CWP-195
December 1995

A Comparison of Moveout-Based Approaches to Suppression of Ground-roll and Multiples

Gabriel Alvarez De La Hoz

— Master's Thesis —
Department of Geophysics

Center for Wave Phenomena
Colorado School of Mines
Golden, Colorado 80401
303/273-3557

ABSTRACT

One of the most important tasks of seismic data processing is the successful identification and suppression of seismic energy that does not correspond to primary body waves. These include such energy as surface waves, which do not penetrate deep into the earth and so do not carry useful information about the deeper subsurface, and multiples, seismic energy that has bounced more than once from acoustic subsurface interfaces before its recording on the surface of the earth.

Traditional approaches for ground-roll suppression, such as low-cut frequency filtering, F-K filtering, and τ-p filtering, are reviewed and compared in this study. τ-p filtering is shown to be the most effective of these three methods for ground-roll suppression. I also compare the performance of a signal-noise separation algorithm that uses a statistically-derived threshold amplitude that identifies and extracts, in the τ-p domain, those samples more likely to correspond to the signal (Harlan et al., 1984). Harlan's method uses local slant stacks to "focus" (that is, decrease to the minimum the number of parameters required to describe) the desired reflections while defocusing any unwanted noise. In this study, Harlan's method is applied to data sorted into CMP gathers that are NMO-corrected so that linear or parabolic global slant stacks can be used instead of local slant stacks. Under certain conditions, this modified Harlan's approach can provide a refined extraction of the reflections in the presence of the noise; moreover, the modifications allow an order of magnitude increase in computation speed over Harlan's use of local slant stacks.

Since the performance of the τ-p filtering and Harlan's approach depend critically on the ability of the τ-p transform to focus the reflections and to provide an accurate inverse transform, a comparative study is carried out to investigate which implementation is the best for this application. Although simple t-x or F-K implementations of the transform are fast, their focusing power is limited, and artifacts such as edge effects due to the finite extent of the data arise in the inversely transformed data. Beylkin's F-X implementation, is shown to provide better focusing and yields inverse transformed data with edge effects minimized. Beylkin's approach, however, is at least four times more expensive than either the t-x or F-K implementations.

F-K and τ-p filtering (Hampson, 1986) can also be used for suppression of multiples. Harlan's signal-noise separation algorithm, in its original form, cannot be used to suppress multiples because the transform focusing the NMO-corrected reflections will also focus the multiples. Instead, I devise a hybrid approach that combines τ-p filtering (Hampson's method) to suppress the most identifiable multiple energy, with Harlan's signal-noise separation algorithm, to suppress portions of residual multiple energy that overlap the moveout region of the primaries in the τ-p domain. Comparison of the three methods shows that F-K filtering provides only moderate multiple rejection and, worse, the level of multiple rejection is offset-dependent. Hampson's method is more effective for multiple suppression and, if the differential moveout between the primaries and the multiples is enough to map them to separate

regions in the τ-p domain, it provides offset-independent rejection of the multiples. When the separation is imperfect, however, multiple energy remains that contaminates the amplitudes of the extracted primaries at the short offsets. For constant trace-to-trace amplitudes, the hybrid approach is shown to provide an improved multiple rejection, at the expense of being about 50% more computer intensive.

The focus of the thesis is a study of the relative performance of the F-K filtering method, Hampson's method and the hybrid method, in terms of AVO preservation and degree of improvement in primary-to-multiple amplitude ratio after the data are CMP-stacked. When the amplitude of the multiples is high, multiple energy after multiple suppression can be expected to remain in the data. As a result, AVO behavior is distorted because the contribution of the residual multiple energy alters the apparent amplitude of the primaries on the short-offset traces. Tests with modeled data demonstrate that when primary-to-multiple amplitude ratio in the CMP-stacked data is the overriding consideration, and especially if the primary-to-multiple amplitude ratio in the input data is low, the hybrid approach gives the best result. If AVO analysis is expected to be performed with the data, then Hampson's approach is probably better because of its better preservation of the offset dependence of amplitudes. When polarity changes with offset, the performance of all of the algorithms is degraded and perhaps an alternative approach should be sought and comparably tested.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I am very grateful to the national oil company of Colombia Ecopetrol (Empresa Colombiana de Petróleos), and its research and development center ICP (Instituto Colombiano del Petróleo), for their financial support and for the opportunity to carry out this study. I especially acknowledge the decisive support of Dr. Kurt Bayer, head of the exploration and exploitation division of ICP.

I want to give my deepest appreciation to Dr. Ken Larner under whose advising I had the privilege to work during the realization of this study. Dr. Larner's keen observations and always enlightening suggestions and questions were a source of inspiration for me.

I also want to thank Drs. Tom Boyd and John Scales for serving in my committee and for providing useful suggestions. It is my pleasure to acknowledge the great help that I received from Dr. William Harlan to guide me through some of the key computer implementation details of his outstanding algorithm for signal-noise separation.

I want to thank John Stockwell, who initiated me into the mysteries of SU programming. It is a great advantage to be backed up by such a knowledgeable, remarkable individual.

My gratitude also goes to all professors, staff and fellow students in the Center for Wave Phenomena of the Colorado School of Mines for creating an environment that fosters and encourages the type of basic and applied research that our industry requires. Working with such a remarkable group of people is in itself a reward, and for that I am very grateful to Dr. Norman Bleistein, CWP's director, who allowed me to be a part of it.

Last, but not least, I want to thank my family for their support, especially my mother who taught me that I could always reach my goals if only I worked hard enough for them. Thanks to my wife Nancy and my son Cesar Augusto whose love and quiet confidence in me made it all the more easy.

Chapter 1

INTRODUCTION

The interpreter of seismic data is primarily concerned with the recognition and mapping of seismic energy that has been reflected only once by an acoustic or elastic interface back to the surface. This energy, however, is not the only type present in recorded short records; other types of energy may also be present, such as surface waves, which do not penetrate deep into the earth and thus carry no useful information of the deeper subsurface, and multiples, waves that have been multiply reflected before they are recorded at the surface of the earth. These two types of waves represent coherent noise since they have distinct, predictable moveout and are typically strong events that obscure recognition of the sometimes weak genuine reflections lying underneath (see Figure 1.1).

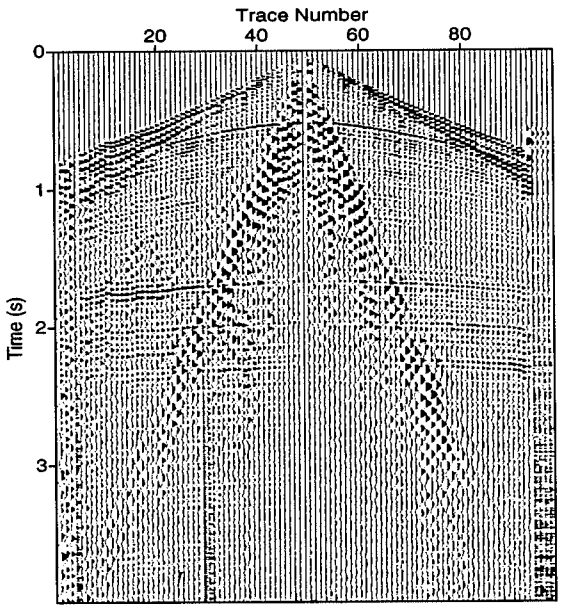

FIG. 1.1. Land shot record showing strong, aliased, dispersive ground-roll. Note that where the noise is present, it obscures the reflections underneath.

1.1 Ground-roll suppression

One of the most important tasks of seismic data processing is the successful elimination of coherent noise, without harming the information contained in the reflections that we want to uncover. Surface waves, loosely called ground-roll, are a serious problem in most datasets, particularly in land data, because they not only obscure the recognition of useful reflections, but also conspire against the success of other important processing steps such as deconvolution.

Traditional approaches to ground-roll elimination include low-cut frequency filtering, which relies on ground-roll's lower frequency content and moveout filtering, which exploits the characteristic linear moveout of ground-roll for its discrimination and suppression.

1.1.1 Low-cut frequency filtering

In traditional exploration seismology, the frequency content of the ground-roll usually ranges from the lowest frequency recorded, up to about 20 Hz, whereas the frequency content of the signal starts at about 5-10 Hz and can go up to perhaps half the Nyquist frequency (see Figure 1.2). It seems natural, therefore, to apply a low-cut frequency filter to the data in

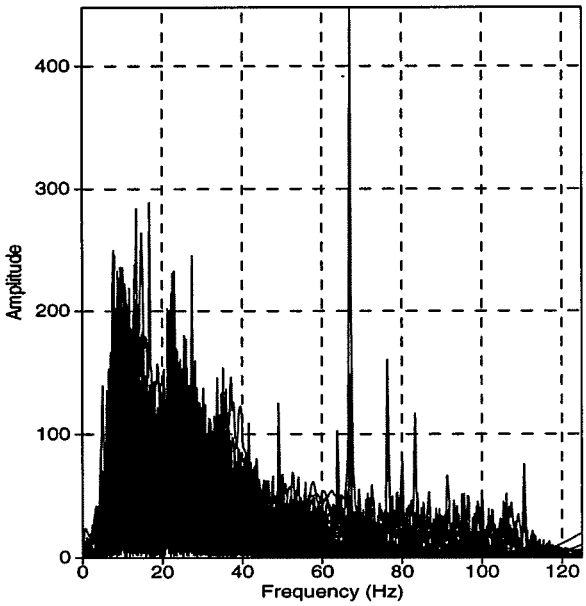

FIG. 1.2. Amplitude spectrum of all traces shot record in Figure 1. Note the high amplitudes of ground-roll for frequencies up to about 16 Hz.

order to remove the ground-roll. Although this approach does indeed suppress the ground-roll (as will be shown in Chapter 4), it also can suppress overlapping low-frequency components

of the signal that are critical for vertical and horizontal resolution and for lithologic inversion. These low frequencies carry information about the general increase of velocity with depth due to compaction and related processes. Low-cut frequency filtering, therefore, is a good choice for suppressing only those frequencies in the ground-roll that are clearly below the lowest useful signal frequencies. Thus, at best, its action at enhancing the signal to ground-roll amplitude is incomplete.

1.1.2 Moveout filtering

In the frequency-wavenumber (F-K) domain, we can in principle separate the noise from the signal based on their distinctly differing moveouts, because the moveout of the noise is linear with large slope (low velocity), while the moveout of the signal is hyperbolic with relatively smaller slopes (high apparent velocities), as seen in Figure 1.3, computed using all the traces and all the time samples in the shot record in Figure 1.1. Note in this figure that the ground-roll has been mapped to fairly narrow linear regions up to about 16 Hz, whereas the signal has been mapped close to the vertical axis, (that is, high apparent velocities as mentioned before).

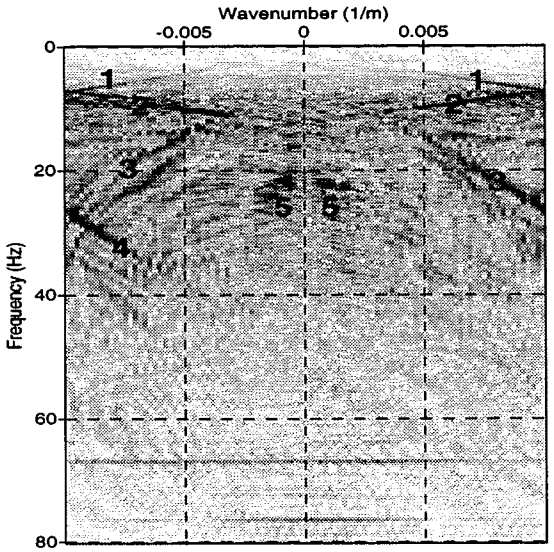

FIG. 1.3. F-K amplitude spectrum of shot record in Figure 1. 1: ground-roll, 2: spatially aliased ground-roll, 3: first arrivals, 4: spatially aliased first arrivals, 5: reflections.

The separation, between the signal and the noise, however, is possible only if the noise is not aliased, so that the noise and the signal occupy different regions in that domain, and

if we are ready to accept the price of sacrificing features with large moveout such as wide-angle signal, diffraction tails, and fault-plane reflections. In Chapter 4, I present results of applying this method. Even if no aliasing is present, however, moveout filtering introduces undesired artifacts such as spatial correlation of background noise, loss of fault clarity, and a Gibb's phenomenon associated with the cutoff moveouts of the filter. Typically, because of the fan shape of the filter, the background noise becomes organized so that it is high in wavenumber and low in frequency (Kirchheimer, 1985). These problems combine to produce seismic sections of "mixed" or "wormy" appearance, which can be potentially misleading for the seismic interpreter.

We see, then, that traditional approaches to ground-roll elimination, despite their conceptual simplicity, are seldom completely effective in removing the noise in all but the most favorable situations. A number of techniques, based on refinements of the ones described above or on the identification and exploitation of additional distinguishing attributes for signal and noise discrimination, have been developed in the past decade.

1.1.3 τ-p filtering

In Chapter 2, I present a detailed overview and comparison of the main ideas and relevant computer implementation details of the most commonly used algorithms to compute linear and parabolic τ-p transforms (i.e., slant stack) and more generally the discrete Radon transform. This comparison is intended to provide the reader with practical criteria to help decide on which implementation to choose for a given application, on the basis of the transform's focusing power, the presence or absence of edge effects, the ability of the inverse transform to recover the input data in terms of phases and amplitudes, and computational cost.

The τ-p transform can be used to suppress ground-roll in the following way: the data are sorted into CMP gathers and NMO-corrected, such that the reflection hyperbolas are flattened; a linear or parabolic τ-p transform is then applied to focus them, concentrating their energy to small regions of the τ-p plane (Hampson, 1986). The range of slopes for the τ-p transform is chosen so that the ground-roll does not show any discernible coherence within this range of slopes, and so it is not modeled by the transformation. An inverse τ-p transform then maps the modeled data (the NMO-corrected reflections) back to the time-offset domain without the ground-roll. This method, therefore, exploits the differential moveout between the NMO-corrected reflections and the ground-roll. Results of applying this method to suppress ground-roll in a field shot record will be shown in Chapter 4.

1.1.4 Signal-noise separation algorithm

A method, based on a statistical approach to signal-noise separation developed by Harlan et al. (1984) is emphasized in this study. This method allows the separation of events considered to be signal from those considered to be noise, on the basis of their difference in moveout pattern. Although the method can be used in a variety of applications, it has not had wide-spread use in seismic data processing partly because some of its practical computer implementation details have not been well understood. In Chapter 3, I present an overview

of the method, along with relevant computer implementation details. This is expected to provide the reader with a relatively easy way to implement the algorithm, if desired.

For ground-roll suppression, this method, like the previous one, uses the τ-p transform to focus the desired reflections (that is, to concentrate their energy to a small region of the τ-p plane). As before, the ground-roll energy is not modeled by the transformation because its slopes are not included in the range of slopes used to compute the τ-p transform. The presence of the ground-roll, however, can create spurious local alignments (particularly if the data are spatially aliased) that the transform models as background noise. The method then uses a statistically developed amplitude threshold to separate the samples focused by the transformation from those not focused by it, thereby allowing the extraction of the reflections from the background noise. Details on how this amplitude threshold is computed and used are given in Chapter 3.

Harlan's original method for ground-roll suppression approximates the reflection hyperbolas by small segments of linear events that are then focused with the use of local slant stacks. That method is computationally intensive. To reduce cost, I have modified Harlan's method by sorting the data into common midpoint (CMP) gathers and applying normal-moveout (NMO) correction so that the reflection hyperbolas will be flat. After the NMO correction, a simple linear or parabolic Radon transform can be used to focus the reflection events with residual moveout, rather than the slower local slant stacks that have less focusing power and are more likely to focus spurious alignments. With this modification, the algorithm runs about ten times faster and the results are better, for ground-roll suppression, as is shown in Chapter 4.

For ground-roll suppression, both the direct τ-p filtering and the signal-noise separation algorithm rely on the large moveout difference between the reflections and the ground-roll, but the latter adds the possibility of further discrimination between the reflections and background noise in the τ-p domain, on the basis of focused amplitude, perhaps producing a cleaner extraction of the reflections. Whether this additional discrimination power is significant or not, compared to the result of applying either of the previous two methods, is studied in Chapter 4, where the result of applying these methods to a field shot record is examined.

The results presented in Chapter 4 are qualitatively graded and compared, on the basis of quality of signal preservation, level of noise rejection and computational cost. This comparison should help the reader decide which method to use for ground-roll suppression, according to which of the above criteria is the priority.

1.2 Multiple suppression: traditional approaches

Multiply reflected energy can be a serious noise problem, particularly in marine data, because sea-floor reflections tend to be very strong (high acoustic contrast between the water layer and the bedrock) so the water layer acts as a wave guide to the seismic reflections. Strong multiples can also arise when a high-velocity layer is sandwiched between layers of lower velocity above and below. The energy from multiples can have a detrimental impact on the ability of the interpreter to identify genuine reflections because both have hyperbolic moveout. It is, therefore, possible to misinterpret multiples as genuine reflections.

The suppression of multiples depends on their particular characteristics; the methods used most are: stacking, deconvolution, and velocity filtering in both the F-K and τ-p domains.

1.2.1 Deconvolution

The attenuation of short-period multiples is achieved with the use of predictive deconvolution, in which the periodicity of the multiples can be used to design an operator that will identify and extract the predictable part of the wavelet (multiples), leaving only the non-predictable part of it (signal), assuming that genuine reflections come from an earth reflectivity series that can be considered random. The success of predictive deconvolution strongly depends on the ability of the process to identify any periodicity in the wavelet and on this periodicity being associated exclusively with the presence of multiples (Yilmaz, 1987). In general, for other than short-period multiples, only moderate success can be achieved with this procedure.

In principle, deterministic deconvolution can be applied to remove water-bottom reverberations in situations in which the exact depth and speed of sound of the water layer are known. Since these conditions are rarely met, deterministic deconvolution is not widely used, despite the elegance of its closed, exact mathematical formulation.

1.2.2 Moveout filtering

As is the case for ground-roll suppression, the performance of an F-K filter in suppressing multiples strongly depends on the transform's ability to map genuine reflections and multiples to separate regions of the F-K plane. This is in general the case for events on far-offset traces, for which the difference in moveout can be large, but not for those on short-offset traces for which the difference in moveout is small. The performance of applying F-K filtering, therefore, is not uniform for all offsets but is rather good for far offsets and poor for small offsets. This shortcoming, together with the problems mentioned above in relation to ground-roll suppression, usually make F-K filtering an undesirable option for multiple elimination.

1.2.3 CMP stacking

CMP stacking is done routinely to take advantage of the redundancy in normal seismic data acquisition. Stacking is itself a powerful tool for the suppression of both non-coherent events and coherent events that depart significantly from horizontal after NMO correction. Since this is the case for multiple energy, the stack itself can prove powerful for multiple suppression. The performance of this process, however, is limited and some residual energy survives the stack. Moreover, it is desirable to suppress multiples on unstacked data to enhance the chance of success for data-dependent processes such as velocity analysis and statics estimation. Likewise, interpretation of unstacked data, such as in amplitude-versus-offset (AVO) analysis, benefits from suppression of multiples prior to CMP stacking.

1.2.4 Hampson's parabolic Radon transform filtering

Similar to the τ-p approach to suppression of ground-roll, we can NMO correct the data in CMP gathers and then apply a parabolic Radon transform to map the flattened reflection hyperbolas and the undercorrected multiples to different concentrated regions of the τ-p plane (Hampson, 1986). This separation allows the application of a mute (of selected slopes p) to remove the multiple energy. Provided that the separation between the primary and multiple energy regions is large enough, accurate extraction of the primary energy for all offsets can be achieved. Again, a separate chapter, Chapter 5, is devoted to the detailed description of this method.

1.2.5 Signal-noise separation

Harlan's approach, in its original form, cannot be used to separate primaries and multiples in the τ-p domain because they are both focused by the transformation. In Chapter 5, however, I show that a hybrid approach can be devised in which the most identifiable multiple energy is first removed simply by applying a τ-p mute, in the same way as with Hampson's approach. The signal-noise separation algorithm is then used to discriminate, on the basis of a statistically derived amplitude threshold, between the focused primaries and any residual multiple energy that may have "leaked" into the primary energy zone. The expectation is that this approach will yield a refined multiple suppression. How significant is this refinement will be discussed in Chapter 5.

In Chapter 6, I demonstrate, through comparison processing of four model datasets, the important issues of how the different methods for multiple suppression influence the quality of the stack and the amplitude variation with offset of the primaries. The results of the tests presented in Chapters 5 and 6 are aimed at learning which method best suppresses multiples in stacked data, which method causes least distortion of amplitude variation with offset of the primaries, and what are the relative computational cost of the various methods.

Chapter 2

THE GENERALIZED DISCRETE RADON TRANSFORM

2.1 Introduction

Since the discrete Radon transform is a critical tool for the implementation of the algorithm proposed in this study, in this chapter I present an overview of its most important features, along with some details on computer implementation.

2.2 Theoretical overview

2.2.1 Continuous Radon transform

The classical Radon transform of a function z of two variables (x, t) where, for most applications, x represents distance and t time, is defined as (Claerbout, 1985; Foster and Mosher, 1992)

$$y(p, \tau) = \int_{x_{min}}^{x_{max}} z(x, t = \tau + px)dx, \tag{2.1}$$

where x_{min} and x_{max} are the minimum and maximum offsets in the data. This equation can be interpreted mathematically as the linear transformation that assigns, to every point in the τ-p domain, the sum of the energy in $z(x, t)$ along a straight line of slope p and intercept time at $x = 0$, τ, with the summation carried out over all possible values of x (that is, for all traces). In the continuous, infinite case, this equation is the so-called τ-p transform. From a geophysical point of view, the equation can be interpreted as a decomposition of the reflected seismic wavefield $z(x, t)$ into plane-wave components with propagation directions related to the slope direction p.

The inverse of this transformation, which maps the energy from the τ-p domain back to the t-x domain is given by (Claerbout, 1985)

$$z(x, t) = \rho(t) * \int_{p_{min}}^{p_{max}} y(p, \tau = t - px)dp, \tag{2.2}$$

where p_{min} and p_{max} are the minimum and maximum slopes of the data in the τ-p domain. This inverse transformation is basically another τ-p transform with the sign of p changed and followed by a convolution with the so-called ρ (rho) filter, equivalent to multiplying every frequency component in the data by the absolute value of its frequency, thus correcting for high-frequency losses when the transform was performed. The expressions for this ρ filter in time and frequency domain are (Claerbout, 1985; Tatham, 1984)

$$\rho(t) = \frac{1}{t^2}. \tag{2.3}$$

$$\rho(\omega) = (-i\omega)isgn(\omega) = |\omega|. \tag{2.4}$$

Equation (2.1) can be generalized such that the data are summed along any defined curve $g(x)$. In this case, the continuous generalized Radon transform is given by

$$y(p,\tau) = \int_{x_{min}}^{x_{max}} z(x, t = \tau + pg(x))dx. \tag{2.5}$$

Equation (2.1) is a special case (with $g(x) = x$) of equation (2.5), which can be interpreted as a family of transforms.

2.2.2 Discrete Generalized Radon Transform

The seismic data are recorded digitally in time and at discrete positions in space, so they are continuous in neither time nor space. We can write the discrete counterparts of the equations given in the previous section just by changing integrals to summations carried out over all possible values of the integration variables x and p. The equations become

$$y(p_i,\tau) = \sum_{k=0}^{Nx-1} z(x_k, t = \tau + p_i g(x_k))\Delta x \tag{2.6}$$

and

$$z(x_k,t) = \rho(t) * \sum_{i=0}^{Np-1} y(p_i, t = \tau - p_i g(x_k))\Delta p. \tag{2.7}$$

Here, N_x and N_p are the number of traces and slopes, respectively.

2.3 Computer Implementation of Radon Transform

Several different computer implementations of the discrete Radon transform have been devised. In describing them, I demonstrate their comparative actions on model dataset 1, shown in Figure 2.1. This dataset consists of 128 traces with 1000 samples each, at 4-ms sample interval and 50-m trace interval.

2.3.1 Time-Domain Implementation

The simplest way to implement the Radon transform is the direct use of its definition for discrete data, given by equation (2.6). The procedure is straightforward: for each point in the τ-p domain, say τ_0 and p_0, we sum the energy in the original data along the straight line of slope p_0 (or along the curve associated with the parameter p_0 in the case of the generalized Radon Transform) and zero intercept time τ_0. This procedure is repeated for all desired values of τ and p. Although doing this is simple enough, a couple of problems may arise. First, since the data in t-x space exist only in a finite grid of points, it is highly unlikely that we will hit one of those grid points when doing the summation, so some sort of interpolation between grid points is required; frequently, a simple linear interpolation is adequate since the time sample interval is usually sufficiently small. The second problem is space and slope

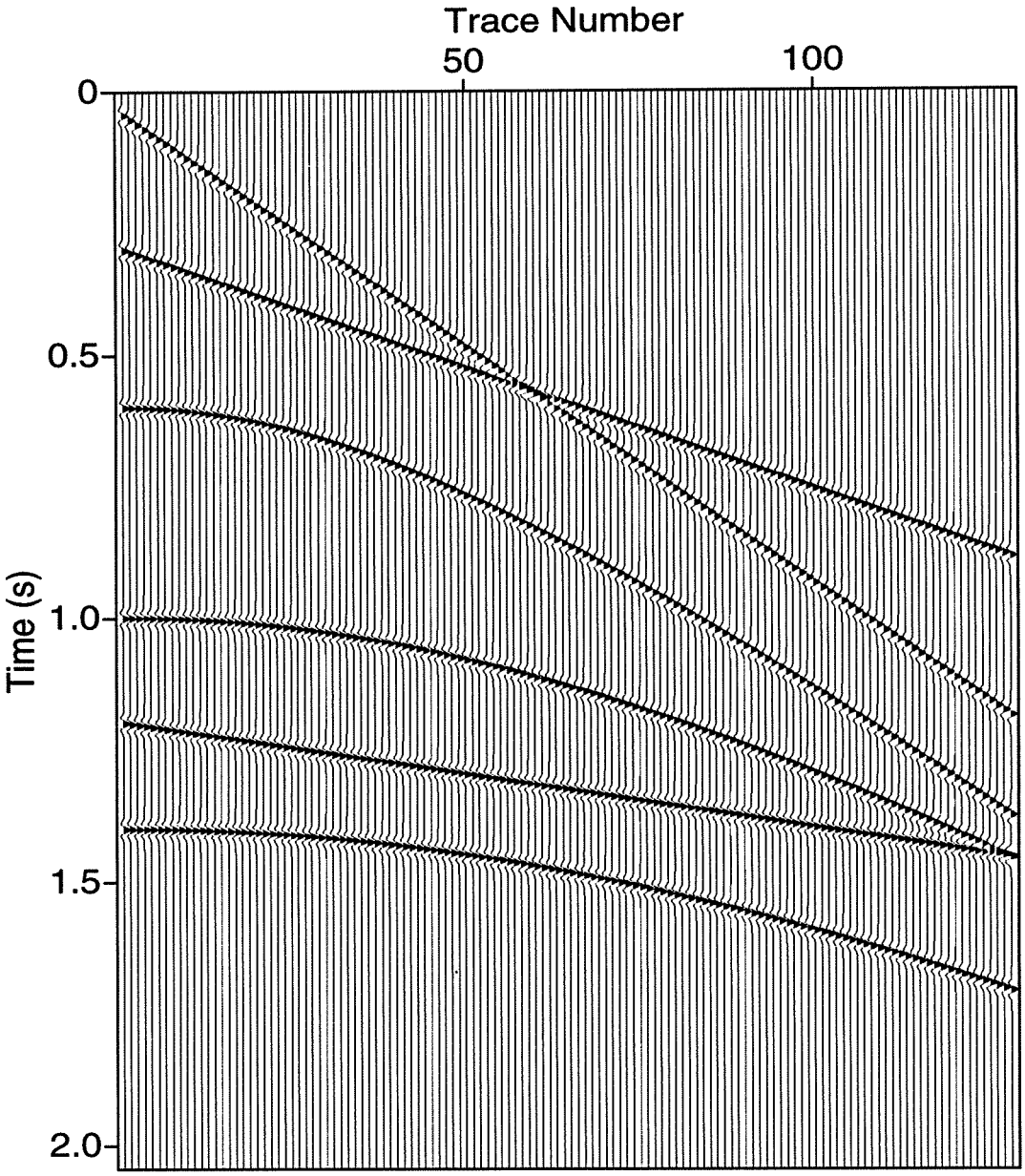

FIG. 2.1. Modeled test dataset 1 containing events with hyperbolic and linear moveout. Each reflector is a characterized by a Ricker wavelet with a dominant frequency of 37 Hz. The amplitudes of every reflection on every trace are the same.

aliasing, which forces us to use small enough sampling intervals in space and slope to avoid distorting the energy at high slopes and large offsets in pretty much the same way that a fine enough sampling in time is required to avoid distortions of the highest time frequencies. Since this is a problem for any implementation, it will be discussed further in a later section of this chapter.

The time-domain computer implementation can then be carried out simply as a cascade of three loops, the outermost one over all slopes of interest, the middle one over all possible time intercepts (τ values) and the innermost to carry out the summation over x, taking into account the required time interpolation. The result of applying this time domain τ-p algorithm to our modeleded data is shown in Figure 2.2. Straight lines have been mapped to small regions in the τ-p plane roughly centered around their true slope and intercept values, while the hyperbolas have been mapped to ellipses, as they should. The mapping has not transformed linear events to points because the linear events in the data do not have infinite extent. The inverse transform is performed in the same way, except that the sign of p is changed in the summation and the roles of p and x are exchanged (that is, the same subroutine used to compute the forward transform is used to compute the inverse by just calling it with negative values of p) and the result is convolved with the rho filter to correct for losses in high-frequency energy that arose during the forward τ-p transform.

The time-domain inverse τ-p transform can be expected to recover the shape and position of the events in the input data but not amplitudes because a perfect inverse discrete Radon transform is possible only if the input data can be considered periodic (Beylkin, 1987). Some form of scaling is required, therefore, to bring the inverted data back to the amplitude level of the input data. Figure 2.3 shows the result of computing the inverse transform of the data in Figure 2.1, with a least-squares amplification applied. The inverse transform has been successful in recovering the shape and locations of the events as is emphasized in Figure 2.4, which shows the difference between the input and the amplified τ-p inverted data. The larger differences for the largest and smallest offsets result from edge effects due to the fact that we are approximating data in a finite number of traces with an infinite-length transform. Figure 2.5 shows a flow diagram of the τ-p implementation in this domain.

2.3.2 F-K Implementation

A faster algorithm can be implemented in the F-K domain in the following way (Wade and Gardner, 1988; Claerbout, 1985).

The data are first transformed in time and space to the F-K domain via a 2-D Fourier transform. Next, a change of variables from k to p is performed using the relation

$$k = fp, \tag{2.8}$$

so

$$p = k/f. \tag{2.9}$$

This change of variables can be interpreted mathematically as a mapping of radial lines in the F-K domain given by equation (2.8) to hyperbolas in the f-p domain given by equation (2.9)

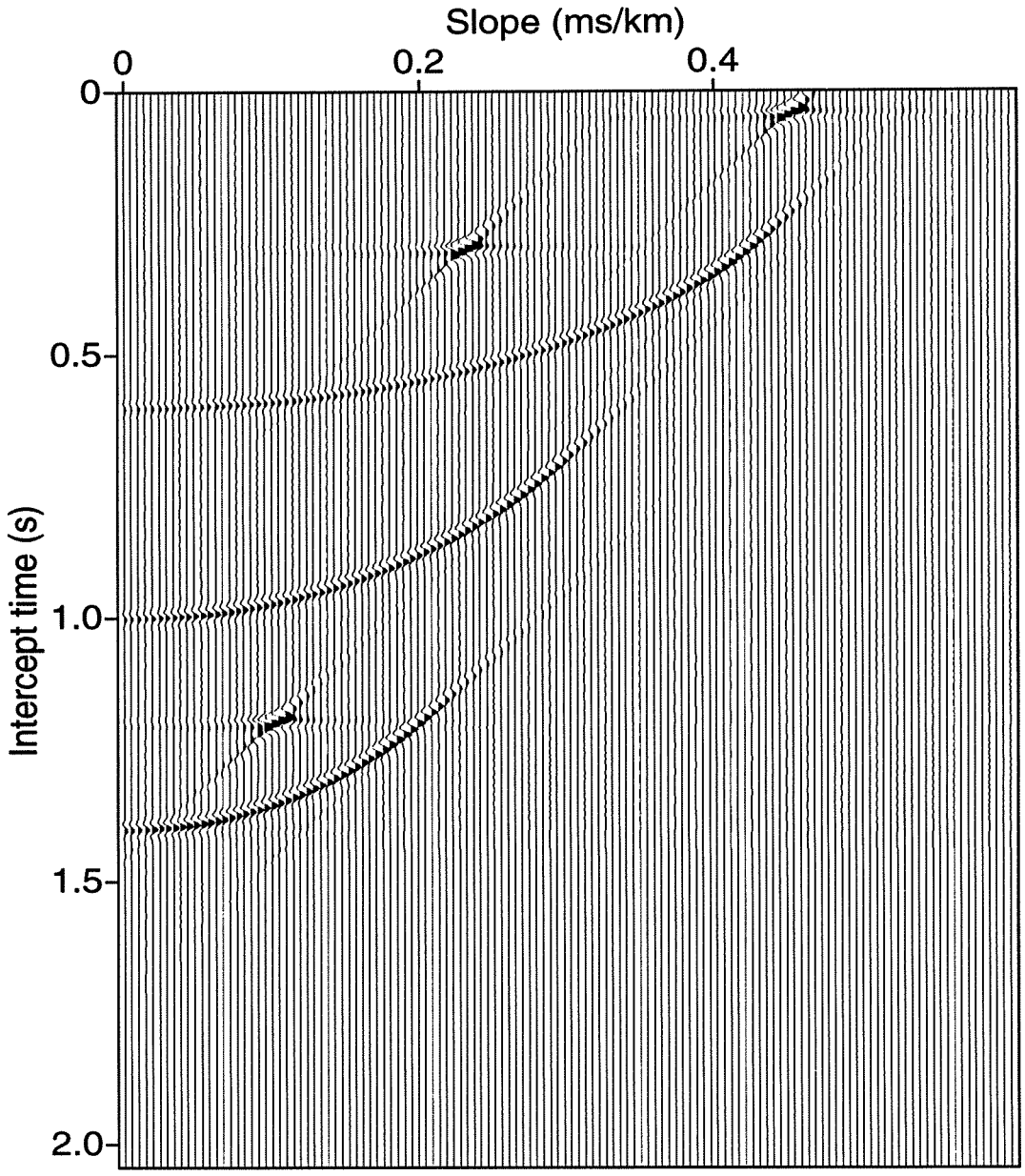

FIG. 2.2. Forward time-domain τ-p transform of the modeled test dataset 1. Note that the three linear events have been mapped to small regions in this domain.

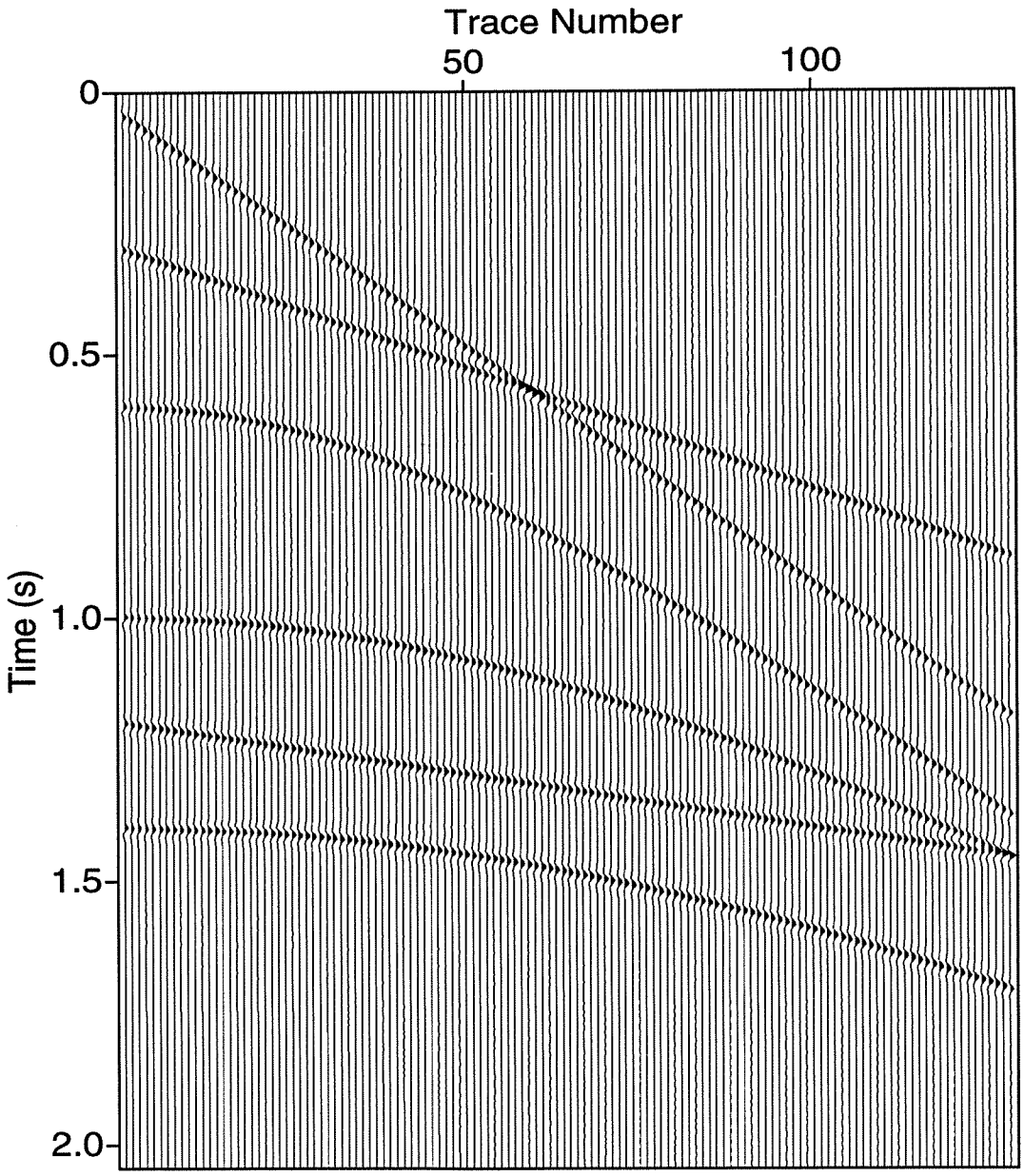

FIG. 2.3. Inverse time-domain τ-p transform of the modeled test data.

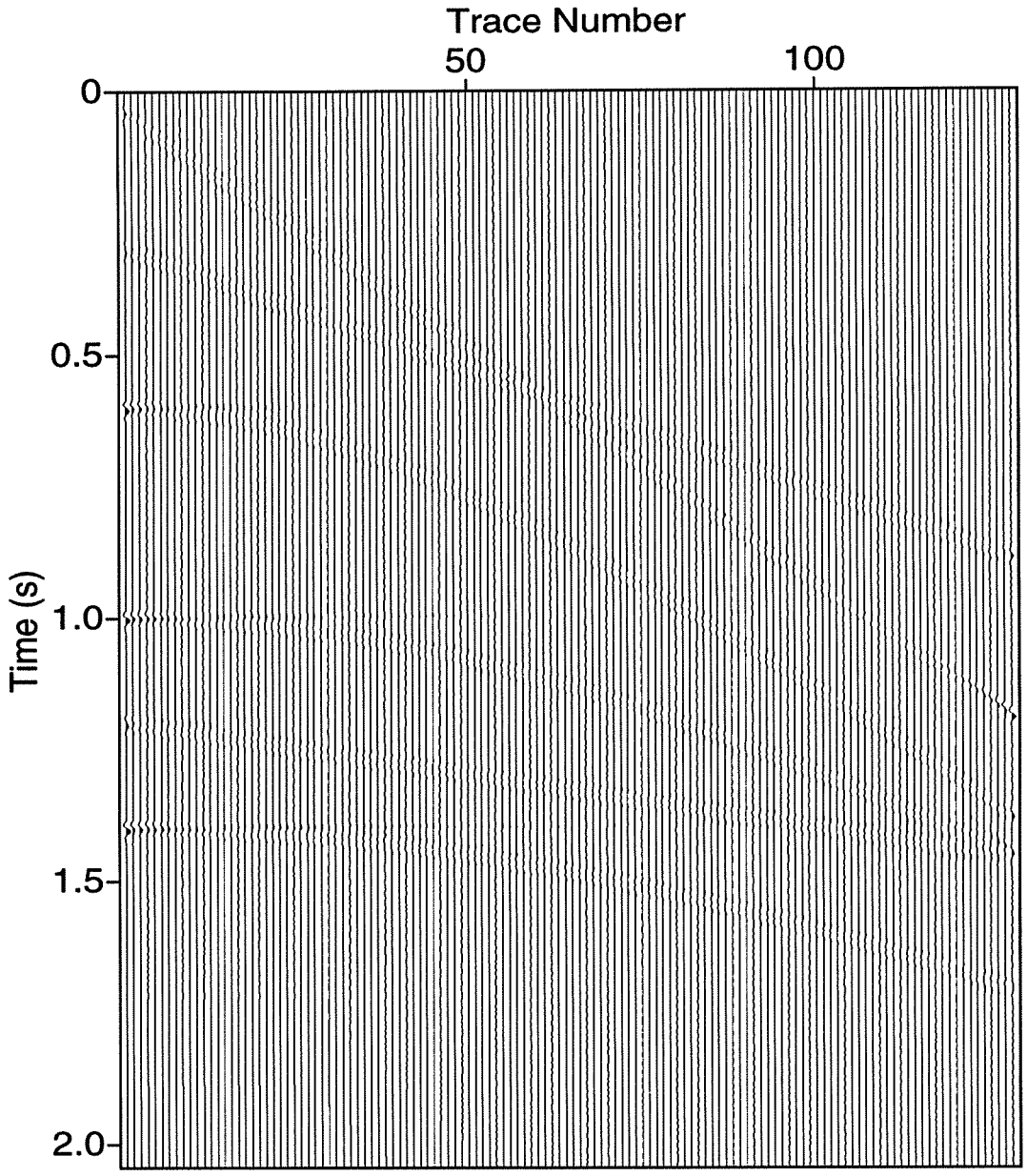

FIG. 2.4. Difference between time-domain inverse τ-p transform and original data. Note the residual energy at the edges due to truncation effects associated with the finiteness of the input data.

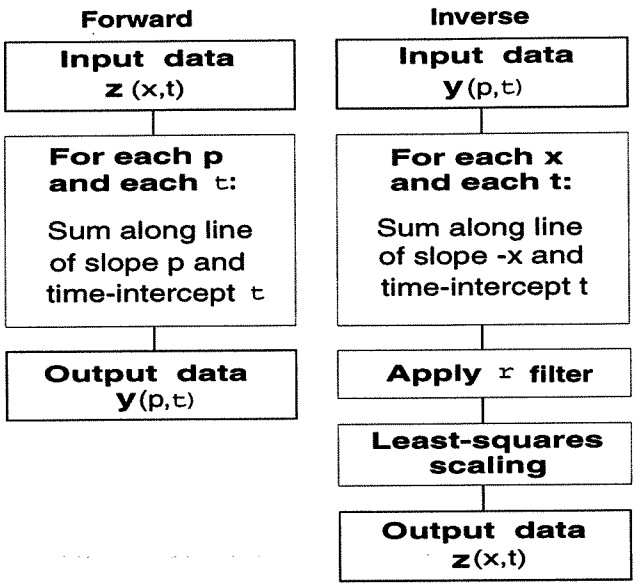

FIG. 2.5. Flow diagram for the time-domain implementation of the forward and inverse τ-p transform.

(Wade and Gardner, 1988). This change of variables can be regarded as a data stretch and is similar to the one applied in Stolt migration (Stolt, 1978).

The problem that arises is that although k and f are regularly sampled, the p values computed by this relation will not be regularly spaced and may not correspond to the range of values that we want for p. The way around this difficulty is to choose the values of p that we desire and then interpolate the corresponding values of k. The interpolation, however, is in the 2-D F-K domain and so is no longer a simple 1-D linear interpolation as it was with the time-domain implementation (Claerbout, 1985). Besides, the interpolation must be carried out with complex numbers. The best interpolation choice is a truncated sinc interpolation (Rosenbaum and Boudreaux, 1981). Wade and Gardner (1988) have given an explicit, efficient formula for that interpolation.

Once the data are in the f-p domain, all that is left is to apply an inverse Fourier transform over frequency to get the data in the desired τ-p domain. Figure 2.6 shows the result of applying the transform with this implementation. The result is almost identical to the one obtained by the time-space domain approach (Figure 2.2).

The inverse transform can be carried out with basically the same algorithm. The data are first Fourier transformed in time from τ to f, an inverse mapping is then performed to go from the hyperbolas in the f-p domain to the radial lines in the F-K domain. With this mapping, the data are mapped from a rectangle in the τ-p domain to a triangle in the F-K domain (see Figure 2.7). Finally, the data are then inversely transformed back to the t-x

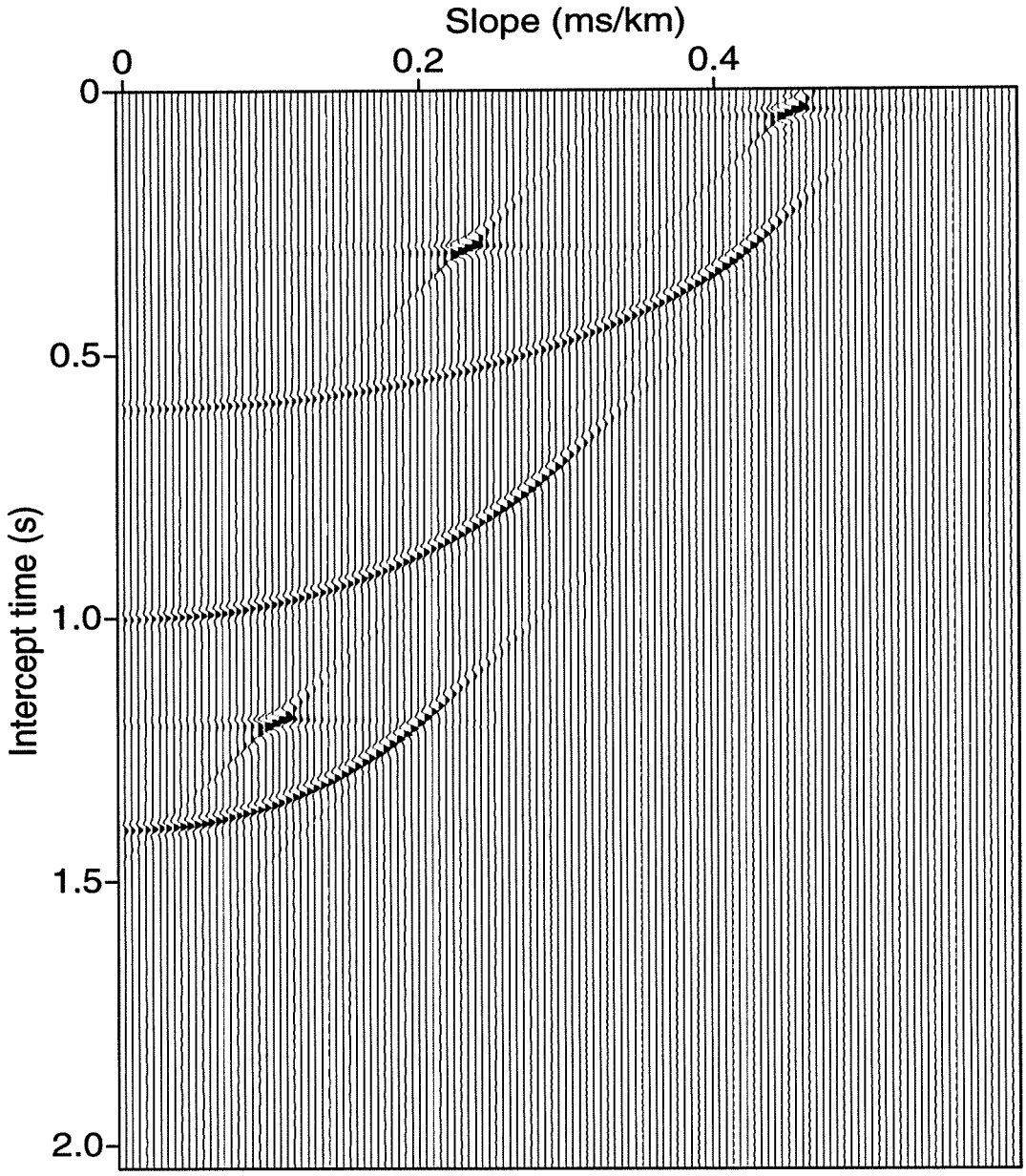

FIG. 2.6. Forward F-K domain τ-p transform of modeled test dataset 1. Again, the three linear events have been mapped to small regions in this domain.

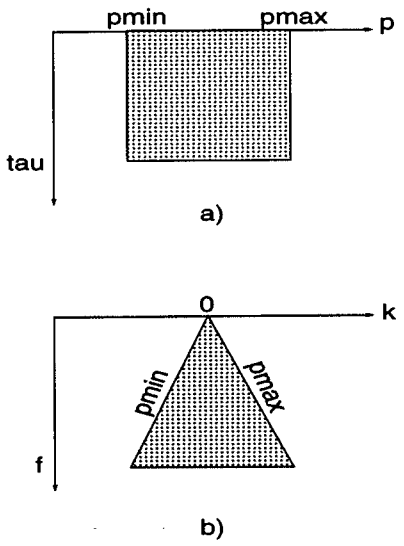

FIG. 2.7. Mapping from the F-K to the F-p plane for the F-K τ-p implementation.

domain with a two-dimensional Fourier transform. Through this stage, the inverse transform and the forward transform are basically the same operation. The data are then least-squares fitted with the input data to obtain a scale factor to apply to the result.

Figure 2.8 shows the result of this inverse F-K transform. As with the result of the t-x implementation, the positioning and shape of the events are good, but the amplitudes for the near and far offsets are lower than those for the intermediate ones. This, again, is due to the finite extent of the data and the inappropriateness of modeling them with a transform that assumes an infinite number of traces. Figure 2.9 shows the difference between this inverse and the input data. As with the t-x domain implementation, the transform has done a better job at recovering the intermediate offsets than at recovering the near and far ones. This edge-effect problem, which is worse for the F-K domain implementation, would be a serious problem for AVO analysis.

Note that this F-K implementation is possible only if the space dimension is uniformly sampled. Figure 2.10 shows a flow diagram for this algorithm.

2.3.3 F-X Implementation

In order to overcome the problems mentioned above with regard to the amplitude level of the inverse τ-p transform when applied in either t-x or F-K domain, and in particular the fact that the inverse transform is not capable of completely recovering the near and far offsets, an alternative to the above τ-p transform is required. Beylkin (1987) devised an algorithm for implementing the Radon transform in the frequency-space (F-X) domain that has the

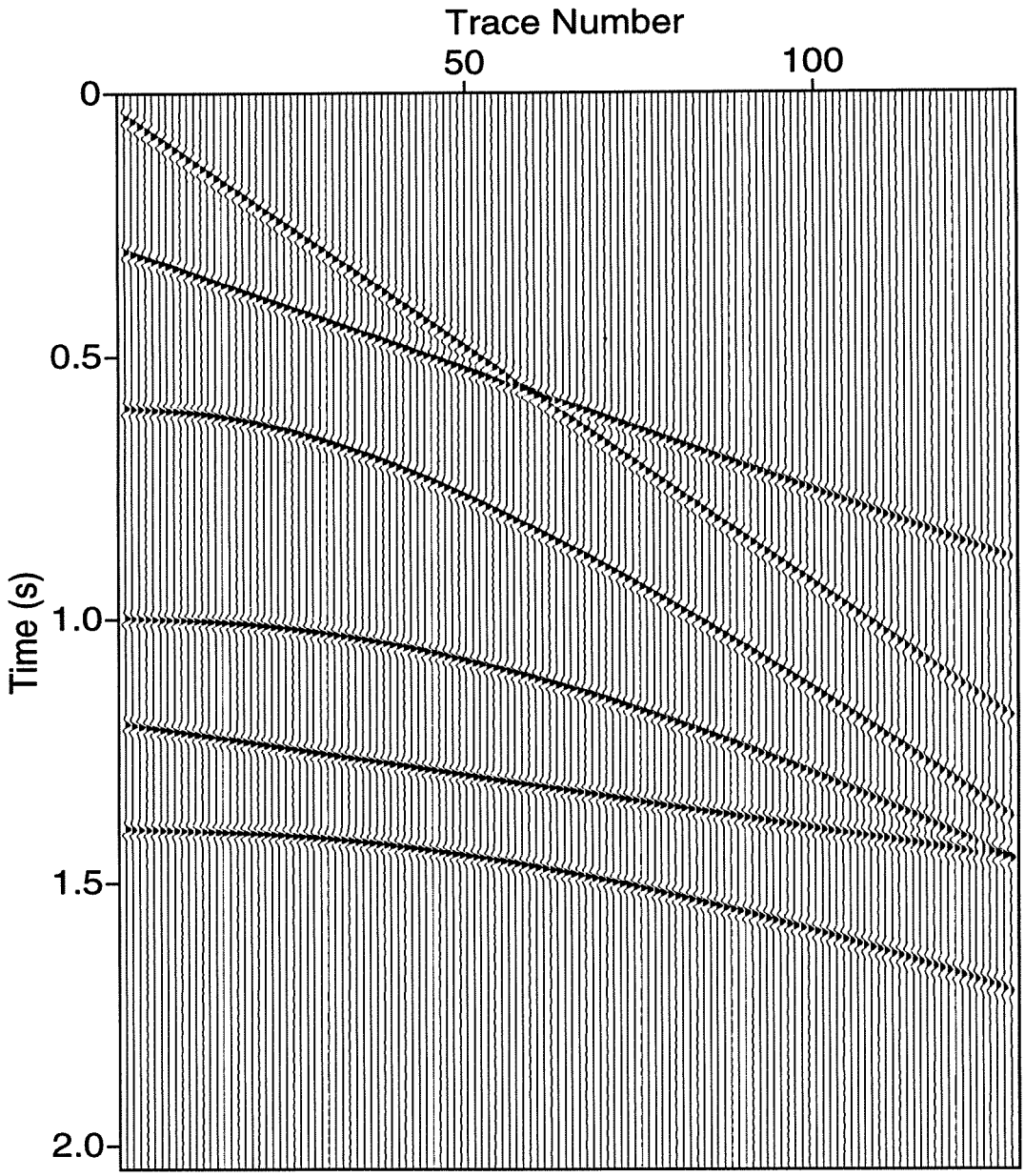

FIG. 2.8. Inverse F-K domain τ-p transform of modeled test data. Note that the amplitudes at the near and far offsets are weaker than those at the intermediate offsets.

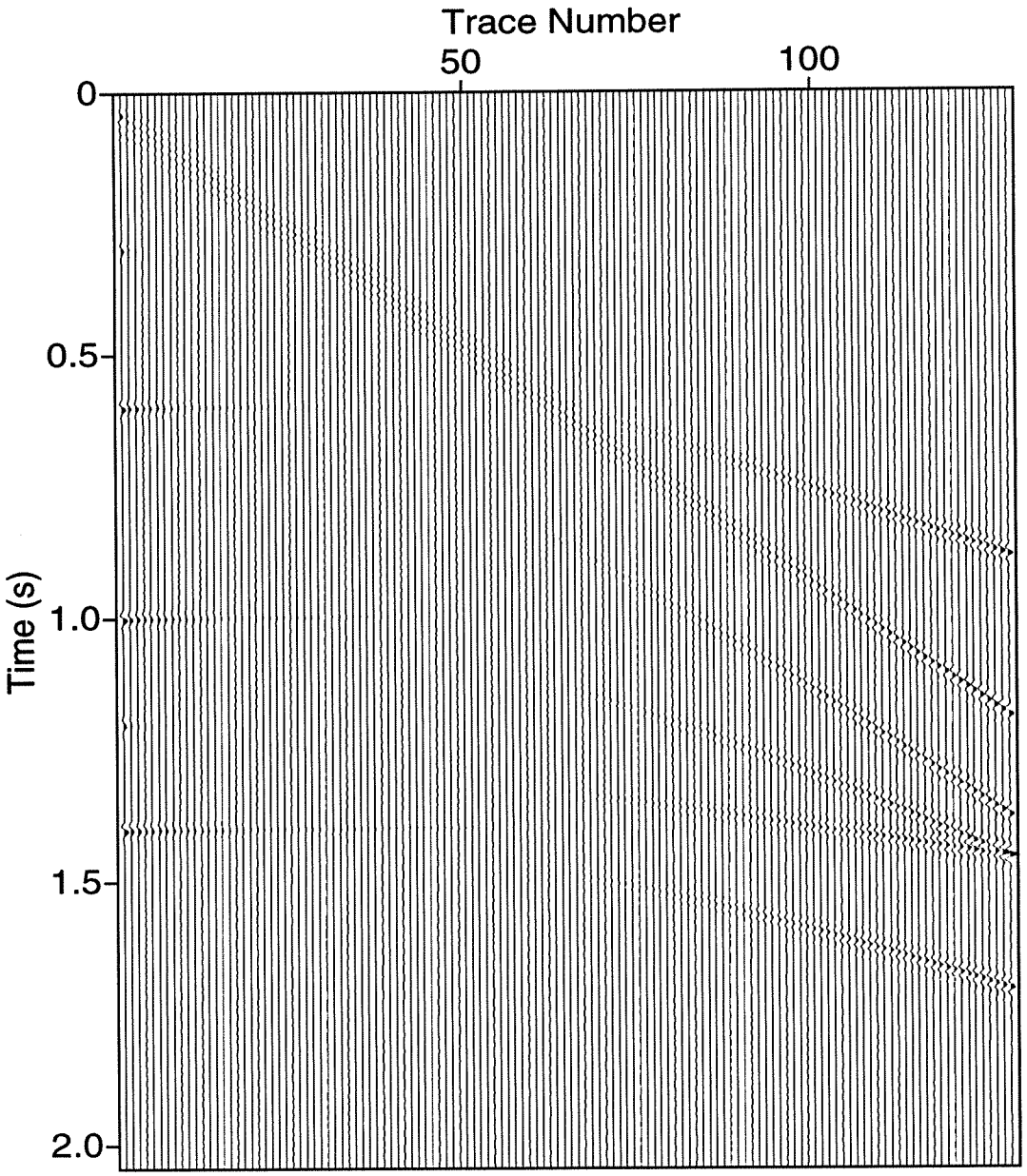

FIG. 2.9. Difference between inverse F-K domain τ-p transform and input data. Note that the near and far offsets were not recovered as well as were the intermediate ones.

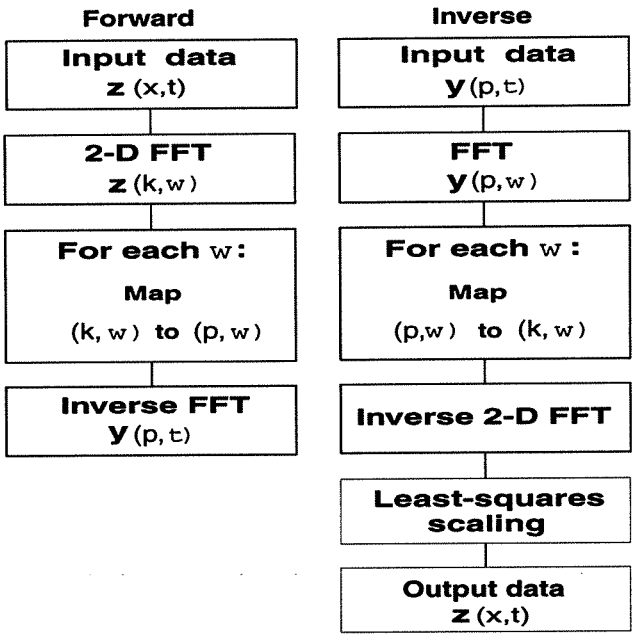

FIG. 2.10. Flow diagram for the F-K implementation of the forward and inverse τ-p transform. The ρ filter is implicitly applied in the frequency domain.

desirable characteristic of modeling the data in terms of linear events that are truncated to match the lateral extent of the window of data being transformed. In Beylkin's method, we start by Fourier transforming with respect to time both sides of equation (2.6) to get

$$y(p,\omega) = \sum_{k=0}^{N_x-1} z(x_k,\omega) e^{i\omega p g(x_k)} \Delta x. \tag{2.10}$$

Here, N_x is the number of traces to be τ-p transformed. This equation illustrates that the summation along curved lines $g(x)$ in the time domain can be performed as a summation of phase shifts over all possible values of x in the frequency domain. This equation can be written in matrix form as

$$\mathbf{y}_\omega(p_j) = \mathbf{R}\mathbf{z}_\omega(g(x_k)), \tag{2.11}$$

where p_j ranges over all desired slopes N_p, k over all traces N_x and R is an N_pxN_x dimensional matrix with elements given by

$$R_{jk} = e^{\omega p_j g(x_k)} \Delta x. \tag{2.12}$$

Note that this matrix is data-independent. The forward Radon transform can thus be computed directly by carrying out the multiplication in equation (2.11), followed by an inverse Fourier transform over time to take the data from the p-ω to the p-τ domain. It is important to note that in order to compute the forward and inverse τ-p transform with this algorithm, we do not require that the space sampling be uniform since no Fourier transform is taken

over the space coordinate. To get the inverse transform, equation (2.11) can be solved, in the least-squares sense, for the ω-x data as (Beylkin, 1987)

$$\mathbf{z}_\omega(g(x_k)) = [\mathbf{R}^\dagger\mathbf{R}]^{-1}\mathbf{R}^\dagger\mathbf{y}_\omega(p_j). \tag{2.13}$$

In this equation, the dagger (†) symbol denotes the adjoint matrix or complex conjugate transpose and, again, an inverse time Fourier transform returns the data to the t-x domain. In this way, we have an exact forward transform to go from the t-x to the τ-p domain and a least-squares inverse to get back to the t-x domain. As mentioned above, an exact inverse is not possible unless the input data are periodic in x (Beylkin, 1987). No interpolation is required, but the solution of equation (2.13) is time consuming if the receiver positions are irregular, which in general they are. The size of the $\mathbf{R}^\dagger\mathbf{R}$ matrix is N_pxN_p, typically of the order of 100x100.

As an alternative way to formulate the problem (Foster and Mosher, 1992), start with the inverse Radon transform, the operation that goes from the model in the τ-p domain to the data in the t-x domain. Then formulate the forward Radon transform from the data to the model space as a least-squares inverse problem. With this perspective, we start with

$$z(x,t) = \sum_{k=0}^{N_p-1} y(p_k, \tau = t - p_k g(x))\Delta p. \tag{2.14}$$

Fourier transformation over time gives,

$$z(x,\omega) = \sum_{k=0}^{N_p-1} y(p_k,\omega)e^{i\omega p_k g(x)}\Delta p. \tag{2.15}$$

Again, for discretely sampled values x_k, this can be written in matrix form as

$$\mathbf{z}_\omega(g(x_k)) = \mathbf{R}^\dagger\mathbf{y}_\omega. \tag{2.16}$$

The inverse least-squares solution for $\mathbf{y}_\omega$ is

$$\mathbf{y}_\omega(p_j) = [\mathbf{R}\mathbf{R}^\dagger]^{-1}\mathbf{R}\mathbf{z}_\omega(g(x_k)). \tag{2.17}$$

In this way, the least-squares inverse generates the forward Radon transform. It seems at this point that we are in exactly the same situation as before, but we are not, because the operator $\mathbf{R}\mathbf{R}^\dagger$ is shift invariant whereas $\mathbf{R}^\dagger\mathbf{R}$ is not. To see this, let us choose the slopes to be uniformly sampled so that

$$p_k = p_{min} + k \cdot \delta p. \tag{2.18}$$

With this restriction, the matrix $\mathbf{C} = \mathbf{R}\mathbf{R}^\dagger$ becomes Hermitian Toeplitz and an efficient algorithm can be employed to solve for the least-squares inverse. To see that this matrix is Hermitian Toeplitz, first note that, for any matrix, the product $\mathbf{R}\mathbf{R}^\dagger$ is Hermitian, i. e.

$[\mathbf{RR}^\dagger]^\dagger = \mathbf{RR}^\dagger$. To see that it is Toeplitz, let us examine its components (Anderson, 1993)

$$C_{jk} \equiv \sum_{l=0}^{N_x-1} R_{lj}^\dagger R_{lk}. \tag{2.19}$$

With p given in equation (2.18), we get

$$C_{jk} = \sum_{l=0}^{N_x-1} e^{-i\omega p_j g(x_l)} e^{i\omega p_k g(x_l)} = \sum_{l=0}^{N_x-1} e^{i\omega(p_k-p_j)g(x_l)}, \tag{2.20}$$

and, since $p_k - p_j = (k-j)\delta p$,

$$C_{jk} = \sum_{l=0}^{Nx-1} e^{i\omega(k-j)\delta p g(k_l)}. \tag{2.21}$$

So, $C_{jk} = C_{k-j}$ making the matrix Toeplitz; i. e. all elements are identical along any given diagonal. This fact allows us to use the efficient complex Levinson recursion algorithm for the inversion; thus, the number of operations required is proportional to n^2 rather than to n^3 for regular matrix inversion. It is important to note that the space variable need not be evenly sampled; only the slopes need be so. The trade of the need for regular trace spacing for the need for regular slope sampling is advantageous since the increment of p is at our choice whereas the increment in x is dictated by limitations in the field. Note also that if we do not insist on having the advantage of a Toeplitz matrix, then we do not require regular sampling in either space or slope, and the method becomes completely general, but more expensive. Another important point to note is that in order to have the advantage of the Toeplitz matrix, we require $N_p \leq N_x$. Otherwise, we will not have a Toeplitz system matrix and so are forced to go back to the general algorithm (Anderson, 1993). Normally, I have chosen N_p equal to N_x.

If $N_p > N_x$, we can no longer compute the inverse using equation (2.17) because the matrix $\mathbf{R}$ is not full rank and so is not invertible (Kostov, 1990). Ideally, in this case the least-squares inverse should be computed as

$$\mathbf{y}_\omega(p_j) = \mathbf{R}^\dagger[\mathbf{RR}^\dagger]^{-1}\mathbf{z}_\omega(g(x_k)). \tag{2.22}$$

but the Levinson recursion is lost since $[RR^\dagger]$ is not Toeplitz as mentioned before.

In practice, however, we can still compute the least-squares inverse using equation (2.17) when $N_p > N_x$ if we prewhiten the inversion step adding a small value to the diagonal elements of the matrix $\mathbf{R}$. The prewhitening avoids the possibility for two or more columns to be identical, thereby making the matrix full rank, with rank equal to N_x (Anderson, 1993; Kostov, 1990). Speed is not the major benefit of defining the forward transform from the τ-p to the t-x domain. More important, we gain in the ability of the transform to focus finite-extent linear events (or, in general, events defined by the offset function $g(x)$) as close as possible, in the least-squares sense, to points. This is an important feature for most applications of the transform for which the discrimination between focused and unfocused

energy becomes critical. We will explore this feature when we review some of the applications in later chapters.

If we defined the Radon transform from t-x to τ-p, then the least-squares-inverse step would be striving to focus a line in the τ-p domain to a point in the t-x domain, which is also a nice feature but less valuable than the previous one for the applications investigated in this study. Also, this focusing ability can be lost or severely reduced if we use too large a value for the prewhitening factor mentioned above. In fact, the prewhitening value should be chosen as the smallest value that ensures stability of the inverse transform. Normally I have used a value between 0.01 and 0.1%.

The result of applying this transform to the modeled test data in Figure 2.1 is shown in Figure 2.11. Compared with the results Figure 2.2 and Figure 2.6, the Beylkin approach (Figure 2.11) shows a sharper focusing of the linear events. The focusing power of the filter term $\mathbf{R}^\dagger\mathbf{R}$ arises because the method describes finite-aperture data with finite-length straight lines rather than with infinite-length ones. We can expect the performance of algorithms that depend on the focusing property of the Radon transform to perform better with this implementation of the transformation.

In summary, the algorithm for the forward transform to go from the data t-x space to the model τ-p space is implemented in the following way: For each trace, apply a forward time FFT to take the data $\mathbf{z}(x,t)$ to the f-x domain $\mathbf{z}_\omega(g(x_k))$. For each frequency, collect the information from each trace and solve the system $\mathbf{Ry} = \mathbf{z}$ as in equation (2.17) for $\mathbf{y}$, using a complex form of Levinson recursion. For each value of p, gather the information from each frequency and apply an inverse FFT to take the data to the τ-p domain. Note that since each frequency is processed independently of the others, the algorithm is well suited for parallelizing.

For the inverse transform, the same is done, exchanging the roles of p and x, except that instead of solving a system of linear equations a simple matrix multiplication is required. The result of applying the inverse transform is shown in Figure 2.12. Again, the events are in the right places, but now the amplitudes are close to correct, as seen in Figure 2.13, which shows that the difference between the input and the inverted data is small. Amplification of this difference by a factor of 10, as shown in Figure 2.14, shows that the waveforms of the inverse τ-p transformed data differ slightly from the input data. Nevertheless, the result for this F-X implementation is far superior to those in Figures 2.4 and 2.9. No special least-squares effort at deriving scale factors is required this time because the scaling is accommodated in the least-squares Beylkin approach. Note also that edge problems no longer exist because, as mentioned before, the data are now represented in terms of finite linear segments. The ρ filter is also implicitly applied in this method.

Figure 2.15 shows a flow diagram illustrating the implementation of Beylkin algorithm.

2.4 Local Slant Stacks

In some applications, it is necessary to focus energy along curves that are not easily expressed as a simple function $g(x)$. In those cases, it is advantageous to approximate such a curve locally as a series of segments of simple shape, the simplest being a series of

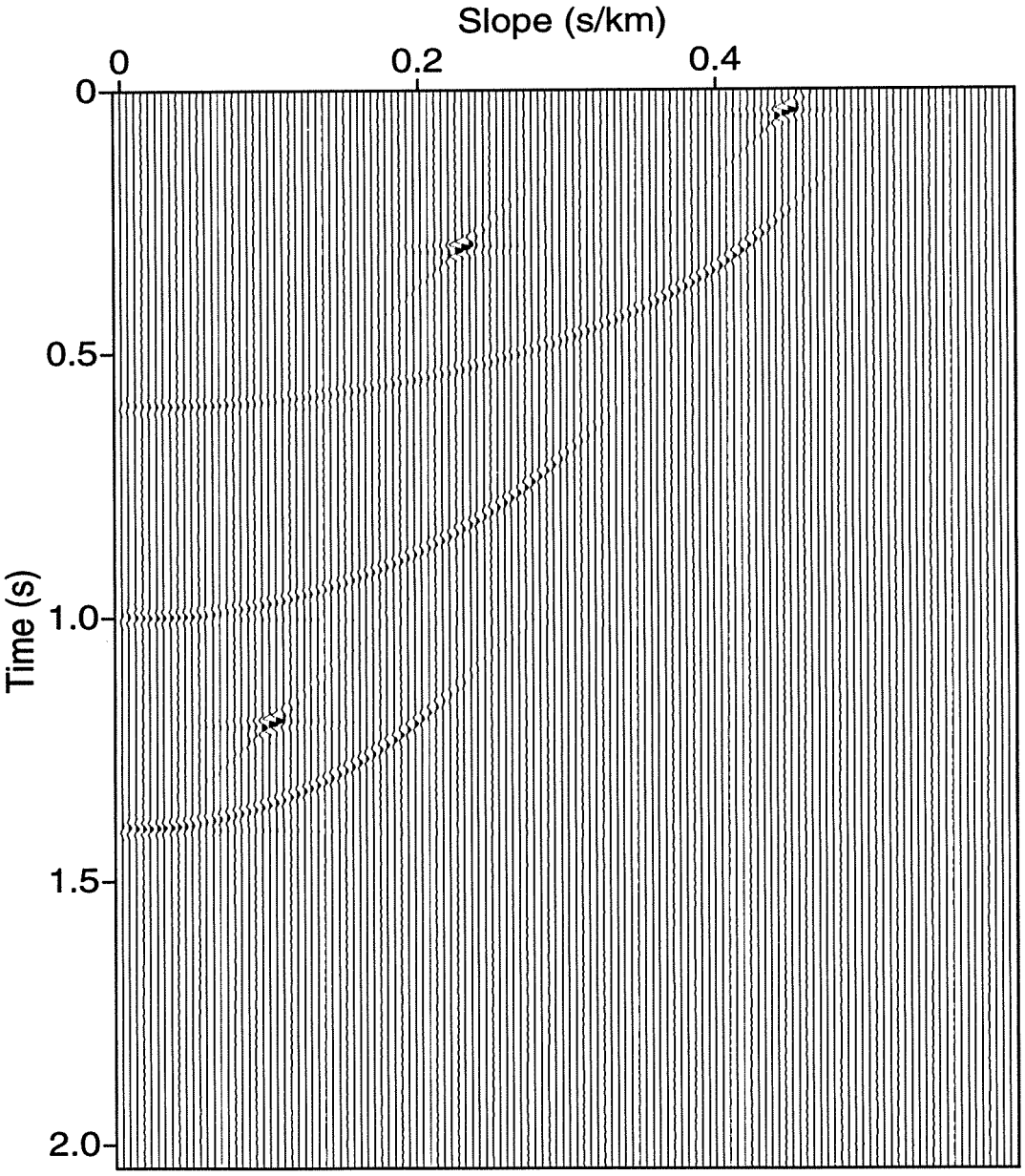

FIG. 2.11. Forward F-X τ-p transform of modeled test dataset 1.

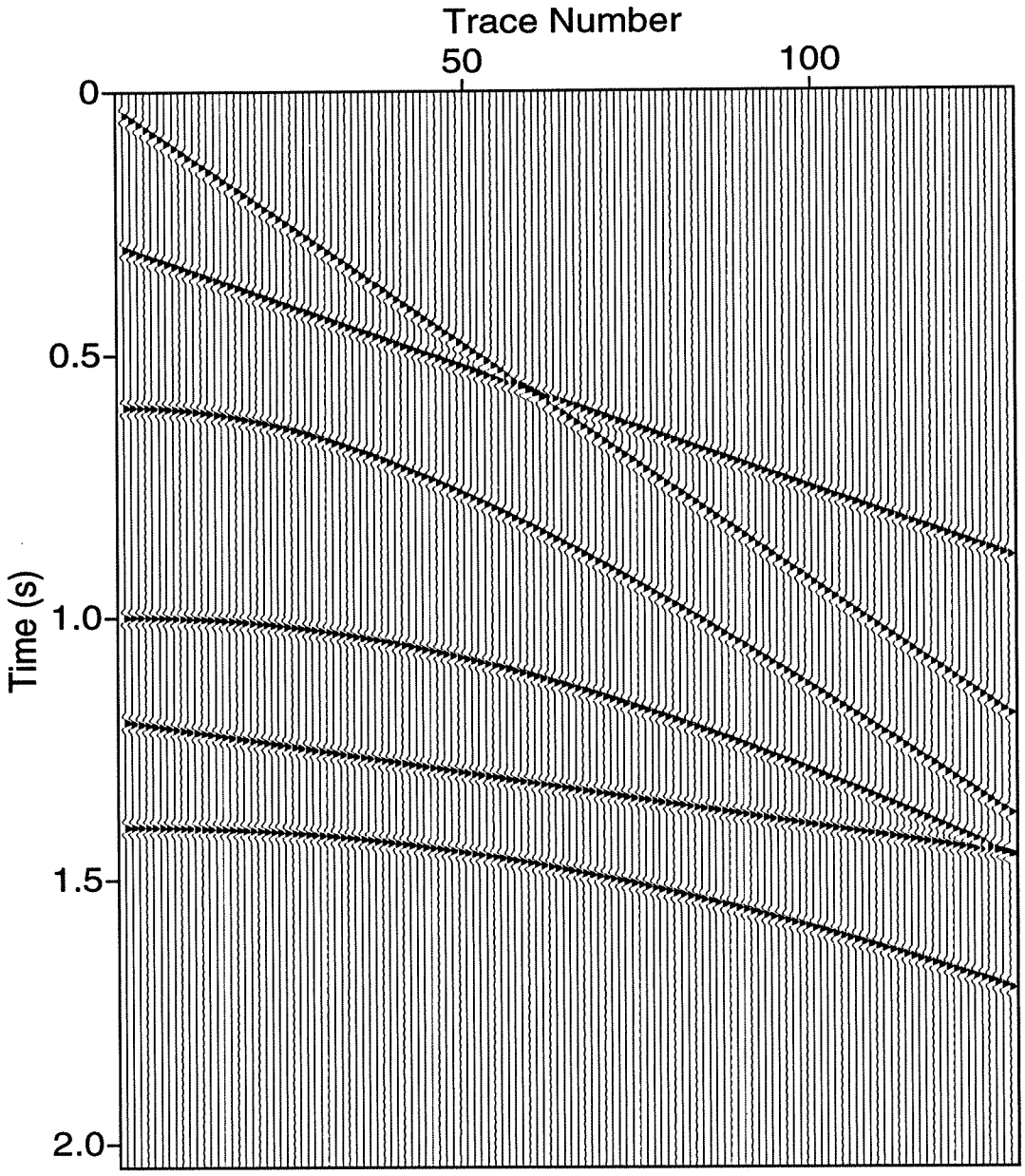

FIG. 2.12. Inverse F-X τ-p transform of modeled test data 1. The shape and position of the events have been well recovered.

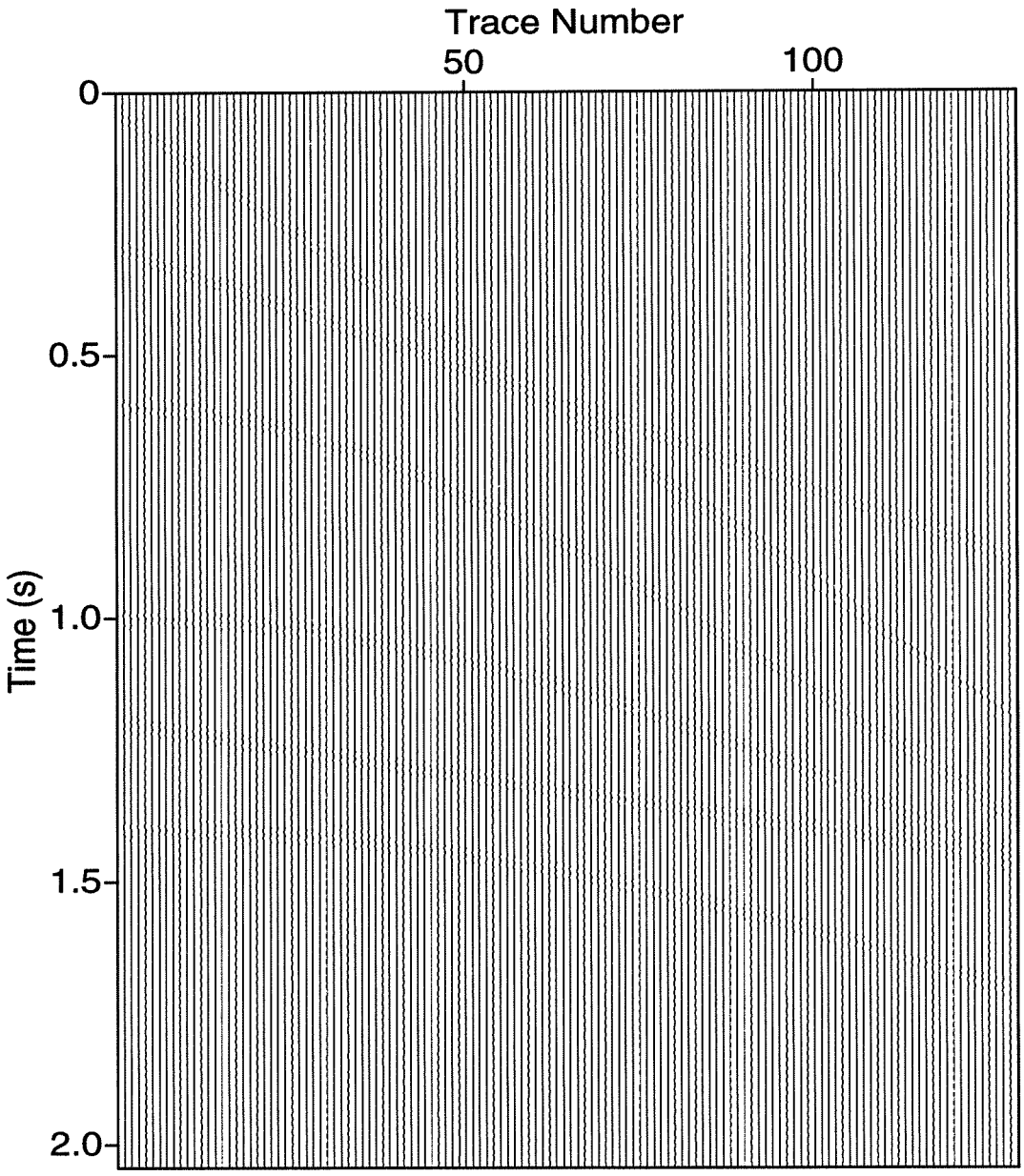

FIG. 2.13. Difference between F-X inverse τ-p transformed and input modeled test dataset 1.

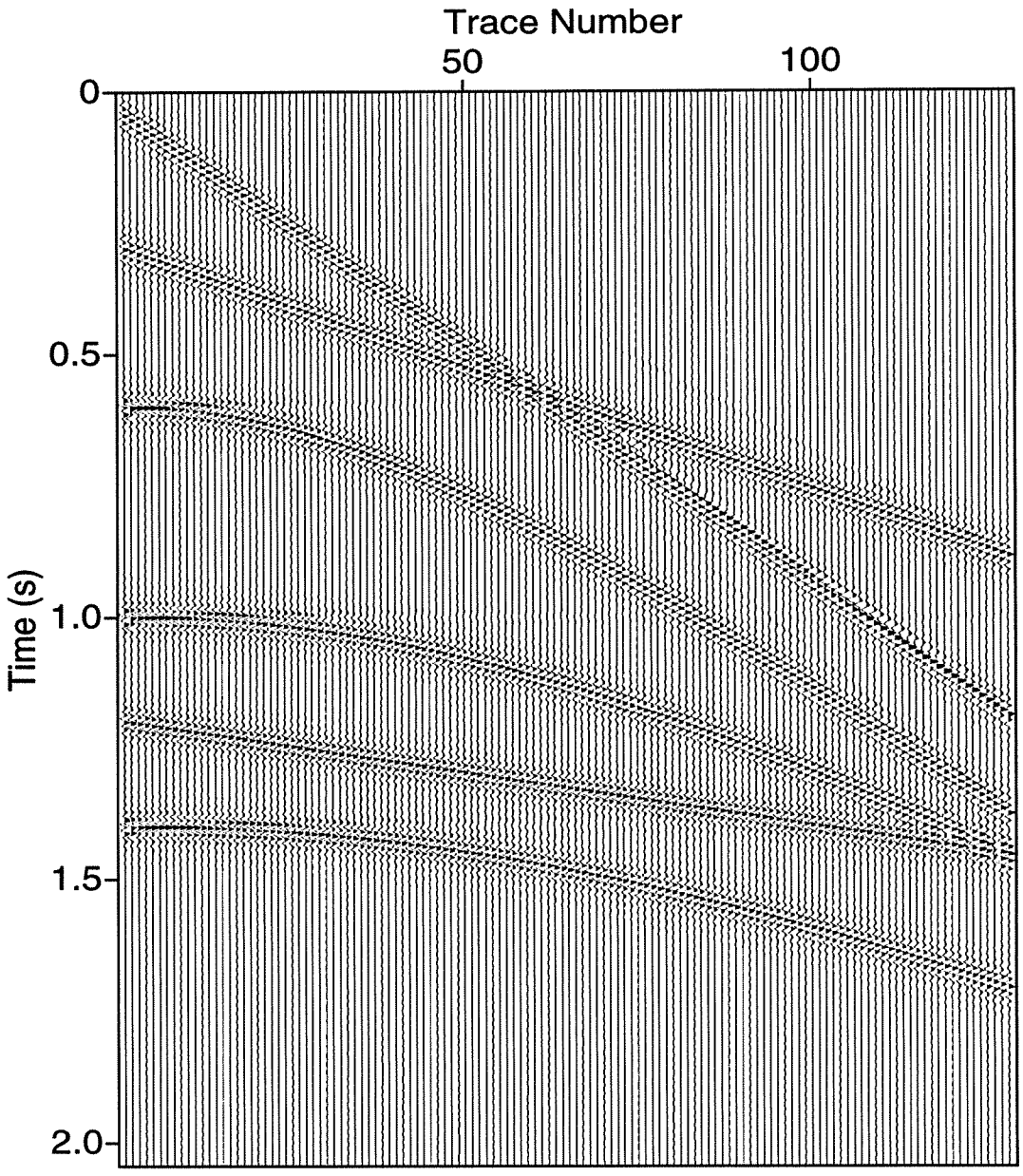

FIG. 2.14. Difference between F-X inverse τ-p transformed and input modeled test dataset 1, amplified by a factor of 10.

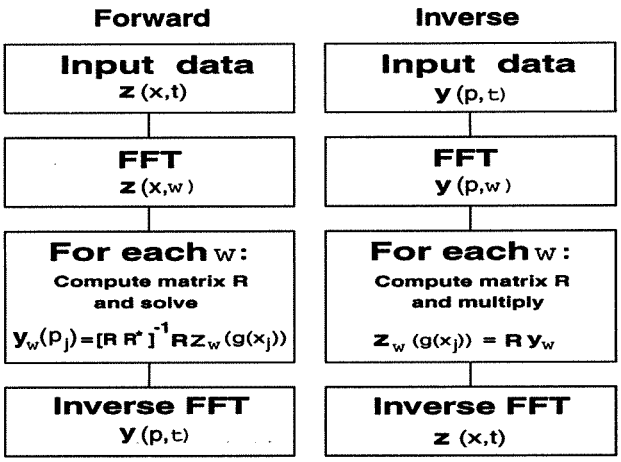

FIG. 2.15. Flow diagram for the F-X implementation of the forward and inverse Radon transform.

straight lines. The corresponding transform, called a local slant stack, maps data from a two-dimensional (x,t) space to a three-dimensional space (p,τ,x) with the transformation (Harlan et al., 1984)

$$y(p,\tau,x_c) \equiv \{\int\int W(\frac{x-x_c}{X_w}) z[x, t=\tau+p(x-x_c)]dx\} * \rho(\tau). \tag{2.23}$$

Here, $W(x)$ is a suitable spatial windowing function, in the simplest case a rectangle; x_c is the x-coordinate of the window center and X_w is the window width, normally just a few traces. In their original work Harlan et al. included the ρ filter in the forward Radon transform instead of the inverse transform (the choice is a matter of convenience), for the local slant stacks computed in the t-x domain. With the posterior development of Beylkin's method, the ρ filter is no longer needed.

The procedure to compute the forward slant stack is then to select traces in a window by the use of the windowing function $W(x)$ and apply a regular τ-p transform to the data in this window. The windows are chosen so that for each input trace there is one such window with that trace as the center trace (some roll-in and roll-out must be applied at the edges of the data, obviously). The type of window function to use is also a matter of convenience. Most times a simple rectangular window is used. If an F-K implementation is used to compute the slant stacks, however, the window function should be smoothly varying, the gaussian, $W(x) = e^{-\pi x^2}$, being perhaps the best choice.

Figure 2.16 shows one of these windows, 21 traces wide, the one centered around trace 65. A simple rectangular window was applied; note that the reflection hyperbolas look pretty much linear over this small window. Figure 2.17 shows the corresponding local slant stack, with the quasi-linear events well focused.

This procedure yields displays similar to the sonograms employed by Rieber (1936) to decompose seismic data into planes containing narrow ranges of dip (Harlan et al. 1984).

The inverse slant stack is now easily computed, since we have decomposed the wavefield recorded by a narrow range of traces centered at a given trace into plane-wave components. The information contained in that narrow window of data is recovered simply adding back those plane-wave components. The procedure is repeated for all traces (Harlan et al., 1984)

$$z(x,t) = \int y(p, x_c = x, \tau = t)dp. \tag{2.24}$$

Figure 2.18 shows the inverse slant stack of the input data, and Figure 2.19 shows the difference between the inverse and the input data, amplified by a factor of 10. Again, although the inverse transform is not perfect, the amplitude level of the difference is an order of magnitude less than that of the input. Note however, that the short offsets show a higher error for the hyperbolas, probably due to a combination of the roll-on for the windows at the edge of the data and the fact that the hyperbolas have their largest change in curvature in that region.

Note that although the local slant stacks allow us more flexibility in the approximation of arbitrary smooth curves by local segments of straight lines or other curves, they also imply a considerable increase in computational cost, depending on the number of traces used for each window and the domain of application of the transform. This increase in cost can be up to ten times that of a global transform if both the local slant stacks and the global global slant stack are computed using the preferred Beylkin's approach.

2.5 Aliasing, Stationarity and Stability

Let us look at three further critical features having to do with the efficiency and performance of the Radon transform: aliasing, stationarity and stability.

2.5.1 Aliasing

As with any transformation that deals with discretely sampled data, the issue of how finely it is necessary to sample the data in order to avoid aliasing artifacts is critical to obtaining good results with any implementation of the Radon transform. Space aliasing tends to introduce noises in the τ-p transform and to produce an imperfect inverse transform. Likewise, too coarse sampling in slopes will introduce artifacts in space (Yilmaz, 1987).

For the linear Radon transform, Turner (1990) shows that to avoid having aliasing in

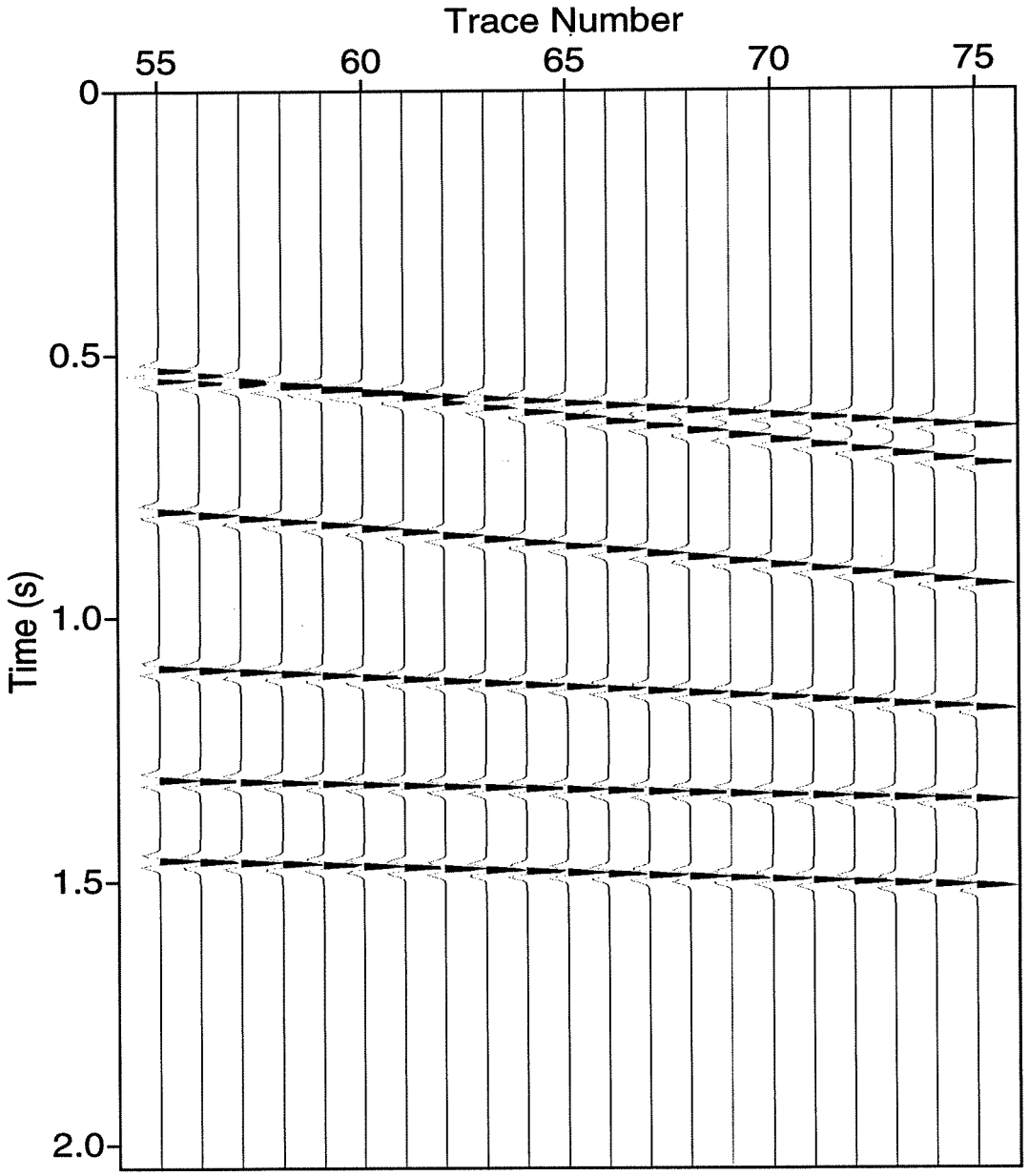

FIG. 2.16. Data window centered at trace 65 of modeled test dataset 1. Note that the segments of hyperbolas look like linear events.

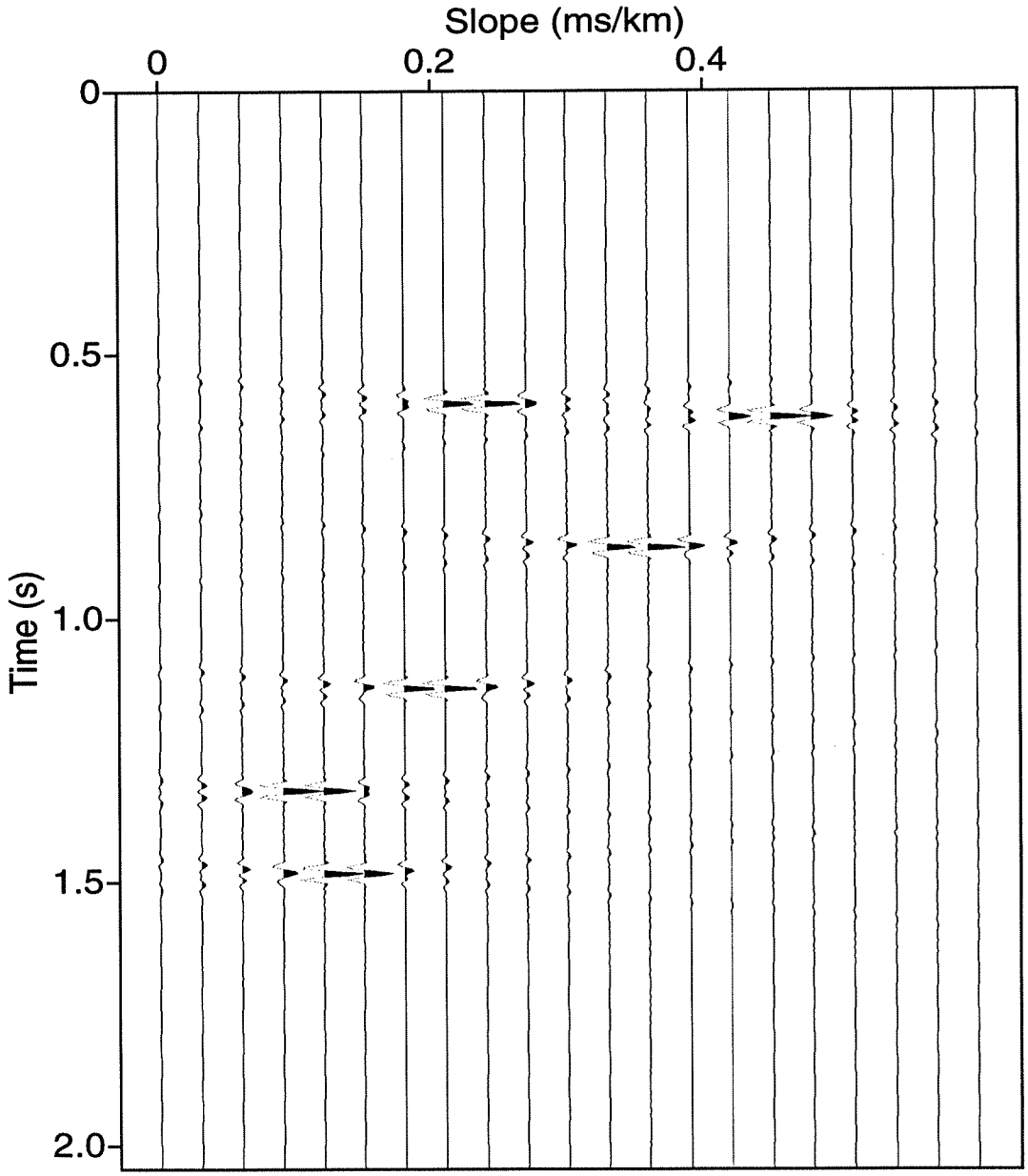

FIG. 2.17. Local F-X domain τ-p transform of windowed data. All events in Figure 2.16 have been focused.

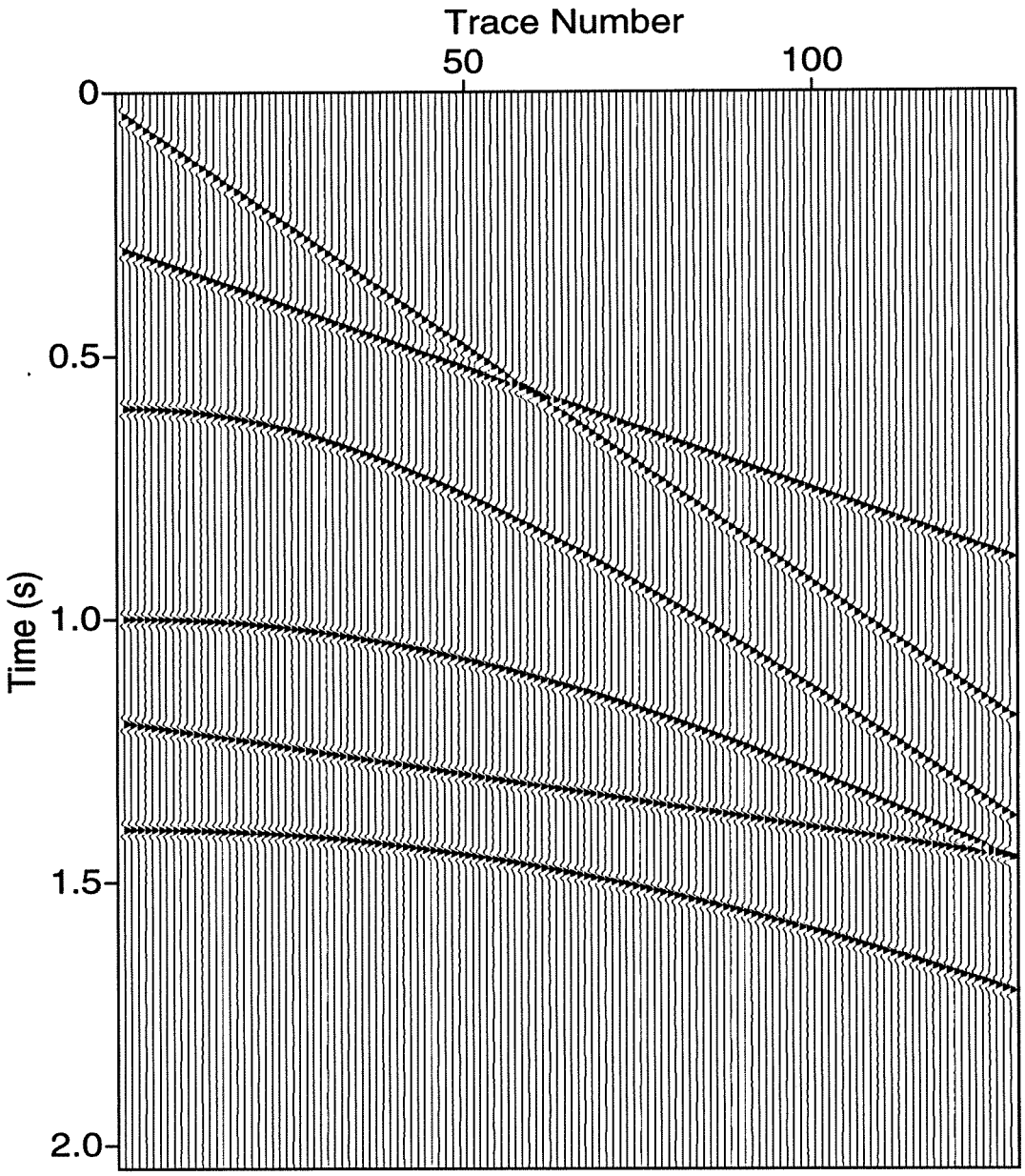

FIG. 2.18. Inverse local linear F-X τ-p transform of modeled test dataset 1.

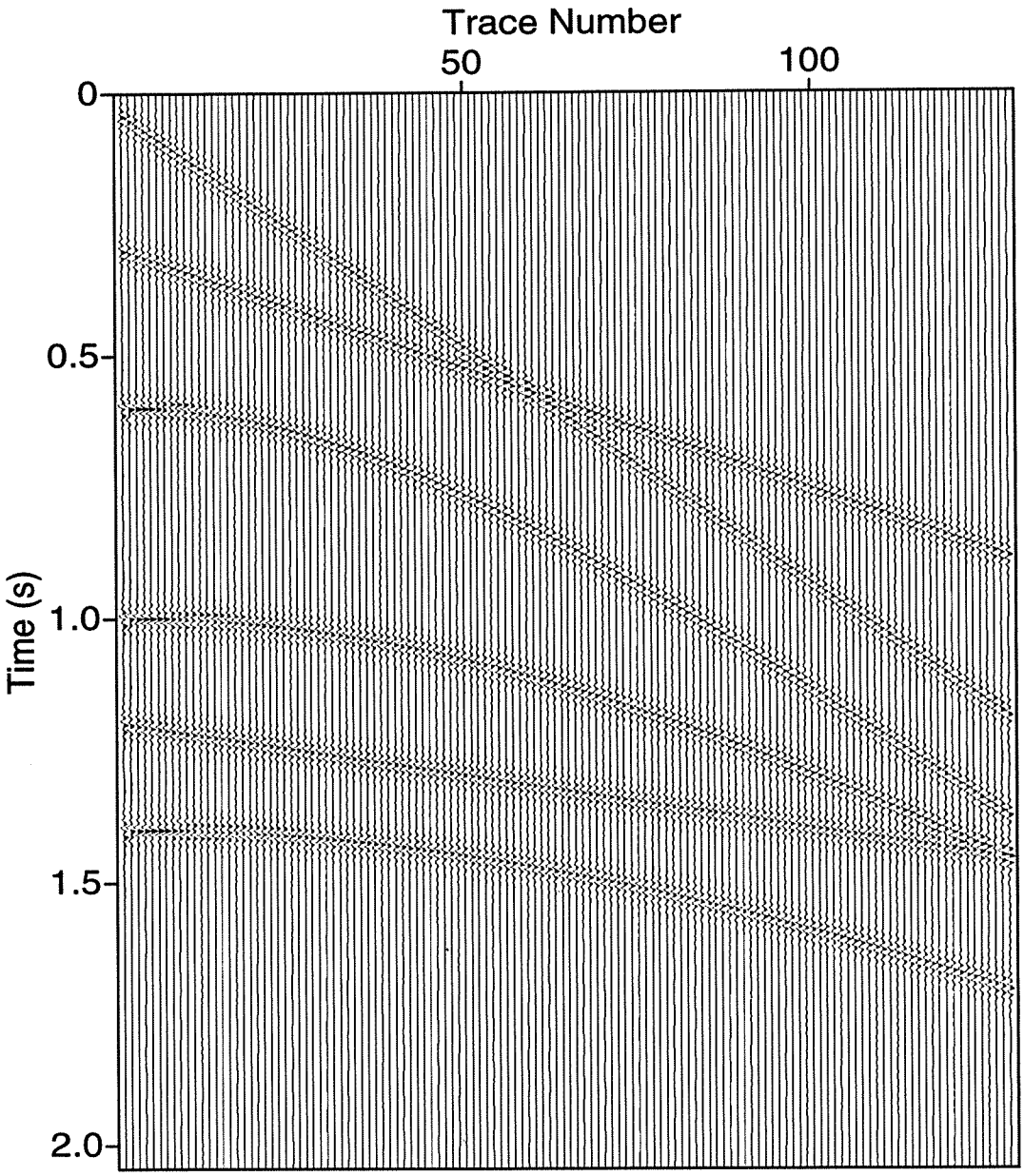

FIG. 2.19. Difference between inverse local F-X τ-p and input data, amplified by a factor of 10.

the τ-p domain, the p and x sampling intervals should be

$$\Delta p < \frac{1}{f_{max}(x_{max} - x_{min})}, \tag{2.25}$$

and

$$\Delta x < \frac{1}{f_{max}(p_{max} - p_{min})}, \tag{2.26}$$

where x_{max} and x_{min} are the minimum and maximum offsets, p_{min} and p_{max} are the minimum and maximum slopes and f_{max} is the maximum frequency present in the data. Equation (2.25) is sometimes expressed as (Harlan et al., 1984)

$$\Delta p < \frac{2\Delta t}{N_x \Delta x}, \tag{2.27}$$

where Δt is the time sampling interval, Δx is the space sampling interval and N_x is the number of traces. This equation, however, is equivalent to the expression given in equation (2.25) only if the maximum frequency present in the data is the Nyquist frequency.

More recently, Hugonnet and Canadas (1995) gave equivalent expressions, appropriate for the parabolic Radon transform (that is, the transform when $g(x) = x^2$)

$$\Delta p < \frac{1}{f_{max} \cdot [(x_{max})^2 - (x_{min})^2]}, \tag{2.28}$$

and

$$\Delta x < \frac{1}{2 f_{max} \cdot |x|_{max} \cdot [p_{max} - p_{min}]}. \tag{2.29}$$

In this equation, $|x|_{max}$ is the absolute value of the maximum offset in the data. The slope and space sampling interval are chosen to comply with these equations, and the number of slopes N_p is chosen so that the range of slopes in the data is spanned in applying the τ-p transform. As mentioned before, if the F-X implementation is used, it is better to use $N_p \leq N_x$ for the sake of computational efficiency.

In order to satisfy this and the aliasing condition, we may have to high-cut filter the data, restrict the range of offsets, or do a combination of the two.

2.5.2 Stationarity

Recall that efficient implementation of the generalized Radon transform does not require that the data be regularly sampled in space, and we can perform the transform with the same efficiency regardless of the shape of the stacking curve $g(x)$. This is true, however, only if the shape of the curves does not change with time, that is, if the stacking curves are stationary. If, as with NMO hyperbolas, this is not the case, the advantage of the F-X computation is lost because then the Fourier transform over time cannot be computed efficiently. One could still use a time-domain implementation, but it is not accurate as was shown before; thus, it is almost never used in practice.

A way to circumvent this problem is to NMO-correct the data so that any residual moveout can be approximated by parabolas (Hampson, 1986) or time-independent hyperbolas (Foster and Mosher, 1992). The first of these approaches was employed to obtain the results shown in later chapters.

2.5.3 Stability

Although efficient implementation of the generalized Radon transform is stable for all frequencies of interest, problems arise for some frequency components, such as at 0 Hz, or at frequencies where aliasing artifacts occur (Gulunay, 1990). In these cases, the matrix inversion cannot be successfully applied since the problem becomes undetermined: the matrix is no longer diagonally dominant, and therefore is not invertible. The solution, again, is to add a prewhitening factor to the diagonal elements of the matrix to stabilize the inverse. In the limit of an infinitely large prewhitening factor, the matrix becomes diagonal and the result is the same as that obtained by a traditional τ-p implementation (Anderson, 1993). As this factor is made smaller, the quality of the inverse improves, provided that it remains stable. Ideally, this factor should be chosen to be frequency dependent, so that a large value is used for only those frequencies likely to produce unstable results. In practice, however, a single value is employed for ease of implementation. For all of my tests, I have used a single prewhitening factor of 0.01%.

2.5.4 Additional optimization

When a large number of shot or CMP records must be transformed, an additional reduction in amount of computation can be done if the trace spacing is regular or if we can regularize it, for instance by stacking adjacent CMP gathers. This optimization is based on the observation that the matrix $\mathbf{R}$ is data-independent, so that it, along with the convolution operator $\mathbf{R}\mathbf{R}^{\dagger}$, need be computed only once (once for every frequency, obviously), stored, and used for all the CMP records. This will save a significant amount of computer time, but at the expense of requiring a large amount of memory to store these matrices.

2.6 Relative cost

As the data comparisons in this chapter have shown, the preferred way to compute the τ-p transform is by the use of the F-X (Beylkin's) implementation. This is due to the better focusing nature of the transform, its better handling of the amplitudes in the inverse transform and the absence of edge effects. These factors are associated, as mentioned above, with the fact that the transform approximates the limited aperture data with finite-length segments. It may seem, therefore, that there is no reason to even consider the other approaches, but there is: computational cost. The F-K implementation of the τ-p transform is the fastest because it only involves FFT's and interpolation. The t-x implementation could be even faster than the F-K implementation if the data consisted of just a few traces. The F-X approach is the slowest because it involves solving a linear set of equations, which can be pretty large, for each frequency component in the data.

For model test dataset 1 used in this chapter, the t-x implementation is slower than the F-K implementation by about 30%, and the F-X implementation is about four times as expensive as the F-K implementation.

Beylkin's (F-X) implementation of the generalized τ-p transform will be used in the remainder of this study for both Hampson's parabolic τ-p filtering and Harlan's signal-noise separation. Harlan's original work preceded Beylkin's development of the F-X algorithm so he did not originally use it. Harlan (personal communication) has since acknowledged that Beylkin's approach is the best available option because of its power to focus events of interest.

Chapter 3

THE SIGNAL-NOISE SEPARATION ALGORITHM

3.1 Introduction

Harlan et al. (1984) developed a statistical approach to separating signal and noise on the basis of the difference of their patterns as they appear in an appropriate domain, where data in that domain can be obtained by an invertible linear transformation of the input data. Here I give a description of the algorithm as applied to the separation of linear reflectors from both events with hyperbolic curvature and random noise. To illustrate the method, I generated a dataset (see Figure 3.1) by zero-offset modeling of a series of truncated linear horizontal reflectors and three isolated point diffractors. The dataset contains 100 traces with 600 samples each, at 4-ms sampling interval and trace spacing of 20 m. Of particular importance for this dataset is to evaluate the ability of the method to extract even the extremely weak diffractions off the edges of the truncated horizontal reflectors.

3.1.1 Focusing the data by a linear, invertible transformation

The signal-noise separation algorithm must be capable of recognizing and exploiting the characteristics that differentiate the signal pattern from the noise pattern after the application of a suitable linear transformation. Although the type of transformation used depends on the data and the application at hand, let us concentrate in this chapter on the identification and extraction of linear coherent events from either incoherent events or coherent events with moveout other than linear. Linear events require only two parameters for their complete specification: their slope and their zero-offset intercept time. In a seismic section, however, reflection events are seen as amplitudes on every trace; in other words, their patterns are overspecified. Incoherent noise, on the other hand, not being laterally coherent, cannot be specified by any fewer number of parameters. Therefore, if we apply a linear τ-p transform to the data, we focus the linear events (that is, decrease to the minimum the number of parameters required to describe them) without focusing the noise. Obviously, coherent nonlinear events will not be focused by the transformation and will, for the current purpose, behave as noise (they could be focused by a different transform if that were desired). Figure 3.2 shows the τ-p transform of the model dataset 2, using the F-X (Beylkin's) approach. Note that the linear events have been focused to small regions of the τ-p plane, although not to points because of the limited number of traces. The diffraction hyperbolas, on the other hand, were not focused by the transformation, but were mapped to ellipses that are barely visible in that figure.

From a mathematical point of view, the transformation increases the statistical independence of the data samples, by decreasing their lateral predictability. This offers an important

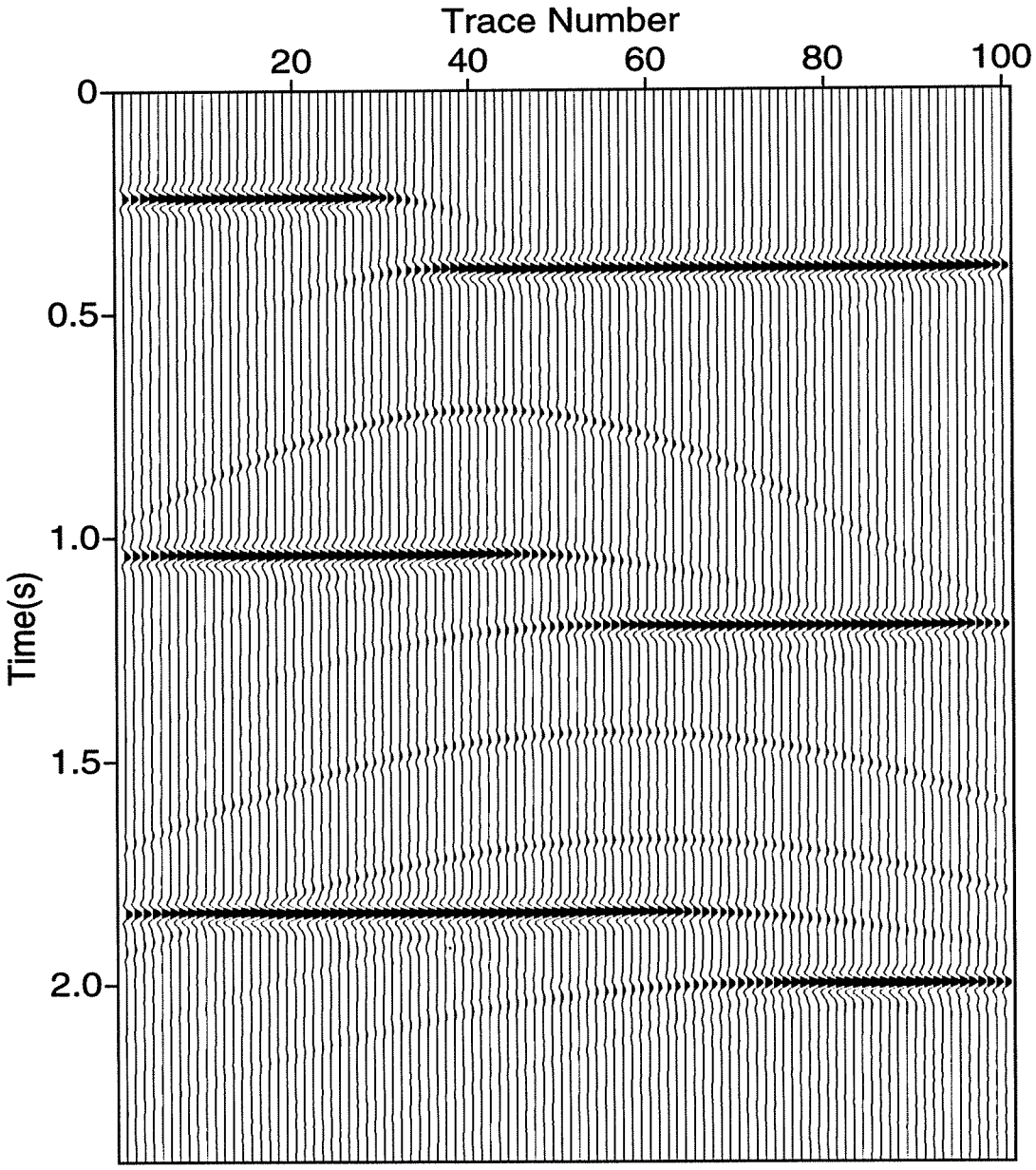

FIG. 3.1. Modeled dataset 2, obtained by zero-offset modeling of truncated linear reflectors and three point diffractors.

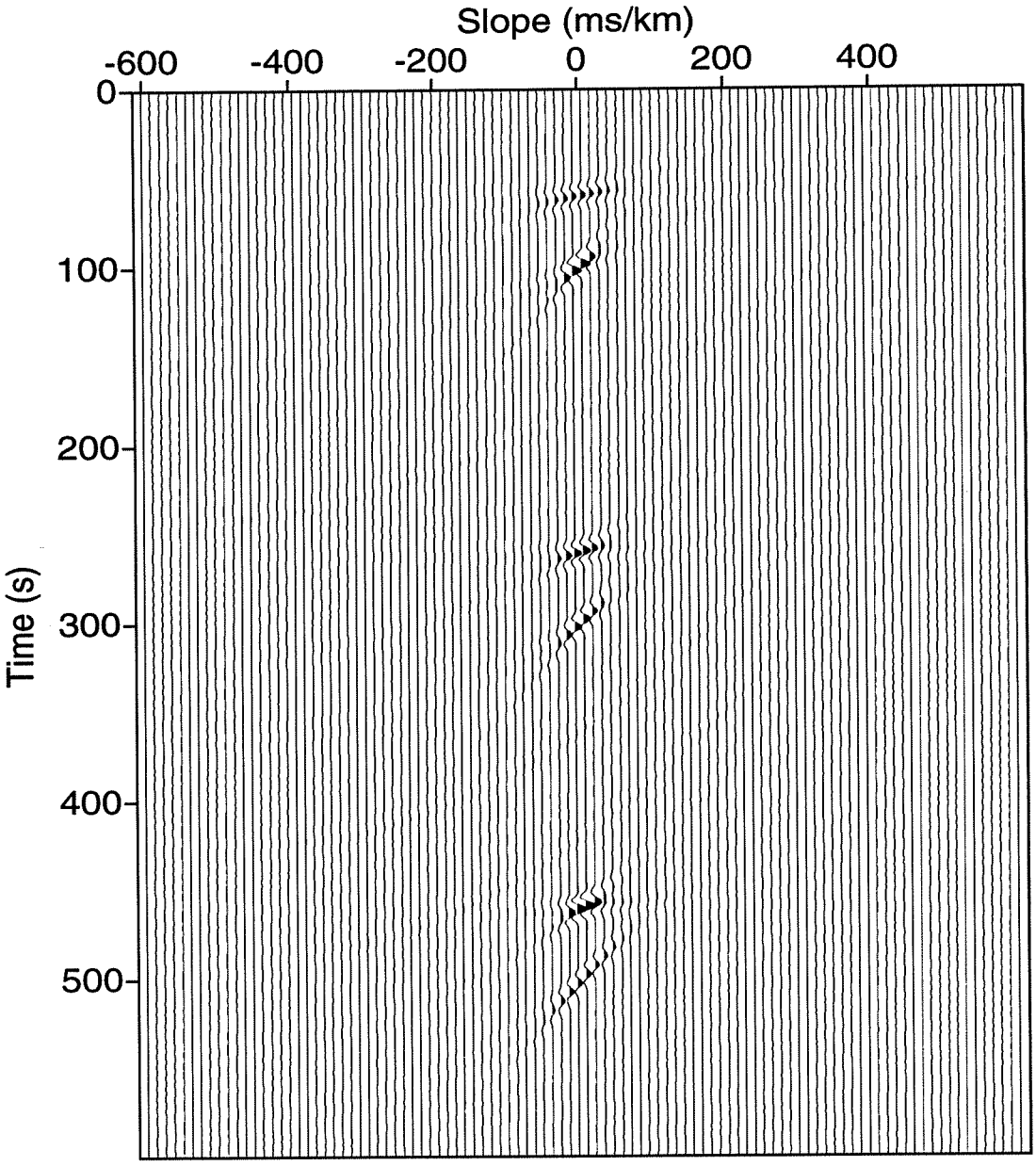

FIG. 3.2. Linear F-X τ-p transform of model dataset 2. Note the focused energy corresponding to the linear events.

advantage for the way we can handle the statistical characteristics of the data. In their original domain, the samples of the signal (the linear events) are far from being statistically independent, so we cannot describe them using marginal probability density functions (which require the samples to be statistically independent); we need to deal with their joint probability density functions. These joint probability density functions, however, cannot be computed with just one realization of the data, so the recorded data do not have enough redundancy for their computation. If we apply an appropriate transformation, as explained before, the data samples that correspond to the signal are focused and so its samples become statistically independent (or at least we will assume that this is the case) and we can now work with marginal probability density functions, which can be computed relatively easily from one-dimensional histograms, provided that we can neglect any correlation of the signal or noise samples after the transformation.

The first and most critical step in the application of the algorithm, is to choose the appropriate transformation that will focus the data without focusing the noise, with our working definition of noise as anything that is not focused by the given transformation.

3.1.2 Basic Assumptions

Some of the basic assumptions of the algorithm may be obvious from the previous discussion, but will be repeated here for the sake of reaffirming their importance. In order for the method to work, three basic assumptions must be satisfied

- An appropriate transformation must exist that, ideally, will focus the signal to the smallest possible number of parameters required to describe it.
- The same transformation must defocus the noise (or at the very least not focus it). Equivalently, we redefine noise as any event that is not focused by the transformation.
- After the transformation, the signal and the noise are considered statistically independent, uncorrelated random variables. This may be debatable, but it is a basic assumption that seems to be satisfied by most datasets.

3.2 Description of the Method

I now describe, step by step, the different computations required to implement the method of Harlan et al. (1984), leaving some particular, yet crucial, computer-implementation details for a later section after the basic ideas about the method have been explained.

3.2.1 Computation of the probability density functions

As we shall see, the essence of the algorithm is the statistical separation of the signal and the noise based on estimates of their probability density functions. In this section, we see how estimates of these functions are computed.

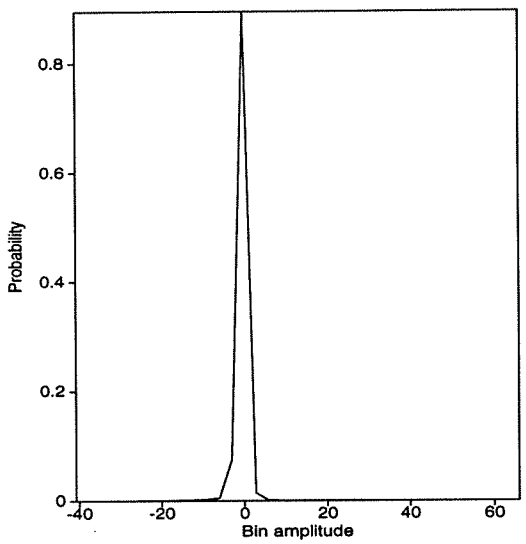

FIG. 3.3. Data probability density function. Note that most samples fall in the smallest-amplitude bins.

Data probability density function With the assumptions mentioned above, the data probability density function is computed from the transformed data in a straightforward way by the computation of an amplitude histogram. The only parameter of concern is the number of bins to use in computing the histogram. I have found this not be a critical parameter; any number between 50 and 100 seems to be adequate. Figure 3.3 shows the computed data histogram for the transformed model dataset 2 (Figure 3.2), with the ordinate re-scaled from frequency of occurrence to probability density values by ensuring that the area under the curve is unity. From this figure, it is obvious that the vast majority of the samples have amplitudes close to zero (that is, they correspond to samples that were not focused by the transformation), and a relatively few correspond to the high-amplitude, focused samples. That some higher amplitudes are present is better seen in Figure 3.4, which is a detailed close-up of the low-probability region of Figure 3.3.

Noise probability density function The noise probability density function is not so straightforward to compute because we do not have information as to what exactly is the noise. (Noise is one of the components we are trying to isolate.) The solution to this difficulty lies on our working definition of noise: anything that is not focused by the linear transformation. What this means is that, although there is no way to know the exact probability density function of the noise, we can get at least a pessimistic estimate of it, assuming that everything in the data is noise (it is not, of course; that is why the estimate is pessimistic). This pessimistic estimate means that we are being conservative about our signal extraction, desiring to prevent any noise from appearing as signal. In a subsequent iteration of the

process, we can compute a better estimate of the noise distribution after we have computed and extracted a first estimate of the signal present in the data; this will be demonstrated in a later section. A mathematical justification for these facts is given in Appendix B of Harlan et al. (1984). From our previous discussion, the way to make the data behave as noise is to artificially destroy the linear coherence of the events. This can be done in many different ways. Here are some

- randomly reverse the polarity of the traces
- randomly introduce static shifts to the data
- randomly alter the order of the traces

The simplest one to implement is the first one, and this is the approach used in this study (see Figure 3.5).

We next apply the transformation to the disrupted data and obtain as a result that the energy is not focused, but instead is rather spread out all across the transformed section (see Figure 3.6), as we wanted it to be. We can now compute the probability density function of the "noise" with a histogram, in exactly the same way that we did for the data. Figure 3.7 shows the estimated probability density function, and Figure 3.8 shows a detailed view of the low-probability region. These figures show that we still have a predominance of samples with low amplitudes. No significant energy is found with the high amplitudes because no significant energy was focused by the transformation. One point to note here is that the same amplitude intervals have to be used for both the data and the noise histograms.

Signal probability density function We cannot directly compute the signal probability density function for the same reason that we could not with the noise: at the outset, we do not know what is the signal. At this point, we make use of the third assumption listed above, that the signal and the noise are (or at least we assume that they are) independent random variables. For every sample of data in the transformed domain, we can write

$$d = s + n, \tag{3.1}$$

where d means data, s signal and n noise; that is, every data sample is the result of a contribution from the noise and a contribution from the signal (although the amplitude of one or both of them could be zero for any particular sample). For signal and noise independent of one another, the probability density function of the signal and of the noise are related to that of the data by (Papoulis, 1965)

$$p_d(x) = p_s(x) * p_n(x), \tag{3.2}$$

where the asterisk denotes convolution. Therefore, provided that signal and noise are independent random variables, the knowledge of any two of these probability density functions determines the other. In particular, since we have estimated $p_d(x)$ and $p_n(x)$, we can, in principle, deconvolve equation (3.2) to get $p_s(x)$. This may sound trivial, but an ordinary

deterministic deconvolution will not work since $p_s(x)$, being a probability density function, has to satisfy two restrictions:

- all samples of $p_s(x)$ have to be non-negative
- the area under the curve of $p_s(x)$ has to be unity

What this means, is that some sort of constrained deconvolution is required in order for the result $p_s(x)$ indeed to be a probability density function. Appendix B of Harlan 1988, shows that a valid form of $p_s(x)$ can be obtained by posing the problem as an optimization problem: Given $p_d(x)$ and $p_n(x)$, find the $p_s(x)$ that minimizes the error between $p_d(x)$ and $p_n(x) * p_s(x)$ in the least-squares sense. This can be done using any iterative technique such as conjugate gradient or steepest descent, with the two constraints mentioned above introduced as parameters via Lagrangian multipliers.

As with any optimization problem that requires an iterative solution, we are still faced with the questions of what to use as the starting model for $p_s(x)$ and how to determine when to stop the iterations. In order to answer the first question, let us think of the data in the transformed domain. We have focused the signal (in the present example, the linear events) to small areas (they would be points if we had an infinite number of traces) and we have spread out the noise (defocused it) all over the transformed domain. This means that we expect our signal probability density function to be close to a spike, with a large number of samples having low absolute amplitudes (that is, most of the points in the transformed domain do not have any significant contribution from the focused signal) and just a few points (those focused by the transformation) having large absolute values. A spike centered at the zero-amplitude bin is therefore the natural choice for a starting guess. The answer to the second question is less precise. We stop the iterations when either the difference between $p_d(x)$ and $p_n(x) * p_s(x)$ falls below a given threshold or when the change in the difference from one iteration to the next becomes small. I have used the first stopping criterion and found this threshold to be a parameter that does not need to be finely tuned. I normally get good convergence within a few (less than ten) iterations.

In summary then, we can directly compute an estimate of $p_d(x)$, compute an admittedly pessimistic estimate of $p_n(x)$, and then use the two to estimate $p_s(x)$ as an optimization problem. As seen in Figures 3.9 and 3.10, the result of doing this computation, the computed $p_s(x)$, departs only slightly from the starting spike. That slight departure, however, is the key to the computation, because for the highest amplitudes, we have basically no contribution from the noise. Moreover, since the main contribution to those large amplitudes is from the signal, those large values are the most valuable samples in our signal extraction; that is, there is near-zero probability that those samples are noise. It should also be noted that Figure 3.10 shows more samples with large positive amplitudes than those with high negative amplitudes. This is a consequence of the fact that the model data were created with only positive reflection coefficients, so that the negative amplitudes appear only because of the lobes of the Ricker wavelet used in generating the data.

Now, how close are we able to recover $p_d(x)$ from the convolution of $p_s(x)$ and $p_n(x)$? Figure 3.11 answers that question, and the answer is that we cannot exactly recover $p_d(x)$,

but, as will be seen later, we can get close enough for the method to work. Note in particular that we have been able to recover the highest amplitudes, which is good because they are the most important ones for our purpose of extracting signal.

Once we have these probability density functions, what remains to be done is to use them to estimate how much signal is present in a given sample of data and to zero out those samples that are judged to contain less than a prescribed amount of signal. That is, the next step is to develop a criterion for selecting a threshold amplitude above which data focused in the τ-p domain are considered signal, and below which data are treated as noise.

3.2.2 A Bayesian Signal Estimator

Following Harlan et al. (1984), we compute the expected value of the signal given the amplitude of the data sample. This Bayesian estimator, $E(s|d)$, is the amount of signal estimated to be in a sample of data with amplitude d. It is given by (Harlan et al., 1984; Papoulis, 1965)

$$E(s|d) \equiv \int x p_{s|d}(x|d)dx = \frac{\int x p_s(x) p_n(d-x)dx}{p_d(d)}. \tag{3.3}$$

In this equation, x is the bin center corresponding to a sample with amplitude d (in the transformed domain), and the summation is carried out over all histogram bins. The quantity $p_{s|d}(x|d)$ represents the conditional probability of the signal given the data, and is a function of the bin amplitudes x and the sample amplitude d. The numerator in equation (3.3) can be easily computed by recognizing that it is just a convolution of $xp_s(x)$ with $p_n(x)$. The denominator is not the original data probability density function but rather its estimate computed as a convolution of the estimated signal and noise probability density functions.

This equation has a remarkable interpretation. It is telling us that, for every sample of data, we can directly estimate what is the expected value of the signal contribution (and thereby of the noise contribution) and so we have a direct way of deciding which samples are significantly contaminated by noise and thus should be zeroed out. Two important extreme cases can be used to illustrate this. If all of our data consist of signal, that is, if no noise is present, then $p_n(x) = \delta(x)$ and $p_d(x) = p_s(x)$, so we have:

$$E(s|d) = \frac{\int x p_s(x)\delta(d-x)dx}{p_d} = \frac{d \cdot p_d(d)}{p_d(d)} = d. \tag{3.4}$$

That is, in this noiseless case, all the data samples are judged to be pure signal, i.e., they are 100% reliable. In the other extreme, if the data consist entirely of noise (remember, noise is not necessarily random, it is just anything not focused by the transformation, including linear events whose slopes are outside the range of values allowed for the τ-p transformation), we have $p_s(x) = \delta(x)$ and $p_d(x) = p_n(x)$, so:

$$E(s|d) = \frac{\int x\delta(x) p_n(d-x)dx}{p_d(d)} = 0. \tag{3.5}$$

This result states that, in this case, none of the samples would be found to contain any signal

and so are 100% unreliable. In all intermediate situations, we get intermediate results; thus, we have a way of discriminating signal and noise for every data sample in the transformed domain on the basis of their computed $E(s|d)$ value.

Figure 3.13 shows this estimator for our example. Except for the low-amplitude samples, the amplitude of the signal is roughly equivalent to the amplitude of the data (i.e., the curve of $E(s|d)$ versus d is close to the dashed line with unit slope), telling us that these samples are unlikely to contain any noise.

3.2.3 Reliability measure

From the previous discussion, we might think that once we have computed the Bayesian estimator we are basically home free, having achieved our goal of discriminating the signal from the noise. It is now possible to set a threshold value for ascertaining whether or not a sample should be considered as signal; samples that have an $E(s|d)$ value larger than, say, 80% of the data amplitude will be considered to correspond to the signal we wish to extract, and those with $E(s|d)$ values less than that will be considered to correspond to the noise we wish to suppress and so will be zeroed out. There is a problem, however, that we have not addressed: how reliable is the Bayesian estimate $E(s|d)$; that is, how much can we trust that if the estimator indicates that a given data sample is, say, 80% signal and 20% noise, it is not in fact, say, 75% signal and 25% noise?

In order to assess the reliability of the Bayesian estimator, we can define a reliability indicator as the conditional probability that the estimated signal (that is, the value given by the Bayesian estimator) is within a certain percentage of the true (unknown) value. Its mathematical expression is given by (Harlan, 1988)

$$\begin{aligned} \textit{reliability} &\equiv P\left[(-c\hat{s} < s - \hat{s} < c\hat{s}) \,|d\right] && (3.6) \\ &= \frac{\int_{\hat{s}-c\hat{s}}^{\hat{s}+c\hat{s}} p_s(x) p_n(d-x)dx}{\int_{-\infty}^{\infty} p_s(x) p_n(d-x)dx}. && (3.7) \end{aligned}$$

Here, as before, s is the signal amplitude, $\hat{s}$ is the estimated signal amplitude given by the Bayesian estimator, d is the data amplitude and c is a measure of how much uncertainty we are willing to accept in the Bayesian estimate of signal. Normally, c is chosen to be a small value, less than say 5%. The numerator of this equation is a bounded convolution that represents the contribution of the signal to the data sample, whereas the denominator, as before, is the estimated probability density function of the data. If a given data sample were all signal, the numerator and the denominator would be the same and the reliability of that sample would be 1, that is 100% reliable. The data samples with the highest amplitude (in the transformed domain), are therefore the ones most likely to correspond to signal with the highest reliability. This reliability indicator can then be used to identify the signal samples as those with the highest reliability.

The question of how high is high reliability depends on how well the transform was capable of focusing the signal. It is a processing parameter that is data-dependent and requires testing for each dataset and each application of the algorithm. As a matter of fact,

it is *the* parameter to use in deciding how much signal is to be extracted. The parameter c is not so critical and I have found that any value between 1 and 5% works well.

Figure 3.14 shows the result of the calculation of the reliability parameter for model dataset 2. As expected, the low amplitudes have basically zero reliability, and this reliability increases rapidly with increasing amplitude.

For this particular dataset, since the amplitudes of focused events are much larger than those of defocused events, a small value of reliability is used to discriminate between reliable and unreliable samples. Figure 3.14 shows that the transition between unreliable samples (for which the reliability values is almost zero) and reliable ones for which the reliability indicator is significantly greater than zero, is very steep. This is a consequence of the high amplitudes of the focused energy and the very low amplitudes of the unfocused energy (the barely seen ellipses in Figure 3.2) which makes only the very low-amplitude samples unreliable. Thus, for this dataset, the reliability value necessary to discriminate between reliable and unreliable samples needs only be able to detect for which bin amplitude the reliability curve (Figure 3.14) abruptly increases upward from near zero (again, the unquestionably unreliable samples). A value of 0.001 did the work for this dataset. This may seem too small a value for a parameter that represents the reliability of the data samples but, as implied above, it is just telling us that, for this dataset, only the samples that fall in the very low-amplitude bins (from about -15 to 15) are unreliable, and that larger-amplitude samples, although having small reliability values, are in fact reliable enough to contribute to the solution. For other datasets, in which the transition from small to large reliability is not so steep, a much larger value of reliability may be necessary to identify the reliable samples. As mentioned before, the Bayesian estimator itself could be used to extract the reliable samples, but it is not so accurate, in particular because it may be difficult to numerically compute $E(s|d)/d$ for small values of d for which both the numerator and the denominator tend to zero. Other ways to discriminate between reliable and unreliable samples may also be possible. Perhaps even the direct comparison of the data, noise and signal histograms could give us a good idea of what is the minimum sample amplitude to consider samples as reliable. In all my tests, however, I used the reliability measure as defined above to do the extractions.

Once the unreliable samples are identified, they are zeroed out, so that what remains constitutes what are judged to be signal samples that can be inverse transformed to get the extracted signal back in the time domain. The extracted noise can then be computed by subtracting the extracted signal from the original data.

Results of this process for the test data in Figure 3.1 are presented in Figure 3.15 and Figure 3.16. The algorithm was successful in identifying and extracting even the weakest diffractions despite the fact that their amplitudes were much smaller than those of the reflections (to the point that they were virtually invisible in the model plot, Figure 3.1) and that they were coincident with the signal at some points. Some linear coherent energy from the signal still remains, however, indicating that the extraction was not perfect. A second iteration of the algorithm can be made, with the noise estimated from Figure 3.16 rather than from the input data. This is important and will be discussed in a later section.

3.3 Computer-implementation details

Having reviewed how the approach of Harlan et al. uses the focusing power of the transformation, along with a statistically derived threshold amplitude detection to separate signal and noise, some critical implementation details need to be addressed.

3.3.1 Analytic envelope

The approach described above should work well for a sample-to-sample extraction, that is, assuming that the extraction of a sample is completely independent of that of neighboring samples. This would be the case if the transformed data consisted of spikes, but seismic data consists of band-limited wavelets. Thus, we want to be able to extract, as reliable, the complete wavelet and not just its peak. In other words, for a given wavelet, we not only want, but require, that the program be able to identify as comparably reliable all of its samples, whether peak, trough, zero crossing or values in between. This obviously means that we cannot solely use the sample amplitude to do the extraction, but instead need a more global measure of the reliability of amplitudes, one that looks at the larger picture and is not fooled into believing that the zero crossings of a reliable wavelet, being zero amplitude, are unreliable. A useful tool that provides such a global measure of amplitude is the analytic envelope of the transformed traces. Harlan (1985) suggests using this analytic envelope for the extraction. Given a time series $z(x)$, its analytic envelope is defined as (Bracewell, 1967)

$$y(x) = \sqrt{z^2(x) + (H[z(x)])^2}, \tag{3.8}$$

where the capital H means the Hilbert transform of the time series $z(x)$.

In order to be able to extract the complete wavelet, we compute the analytic envelopes of the transformed traces and use their samples as "d" in computing the reliability measure, equation (3.7). The sign of the analytic envelope samples is that of the transformed data samples. Analytic envelopes were used in the computation of reliability for all the results shown in this study.

3.3.2 Iteration

A refinement included in the method of Harlan et al. consists of applying a second iteration using the extracted noise as a better estimate of the noise present in the data than that obtained assuming that everything in the data was noise. It is still necessary to destroy any residual coherency that might be left in the estimated noise by randomly reversing the polarity of its traces, since we prefer to overestimate the noise to guarantee that no signal is lost in the extraction.

I have found this second iteration to be more needed for noiseless modeled data than for field data probably because field data have a significant amount of random noise to begin with, so the initial estimate of the noise is not too far away from the true noise character.

Figures 3.17 and 3.18 show results obtained with this second iteration of the method. Comparison of these with their counterparts in Figures 3.15 and 3.16 shows that, here, a

significant improvement was obtained by this second iteration, in particular for the shallow events; not only is the level of noise rejection better, but also, less noise was allowed in the extracted diffractions (the arrows in Figure 3.18 point to places with improvement over that in Figure 3.16).

3.3.3 Smoothing

It is desirable to have a smooth transition between accepted and rejected samples because a sharp transition will distort the spectrum of the extracted signal, which will be evidenced as ringiness in its Fourier transform. This smooth transition can be done by applying a two-dimensional smoothing to the extracted data, or equivalently, flagging samples judged reliable with ones, those judged unreliable with zeros and smoothing this array of ones and zeros vertically and horizontally prior to multiplying it with the data. This will have the desired effect of providing a smooth transition between samples accepted as reliable from those rejected as unreliable. The level of smoothing is a data- and transform-dependent parameter that in principle is necessary to test for every application. I have found this smoothing parameter, however, not to be critical in the applications that I have investigated. A vertical smoothing over 10 time samples and a horizontal smoothing over two slope values was applied in all the tests that I have done.

To show how important is this smoothing, Figure 3.19 shows the result of applying the algorithm, with exactly the same parameters used to generate Figure 3.18, but without the smoothing. The lack of smoothing has resulted in a much noisier estimate of the diffractions.

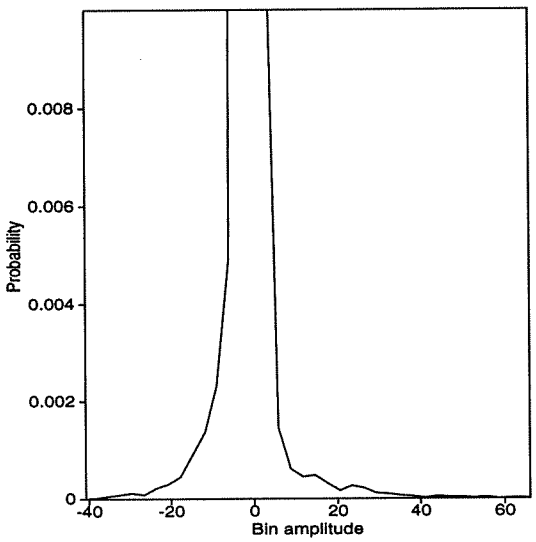

FIG. 3.4. Close-up of the low-probability region of the estimated data probability density function to show that there are, in fact, a small number of samples with high amplitudes.

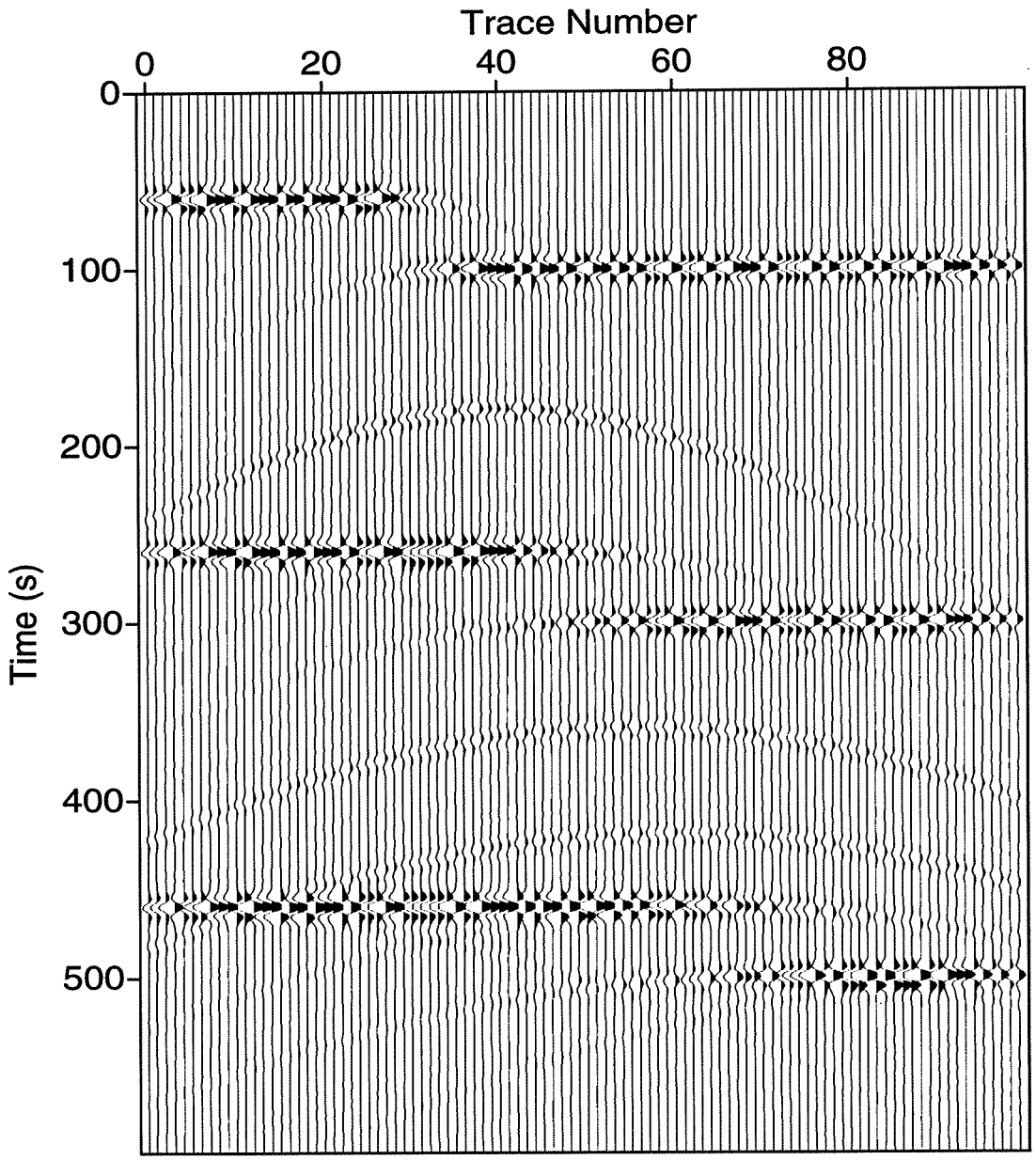

FIG. 3.5. Randomly reversed input data used to estimate the noise probability density function.

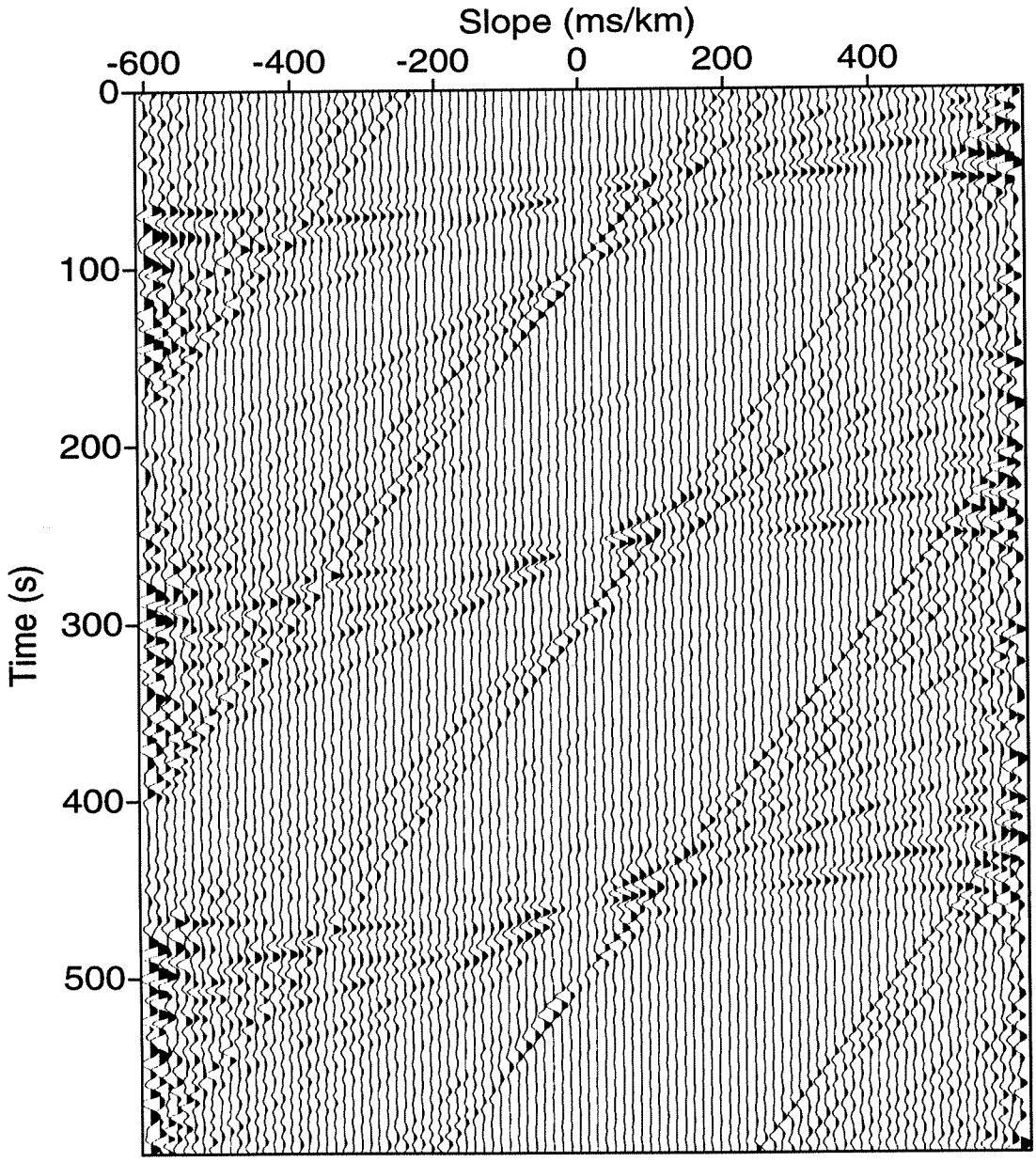

FIG. 3.6. τ-p transform of randomly reversed input data, amplified by a factor of 5.

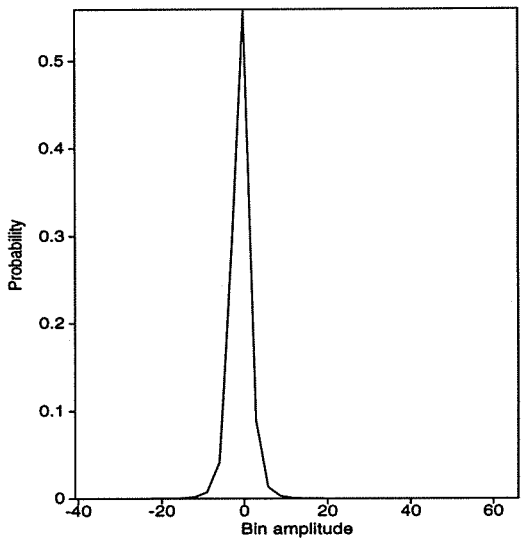

FIG. 3.7. Estimated noise probability density function.

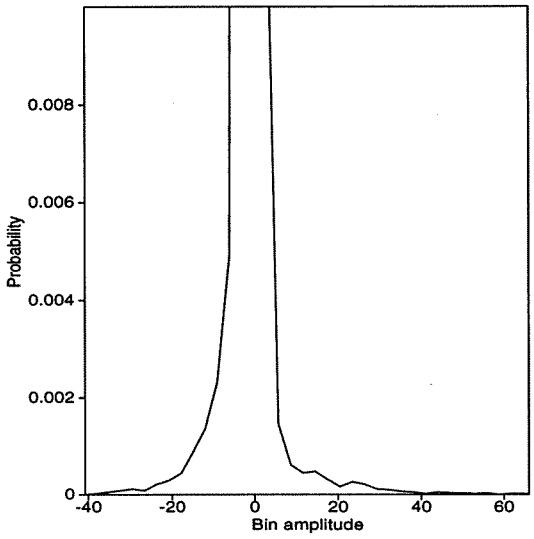

FIG. 3.8. Close-up of the low-probability region of the estimated noise probability density. There are no samples with high amplitudes.

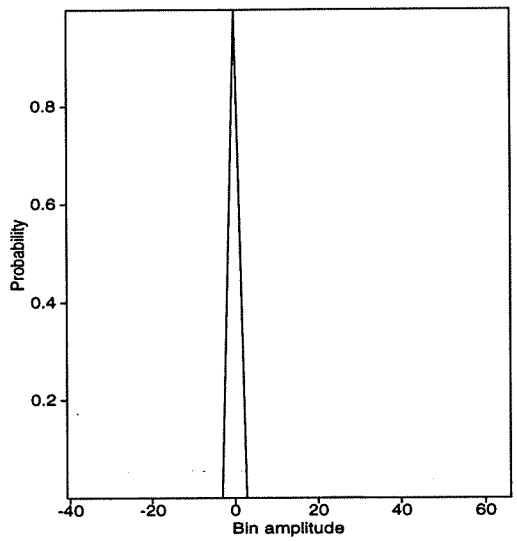

FIG. 3.9. Estimated signal probability density function.

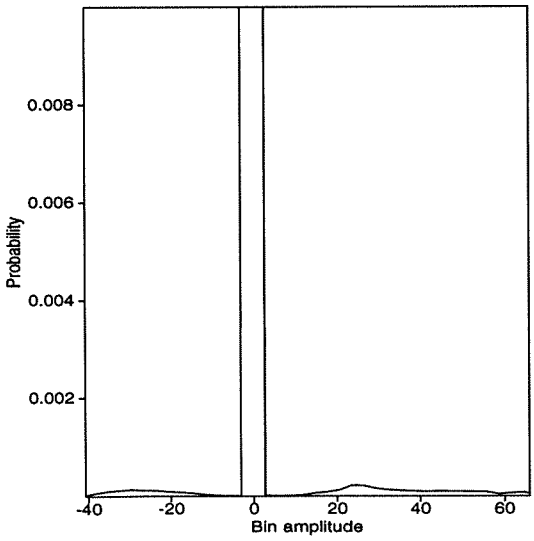

FIG. 3.10. Close-up of the low-probability region of the estimated signal probability density function showing that some samples indeed have high amplitudes.

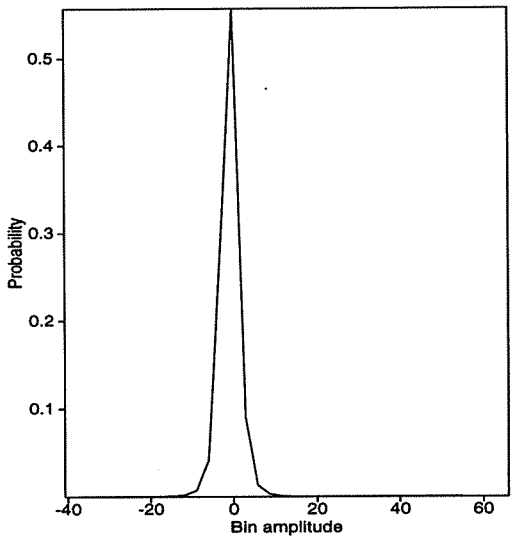

FIG. 3.11. Estimated data probability density function computed as $p_s * p_n$.

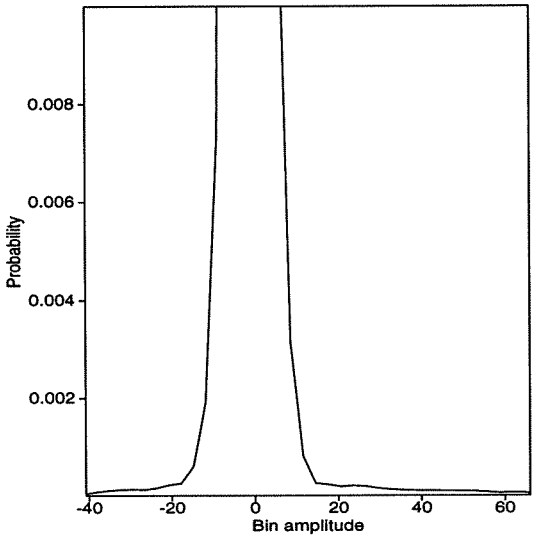

FIG. 3.12. Close-up of the low-probability region of the data probability function computed as $p_n(x) * p_s(x)$. Note that we recovered the samples with the high amplitudes.

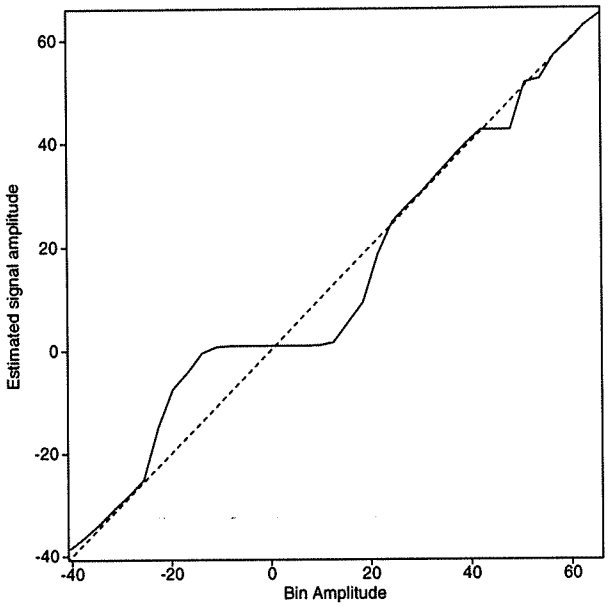

FIG. 3.13. Bayesian estimator, showing that the signal and the data amplitudes are about the same, i.e., close to the dashed line with unit amplitude through the origin, except for the very small amplitudes for which the estimated signal is nearly zero.

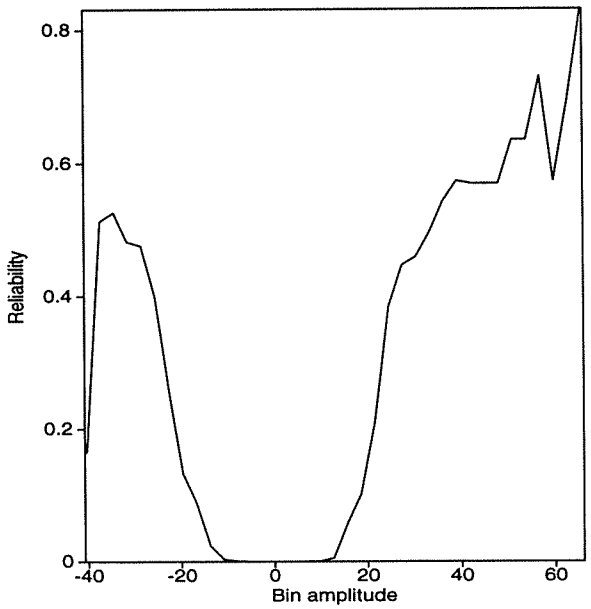

FIG. 3.14. Reliability indicator, showing that the high amplitudes are very reliable, whereas the lowest ones are not.

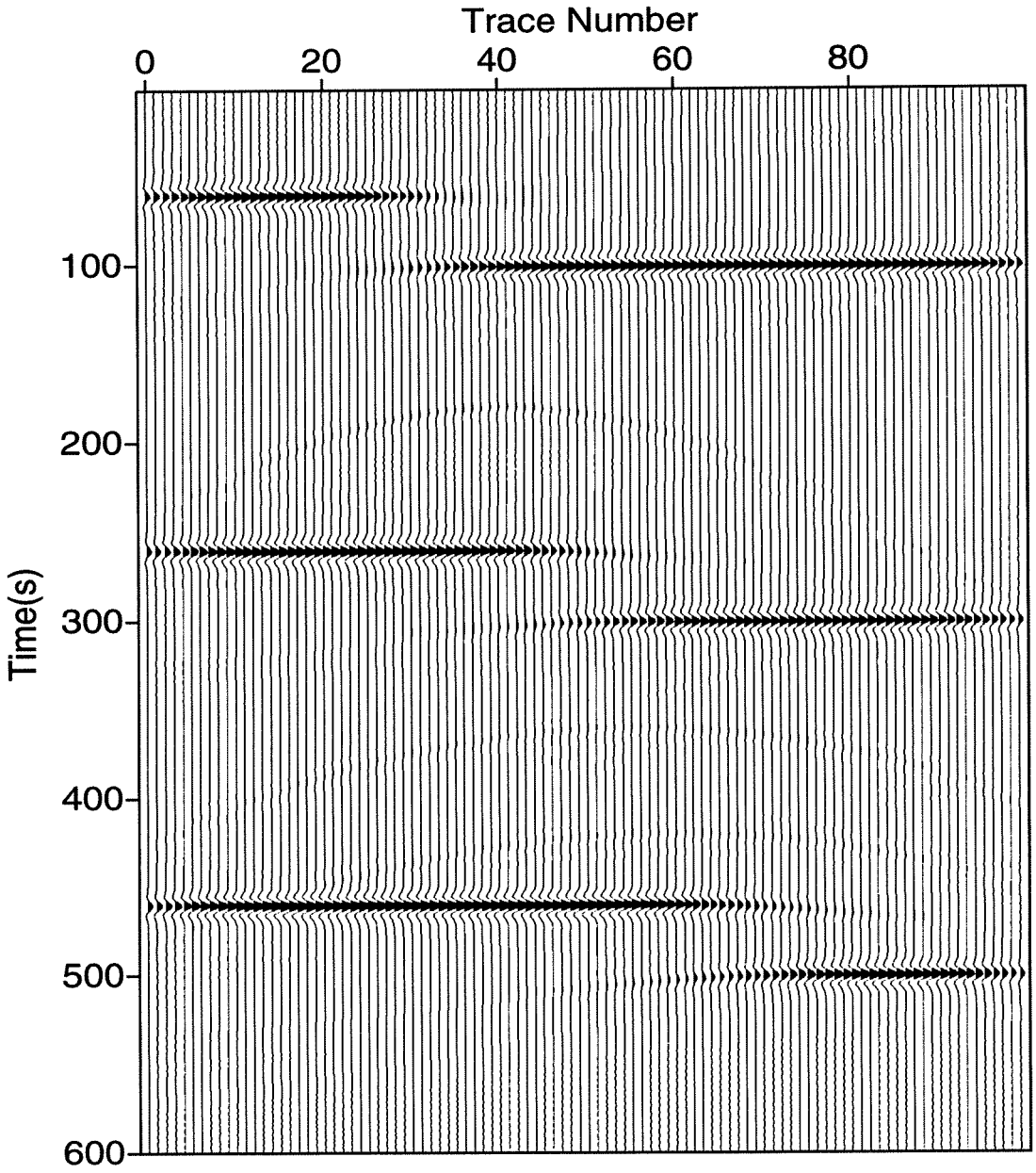

FIG. 3.15. Extracted signal. Some energy from the diffractions remains.

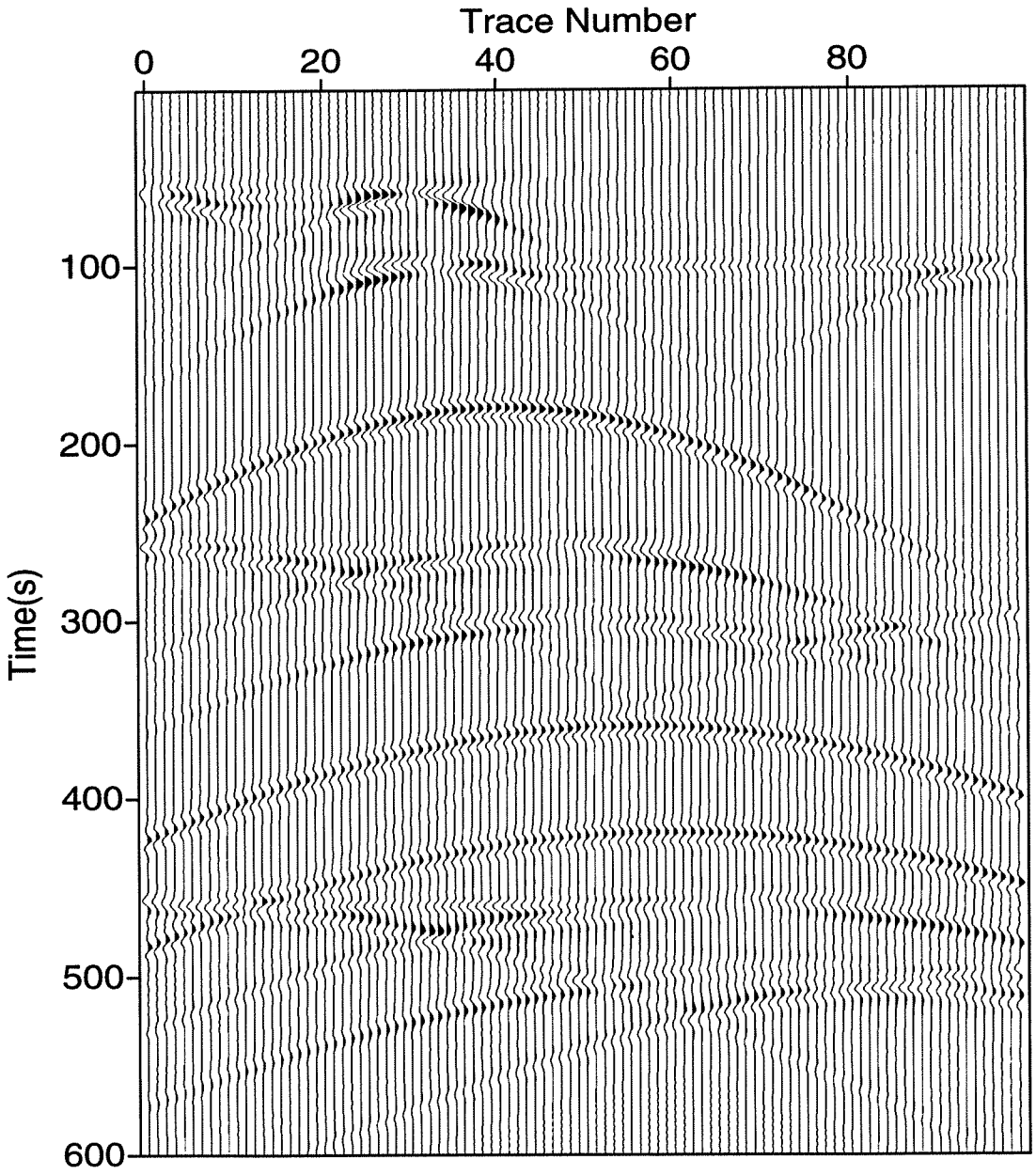

FIG. 3.16. Extracted noise. Some energy from the reflections remains. Amplification by a factor of 4.0 was applied for display.

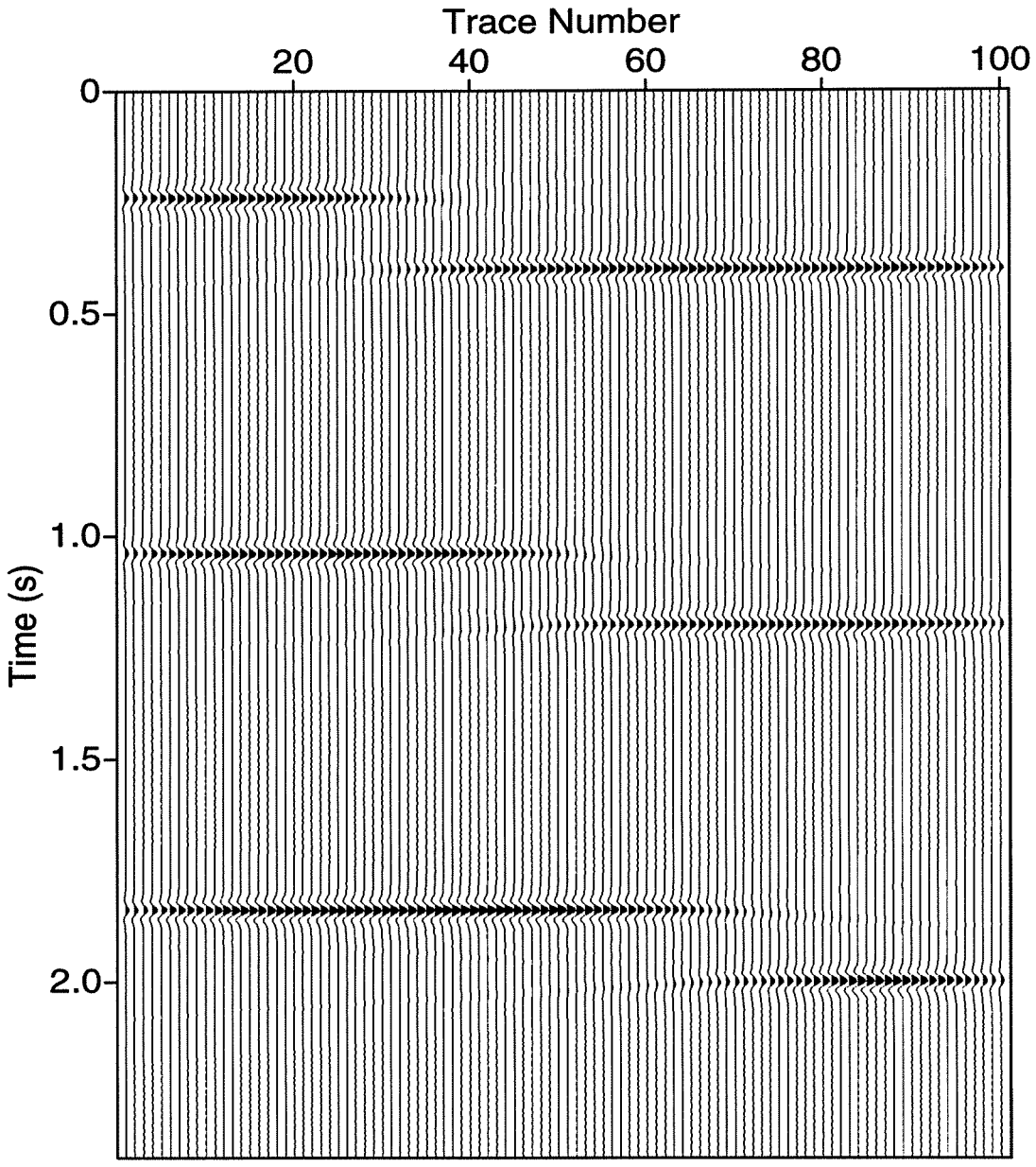

FIG. 3.17. Extracted signal, second iteration. Note that very little energy from the diffractions was extracted this time.

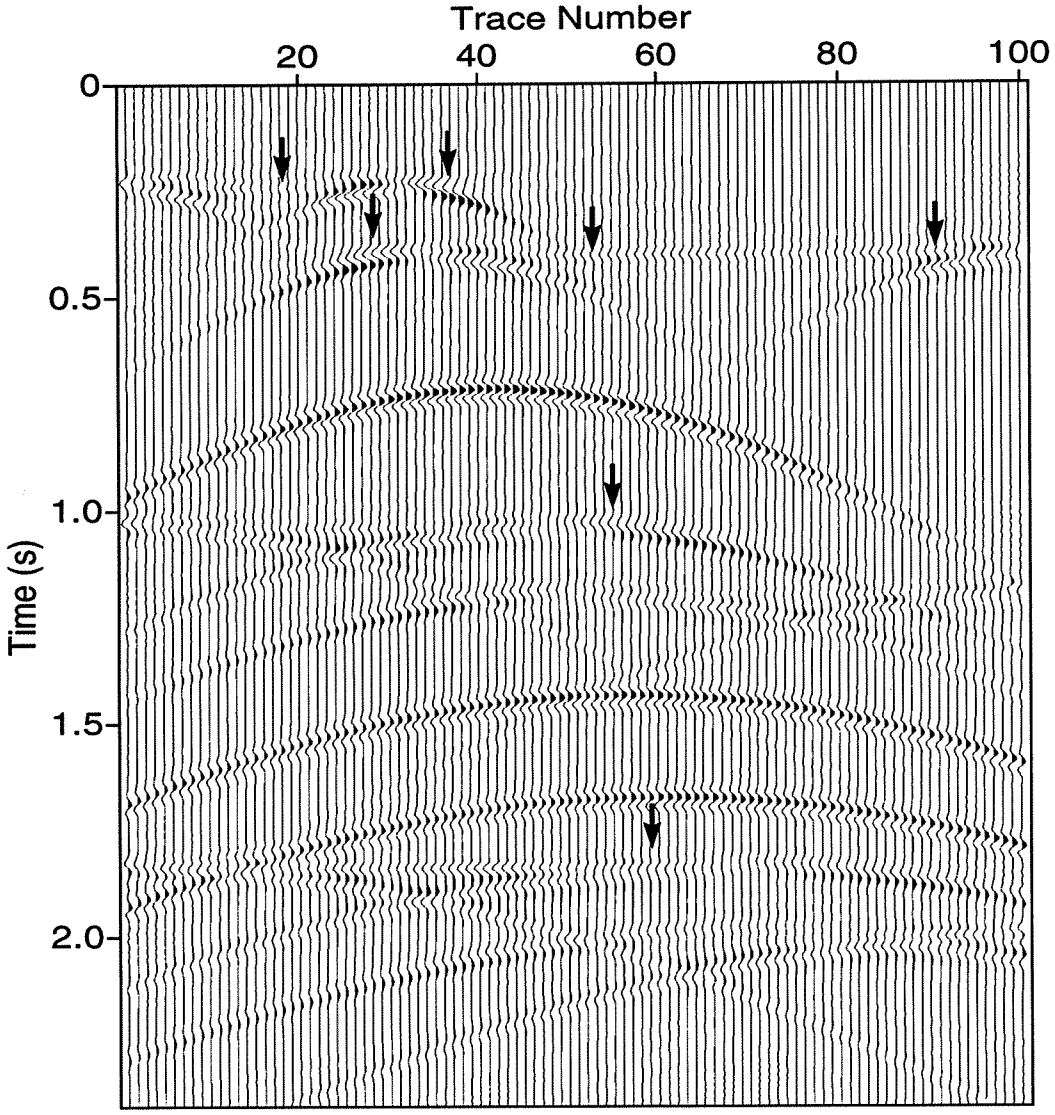

FIG. 3.18. Extracted noise, second iteration. The result has improved relative to that of the first iteration (Figure 3.16), in particular for the shallow events (less than 500 ms). Arrows point to regions where differences can be found. Amplification by a factor of 4.0 was applied for display.

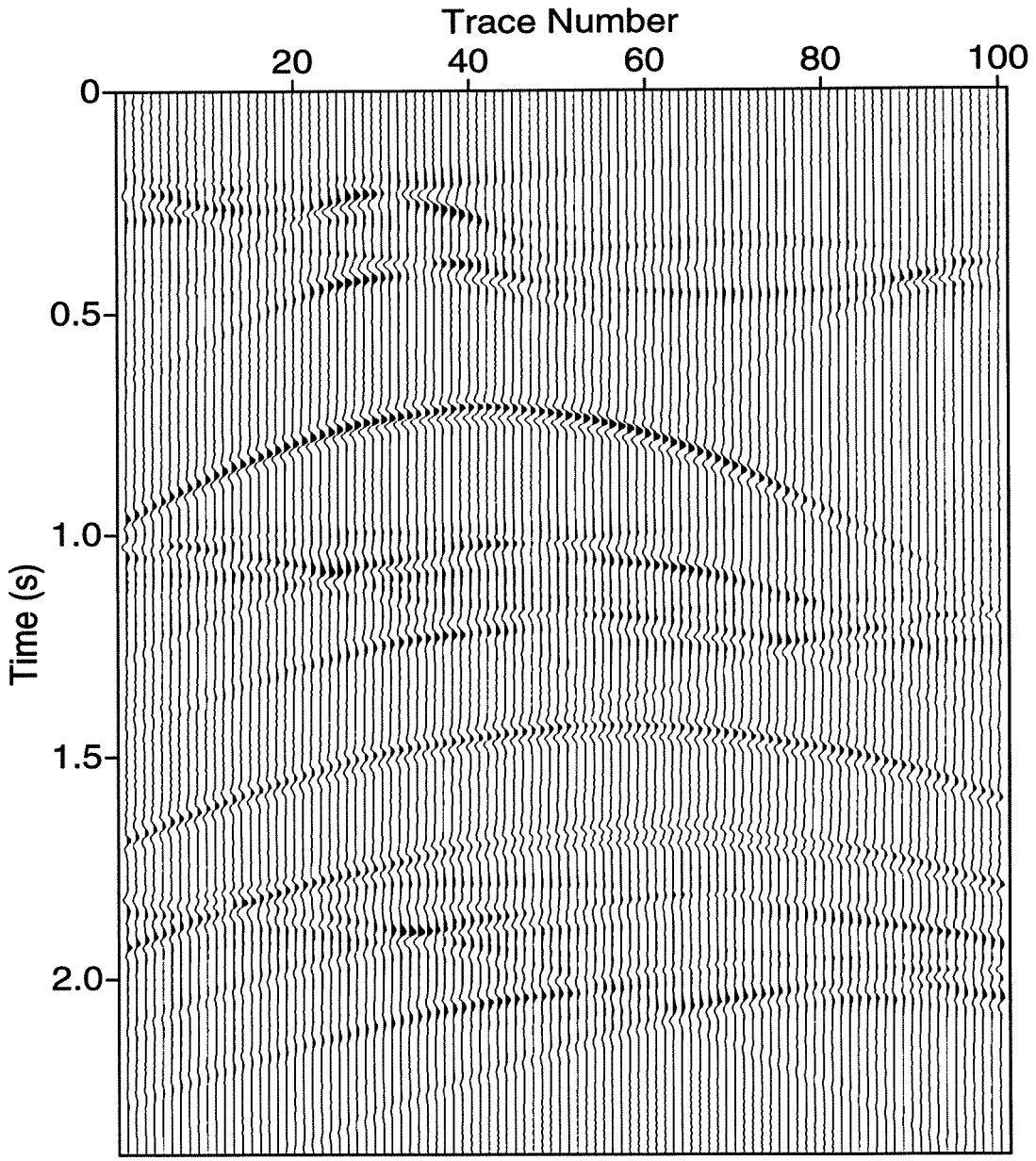

FIG. 3.19. Extracted noise, with no smoothing applied during the extraction. Note the processing noise that has been superimposed on the diffractions. Amplification by a factor of 4.0 was applied for display.

Chapter 4

GROUND-ROLL SUPPRESSION

4.1 Introduction

I now explore the application of the signal-noise separation algorithm described in the previous chapter to the suppression of ground-roll noise. Ground-roll is a high-amplitude, coherent, low-frequency, dispersive wavetrain that travels near the surface of the earth and is a serious problem with most land datasets. First, I review traditionally-used techniques for its suppression and then describe the proposed method, and show results of applying it to a field shot record. Figure 4.1 shows this shot record; it consists of 44 traces, with 1000 samples per trace at 4-ms sample interval and 50-m trace interval.

4.2 Traditional techniques

Traditional techniques for ground-roll suppression have relied on using geophone arrays on the field, and using a combination of F-K filters, low-cut frequency filters and deconvolution during processing. Each is simple to apply, but while these various approaches can be effective, where the ground-roll amplitude is particularly strong and not sufficiently different in frequency or wavenumber content from the signal, the resulting suppression can be inadequate. In particular, the noise suppression can distort the variation of signal amplitude with offset on unstacked traces compromising interpretation of amplitude variations with offset. Therefore, where the quality of extracted signal is particularly important, such approaches may be inadequate. Before applying the method of this thesis, let us examine some basic characteristics of these methods by applying them to the field data set.

4.2.1 Low-cut frequency filtering

Ground-roll noise is often distinctly low frequency compared with the body-wave energy that we wish to preserve. Where the amplitude spectra of the two do not overlap, a low-cut frequency filter can be applied to suppress the ground-roll frequency band. The shot record in Figure 4.1 shows strong, dispersive and spatially-aliased ground-roll, and Figure 4.2 shows the amplitude spectra of all the traces in this shot record. For frequencies below about 16 Hz, the data are swamped by the noise. Figure 4.3 shows the corresponding amplitude spectra after application of a low-cut trapezoidal frequency filter with a low-cut frequency of 16 Hz and a low-pass frequency of 20 Hz. The filter was applied over the entire shot record and not just in the region of the noise.

Figure 4.4 shows the extracted signal, and Figure 4.5 the extracted noise. The low-cut filter did indeed suppress much of the ground-roll, but, as seen in Figure 4.5 the low-frequency

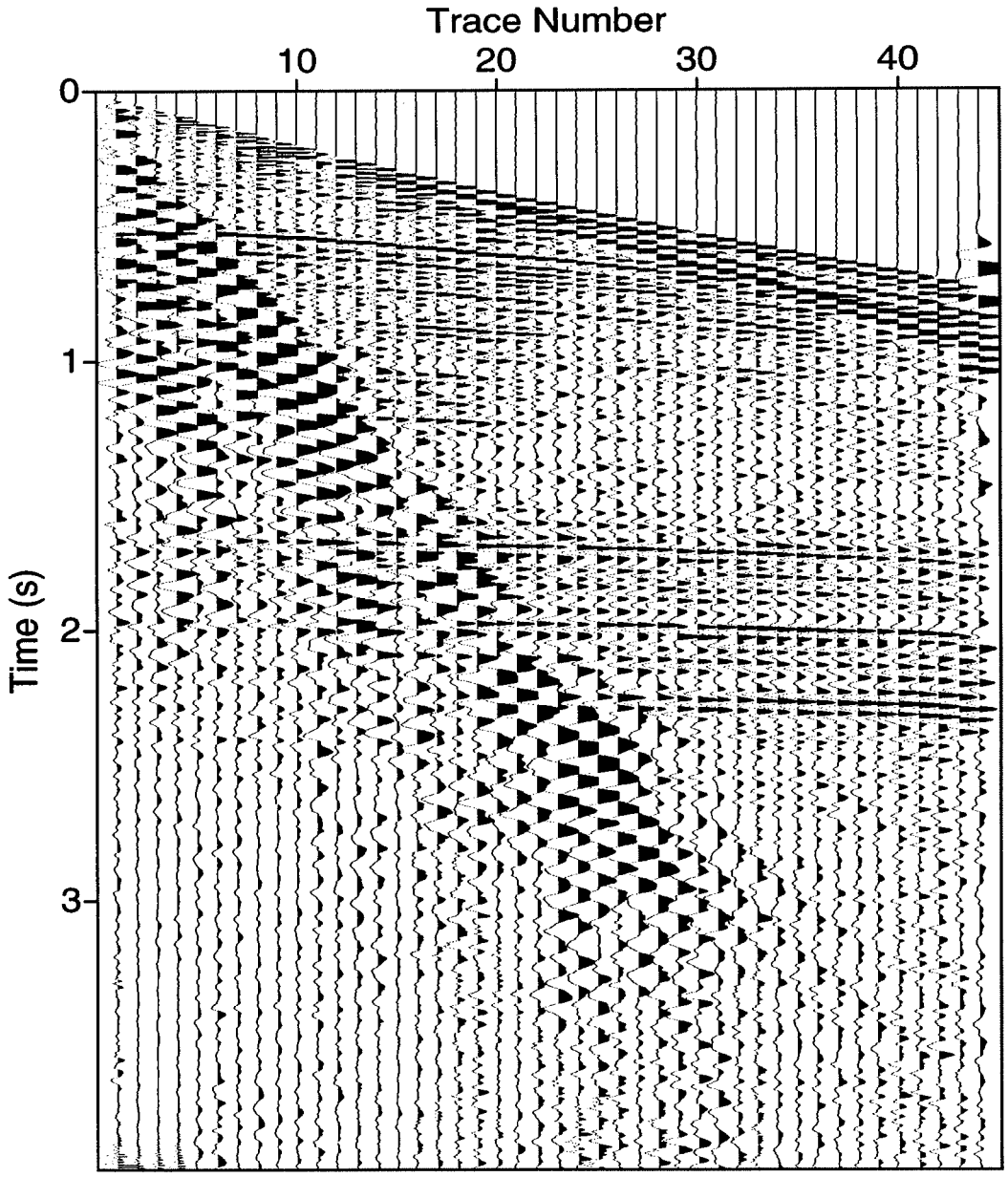

FIG. 4.1. Field shot record with strong, spatially-aliased, low-frequency ground-roll.

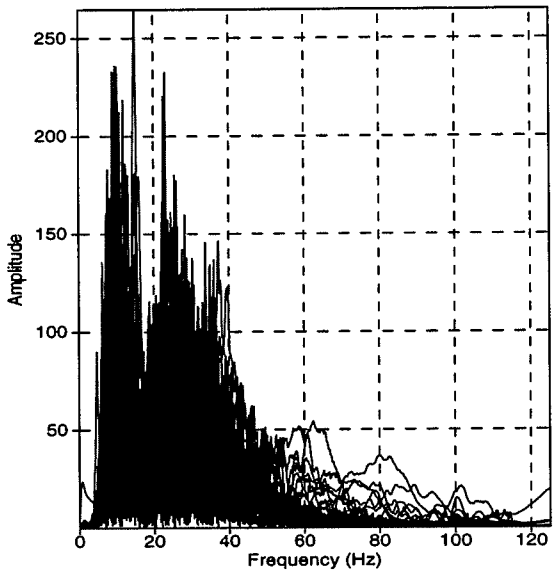

FIG. 4.2. Amplitude spectrum of the test shot record.

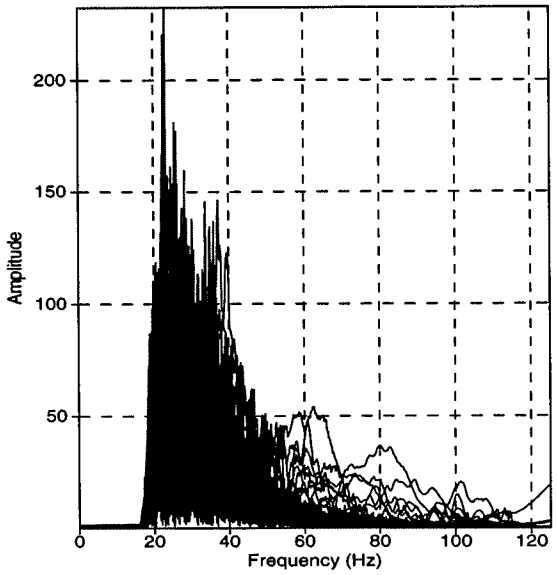

FIG. 4.3. Amplitude spectrum of low-cut filtered shot record with low-cut frequency of 16 Hz and low-pass frequency of 20 Hz.

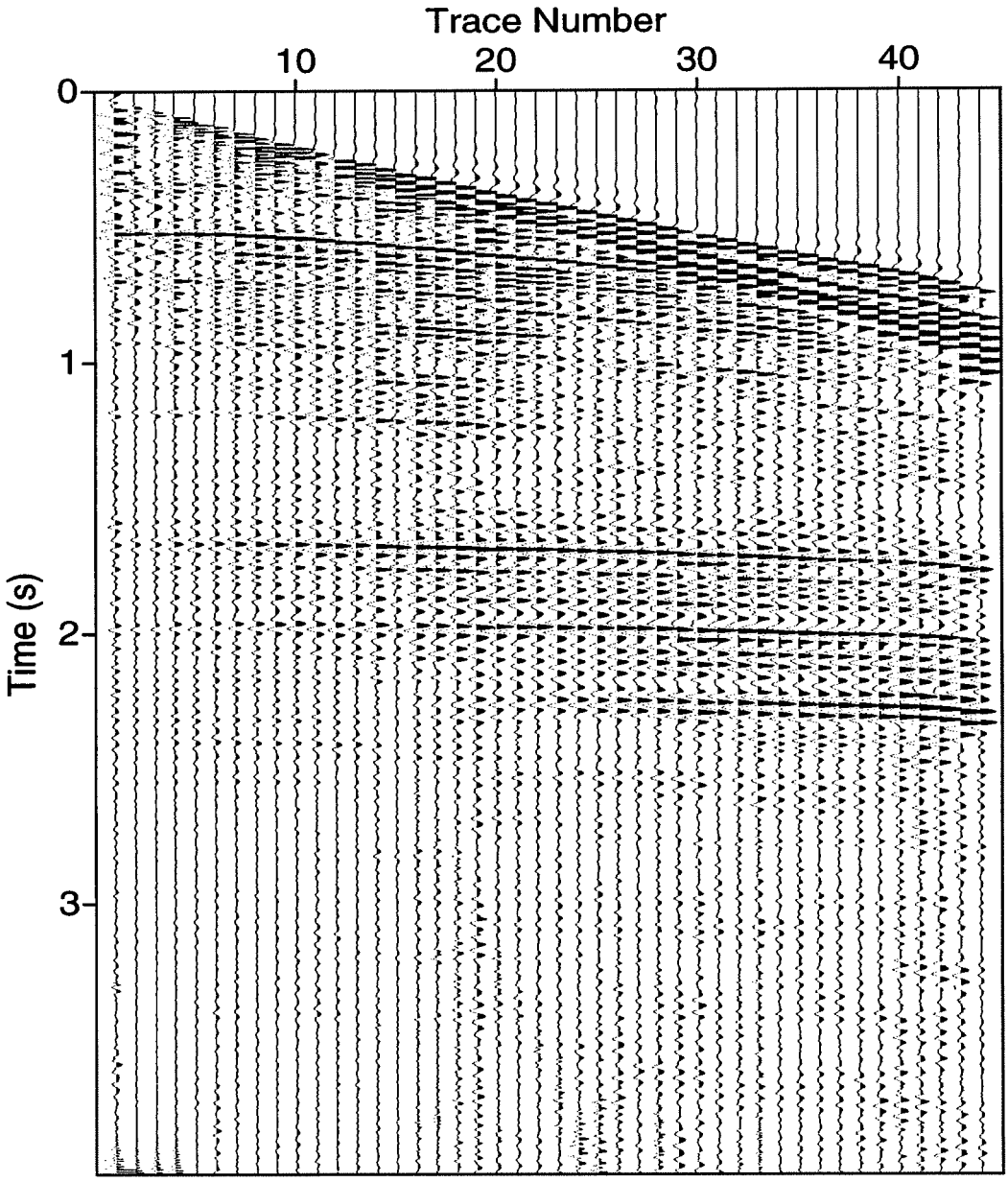

FIG. 4.4. Low-cut filtered test shot record. Note that all the ground-roll is gone.

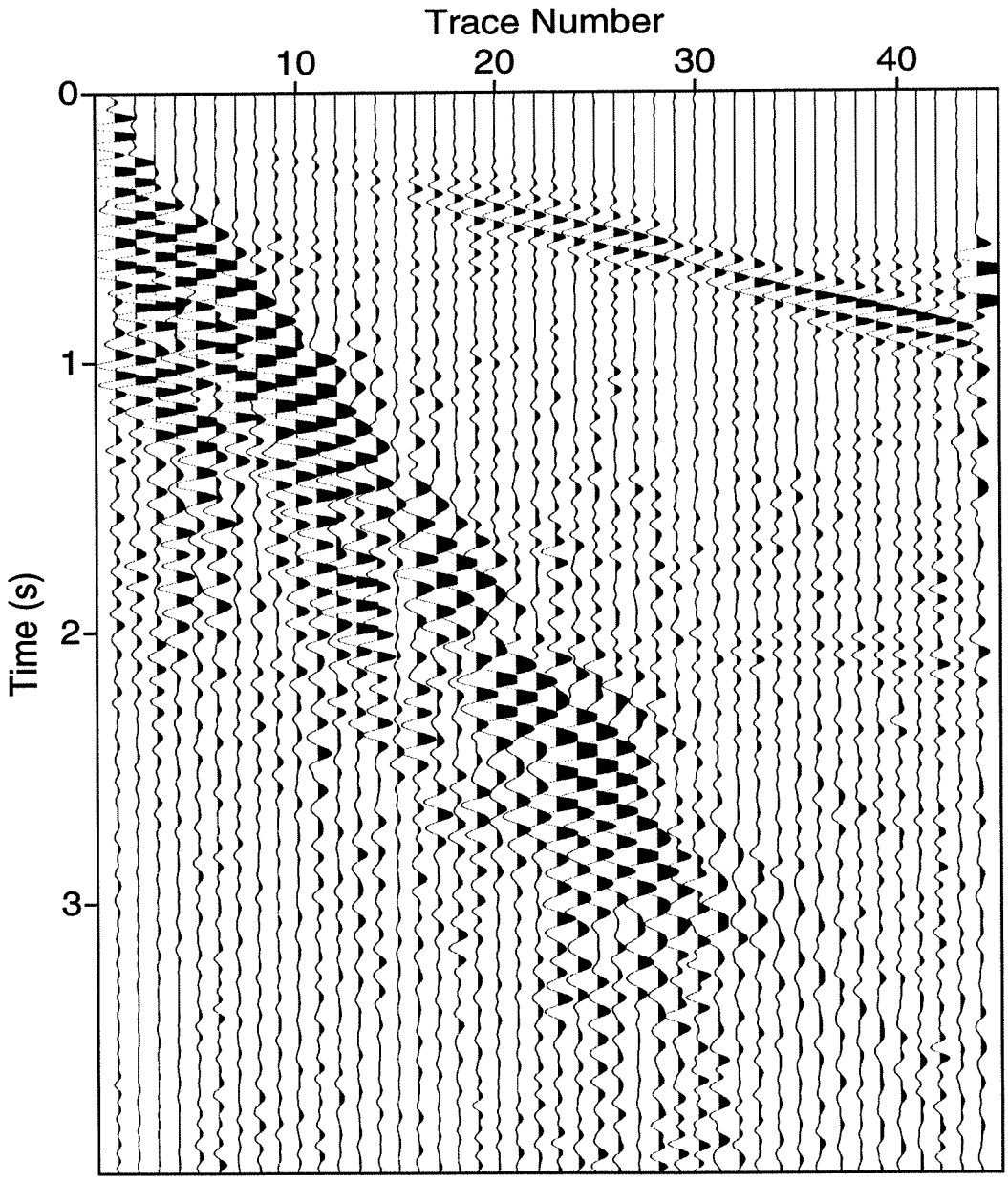

FIG. 4.5. Extracted ground-roll after applying a low-cut filter.

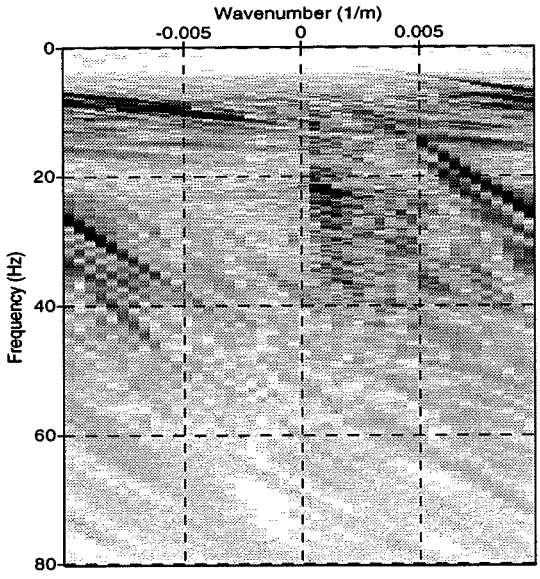

FIG. 4.6. F-K spectrum of test shot record. The first arrivals and the noise are spatially aliased.

component of the signal was suppressed along with it. While this may be acceptable for a structural interpretation, it is not if a detailed stratigraphic interpretation is desired since the low-frequency components contribute to the resolution of the resulting wavelet. The limitation is important when the data are to be used for lithological inversion, since the low-frequency components of the data carry useful information about geology such as compaction trends or deposition cyclicity. Sometimes the low-cut frequency filter is applied not to the entire shot record but only to the "noise cone", or the zone in which the ground-roll is present. This has the advantage that only the data inside that cone is filtered, but has a serious problem in that the amplitude of the wavelet is changed in an offset-dependent way, thereby precluding the utilization of any AVO analysis.

4.2.2 F-K filtering

In this method, the data are transformed to the F-K domain by the application of a 2-D Fourier transform and then a slope filter, aimed at removing any feature within a prescribed range of slopes, is used to separate the signal from the noise.

Since the ground-roll waves travel along the spread with much lower velocity than do the body waves, they are mapped to a different part of the F-K domain than is the signal. Figure 4.6 shows the F-K transform of the data in Figure 4.1. Note that the ground-roll (the linear events with frequencies less than about 16 Hz) is aliased and so are the first breaks, (the strong amplitude linear events for higher frequencies). A fan shaped slope filter might then be applied to separate the signal from the noise.

In practice two problems frequently conspire against the success of the algorithm.

Aliasing: Spatial aliasing of either the signal or the noise, due to insufficient sampling in space (i. e., fewer than two samples per wavelength for the shortest wavelengths present in the data), makes events appear to have moveout that can differ greatly from the true moveout, causing events to appear in the wrong part of the transform domain. As a result, it may be difficult, if not impossible, to apply the desired mute in the F-K domain. Figure 4.6 should show no energy in its left hand-side half since all the events in the input data are dipping to the right, but it does because of the aliased events. Although there are ways to alleviate this problem, it is always detrimental to the quality of the F-K filtering action.

Tapers: If the boundary between the pass zone and the reject zone of moveouts is too sharp, moveout filters suffer from a two-dimensional version of the Gibb's phenomenon, familiar when the pass-reject boundary of one-dimensional frequency filters is too sharp. Therefore, a transition zone with a taper must be applied to go gently from the slopes that are to be fully rejected to those that are perfectly accepted. Since normally the moveouts of signal and noise are not clearly separated, a tradeoff is necessary: short transition zones mean more effective rejection but the possibility of introducing Gibb's phenomenon noise with slopes comparable to the cutoff slopes of the moveout filter. Broader taper zones, on the other hand, mean a less severe Gibb's phenomenon noise, but also less perfect noise rejection.

Figure 4.7 shows the F-K spectrum after the application of such a fan filter, with smallest reject slope of -0.218 ms/trace, (that is, energy with slopes less than that are multiplied by zero), smallest pass slope of 0.0 ms/trace (a linear taper is applied for slopes between -0.218 and 0 ms/trace), largest pass slope of 5.2 ms/trace (that is, energy between 0 and 5.2 ms/trace are multiplied by one) and largest reject slope of 6.2 ms/trace (that is, energy for slopes larger than this value are multiplied by zero; a linear taper is applied for slopes between 5.2 and 6.2 ms/trace). Figure 4.8 shows the extracted signal. Some ground-roll still remains, in particular between 700 and 1200 ms. F-K filtering, while useful for suppressing ground-roll, is thus imperfect in separating the signal and noise. Figure 4.9 shows the extracted ground-roll. Note that no discernible energy from the primaries was extracted.

4.3 Signal-noise separation method

In order to apply the signal-noise separation algorithm described in the previous chapter, we need to design an invertible linear transformation that will focus the signal and, at the same time, defocus the noise. If we consider the signal to be the reflection hyperbolas, and the noise to be either random noise or ground-roll, then our transform must be able to focus hyperbolas while defocusing non-hyperbolic events. Let us now compare the action of several alternative approaches to focusing the hyperbolas.

4.3.1 Local Slant Stacks

As mentioned above, Harlan et al. (1984) proposed the local slant stack method for extracting the ground-roll noise present in a shot record by approximating the reflection hyperbolas with a series of small, linear segments. Recall that the idea is that a hyperbola or any smooth curve, can be locally approximated with a short linear segment. For each input trace, the energy in a window of a few traces centered on that trace will be locally slant stacked so that a 2-D τ-p representation is obtained for the chosen range of slopes in the transform. Events that are linear (or can be reasonably approximated as linear) over the window of data used in the transform are focused, whereas those that do not show that coherence are not focused. The window moves up by one trace and the process is repeated for all traces in the shot record.

The algorithm described in the previous chapter can be applied to each of these τ-p planes so that the linear segments can be extracted. To the extent that the number of traces for the local slant stacks is appropriate (not so many that the linear approximation breaks down, and not so few that τ-p transform is not reliable) and that an appropriate range of slopes for the τ-p transform is chosen, the signal can be retrieved from the stronger noise. Figure 4.10 shows a window of data with 21 traces centered on trace 22. Note that some linear reflections are clearly visible between 400 and 600 ms (in reality, these are segments of hyperbolas that look linear because of the small range of offsets). Figure 4.11 shows the result of applying the local slant stack to that window of data. Again, the linear and nearly linear events are focused near zero moveout, with strongest amplitudes between 400 and 600 ms.

While the linear and nearly linear hyperbolic events were focused, as desired, the focusing is weak; that is, the energy in the linear events is not collapsed to a point because of the small number of traces in the transform, which make nonlinear events show spurious coherency. Nevertheless, the focusing is sufficient to get a good separation from the ground-roll, as shown in Figure 4.12. The ground-roll has been largely suppressed, except for a remnant at about 800 ms. Moreover, no significant coherent energy has been extracted as noise, as shown in Figure 4.13. These figures should be compared with their counterparts in Figures 4.4 and 4.5 for the low-cut filter and Figures 4.8 and 4.9 for the F-K filtering. The ground roll has been better suppressed and no significant primary energy has been lost.

4.3.2 Hyperbolic slant stack

The local slant stack approach is effective in separating the signal from the ground-roll but it is also expensive because the local slant stacks themselves are and because the signal-noise separation algorithm has to be applied to each window of data. An alternative to the method, as described previously, is to consider focusing the hyperbolic events not as a series of linear events, but directly as the hyperbolas that they are; after all, we know the equations of those hyperbolas. Although this can be done (Thorson and Claerbout, 1985), the hyperbolas are not time invariant (they tend to be flatter for deeper events). Therefore, the efficient F-X Radon implementation described in Chapter 2, cannot be used. One could use a time-domain implementation, but in practice the code is highly inefficient. Also, if the

reflections are from dipping reflectors, then the apexes of the hyperbolas will be shifted from the zero-offset trace, making the flattening of these events a more complicated endeavor. This latter issue can be addressed by sorting the data into CMP gathers, so that reflections are approximately hyperbolic, with apexes at the origin. After the sorting, the moveout of the ground-roll remains sufficiently different that it will not be focused by the τ-p transform.

4.3.3 Parabolic slant stack

To overcome the problem of the non-stationarity of the NMO hyperbolas, I propose to apply an NMO correction to the CMP gathers, aimed at flattening the hyperbolas so that they can be approximated by time-independent parabolas (if the NMO correction were perfect, they could even be approximated with straight lines as before). Hampson (1989) proposed this idea for suppressing multiples, and Foster and Mosher (1992) extended it to use time-independent hyperbolas in an effort to better approximate undercorrected or overcorrected hyperbolas. I will apply the algorithm here to a shot record because the geology is sub-horizontal so that the reflections are approximately symmetric, as they would be in CMP gathers. Figure 4.14 shows the NMO-corrected data. In general, all the reflections have been well corrected while the noise continues to exhibit large moveout. Next, we transform the data with a parabolic τ-p transform (the transform could be linear if the NMO-correction was perfect) and use the signal-noise separation algorithm to identify and extract the focused energy, that is, the NMO-corrected hyperbolas, as we did in Chapter 3. The difference between the extracted signal energy and the original data represents noise, which consists not only of ground-roll, but any events that were not focused by the transformation, such as head waves and random noise. The extracted signal can then be inverse NMO-corrected, and other processing steps applied as usual.

Figure 4.15 shows the parabolic τ-p transform of the shot record. The energy from the reflections has been mapped to near zero slope, and background noise has been generated from spurious alignments (spatial aliasing) of either the ground-roll or the signal. The focusing of the energy is not very sharp because the p-sampling interval is small (in order that $N_p = N_x$ and to avoid p-aliasing).

Figure 4.16 shows the signal extracted with this modified signal-noise algorithm using the parabolic transform. Note that no trace of the ground-roll is left. Figure 4.18 shows the noise that was removed, not only the ground-roll but also high-frequency incoherent noise. Very little signal has been removed as noise. Figure 4.17 shows the preserved signal after an inverse NMO has been applied.

Comparing Figure 4.17 with the input data in Figure 4.1 shows that not only has the ground-roll been completely suppressed, but no apparent distortion has been introduced in the wavelets.

Filtering of ground-roll in the τ-p domain was mentioned in Chapter 1; recall that it consists of directly filtering the ground-roll in the τ-p domain via the application of a suitable tapered mute to remove the noise energy. In Figure 4.15 the dashed vertical line represents the maximum p-value fully preserved in the data and the dashed slanted line the minimum p-value totally rejected as noise. A linear taper is applied between the two. Here, the rejected

noise is not the ground-roll since its p-values are outside the range of the transform, and so was not modeled by it. It represents noise associated with spurious alignments and noise present in the data, as mentioned before. Figure 4.20 shows the signal extracted by this method; no trace of the ground-roll is present. Figure 4.19 shows the suppressed noise. Note that although the method was successful in removing all the ground-roll, some signal was removed along with it. This is clear in particular between 1.6 and about 2.6 s.

4.4 Practical Aspects

Once the idea behind the discrete generalized Radon transform and the signal- noise separation algorithm are understood, it is relatively simple to apply the method for ground-roll extraction, as described in the previous section. A few details, however, must be taken into account in order to get the best results.

Linear versus parabolic τ-p transform: It was previously mentioned that if the NMO correction is perfect, we can apply a linear τ-p transform to the data, but since we are in general unlikely to get a perfect correction, a parabolic τ-p is a better option.

Range of slopes: Limiting the range of slopes of the τ-p transform (whether linear, parabolic or whatever), is the simplest way to apply a τ-p domain filtering since those events with slopes outside the range of the transform slopes will not be modeled by the transformation. This is, however, a somewhat crude approach to the problem because if spurious alignments occur, the transform will model and extract them as signal. With the signal-noise separation algorithm, we can in principle choose any reasonable range of slopes, since the method is not based on slope filtering. It is wise, nevertheless, to restrict the range of slopes used to those that will correctly describe the events that we consider signal and allow the use of a small sampling interval in slope, thereby avoiding the possibility of aliasing noise. In this sense, the modified Harlan's method may be regarded as a hybrid of Harlan's signal-noise separation algorithm (based on zeroing out amplitudes below a statistically derived threshold value) and Hampson's τ-p filtering (based on muting of regions deemed to be noise). It is also advisable to use a number of slopes that is less than or equal to the number of traces because that guarantees the best performance of the F-X Radon transform, as was mentioned in Chapter 2.

Preprocessing: Since the method strongly depends on amplitude histograms derived from the data, it is important that all the traces be brought to about the same amplitude before applying the method. Traces with abnormally weak amplitude level, due perhaps to poor geophone coupling or any other reason, should be brought to the same amplitude level of the others, so that no amplitude variations occur that are not related to the signal and noise content of the data. This can be done by applying some trace balancing. This, however, should be done with care (making sure that the balancing is not systematic with offset), if AVO analysis is expected to be performed with the filtered data. A geometrical spreading correction is also convenient to bring up the late energy.

4.5 Comparison of Results

Having compared the action of the five methods presented in this chapter (low-cut frequency filtering, FK moveout filtering, signal-noise separation with local slant stacks, signal-noise separation of NMO-corrected CMP gathers, and τ-p moveout filtering) for the suppression of ground-roll in the test field record, I now summarize a comparison of their relative performances on the basis of three criteria: level of ground-roll suppression, level of signal preservation and cost.

The comparison of the performance of these methods in terms of signal preservation is based on the results shown in Figure 4.4 for the low-cut frequency filtering, Figure 4.8 for the FK moveout filtering, Figure 4.12 for the signal-noise separation algorithm using local slant stacks, Figure 4.16 for the signal-noise separation algorithm applied to NMO-corrected CMP gathers, and Figure 4.20 for Hampson's τ-p filtering method.

In terms of their performance to noise rejection, the comparison involves Figure 4.5 for the low-cut frequency filtering, Figure 4.9 for the FK moveout filtering, Figure 4.18 for the signal-noise separation algorithm using slant stacks, Figure 4.13 for the signal-noise separation algorithm for NMO-corrected CMP gathers, and Figure 4.19 for Hampson's τ-p filtering method.

The results in all these figures are graded good, fair or poor on the basis of what an ideal hypothetical method would do, and are tabulated in Table 4.1. Cost is graded taking the fastest method (low-cut filtering) as a reference, and assigning the other costs relative to it. Both the τ-p filtering and the signal-noise separation (applied to NMO-corrected CMP gathers) are more expensive because they use the more costly F-X implementation. Cost could be reduced in these approaches if a t-x or an F-K implementation of the transform were used, but implementations in these domains have the unacceptable shortcomings of limited focusing capability and edge effects as was mentioned in Chapter 2. The signal-noise separation algorithm with many local slant stacks is particularly expensive because so many local slant stacks need to be computed. Again, cost could be reduced by using implementations of the transform other than that in the F-X domain, but at the expense of sacrificing accuracy of the results.

This table indicates that, on the basis of results for the tests on the field shot record in Figure 4.1, the signal-noise separation algorithm, with the modification introduced in this chapter, is perhaps the best option for ground-roll suppression if we are willing to pay the extra cost.

Table 4.1. Comparison between the different methods for ground-roll suppression presented in this chapter. The numbers in parenthesis on the last column represent rough computational cost relative to that of the low-cut frequency filtering method. True relative cost, of course, depends on details of software implementation and hardware used.

Method	**Signal preservation**	**Noise suppression**	**Cost**
Low-cut frequency filtering	Poor	Good	Good (1)
FK moveout filtering	Fair	Poor	Good (2)
Signal-noise separation (Local slant stacks)	Good	Fair	Poor (60)
Signal-noise separation (NMO-corrected CMP gathers)	Good	Good	Fair (6)
τ-p filtering	Fair	Good	Fair (4)

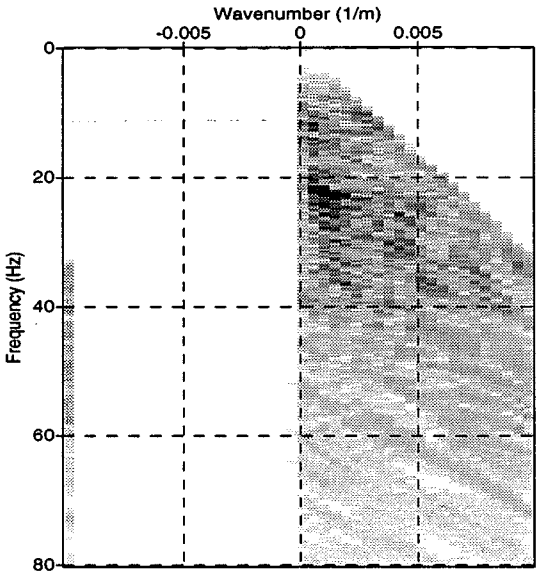

FIG. 4.7. F-K spectrum of the F-K filtered shot record. The filter slopes have been chosen so as to zero out the large-moveout noise region and the region of aliased noise.

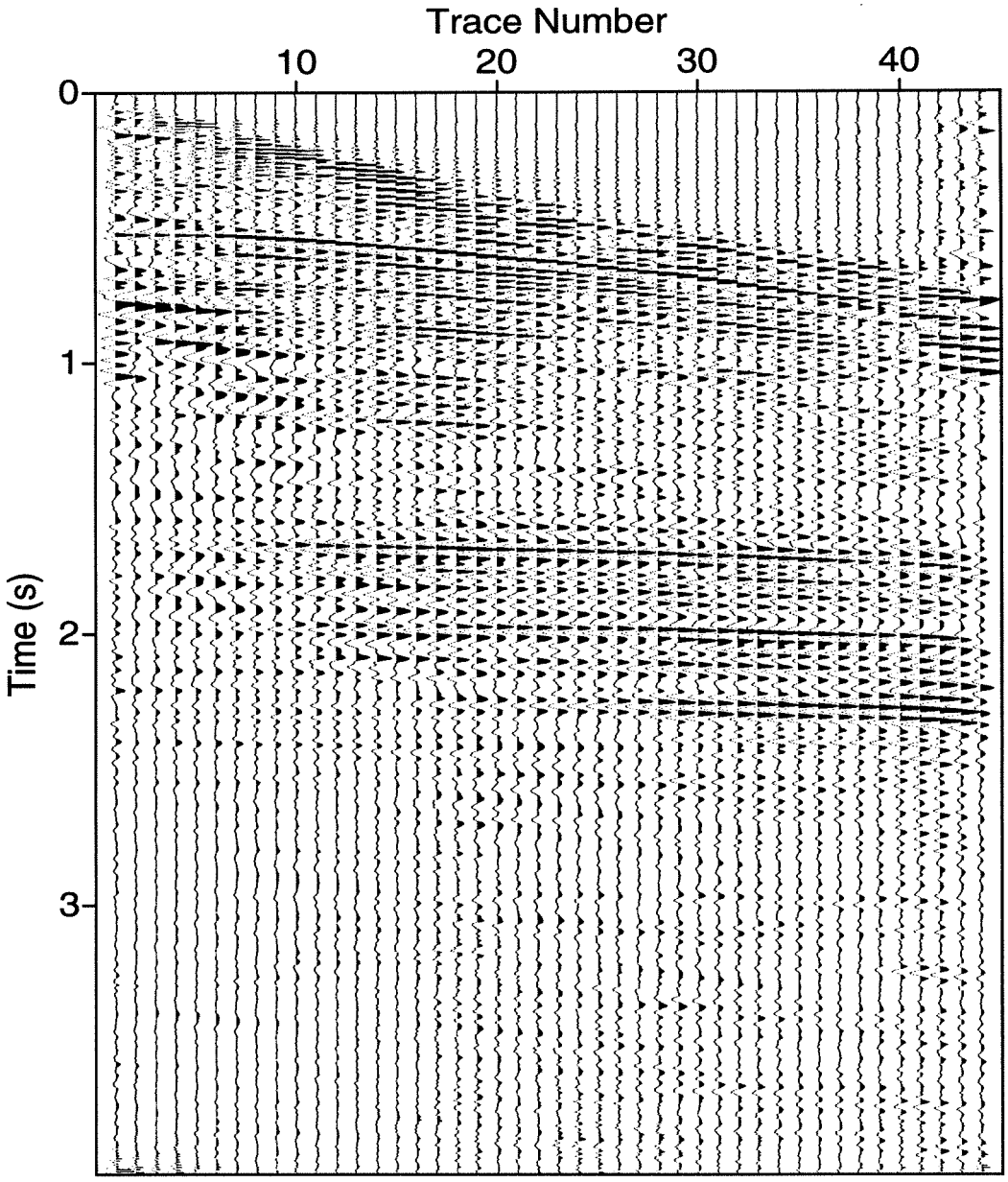

FIG. 4.8. F-K filtered shot record. Some ground-roll noise still remains.

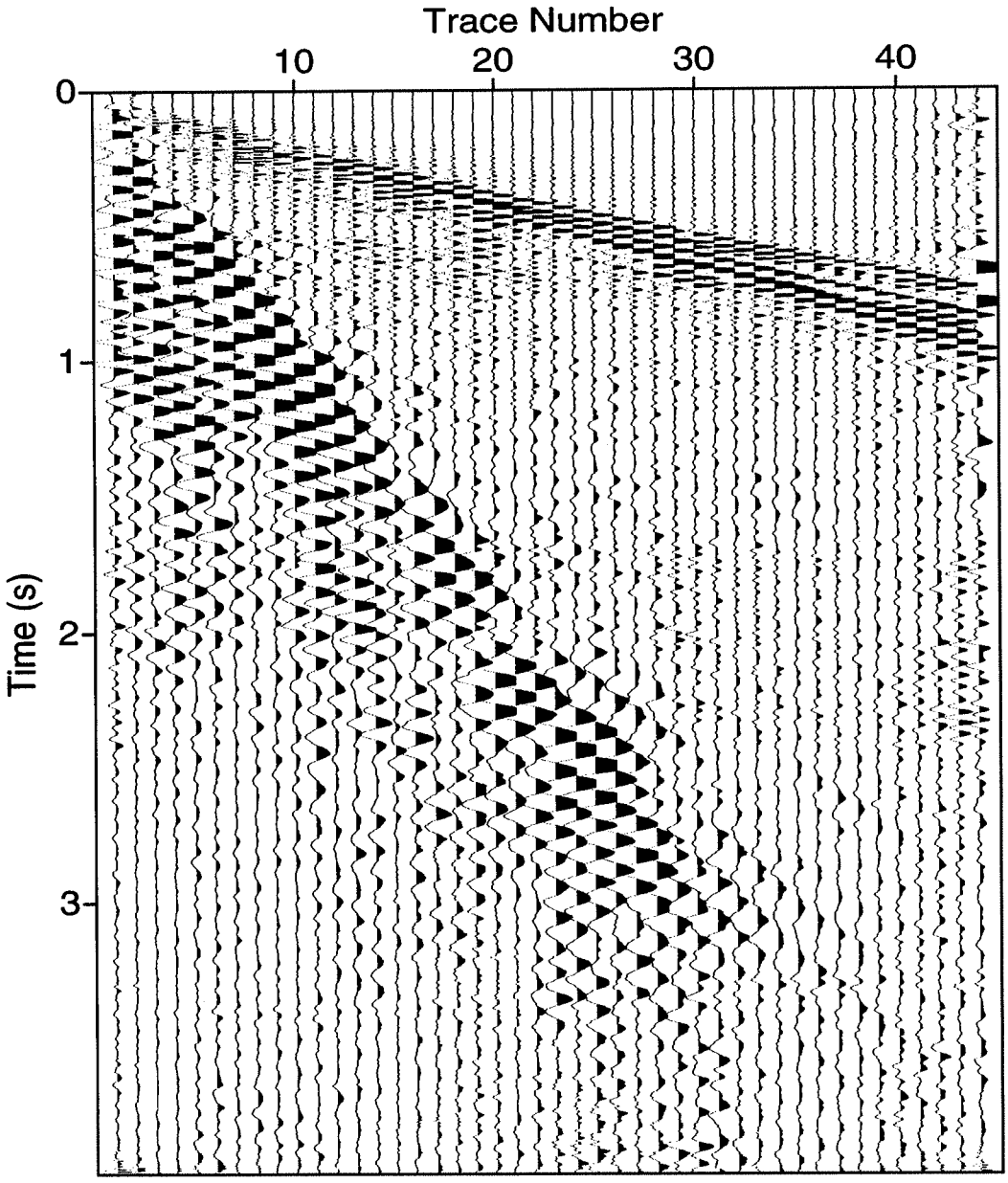

FIG. 4.9. F-K extracted ground-roll. Virtually no signal was extracted along with the noise.

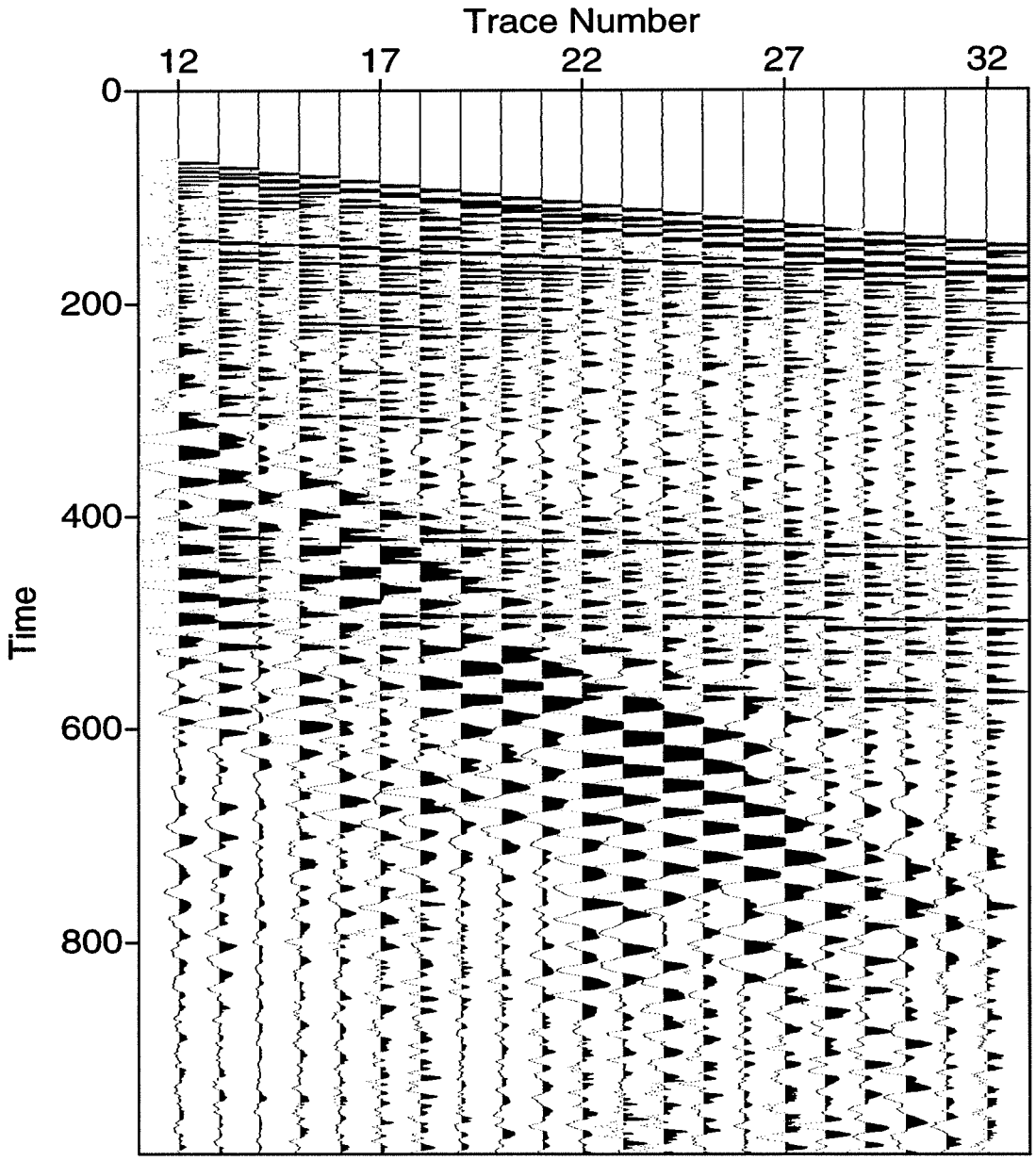

FIG. 4.10. A window of the shot record, 21 traces wide and centered at trace 22.

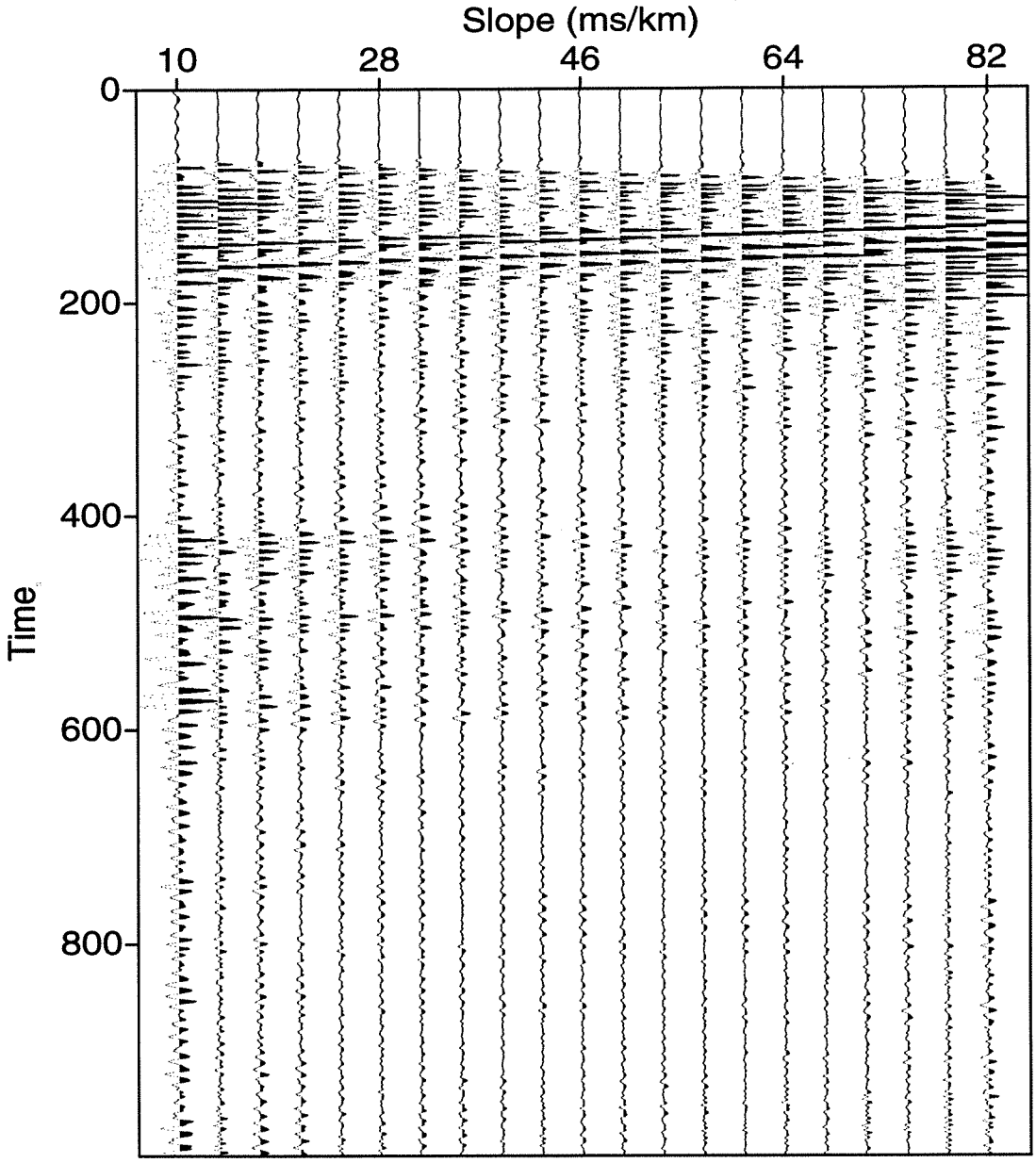

FIG. 4.11. Slant stack of the windowed data. The segments of hyperbolas, being almost linear, were focused by the transformation at times between 400 and 600 ms. Because of the small number of traces, this focusing is not very sharp.

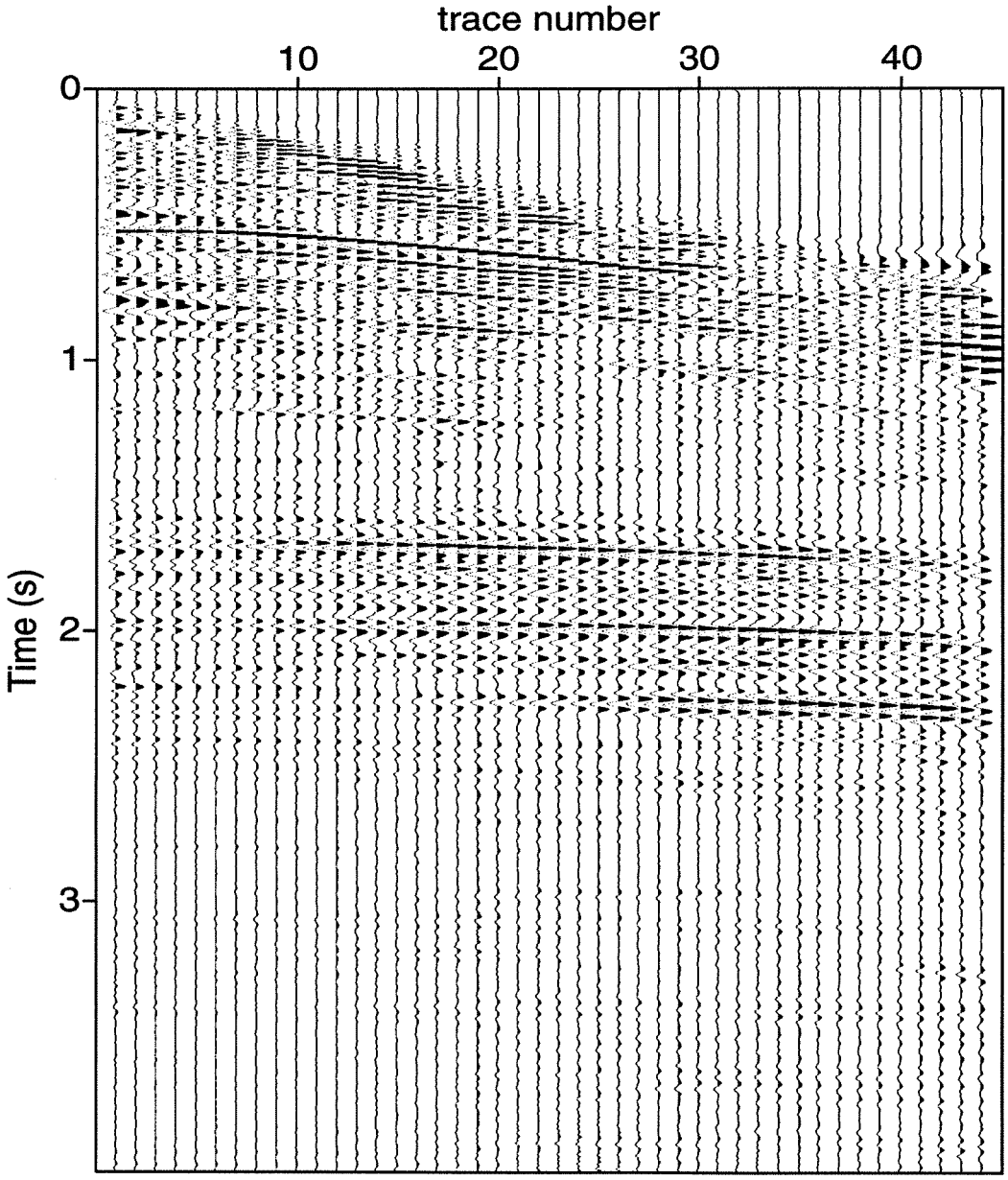

FIG. 4.12. Extracted signal using the local linear slant stack approach. Most of the ground-roll was extracted, but some remainder can be seen at about 800 ms.

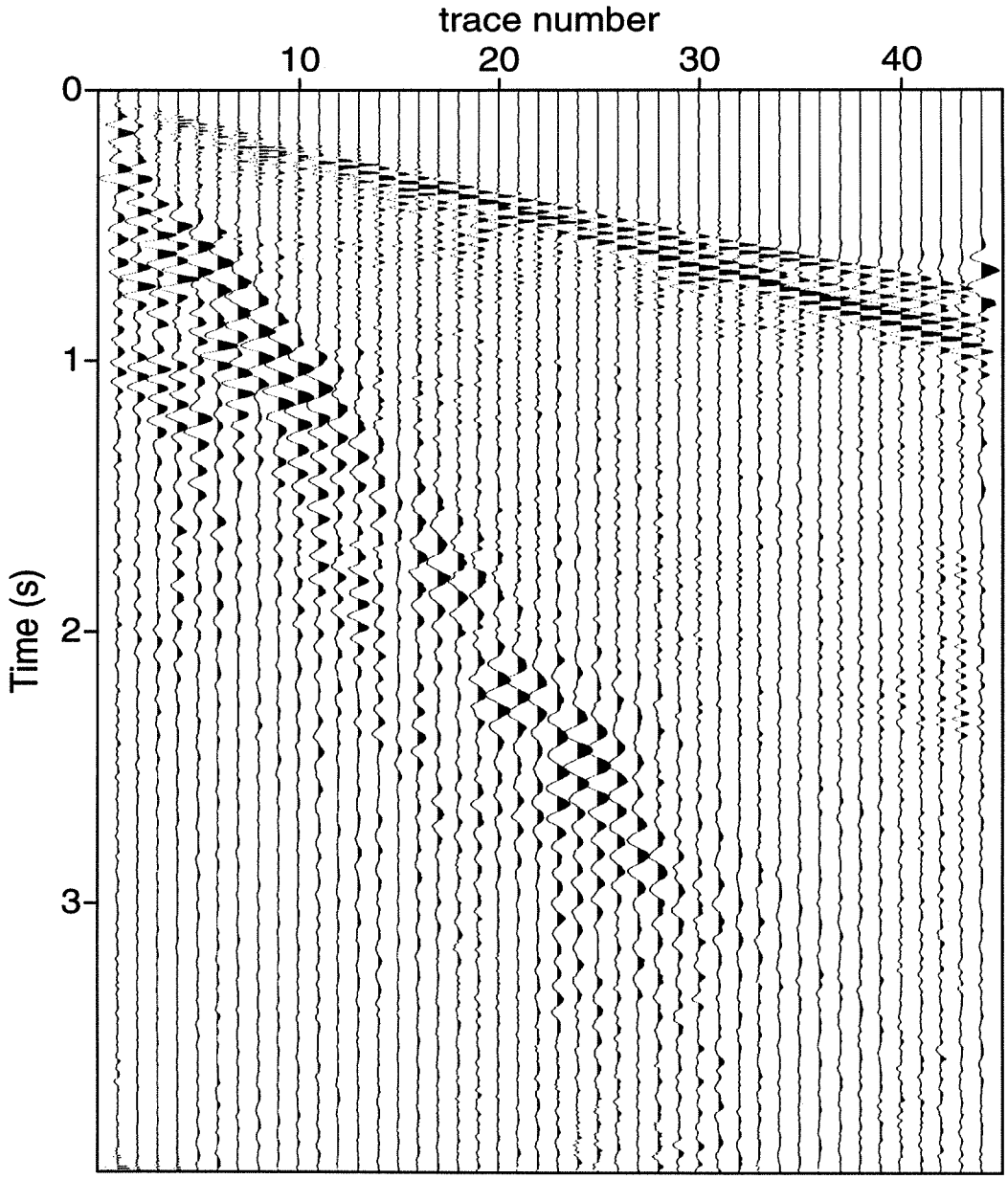

FIG. 4.13. Extracted noise using the local linear slant stack approach. Virtually no energy from the signal was extracted.

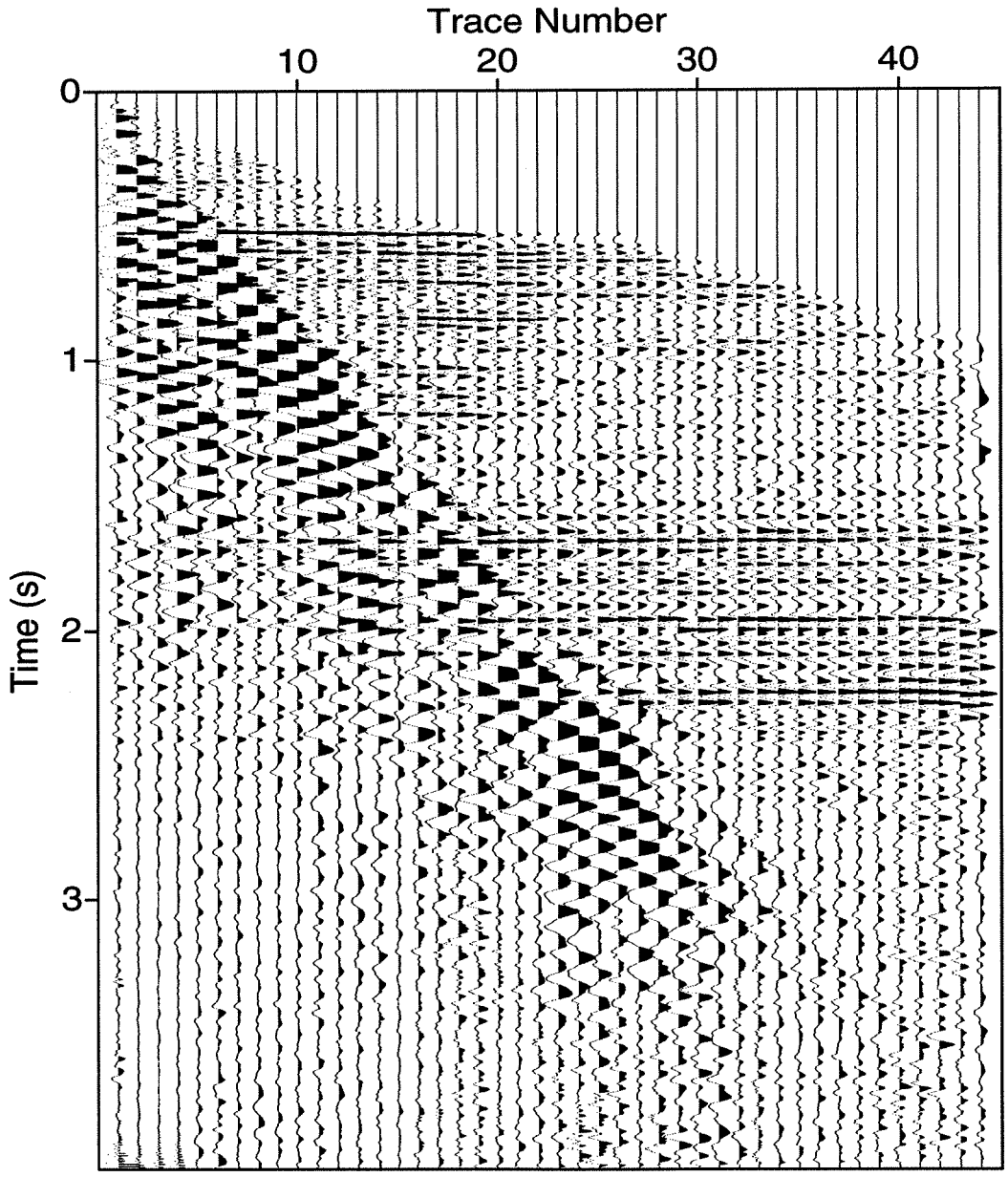

FIG. 4.14. NMO-corrected input shot record.

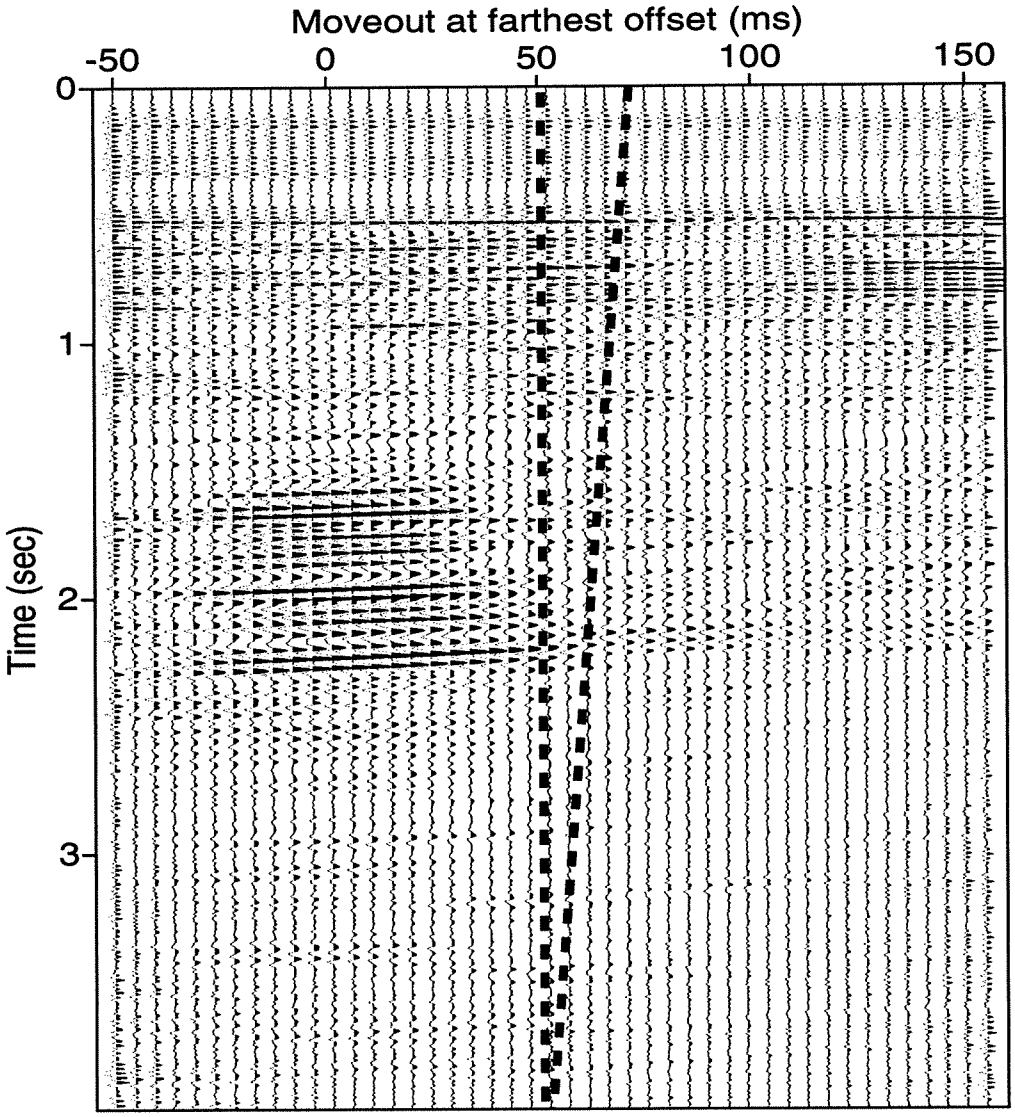

FIG. 4.15. Parabolic τ-p transform of field shot. The NMO-corrected reflections have been mapped close to zero p-value. The vertical dashed line represents the maximum p-value fully preserved and the slanted dashed line the minimum p-value totally rejected. A linear amplitude taper is applied between the two.

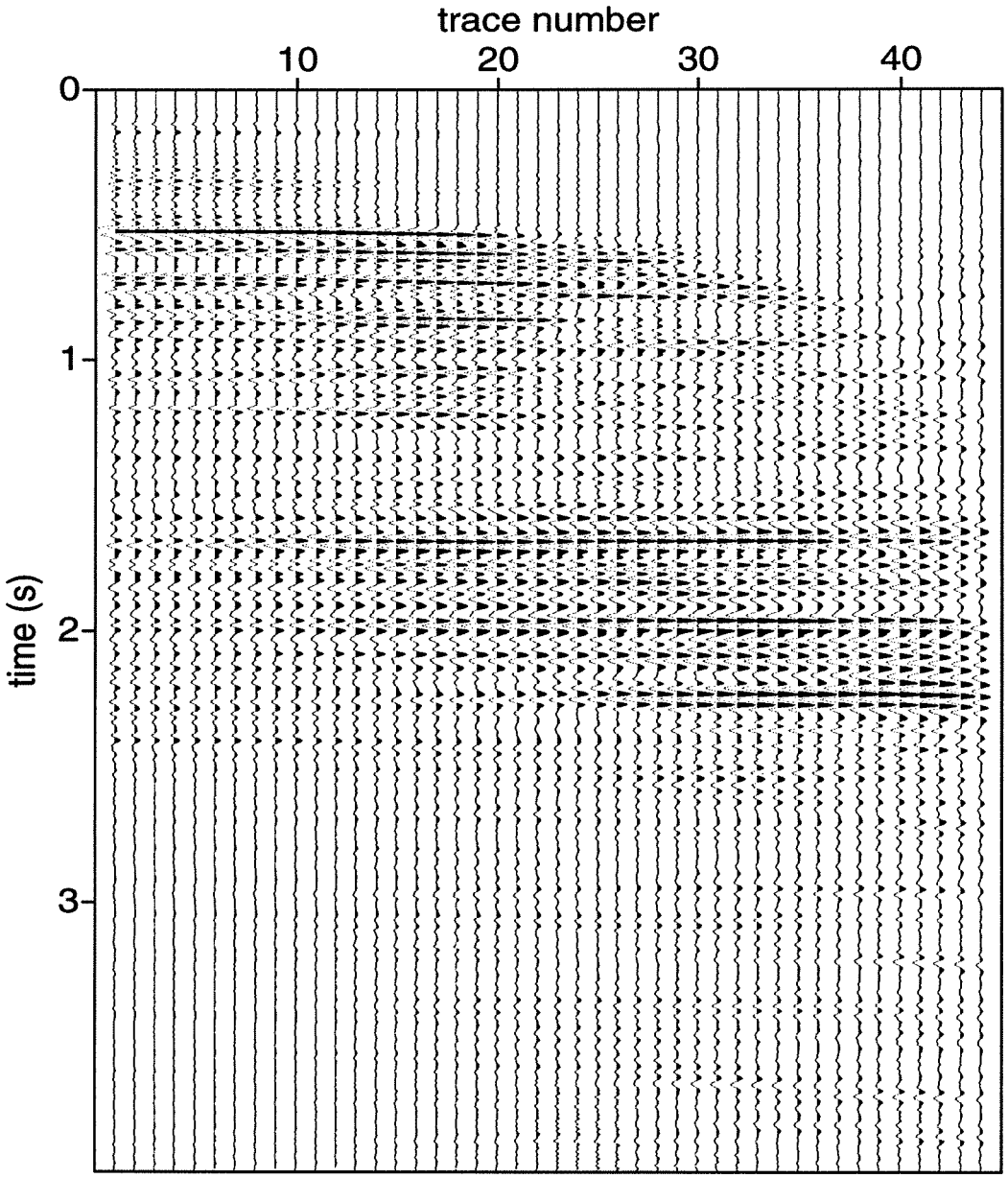

FIG. 4.16. Extracted signal with the modified signal-noise separation algorithm. Note that no ground-roll has been left and the reflections have been recovered even for the shortest offsets.

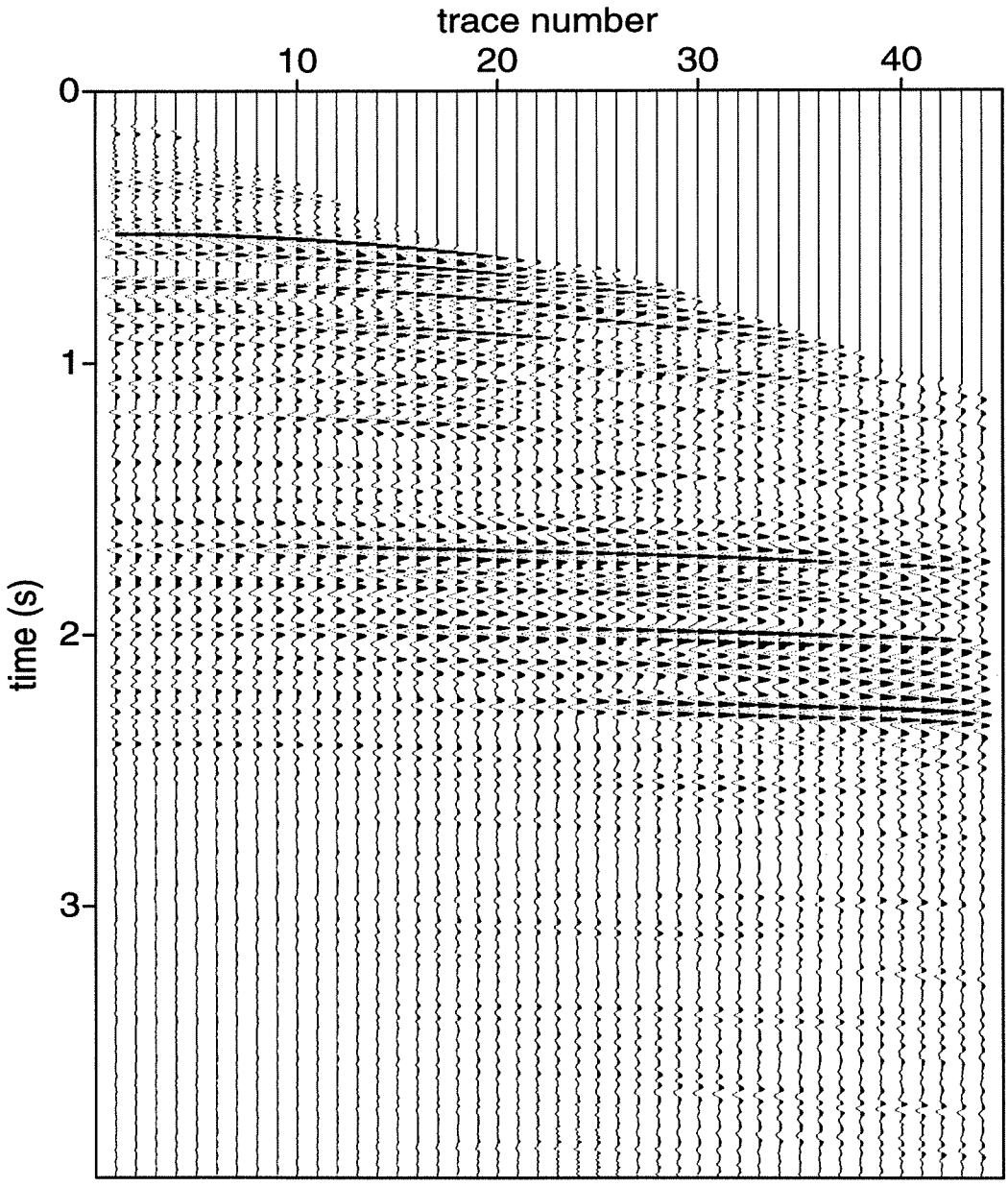

FIG. 4.17. Extracted signal with the modified signal-noise separation algorithm after the application of an inverse NMO correction.

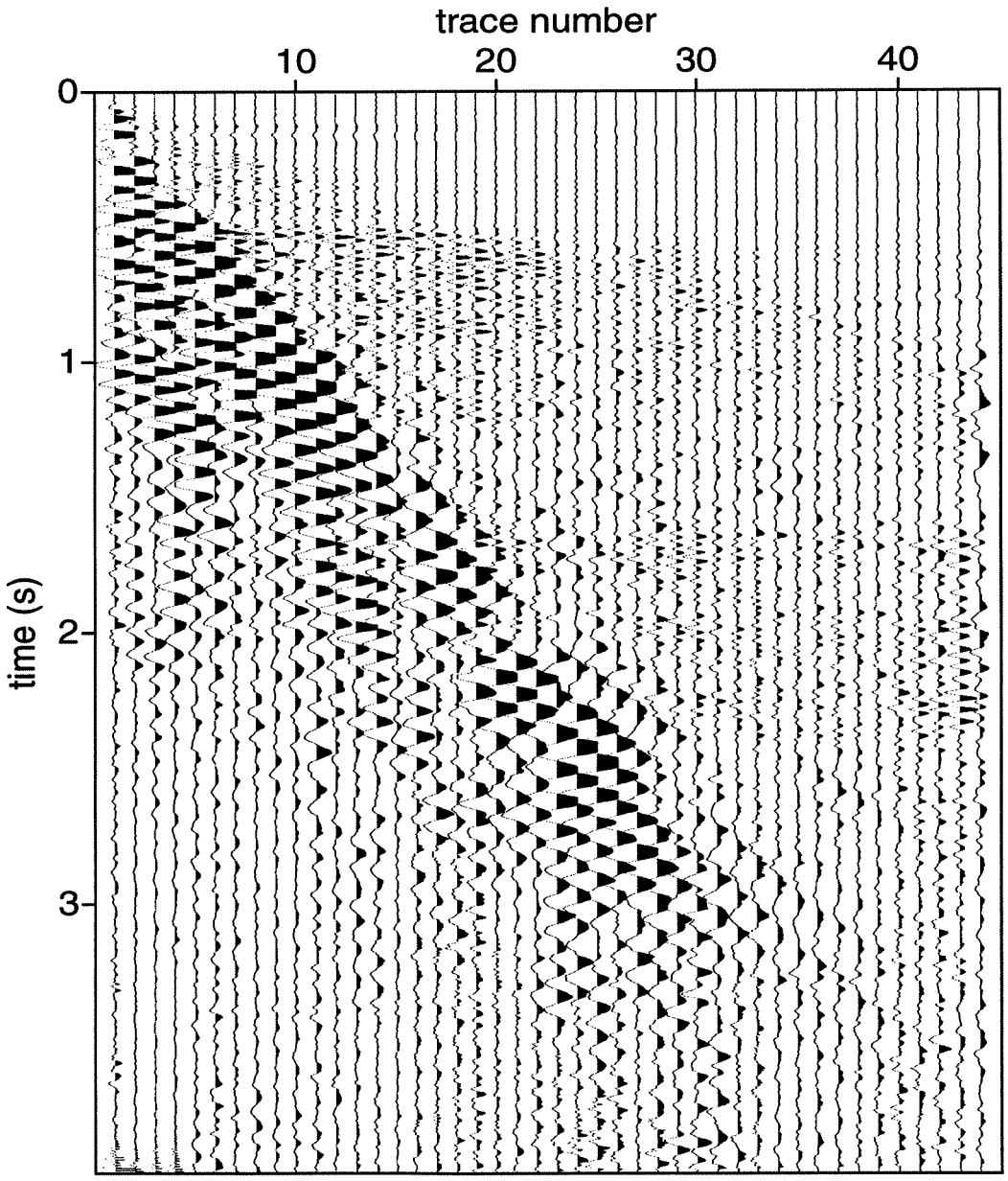

FIG. 4.18. Extracted noise with the modified signal-noise separation algorithm. Very little signal has been extracted as noise.

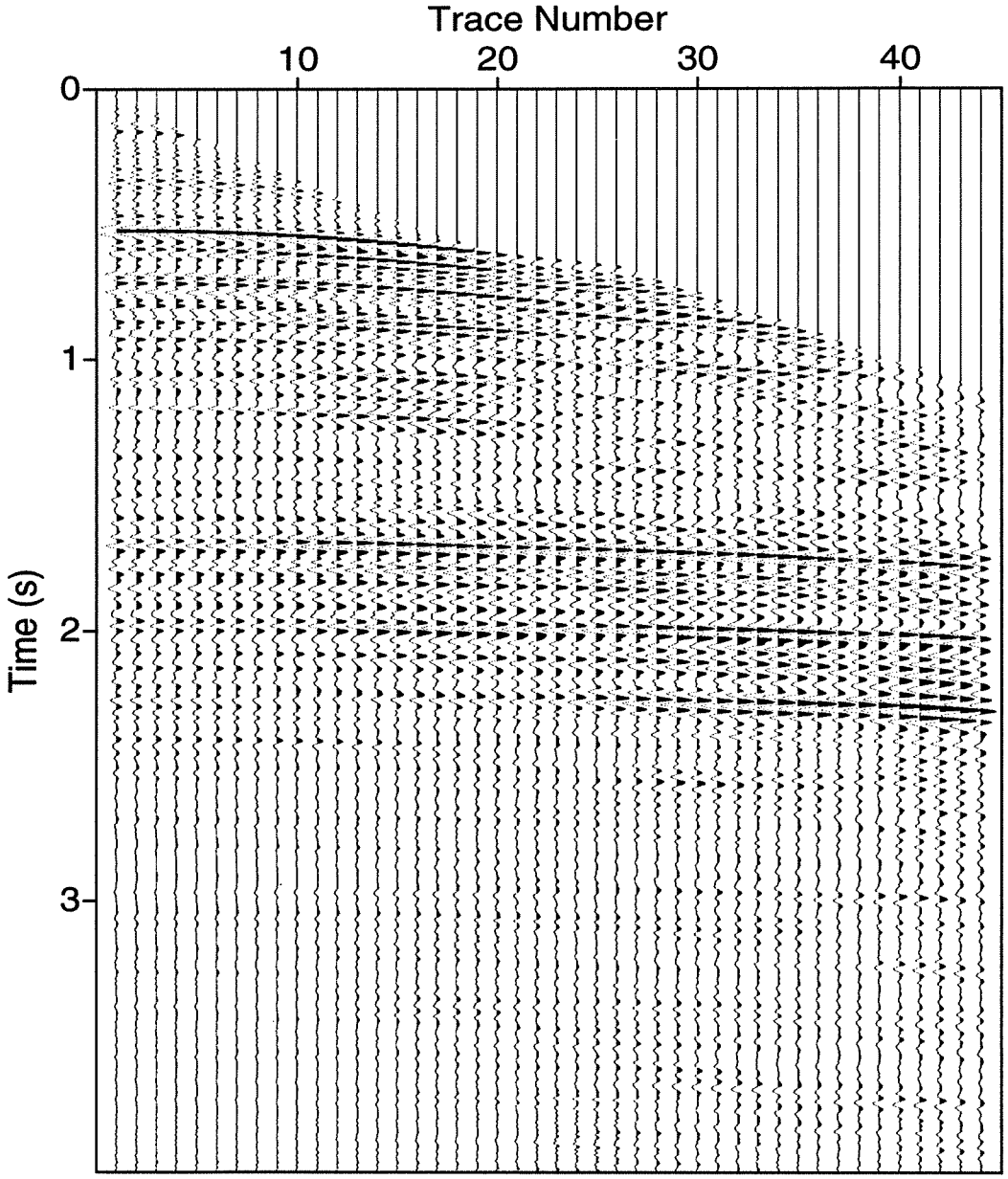

FIG. 4.19. Signal extracted with a parabolic Radon transform and a mute. This method was also successful in suppressing the ground-roll noise.

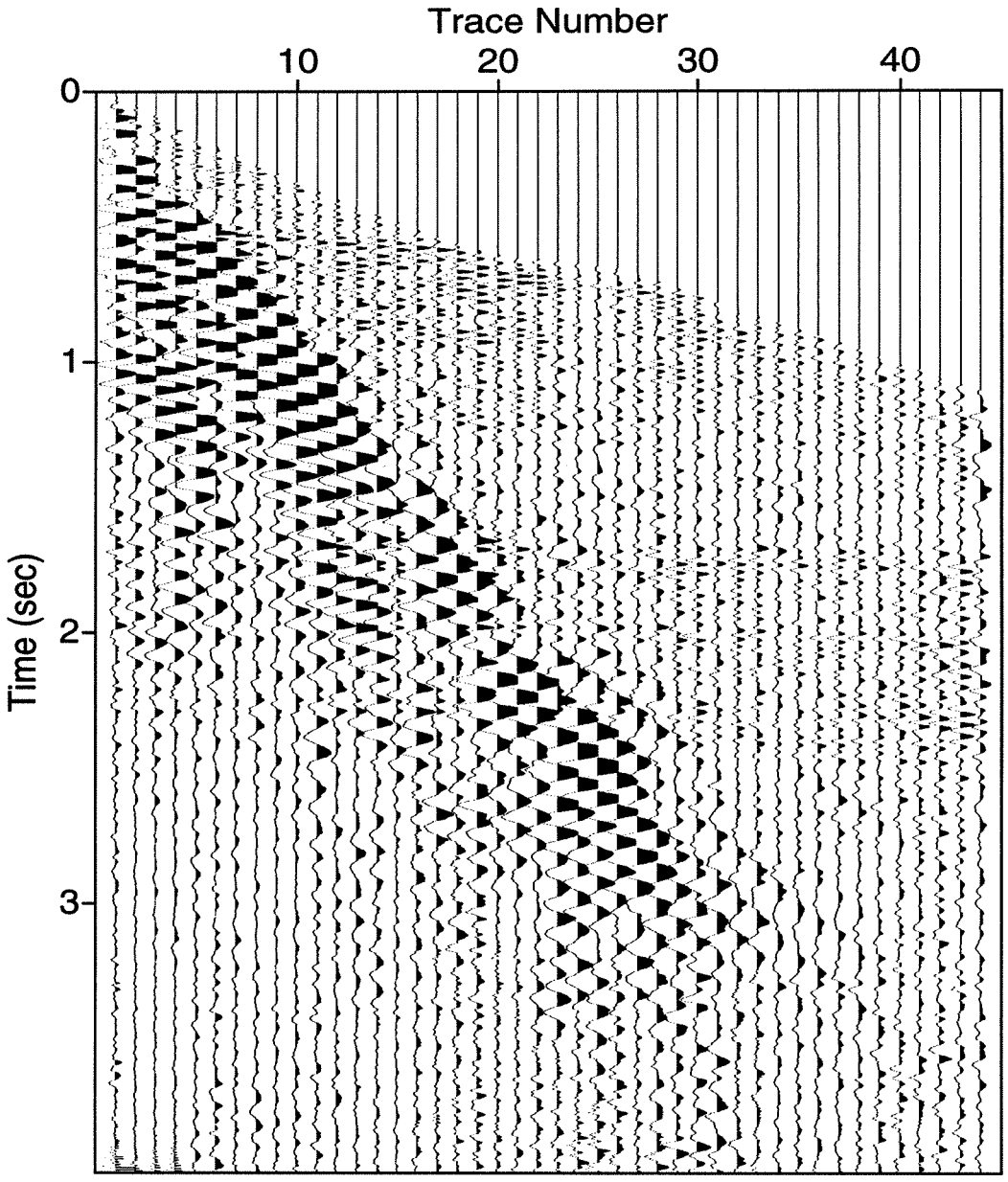

FIG. 4.20. Noise suppressed with a parabolic Radon transform and a mute. Note the residual extracted signal between 1.6 and 2.2 s.

Chapter 5

MULTIPLE SUPPRESSION

5.1 Introduction

Multiples are seismic energy that has reflected from more than one interface before being recorded at the surface of the earth. Energy from multiples is an undesirable feature in seismic records and in some cases can so dominate over primary reflections that they preclude the identification and interpretation of the primaries, especially in marine data. Common multiple-suppression methods can be roughly divided into two main categories: those that exploit the periodicity of the multiples (mainly predictive deconvolution) and those that exploit the difference in moveout between the primaries and the multiples (these include F-K filtering and τ-p filtering). Here, I review these methods and illustrate a way in which the signal-noise separation algorithm presented in Chapter 3 can be used to suppress multiples.

Figure 5.1 shows a test dataset with six horizontal events, simulating NMO-corrected primaries, and six low-moveout hyperbolas, simulating the under-corrected multiples. The amplitude ratio of the multiples to the primaries is 4:1. The data set consists of 100 traces, with 1000 samples each, at 4-ms sampling interval and trace spacing of 30-m. It is important to note that the amplitude of the signal in these modeled data is the same on all traces. Therefore, one measure of the success of any multiple-suppression algorithm will be that the output data exhibit amplitudes that are independent of offset.

5.2 F-K filtering

The use of F-K filtering to suppress multiples is based on the same idea described in the previous chapter for ground-roll suppression. Specifically, after a 2-D Fourier transform, the NMO-corrected primaries and the multiples, having different slopes, will map to different regions of the F-K plane, thereby making it possible to apply a mute to remove the energy in the region to which the multiples mapped, keeping that of the primaries. As with the ground-roll, the success of the method is compromised if the signal or the noise or both are spatially aliased or if the difference in moveout between the primaries and the multiples is not sufficiently large. Even if the data are not spatially aliased, we still must deal with the inherent trade off between the required degree of rejection and the need for a taper to avoid introducing Gibb's phenomenon noise.

To complicate the situation, we cannot achieve the same degree of rejection for all offsets because the energy associated with the smallest offsets, corresponding to the apex of the hyperbolas, is mapped to much the same region as that of the signal and so the two cannot be separated. The inevitable offset-dependence of multiple suppression might render the

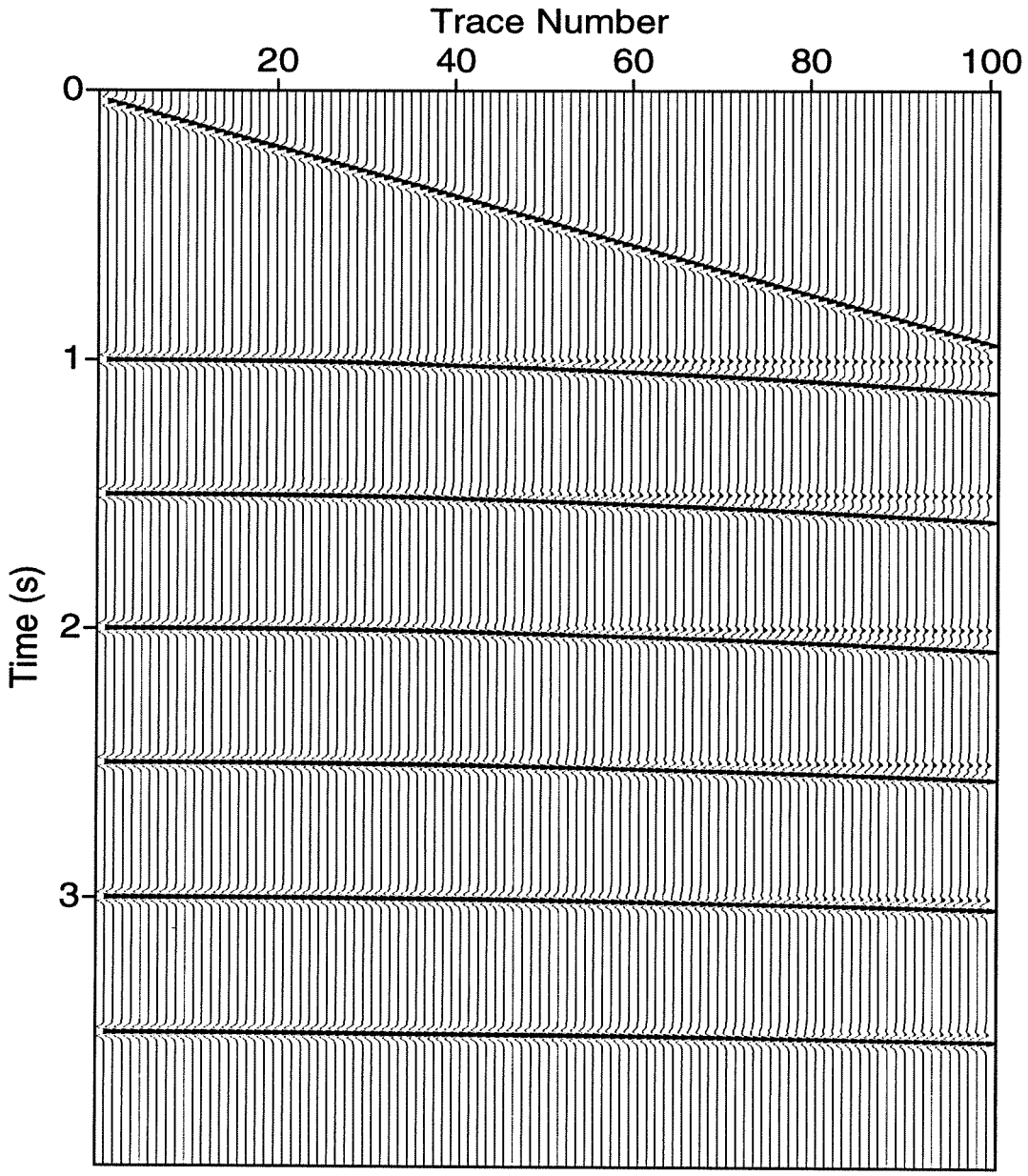

FIG. 5.1. Modeled test dataset 3. The horizontal events are the primaries, and the hyperbolic ones the undercorrected multiples. The multiple to primary amplitude ratio is 4:1.

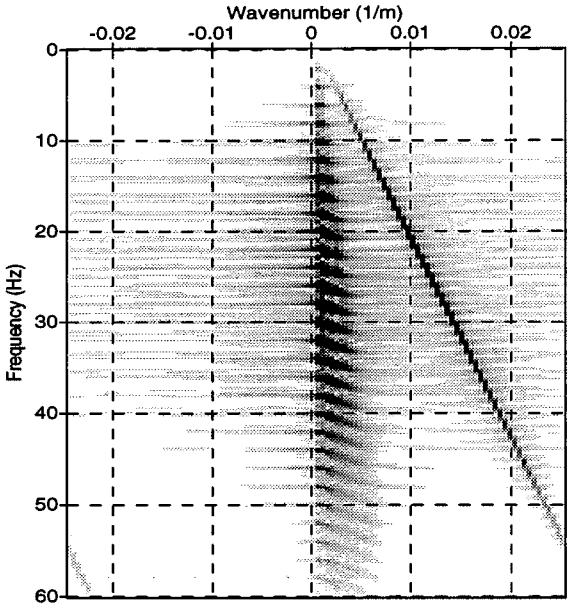

FIG. 5.2. F-K spectrum of test data. Primaries are mapped near the vertical. Multiples are mapped just away from the vertical. The linear event corresponds to the first arrivals. All the data were used in computing the F-K transform.

processed data suspect for AVO, as will be shown quantitatively in the next chapter.

Figure 5.2 suggests that in this case, even though the data are not spatially aliased, the separation between the primaries, mapped to the vertical axis (infinite apparent velocity) and the multiples, mapped to close to the vertical, may not be large enough for the method to perform well.

Figure 5.3 shows the corresponding F-K spectrum after the application of a fan-shaped mute with smallest cutoff slope of -0.218 ms/trace (events with slope less than this value will be multiplied by zero), smallest pass slope of 0.0 ms/trace (events with slope between -0.218 and 0 ms/trace are linearly tapered), largest pass slope of 0.394 ms/trace (events with slope between 0 and 0.394 ms/trace are multiplied by unity), and largest cutoff slope of 0.446 ms/trace (events with slope larger than that are multiplied by zero; those with slope between 0.394 and 0.446 ms/trace are linearly tapered). The low-frequency components of the primaries may be compromised by the rejection process because the width of the taper zone of the fan-shape filter approaches zero and the moveout difference between the primaries and the multiples is small for low frequencies. As suggested above, the result of applying this method normally shows good rejection for the large offsets and poor rejection for the small ones. Figure 5.4 and Figure 5.5 exhibit this problem. The level of noise rejection is good for the far offsets, but poor for the near ones. This is of course a most undesirable feature because the remaining multiple energy, being almost flat, will contribute to the stack and will

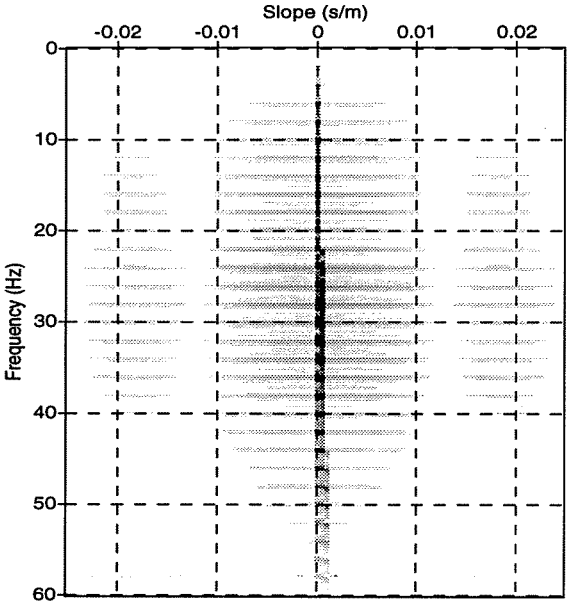

FIG. 5.3. F-K spectrum of the F-K filtered modeled test dataset 3.

thereby distort it. Also, this non-uniformity in the level of multiple suppression with offset, will distort the interpretation of AVO analysis.

5.3 Hampson's parabolic τ-p filtering

Recall that Hampson (1986) developed an algorithm, called inverse velocity stacking, to achieve multiple suppression for NMO-corrected CMP gathers. Its main features are as follows. The shot records are sorted into CMP gathers so that the reflection hyperbolas will be symmetric even if the reflectors are not horizontal. The CMP gathers are then NMO-corrected with the primary velocities so that the multiples will remain undercorrected. Hampson's key observation, which enabled him devise his efficient algorithm, was that the moveout of the undercorrected multiples could be approximated by parabolas as long as the undercorrection was not too severe. This observation made it possible to use the efficient F-X discrete generalized Radon transform described in Chapter 2 to stack the data along parabolas with moveout spanning the range of residual moveouts in the undercorrected multiple events.

Once the parabolic transform has been applied, the NMO-corrected primaries should map to or near zero parameter p and, because of their moveout, the multiples should map to larger p values. Then, provided that moveout difference between the primaries and the multiples is sufficiently large, a mute can be applied to separate the data mapped to the smallest p values (primaries) from those mapped to larger p values (multiples). As with the F-K filtering, it is advisable to apply some taper to provide a smooth transition between the

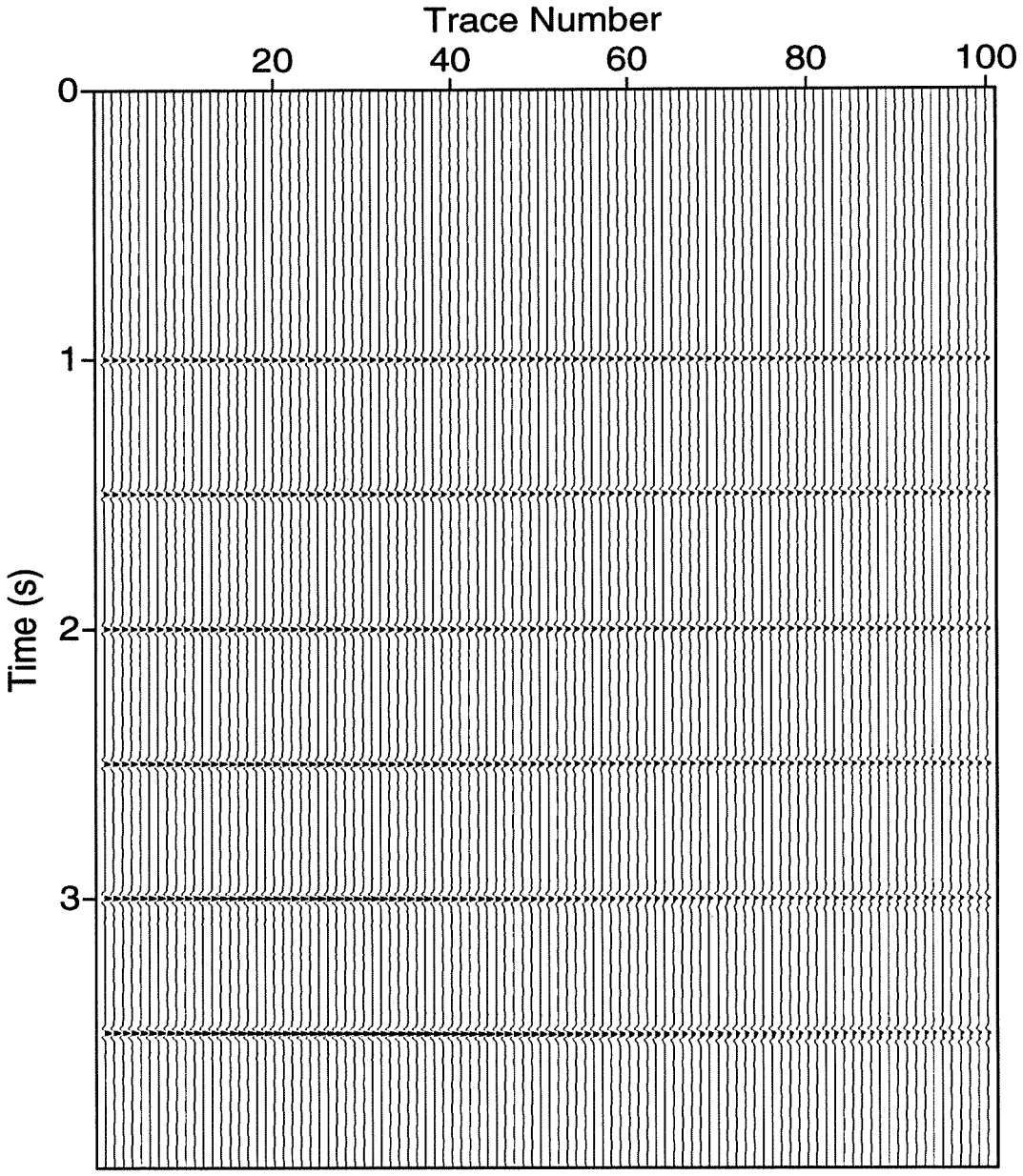

FIG. 5.4. Modeled dataset 3, F-K filtered to preserve signal. Strong energy from the multiples remains, in particular at short offsets and large times. Variation of amplitudes with offset can also be seen.

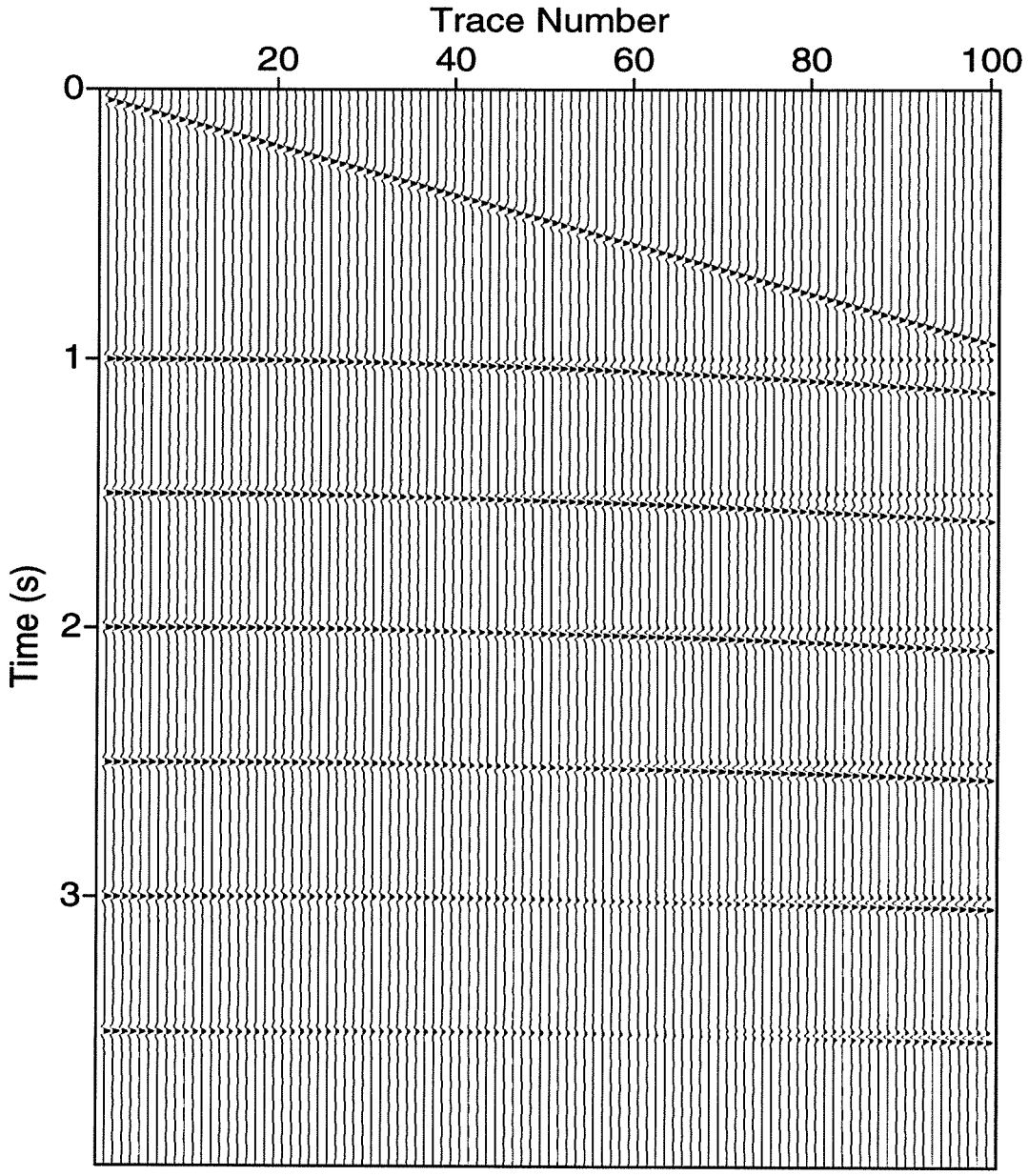

FIG. 5.5. Multiples and direct arrivals suppressed by the F-K filtering method. Note the poor suppression at short offsets, specially for deeper events. Also, some signal has been erroneously suppressed, at far offsets.

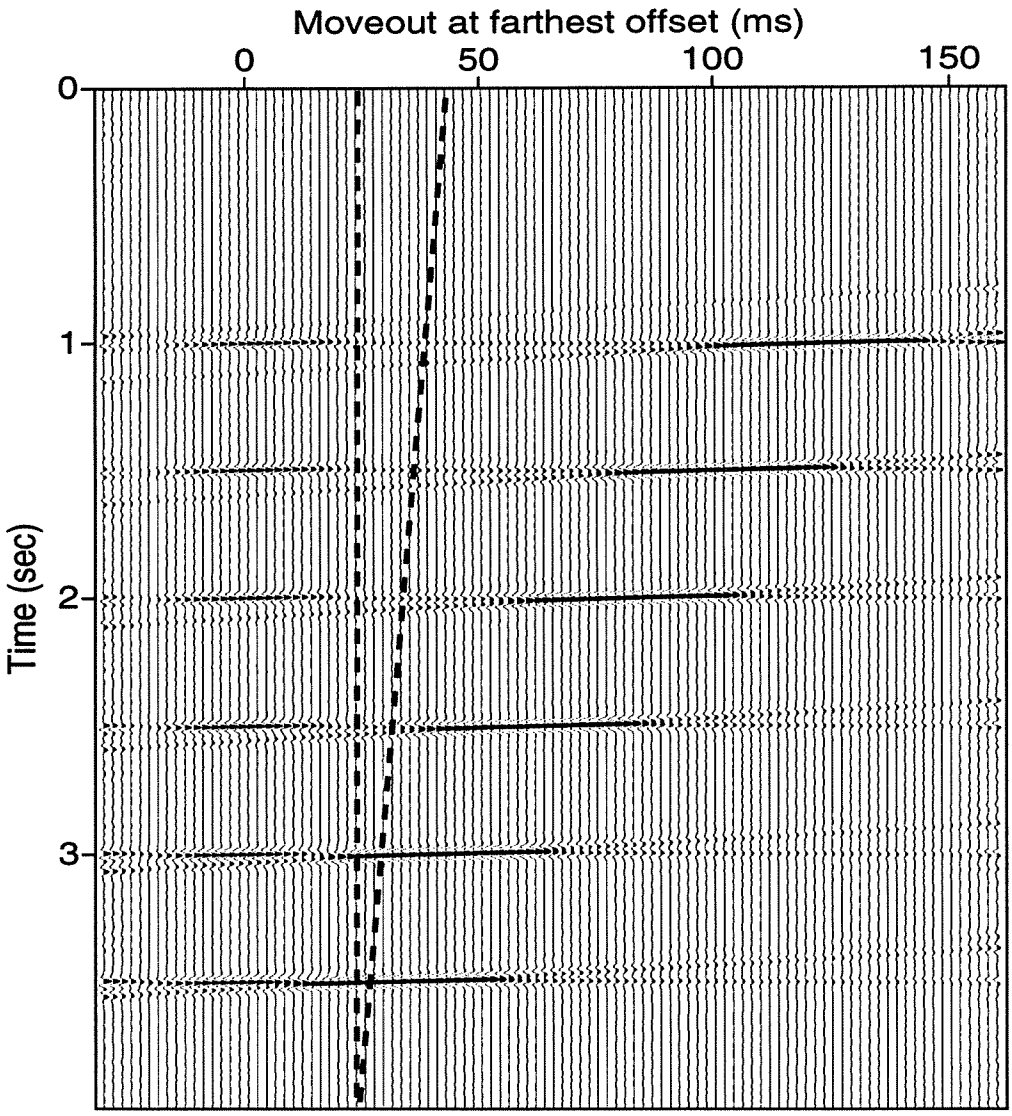

FIG. 5.6. Parabolic τ-p transform of modeled test dataset 3. The primaries and the multiples have been mapped to different regions in this domain. The dashed vertical line represents the maximum p-value fully accepted as primary, and the slanted dashed line the minimum p-value completely rejected as multiples. A linear taper of amplitudes is applied between the two.

accepted and the rejected regions. An inverse τ-p transform is then applied to the extracted signal and, if desired, the multiples can be computed by subtracting the primaries from the original data. Also, an inverse NMO correction is usually applied to prepare the data for subsequent velocity analysis.

Figure 5.6 shows the result of applying the parabolic transform to the modeled dataset 3 of Figure 5.1. In this figure, because the transform is not linear, the important parameter p is not slope, but rather the curvature of the corresponding parabolas. The parameter p is sometimes relabeled q for parabolic transforms. The units in Figure 5.6 are moveout at farthest offset because this is the parameter that most directly relates to what we really want to see, difference in moveout between primaries and multiples at the farthest offset. The primaries and the multiples have been mapped to separate regions of the τ-p plane and, except for the two deepest reflectors, this separation seems enough to allow for the separation of the two types of events. Note, however, that because of the strong amplitude of the multiples, part of their energy has "leaked" to the primary zone. This is clearly seen below each of the focused primary energy regions. The vertical dashed line represents the largest p-value of the primaries, and the slanted dashed line the smallest p-value of the multiples. A linear taper of amplitudes is applied between the two. The most important advantage of this method over the previous one is that we can achieve the same level of multiple rejection for all offsets, so that if the moveout difference is sufficient, in principle the multiple suppression will be independent of offset. This issue will be investigated further in the next chapter.

The result of applying this technique to our modeled test dataset 3 is shown in Figure 5.7. The preserved primaries show residual energy from the multiples, particularly for the deepest reflectors and far offsets. This is due to the small moveout difference between the deepest multiples and primaries, which does not allow a clear separation between them (see Figure 5.6). Comparison with the primaries extracted with the F-K filtering method (Figure 5.4) shows a much better extraction in particular in terms of amplitude variation with offset. Figure 5.8 shows the suppressed multiples. Almost no signal has been extracted, and the overall result is good; in particular, the strength of the suppressed multiples is much more uniform with offset than is the case for the F-K filtered result in Figure 5.5. Note, nevertheless, that, as mentioned with regard to the previous figure, the deepest multiple was not very effectively removed.

5.4 Signal-noise separation

In Chapter 3, Harlan's signal-noise algorithm was described as a means to separate events with different patterns of moveout, provided that an appropriate invertible linear transformation could be devised that would focus the events considered signal while at the same time defocusing those considered noise. In this chapter, we consider use of this algorithm to separate primaries from multiples on the basis of their differing moveouts after an NMO-correction has been applied to flatten the primaries while leaving the multiples undercorrected. The situation here, however, differs significantly from that on Chapters 3 and 4 because now, if a linear τ-p transformation is applied to focus the corrected primaries, non-neglegible focusing of the undercorrected multiples is obtained, since their moveout is close to flat. An

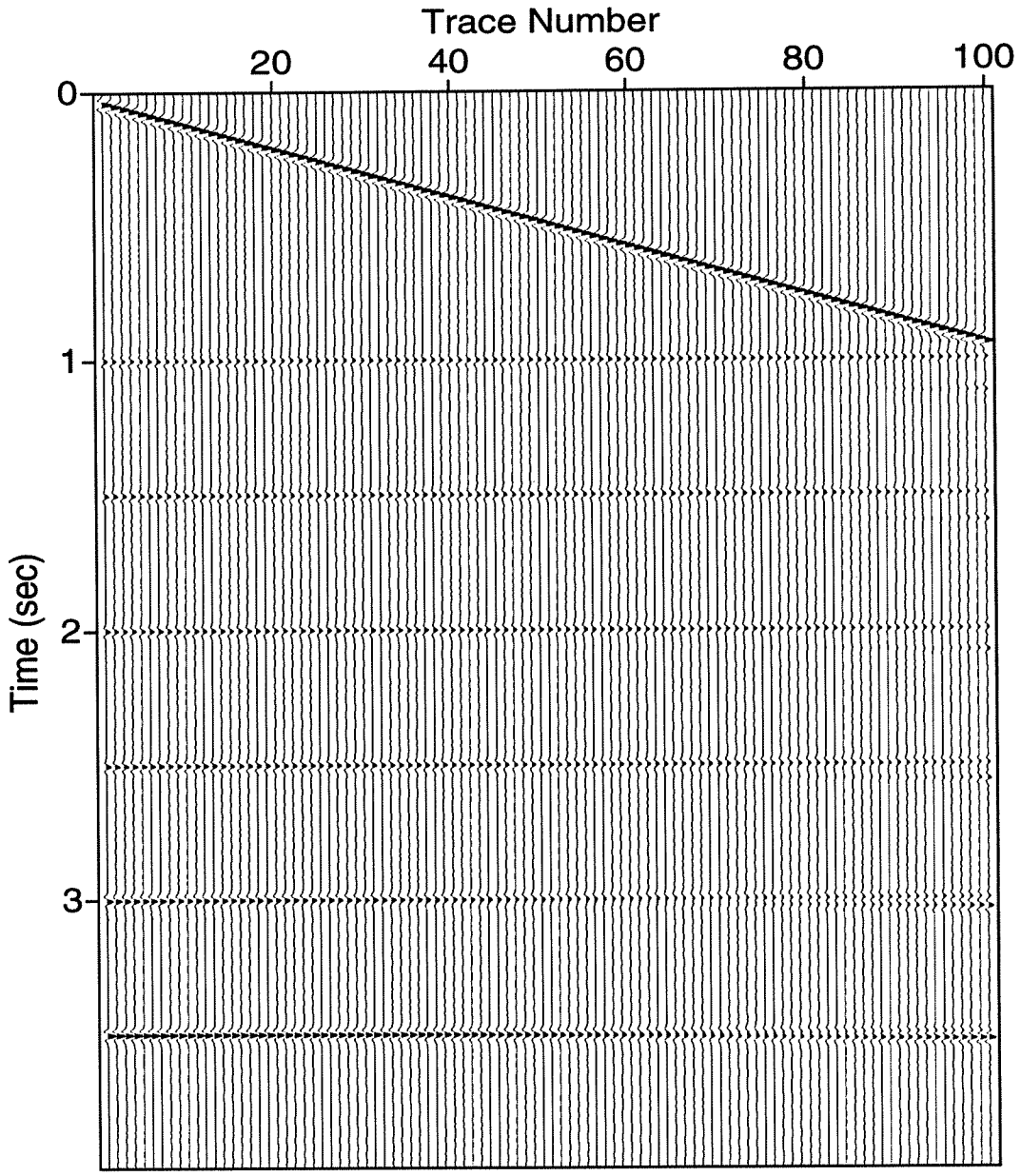

FIG. 5.7. Parabolic τ-p filtering of modeled test dataset 3 to preserve primaries and suppress multiples. Multiple energy was preserved along with the primaries, in particular for the deepest multiple.

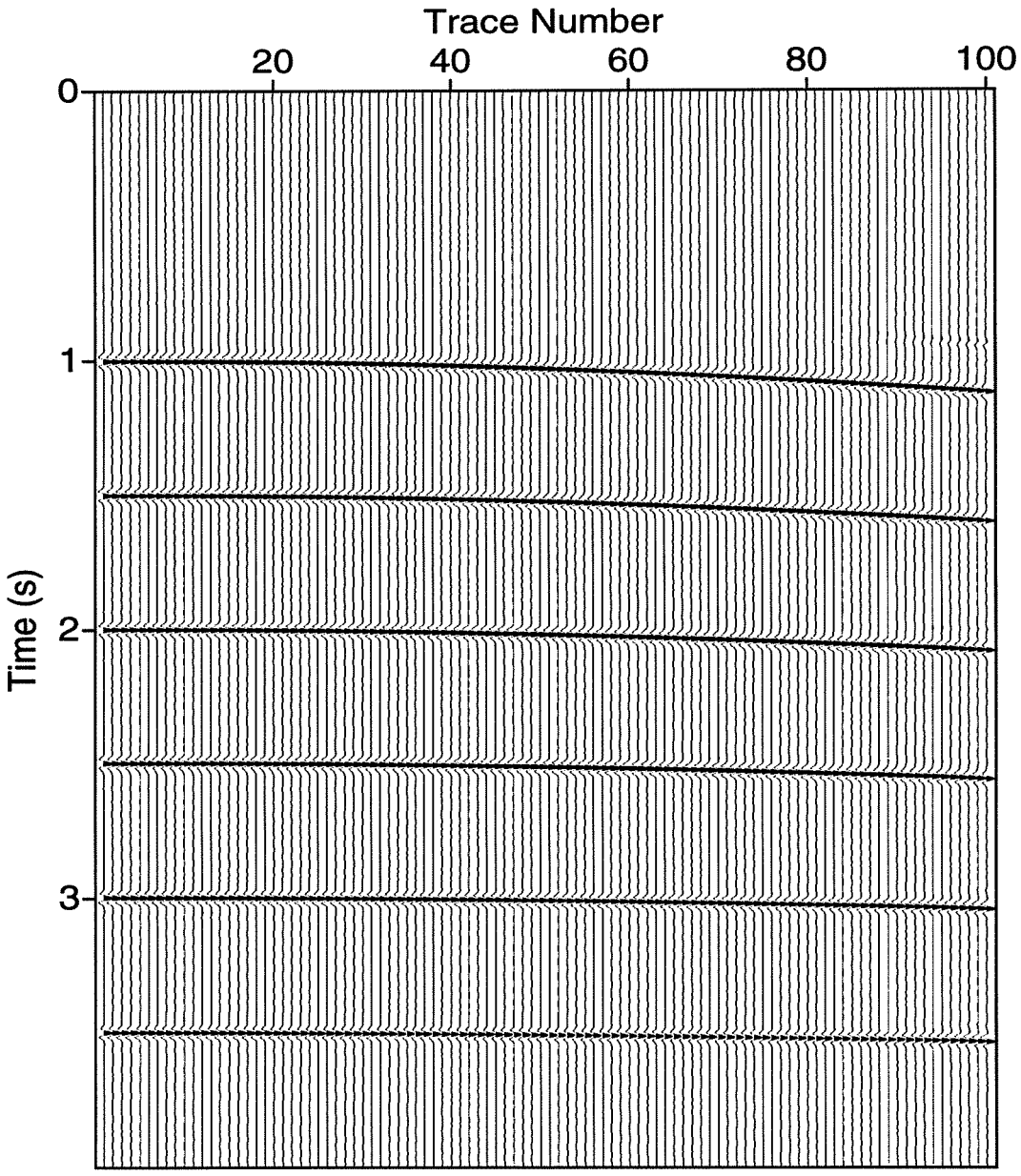

FIG. 5.8. Multiples suppressed with the parabolic τ-p transform approach. The deepest multiple was not very effectively removed.

alternative approach is to try focusing the multiples using a parabolic τ-p transform, but there is a problem. Namely, the flat events are a special case of parabolic events with zero p-value and so will also be focused. If the relative amplitudes of the primaries are of the same order of magnitude of those of the multiples, it is not possible to directly apply the signal-noise separation algorithm to discriminate between the two.

In order to use the signal-noise separation algorithm for this application, I propose a hybrid approach between Hampson's p-filtering and Harlan's signal-noise separation algorithm. The energy that is unquestionably associated with the multiples (that is, to the right of the slanted dashed line in Figure 5.6) is zeroed out and the signal-noise separation algorithm is used to refine the extraction of the primaries in the primary zone (that is, to the left of the dashed line in Figure 5.6), eliminating any residual contribution by the multiples. The multiples are then computed by subtraction from the original data, after the extracted primaries have been inversely τ-p transformed.

In this application, therefore, the goal of the signal-noise separation algorithm is not to discriminate between the focused primaries and the focused multiples (which it is unable to do), but between the former and the unfocused residual multiples. Most of the multiple energy, therefore, is removed via τ-p filtering just as Hampson's approach does and its residual (if any) is removed via the signal-noise separation algorithm. The stronger the multiples in comparison with the primaries, the more residual multiple energy will be left for the signal-noise separation algorithm to remove. This issue will be studied in more detail in the next chapter.

Figure 5.9 shows the primaries extracted with the hybrid approach. Note that in this case, less residual energy from the multiples can be seen, in particular for the shallower events, than in Figure 5.7. This is a consequence of the residual multiple extraction provided by this method. The improvement is not so great for the deepest multiples because, as mentioned before, their differential moveout with respect to the primaries at the same times is not large enough to allow a clear separation between them. The amplitude level of the primaries, however, looks weaker than in Figure 5.7, indicating a tradeoff between level of multiple rejection and primary extraction. In the next chapter I quantify both these effects, the level of improvement in multiple rejection against the weakening of primary extraction.

Figure 5.10 shows the removed multiples with the hybrid approach described in the previous paragraph. The main contribution to the removal of multiples comes from the τ-p filtering, with a residual suppression from the signal-noise separation algorithm. Figure 5.10 is almost identical to Figure 5.8, which shows the extracted multiples with Hampson's approach. This of course is due to the fact that the residual multiple energy extracted with the signal-noise separation is much weaker than the bulk of the multiple energy removed with the τ-p filtering portion of the hybrid algorithm.

To end this section, I would like to point out that the extra flexibility afforded us by the signal-separation algorithm in removing the residual multiple energy does not come free, the algorithm is about 50% more expensive than the simpler Hampson's τ-p mute. Whether or not this extra expense is justified depends on whether or not we can tolerate some residual multiple energy in the extracted primaries and perhaps more important, in the way this residual multiple energy may distort amplitude variations with offsets. This important issue

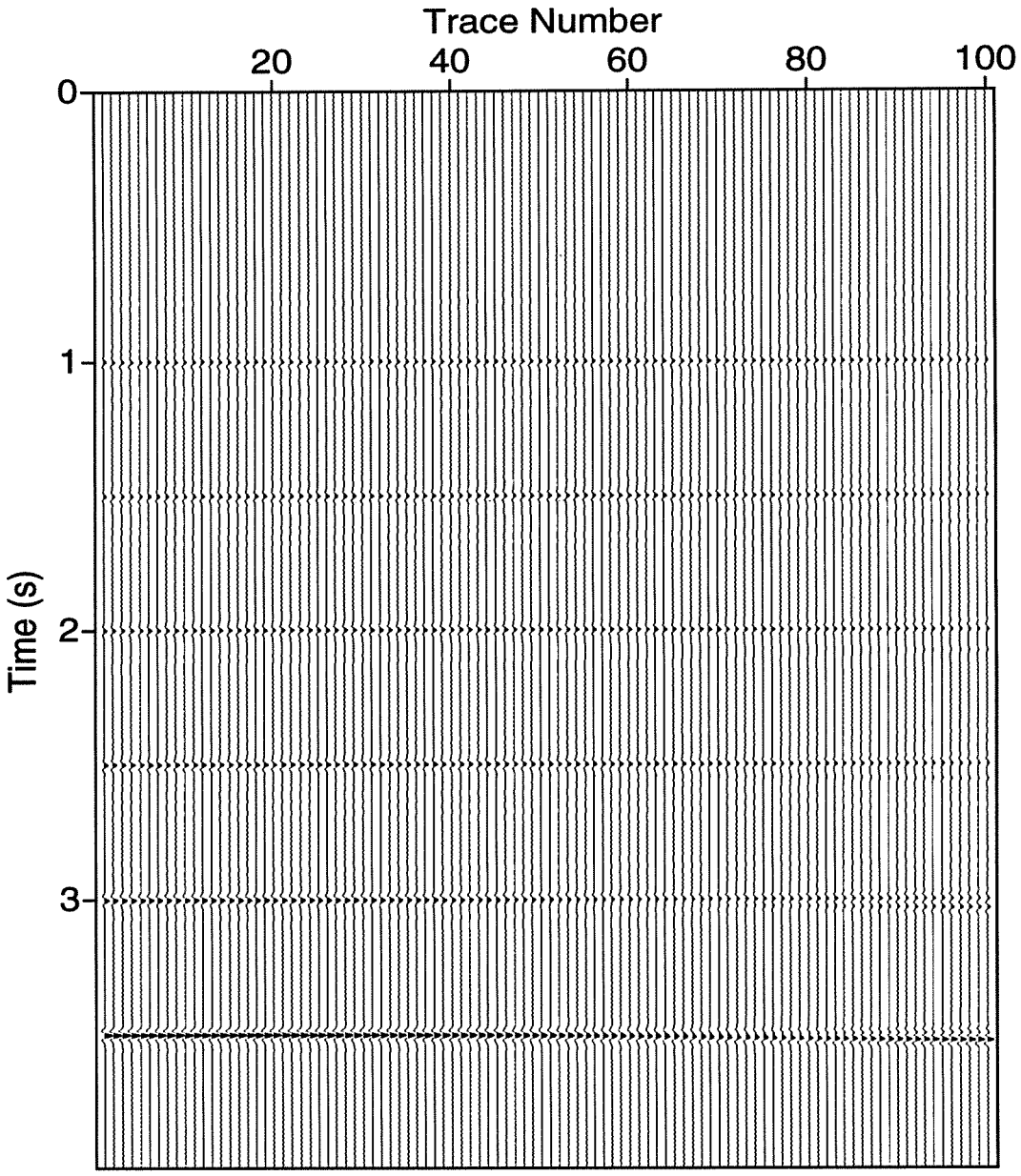

FIG. 5.9. Primaries extracted with the hybrid approach of the τ-p filtering and the signal-noise separation algorithm. Note that the multiple rejection is better than in Figure 5.7 for the shallow events, for which the differential moveout between the primaries and the multiples is largest.

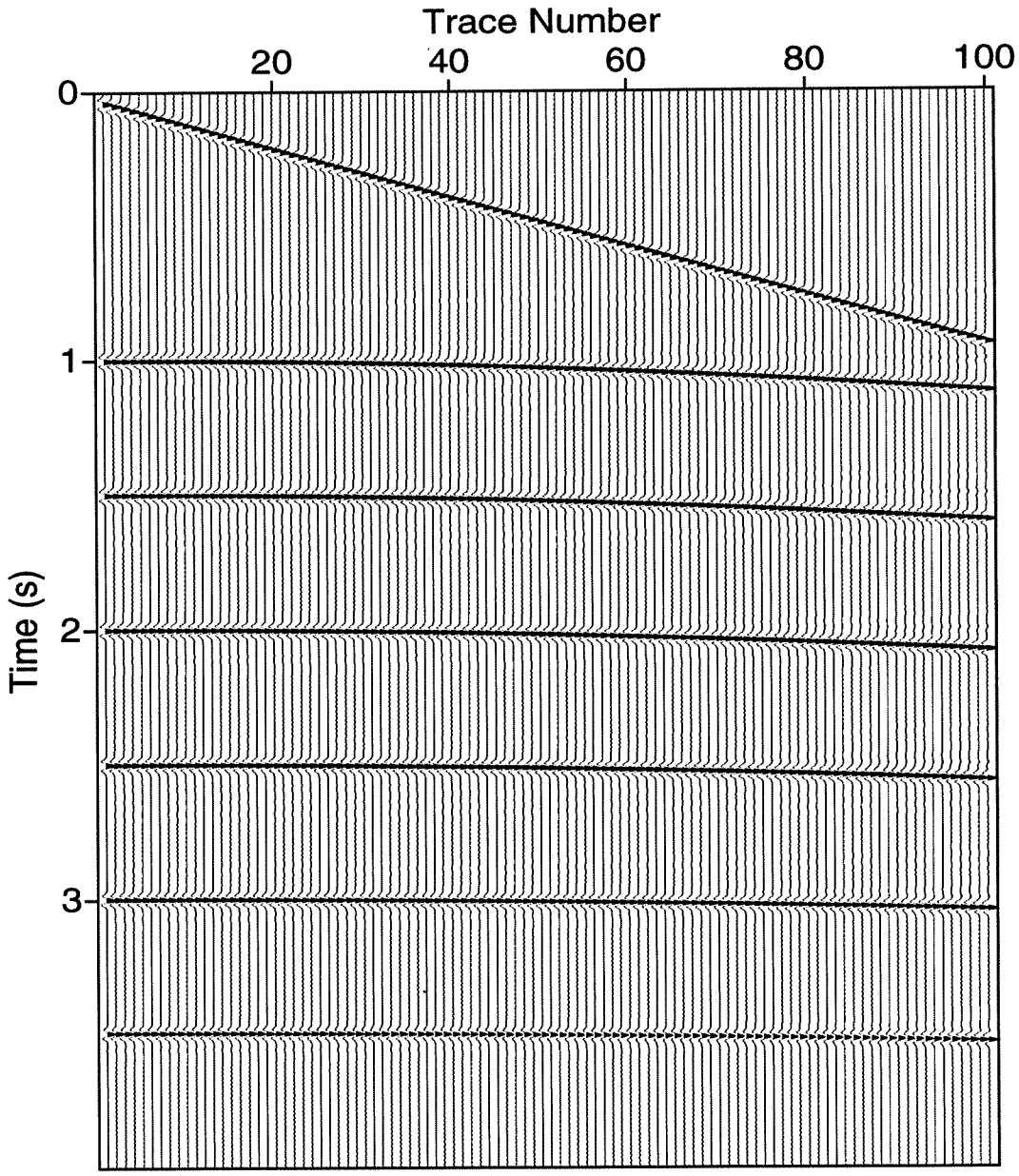

FIG. 5.10. Multiples removed with the hybrid approach of the τ-p filtering and the signal-noise separation algorithm. Virtually no energy from the signal was extracted. Note the poor removal of the deepest multiple due to the small moveout difference with the corresponding primary.

is postponed until next chapter.

5.5 Simultaneous suppression of multiples and ground-roll

We have now seen that the signal-noise separation algorithm can be effectively used to suppress either ground-roll or multiples. In this section, I apply the method to model data contaminated by both problems. Although multiples are by far a more serious problem with marine data for which ground-roll is not usually a problem, many examples exist of land data in which multiples, although not completely dominating the primary energy, interfere to such degree with the primaries that their elimination is essential to delineating subtle stratigraphic features in CMP stacks and to performing AVO analysis for lithologic purposes.

In principle, the presence of the ground-roll should pose no additional strain on the performance of the algorithms, since its moveout is far larger than that of either the primaries or the multiples. To demonstrate this, I added a synthetic train of dispersive waves to the modeled test dataset 3 in Figure 5.1 to simulate the ground-roll. The resulting data, modeled test dataset 4, are shown in Figure 5.11. The dispersive wavetrain was generated using an algorithm that propagates a dispersive wave using a supplied dispersion relation, with velocities between 250 and 1000 m/s and a frequency range from 5 to 25 Hz.

5.5.1 Hampson's parabolic τ-p filtering

The primaries can be separated from the multiples and the ground-roll using Hampson's approach, in the same way described in the previous section. Here, I applied the algorithm a second time, this time to separate the multiples from the ground-roll. In practice this second pass probably would not be necessary, since all we normally wish to do is to extract the primaries.

Figure 5.12 shows the preserved primaries, with just a little residual from the multiples. This result is virtually the same as that in Figure 5.7 corroborating that the presence of the ground-roll is not a threat to the performance of the algorithm. Figure 5.13 shows the extracted multiples, with almost no removed signal. Again, this result is almost the same as that obtained in the absence of ground-roll (see Figure 5.8). Finally, Figure 5.14 shows the removed ground-roll. No residual energy from either the primaries or the multiples is seen.

5.5.2 Signal-noise separation

The extraction of the primaries, using the signal-noise separation algorithm, proceeds in the same way described in the previous section. Again, in practice we probably just want to extract the primaries, leaving the multiples and the ground-roll together, but here I apply the algorithm once again to separate the ground-roll from the multiples.

Figure 5.15 shows the primaries extracted with the hybrid τ-p filtering and signal-noise separation algorithm, as described in the previous paragraph. As expected, the presence of the ground-roll poses no particular threat to the ability of the algorithm to identify and extract the primaries; they have been extracted the same as before (Figure 5.9), with energy from the multiples present only on the deepest reflectors. In a situation with field data, however, the

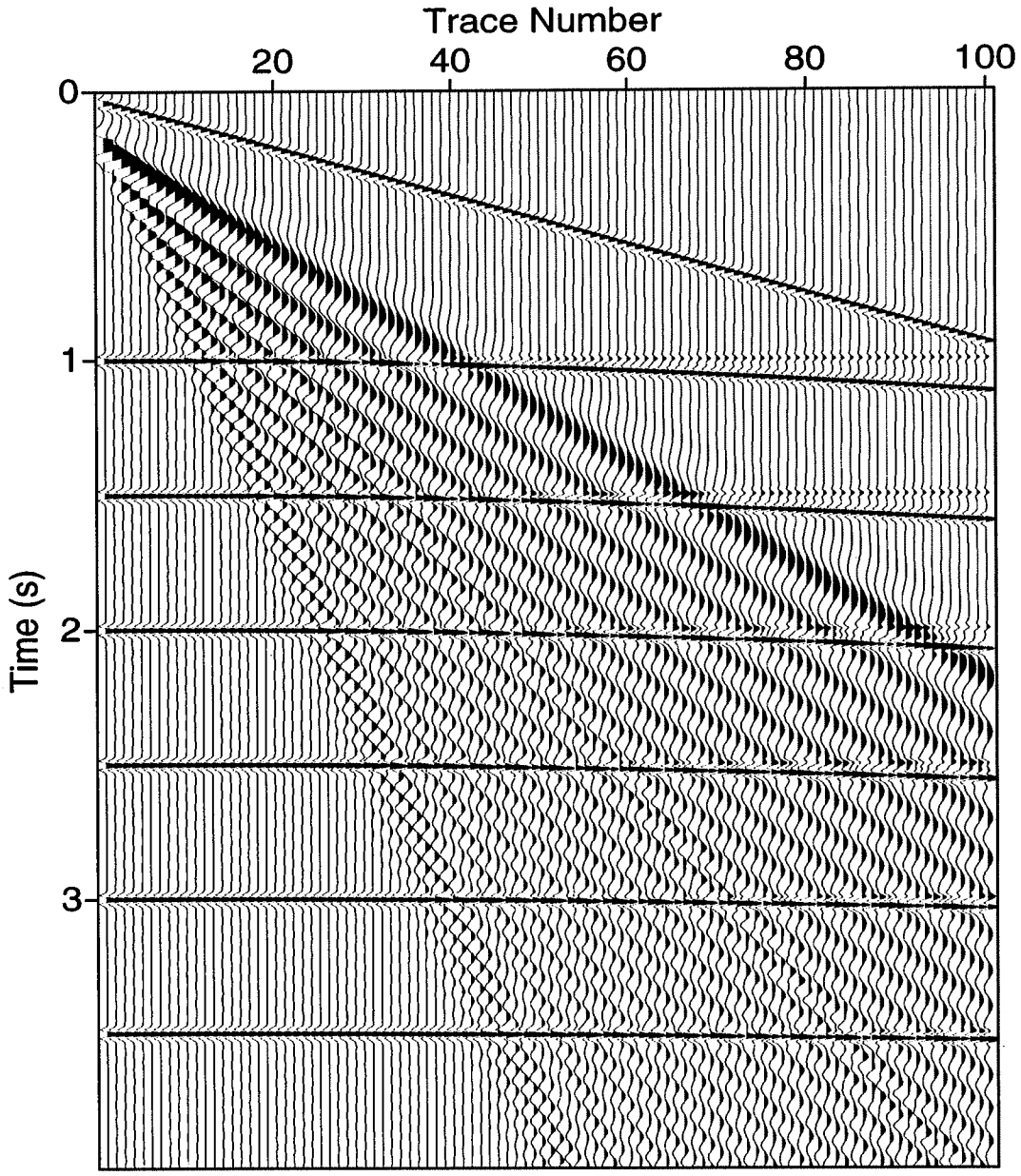

FIG. 5.11. Modeled test dataset 4; modeled test dataset 3 with a train of dispersive waves added to simulate ground-roll.

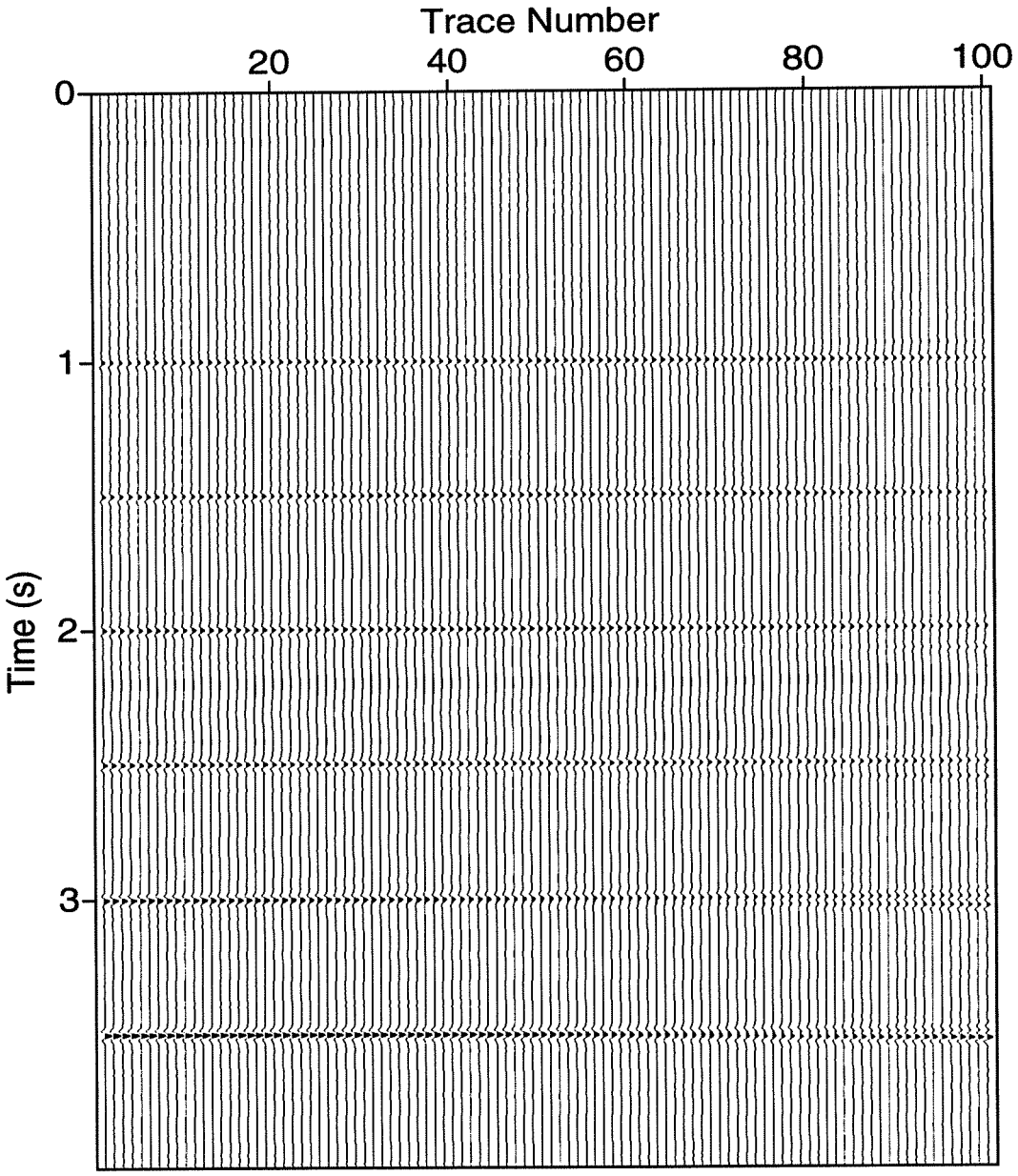

FIG. 5.12. Primaries extracted with the parabolic Radon transform and a mute for the data with ground-roll. This result is almost the same as that in Figure 5.7, indicating that the presence of the ground-roll is no threat to the suppression of multiples.

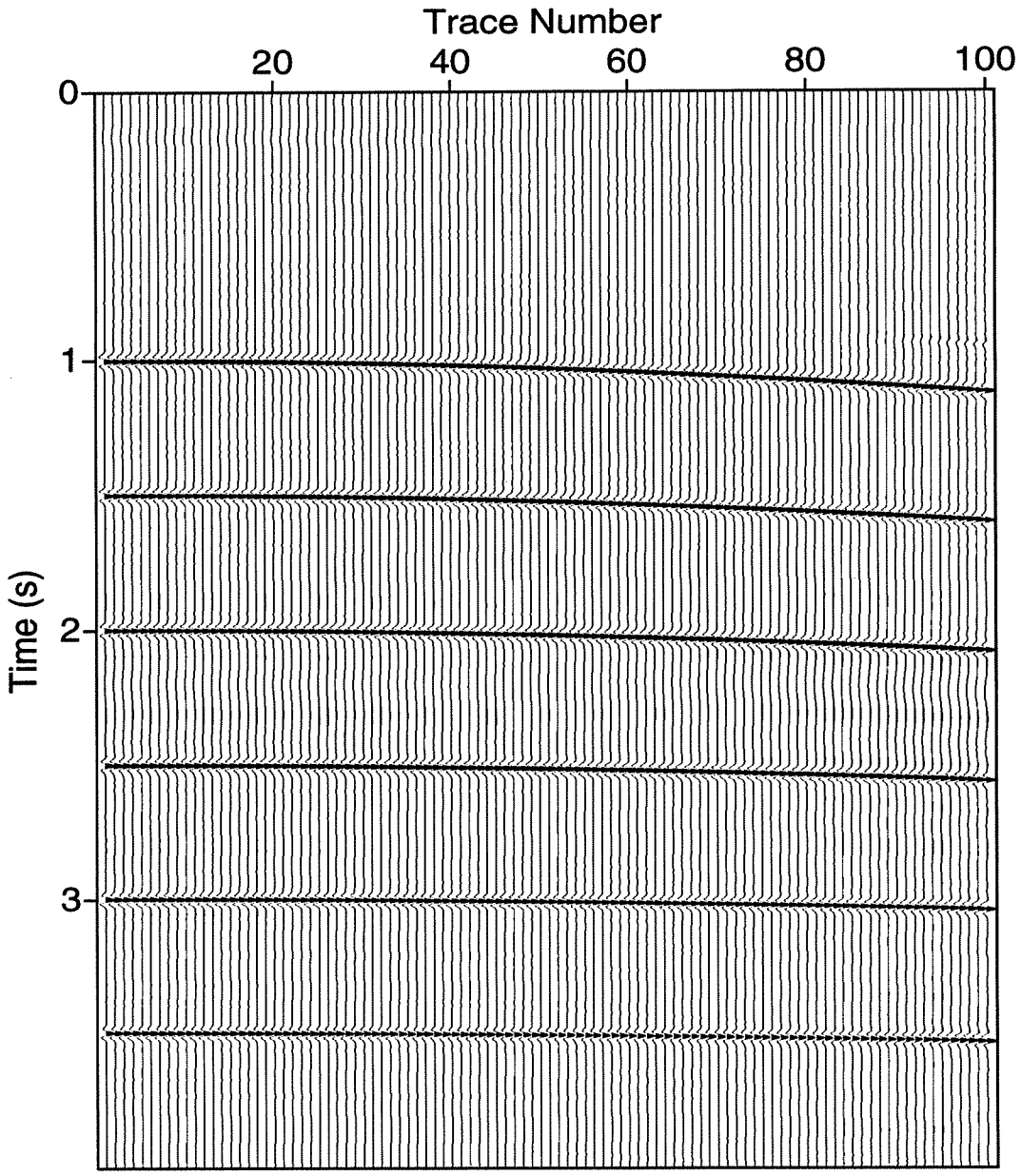

FIG. 5.13. Multiples removed with the parabolic Radon transform and a mute for the data with ground-roll. Again, the presence of the ground-roll did not degrade the performance of the algorithm (compare with Figure 5.8).

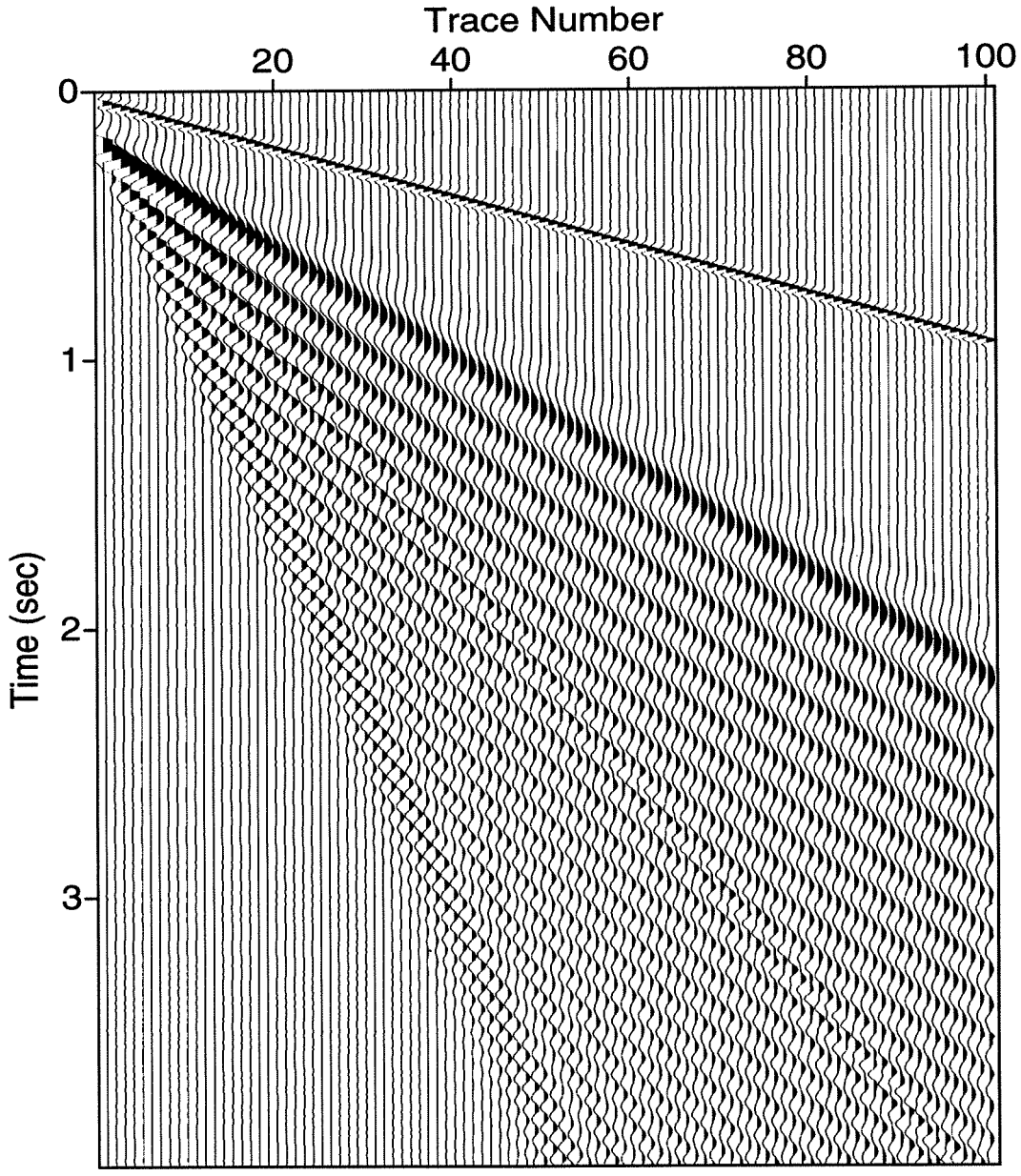

FIG. 5.14. Removed ground-roll with a parabolic Radon transform and a mute. No energy from primaries or multiples remains.

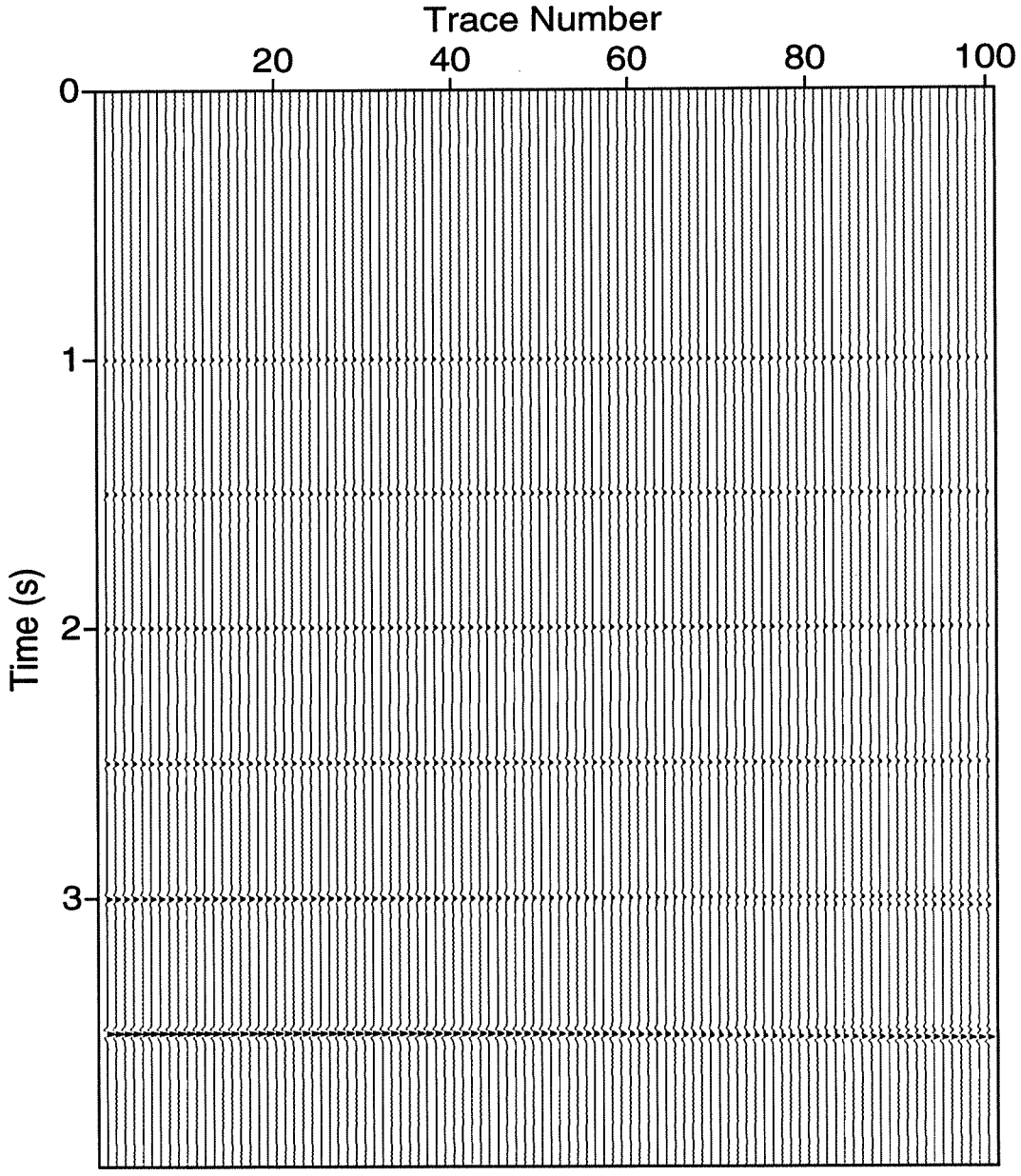

FIG. 5.15. Primaries extracted with the hybrid τ-p filtering and signal-noise separation algorithm for the data with ground-roll. Compare with Figure 5.9 for data without the ground-roll.

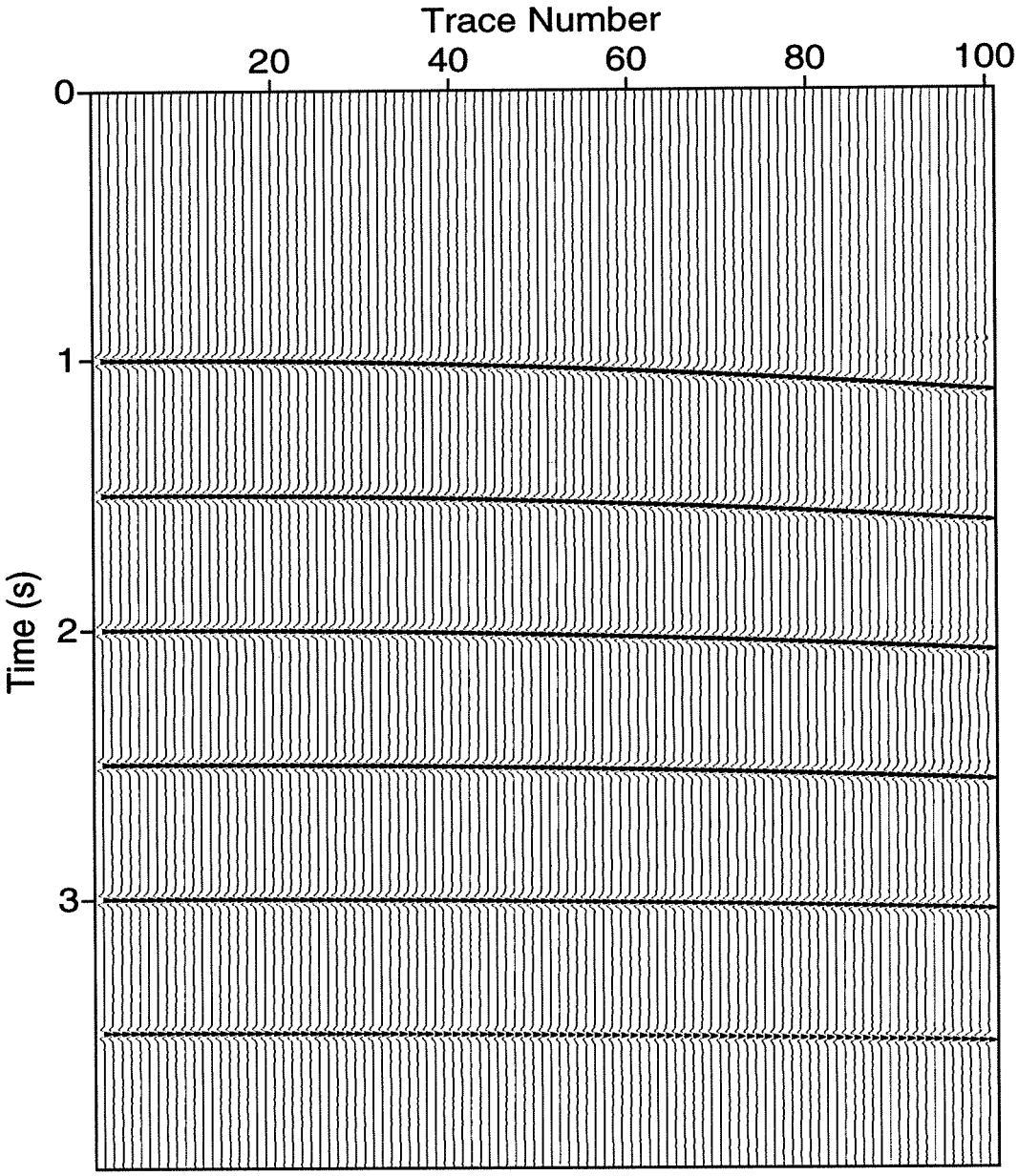

FIG. 5.16. Multiples removed the presence of ground-roll using the hybrid τ-p filtering and signal-noise separation algorithm. Compare with Figure 5.10 for data without ground-roll.

presence of the ground-roll could compromise the quality of the primary velocity estimation, thereby perhaps jeopardizing the performance of the algorithm.

Figure 5.16 shows the multiples removed with this hybrid algorithm. Again, this result is almost the same as that in Figure 5.10 for the data without the ground-roll. Figure 5.17 shows the removed ground-roll with almost no energy from either primaries or multiples.

In summary, then, I have demonstrated, with this example, that the performance of either method for multiple suppression is about the same with or without ground-roll. The same will happen with any other type of noise whose moveout is sufficiently different from that of the primaries or the multiples.

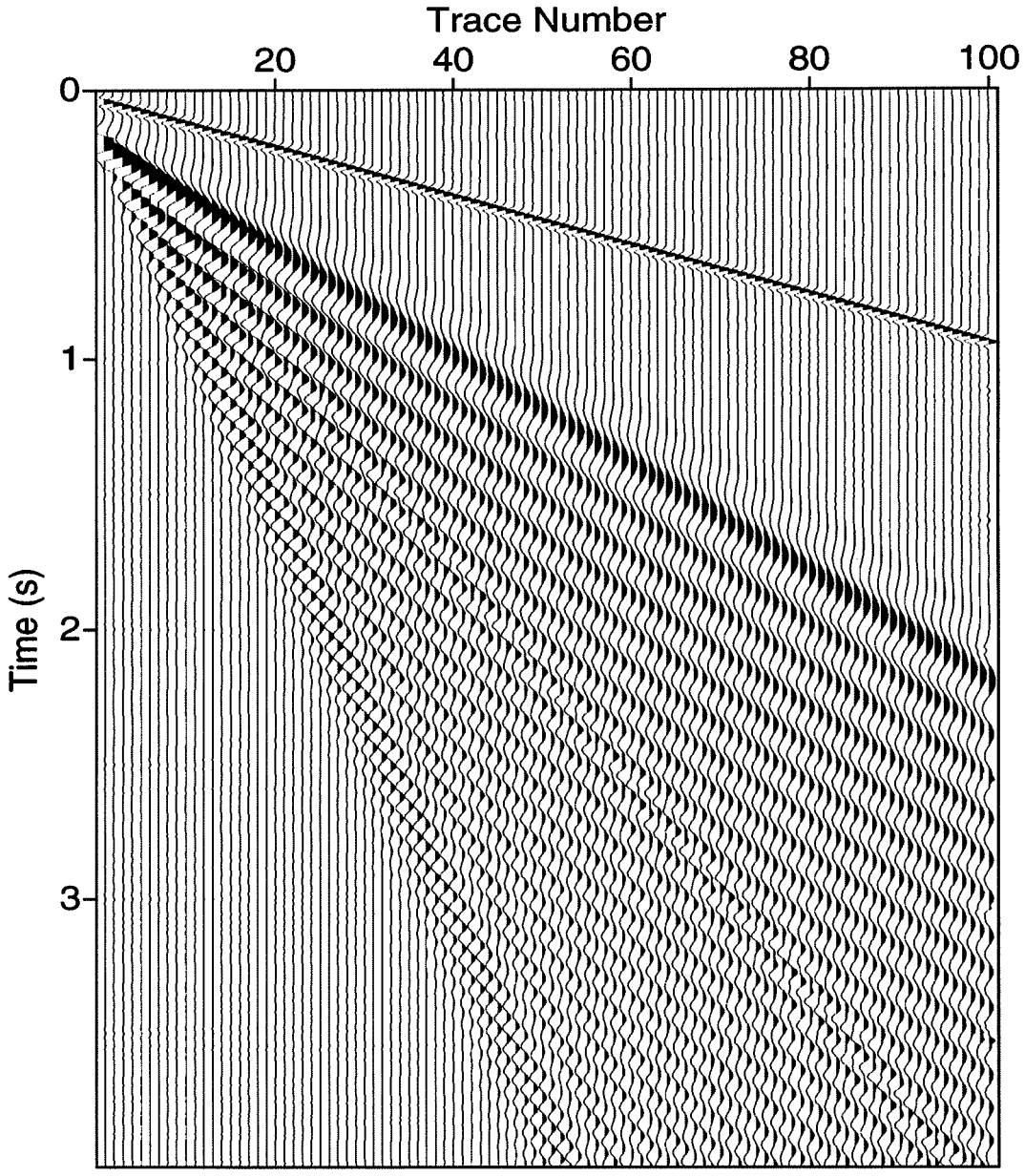

FIG. 5.17. Ground-roll removed with the hybrid τ-p filtering and signal-noise separation algorithm. Some weak residual energy from signal and multiples can be seen.

Chapter 6

IMPLICATIONS OF MULTIPLE SUPPRESSION FOR AVO ANALYSIS AND CMP-STACKED DATA

In Chapter 5, the removal of strong multiple energy in the presence of weak primary energy was accomplished with three different methods: F-K filtering, Hampson's τ-p filtering, and a hybrid approach combining Hampson's τ-p filtering with Harlan's signal-noise separation. Here, I study more closely the relative performance of these approaches to multiple suppression in terms of the preservation of signal amplitude with offset, a primary reason for attempting to suppress multiples on unstacked data, and of the signal-to-noise enhancement as seen after CMP stacking, which itself is powerful in suppressing multiples. In order to do this, I created the four model datasets described in the next section.

6.1 Description of the test model datasets

All test model datasets in this chapter have 100 traces with 1000 samples each, at 4-ms sample interval and 30-m trace interval. Each dataset consists of four primary reflections at zero-offset times of 1.0, 1.9, 2.5 and 3.0 s and corresponding velocities of 2500, 3000, 3500 and 5000 m/s, and four multiple reflections at zero-offset times of 1.0, 2.0, 2.5 and 3.1 s and corresponding velocities of 2300, 2600, 2800 and 3000 m/s. The differences between the four datasets are aimed at emphasizing the influence of multiple-to-primary amplitude ratio and variation of amplitude and polarity with offset on the implications that each multiple-suppression method has for AVO analysis and the output of CMP stacking.

6.1.1 Test model dataset 5a

Test model dataset 5a has no trace-to-trace amplitude variation (prior to NMO-correction) and a multiple-to-primary amplitude ratio of 4:1, and Figure 6.1 shows this dataset after an NMO-correction has been applied with the velocity of the primaries. Figure 6.2 shows its corresponding parabolic τ-p transform. For such strong multiples a substantial amount of imperfectly focused energy from the multiples is mapped to the same moveout region as that of the primaries. This is clearly seen as the amplitudes below the primary events in Figure 6.2. The vertical dashed line represents the maximum p-value that is to be fully preserved in the data. The slanted dashed line represents the minimum p-value to be completely suppressed from the data. A linear amplitude taper is applied between the two.

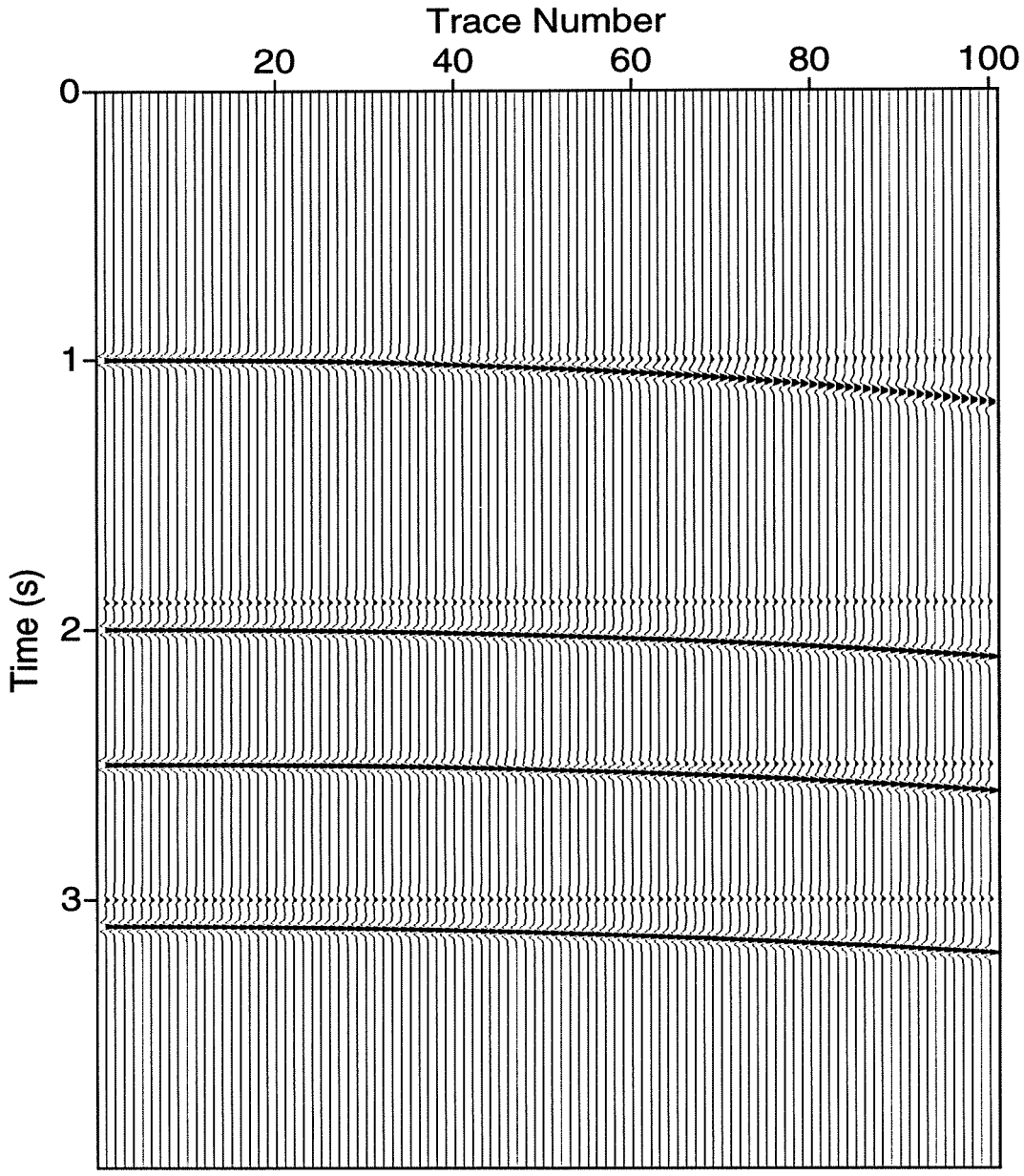

FIG. 6.1. NMO-corrected modeled test dataset 5a. The amplitude of the multiples is four-times that of the primaries. Neither signal nor multiple has any trace-to-trace amplitude variation.

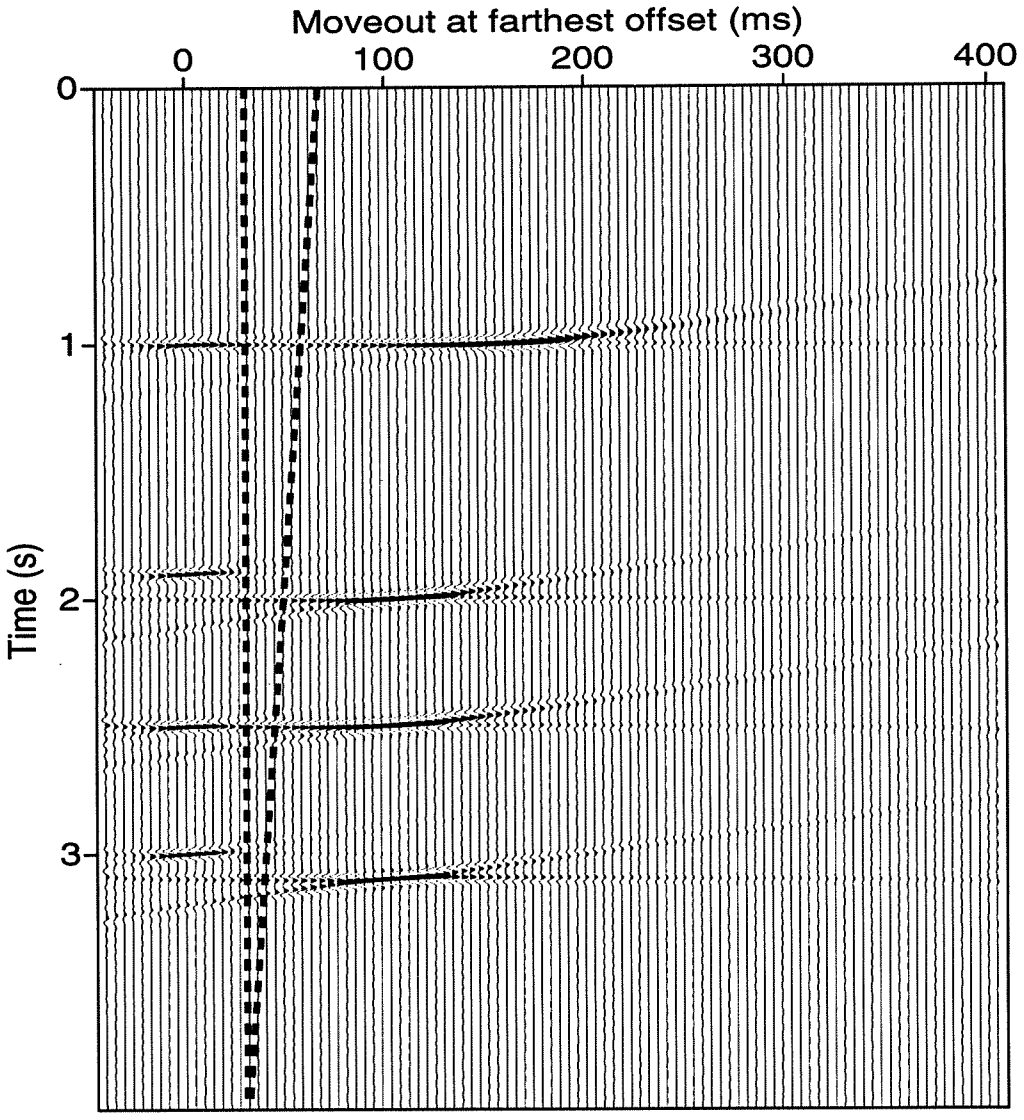

FIG. 6.2. τ-p transform of NMO-corrected modeled test dataset 5a. Because of the relative strength of the multiples, substantial residual energy from the multiples is mapped to the same slope region as the primaries. The vertical dashed line represents the maximum p-value fully preserved in the data. The slanted dashed line denotes the minimum p-value completely suppressed. A linear amplitude taper is applied between the two.

6.1.2 Test model dataset 5b

This dataset is the same as the previous one except that the multiple-to-primary amplitude ratio is 1:1. This model, together with the previous one, is intended to assess the influence of the multiple-to-primary amplitude ratio on the relative performance of the algorithms for both AVO analysis and CMP stacking. The model dataset is shown in Figure 6.3, and Figure 6.4 shows the corresponding parabolic τ-p transform. The 1:1 amplitude ratio between the primaries and the multiples makes them both appear with the same degree of focusing in the τ-p domain. Thus, with this lower level of multiple energy, the primary zone in the τ-p domain is less severely contaminated by imperfectly focused multiples. The dashed lines have the same meaning as in Figure 6.2.

6.1.3 Test model dataset 5c

This dataset is the same as that in Figure 6.1 (4:1 multiple-primary amplitude ratio), except that a linear amplitude variation with offset is introduced. The amplitude of any given event at the farthest offset is only one half its amplitude on the nearest offset. The model dataset is shown in Figure 6.5. Due to the angular dependence of reflection coefficients, geometric spreading, and array-directivity, we can expect signal amplitude to vary with offset, although not of course in this simplistic manner. Likewise, we certainly would not expect the offset dependence for late reflections to be the same as that for early reflections. Nevertheless, these tests should mimic major issues related to amplitude variation with offset.

In the corresponding parabolic τ-p transform (Figure 6.6), note the weak decrease of focusing power caused by the variation in amplitude, (compare with Figure 6.2, the case with no amplitude variation with offset).

6.1.4 Test model dataset 5d

This dataset (Figure 6.7) is the same as in Figure 6.1 except that the polarity of events changes at larger offsets as may happen for reflections from thin beds in elastic media. Polarity reversals arise, for example, when the elastic parameters of the layers change strongly across an interface. Rüger (1995) has given an equation that shows relative values of P- and S-wave velocities and densities for which phase changes and polarity reversals can be present in a CMP gather. The model used to describe this situation here is extreme because the change of polarity with offset can be expected for wider angle reflections than those typically used in AVO analysis. Certainly, the offsets at which polarity can be expected to change will typically increase with increase in reflector depth. This model, therefore, is more severe than those that can be expected for field data.

The parabolic τ-p transform of the data in Figure 6.7 is shown in Figure 6.8. The presence of the polarity reversal has seriously decreased the focusing of both primaries and multiples (compare with Figure 6.2 for model dataset 5a) to the point that the primaries are virtually unfocused. The reason is that the parabolic transform describes the data in terms of events with parabolic moveout whose amplitude and phase are independent of offset.

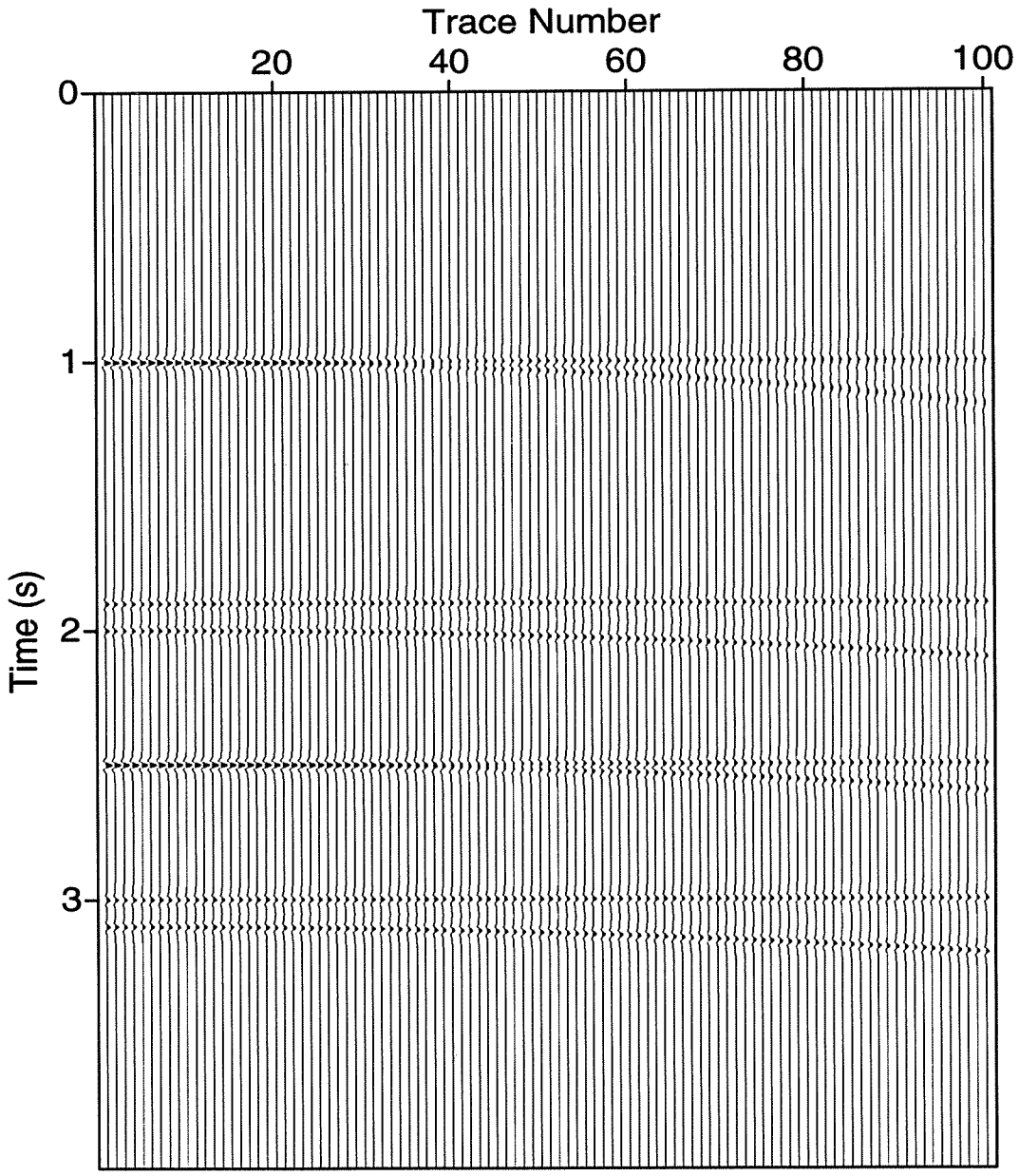

FIG. 6.3. NMO-corrected modeled test dataset 5b. The amplitude of the multiples is the same as that of the primaries. Amplitudes again do not change with offset.

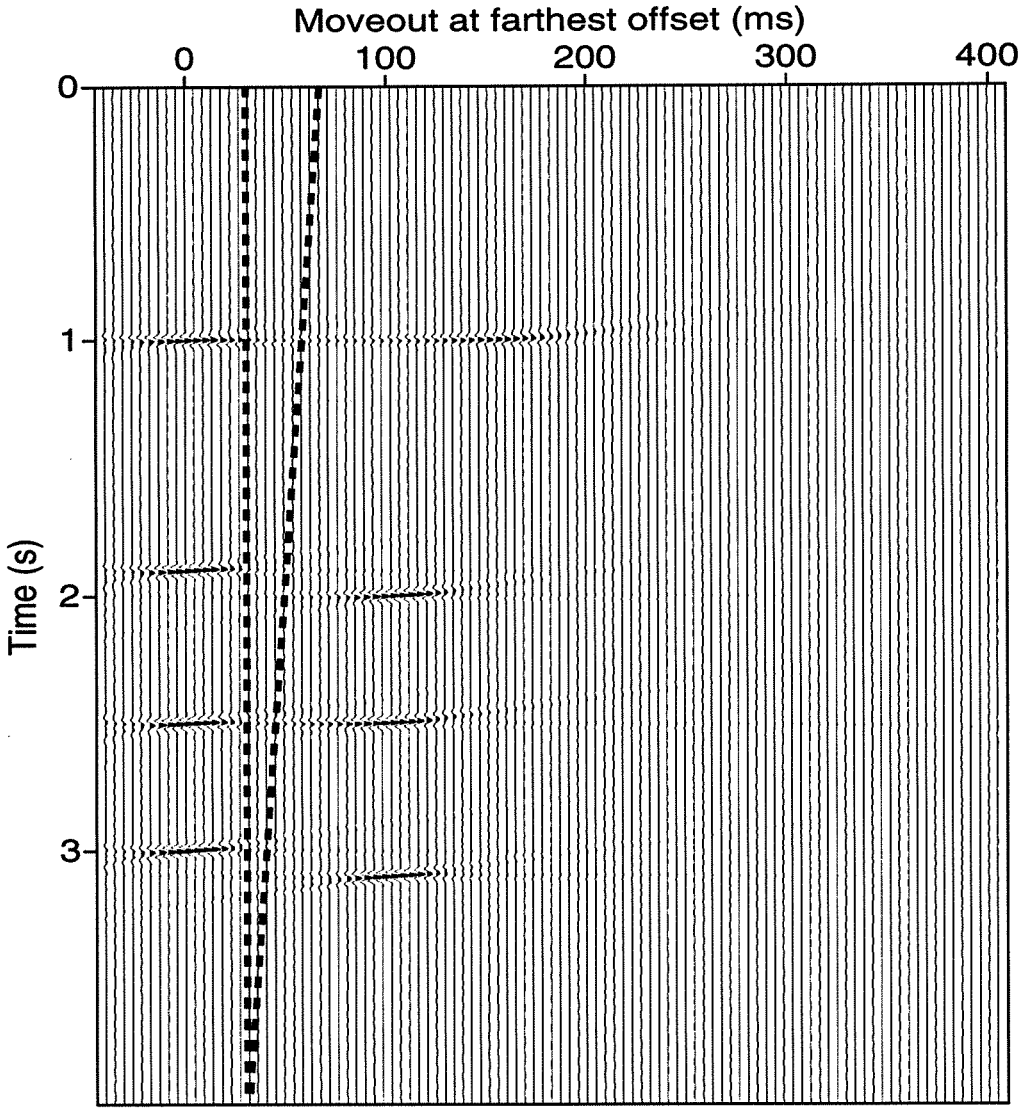

FIG. 6.4. τ-p transform of NMO-corrected modeled test dataset 5b.

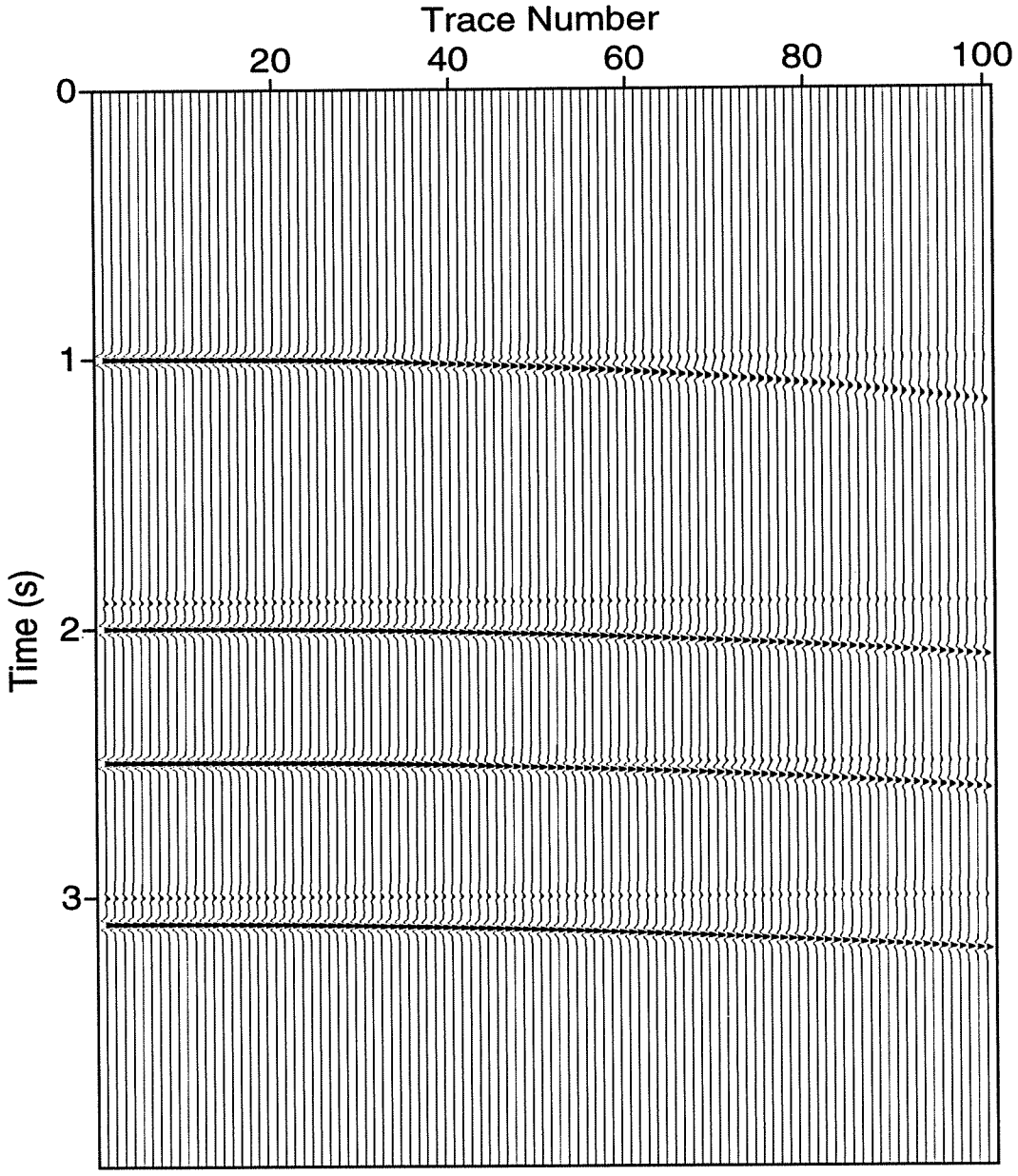

FIG. 6.5. NMO-corrected modeled test dataset 5c. The amplitude of the multiples is four times that of the primaries, and amplitudes decrease linearly with offset such that the amplitudes at the far offset are half those at near offset.

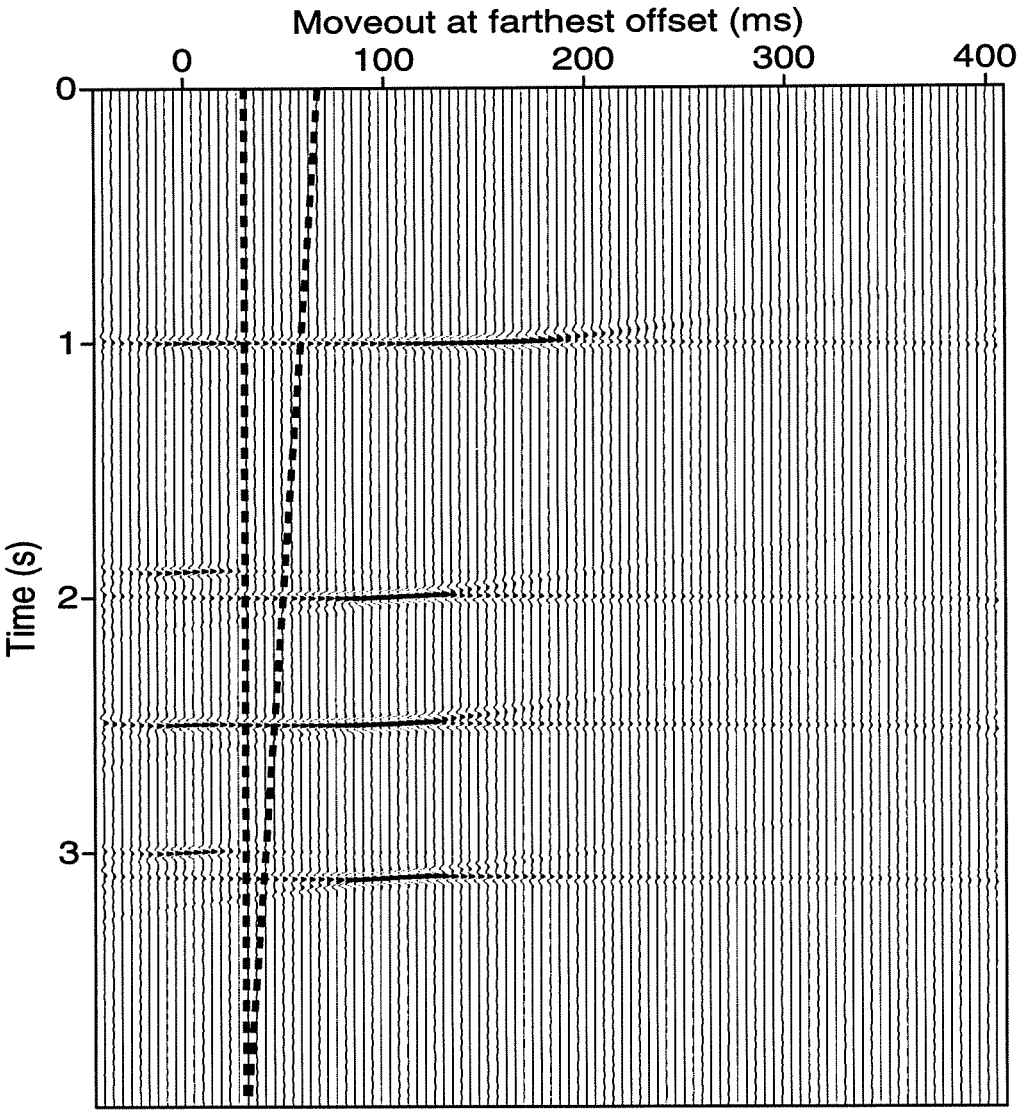

FIG. 6.6. Parabolic τ-p transform of NMO-corrected modeled test dataset 5c. As in Figure 6.2, strong unfocused energy from the multiples appears into the primary zone.

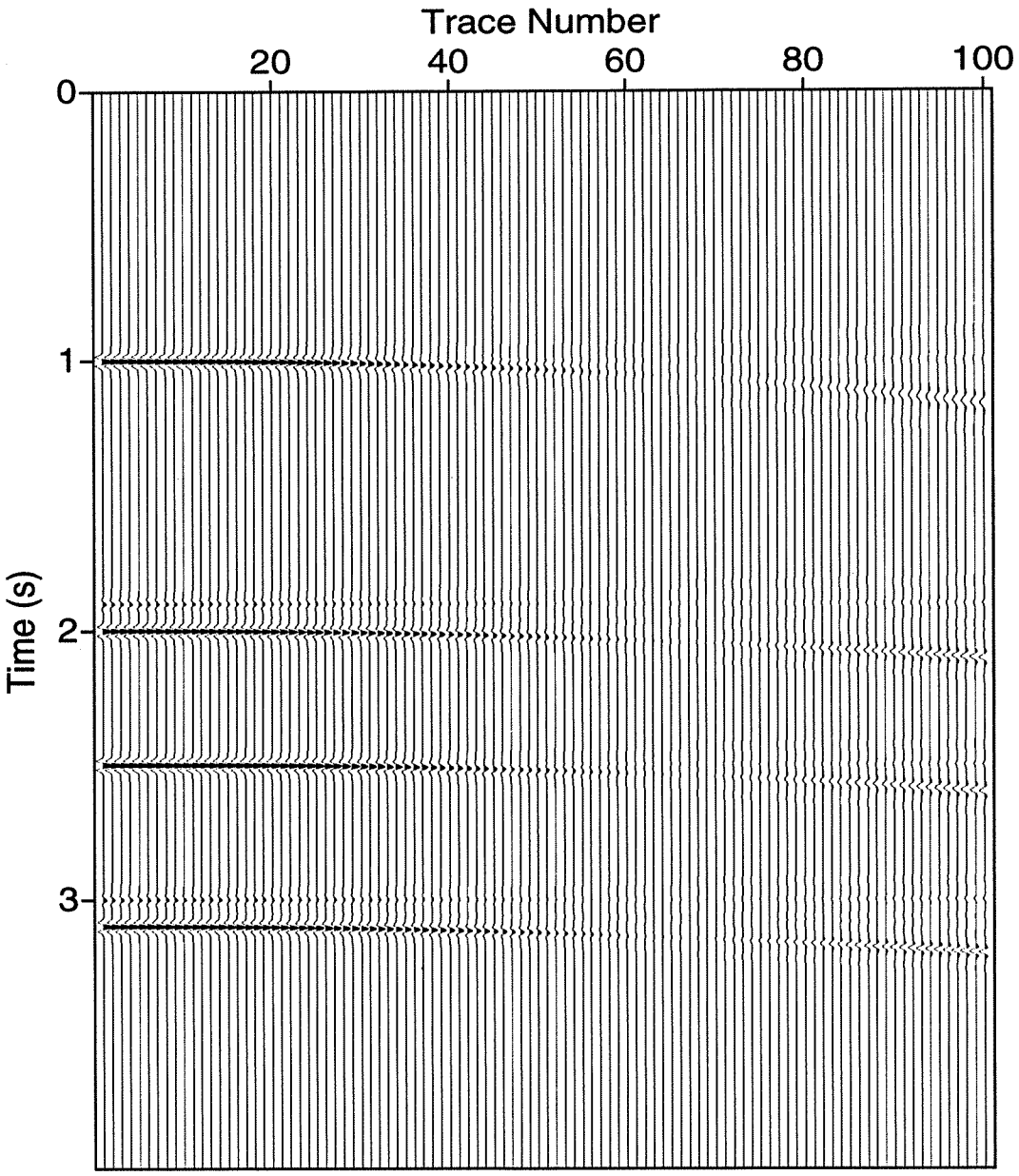

FIG. 6.7. NMO-corrected modeled test dataset 5d. The amplitude of the multiples is four times that of the primaries, and amplitude decreases linearly with offset so that the amplitudes on the last trace are -0.5 times those on the first trace.

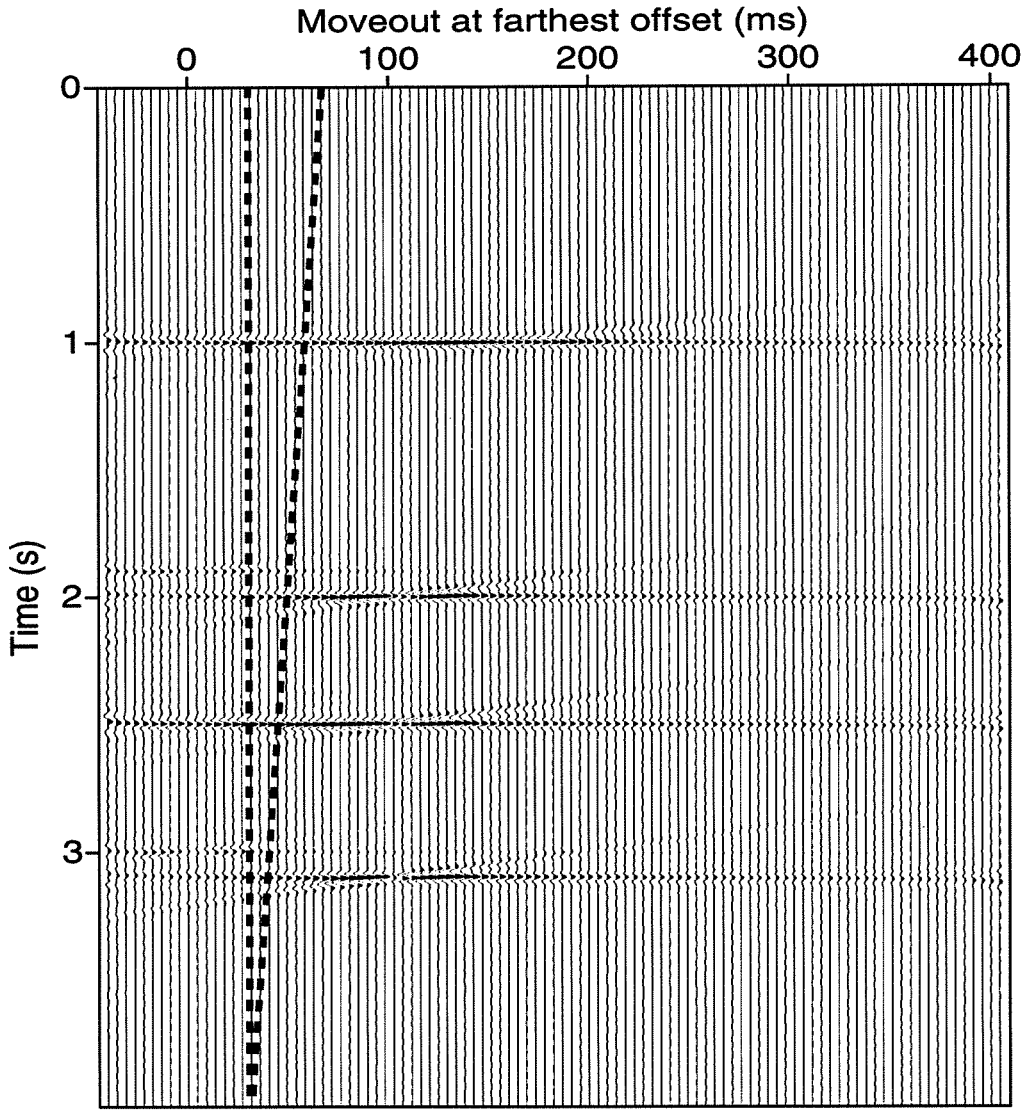

FIG. 6.8. τ-p transform of NMO-corrected modeled test dataset 5d. The introduced polarity reversal has decreased the focusing of both the primaries and the multiples.

6.2 Multiple suppression for the four datasets

In this section, I show the result of applying each of the three methods for multiple suppression described in the previous chapter to each one of the four modeled test datasets described above. The results are presented separately for each dataset.

6.2.1 Model dataset 5a

Figure 6.9 shows the primaries extracted with the F-K filtering method for dataset 5a (primary-to-multiple amplitude ratio of 4:1 and constant amplitudes for all traces). A substantial part of the multiples remains. In particular, because moveout differential between multiples and primaries approaches zero for small offset, multiple energy is not well suppressed at zero offset. Where the multiples and primaries have the same zero-offset time, the residual multiple energy gives the primaries the appearance of having amplitude that varies strongly with offset. This is more clearly shown in Section 6.3.

Figure 6.10 shows the primaries extracted with Hampson's approach. The earliest multiple was well removed because it and the primary exhibited enough differential moveout at the same zero-offset time to allow a clear separation between the two. The multiples at times 2.0, 2.5 and 3.1 s were not so clearly removed because their differential moveout was smaller. More specifically, the tails of the multiples that overlap the primary region in τ-p space are associated with the smaller moveouts on the short-offset traces. The influence of the multiple at 2.5 s is seen as an increase in apparent amplitude of the primary for the short-offset traces.

Figure 6.11 shows the primaries extracted with the hybrid approach of Chapter 5. The hybrid approach has helped further remove the residual multiple energy, although at the expense of weakening the extracted primaries, as a result of the algorithm's effort to produce the stronger degree of residual multiple rejection. See, in particular, those primaries at 1.9 and 3.0 s. The trade off between added multiple suppression and loss of signal is controlled by the reliability threshold used to discriminate between signal and noise on the assumption that high amplitudes represent well-focused signal. The degree and importance of this weakening of the primaries will be quantified in Section 6.3.

6.2.2 Model dataset 5b

Figure 6.12 shows the primaries extracted with the FK filtering approach. Because of the linearity of the FK filtering operation, the strength of the multiple-to-primary amplitude ratio does not influence the performance of the algorithm (compare the extracted primaries at 1.9 and 3.0 s in Figures 6.9 and 6.12). The ratio of the amplitudes of the residual multiples in Figures 6.9 and 6.12 at any given offset is the same as that on the corresponding input datasets. Thus, while the extracted primaries are again contaminated with residual multiple energy, since the multiples are much weaker in this dataset, offset-dependent contamination of primaries by residual multiple energy is less severe than that for dataset 5a.

Figure 6.13 shows the primaries extracted with Hampson's τ-p filtering approach. The muting operation in this dataset is the same as in the previous dataset, so because of the

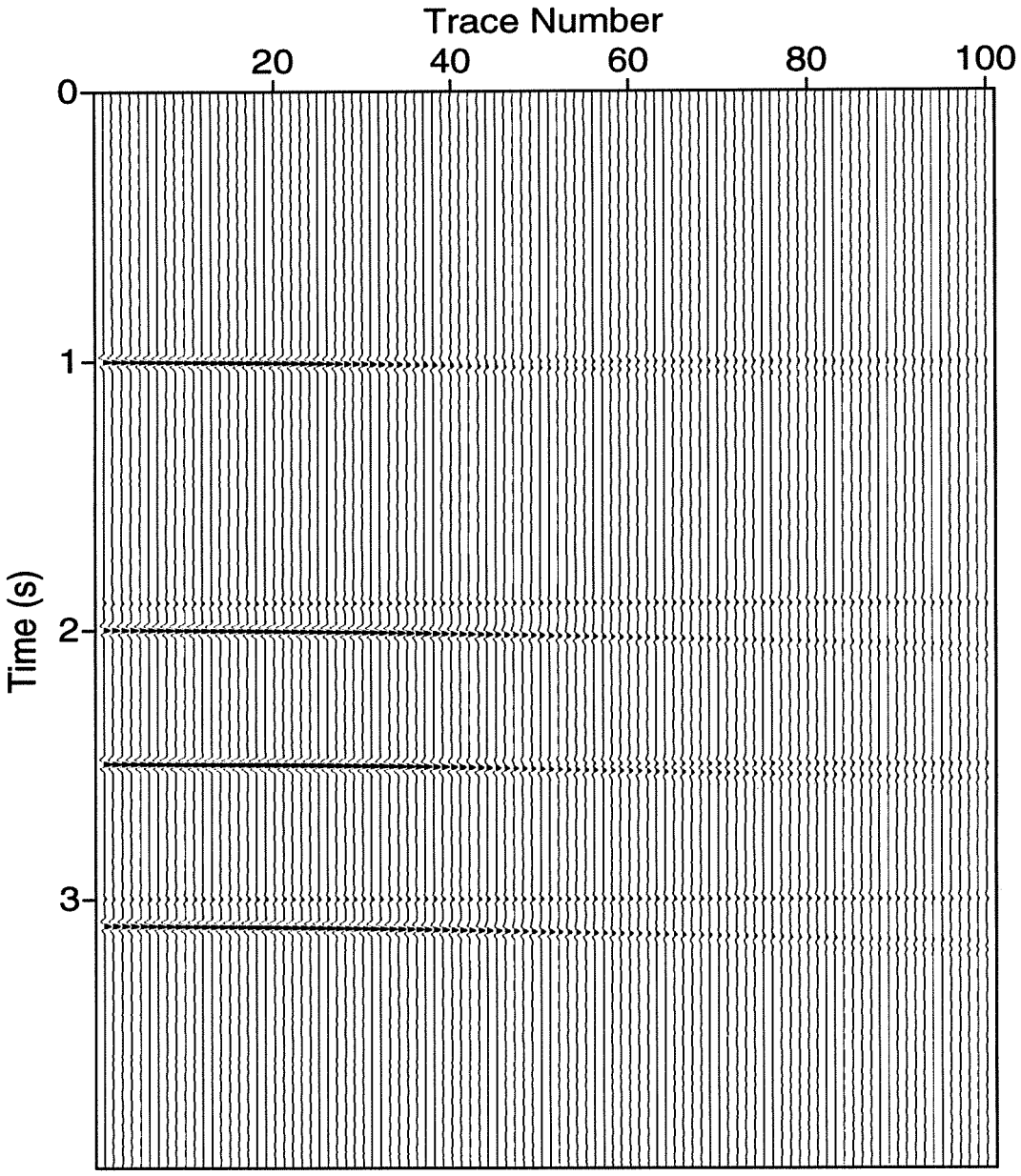

FIG. 6.9. Primaries in model dataset 5a extracted with the F-K filtering approach. The substantial residual energy from the multiples, which is much better suppressed on the larger offset traces, makes the amplitude of the interfered-with primary events appear to change with offset as well.

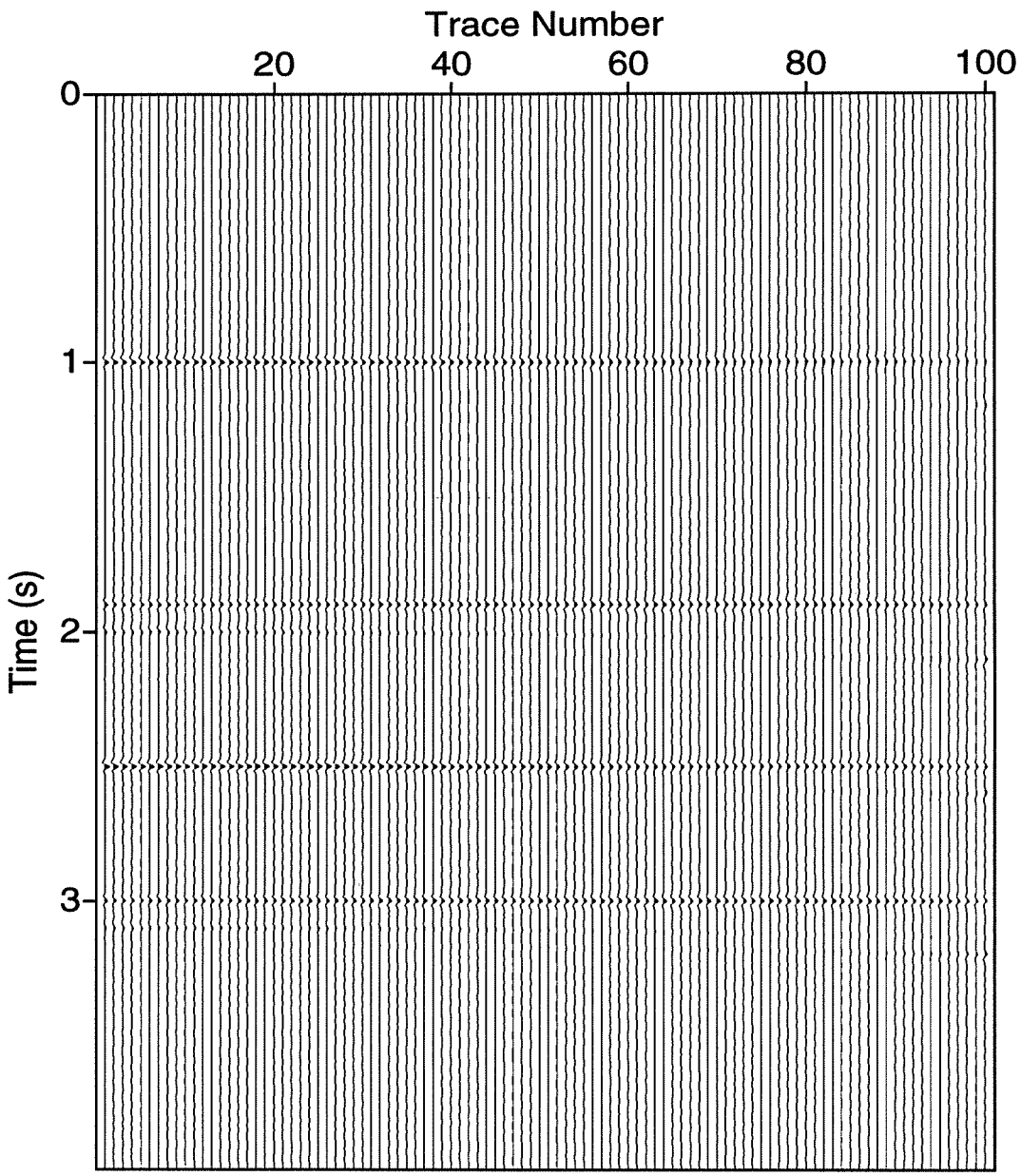

FIG. 6.10. Primaries in model dataset 5a extracted with Hampson's approach. Some residual energy from the multiples is clearly seen on the edges of the section, but here the multiple energy is weak enough that the amplitudes of primaries do not seem to change so much with offset.

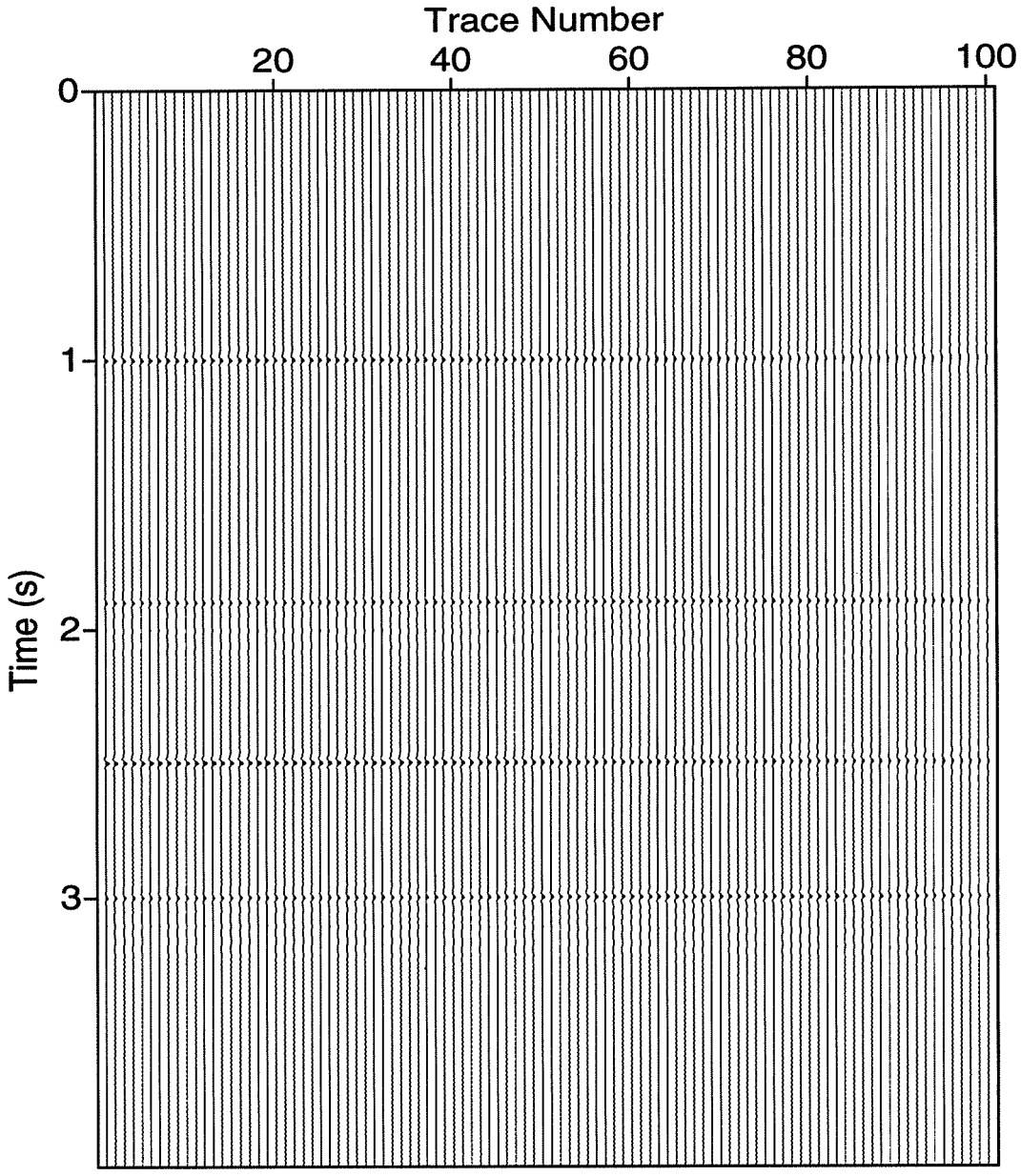

FIG. 6.11. Primaries in model dataset 5a extracted with the hybrid approach of Chapter 5. Much of the residual multiple energy is gone. The amplitudes of the extracted primaries have been weakened, in particular for the primaries at 1.9 and 3.0 s.

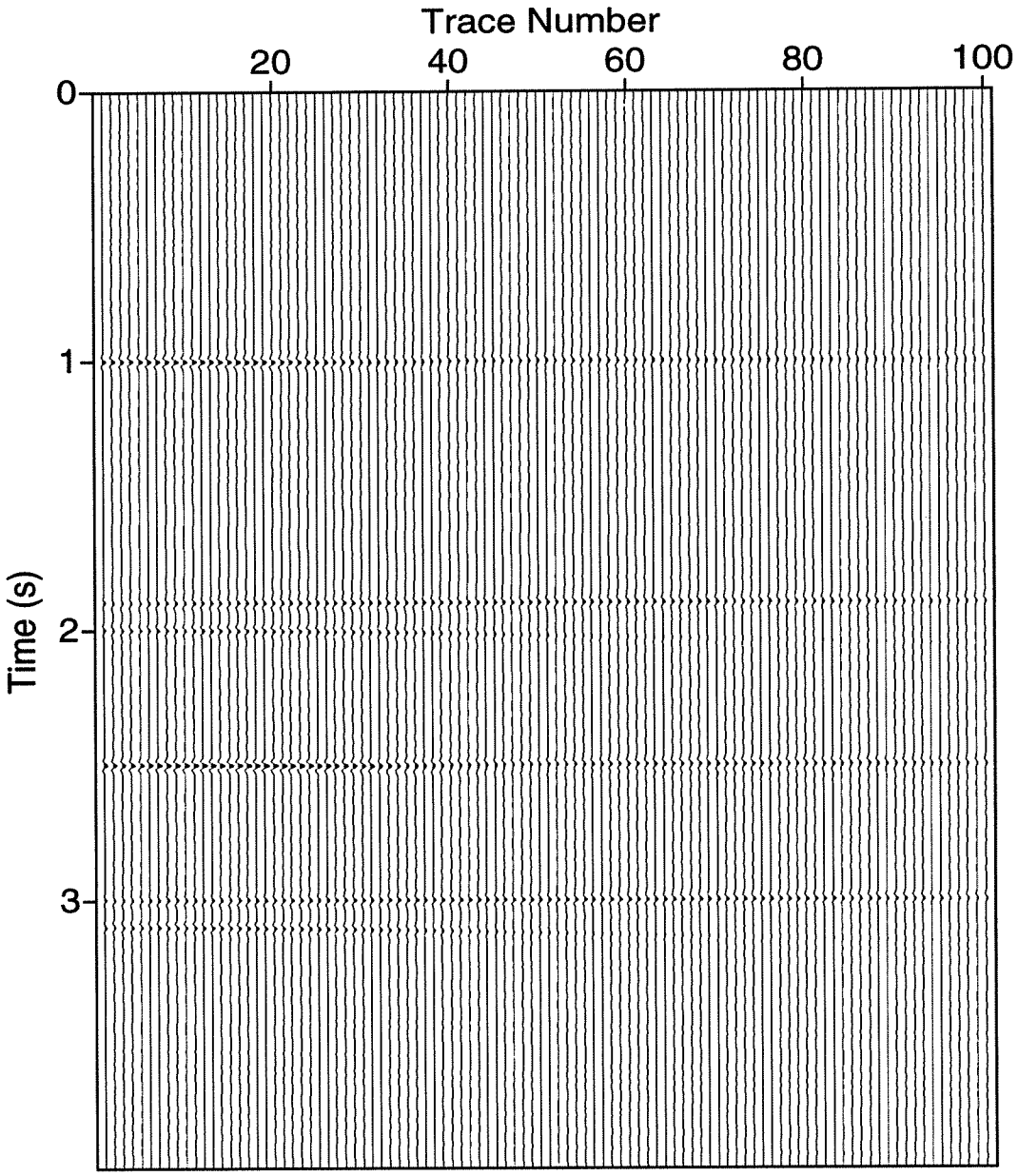

FIG. 6.12. F-K extracted primaries for model dataset 5b. As a consequence of the linearity of the FK filtering operation, the relative action on signal and multiples is identical as that in Figure 6.9.

linearity of the τ-p transform, relative treatment of primaries and multiples is the same as for dataset 5a. Again, due to the lower multiple-to-primary amplitude ratio on input, less multiple energy can be seen as compared with Figure 6.10.

Figure 6.14 shows the primaries extracted with the hybrid approach of Chapter 5. This time, no residual energy from the multiples appears with the extracted primaries, although the primaries again show weaker amplitudes. Potentially, the hybrid algorithm provides an additional degree of multiple rejection by discriminating between the focused primary energy and the residual multiple energy. Since the multiple energy was much weaker for this than for the previous dataset, virtually no residual multiple energy can be seen in the extracted primaries. The price for the extra degree of multiple rejection over and above that obtained by Hampson's method, is again a weakening of the amplitudes of the extracted primaries. In section 6.3, where I compare output signal-to-multiple ratio, rather than separate treatment of signal and multiples for the various processes, I show that the improvement in primary-to-multiple amplitude ratio with this method is about the same for both datasets 5a and 5b.

6.2.3 Model dataset 5c

Figure 6.15 shows the primaries extracted with the FK filtering approach. The linear decrease of amplitudes with offset introduced in this model (recall that the amplitudes on the last trace are only 0.5 those on the first trace) can be seen in the extracted primaries at 1.9 and 3.0 s, which are not coincident with multiples at zero-offset time. The residual multiple energy, however, makes the amplitudes in the short-offset traces of the other two events appear much stronger than are the true signal amplitudes, as was the case for model dataset 5a in Figure 6.9.

Figure 6.16 shows the primaries extracted with Hampson's approach. Comparison with Figure 6.10 shows that the primary extraction is about as good with this model as it was with the constant-amplitude model. The linear amplitude variation with offset introduced in this model apparently has posed no additional strain in the ability of Hampson's approach to identify and extract the primary energy. Although the focusing of both primaries and multiples in the τ-p domain is not so strong as with dataset 5a, no additional overlapping occurs (compare Figures 6.2 and 6.6) between the focused primaries and multiples, so the mute operation works well. Somewhat surprising, considering the substantial amplitude variation across the traces in dataset 5c, the residual multiple on the smaller-offset traces is comparable with that in Figure 6.10.

Figure 6.17 shows the primaries extracted with the hybrid approach of Chapter 5. The residual multiple energy has been reduced considerably, but not so much as in Figure 6.11, for which input data had no amplitude variation with offset. The reason is that the weaker amplitude in the far offset traces of the primaries reduces the degree of focusing of the primary energy (compare Figure 6.6 with Figure 6.2 for the constant amplitude case). The focusing of the multiples is also weaker in Figure 6.6 than it was in Figure 6.2 but the relative strength of the residual multiples is about the same in both figures. This combination makes it more difficult for the signal-noise algorithm to discriminate between the primary and the

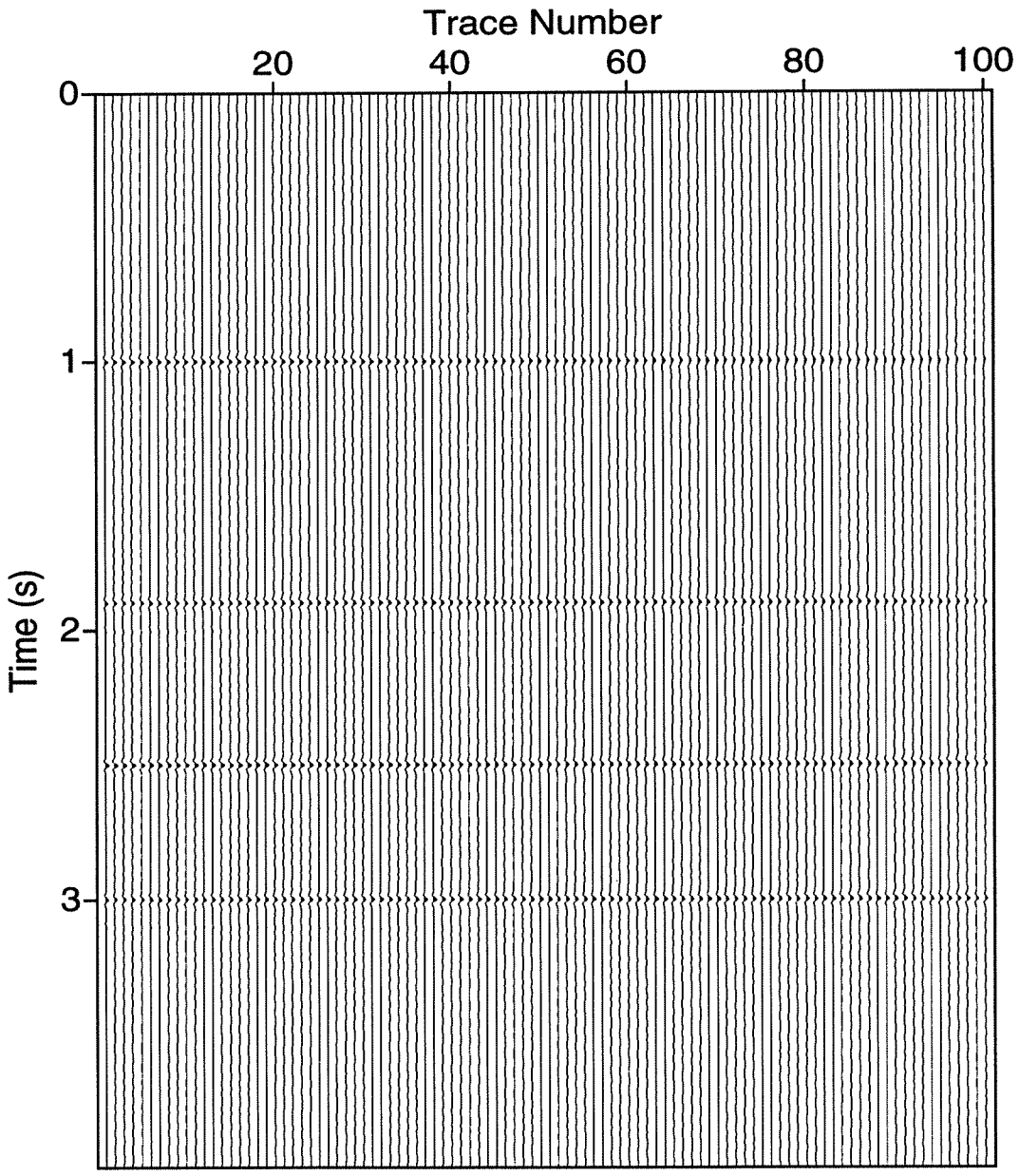

FIG. 6.13. Primaries in model dataset 5b extracted with Hampson's approach. Only a small residual from the multiples can be seen.

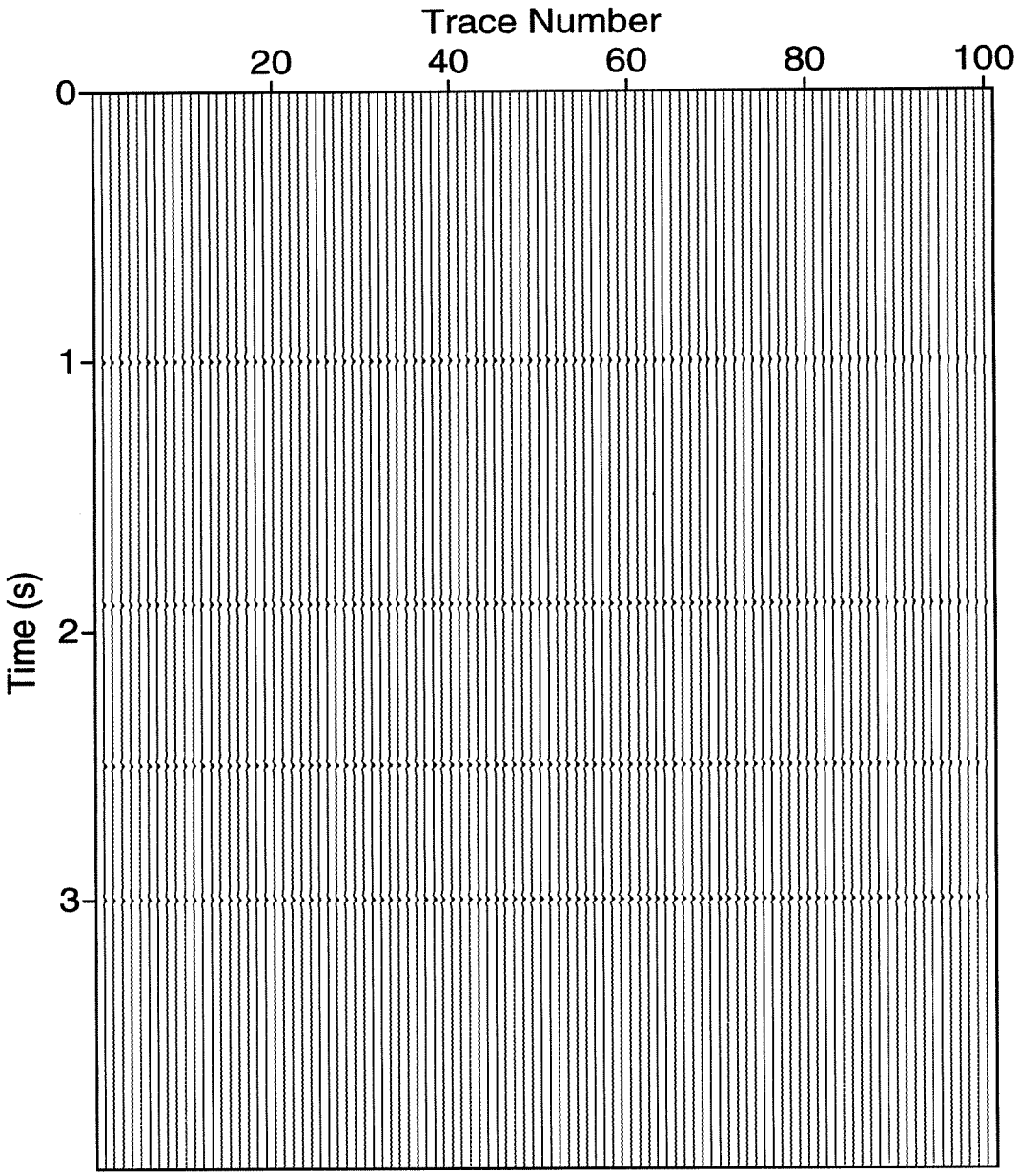

FIG. 6.14. Primaries in model dataset 5b extracted with the hybrid approach of Chapter 5. This time, the algorithm eliminated virtually all residual multiple energy. Weakening of the primaries (compare with Figure 6.3) has been the price.

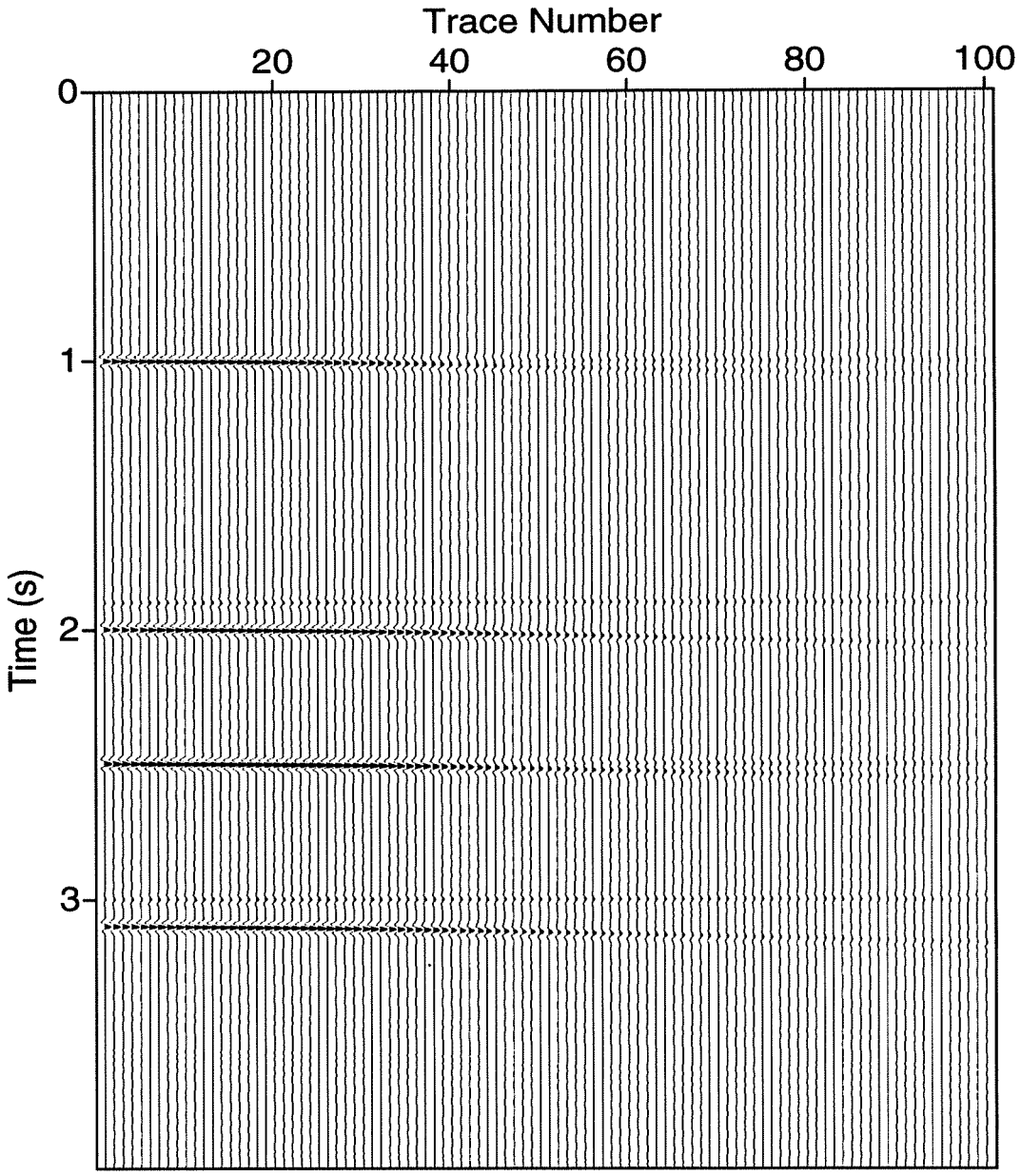

FIG. 6.15. F-K extracted primaries for model dataset 5c. The results are similar in level of multiple rejection to those for dataset 5a (Figure 6.9) despite the linear decrease of amplitude with offset.

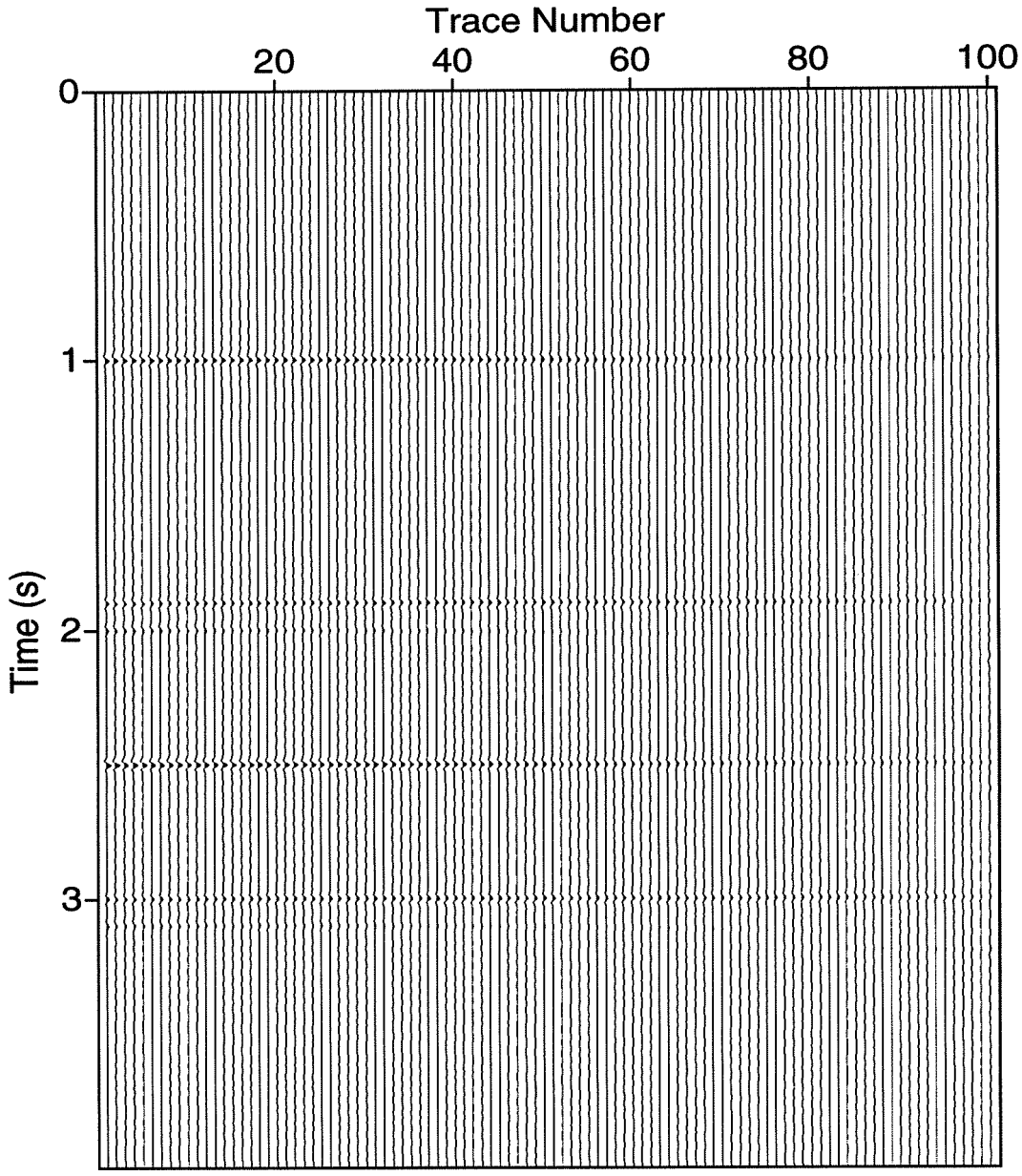

FIG. 6.16. Primaries in model dataset 5c extracted with Hampson's approach. Residual energy from the multiples can be seen at the short offsets. The linear decrease of primary amplitude with offset has been largely honored.

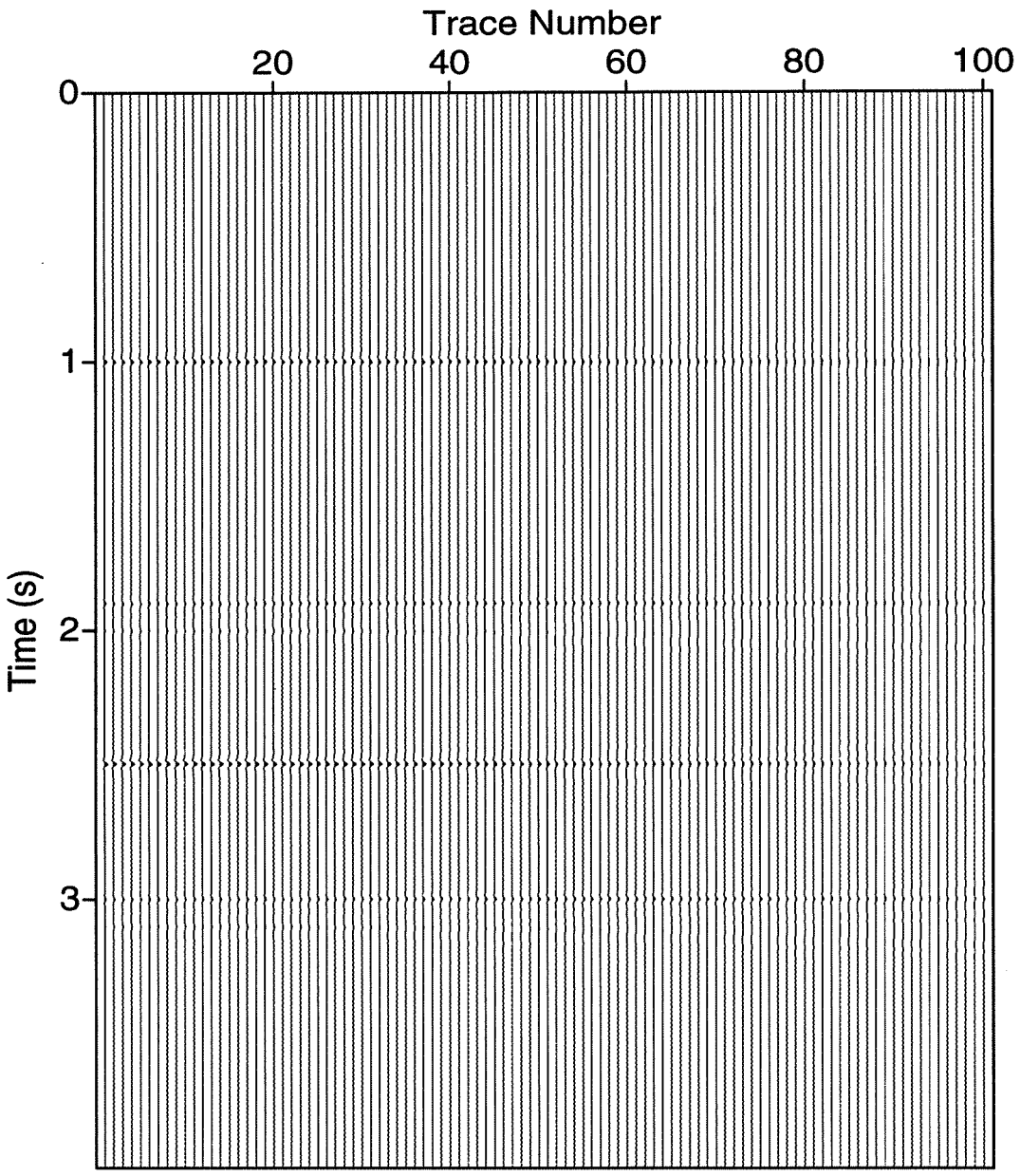

FIG. 6.17. Primaries in model dataset 5c extracted with the hybrid approach of Chapter 5. The multiple energy has been reduced but not eliminated. Signal amplitudes still show the decrease of amplitude with offset, but overall the primary amplitudes on all traces have been reduced.

residual multiple energy, which results in a lesser degree of multiple rejection and an increased weakening of the extracted primaries, compared with the results in Figure 6.11.

6.2.4 Model dataset 5d

Figure 6.9 shows the primaries in dataset 5d (the data with a polarity reversal), extracted with the FK filtering algorithm. As seen in the primaries that are not contaminated by multiples (those at 1.9 and 3.0 s), while the amplitudes of the primaries are weakened, and, as with all previous datasets, too much energy from the multiples remains.

Figure 6.19 shows the primaries extracted with Hampson's method. Although again better than that with the F-K filter method, the degree of multiple suppression is less for this dataset than it was for dataset 5a (compare with Figure 6.10). Recall that the two datasets are identical except for the introduced polarity reversal. The multiples are stronger, which can be seen in particular at 2.0 and 3.1 s. To make things worse, the amplitude of the extracted primaries is weaker because of the polarity reversal, as can be seen at 1.9 and 3.0 s. The reason for these results lies in the comparison between Figures 6.2 and 6.8, which shows that the polarity reversal resulted in much energy from the multiples being smeared into the primary zones, thereby making it difficult for the algorithm to separate multiples and primaries.

Finally, Figure 6.20 shows the primaries extracted with the hybrid algorithm of Chapter 5. Comparison of Figures 6.20 and 6.19 shows that the refinement of the signal-noise separation to extract the residual multiple energy did almost nothing beyond the action of the Hampson method. The reason for this result lies again in the comparison between the τ-p transforms of the data with and without the polarity reversal (Figures 6.8 and 6.2), as explained in the previous paragraph.

6.3 AVO implications of multiple suppression

Having qualitatively compared the relative performance of each of the three methods for multiple suppression with each of the test datasets, here I quantitatively compare their relative implications for AVO analysis. In order to do this, I plot the amplitude of each of the extracted primary reflections as a function of offset for every model and every multiple removal method. The amplitudes are measured as the peak of the wavelet at the two-way traveltime corresponding to each primary (which, at times, may be contaminated by residual multiple). In the plots below, a solid black line indicates the amplitude of a primary (in the absence of multiples) in the input data. Any departure from this curve for the extracted primaries, indicates amplitude variation with offset introduced by the multiple-suppression algorithm. That variation could be a combination of the amplitude variation of the extracted primary and contamination from residual multiples. A finely dotted line indicates the amplitude of the primary in the input data including the presence of the multiples. For primaries not coincident with multiples, the two curves are the same. A dashed black line represents the primary extracted with the FK filtering algorithm, a solid gray line the primary extracted with Hampson's approach, and a dashed gray line the primary extracted with the hybrid

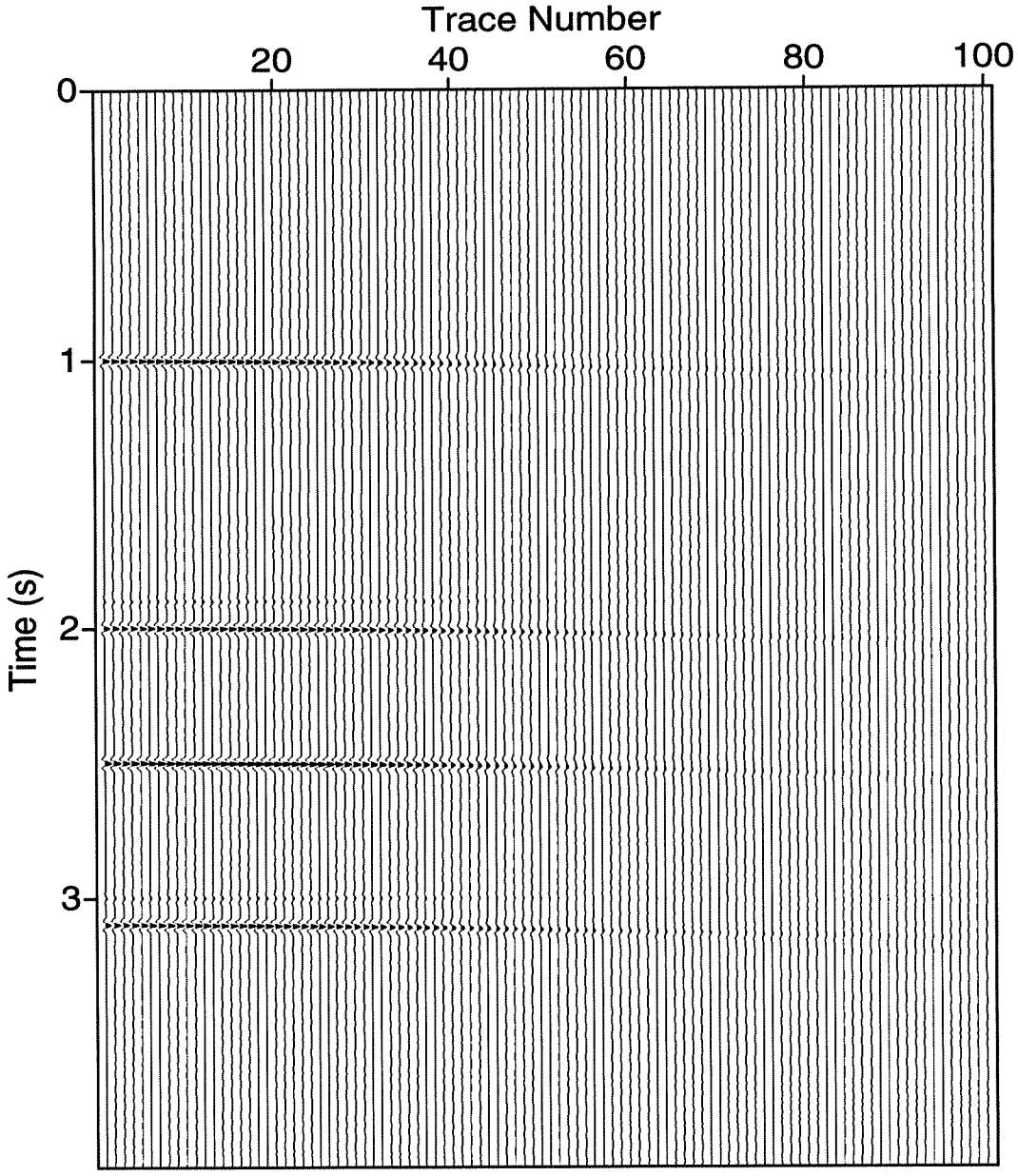

FIG. 6.18. F-K extracted primaries for model dataset 5d. The extraction has been unsuccessful for this dataset. The residual multiple energy is stronger than the primaries.

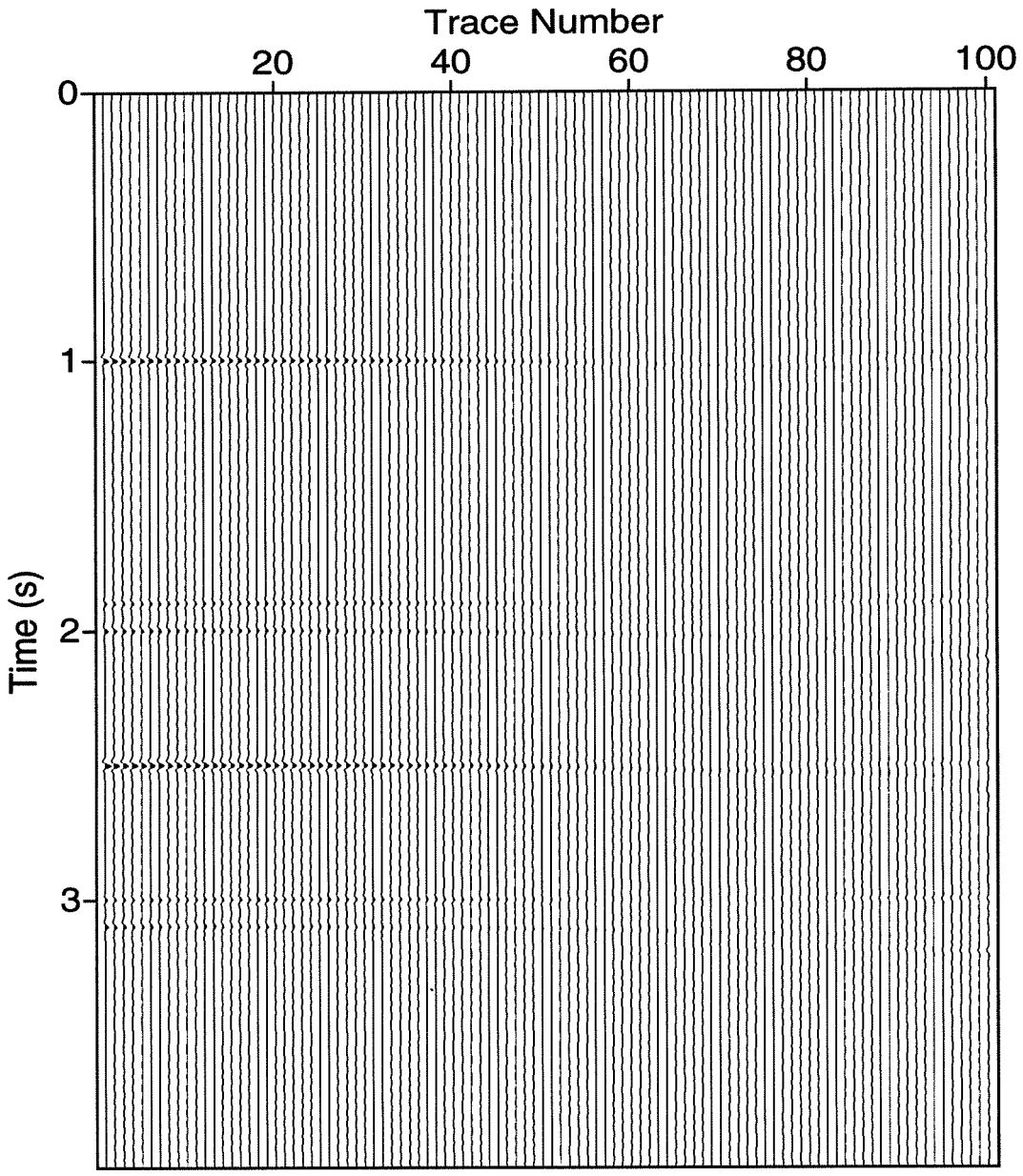

FIG. 6.19. Primaries in dataset 5d extracted with Hampson's approach. For this dataset, although the signal-to-multiple ratio was improved over that of the input data, the extraction of the primaries was not nearly as successful as in the previous datasets; strong energy from the multiples remains and the primaries have been weakened (compare with Figure 6.7).

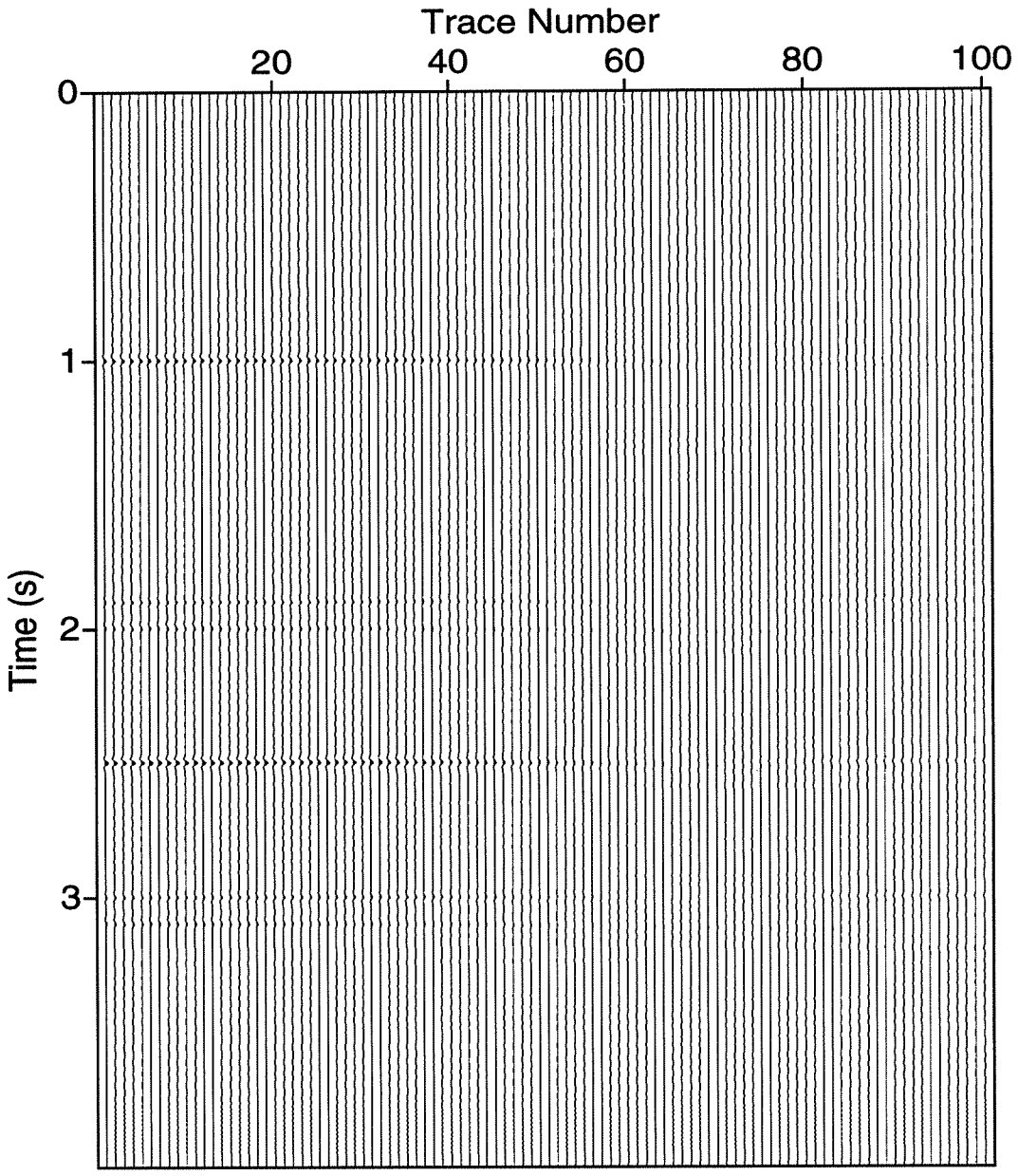

FIG. 6.20. Primaries in model dataset 5d extracted with the hybrid approach of Chapter 5. For this dataset, the method did little to increase the rejection of the multiples over and above the performance of Hampson's approach.

approach of Chapter 5. A dashed vertical line indicates the offset that equals the depth of the reflector. This line is relevant because normally offsets approaching the depth of a reflector are not included in AVO analysis. Amplitude variations for offsets smaller than the one corresponding to that vertical line, therefore, are the most important ones to consider in this study.

Figure 6.21 shows the four sets of curves described above, corresponding to each of the four primaries in modeled test dataset 5a (multiple-to-primary ratio of 4:1, with amplitudes independent of offset). First, note that where primaries are contaminated by multiples (plots a and c in Figures 6.21 to 6.24), the contamination is so severe for dataset 5a (and for the others as well) that AVO analysis would be hopeless. Because the FK filtering algorithm (dashed black line) has not suppressed the strong multiples sufficiently for the smaller offsets of the primaries that are contaminated by multiples (compare the solid and dotted thin black lines in Figures 6.21a and 6.21c), those primaries show insufficient improvement of measured amplitude on the small-offset traces and thus are again useless for AVO study. For those primaries that are not coincident with multiples (Figures 6.21b and 6.21d), the extracted primary amplitudes are closer to the ideal behavior but are distorted at the very short and long offsets, due to edge effects. Thus, although the distorsion is less severe for the primaries that are not contaminated by multiples, nevertheless the FK algorithm is unsatisfactory where the goal is to analyze AVO behavior.

Hampson's approach (solid gray line) performed well for intermediate offsets for all primaries, but showed departures for long- and, of more importance, short-offsets. Although not as severe as the departures when the FK filter approach was used, the amplitude variations will distort AVO analysis. For the third primary, for example, the departure from the true amplitude is almost 50% in Figure 6.21c. More important, a strong, roughly linear amplitude variation with offset is present for offsets between about 800 to 1600 m. The results in Figures 6.21b and 6.21d indicate that the near-offset primary amplitudes were reduced in the process of suppressing the multiples, and that the increase in amplitude for short offsets for the other two reflections is due to the presence of the residual multiples. For AVO analysis, however, the results are good for a large range of the shorter offsets.

The hybrid approach of Chapter 5 performed the best for the short offsets of the primaries contaminated by the multiples because of the level of residual multiple present is much smaller than that for the other two methods. For those primaries not coincident with the multiples, use of the algorithm resulted in a general loss of amplitudes, (the reliability value was chosen to get an improved level of noise rejection even at the expense of losing some signal). That loss of amplitude, however, seems to have been more or less uniform for all offsets compared with Hampson's results (which makes the curves from this two methods roughly parallel). Since AVO analysis depends more on the relative variation of amplitudes with offsets than on the actual amplitude values, the result of the method in terms of AVO can be considered at least equal to that of Hampson's. The algorithm did not do very well for the far offsets, for which the departure from the true amplitudes varied rapidly to values as high as 50% in Figure 6.21c, probably due to an edge effect. Such large-offset data, however, are not used in AVO analysis.

Figure 6.22 shows the four sets of curves corresponding to the four primary reflections

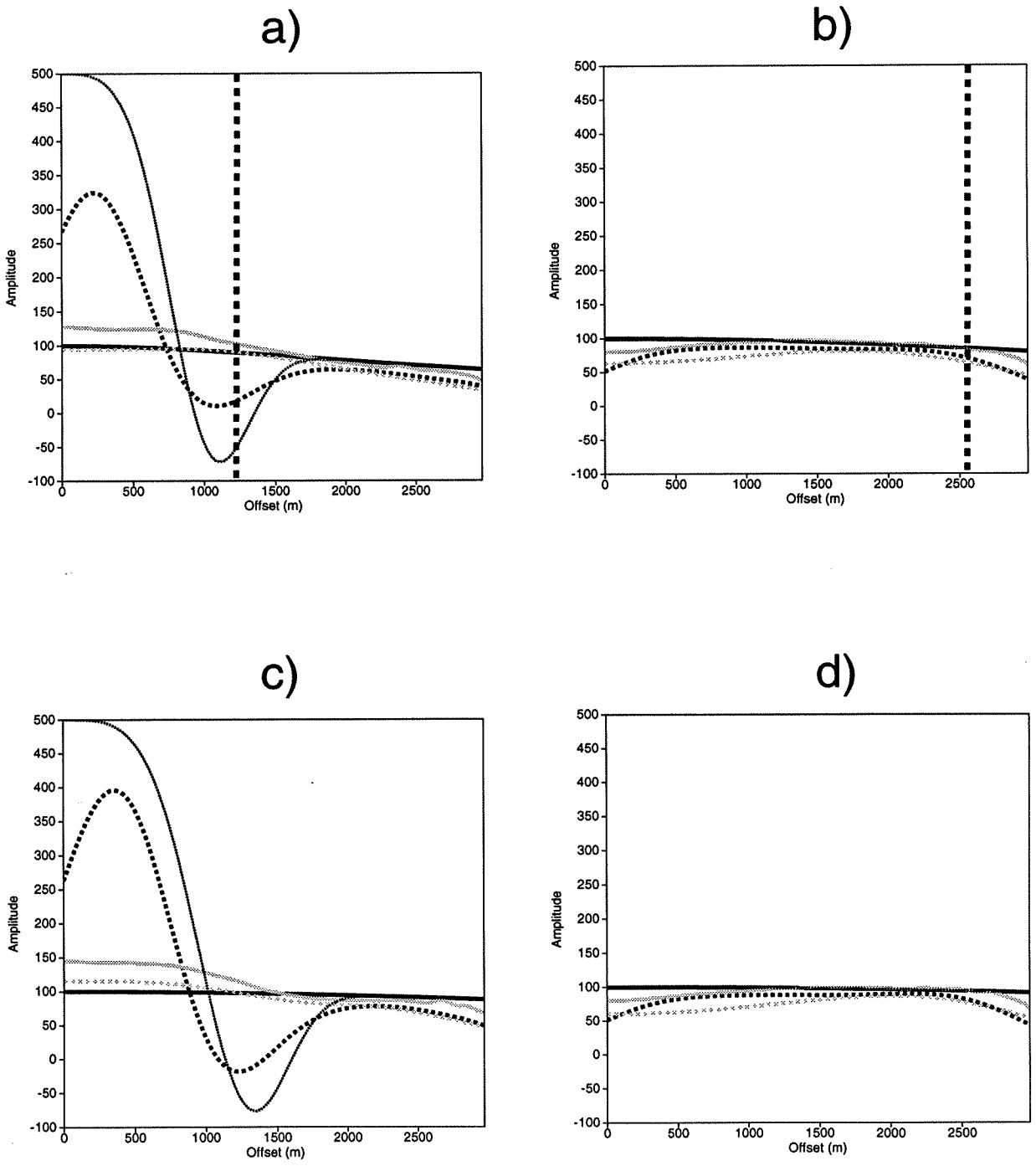

FIG. 6.21. Amplitude variation with offset of extracted primaries in modeled test dataset 5a. a) through d) represent the primaries from the shallowest to the deepest. The solid black line is the input (in the absence of multiples), the thinly dotted line the input (in the presence of multiples) and the dashed black, solid gray and dashed gray lines the extracted primary with FK, Hampson's and hybrid approaches respectively. The dashed vertical black line represents offset equal to the depth of the reflector.

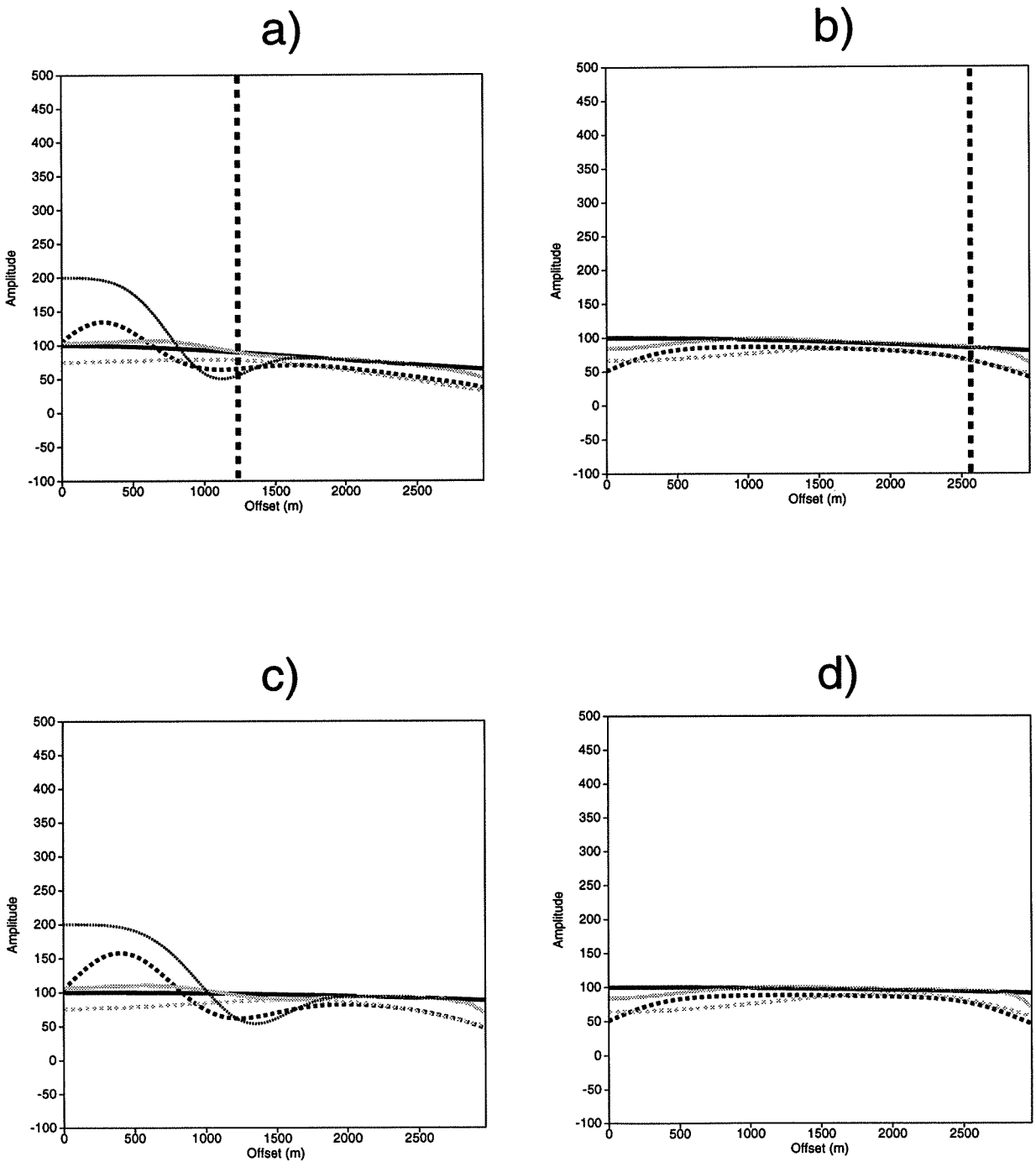

FIG. 6.22. Same as Figure 6.21, but for modeled test dataset 5b.

in modeled test dataset 5b (multiple-to-primary amplitude ratio of 1:1 and constant trace-to-trace amplitudes). As mentioned before, the linearity of the FK filtering operation makes the extraction of the primaries for this dataset equal to that for the previous one. This can be particularly seen in the comparison of Figures 6.22b and 6.22d (primaries non-coincident with multiples) with Figures 6.21b and 6.21d. For primaries coincident with multiples (Figures 6.22a and 6.22c), the result looks different from Figures 6.21a and 6.21d only because the multiples started out weaker in this dataset. The AVO implication of the FK algorithm for this dataset, however, remains the same as that for the previous dataset: results would be meaningless because of the poor rejection of multiples at the short offsets.

Hampson's results for this dataset are also the same as those in the previous dataset because of the linearity of the τ-p filtering operation. Again, the only difference is in the much weaker amplitude of the multiples in this dataset. This weaker amplitude makes the amplitudes for the short offsets of the extracted primaries that are contaminated by multiples (Figures 6.22a and 6.22d) closer to the true amplitudes than in Figures 6.21a and 6.21c for the previous dataset. The results for AVO, therefore, would be commensurately better than for dataset 5a. Where primaries are not contaminated by multiples, the results are the same as for dataset 5a.

The hybrid approach of Chapter 5 is about as successful here as it was with dataset 5a even though the residual multiple energy inside the primary region was comparatively lower than for dataset 5a. The fact that the primary amplitudes are lower in Figures 6.22a and 6.22c compared with Figures 6.21a and 6.21c is somewhat misleading. As shown in plots b and d of Figures 6.21 and 6.22, uncontaminated primaries are weakened by the hybrid process. Contaminated primaries (plots a and c) are likewise weakened, but the measured amplitudes are higher because they include the contribution from the superimposed multiples. In Figure 6.21, the additional multiples fortuitously compensated for the loss of primary amplitudes. Because the residual multiples are weaker, the net amplitudes in plots a and c of Figure 6.22 appear low. The details in the shape of the curves are related to the specific characteristics of the model dataset (i.e., the relative reflection times of primaries and multiples), and so are not as important as is the general trend that they show. Figures 6.22b and 6.22d are virtually the same as Figures 6.21b and 6.21d since those primaries are not contaminated by multiples, the primary is the same in both datasets, and the reliability threshold value used is about the same for both datasets.

It is important to note that, however, just as for dataset 5a, even though the extracted primaries are in all cases weaker than they should be, and weaker than with Hampson's method, from the point of view of AVO the two results are comparable because the curves for the two methods are roughly parallel. It will be shown in Section 6.4 that even with the loss of primary amplitude, the extra rejection of the multiples obtained here could perhaps justify the extra cost of this algorithm compared with Hampson's for this dataset.

Figure 6.23 shows the amplitude curves corresponding to the four primaries in modeled test dataset 5c (with amplitudes decreasing with offset and a multiple-to-primary amplitude ratio of 4:1). As with the previous datasets, the FK filtering algorithm exhibits edge effects at short and far offsets. The departure from the true amplitudes at short offsets, as mentioned before, makes the algorithm inappropriate for AVO analysis. The departure at far offsets is

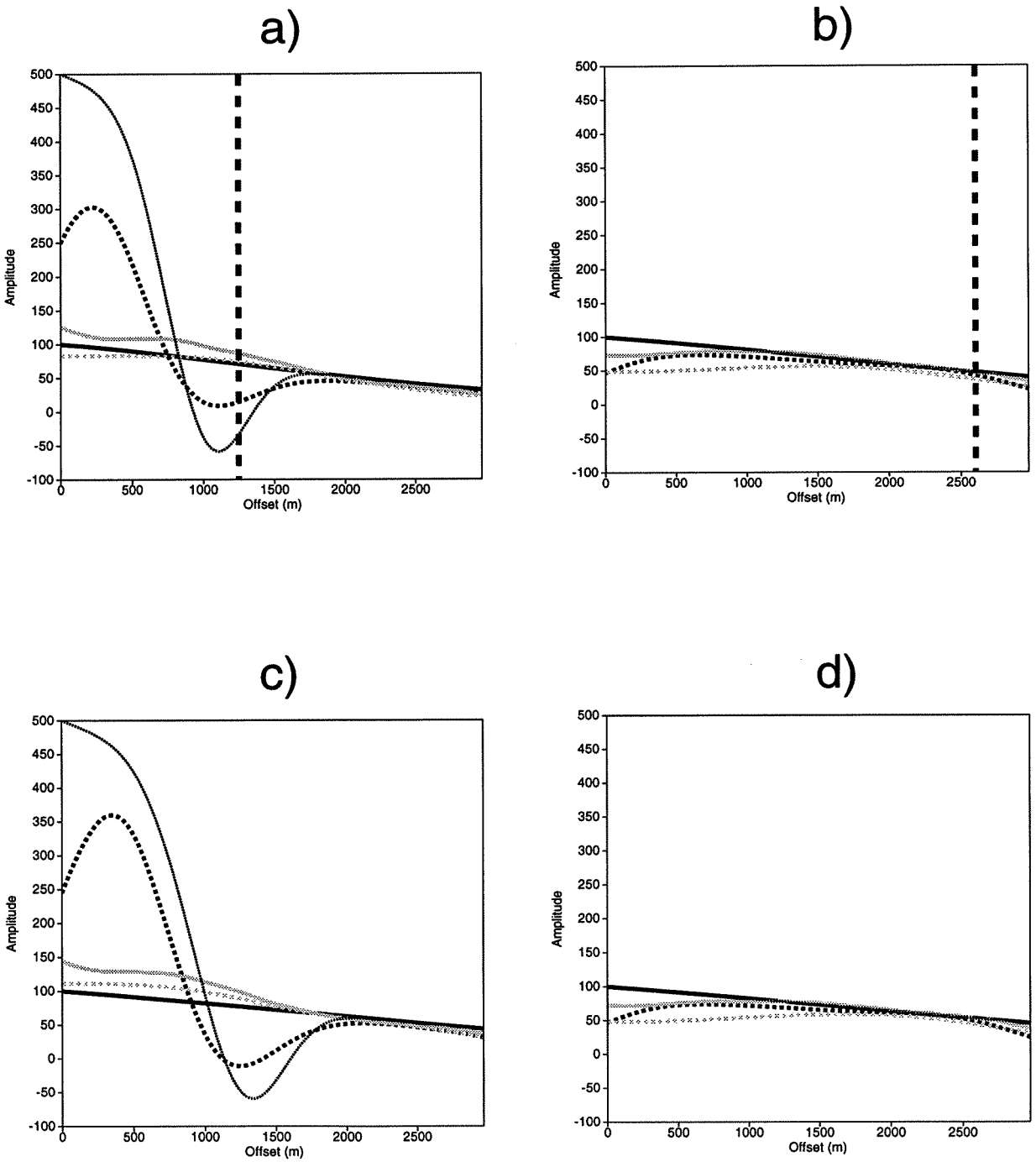

FIG. 6.23. Same as Figure 6.21, but for modeled test dataset 5c.

not as pronounced as with dataset 5a because the linear decrease of amplitude with offset acts as a taper to lessen the edge effect of the FK transform, but this difference is not significant since far offsets are not important for AVO.

Both, Hampson's and the hybrid method performed about the same with this dataset as they did for dataset 5a (recall that the only difference between the two datasets is the linear variation of amplitudes with offsets) for primaries not coincident with multiples (compare Figures 6.23b and 6.23d with Figures 6.21b and 6.21d). The only essential difference is that the decrease of amplitudes with offsets (as with the FK algorithm), helped reduce the edge effects at large offsets. For primaries coincident with multiples, however, some slight distortion is present at the short offsets (compare Figures 6.23a and 6.23c with Figures 6.21a and 6.21c). This distortion, for both of these methods will compromise AVO analysis more for datasets in which the amplitudes decrease with offset than for datasets for which the amplitudes are constant with offset.

Figure 6.24 shows the four sets of curves corresponding to the four primaries in modeled test dataset 5d (when a polarity reversal is introduced in the data). For this dataset, none of the methods performed well from the stand point of fidelity of amplitude variations with offset. The FK algorithm is unable to suppress much of the multiple energy, and even when the primaries are not coincident with the multiples, the edge effects are so strong that the result, as with the previous datasets, is inappropriate for AVO analysis.

The polarity reversal makes the focusing of the primaries and multiples weaker, and even worse, smears the primary and multiple energy in the τ-p domain to such degree that the separation between the two becomes poor (Figure 6.8). This reflects in the inability of either Hampson's or the hybrid approach to honor the steep linear decrease of amplitudes with offset in this dataset. Note also that the results with Hampson's and the hybrid approach are similar for Figures 6.24b and 6.24d but very different for Figures 6.24a and 6.24c because of the failure of both algorithms to suppress a significant part of the multiple energy that was mapped to the primary zone. This implies that AVO analysis of a dataset such as this could be significantly compromised. Perhaps some other method not discussed here might perform better. One that may have the potential for preservation of AVO features, is based on an iterative time-domain conjugate gradient scheme to find a velocity scan that "fits" the associated CMP gather to within a few percent misfit energy, when a hyperbolic forward modeling operator is applied to that inverted velocity scan (Lumley, 1995). It remains to be seen whether or not such an approach is superior to the τ-p methods discussed here.

All tests in this section were created with the same polarity for primaries and multiples. With field data, both positive and negative polarity for both primaries and multiples are likely. While these differences in polarity will surely change the particular shape of the curves shown in these tests, the general conclusions drawn from them will probably be the same. Where multiple suppression is enough, the data remain acceptable for AVO analysis.

6.4 Influence of the multiple extraction on the quality of the CMP stack

Having compared the relative performance of the three methods for multiple suppression for use in AVO analysis, let us now compare their performance in terms of the improvement

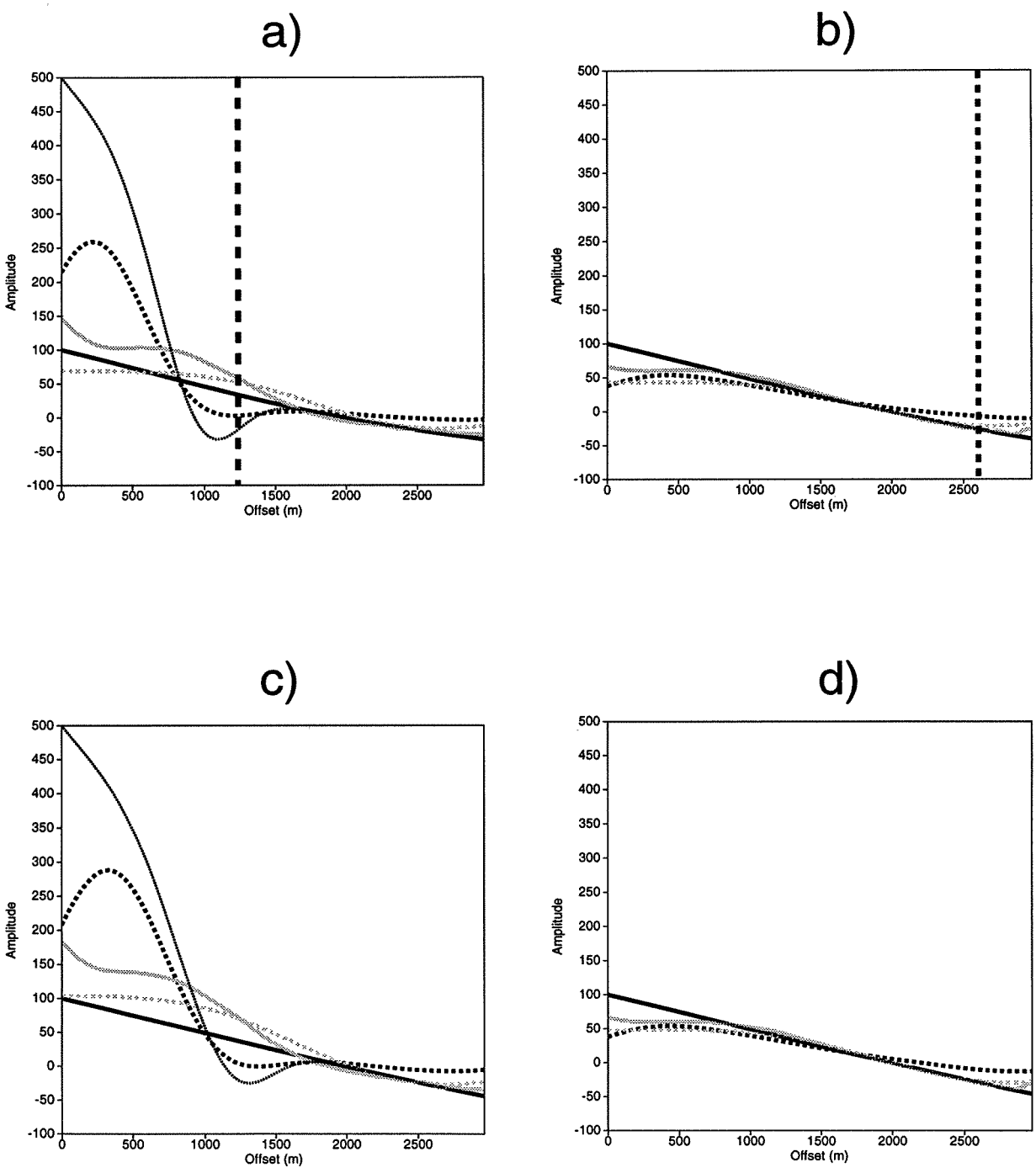

FIG. 6.24. Same as Figure 6.21, but for modeled test dataset 5d.

in primary-to-multiple amplitude ratio in a stacked trace. For this, and for each of the four test models, I stacked the NMO-corrected data, the NMO-corrected input primaries-only and the NMO-corrected input primaries extracted with each of the methods. The stacked traces are plotted side-by-side for comparison such that the first is the stacked trace of the input data, the second is the stacked primary-only input data (ideal) and the next three are the extracted primaries with the FK filtering method, Hampson's method and the hybrid method.

Figures 6.25, 6.26, 6.27 and 6.28 show the stacked traces for test datasets 5a to 5d. From these figures, I measured the primary-to-multiple amplitude ratio (P/M) for the stacked traces after the application of each of the multiple suppression methods. The P/M was computed as the quotient of the amplitude of the extracted primary divided by the amplitude of the residual multiple. For primaries coincident with multiples, the amplitude of the residual multiple was estimated as the difference of the amplitude of the primary (which has a contribution from the multiple) and the amplitude of an adjacent primary with no contribution from multiples. Since the P/M ratio varies slightly for the four sets of primaries and multiples in each of the modeled datasets, I compute a simple arithmetic average for each dataset.

The results of these P/M computations appear in Table 6.1. The first column corresponds to the primary-to-multiple amplitude ratio of the input data and the next the result obtained with the CMP stack alone and with each of the three methods for multiple suppression studied here.

Table 6.1. Comparison between the different methods for multiple suppression presented in this chapter in terms of primary-to-multiple amplitude ratio in a CMP stacked trace. The numbers correspond to peak-to-peak amplitude ratio between primaries and multiples.

Model dataset	**Input**	**CMP stack**	**FK filter**	**Hampson's**	**Hybrid**
Dataset 5a	0.25	1.0	1.0	5.5	10.2
Dataset 5b	1.0	4.0	4.3	18.0	40.0
Dataset 5c	0.25	0.8	0.85	3.9	4.4
Dataset 5d	0.25	0.31	0.33	0.9	0.87

Analysis of the results in Table 6.1 shows that the CMP stack itself was able to provide a P/M ratio improvement of up to 4.0, except for dataset 5d (polarity reversal) for which its improvement was marginal.

The FK filtering approach yielded only marginal P/M ratio improvement over and above what the CMP stack itself did. Since it was shown in Section 6.3 that it also performed poorly in terms of AVO preservation, this method is not adequate for the type of differential moveouts and primary-to-multiple amplitude ratios tested in this chapter.

Hampson's method yielded an improvement in P/M ratio over the CMP stack that ranged from about 3:1 for model dataset 5d to more than 5:1 for model dataset 5a. In

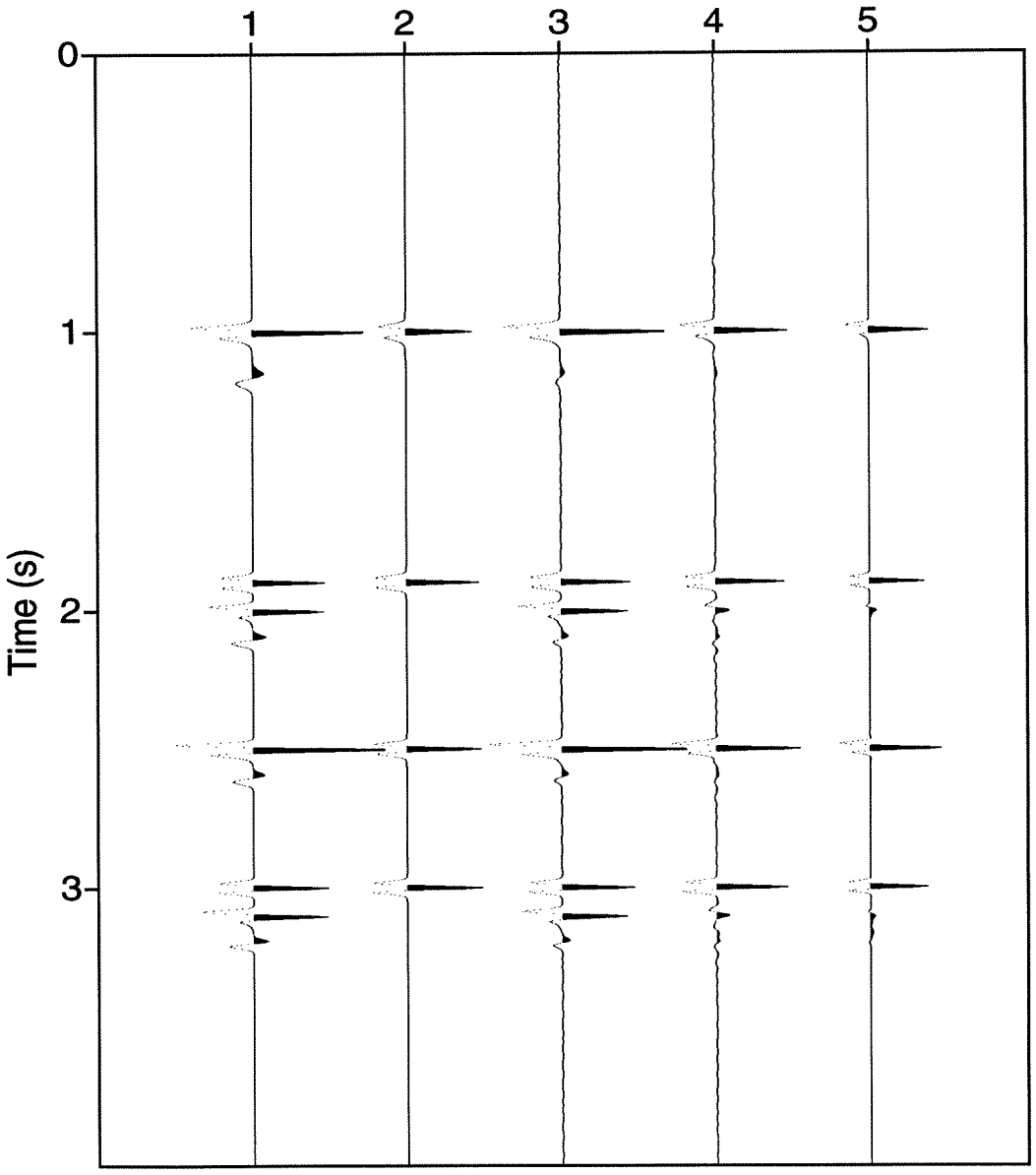

FIG. 6.25. CMP stacked traces for modeled test dataset 5a. The first trace corresponds to the stack of the NMO-corrected input data, the second the stack of the NMO-corrected primary-only input data and the last three traces the extracted primaries with FK filtering, Hampson's and the hybrid method.

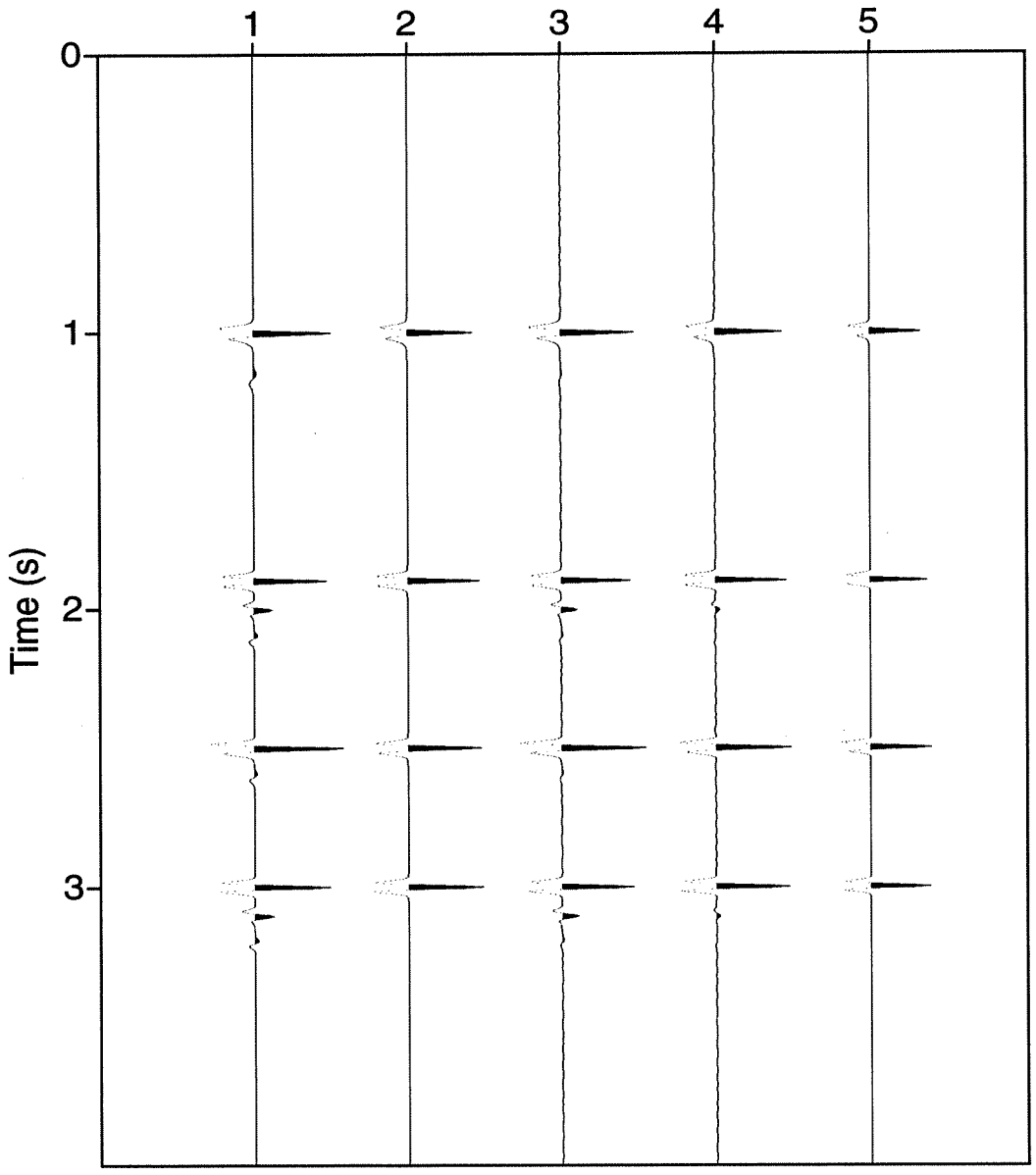

FIG. 6.26. Same as Figure 6.25, except that the CMP-stacked traces are for modeled test dataset 5b.

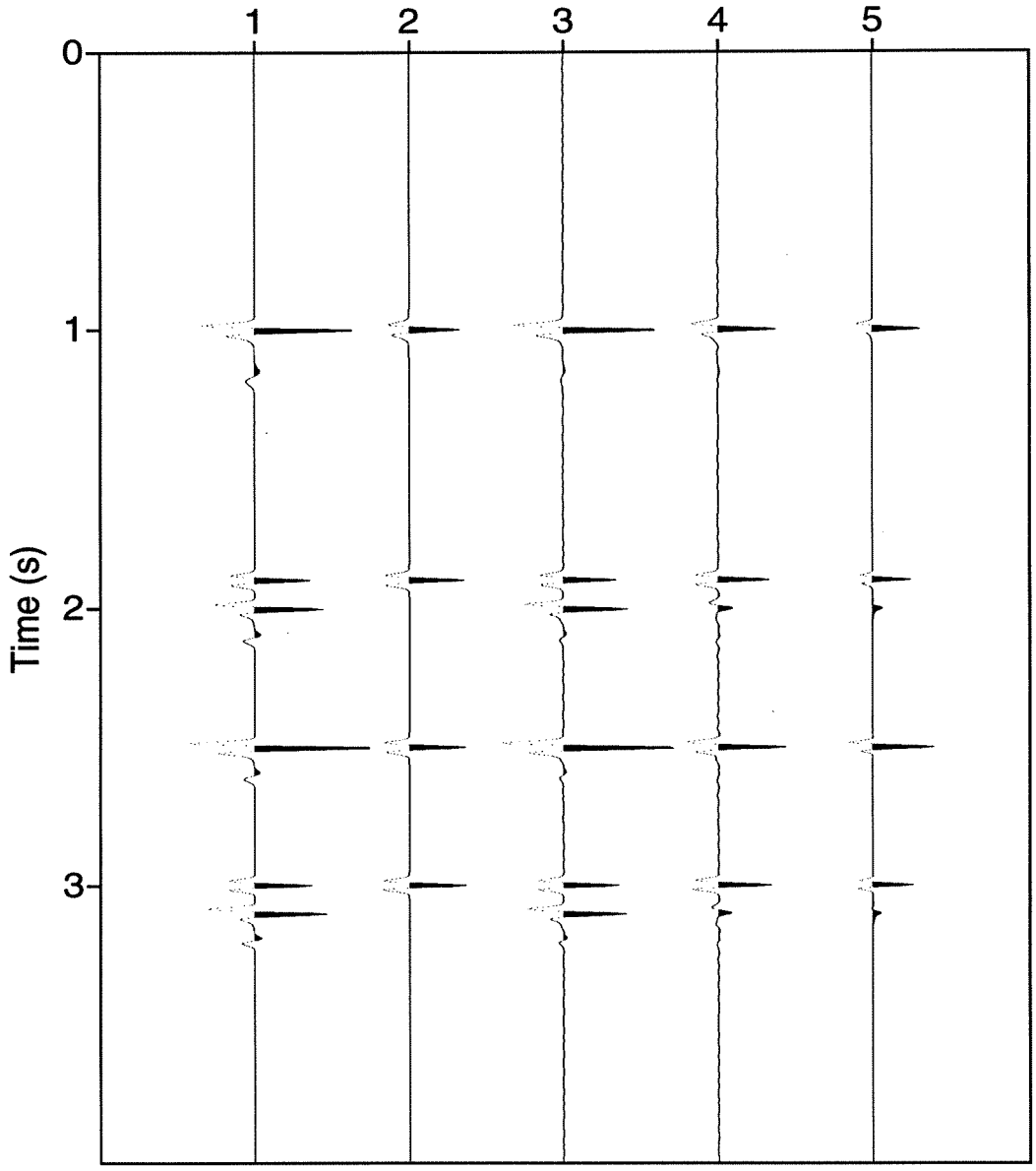

FIG. 6.27. Same as Figure 6.25, except that the CMP-stacked traces are for modeled test dataset 5c.

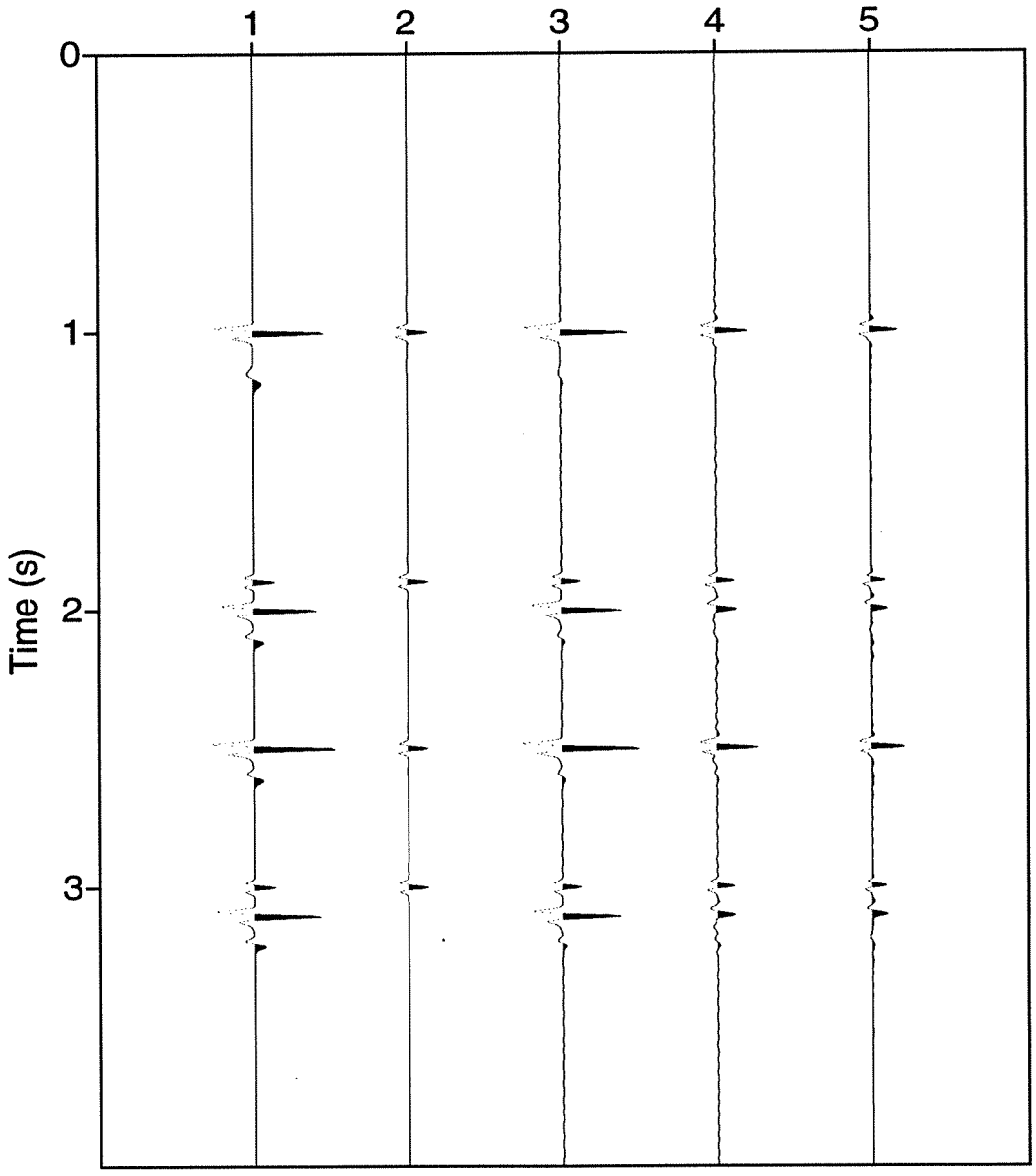

FIG. 6.28. Same as Figure 6.25, except that the CMP-stacked traces are for modeled test dataset 5d.

general, for all the datasets, the improvement in P/M ratio was significant.

The hybrid approach, as expected, yielded further improvement of P/M ratio for the first and second datasets (those for which the trace-to-trace amplitudes were constant). For the fourth dataset, for which the polarity reversal reduced the focusing of the primaries and increased the smearing of multiple energy in the τ-p domain to a point that the algorithm could not discriminate between the primaries and the residual multiples, the result of the hybrid method was actually poorer than that of Hampson's method.

In summary, then, from this quantitative analysis carried out with the four datasets, I consider that if improvement in P/M ratio is the overriding factor, then the extra cost of the hybrid approach is justified for datasets 5a and 5b. For dataset 5c (amplitudes decreasing linearly with offset), the extra multiple rejection is perhaps not enough to justify the extra cost (50%) of the hybrid algorithm compared with Hampson's method. For dataset 5d (polarity reversal with offset), none of the methods was able to produce the level of multiple rejection achieved with the previous datasets. Perhaps another method should be sought, that could produce a larger P/M ratio. Wave-theory-based methods (Verschuur et al., 1992; Verschuur and Berkhout, 1994; Verschuur and Berkhout, 1995) have recently been proposed that could be tested using wave-theoretical data exhibiting characteristics similar to those here. Also, as mentioned above, Lumley (1995) proposed a method especially tailored to preserve AVO features. It is not known whether or not these methods can suppress multiples more effectively than either the Hampson or the hybrid method, for example, in order to improve the accuracy of AVO analysis. For CMP stacking, they must perform on data of the sort studied here at least as well as these two methods.

6.4.1 A closer look

Finally, let us take a more detailed look at the stacking of each primary in order to get a better feeling for the comparative result of the CMP stack tests. To this end, the data are windowed in time so that only one primary is included in each window. The window lengths are 500 ms, centered on the corresponding primary.

Figure 6.29 shows such plots for the primaries in model dataset 5a. This figure shows that Hampson's method yielded the most accurate representation of the primaries. The hybrid approach, on the other hand, did best at suppressing the multiples as was mentioned in the analysis of Table 6.1. In this figure, the small decrease in amplitude of the primaries with the hybrid approach is more than compensated by the increase in multiple rejection power. The FK method performed well in the extraction of the primaries, except for the first primary, but did very poor on the level of multiple rejection.

Figure 6.30 shows similar plots for the primaries in model dataset 5b. The results for this dataset are similar to those for the previous one. As mentioned above, this could be expected from the linearity of the FK and τ-p filtering operations. Hampson's method performed best in extracting the correct primary amplitudes but allowed some small residual multiple energy. The hybrid approach did best in suppressing the multiple energy but at the expense of decreasing the amplitude of the extracted primaries. It was shown in the previous section, nevertheless, that the increase of primary-to-multiple amplitude ratio was likely enough to

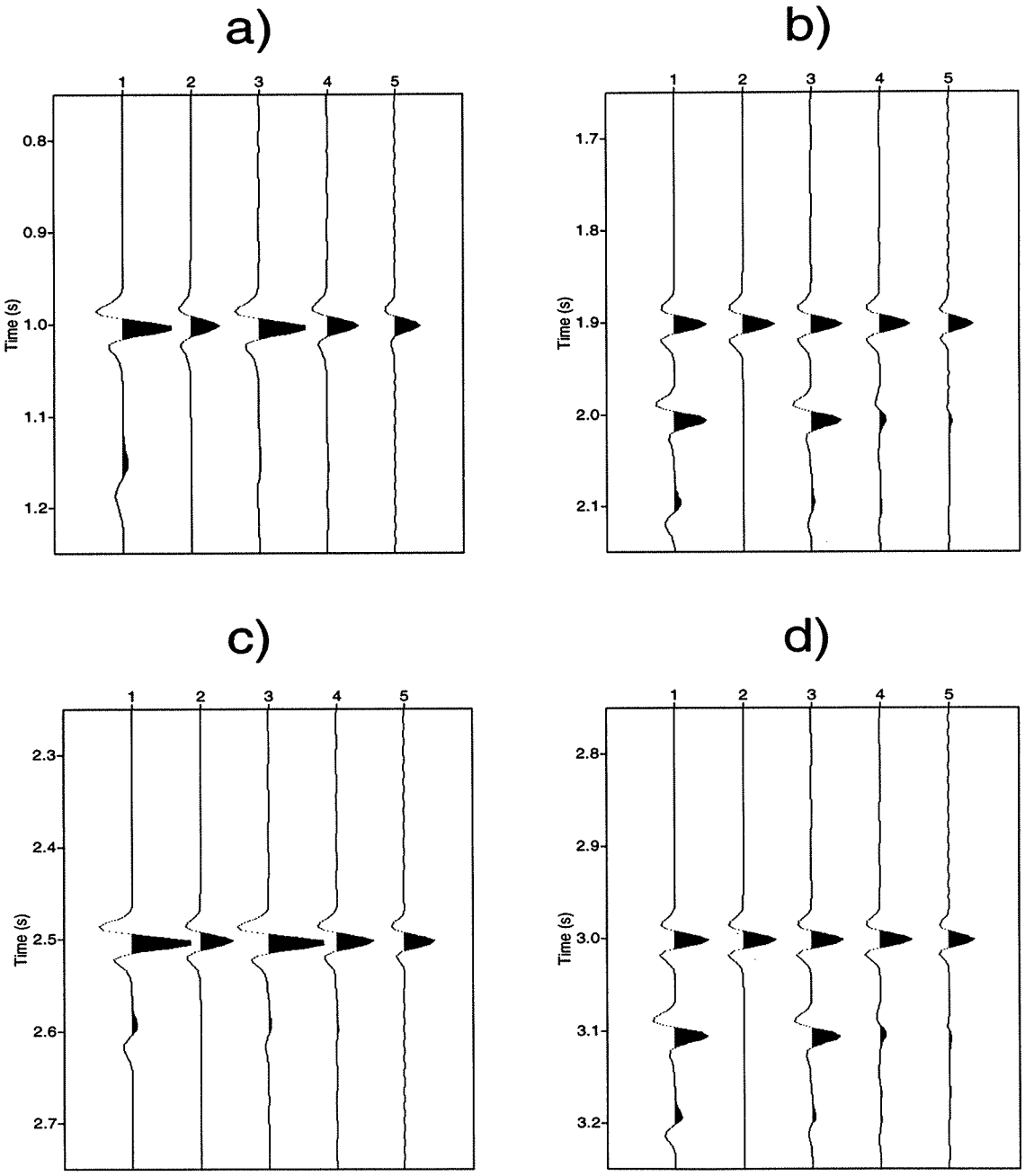

FIG. 6.29. Details of stacked traces for modeled test dataset 5a. The order of the traces is the same as that on the previous set of figures. a) to d) represent the primaries from the shallowest to the deepest.

justify the extra expense (about 50% more than Hampson's approach) of applying the hybrid method to this dataset.

Figure 6.31 shows the detailed stack for the primaries in model dataset 5c (amplitudes linearly decreasing with offset). In this case, the level of primary-to-multiple amplitude ratio was good enough with Hampson's approach. The extra degree of rejection afforded by the hybrid method is not significant and would not justify the extra cost over and above Hampson's approach. Note that this dataset does not meet the assumption under which the τ-p transform of the data works best; specifically, the amplitude variation with offset makes it inaccurate to represent the data as a superposition of parabolas whose amplitudes do not change with offset.

Figure 6.32 shows the detailed stack for the primaries in model dataset 5d (polarity reversal with offset). For this dataset, the lack of focusing of the primaries and the increased smearing of energy for the multiples in the τ-p domain cause both Hampson's method and the hybrid method to be inappropriate. They both reduce the multiple energy to about one third of its value when the data are simply CMP stacked. The residual multiple energy, however, is strong enough to distort the apparent amplitude of the extracted primary, in particular for the first and third primary (primaries coincident with multiples). The extracted amplitude is more than twice its correct value. Here again, this result is not surprising since the data with the polarity reversal cannot be modeled accurately with constant amplitude parabolas.

6.5 Summary

The results of the above tests, are summarized on Tables 6.2, 6.3, 6.4 and 6.5 for each of the datasets. The two attributes discussed here (AVO and CMP stack quality) are used to compare the relative performance of the different methods for each primary of each dataset. A grade of very good, good, fair or poor is used based on the analysis of the previous two sections. Recall that the first and third primary in each dataset are contaminated by a multiple, whereas the second and fourth are not.

The results on Table 6.2 indicate that for dataset 5a (for high multiple-to-primary amplitude ratio) the extra cost of the hybrid algorithm is justified and it should probably be chosen unless cost is a primary concern.

Table 6.3 shows that for dataset 5b (that is, multiple-to-primary amplitude ratio of 1:1) even though the multiple energy was not so strong, the extra cost of the hybrid algorithm is perhaps justified, due to the increased primary-to-multiple amplitude ratio obtained in the stacked trace.

Table 6.4 indicates that for dataset 5c (linear trace-to-trace amplitude variation with offset and multiple-to-primary amplitude ratio of 4:1), the extra cost of the hybrid algorithm is perhaps not justified, because the added improvement in primary-to-multiple amplitude ratio is not large enough. Thus, Hampson's method is perhaps the best choice for this dataset.

Finally, Table 6.5 indicates that for dataset 5d, none of the methods really provides a very good result, and as previously mentioned, other methods should be considered. Of the

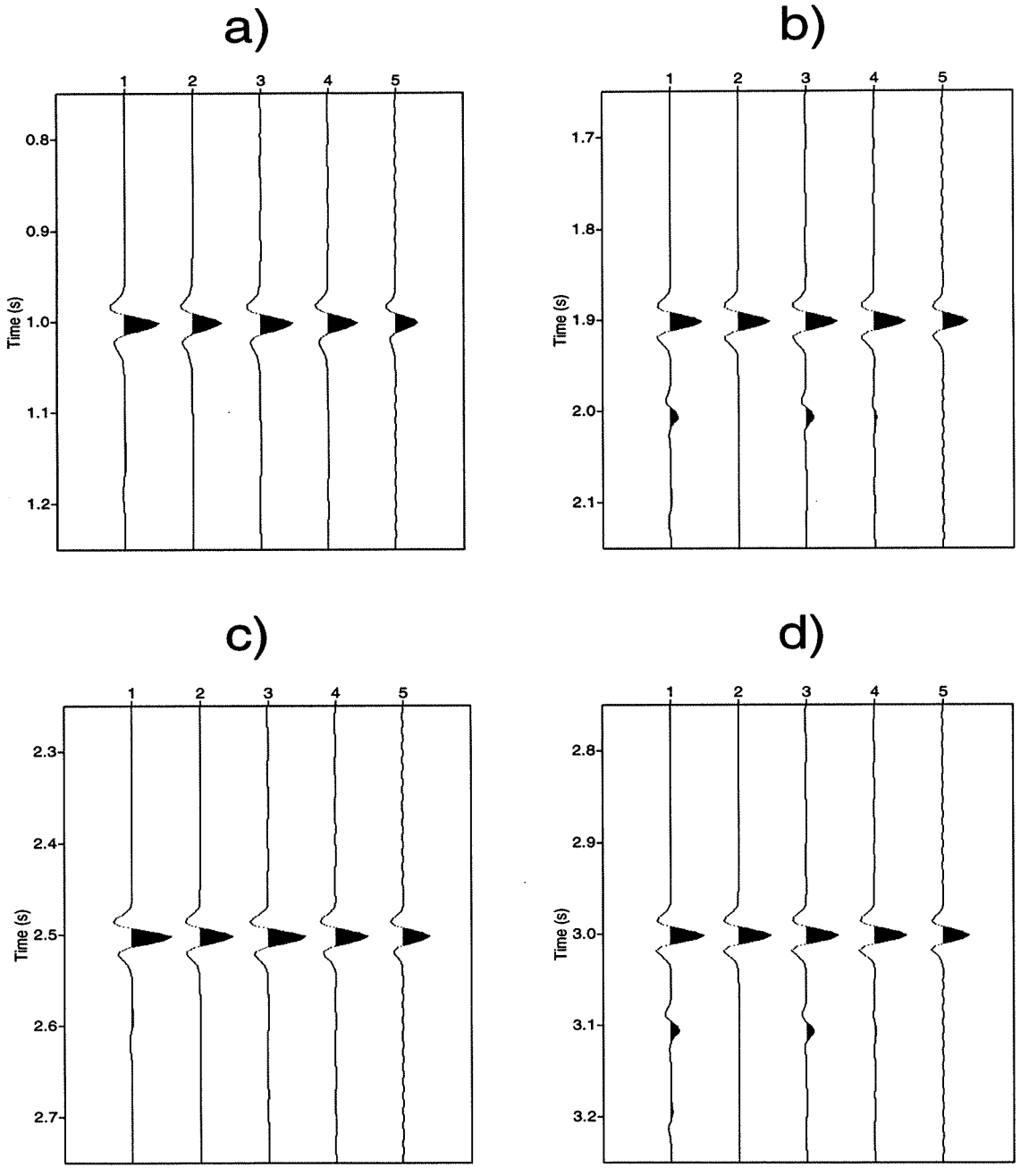

FIG. 6.30. Details of stacked traces for modeled test dataset 5b.

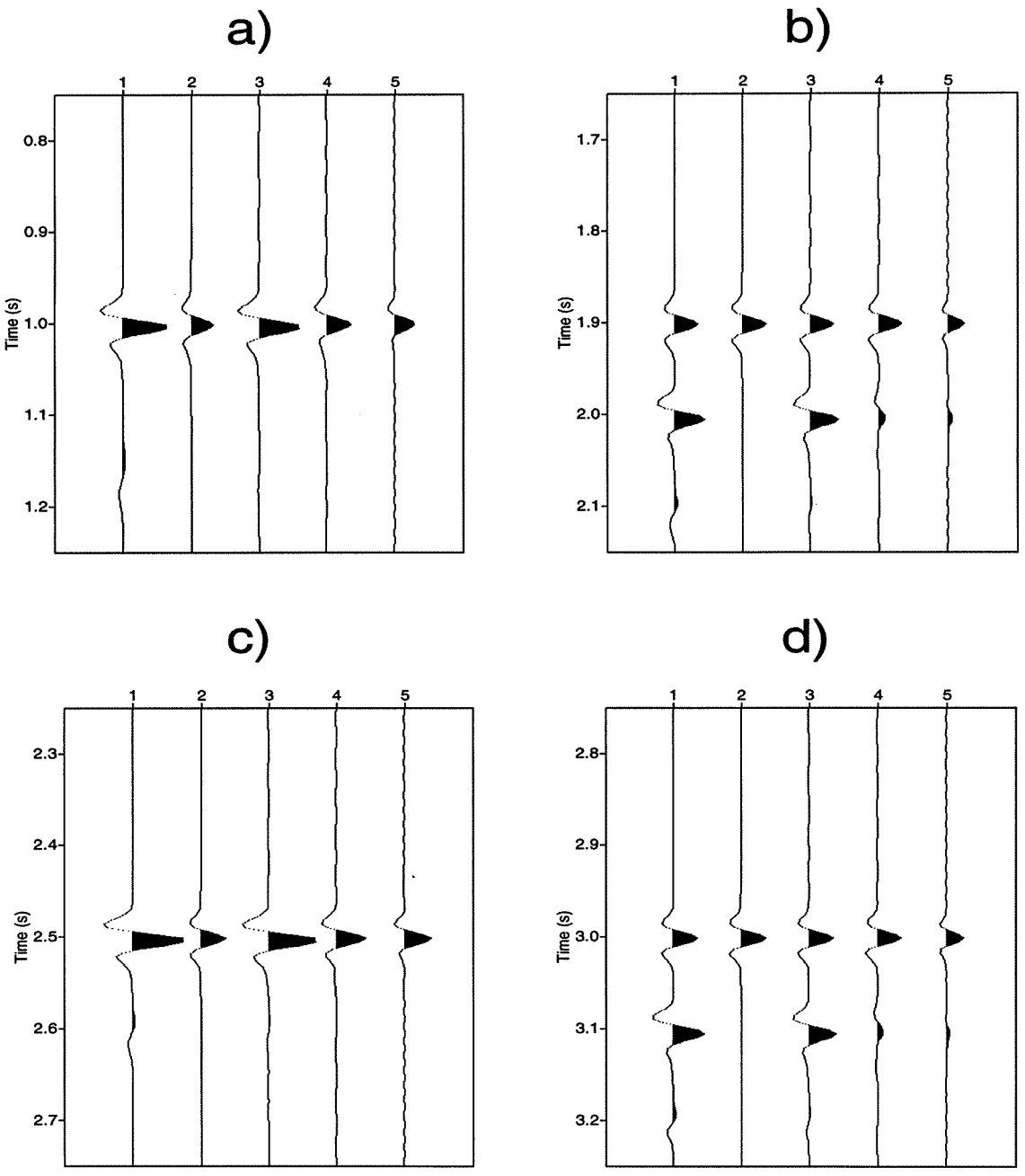

FIG. 6.31. Details of stacked traces for modeled test dataset 5c.

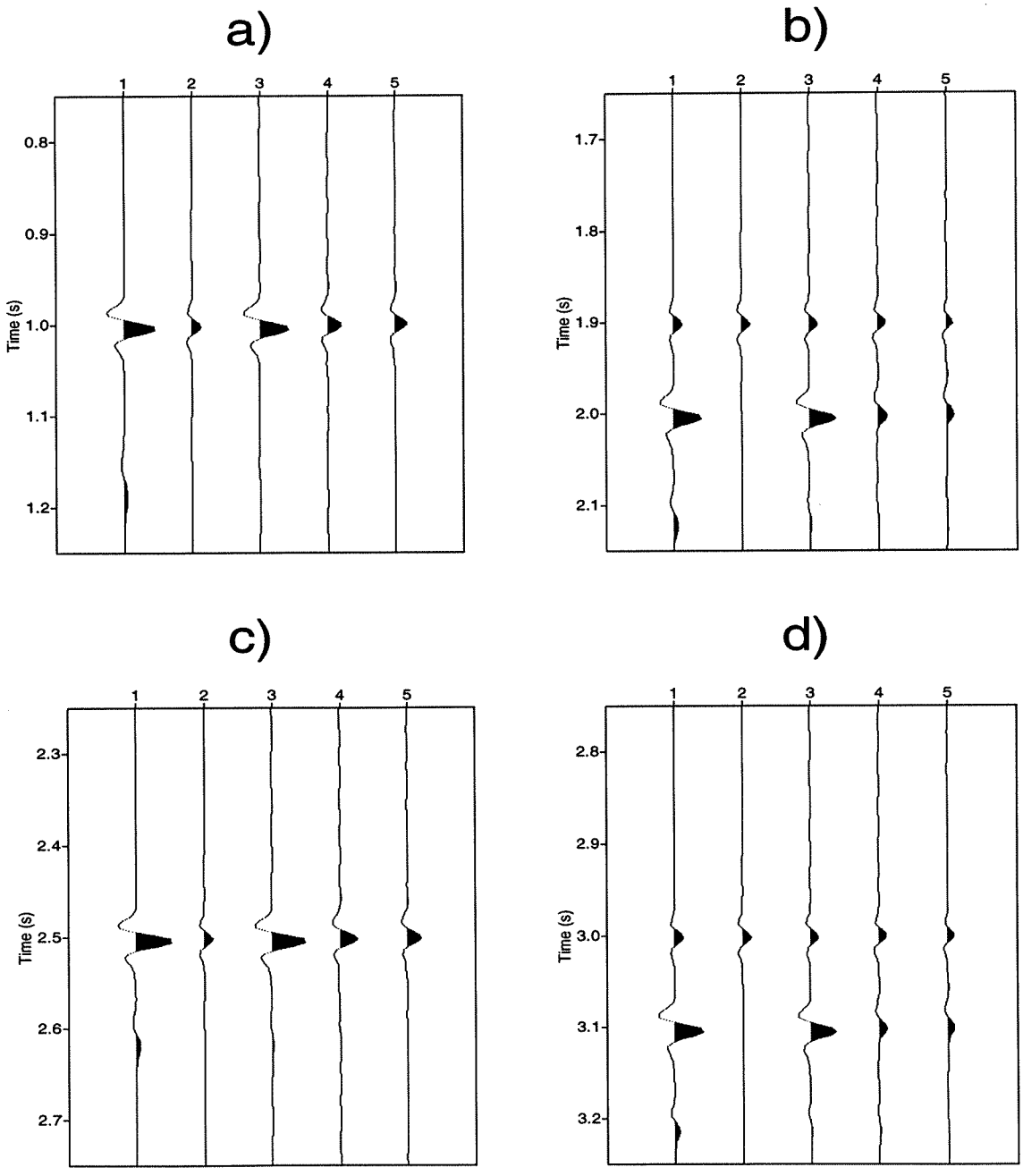

FIG. 6.32. Details of stacked traces for modeled test dataset 5d.

Table 6.2. Performance comparison between the different methods for multiple suppression presented in this chapter applied to dataset 5a (multiple-to-primary amplitude ratio of 4:1 and constant amplitudes with offset).

Method	**AVO preservation**	**Quality of CMP stack**	**Cost**
FK moveout filtering (first and third primary)	Poor	Poor	Good
FK moveout filtering (second and fourth primary)	Fair	Poor	Good
Hampson's method (first and third primary)	Fair	Good	Fair
Hampson's method (second and fourth primary)	Good	Good	Fair
Hybrid approach (first and third primary)	Good	Very good	Poor
Hybrid approach (second and fourth primary)	Fair	Very good	Poor

methods studied here, Hampson's is clearly the strongest one for this dataset. Optimum CMP stacking (Budihardjo, 1995) is yet another alternative, although obviously not for AVO analysis. Optimum CMP stacking can often provide improvements in primary-to-multiple amplitude ratio of more than 4:1 compared with a conventional CMP stacking, but at some loss in attenuation of incoherent noise.

It is also important to note in this summary that the coincidence or lack of it for primaries and multiples at zero-offset time makes a difference in the relative implications of the approached for use in AVO study, as is evident from the results in the previous two sections. Since in practice, however, we are likely to encounter multiples whose polarity and zero-offset time relative to that of primaries varies from one reflection to another, the judgement must consider the likely occurrence of all such situations. Fundamentally, the relative implications for AVO analysis and for CMP stacking depend on the relative performance in suppressing multiples on the one hand while preserving primaries on the other. Where multiples are a problem in field data, their amplitudes can range widely, readily exceeding four times that of primary. This suggest some promise for practical use of the hybrid method.

I finish this chapter by stating that whenever the goal is maximum degree of multiple rejection in the stacked data and AVO is not an issue, then it is better to try the hybrid approach, if the multiple-to-primary amplitude ratio is large and offsets are restricted such that the primary wavelet does not reverse polarity as offset increases. On the other hand, when the overriding concern is AVO and some residual multiple energy can be tolerated in the

Table 6.3. Performance comparison between the different methods for multiple suppression presented in this chapter applied to dataset 5b (multiple-to-primary amplitude ratio of 1:1 and constant amplitudes with offset).

Method	**AVO preservation**	**Quality of CMP stack**	**Cost**
FK moveout filtering (first and third primary)	Poor	Fair	Good
FK moveout filtering (second and fourth primary)	Fair	Poor	Good
Hampson's method (first and third primary)	Fair	Good	Fair
Hampson's method (second and fourth primary)	Good	Good	Fair
Hybrid approach (first and third primary)	Fair	Very good	Poor
Hybrid approach (second and fourth primary)	Fair	Very good	Poor

stack, then Hampson's approach is perhaps the best bet. Notice, however, that this method yields some variation of apparent primary amplitude at the short offsets when the primaries and the multiples coincide at the same zero-offset time. This distortion of the short-offset amplitudes is an undesirable characteristic for AVO analysis.

Table 6.4. Performance comparison between the different methods for multiple suppression presented in this chapter applied to dataset 5c (multiple-to-primary amplitude ratio of 4:1 and amplitudes linearly decreasing with offset).

Method	**AVO preservation**	**Quality of CMP stack**	**Cost**
FK moveout filtering (first and third primary)	Poor	Poor	Good
FK moveout filtering (second and fourth primary)	Fair	Poor	Good
Hampson's method (first and third primary)	Poor	Good	Fair
Hampson's method (second and fourth primary)	Good	Good	Fair
Hybrid approach (first and third primary)	Fair	Good	Poor
Hybrid approach (second and fourth primary)	Fair	Good	Poor

Table 6.5. Performance comparison between the different methods for multiple suppression presented in this chapter applied to dataset 5d (multiple-to-primary amplitude ratio of 4:1 and polarity reversal).

Method	**AVO preservation**	**Quality of CMP** stack	**Cost**
FK moveout filtering (first and third primary)	Poor	Poor	Good
FK moveout filtering (second and fourth primary)	Poor	Poor	Good
Hampson's method (first and third primary)	Poor	Fair	Fair
Hampson's method (second fourth primary)	Fair	Fair	Fair
Hybrid approach (first third primary)	Poor	Fair	Poor
Hybrid approach (second fourth primary)	Fair	Fair	Poor

Chapter 7

CONCLUSIONS

In this study, I reviewed the most commonly used computer implementations of the τ-p transform, reviewed the signal-noise separation algorithm of Harlan et. al. (1984), and showed that a variation of it can be successfully used in the problem of the identification and suppression of ground-roll and multiples. I reviewed the more traditional methods such as F-K and Hampson's τ-p filtering and analyzed the implications of the multiple-removal process with each of the aforementioned methods in terms of AVO preservation and degree of multiple suppression on CMP stacked traces. Following are the main conclusions.

The discrete Radon transform: I showed in Chapter 2, by means of an example, that the best of the currently available computer implementation of the τ-p transform and in general of the discrete Radon transform, is Beylkin's (F-X) implementation. This method produces the best focusing of the desired events in the τ-p domain, accurately reproduces the amplitudes as well as the phases of the events after an inverse transformation and does not introduce edge effects at the nearest and farthest offsets of the inversely transformed data. The other two commonly used implementations (in FK and t-x domains) are faster by a factor of at least four, but they have less focusing power and have edge effects.

Signal-noise separation algorithm: This algorithm is powerful in its ability to separate focused events from unfocused ones after the application of an appropriate invertible linear transformation. The method, with some modifications, can be effectively used to suppress ground-roll (with an order of magnitude improvement in computer efficiency with respect to the original Harlan's method) and to provide an increased degree of rejection of multiple energy (compared with Hampson's method). This increase in multiple rejection power is necessary especially when the moveout difference between primaries and multiples is not large and the multiple-to-primary amplitude ratio is high. Other applications of the method may be envisioned, whenever effective discrimination between events with different moveout patterns is required, provided that an invertible linear transformation can be devised that will focus the desired events (signal) while defocusing the undesired ones (noise). Harlan's method has not been used much probably because some of its implementation details have not been clearly understood and the potential of the method has not been realized. For ground-roll suppression, the computer cost of the algorithm in its original formulation was too expensive. This study helps in both counts making all the relevant computer implementation details explicit and modifying the ground-roll suppression application of the algorithm faster.

Ground-roll suppression: Both Hampson's and the modified signal-noise separation algorithm provide an adequate degree of suppression of the ground-roll energy. The signal-noise separation algorithm, however, can provide an improved preservation and cleaner extraction of the reflections (compared to τ-p filtering), because it can suppress energy from spurious alignments in the transformed domain, especially for strongly spatially aliased data. The algorithm, however, is about 50% more expensive than Hampson's method.

AVO and CMP-stack implications of multiple suppression: The suppression of multiple energy is a necessity with many seismic datasets, particularly marine, but the choice of implementation can have serious implications in distorting AVO behavior and in distorting the amplitudes of the stacked traces. Hampson's approach works well in preserving AVO as long as the primary-multiple moveout separation is sufficiently large and the multiple-to-primary amplitude ratio is not very large, so that not much residual multiple energy can be expected in the extracted primaries. If this is not the case, however, the residual multiple energy will tend to increase or decrease the amplitudes of the short offsets of the extracted primaries, depending on the relative polarity of the two. The hybrid approach, on the other hand, depending on the choice of reliability value, can provide increased multiple suppression but can decrease the amplitude of the extracted primaries in the process. Details of the trade off govern whether or not this approach to multiple suppression will result in data that are appropriate for AVO analysis.

Future work: The comparisons carried out in this study in terms of ground-roll and multiple suppression are not exhaustive. Future similar such numerical tests include other methods for ground-roll suppression such as depth filtering (McMechan and Sun, 1991), dispersion filtering (Beresford-Smith and Rango, 1988), matched filters (Saatçilar and Canitez, 1988) and others. The comparison could, in particular, test the tolerance of the methods to spatially aliased data and could be applied to 3-D data. In terms of multiple suppression, the methods of Lumley (1995), Verschuur et al. (1992) and Verschuur and Berkhout (1994, 1995) could be tested to evaluate their performance in terms of AVO preservation and primary-to-multiple amplitude ratio improvement via CMP stacking, as was done with the methods used in this study.

REFERENCES

Anderson, J. E., 1993, Parabolic and linear 2-D τ-p transforms using the generalized Radon transform, CWP project review, 109-129.

Beresford-Smith, G., and Rango R., 1988, Dispersive noise removal in t-x space: application to arctic data. Geophysics **53**. 346-358.

Beylkin, G., 1987, The discrete Radon transform: IEEE Transactions of Acoustics, Speech, and Signal Processing, **35**, 162-172. In Slant Stack Processing, Geophysics reprint series No. **14**, Society of Exploration Geophysicists, 1991.

Bracewell, R. N., 1978, The Fourier transform and its applications: McGraw-Hill Book Co.

Budihardjo, S., 1995, Quantifying limitations on signal-to-noise ratio improvement in the stacking of seismic data. M.Sc. thesis, Colorado School of Mines.

Claerbout, J. C., 1985, Imaging the earth's interior: Blackwell Scientific Publications, London.

Foster, D. J., and Mosher, C.C., 1992, Suppression of multiple reflections using the Radon transform: Geophysics, **57**, 386-395.

Gulunay, N., 1990, F-X domain least-squares Tau-P and Tau-Q: SEG Expanded Abstracts 1990. Society of Exploration Geophysicists. 1607-1610.

Hampson, D., 1986, Inverse velocity stacking for multiple elimination: J. Can. Soc. Expl. Geophys., **22**, 44-55. In Slant Stack Processing, Geophysics reprint series No. **14**, Society of Exploration Geophysicists, 1991.

Harlan, W. S., Claerbout, J. F., and Rocca, F., 1984, Signal/noise separation and velocity estimation: Geophysics, **49**, 1869-1880. In Slant Stack Processing, Geophysics reprint series No. **14**, Society of Exploration Geophysicists, 1991.

Harlan, W. S., 1986, Signal/noise separation and seismic inversion, Ph. D. thesis, Stanford University.

Harlan, W. S., 1988, Separation of signal and noise applied to vertical seismic profiles: Geophysics, **53**, 932-946.

Hugonnet, P. and Canadas, G., 1995, Aliasing in the parabolic radon transform: SEG Expanded Abstracts, 1995. Society of Exploration Geophysicists. 1366-1369.

Kirchheimer, F., 1985, On some further aspects of fan filtering. SEG Expanded Abstracts 1985. Society of Exploration Geophysicists. 635-638.

Kostov, C., 1990, Toeplitz structure in slant-stack inversion: SEG Expanded Abstracts 1990. Society of Exploration Geophysicists. 1618-1621. In Slant Stack Processing, Geophysics reprint series No. **14**, Society of Exploration Geophysicists, 1991.

Lumley, D., 1995, Amplitude-preserved multiple suppression. SEG Expanded Abstracts 1995. Society of Exploration Geophysicists. 1460-1463.

McMechan, G., and Sun, R., 1991, Depth filtering of first breaks and ground-roll. Geophysics, **56**. 390-396.

Papoulis, A., 1965, Probability, random variables and stochastic processes: McGraw-Hill Book Co.

Rieber, F., 1936, A new reflection system with controlled directional sensitivity: Geophysics, **1**, 97-106. In Slant Stack Processing, Geophysics reprint series No. **14**, Society of Exploration Geophysicists, 1991.

Rosenbaum, J. H., and Boudreaux, G.F., 1981, Rapid convergence of some seismic processing algorithms: Geophysics, **46**, 1667-1672.

Rüger, J. E., 1995, P-wave reflections and azimuthal depedence of AVO in transversely isotropic media, CWP project review, 285-308.

Saatçilar, R., and Canitez, N., 1988, A method of ground-roll elimination. Geophysics, **53**. 894-902.

Stolt, R. H., Migration by Fourier transform: Geophysics, **43**, 23-48. In Migration, Geophysics reprint series No. **4**, Society of Exploration Geophysicists, 1988.

Thatam, R. H., 1984, Multidimensional filtering of seismic data: Proceedings of the IEEE, Vol **72**, No. 10. In Slant Stack Processing, Geophysics reprint series No. **14**, Society of Exploration Geophysicists, 1991.

Thorson, J. R., and Claerbout J., 1985, Velocity-stack and slant-stack stochastic inversion: Geophysics, 1985. **50**, 2727-2741. In Slant Stack Processing, Geophysics reprint series No. **14**, Society of Exploration Geophysicists, 1991.

Turner, G., 1990, Aliasing in the tau-p transform and removal of spatially aliased coherent noise: Geophysics, **55**, 1496-1503.

Verschuur, D., Berkhout, A., and Waapenar, C., 1992, Adaptive surface-related multiple elimination. Geophysics, **57**, 1166-1177.

Verschuur, D., Berkhout, A., 1994, Multiple technology, part I: Estimation of multiple reflections: Expanded Abstracts, 1994. 1493-1496.

Verschuur, D., Berkhout, A., 1995, Multiple technology, part II: Examples on marine and land data: Expanded Abstracts, 1995. 1470-1473.

Wade, C. J., and Gardner, G. H., 1988, Slant-stack inversion by hyperbola extraction in the Fourier domain: SEG Expanded Abstracts. Society of Exploration Geophysicists. 676-679. In Slant Stack Processing, Geophysics reprint series No. **14**, Society of Exploration Geophysicists, 1991.

Yilmaz, O., 1987, Seismic data processing: Society of Exploration Geophysicists. Tulsa.