Geophysical Prospecting 35, 12–32, 1987.

SIGNAL-TO-NOISE RATIO ENHANCEMENT IN MULTICHANNEL SEISMIC DATA VIA THE KARHUNEN–LOÉVE TRANSFORM*

I.F. JONES** and S. LEVY***

ABSTRACT

JONES, I.F. and LEVY, S. 1987, Signal-to-Noise Ratio Enhancement in Multichannel Seismic Data via the Karhunen–Loéve Transform, Geophysical Prospecting 35, 12–32.

The Karhunen–Loéve transform, which optimally extracts coherent information from multichannel input data in a least-squares sense, is used for two specific problems in seismic data processing.

The first is the enhancement of stacked seismic sections by a reconstruction procedure which increases the signal-to-noise ratio by removing from the data that information which is incoherent trace-to-trace. The technique is demonstrated on synthetic data examples and works well on real data. The Karhunen–Loéve transform is useful for data compression for the transmission and storage of stacked seismic data.

The second problem is the suppression of multiples in CMP or CDP gathers. After moveout correction with the velocity associated with the multiples, the gather is reconstructed using the Karhunen–Loéve procedure, and the information associated with the multiples omitted. Examples of this technique for synthetic and real data are presented.

INTRODUCTION

In many physical problems it is desirable to separate one type of coherent waveform (the 'signal') from a different coherent waveform, or from incoherent energy. F–K filtering and coherency stacks have been traditionally used to achieve this separation. We consider the Karhunen–Loéve (KL) transformation as an alternative method.

Applications of the KL method to seismic signals were considered by Hemon and

* Received July 1985, last material April 1986.

** Department of Geophysics and Astronomy, University of British Columbia. Present address: Compagnie Générale de Géophysique, 47–55 The Vale, Acton, London W3 7RR, UK.

*** Department of Geophysics and Astronomy, University of British Columbia. Present address: IT&A, Inverse Theory and Applications Inc., 2162 Western Parkaway, Vancouver, BC V6T 1V6 Canada.

Mace (1978), and more recently by Jones (1985), and Ulrych, Levy, Oldenburg and Jones (1983). Levy, Ulrych, Jones and Oldenburg (1983) introduced an extension of the transform to complex signals to address the problem of trace-to-trace phase variation. In the field of image processing, the KL transform has been widely applied to data transmission and analysis (Mallick and Murthy, 1984) and digital image enhancement (Ready and Wintz 1973, Ahmed and Rao 1975, Huang and Narendra 1975, and Andrews and Patterson 1976a, 1976b).

We expand upon the applications introduced by Levy et al. (1983) and Ulrych et al. (1983), and look at two specific problems, namely:

(a) the separation of signal from incoherent and dipping coherent noise in stacked seismic data,
(b) the suppression of multiples in common-depth-point (CDP) or common-mid-point (CMP) gathers by isolating coherent energy associated with a particular velocity, from other coherent energy.

Mathematical Background

There are many ways in which the KL transformation can be derived (Ulrych et al. 1983). The approach of Kramer and Mathews (1956) is, however, very straightforward and revealing. We first present the essence of that paper:

For data compression, we consider the problem as follows. Given a set of n real signals $x_i(t)$ $(i = 1 \cdots n)$, we define a transformed set $\psi_j(t)$ and a transformation (rotation) matrix $\mathbf{A}$ (yet to be defined) such that

$$\psi_j(t) = \sum_{i=1}^{n} a_{ij} x_i(t) \qquad j = 1 \cdots n, \tag{1}$$

where a_{ij} are the elements of A (see appendix A). The signals $\psi_j(t)$ are chosen such that they form an orthogonal basis, so that each signal $x_i(t)$ can be expressed (exactly) as

$$x_i(t) = \sum_{j=1}^{n} b_{ij} \psi_j(t) \qquad i = 1 \cdots n,$$

or approximately as

$$\tilde{x}_i(t) = \sum_{j=1}^{m} b_{ij} \psi_j(t) \qquad i = 1 \cdots m; \quad m < n, \tag{2}$$

where $\tilde{x}_i(t)$ is the ith reconstructed signal, b_{ij} are the elements of $\mathbf{B}$, the inverse transformation matrix, and m is the number of basis functions used in the truncated expansion.

The objective at this point is to reconstruct $x_i(t)$ to within a given error using the smallest possible number of basis signals. For a given m, we require that the transformation matrices $\mathbf{A}$ and $\mathbf{B}$ be those that minimize the least-squares error

$$\phi(m) = \sum_{i=1}^{n} \int_0^T (x_i(t) - \tilde{x}_i(t))^2 \, \mathrm{d}t. \tag{3}$$

Kramer and Mathews (1956) showed that the rows of the matrix **A** consisted of the normalized eigenvectors of the covariance matrix **Γ**, defined as

$$\gamma_{ij} = \int_0^T x_i(t)x_j(t)\,\mathrm{d}t, \tag{4}$$

where γ_{ij} are the elements of **Γ**, and that **B** = **A** (in the notation of Kramer and Mathews (1956), our **A** would be transposed). The covariance matrix **Γ** is symmetric and positive semidefinite and hence is decomposable

$$\mathbf{\Gamma} = \mathbf{R}\Lambda\mathbf{R}^{\mathrm{T}}.$$

Here $\Lambda = \mathrm{diag}\,(\lambda_1, \lambda_2 \cdots \lambda_n)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n$. The columns of **R** contain the normalized eigenvectors $\mathbf{r}_j$ where $\mathbf{\Gamma r}_j = \lambda_j \mathbf{r}_j$. When **A** = **R**, the rotated signals $\psi_j(t)$ form an n-dimensional subspace of a Hilbert space. Using these basis elements defined in (2), the truncation error from (3) is

$$\phi(m) = \sum_{j=m+1}^{n} \lambda_j. \tag{5}$$

Since the eigenvalues of **Γ** are arranged in descending order it follows that the first basis function can be used to reconstruct more of the total signal energy than any other basis function. Thus, it is called the first principal component. Similarly, the second basis function can be referred to as the second principal component, etc.

For the applications to be described, it is important to note:

(a) The Karhunen–Loéve transform produces a set of uncorrelated (orthogonal) principal components from the data set; the size of the jth eigenvalue is a measure of the amount of coherent energy present in the jth principal component (appendix A). Hence, reconstructing the original signals using only those principal components which are associated with 'fairly large' eigenvalues amounts to reconstruction of the coherent energy present in the input seismograms.

In the context of this paper, 'coherent' refers to events which are similar horizontally in a trace-to-trace sense. This similarity is later extended to events which appear similar along specified dip directions.

Conversely, reconstructing the original data from those principal components associated with the smaller eigenvalues amounts to reconstructing the less coherent or 'anomalous' parts of the input data.

(b) If the original data $x_i(t)$ consist of scaled versions of some basic signal, then all the eigenvalues of **Γ** with the exception of λ_1 are zero and the first principal component is a scaled version of the same basic signal.

(c) If the original data possess no trace-to-trace coherency (i.e., if they are orthogonal), then **Γ** is diagonal and the data set itself corresponds to the set of principal components. Hence no advantage can be gained by using the transform.

IMPLICATIONS

These points carry a number of implications which suggest various applications in geophysical data processing. Several applications have already been discussed by Jones (1985), Ulrych et al. (1983) and Levy et al. (1983). Here we take advantage of these implications to deal with two further topics in multichannel seismic signal processing. After introducing the applications with reference to the background information, we present examples of both synthetic and real data to demonstrate the efficacy of the method.

Isolating coherent components

The separation of coherent information present in a set of seismic signals is achieved by reconstructing the input data using only those principal components which are associated with the 'larger' eigenvalues of the covariance matrix $\mathbf{\Gamma}$. In practice, a number of reconstructions have to be made, as the choice of the 'best' reconstruction is subjective. In the mode of operation to be followed here, the number of principal components used in the reconstruction is determined by the requirement that a specified amount of the coherent energy present in the input be present in the output. We express this requirement in terms of the percentage reconstruction energy

$$\varepsilon(m) = 100 \sum_{j=1}^{m} \lambda_j \Big/ \sum_{j=1}^{n} \lambda_j, \tag{6}$$

and choose m accordingly.

Isolating steeply dipping events and anomalous features

Given a set of seismograms featuring both flat and dipping events, we note that their covariance matrix $\mathbf{\Gamma}$ contains information pertaining to all the various dips present. However, with the definition of $\mathbf{\Gamma}$ given in (4), the shallowly dipping events will be recognized as the most strongly correlated portion of the section and hence they may be reconstructed faithfully from principal components which correspond to the larger eigenvalues. Conversely, the dipping events are recognized as less coherent and can be reconstructed from the principal components corresponding to the smaller eigenvalues. This later reconstruction is achieved by

$$\hat{x}_i(t) = \sum_{j=m+1}^{n} b_{ij} \psi_j(t) \qquad i = 1 \cdots n; \quad m \leq n. \tag{7}$$

In practice, the index m needed for the above separation scheme is found by trial reconstructions.

The procedure described also holds for anomaly reconstruction. In this application we reconstruct the zone of interest by discarding the most coherent information.

Separating 'Signal' from 'Noise'

Synthetic data examples

Separating steeply dipping and shallowly dipping events. To demonstrate how the KL reconstruction segregates events of different dip, we use a set of synthetic seismograms depicting shallowly dipping events overlain by steeply dipping events.

Twenty-four synthetic seismic traces representing shallowly dipping events superimposed on steeply dipping coherent 'noise' events, such as those introduced into a final stacked seismic section by marine streamer cable motion (Larner, Chambers, Yang, Lynn and Wai 1983) or ground roll, were constructed. We introduced a progressive phase change (from 0 to $\pi/3$ rad) to the wavelets across the events representing the geological horizons, and introduced a vertical fault of offset 36 ms at the 10th trace. We then added 10% white noise (all noise levels are expressed here as a percentage of the maximum trace amplitude).

Figure 1a shows the seismic representation of the basic geological model—phase shifted wavelets in shallowly dipping horizons offset by a vertical fault. Figure 1b is the 'noise' section, i.e., steeply dipping events plus random noise. Figure 1c shows the contaminated synthetic seismic section described above (i.e., the sum of 1a and 1b). In fig. 1d we see a 75% reconstruction of the data—the criterion governing reconstruction was the reconstruction energy, defined in (6). We note that the events which deviate from the flat-lying character of the model have been severely attenuated, leaving the representation of the underlying 'geological' structure basically intact.

Conversely, using (7), we can reconstruct to discard the most common part of our data. In fig. 1e we have reconstructed the portion of the information contained in the remaining principal components not used in fig. 1d. It is evident that the latter reconstruction has brought forth the steeply dipping events which are present in fig. 1c.

The phase character of the signals has been preserved, as has the distinctiveness of the 'fault' edge. However, we note an increase in the background noise in comparison to fig. 1a. This is a residual effect brought about by discarding principal components which contained contributions from both the dipping noise and the flat lying waveforms. The 24 principal components corresponding to the data in fig. 1c are seen in fig. 1f. The first five principal components are characteristic of flat lying events and combine with the appropriate weights to produce the data of fig. 1d. The next few principal components (6–12) show dipping patterns which combine to form the dipping waveforms of the input data. The principal components have been normalized for plotting, otherwise only the first few would have a discernible amplitude.

Noise suppression. The second example we consider is similar to the first, but with 50% white noise added to the first seven and last eight traces. Note that the fault location is between traces 7 and 8 in this example. In figs 2a and b, we see the seismic data and noisy section after bandpass filtering (0–50 Hz). Attempting to isolate the model from the noise, we reconstructed the seismograms in fig. 2b to 72% of the total input energy. Figure 2c shows that this attempt has been

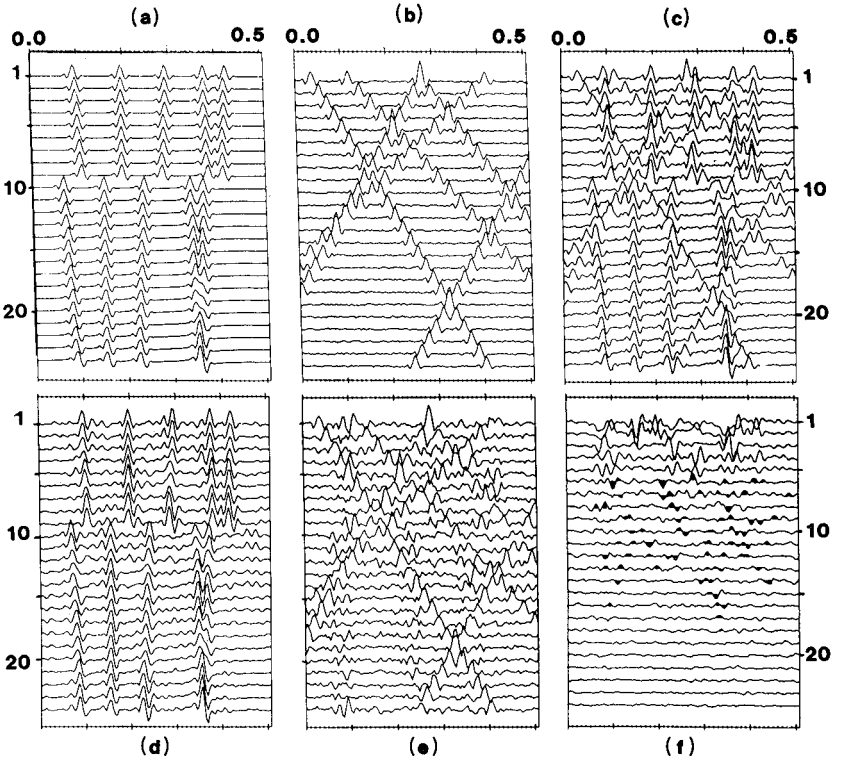


Fig. 1. (a) Seismic representation of the geological model: phase shifted wavelets in shallowly dipping horizons (up to 2 ms per trace), offset by a vertical fault at the 10th trace. (b) The 'noise' section: steeply dipping coherent events (dips between 16 ms and 24 ms per trace, e.g., from marine streamer noise or ground roll) and 10% random noise. (c) A sum of the previous two data sets: this is the input for the processing. (d) A 75% reconstruction of fig. 1c, requiring the first five of the 24 principal components. Note the preservation of phase information, and the clarity of the fault edge. (e) A misfit reconstruction (principal components 6–24) of the data shown in fig. 1c, isolating the dipping noise events. (f) The 24 principal components corresponding to fig. 1c. The first five principal components are dominated by strong peaks characteristic of flat lying structure. Note the smaller amplitude peaks in the band of principal components 6–17 (shaded). These waveforms combine to form the steeply dipping events.

reasonably successful—the noise level has been considerably reduced, and the phase character and fault edge have been well preserved. Mostly, the fault edge is not smeared over adjacent traces but remains distinct.

Slant-KL. If the geological model comprises more steeply dipping parallel (or sub-parallel) beds, then we could use the 'slant-KL' procedure described in appendix B. This modifies the calculation of the covariance matrix so that the segregation of dipping events is biased in favour of those with the specified dip ((B1) in appendix B).

A data set similar to that used in example 1 was constructed, but the dips on the 'geological' horizons were increased by a factor of five. Following the same procedure, we see in fig. 3 (a) the seismic section, (b) the section plus noise and 'streamer noise', (c) the data after a dip of 8 ms per trace has been removed, (d) the 75% reconstruction depicting the underlying structure, (e) a misfit reconstruction isolating the dipping noise events, and finally, (f) the reconstructed data with the dip reinstated. The phase character and fault definition are well preserved, whereas the dipping noise events have been severely attenuated.

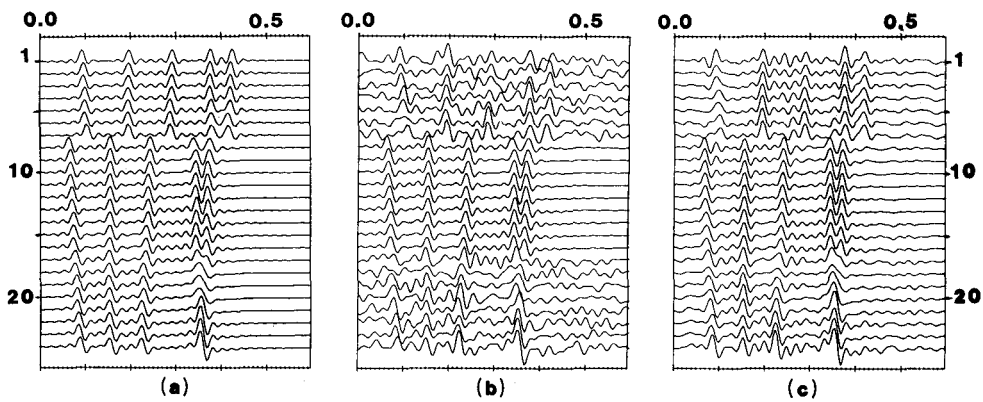


Fig. 2. (a) Synthetic data similar to that of fig. 1a, with a vertical fault between traces 7 and 8, after bandpass filtering (10–50 Hz). (b) The data after addition of 50% random noise to the first 7 and last 8 traces, and bandpass filtering as above. (c) A 72% reconstruction requiring four of the 24 principal components. Note the marked reduction in the background noise level and the preservation of the essential features.

Real data examples

Following the trials on synthetic data, we proceed to exemplify the method's ability to enhance coherency in real stacked data. In the reconstructions shown, it is noted immediately how the uncorrelated background noise level has been reduced and trace-to-trace coherency thus enhanced.

In fig. 4a we have selected a window of 100 traces from a conventionally processed stacked section. The traces were decomposed into their principal components: the first five principal components accounted for 85% of the input energy, and the first twelve principal components for 95% of the energy, i.e., for this data we need only 12% of the 100 principal components to almost perfectly reconstruct the input data. Figures 4b and 4c, respectively, show the 95% and 85% reconstructions. Note the increase in coherency in the zone 1.1–1.5 s, which consists of interbedded sand–shale sequences.

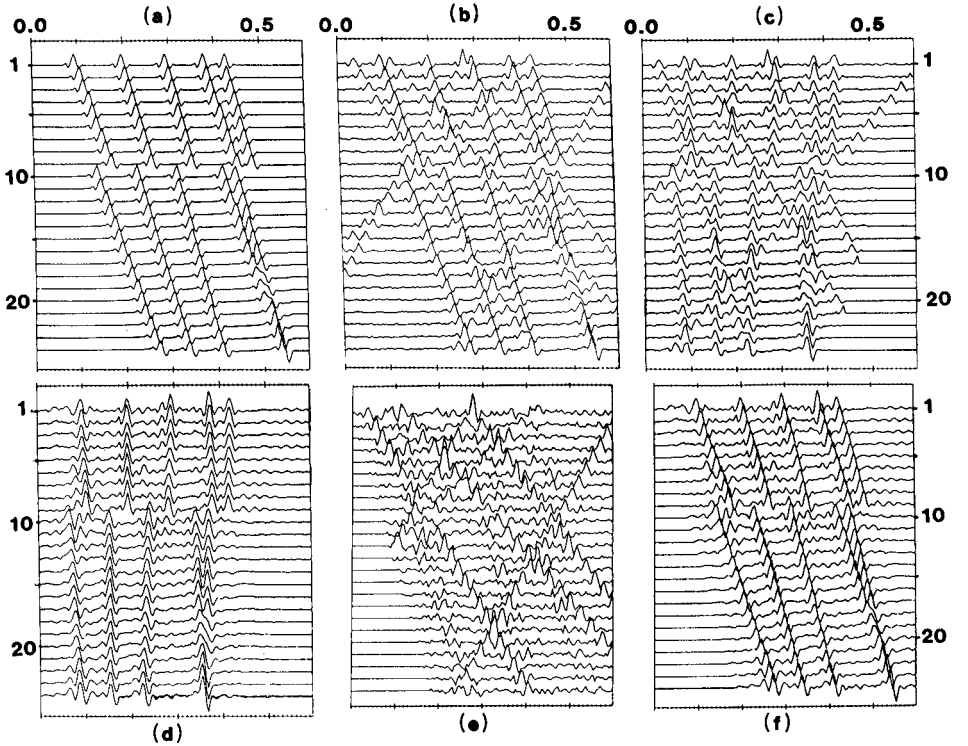
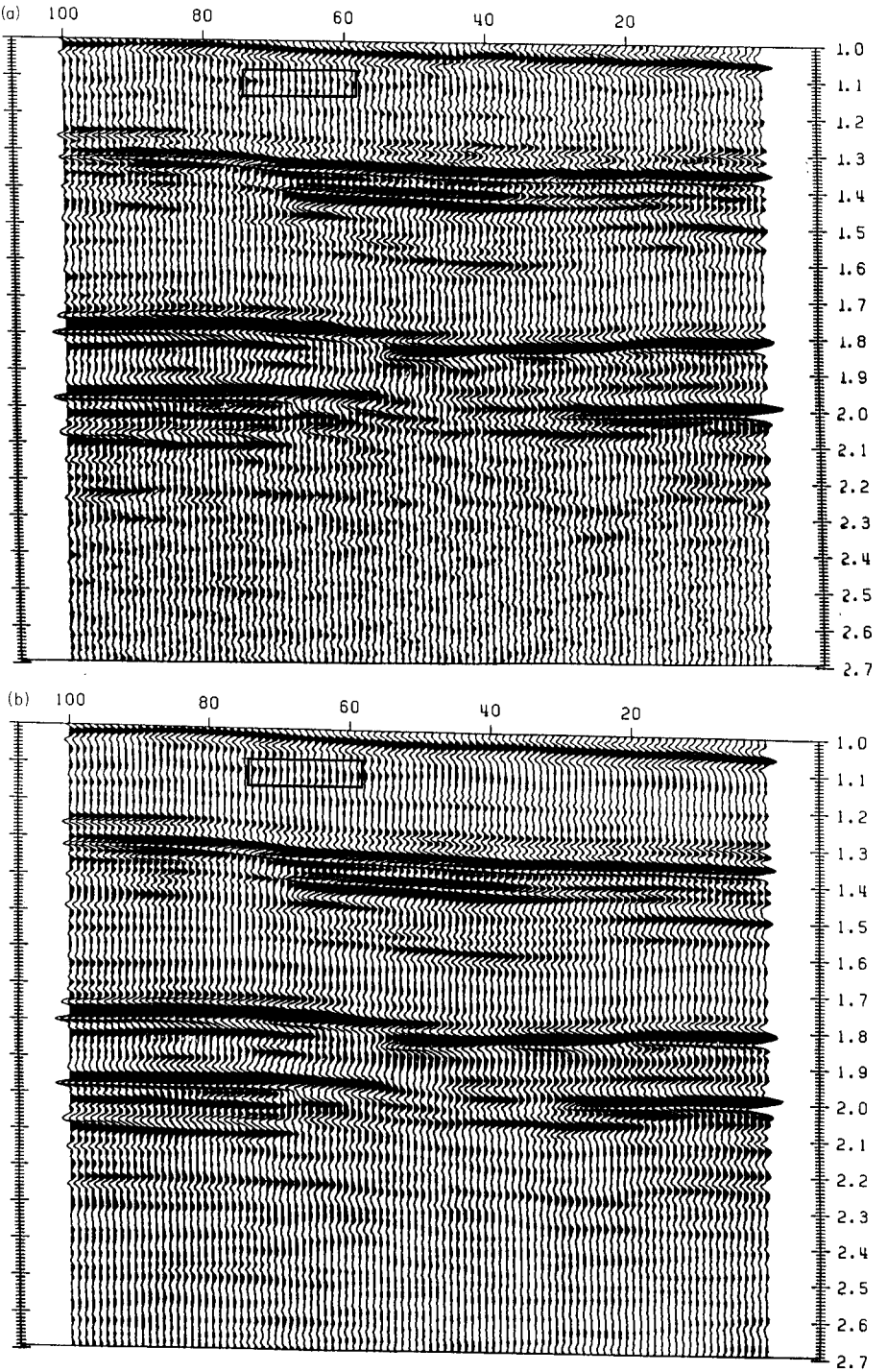


Fig. 3. (a) Steeply dipping (between 7 ms and 10 ms per trace) sub-parallel phase shifted events with a vertical fault. (b) The data plus the dipping 'noise' events of fig. 1b. (c) The flattened contaminated data, i.e., fig. 3b after a dip of 8 ms per trace has been removed. (d) A 75% reconstruction requiring 5 of the 24 principal components. (e) A misfit reconstruction (principal components 6-24) of the data shown in fig. 3c, isolating the dipping noise events. (f) The reconstruction after the dip has been reinstated. Again, the phase character and fault definition are well preserved, whereas the dipping noise events are severely attenuated. The background noise level has increased in places, since we have omitted some principal components which contained both the desired signal and the dipping noise events.

In the 95% reconstruction, there is a general increase in coherency, while small-scale features are also preserved, e.g., the small lens-like feature depicted in the box at about 1.1 s between traces 55–75, is enhanced in the 95% reconstruction. However, discarding more principal components—as done with the 85% reconstruction—leaves the section looking very smooth and continuous, especially over areas with small-scale features. The lens-like feature is obliterated in this reconstruction. The phase structure of the major horizons remains intact in both reconstructions; this can be seen by examining the waveforms along the edges of the section.

To demonstrate the slant-KL procedure, in fig. 5 we consider 96 traces from a conventionally processed stacked section. The events in this section dip to the left


(a)
100
80
60
40
20
1.0
1.1
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9
2.0
2.1
2.2
2.3
2.4
2.5
2.6
2.7
(b)
100
80
60
40
20
1.0
1.1
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9
2.0
2.1
2.2
2.3
2.4
2.5
2.6
2.7

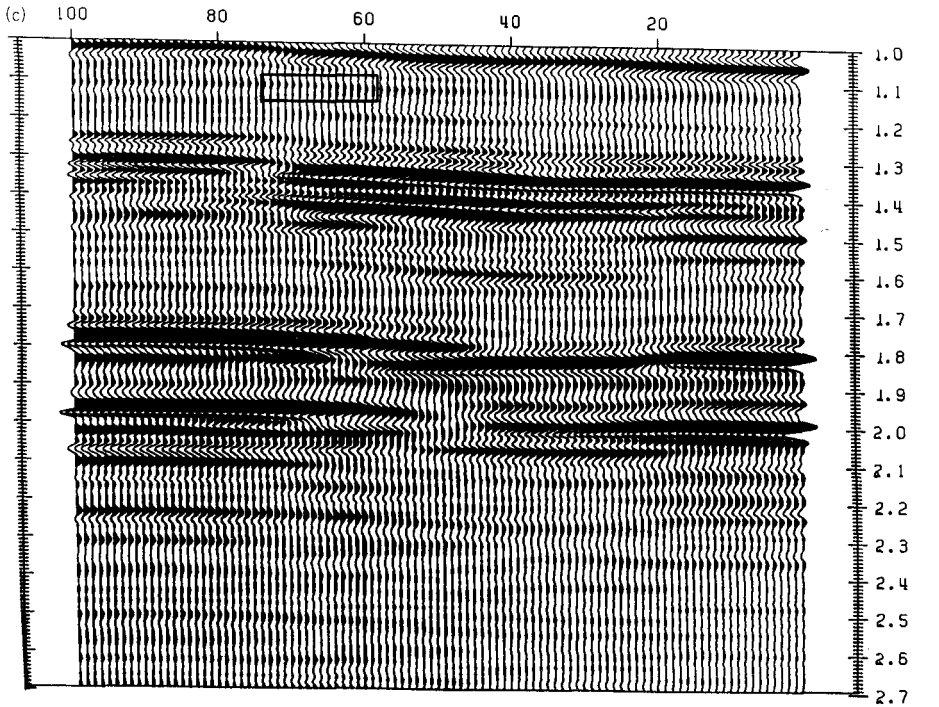
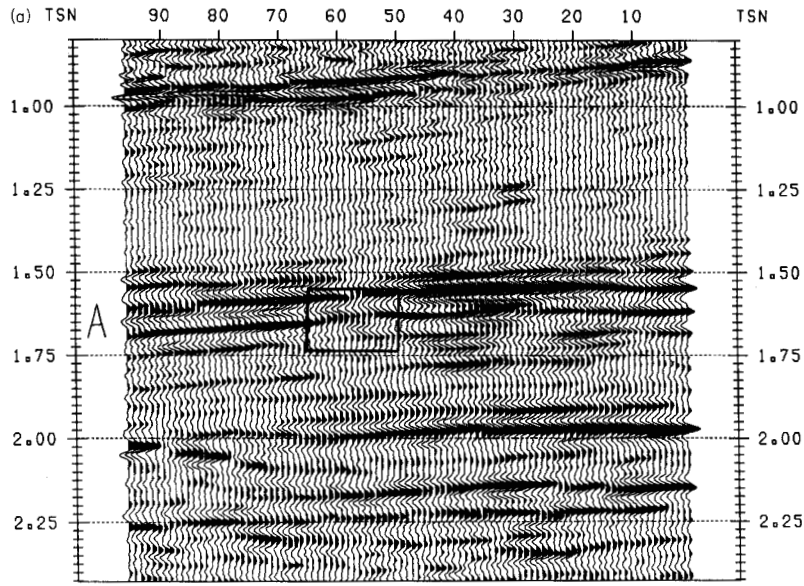
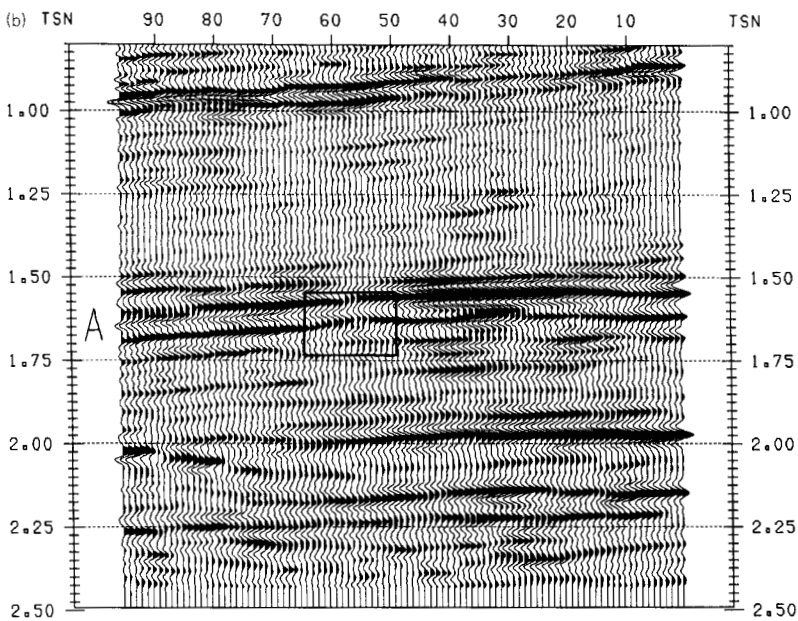


Fig. 4. (a) One hundred traces from a conventionally processed stacked section. Note the discontinuous events in the lower part of the section (2.1–2.7 s) and the small lens-like feature in the box at 1.1 s. (b) The 95% reconstruction requiring 12 of the 100 principal components. In this reconstruction, the background of incoherent energy has been reduced greatly, and small scale features such as the lens highlighted in the box are preserved. (c) The 85% reconstruction requiring 5 of the 100 principal components. In this reconstruction, the background of incoherent energy has also been reduced greatly, but small-scale features such as the lens highlighted in the box have been obliterated, because by this reconstruction we have left only the gross features.

about 0.1 s over the 96 traces, and the computation of the covariance matrix was adjusted to accommodate this dip. The strong pair of arrivals seen above 1.0 s in the input data (fig. 5a) are clarified markedly in the 95% reconstruction (fig. 5b). However, the 85% reconstruction (fig. 5c) loses much of the finer detail. A further example of loss of resolution is the small 'pull-up' type event in the box A. This feature is well preserved in the 95% reconstruction, but lost in the 85% reconstruction.

As a corollary to truncated reconstruction, the following point is important. In the examples we saw how the data could be almost perfectly reconstructed from a relatively small subset of principal components, i.e., the required storage space will be drastically reduced, and data transmission rates from one storage medium to another will be increased (Kramer and Mathews 1956, Ready and Wintz 1973). This


(a) TSN
90 80 70 60 50 40 30 20 10
TSN
1.00
1.25
1.50
1.75
2.00
2.25
A


(b) TSN
90 80 70 60 50 40 30 20 10
TSN
1.00
1.25
1.50
1.75
2.00
2.25
2.50
A

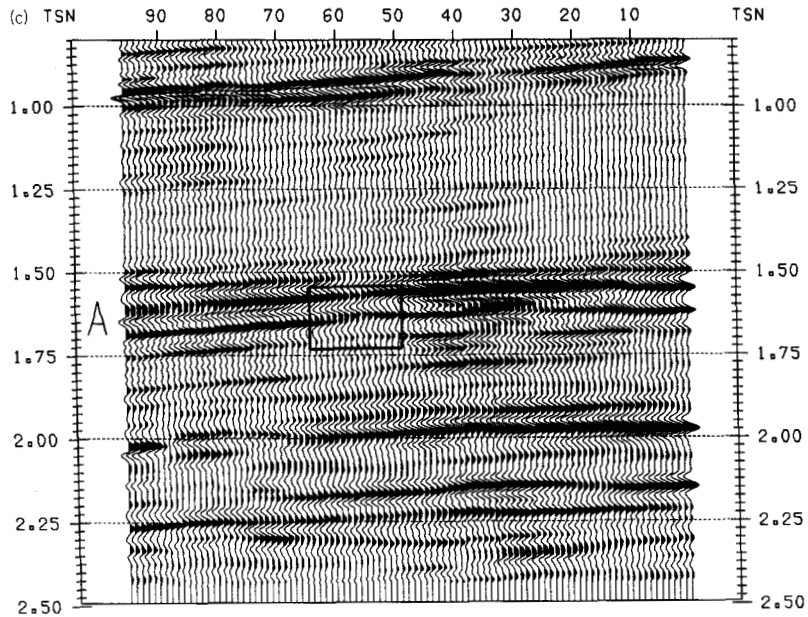


Fig. 5. (a) Ninety six traces from a conventionally processed stacked seismic section. Note the background noise surrounding the pair of dipping events above 1.0 s. Also note the pull-up in box A. (b) Slant-KL 95% reconstruction: the dipping events have been clarified, and the pull-up has been left intact. (c) Slant-KL 85% reconstruction: resolution has been lost, and the pull-up in box A has been smeared-out.

latter point is of concern when data are being shunted from a mainframe computer to an auxiliary machine for processing. In the data examples processed to date, we can typically dispense with 70–90% of our required storage space.

Multiple Suppression in CDP or CMP Seismic Data

For the purpose of multiple suppression, we utilize the energy packing property of the KL transform. The idea here is to segregate the energy associated with the multiples onto a single principal component. A data reconstruction omitting that principal component should be essentially multiple-free.

We proceed in five basic steps:

1. From a velocity analysis, identify the stacking velocity and onset time associated with the observed multiples.
2. Using this stacking velocity, apply a constant velocity normal moveout correction to the data. At this stage the arrivals due to multiples have been more or less flattened, whereas the primary events are under- or over-corrected, and have increased curvature in the section.

3. Compute the KL transform of these moveout-corrected data. Correlated multiple energy in the CMP gather now appears predominantly on the first principal component.
4. Reconstruct the NMO corrected CMP gather from the principal components omitting the first principal component, i.e., leave out any correlated energy associated with the multiples. It may also be advantageous to omit the second principal component when the waveform of the multiples has been severely distorted by interfering primary events, or by the NMO operation.
5. Remove the moveout stretching from the reconstructed data using the same velocity as in step 2.

This method is similar to the standard FK filtering technique. However, the alternative approach described here will not be as prone to the wrap-around problems encountered with the FK method.

Synthetic data examples

In fig. 6a, we show a simple synthetic seismic section representing reflection events from nine flat layers over a half-space, all overlain by water. Included are two events

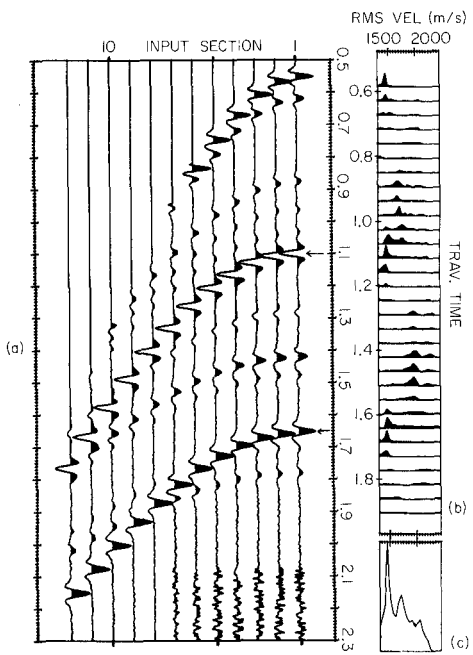


Fig. 6. (a) The seismic representation of reflections from nine layers overlying a half-space, all overlain by water. The data have been muted and AGC'd. Two water bottom multiple events can be seen: the first at 1.10 s and the second at 1.65 s. (b) The velocity analysis of the data. Note the peaks due to the multiples with a velocity of 1450 m/s which dominate the lower portion of the velocity analysis. (c) Stack of the first half of the velocity analysis map (b). As multiple events tend to lie vertically in the velocity analysis map, they will stack to produce a maximum.

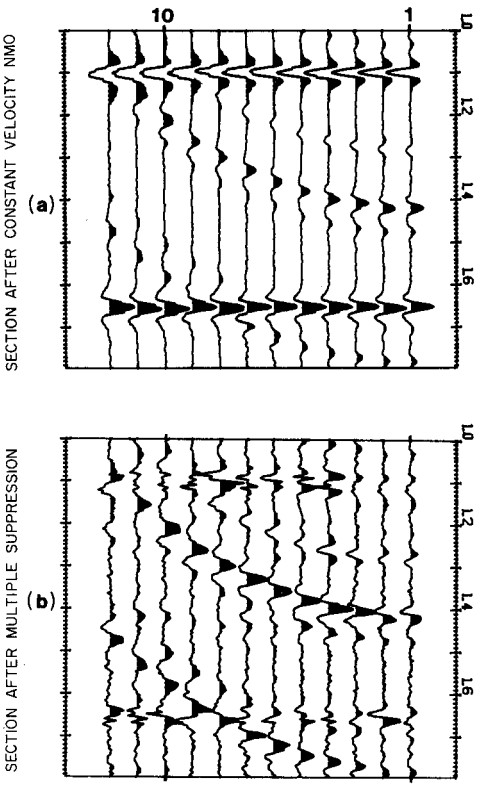


Fig. 7. (a) A data window after move-out correction with a constant velocity of 1450 m/s. Note the flattening of the multiple arrivals. (b) A misfit reconstruction of the NMO corrected data omitting the first two principal components. Some residual noise remains in the locations formerly occupied by the multiple events, but this is largely incoherent, and will not stack to produce a noticeable effect.

due to 'multiple' travel paths in the surficial water layer (indicated by arrows). Figure 6b shows the semblance velocity analysis (Neidell and Taner 1971) of this data. We see 'multiple' energy at 1.10 and 1.65 s with a characteristic velocity of 1450 m/s. In fig. 6c, we show a stack of the velocity map: this highlights the velocity associated with the multiples, and is used to decide upon the NMO velocity for multiple suppression. Figure 7a shows the data after a constant velocity NMO correction. Note that the multiple events at 1.10 and 1.65 s are now flattened.

Applying the KL decomposition to the data window 1.00–1.80 s, and reconstructing (omitting the first and second principal components) gives the results in fig. 7b. The result of this latter procedure is then un-NMO'd to yield the data of fig. 8a. A velocity analysis of the un-NMO'd data after multiple suppression processing is shown in Fig. 8b. By comparison with fig. 6 the multiples have been effectively removed from the data. Figure 9a shows the multiple suppressed CMP gather after

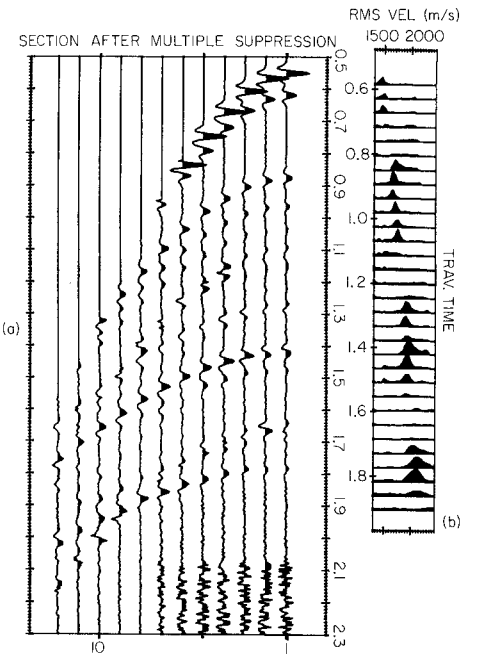


Fig. 8. (a) The processed CDP gather after it has been un-NMO'd. (b) The velocity analysis of the data after multiple suppression. Note that in the absence of multiples the primary stacking velocity profile can be more readily discerned than in fig. 6b.

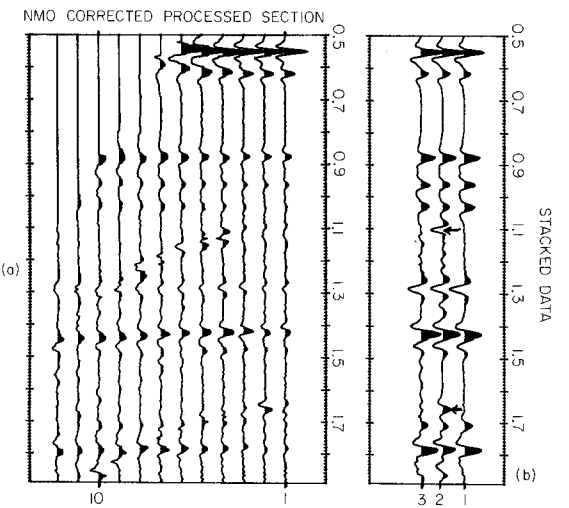


Fig. 9. (a) The multiple-suppressed data after normal moveout correction using a velocity function picked from fig. 8b. The residual events remaining from the multiples can still be seen, but these do not stack constructively. (b) Three versions of a stack of the data: (1) the stack of a synthetic data set produced without multiples: this is our desired, or optimum result; (2) the stack of the data in fig. 6a after NMO correction. Notice the multiple arrivals at 1.10 s and 1.65 s which stacked constructively; and (3) the stack of the multiple suppressed data in fig. 8a after NMO correction. In comparison with (2) note the absence of multiple events.

NMO correction using the stacking velocity profile of fig. 8b. In fig. 9b, we see a comparison of the trace resulting from:

(a) the stack of the ideal data which do not contain multiple arrivals;
(b) the stack of the data with multiples; and
(c) the stack of the data after multiple suppression using the KL technique.

A comparison of the stacked traces in figs 9b(2) and 9b(3) shows that the multiples at 1.10 and 1.65 s have been successfully suppressed. The noise introduced into the section after removal of the first two principal components (fig. 7b) did not stack constructively.

Real data examples

Of greater interest is the performance of the algorithm on real seismic data when multiples pose a problem. Figure 10 shows a CDP gather of marine seismic data. Each gather has 60 traces sampled at 4 ms, but for this example, only every third trace was used. Figure 11a shows a broad-velocity-band velocity analysis of the gather. We see a dominant trend of multiple energy at 1620 m/s starting at 0.50 s. Figure 11b shows a stack of the velocity analysis map, highlighting the presence of the multiples: this information is used to decide upon the 'multiple' stacking velocity. Figure 11c shows the velocity analysis of fig. 10 after multiple suppression using a constant stacking velocity of 1620 m/s. Note the absence of the band of multiple

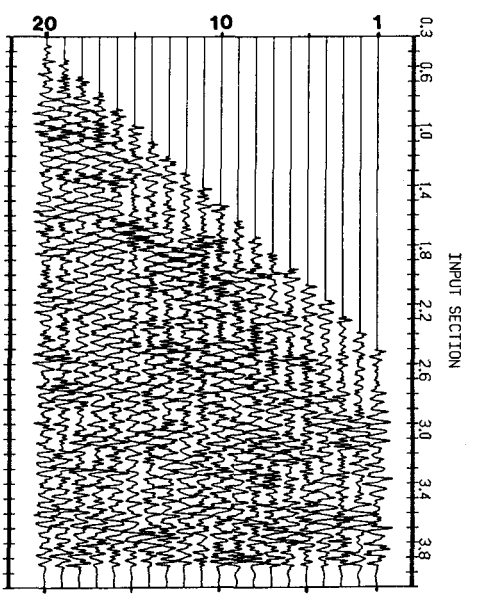


Fig. 10. A marine CMP gather after bandpass (5–55 Hz), muting, and application of an AGC.

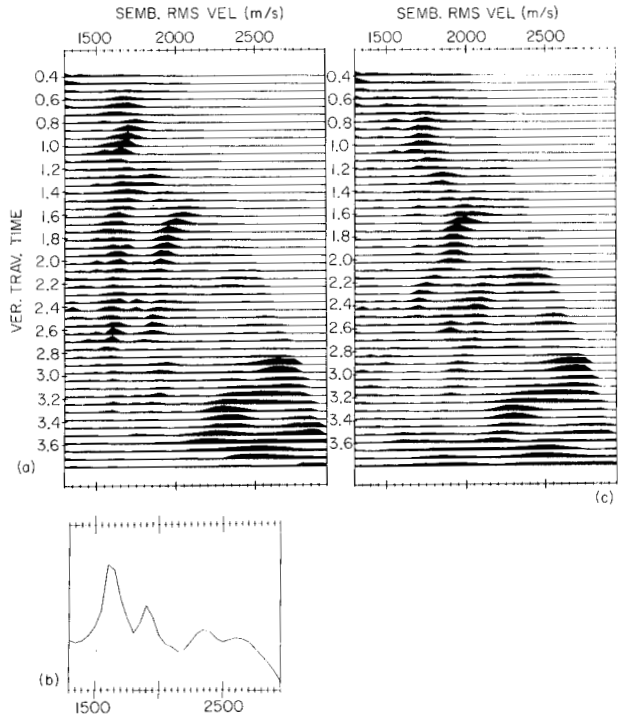


Fig. 11. (a) Velocity analysis of the data in fig. 10. Note the vertical trend of energy at 1620 m/s due to water bottom related multiples. (b) The stack of the velocity analysis map above. The maximum at 1620 m/s was used as the multiple NMO velocity. (c) The velocity analysis map of the section after multiple suppression. Note the absence of energy at 1620 m/s after 0.5 s, and the enhancement of the primary arrival peak at 2.2 s with velocity 2400 m/s.

reflections between 0.50 and 3.40 s. Note also how the event at 2.20 s and 2400 m/s has been enhanced in the multiple-suppressed velocity analysis; it is no longer obscured by arrivals due to multiple travel paths.

Twenty gathers of data were processed in this way, and a comparison of the conventionally stacked data with a stack of the multiple-suppressed data is made in figs 12a and 12b (here, two principal components were discarded in the reconstruction). The most noticeable differences between the multiple-suppressed stack and the conventional stack are the absence of the event at 0.91 s (marked A) in the multiple-suppressed section, which appears as part of a 'doublet' in the unprocessed data, and the changes at 1.35 s (B). Several events are clearer in the multiple-suppressed section, such as that at 0.86 s (C), the trough at 1.10 s (D), and the pair at 2.45 and 2.55 s (E). Subtracting these two sections yields a difference section (fig. 12c) which emphasizes the location and nature of the differences between the processed and unprocessed data. Bands of multiple related energy are seen with 0.44 s spacing.

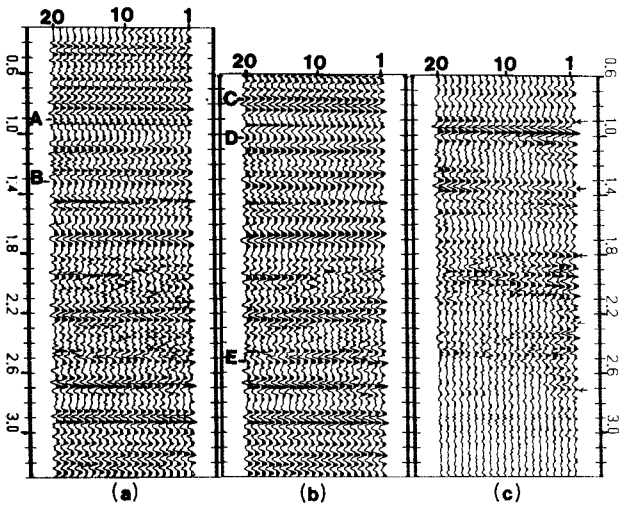


Fig. 12. (a) The stack of 20 gathers. (b) The same stacked section after multiple suppression using the velocity of 1620 m/s. Note the absence of the event at 0.91 s (A) and changes at 1.35 s (B). Also note the enhancement of events at 0.86 s (C), 1.10 s (D), and the pair at 2.45 s and 2.55 s (E). (c) The difference section (stack minus stack after multiple suppression) to emphasize the location and nature of the differences between the processed and unprocessed data. Bands of energy associated with the multiples appear at about 0.44 s intervals, indicated by the arrows.

Discussion

The first section introduced the application of the KL transform to the recovery of coherent information for image enhancement of stacked seismic sections. The stacked section reconstruction technique has proved very successful so far (Jones 1985), and has been applied extensively to real data in industrial processing. Its main advantage is the ability to greatly reduce the background level of incoherent noise, which is often prevalent early in a seismic section. The choice of a suitable reconstruction energy is subjective, but for most of the data examined, 90–95% reconstructions give reasonable results. Dropping the reconstruction energy below 90% results in loss of resolution, as seen in many of the 85% reconstructions presented.

As the covariance matrix is 'tuned' to isolate events with a preferred dip direction, there is a problem when conflicting dips are present. To partially circumvent this problem, the data are broken into blocks of about 100 traces in width, and a few hundred milliseconds in depth, for KL decomposition. The final reconstructed section is a montage of these smaller segments. However, when there are widely varying dips within a small block, the method attenuates events with dips that differ most from the one preferred by the covariance computation.

In data compression for storage and transmission, the transform has a singular usefulness, in that we can represent the data almost exactly with only a fraction of

the information otherwise required. The reduction in cost, especially when transferring data from a mainframe computer to a work station, can be significant.

Multiple arrivals often pose a serious problem in shallow water and certain land environments. Simple and effective methods for the elimination of multiple events can alleviate these problems. Here we present an intuitively simple method for isolating and removing multiple events prior to final moveout correction and stack. On simple synthetic data (layered-earth-model) the method works well. On the small sample of real data analysed, the velocity analysis maps highlight the almost complete absence of multiple-related energy after processing. Primary events which were of small amplitude in the velocity analysis map before processing were also enhanced, because the actual data hyperbola corresponding to the primary event was no longer masked by multiple energy arriving at the same time.

The results on the basis of 'before and after' velocity analysis maps look very promising. Initial tests on 20 gathers show several noticeable differences and their velocity analysis maps show that a significant number of multiple events have been removed. Also important is that the actual waveforms are essentially uncorrupted by the processing.

ACKNOWLEDGMENTS

We thank R.M. Clowes and R.M. Ellis for their support and T.J. Ulrych and D.W. Oldenburg for their helpful discussions.

This work was supported partly by the Canadian Natural Sciences and Engineering Research Council (grants A7707, A2617, A4270, and A1804) and Mobil Oil Canada Ltd. Inverse Theory and Applications (IT&A) Inc., generously made available the computing facilities and some necessary data. We also thank Dr J. den Boer, of Mobil Oil Canada, for kindly supplying data for this project.

APPENDIX A

The energy of the principal components

Given the data matrix $\mathbf{X} = \{x_i(t), i = 1, \ldots, n\}$, where the seismic traces $x_i(t)$ are the rows of X, we firstly compute the covariance matrix $\mathbf{\Gamma}$, for the data

$$\mathbf{\Gamma} = \mathbf{X}\mathbf{X}^{\mathrm{T}} \tag{A1}$$

(it is understood that the ith row of a matrix $\mathbf{X}$ contains the digitized version of a function $x_i(t)$).

We may reduce $\mathbf{\Gamma}$ to its diagonal form by means of a spectral decomposition

$$\mathbf{\Gamma} = \mathbf{R}\mathbf{\Lambda}\mathbf{R}^{\mathrm{T}},$$

where $\mathbf{R}$ is the matrix of column eigenvectors $\mathbf{r}_j$, and the diagonal matrix $\mathbf{\Lambda}$ contains the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_j$, arranged in decreasing size.

Further, using singular value decomposition (SVD) we may write

$$\mathbf{X} = \mathbf{R}\mathbf{\Omega}\mathbf{V}^{\mathrm{T}}, \tag{A2}$$

where $\mathbf{\Omega}$ is a matrix containing the singular values of the decomposition along its main diagonal, with zeroes elsewhere.

Using (A2), (A1) becomes

$$\mathbf{\Gamma} = \mathbf{R\Omega V}^{\mathrm{T}}\mathbf{V\Omega}^{\mathrm{T}}\mathbf{R}^{\mathrm{T}},$$

and as both $\mathbf{R}$ and $\mathbf{V}$ are in general unitary, this becomes

$$\mathbf{\Gamma} = \mathbf{R\Omega\Omega}^{\mathrm{T}}\mathbf{R}^{\mathrm{T}}.$$

Now, we constructed the principal components $\Psi = \{\psi_j(t), j = 1 \cdots n\}$ as

$$\Psi = \mathbf{R}^{\mathrm{T}}\mathbf{X} \tag{A3}$$

and using (A3) and (A2) we also have

$$\begin{aligned}\mathbf{\Psi\Psi}^{\mathrm{T}} &= \mathbf{R}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{R},\\ &= \mathbf{R}^{\mathrm{T}}\mathbf{R\Omega V}^{\mathrm{T}}\mathbf{V\Omega}^{\mathrm{T}}\mathbf{R}^{\mathrm{T}}\mathbf{R},\\ &= \mathbf{\Omega\Omega}^{\mathrm{T}} = \mathbf{\Lambda}.\end{aligned}$$

The matrices $\mathbf{R}$ and $\mathbf{\Lambda}$ are both $[n \times n]$. However, for our n traces each of N points, we have that $\mathbf{X} = [n \times N]$, $\mathbf{\Omega} = [n \times N]$, and $\mathbf{V} = [N \times N]$, but for $n < N$, we only have at most n non-zero eigenvalues in the diagonal matrix $\mathbf{\Omega}$. Hence $\lambda^{1/2}$ can fully represent ω (i.e., the jth element of $\mathbf{\Omega}$, which is $[n \times N]$).

So we see that $\mathbf{\Lambda}$ is simply the covariance matrix of the principal components, i.e., λ_j is the energy content of the jth principal component. Furthermore, trace $[\mathbf{\Lambda}]$ is the total input energy, i.e.,

$$\text{trace } [\mathbf{\Lambda}] = \int_0^T \sum_{i=1}^{n} x_i(t)^2 \, \mathrm{d}t = \text{trace } [\mathbf{\Gamma}].$$

APPENDIX B

The slant-KL transform

For laterally well-correlated data (as in stacked sections corresponding to a horizontally or near horizontally layered earth), we need few principal components to adequately reconstruct the data (i.e., $\varepsilon(m) \simeq 100\%$ with $m < n$ (6)).

This latter statement may be extended to parallel dipping events by a simple modification of the covariance matrix $\mathbf{\Gamma}$, i.e., we adjust $\mathbf{\Gamma}$ so that it represents preferred time lags, corresponding to some desired dips.

This end is achieved by redefining the covariance matrix $\mathbf{\Gamma}$ as

$$\mathbf{\Gamma} = \int_0^T x_i(t)x_i(t - \Delta(i-1)) \, \mathrm{d}t, \tag{B1}$$

or equivalently, by producing a new data set $y_i(t)$ such that

$$y_i(t) = \delta(t - \Delta(i-1)) * x_i(t), \tag{B2}$$

where δ is the Dirac delta function, Δ is the dip of the beds expressed in time samples per trace, and '$*$' denotes convolution. Δ can be positive or negative, depending on the dip direction. We refer to the KL transformation applied to the modified data set $y_i(t)$ as a 'slant-KL' transform.

REFERENCES

AHMED, N. and RAO, K.R. 1975, Orthogonal Transforms for Digital Image Processing, 189–224, Verlag J. Springer.

ANDREWS, H.C. and PATTERSON, C.L. 1976a, Outer product expansions and their uses in digital image processing, IEEE Transactions on Computers 25, 140–148.

ANDREWS, H.C. and PATTERSON, C.L. 1976b, Singular value decomposition and digital image processing, IEEE Transactions on Acoustics, Speech and Signal Processing 24, 26–53.

HEMON, C.H. and MACE, D. 1978, Use of the Karhunen–Loéve transformation in seismic data processing, Geophysical Prospecting 26, 600–626.

HUANG, T.S. and NARENDRA, P.M. 1975, Image restoration by singular value decomposition, Applied Optics 14, 2213–2216.

JONES, I.F. 1985, Applications of the Karhunen–Loéve transform in reflection seismic processing, Ph.D. thesis, University of British Columbia.

KRAMER, H.P. and MATHEWS, M.V. 1956, A linear coding for transmitting a set of correlated signals, IRE Transactions on Information Theory IT–2, 41–46.

LARNER, K., CHAMBERS, R., YANG, M., LYNN, W. and WAI, W. 1983, Coherent noise in marine seismic data, Geophysics 48, 854–886.

LEVY, S., ULRYCH, T.J., JONES, I.F. and OLDENBURG, D.W. 1983, Applications of complex common signal analysis in exploration seismology, Proceedings of the 53rd Annual SEG meeting, Las Vegas, S6.6.

MALLICK, K. and MURTHY, Y.V.S. 1984, Pattern of Landsat MSS data over Zawar lead–zinc mines, Rajasthan, India, First Break 2, 16–21.

NEIDELL, N.S. and TANER, M.T. 1971, Semblance and other coherency measures for multichannel data, Geophysics 36, 482–497.

READY, R.J. and WINTZ, P.A. 1973, Information extraction, SNR improvement, and data compression in multispectral imagery, IEEE Transactions on Communications COM-21, 1123–1130.

ULRYCH, T.J., LEVY, S., OLDENBURG, D.W. and JONES, I.F. 1983, Applications of the Karhunen–Loéve transformation in reflection seismology, Proceedings of the 53rd Annual SEG meeting, Las Vegas, S6.5.