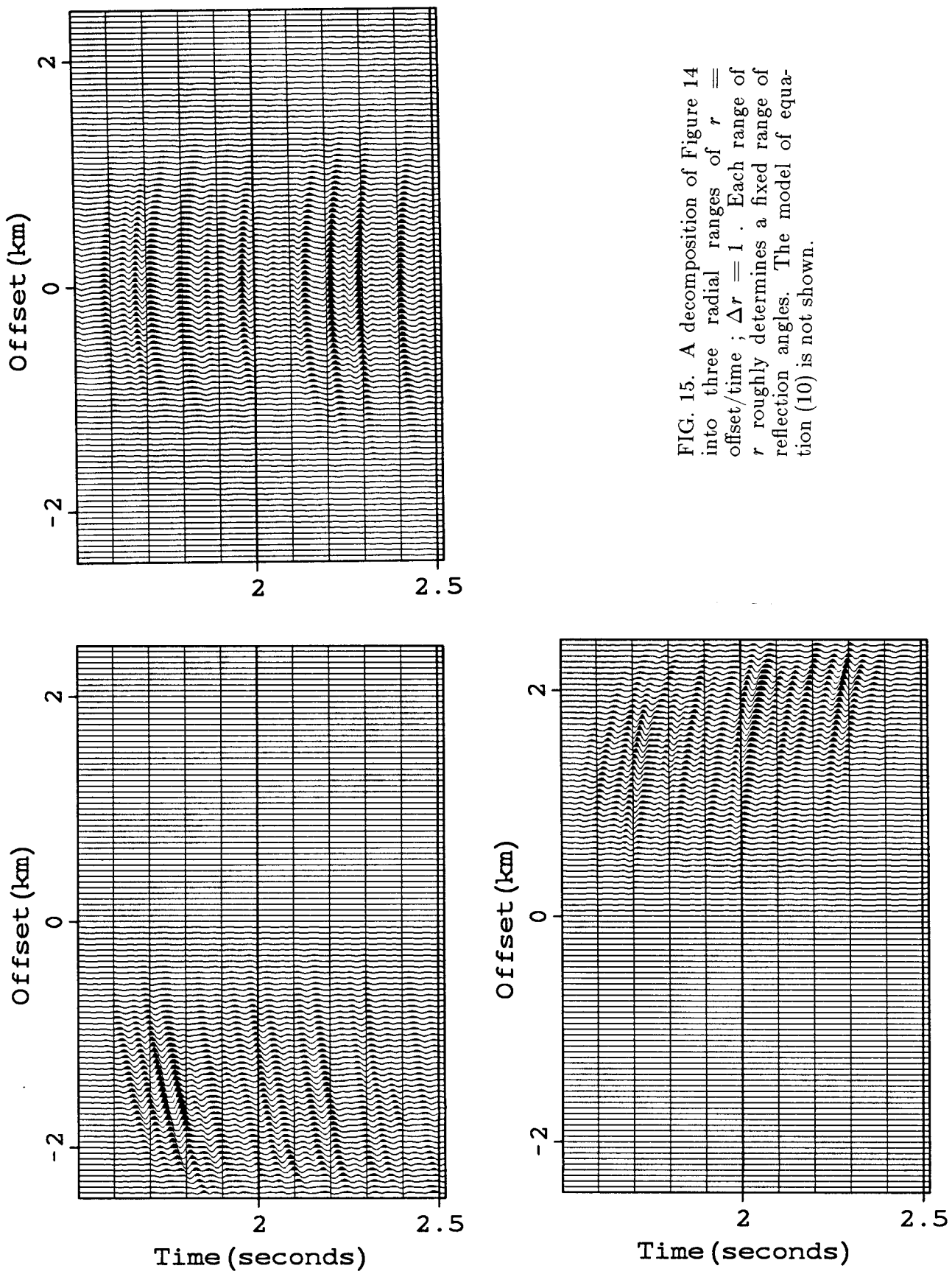


FIG. 15. A decomposition of Figure 14 into three radial ranges of r = offset/time ; $\Delta r = 1$. Each range of r roughly determines a fixed range of reflection angles. The model of equation (10) is not shown.

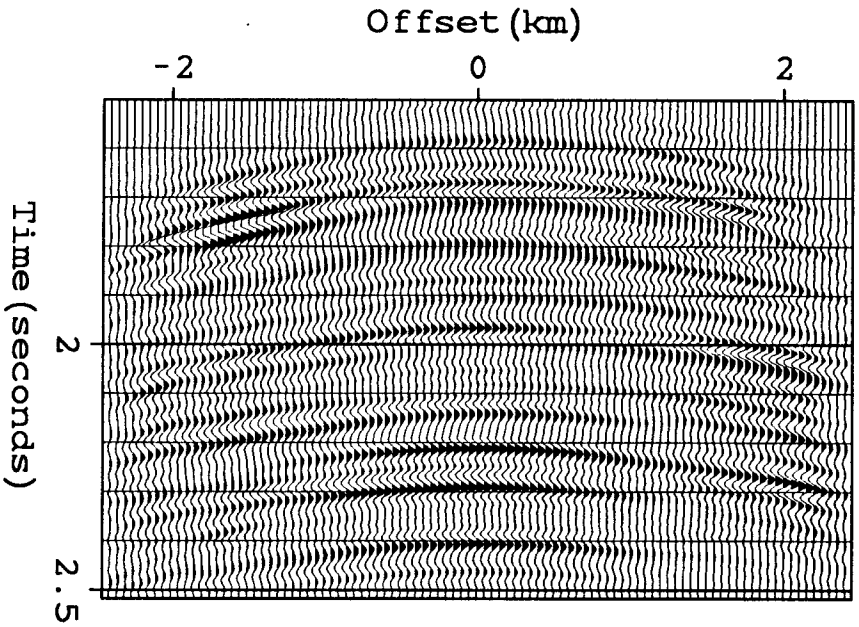


FIG. 16. The three ranges of r do not merge well when the panels of Figure 15 are summed together as in equation (6).

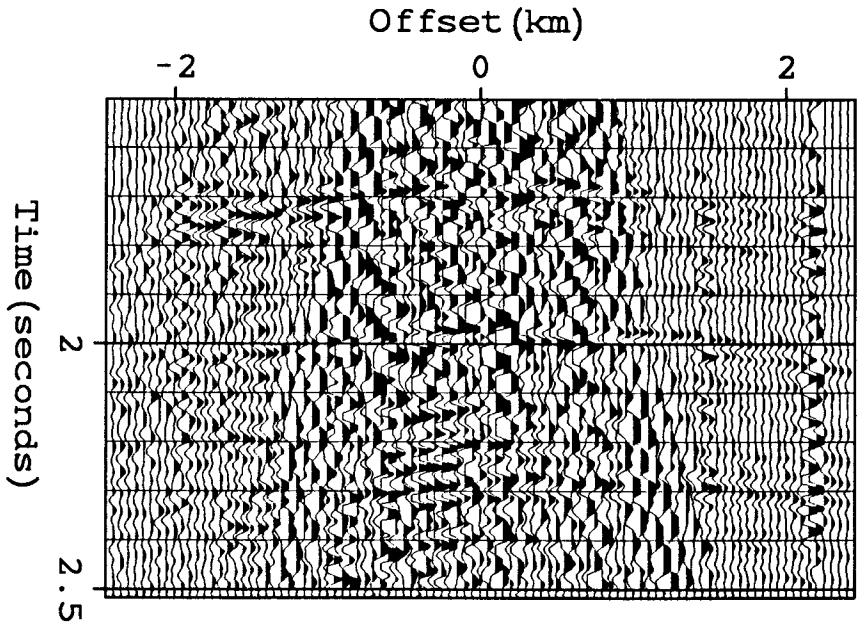


FIG. 17. Many hyperbolic events remain after subtracting the modeled data (Figure 16) from the original data (Figure 14).

considerably more samples in the model.

A offset-local stack

If the object of our stack is a flexible description of the signal for signal/noise separation, rather than an interpretable model, then the size of our model is unimportant. Though physical variations in amplitude and non-hyperbolicity are more or less evenly distributed in the radial coordinate r, the redundancy necessary to discriminate signal and noise is evenly distributed in offset. Let us define an alternative equation for the modeling of the hyperbolic events.

$$data_{x,t} = \sum_{v,\tau} Moveout_{v,\tau,x,t} \sum_{x'} Smooth_{x,x'} \; model_{x',v,\tau} \tag{11}$$

"*Smooth* " is defined as a low pass convolution such as

$$Smooth_{x,x'} = \Lambda(\frac{x' - x}{\Delta x}) \tag{12}$$

Each point in the model maps first to a tapered line segment, then to a hyperbola segment. These hyperbola segments must overlap so continuously that seams will not appear. This model will contain considerably more samples of x' than equation (6) contained of r. Stacks of constant x' will also have less physical basis for interpretation.

Figure 18 shows the result of modeling the data with this equation, using a Butterworth lowpass filter and $\Delta x \approx 10$ traces. Events are indeed more continuous and adaptable to the changing coherence of the hyperbolic signal. No groundroll has been extracted as locally hyperbolic. Figure 19 displays the data minus Figure 18. Considerably fewer hyperbolic events remain than in Figure 17.

We now may regard the groundroll as signal to be extracted and treat all other residuals as noise. We now define the noise to be the former model of equation (11). We extract as signal all samples of the data which cannot be easily described as a sum of hyperbola segments. In order to respect the oscillatory nature of the time samples, extractions were made on analytic traces obtained from Hilbert transformation. The analytic trace was created by zeroing the negative frequencies of the real time function After extraction, the negative frequencies were zeroed again, giving the best fitting analytic trace (in a least-squares sense). The real part was preserved for plotting.

Figure 20 displays the extracted events, containing, in diminishing strength, groundroll, overamplified traces, strong events with static shifts. Subtracting the unwanted events from the original data of Figure 14 exposes much of the previously hidden hyperbolic events (Figure 21). Yet the data has remained untouched in a majority of the

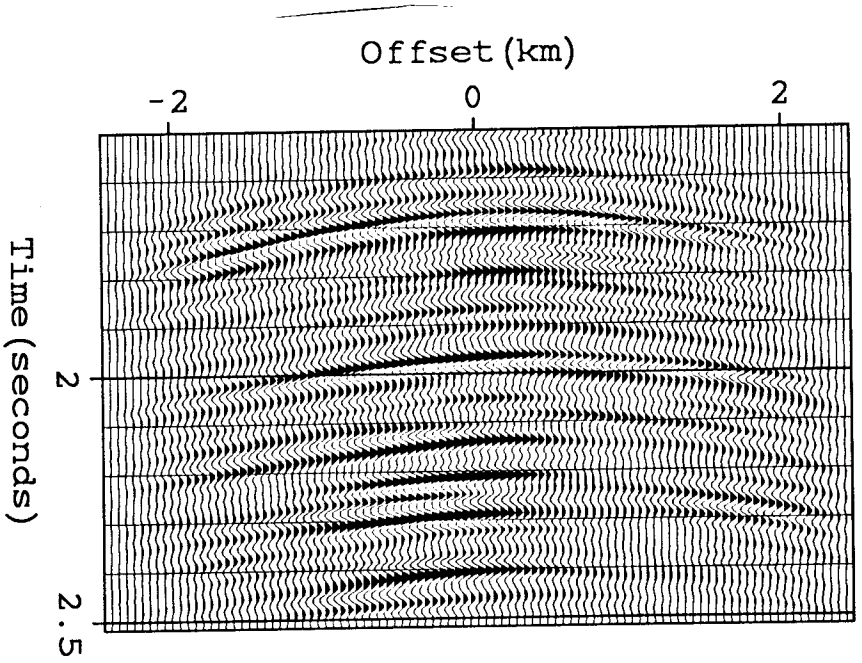


FIG. 18. An offset-local stack (fitting the data with equation [11]) fits the data more smoothly than the radial stack.

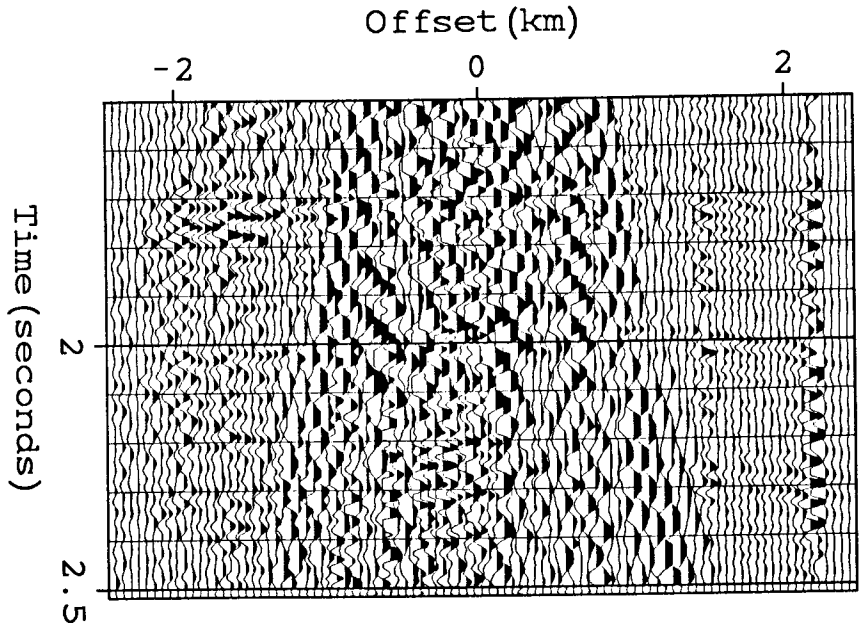


FIG. 19. Considerably fewer hyperbolic events remain than in Figure 14 after subtracting Figure 18 from the data of Figure 17.

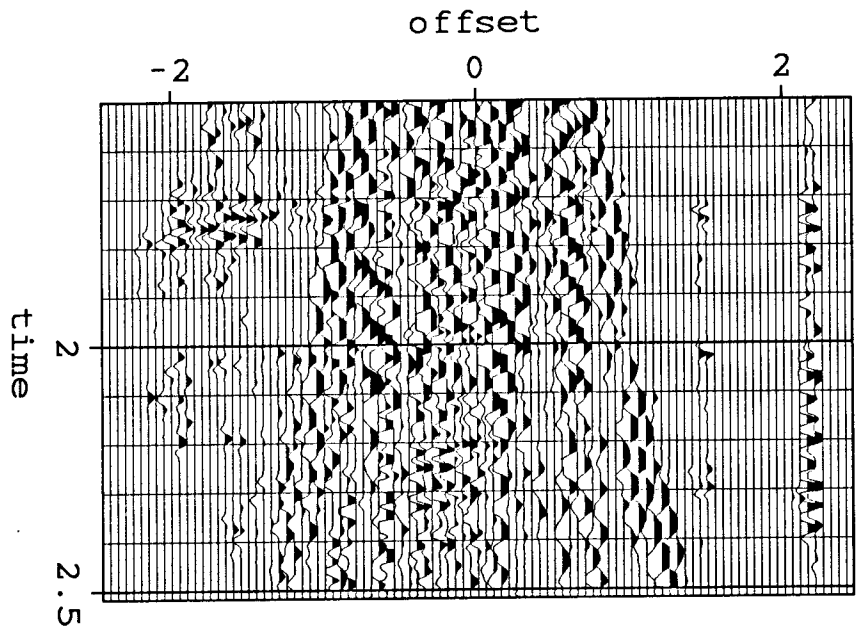


FIG. 20. We extract those samples containing signal in a sufficiently low percentage with a sufficiently high probability. Groundroll, overamplified traces, and strong events with static shifts appear. (Analytic traces were used to represent true amplitudes of the oscillations.)

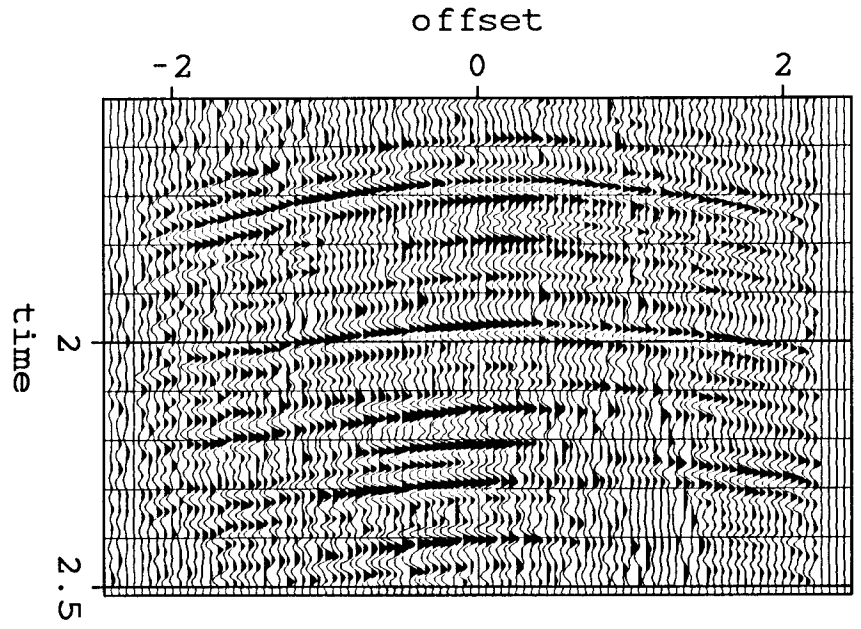


FIG. 21. Subtracting the unwanted events of Figure 20 from the original data of Figure 14 uncovers much of the previously hidden hyperbolic events. The majority of samples remain untouched. The assumptions of the offset-local stack only affect the values interpolated under the extracted noise.

samples. Any bias introduced into the data by the assumptions about the signal in equation (11) can only appear in the interpolated values left after the extraction of the unwanted events.

APPENDIX
THE STATISTICAL TOOLS:
GENERALIZATION TO OTHER TRANSFORMS

We shall now present explicitly the statistical tools necessary for implementing the algorithms outlined in this paper. Other transformations may be easily substituted for that of NMO stacks.

Define the data array as a sum of noise and transformed signal. The samples of $\overline{s}$ and $\overline{n}$ are to represent statistically independent events, as explained in the first section.

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

$$or \quad \overline{d} = \overline{f}(\overline{s}) + \overline{n} \tag{A-1}$$

The transformation $\overline{f}(\cdot)$ may be equations (2), (6), or (11) of this paper, or a non-linear transformation.

The maximum a posteriori (MAP) estimate of $\overline{s}$ maximizes the probability of the data $\overline{d}$:

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

The MAP estimate is a linear function of the data (1) when the *a priori* distributions are Gaussian and (2) when the forward transform is linearized. Equivalently,

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

$$\min_{\Delta\overline{s}} J_2(\Delta\overline{s}) = \sum_i \frac{1}{C_{s_i}^2} (s_i{}^0 + \Delta s_i)^2$$

$$+ \sum_i \frac{1}{C_{n_i}^2} [d_i - f_i(\overline{s}_0) - \sum_j F_{ij}^0 \Delta s_j]^2 \tag{A-3}$$

C's are appropriate standard deviations.

A least-squares perturbation of the model is a linear function of the data residuals.

$$\Delta s_i = d_i' = \sum_j F_{i,j}^{-1} d_j \tag{A-4}$$

So $d_i' = s_i' + n_i'$ where $\overline{s}'$ is a function of $\overline{s}$, and $\overline{n}'$ of $\overline{n}$.

The probability distribution functions convolve:

$$p_{d_i'}(x) = p_{s_i'}(x) \ * \ p_{n_i'}(x) \tag{A-5}$$

To estimate pdf's

(1) Find $p_{d'}(x)$ from a histogram of $\overline{d}'$.

(2) Set $p_{n'}(x)$ equal to the $p_{d'}(x)$ that would result if $p_n(x) = p_d(x)$ without coherence.

To find $p_{s'}(x)$ minimize

$$\min J_3[p_{s'}(x)] = \int p_{d'}(x) \ln[p_{d'}(x)/p_{s'}(x) \ * \ p_{n'}(x)]dx \tag{A-6}$$

Include constraints of positivity and unit area.

A Bayesian estimate is defined as the expected value of signal when the sum of signal and noise is known.

$$\hat{s}' = E(s' \mid d') = \int x \ p_{s' \mid d'}(x \mid d')dx$$

$$= \frac{\int x \ p_{s'}(x) \ p_{n'}(d'-x) \ dx}{p_{d'}(d')} \tag{A-7}$$

Define reliability as the probability that the estimated signal is within a fraction c of the actual value.

$$reliability = p[-c\hat{s}' \leq s' - \hat{s}' \leq c\hat{s}' \mid d']$$

$$= \frac{\int_{-c\hat{s}'}^{c\hat{s}'} p_{s'}(\hat{s}' - x)p_{n'}(x)dx}{\int_{-\infty}^{\infty} p_{s'}(\hat{s}' - x)p_{n'}(x)dx} \tag{A-8}$$

Accept then those events (sample perturbations) that have a sufficiently high reliability.