Geophysical Prospecting, 1997, **45**, 183–205

Characterizing seismic time series using the discrete wavelet transform[1]

H.J. Grubb[1] and A.T. Walden[2]

Abstract

The discrete wavelet transform (DWT) has potential as a tool for supplying discriminatory attributes with which to characterize or cluster groups of seismic traces in reservoir studies. The wavelet transform has the great advantage over the Fourier transform in being able to better localize changes. The multiscale nature and structure of the DWT leads to a method of display which highlights this and allows comparison of changes in the transform with changing data. Many different sorts of wavelet exist and it is found that the quality of reconstruction of a seismic trace segment, using some of the coefficients, is dependent on the choice of wavelet, which leads us to consider choosing a wavelet under a 'best reconstruction' criterion. Location shifts, time zero uncertainties, are also shown to affect the transform, as do truncations, resampling, etc. Using real data, examples of utilizing the DWT coefficients as attributes for whole trace segments or fractional trace segments are given. Provided the DWT is applied consistently, for example with a fixed wavelet, and non-truncated data, the transform produces useful results. Care must be exercised if it is applied to data of different lengths. However, as the algorithm is refined and improved in the future, the DWT should prove increasingly useful.

Introduction

We discuss an approach to summarizing the features of seismic traces using the discrete wavelet transform (DWT). For geophysicists it is important to note at the outset that the term wavelet in this paper does not refer to a seismic wavelet, but rather to a functional shape (the equivalent of which for the Fourier transform is a sine or cosine). The aim here is to provide several informative 'attributes' which characterize lateral changes in the seismic traces.

---

[1]Paper presented at the 56th EAEG meeting, June 1994, Vienna. Received April 1994, revision accepted June 1996.
[2]Dept. of Applied Statistics, The University of Reading, PO Box 240, Reading RG6 2FN, U.K. Formerly at[3].
[3]Dept. of Mathematics, Imperial College of Science, Technology and Medicine, 180 Queen's Gate, London SW7 2BZ, U.K.

Trace characterization

A typical trace-characterization study proceeds in three steps. Firstly, trace segments are extracted between consecutive geological events which have been identified on the sections: these segments thus refer to the same geological interval. Secondly, attributes are derived from each segment by the proposed method. Finally, the attributes are examined for clustering effects; i.e. homogeneous regions of attribute space which are also clustered in location are identified, indicating some form of similarity between traces in a cluster/region which can then be further investigated by a processor or geologist. Such clusters can inform on extrapolation of observed characteristics from wells, or point to regions for more detailed examination. Sometimes areas of poorly processed data are highlighted indicating that reprocessing of the data is required.

Attribute clustering can be carried out in several dimensions, for example with a single attribute, a simple thresholding could be used to split the data. With two attributes, one might look at points to one side of a line on a scatterplot of the attributes, or enclosed within a polygon. As an example, consider Fig. 1.

Figure 1a shows the locations of survey lines in a region from the southern North Sea, each dot or cross locating a particular trace segment. Figure 1b shows a scatterplot of two attributes (derived from the DWT and discussed later). The attributes are split into two groups, dots and crosses, and the locations of the traces from which the attribute values come are identified by the same symbols in Fig. 1a. The views can be dynamically 'linked,' so that changing the clustering of the attributes automatically changes the locations of the corresponding dots and crosses. This example is discussed further in later sections.

In higher dimensions, visualization becomes difficult, although we can use several linked scatterplots of pairs of attributes, an extension of the linking in Fig. 1, to give a view of higher dimensions (see e.g. Cleveland and McGill 1988). To help with this we can use a statistical clustering technique to give a projection of the data on to fewer dimensions: for example projection pursuit projects the attribute space on to one dimension (Jones and Sibson 1987; Lendzionowski, Walden and White 1990; Walden 1992) or two dimensions (Friedman 1987), enabling clusters to be readily perceived. We may also examine outlying points, perhaps indicating data problems or regions where the seismic behaviour is quite different from the rest of the data.

Choice of attributes and wavelet transform

The extraction of useful attributes from seismic traces has been an active research area for a number of years. For example, Hagen (1982) utilized principal component correlation coefficients for clustering, while Bois (1980, 1982) fitted autoregressive models to seismic trace segments, and used the model coefficients as new discriminatory variables. Dumay and Fournier (1988) used various 'conventional' summaries, such as energies and zero-crossings, as discriminatory variables. Lendzionowski *et al.* (1990) demonstrated that instantaneous attributes could successfully identify a gas–water

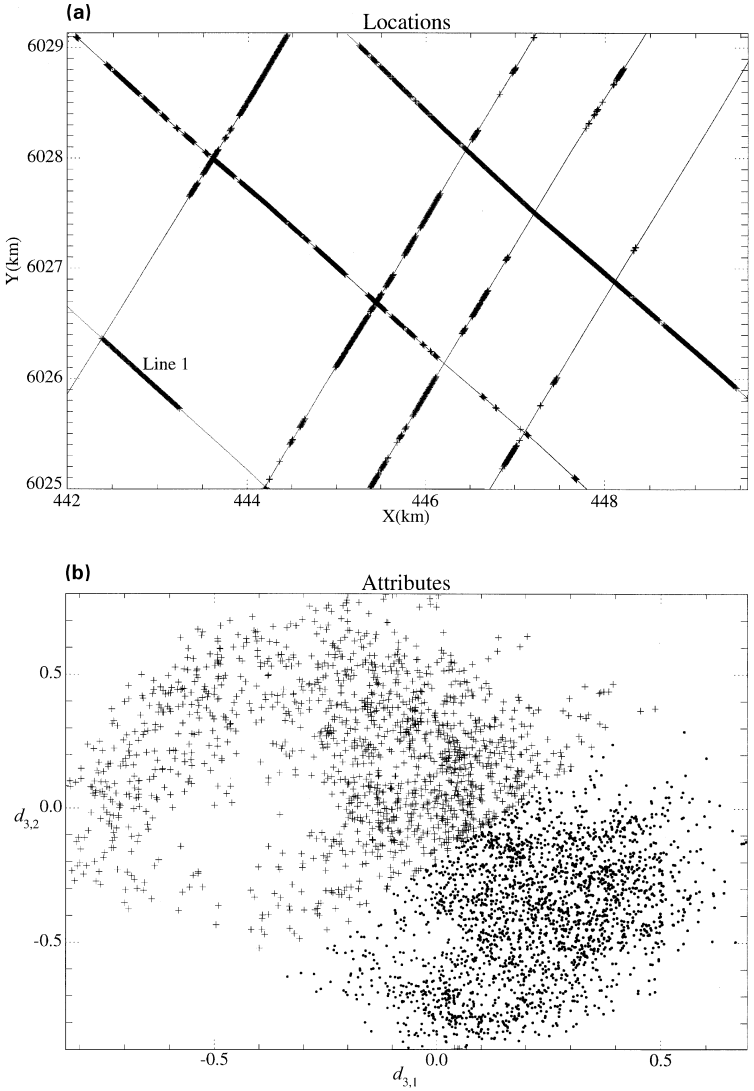

Figure 1. (a) Locations of survey lines in a region. Crosses and dots mark the classification of seismic trace segments according to the split of the scatterplot in (b), which consists of two-dimensional points corresponding to two discrete wavelet transform attributes (see later for details) calculated from the trace segments in (a).

contact. Walden (1994) discussed the use of trace energy, half-time (the fraction of the length of the trace segment required to build up half the total energy) and centre frequency as attributes to be used in improving the positioning of a pinch-out.

What does the DWT have to offer that methods such as those above do not? An important point to notice about all the above attributes is that they are derived from the

whole trace segment. On the other hand, the DWT consists of coefficients which represent features in different sections of each trace segment, from 'all' to a tiny part. These coefficients can hence be used very flexibly without constantly having to re-define and re-extract different trace segments. We could of course represent a sequence in the frequency domain via the Fourier transform. However, this representation is in terms of infinitely long sines and cosines. Hence local changes to a sequence cause many Fourier coefficients to change. The DWT, by contrast, represents the sequence in terms of finitely long quantities, wavelets at different scales, and hence a local change in the sequence manifests itself as a change to only a very few wavelet coefficients. Hence lateral changes to a seismic reflection sequence should be more easily 'seen' using the DWT coefficients.

The wavelet transform has already found uses in exploration geophysics. One of the earliest papers on the subject was that by Goupillaud, Grossmann and Morlet (1984) and further use of the transform was made by Morlet *et al.* (1982a, b) in a study of the propagation of plane waves. While this paper was under review, Pike (1994) showed how the modulus of the wavelet transform could be used towards classification of scatterers in high resolution marine seismic data; he also used the wavelet transform for attenuation estimation. These studies used the continuous rather than discrete wavelet transform. Interestingly, Pike (1994), who used the Morlet wavelet, comments that 'Examination of alternative wavelets that may be more diagnostic for classification of scattering should be pursued.' We examine wavelet choice here. Finally, Olmo, Lo Presti and Spagnolini (1994) showed how the wavelet transform can be utilized for velocity analysis.

In this application we extract trace segments between consecutive geological events which have been identified on the sections: these segments thus refer to the same geological interval. The basic DWT, like the simplest fast Fourier transform, needs a data length which is a power of 2. Hence, for convenience we take all the segments to consist of the same number of points, some convenient power of 2, which keeps all segments within a geological interval defined by top and bottom reference events; all segments start at the top reference event.

The discrete wavelet transform

The discrete wavelet transform (DWT) decomposes an input data sequence $\mathrm{X} = (x_1, \ldots, x_N)^{\mathrm{T}}$ into so-called wavelet coefficients, which capture the behaviour of the sequence at different scales or frequencies. In our application, think of the 'sequence' as the trace segment amplitudes at a particular location; Fig. 2a shows 32 such amplitudes. One way of considering the transform is in the context of 'multiresolution analysis': looking at or approximating the sequence in some way at different scales. The wavelet coefficients then characterize the differences between these successive approximations. The transform can also be considered as a series of filtering operations for extracting information from contiguous octave bands in the frequency-domain representation of the sequence.

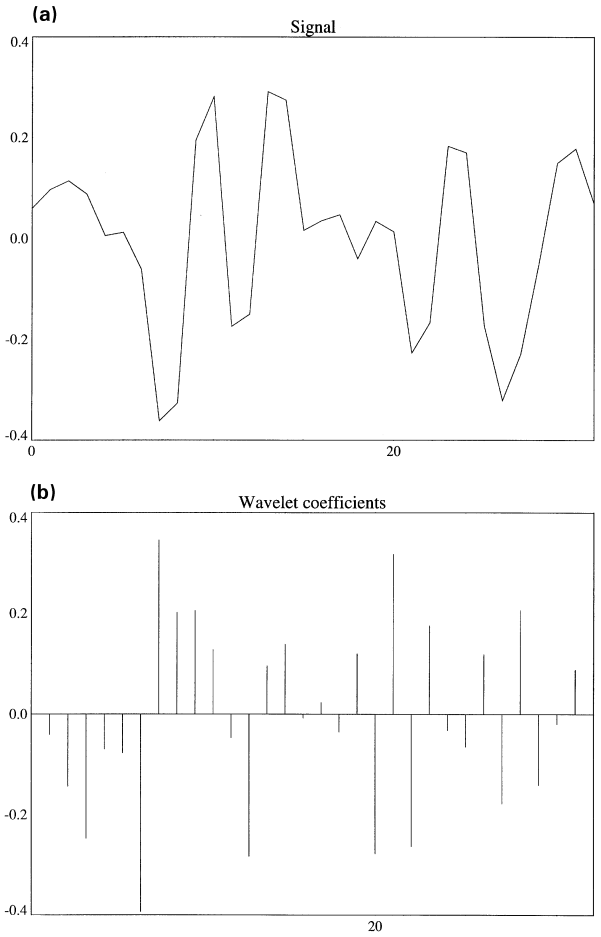

Figure 2. (a) A segment of seismic trace of length $N = 2^5 = 32$ (b) and a naïve display of the resulting DWT coefficients which does not indicate the different scales involved.

The detailed mathematical background to wavelets can be found in Daubechies (1992). An interesting review and comparison with the Fourier transform can be found in Strang (1993). Another application to clustering is that of Bradshaw and Spies (1992).

Here we do not go into mathematical details, but seek merely to explain the structure of the transform and several aspects of its behaviour.

The transform

The DWT can be written as a matrix W operating on an input sequence $\mathrm{X}_p = (x_1, \ldots, x_N)^{\mathrm{T}} \equiv (x_{p,1}, \ldots, x_{p,N})^{\mathrm{T}}$ say, where $N = 2^p$ Note that a subscript p has

been added to the original sequence to emphasize that it is of length 2^p; the original sequence is said to be at scale 1. We denote the transformed sequence by

$$\mathrm{Z} = \mathrm{W}\mathrm{X}_p.$$

The transform used is orthogonal, so that

$$\mathrm{X}_p = \mathrm{W}^{-1}\mathrm{Z} = \mathrm{W}^{\mathrm{T}}\mathrm{Z}.$$

What is the nature of the transform? At each (decreasing) step, starting at $m = p - 1$, the current x-sequence is 'filtered' with a low-pass filter $\{g_l\}$ to produce the next set of scaling coefficients:

$$x_{m,k} = \sum_{l=0}^{L-1} g_l x_{m+1,\, l+2k-1}, \quad k = 1, \ldots, 2^m, \tag{1}$$

where L is the length of the filter. The output is a low-pass or smoothed version of the input, and the number of points output is the number input divided by 2; the x-sequence is said to have been 'decimated', 'down-sampled' or 'down-scaled' by a factor of 2. At each stage the current x-vector component is also 'filtered' with another filter $\{h_l\}$, which this time is a high-pass filter to produce the next set of wavelet coefficients:

$$d_{m,k} = \sum_{l=0}^{L-1} h_l x_{m+1,\, l+2k-1}, \quad k = 1, \ldots, 2^m. \tag{2}$$

There are two important points to note about (1) and (2).

[i] Although $\{g_l\}$ and $\{h_l\}$ are referred to as wavelet filters, the operations in (1) and (2), as commonly used in the literature, are inner products rather than convolutions.

[ii] While, at the first step, $\{d_{p-1,k}\}$ is associated with changes on a scale of one, $\{x_{p-1,k}\}$ is associated with averages on a scale of two: the scales associated with the outputs of the high- and low-pass filters thus differ by a factor of two. The sequence $\{x_{m,k}\}$ is said to be at scale $2^{p-m} = N/2^m$, and $\{d_{m,k}\}$ is said to be at scale $2^{p-m-1} = N/2^{m+1}$.

Eventually, after p steps, we end up with $\mathrm{X}_0 = x_{0,1}$ and all the accumulated detail or wavelet coefficients, so that the transform can be written

$$\mathrm{X}_p = (x_1, \ldots, x_N)^{\mathrm{T}} \rightarrow \left(x_{0,1}, \quad \underbrace{d_{0,1}}_{\text{scale } N/2} \quad \underbrace{d_{1,1}, d_{1,2}}_{\text{scale } N/4}, \ldots, \underbrace{d_{p-1,1}, \ldots, d_{p-1,2^{p-1}}}_{\text{scale1}} \right)^{\mathrm{T}} = \mathrm{Z}, \tag{3}$$

where coefficients of the same scale are bracketed. The nature of the transform is illustrated in Fig. 3.

In order to avoid tedious technicalities we shall ignore filtering 'end effects.' Suffice it to say that we use an implementation with periodic or wrap-around boundary conditions.

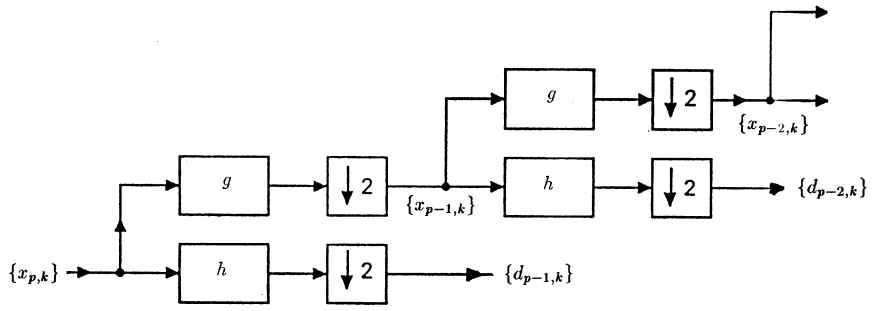

Figure 3. Illustration of the steps involved in the DWT. The down-arrow followed by a 2 indicates decimation or down-sampling by 2.

The filters are the key to the method and the nomenclature. The filters satisfy two-scale difference equations,

$$\phi(t) = 2^{1/2} \sum_l g_l \phi(2t - l),$$

and

$$\psi(t) = 2^{1/2} \sum_l h_l \phi(2t - l),$$

The function $\phi(\cdot)$ is called the scaling function of $\psi(\cdot)$ is the wavelet function. The low-pass filter $\{g_l\}$ is directly related only to the scaling function, so that coefficients derived using the low-pass filter, such as in (1), are called scaling coefficients. The high-pass filter $\{h_l\}$ is related to the wavelet function (and the scaling function), so that coefficients derived using the high-pass filter, such as in (2), are called wavelet coefficients.

Display of the wavelet coefficients

The naïve way to display the results of applying the DWT to a sequence, namely the elements of Z in (3), is shown in Fig. 2b for a particular segment of length $N = 2^5 = 32$ extracted from a seismic trace (Fig. 2a). However, from (3) we can see that each wavelet coefficient $d_{m,k}$ is associated with a particular scale and thus they are 'defined' over different ranges. A suitable way of displaying the coefficients to highlight this is the so-called 'scalogram', an image array where the x-axis references time (location) and the y-axis references scale. In Fig. 4, each row corresponds to a particular scale, with the bottom row being the original sequence. Note that here $p = 5$ so that the finest scale is 1 ($m = 4$) and the coarsest is 8 ($m = 1$); we do not display the value $d_{0,1}$ since it is a constant across the range of the data.

The bottom plot is of the original data $\mathrm{X}_5 = (x_{5,1}, \ldots, x_{5,32})^{\mathrm{T}}$, the next to bottom is of $\mathrm{D}_4 = (d_{4,1}, \ldots, d_{4,16})^{\mathrm{T}}$, and the top is of $\mathrm{D}_1 = (d_{1,1}, d_{1,2})^{\mathrm{T}}$. To represent the ranges of the coefficients we plot them as horizontal bars rather than on a grey scale, one benefit of

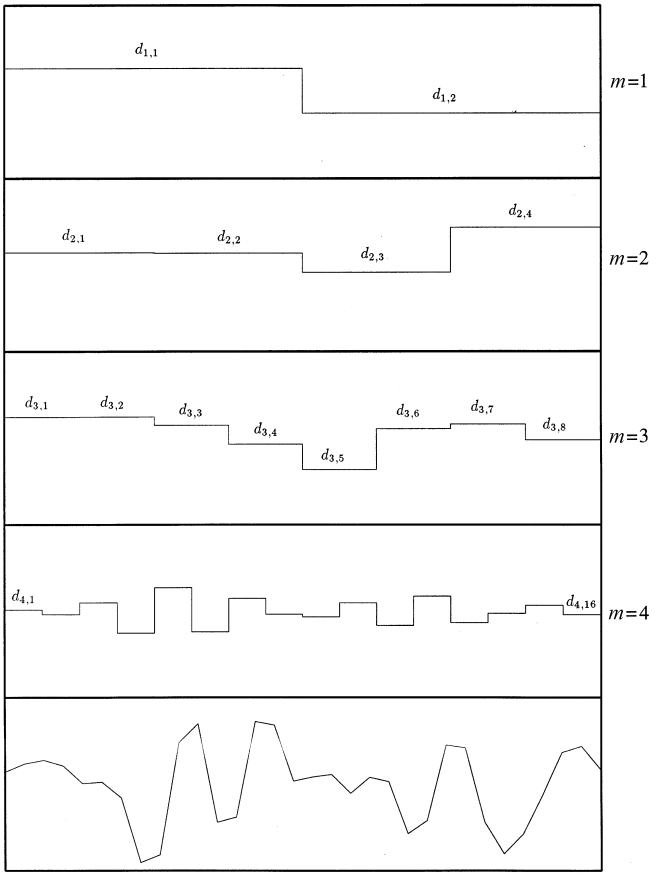

Figure 4. A scalogram. The bottom row shows the same segment of seismic trace as in Fig. 2. Each row shows the detail coefficients, for scale 1 ($m = 4$) up to scale 8 ($m = 1$).

such a display being that coefficients for different series can be plotted on top of one another (see e.g. Fig. 7).

Choice of the wavelet

The elements of the original sequence are related to the wavelet coefficients through

$$x_n = \sum_{m=0}^{p-1} \sum_{k=1}^{2^m} d_{m,k} \psi_{m,n-2^{p-m}k} + x_{0,1} \phi_{0,n-N}. \tag{4}$$

Here the sequences $\{\psi_{m,n}\}$ are called wavelet sequences for scale 2^{p-m-1}. (These sequences are defined recursively using the same low- and high-pass filters, $\{g_l\}$ and $\{h_l\}$. Note that the sequence $\{\psi_{m,n-2^{p-m}k}\}$ is just the same as the sequence $\{\psi_{m,n}\}$ after a shift of $2^{p-m}k$, and hence has the same 'shape.' From (4) we see that a wavelet

coefficient $d_{m,k}$ will be large if the sequence $\{x_1, \ldots, x_N\}$ contains a shape of the form $\{\psi_{m,n-2^{p-m}k}\}$. The sequences $\{\psi_{m,n-2^{p-m}k}\}$, $m = 0, \ldots, p-1$; $k = 1, \ldots 2^m$ play the role of the sines and cosines in Fourier analysis. Together with the scaling sequence $\{\phi_{0,n-N}\}$ they are called basis functions. Under an assumed regularity condition, the wavelet sequences when plotted on an appropriate scale converge, with decreasing m, to the continuous wavelet $\psi(t)$. Hence, roughly speaking, the wavelet coefficient at a particular scale and location will be large whenever the data sequence contains a shape similar to a scaled version of the wavelet $\psi(t)$ at that location. This raises the question of which sort of wavelet filter coefficients to choose in our application.

Clearly, the choice will have an effect on the coefficients of the transform. We have been investigating the use of Daubechies' compactly supported wavelets (Daubechies 1992). These are straightforward to implement, and, due to their compact support, behave well. For example, even a small number of filter coefficients (four or more) guarantees good octave-band concentration at each scale, and as more coefficients are used, the wavelet becomes smoother in a predictable way. These form a family of wavelets with increasing regularities which can locally approximate polynomial functions of increasing degrees. They are defined and indexed by the number of filter coefficients used in their construction, which is also related to the local continuity of functions they can represent, i.e. the wavelet constructed from four coefficients can represent piecewise linear functions; increasing the number of coefficients by 2 adds one degree to the local polynomials they can represent. Furthermore, a Daubechies wavelet corresponding to a filter with L coefficients (L even) has $L/2$ vanishing moments; such vanishing moments ensure a fast decay of the wavelet coefficients, which is desirable for a sparse representation of the seismic trace segment. Figure 5 shows several continuous wavelets $\psi(t)$ from this family (using long sequences in order to show sufficient detail in the plot). Recall that these are the shapes to which the wavelet sequences converge.

While $\{g_l\}$ is a low-pass filter, and $\{h_l\}$ a high-pass filter, so that the amplitude responses of these filters is clear, there is obvious flexibility in the way the phase properties of these filters are chosen. Daubechies (1992) examined both *minimum-phase* and close to *linear-phase* choices. The first eight plots of Fig. 5 are wavelets generated using different numbers of minimum phase filter coefficients; an example filter is given in Fig. 6a. For convenience we shall call these the MP family.

The last seven functions shown in Fig. 5 are wavelets generated from different numbers of close to linear phase filter coefficients; an example filter is given in Fig. 6b. The latter are often called the 'least asymmetric' (LA) family. For the same number of filter coefficients, these wavelets are not as smooth as the MP class, but have better phase properties, i.e. can be made close to zero phase by a time shift.

Different choices of wavelet shape might be better tuned to particular data or to capturing particular features of data. A large wavelet coefficient will occur where the wavelet at that scale matches a similarly shaped waveform present in the trace. Hence, the asymmetric wavelets might not be a good choice if the seismic traces have been zero-phased; the least asymmetric wavelet would seem a more natural choice. Figure 7

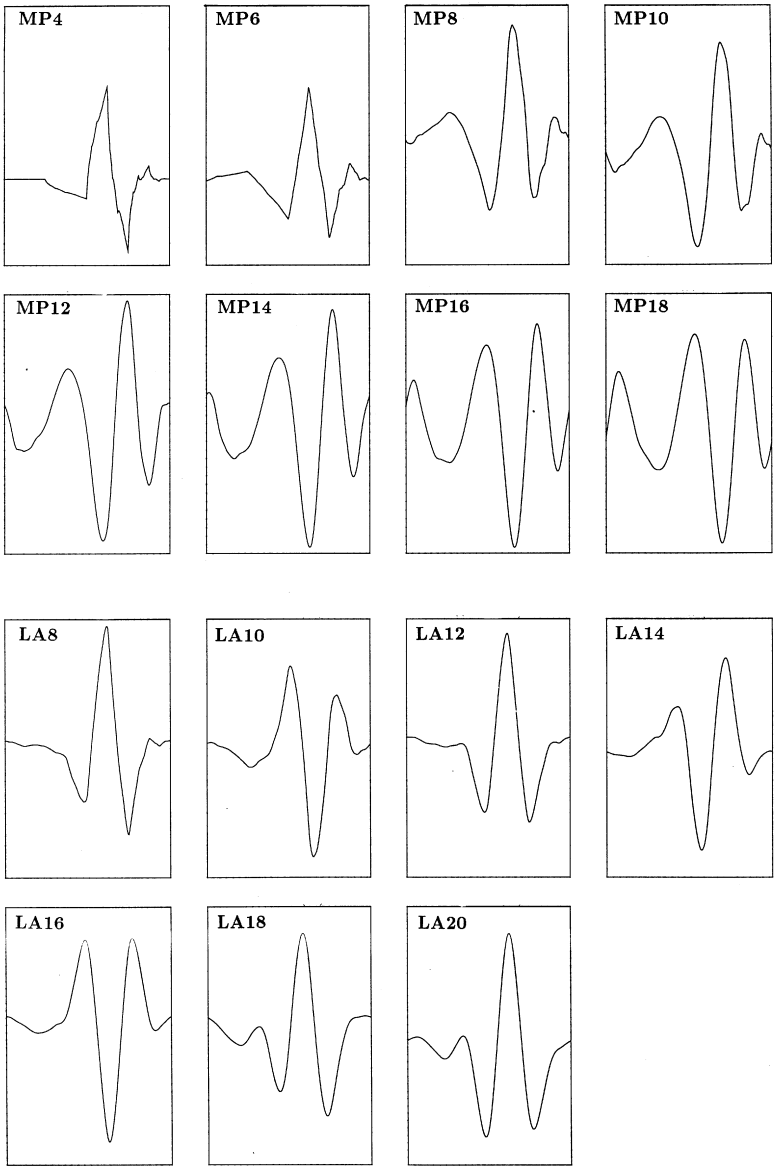

Figure 5. A selection of continuous wavelets $\psi(t)$. The first eight plots (top to bottom, left to right) are MP wavelets; the last seven are LA wavelets. Number of coefficients used in generating are as specified.

shows the transform of the same data sequence for different wavelet filter coefficients; first column: MP 4 and 6 filter coefficients, second column: MP 4 and 8 filter coefficients, third column: MP 4 and LA 16 filter coefficients. In each column the result using MP 4 filter coefficients is shown by the solid line.

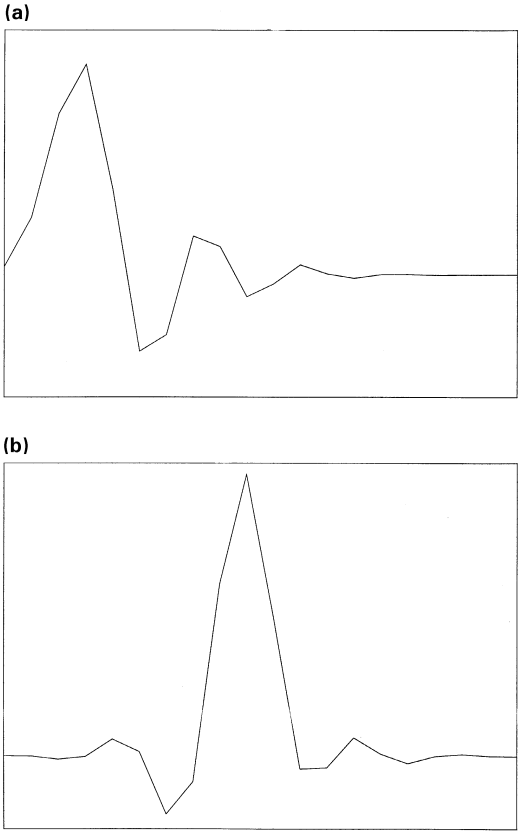

Figure 6. Two examples of generating filters: (a) minimum-phase (MP), and (b) close to linear phase or 'least asymmetric' (LA).

We might like to choose the wavelet filter coefficients based upon the accuracy with which we can reconstruct the data sequence from a fixed number of coefficients, or on how accurately certain features are reconstructed from fewer coefficients. Figure 8 shows the error sum of squares in reconstruction (the seismic traces have been normalized to have a sum of squares of unity) when using 16 (Fig. 8a) and 8 (Fig. 8b) of 32 coefficients for reconstruction for different wavelets. Reconstruction of 1000 traces was carried out.

The error over the 1000 cases is represented by box-plots: the middle horizontal bar marks the median error, the top and bottom of each black 'box' marks the top and bottom of the interquartile range of the error, while the top and bottom bars mark 1.5 times the interquartile range. The odd horizontal bars show a few outlying and atypical results. To distinguish different choices of wavelet filter coefficients we label each with 'LA' or 'MP' together with the number of coefficients *i.* (Note that MP2 corresponds to the special Haar wavelet transform.) This plot suggests that MP12 and MP16 provide the best reconstructions for this data set, although not by very much.

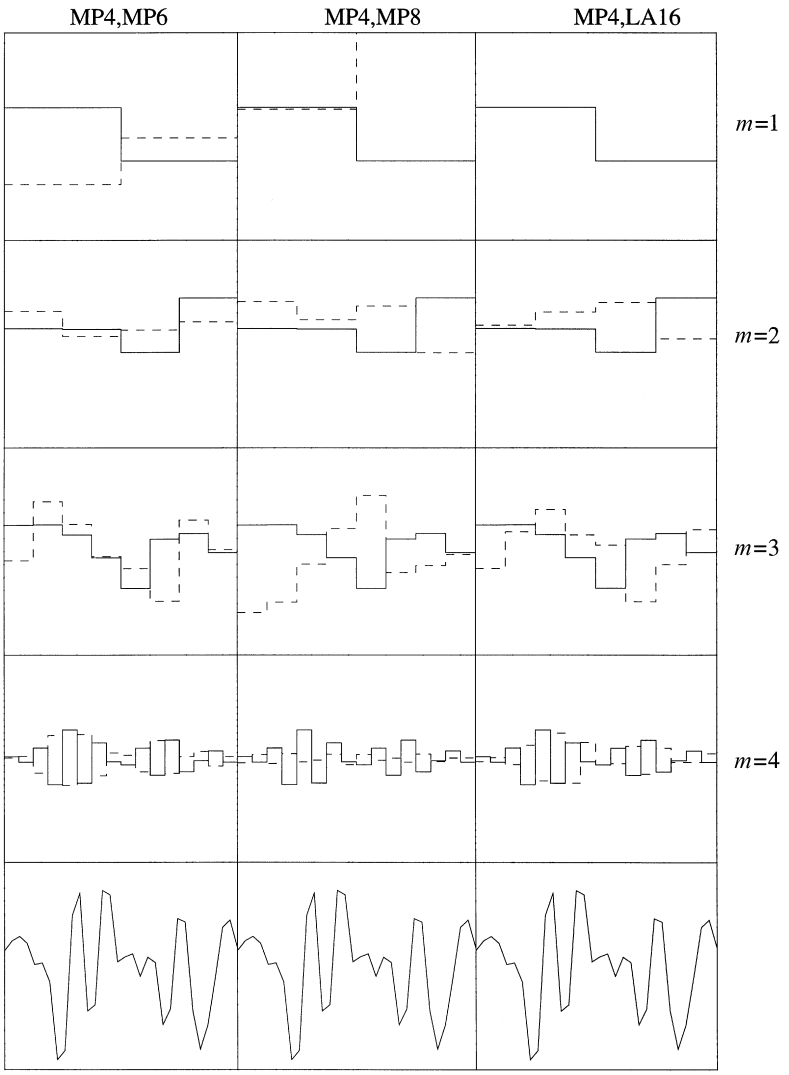

Figure 7. Scalogram, shown by dashed lines, of same data sequence (bottom) for different wavelets, using, from left to right, MP 4 and 6 filter coefficients, MP 4 and 8 filter coefficients, MP 4 and LA 16 filter coefficients. In each column the result using MP 4 filter coefficients is shown by the solid line. The dashed line for the top-right plot is off-the-scale.

Choice of the wavelet is not completely determined; it may be useful to try several different wavelets. A benefit could be gained from constructing a suitable wavelet to match the processing/application, e.g. LA for zero-phased data and MP for minimum-phase data. This is a potentially valuable research area (see e.g. Pike 1994).

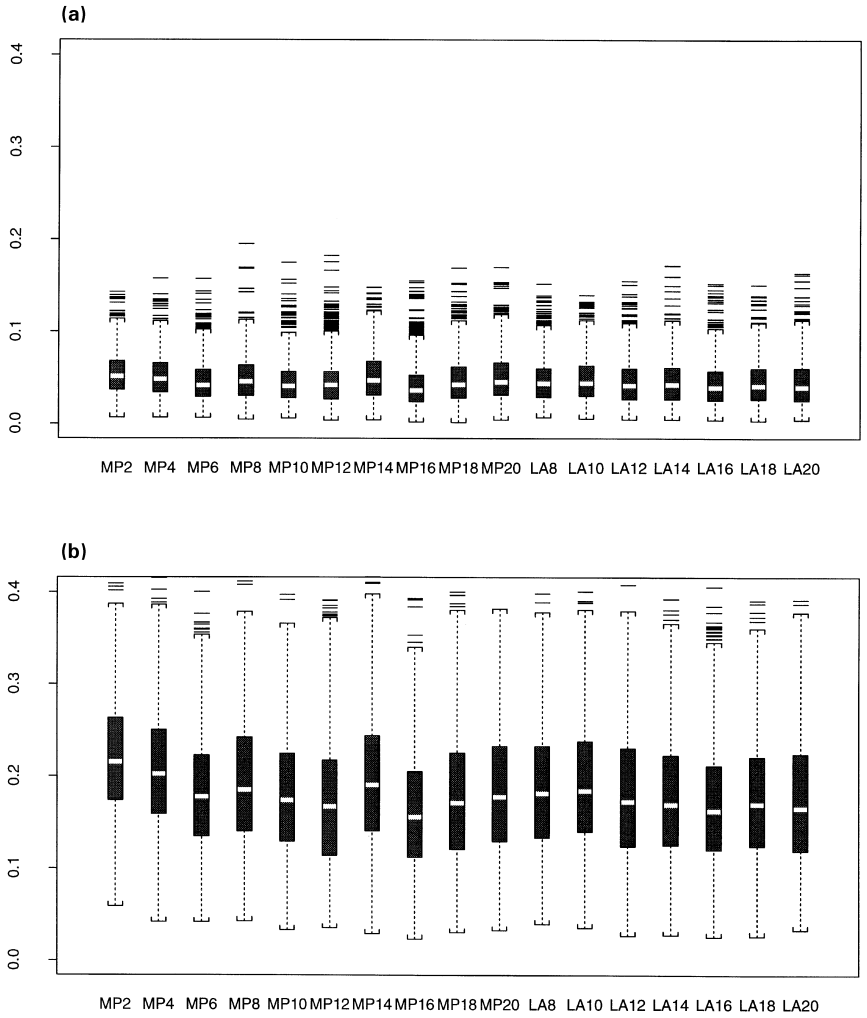

Figure 8. Figure 8. The error sum of squares in reconstruction of seismic traces using the DWT when using (a) 16 and (b) 8 of 32 coefficients for reconstruction for different choices of wavelet filter coefficients. Reconstruction of 1000 traces was carried out. The error over the 1000 cases is represented by box plots (see text for details on interpretation). To distinguish different choices of wavelet filter coefficients, each case is marked by 'MP' or 'LA' together with the number i of coefficients used in the transform. (Note MP 2 corresponds to the special Haar wavelet transform.)

Wavelet properties

We consider the effect on the transform of a linear transformation of the input sequence. If we rescale the input signal so that $X'_p = aX_p$, then $WX'_p = aWX_p = a\,Z$ and all coefficients simply get rescaled. Thus, once we have extracted the original energy of

the data sequence, $\Sigma_{i=1}^{N} x_i^2$, we work with normalized sequences. Note that since the transformation is orthogonal,

$$(\mathrm{W}\mathrm{X}_p)^{\mathrm{T}}\mathrm{W}\mathrm{X}_p = \mathrm{X}_p^{\mathrm{T}}\mathrm{W}^{\mathrm{T}}\mathrm{W}\mathrm{X}_p = \mathrm{X}_p^{\mathrm{T}}\mathrm{W}^{-1}\mathrm{W}\mathrm{X}_p = \mathrm{X}_p^{\mathrm{T}}\mathrm{X}_p = \sum_{i=1}^{N} x_i^2,$$

so that energy is preserved.

If $\mathrm{X}'_p = \mathrm{X}_p + \mathrm{e}$, where $\mathrm{e} = (e_{1,\ldots,}e_N)^{\mathrm{T}}$
then $\mathrm{W}\mathrm{X}'_p = \mathrm{W}(\mathrm{X}_{\mathrm{p}}+\mathrm{e}) = \mathrm{W}\mathrm{X}_{\mathrm{p}} + \mathrm{W}\mathrm{e} = \mathrm{Z} + \mathrm{W}\mathrm{e}$.

To see the properties of such changes we can look at the transform of the change, We. If e is a constant vector, i.e. we shift the signal up by a certain amount, then the only non-zero wavelet coefficient of We is the overall smooth value $\mathrm{e}_{0,1}$ (the transform is such that the row sums $\Sigma_{j=1}^{N} W_{i,j}$ are zero for $i \neq 1$). This should not be an issue for seismic signals which are usually zero mean.

To demonstrate the time-frequency localization we display the transform of a single spike ($\mathrm{e} = (0, \ldots, 0, 1, 0, \ldots 0)^{\mathrm{T}}$) at sample number 77 (dashes) in a sequence of length $256 = 2^8$ Fig. 9. A MP 4 wavelet was used. The spike produces an effect on the coefficients at each scale which is confined to a location close to the original spike, becoming less localized at larger scales. The transform for a spike at sample number 78 (solid line) is also plotted. The difference between these transforms highlights a potential weakness of the DWT, which is a lack of translation invariance: the coefficients are dependent upon the location of the signal. We could capture and remove the effect of alignment in another 'attribute' or variable, using for example, maximum semblance, thus working with aligned signals before calculating their wavelet transform.

In the case of attributes derived from the seismic data set, translation can be thought of as extracting the trace beginning at a different place. For random shifts, as could occur from mispicking, some 'bleeding' of the identified cluster across boundary lines takes place and the change in structure is not usually too great. However, the magnitude of this effect will depend upon the scale of the attributes chosen; if we consider coefficients at higher levels in the DWT, these will be less sensitive to shifts of the data. The lack of translation invariance, demonstrated by Fig. 9, is an active research area (see e.g. Simoncelli *et al.* 1992).

Sequence length

Another issue in this context is that of data sequence length, since we may not always be able to extract lengths which are convenient powers of two. How we deal with this depends upon the nature of the data: if the interval defining the segment is of varying thickness due to, say, erosion, we might expect the signal to be effectively truncated. In this case we could pad it with zeros to make the length a suitable power of 2. Alternatively, the waveform may be compressed, or stretched, between two events, in which case it might be better to capture this in some form of frequency

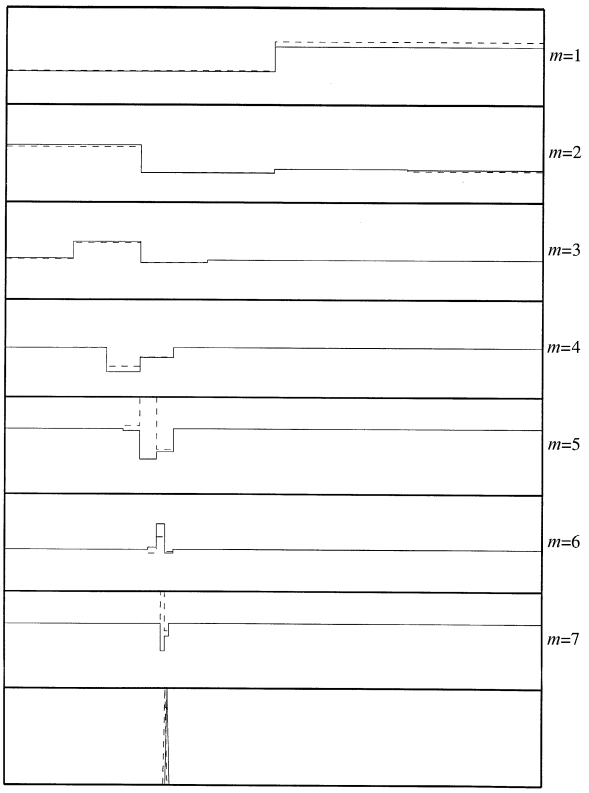

Figure 9. DWT of a spike, displayed as a scalogram with solid lines, and the DWT of another spike at the next sample point, displayed as a scalogram with dashed lines.

variable and resample the trace to a suitable length. In the case when the trace segment is truncated from 32 to 30 samples, Figure 10 shows the effect on the wavelet transform of padding with two zeros (first column) and resampling (second column) using a sinc function. We see that padding with zeros has an effect on the coefficients, particularly at higher scales, while resampling is clearly wrong and has a large effect on all of the coefficients. When the trace is compressed, the third column of Fig. 10 shows the effect of padding with zeros, while the fourth column shows the effect of resampling, which is virtually indistinguishable from the original data when using a sinc function. In each case the solid line gives the wavelet coefficients for the original data sequence. It is clearly important to carry out the most suitable manipulation; geological knowledge will be important here.

Selection of attributes from the discrete wavelet transform

The wavelet transform of a sequence contains the same number of coefficients as the original sequence, all of which are necessary to reconstruct the sequence exactly. There

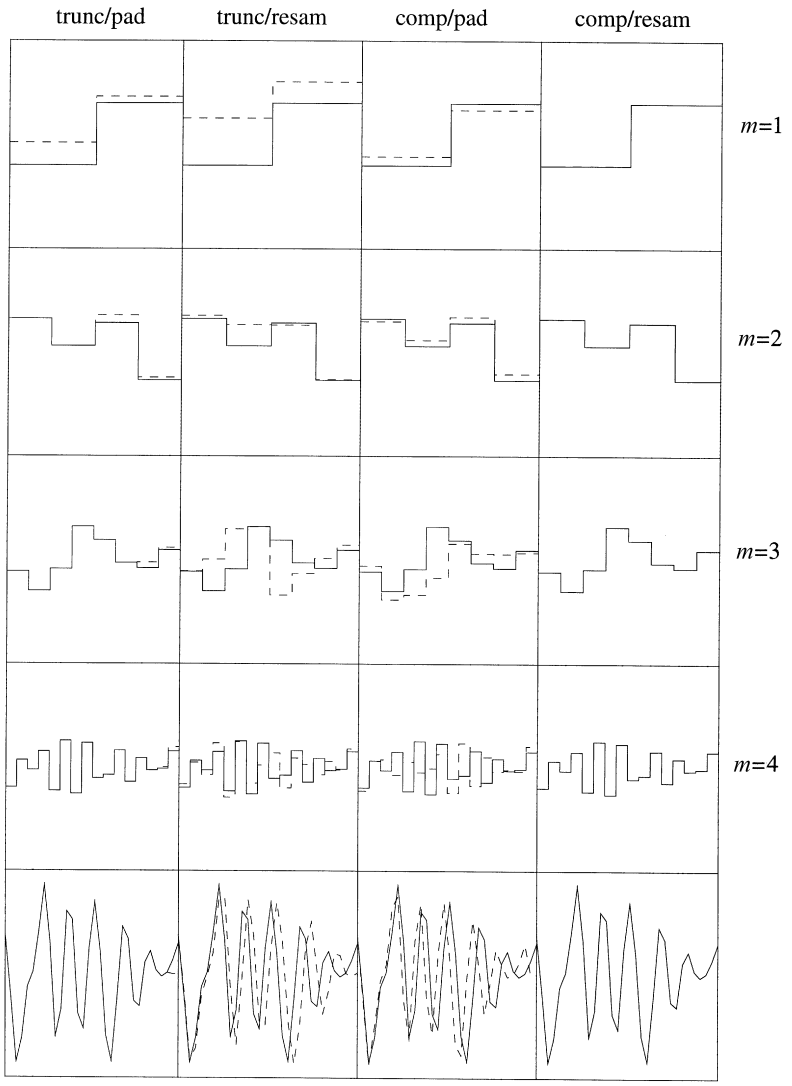

Figure 10. DWT of a seismic trace segment, displayed as a scalogram with solid lines, and of the same after some data manipulation. From left to right: after truncation from 32 to 30 points and padding with zeros to 32 points; after truncation from 32 to 30 points and resampling to 32 points; after compression from 32 to 30 points, and padding with zeros to 32 points, and finally after compression from 32 to 30 points and resampling to 32 points.

is however a potential redundancy in parametrization, so that we may hope to reconstruct the signal reasonably accurately from only a few coefficients of the transform. This is equivalent to hoping that a few coefficients summarize the main features of the data. These can then be used for attribute analysis/clustering.

Energy at different scales

A useful summary of $d_{m,k}$, $k = 1, \ldots, 2^m$ is the energy. This provides a decomposition of the total energy of a signal into that present at different scales, i.e.

$$\sum_{n=1}^{N} x_n^2 = x_{0,1}^2 + \sum_{m=0}^{p-1} \sum_{k=1}^{2^m} d_{m,k}^2 = x_{0,1}^2 + \sum_{m=0}^{p-1} E_m,$$

where E_m is the energy at scale 2^{p-m-1}. The 'variance' at each scale, i.e. the energy normalized by the number of coefficients at that scale, is used by Bradshaw and Spies (1992), and will have the same properties as the energy. For the sequence shown in Fig. 4, we display the decomposition of energy for different choices of wavelet filter coefficients shown in Table 1. Each column contains the energy at increasing scales for a particular choice, indexed by the number of coefficients; the lower part of the table refers to the least asymmetric family, which is not defined for less than eight coefficients. This demonstrates the dependence of energy decomposition on the wavelet used. However, for a fixed wavelet, different sequences will give different energy-by-scale decompositions, which can be used as attributes.

The energy decomposition for the different locations of the spike of Fig. 9 (77 and 78 in a signal of length 256) given in Table 2, shows quite a large change at fine scales (6 and 7), although the change is mainly confined to these scales, as is also evidenced in Fig. 9.

For the case of the series of length 32 (Fig. 10) to which various data manipulations were applied (truncation and padding with zeros or resampling; compression and padding with zeros or resampling) the energy decomposition is as shown in Table 3. By comparing this with the energy decomposition of the original trace, also shown in Table 3, we conclude that, for this example, truncating/padding and compression/resampling give the most accurate energy decompositions (the latter gives extremely good reconstruction). Another recent example of the use of zero padding is given by Wang (1995). Also, resampling is a natural way to 'undo' compression of a signal. Clearly the choice of data manipulation is important in determining an appropriate energy decomposition.

Using selected coefficients at fine or coarse scales

It is possible that a particular scale or certain locations may be important in characterizing lateral changes over segments of seismic traces. Let us look again at Fig. 4 for which $N = 32 = 2^5$. Suppose we choose as attributes the coefficients $(d_{2,1}, d_{2,2})$. These detail coefficients cover the first *half* of the trace sequence at scale $2^{p-m-1} = 2^{5-3} = 4$. Suppose we choose instead $(d_{3,1}, d_{3,2})$; these detail coefficients cover the first *quarter* of the trace sequence at scale *2*. Hence by choosing appropriate coefficients we can look at different sub-intervals of the trace sequence.

Figure 1 uses $d_{3,1}$ and $d_{3,2}$. To see the effect which these appear to be capturing, we consider a single line from Fig. 1 (Line 1) and examine the stacked seismic section

Table 1. Energies $E_m = \Sigma_{k=1}^{2^m} d_{m,k}^2$ for the data sequence shown in Fig. 4. Nine different choices of wavelet filter coefficients of MP type (top) and seven of LA type (bottom) were investigated.

				MP					
m	4	6	8	10	12	14	16	18	20
0	0.004	0.000	0.003	0.006	0.003	0.000	0.007	0.008	0.001
1	0.082	0.127	0.022	0.093	0.170	0.014	0.107	0.151	0.013
2	0.285	0.067	0.308	0.156	0.093	0.282	0.122	0.157	0.218
3	0.211	0.423	0.642	0.313	0.474	0.623	0.317	0.537	0.593
4	0.418	0.383	0.025	0.433	0.259	0.080	0.448	0.148	0.176

				LA					
m	4	6	8	10	12	14	16	18	20
0			0.004	0.002	0.003	0.009	0.002	0.009	0.002
1			0.119	0.097	0.143	0.002	0.159	0.004	0.163
2			0.230	0.126	0.192	0.216	0.149	0.210	0.152
3			0.262	0.750	0.297	0.416	0.350	0.735	0.344
4			0.384	0.025	0.364	0.357	0.340	0.042	0.339

Table 2. Energies $E_m = \sum_{k=1}^{2^m} d_{m,k}^2$ for the two spikes shown in Fig. 9 at locations 77 and 78. MP 4 wavelet filter coefficients were used.

					m			
locn	0	1	2	3	4	5	6	7
77	0.007	0.001	0.014	0.074	0.046	0.028	0.547	0.284
78	0.007	0.001	0.012	0.064	0.030	0.028	0.141	0.717

Table 3. Energies $E_m = \Sigma_{k=1}^{2^m} d_{m,k}^2$ for the sequence shown in Fig. 10. The rows of the table are, respectively, for the original sequence, for the sequence truncated by two with two zeros added, for the sequence truncated by two and resampled to the original length, for the sequence compressed by two points with two zeros added, and finally for the sequence compressed by two points and resampled to the original length.

			m		
data manipulation	0	1	2	3	4
original	0.001	0.017	0.316	0.379	0.287
trunc./pad	0.008	0.007	0.299	0.385	0.302
trunc./resam.	0.004	0.004	0.253	0.486	0.253
comp./pad	0.003	0.012	0.261	0.502	0.222
comp./resam.	0.001	0.017	0.317	0.378	0.287

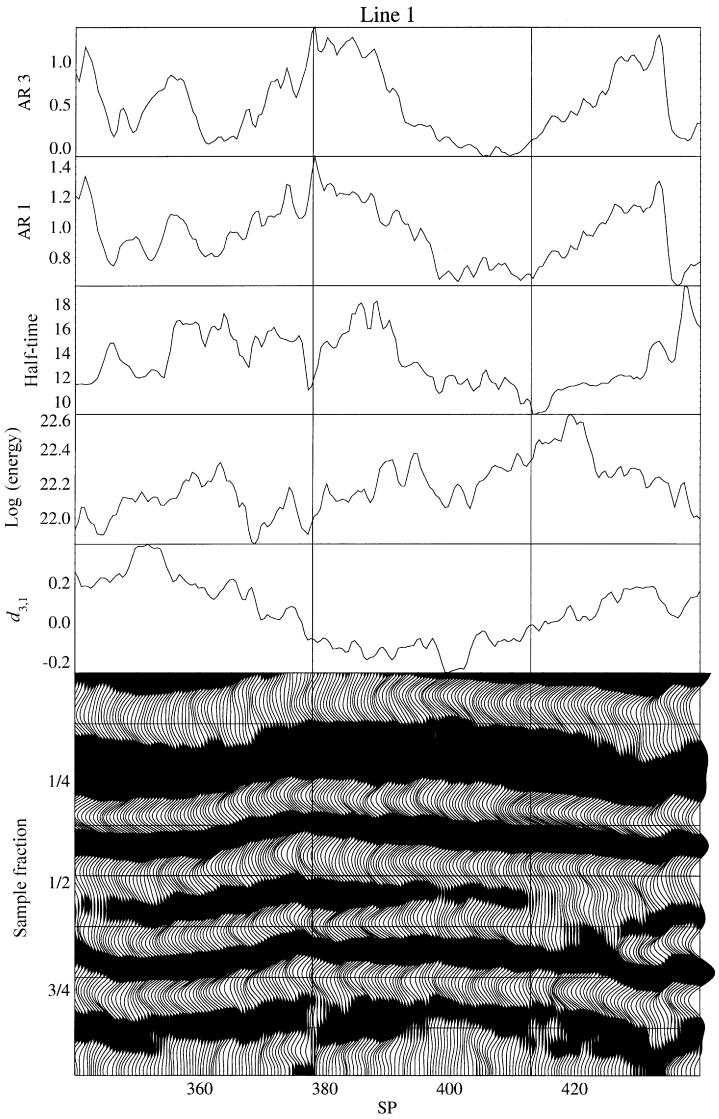

Figure 11. Traces from Line 1 in Fig. 1, together with various attributes plotted above. The region between shotpoints 380 and 410 corresponds to the cluster of crosses on this line.

(Fig. 11) for this line. Notice that instead of 'time' on the *y*-axis of the section, we have split our 32 points into fractions of 32, so that e.g. 1/2 corresponds to the 16th point in the segment of 32 values; this gives a more intuitive correspondence with the idea of the wavelet coefficients covering different fractions of the sequence, as discussed above. Plotted in Fig. 11 are some conventional attributes (see the Introduction), along with

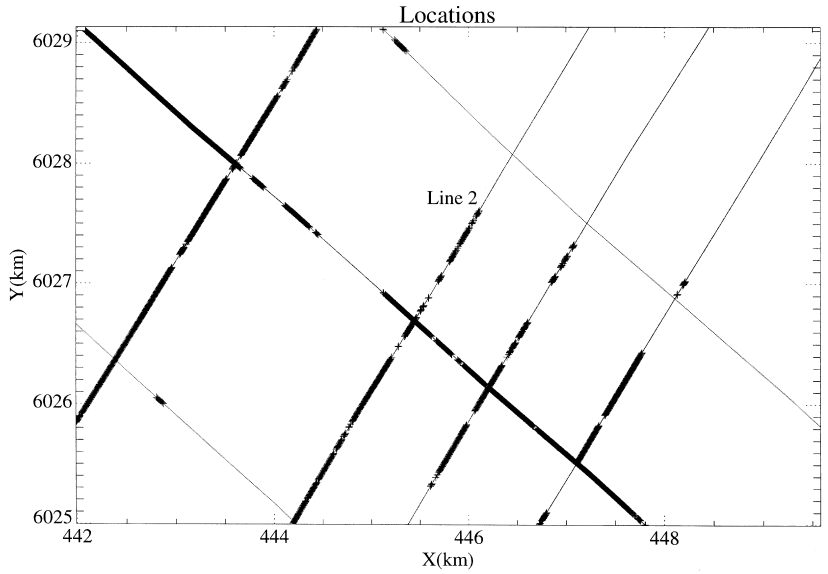

Figure 12. Clustering of survey lines according to high values of first wavelet detail coefficient.

$d_{3,1}$ which covers only the first eighth of each trace ($d_{3,2}$ of course, covers the second eighth of each trace). The cluster of crosses shown in Line 1 of Fig. 1 corresponds to shotpoints 380–410 (the middle panel in Fig. 11). Looking at the section in Fig. 11, we see that the strong event at the top is hardly present in the interval between 380–410, while another strong event moves into the top quarter of the segment and then out again back into the second quarter as we go from left to right. How does this behaviour manifest itself in terms of our chosen attributes? From Fig. 1 we see that in terms of the coefficients, $d_{3,1}$ is relatively low in 380–410 (crosses), while $d_{3,2}$ is relatively high.

None of the conventional attributes appear to capture this localized behaviour in the section so clearly. They all work with the entire trace segments. The autoregressive coefficients (AR 1 and AR 3) from an order-5 autoregressive model are seen to be highly correlated, and energy and half-time are features describing the entire traces. Other attributes can certainly be useful for trace characterization. However those derived from the DWT give us the flexibility to look locally and at different scales.

Figure 12 now shows the same region as Fig. 1, clustered according to high values (crosses) of the first wavelet coefficient $d_{0,1}$. The major break between crosses and dots on Line 2 of Fig. 12 occurs near shotpoint 330 marked in Fig. 13. Crosses are at lower shotpoints, and dots at higher ones. Shotpoints 300–303 in Fig. 13 also correspond to dots on Line 2 in Fig. 12. The coefficient $d_{0,1}$ is a constant across the *whole* segment, and in the sense that it 'represents' the whole segment, it is more like a conventional attribute. The behaviours of $d_{0,1}$ and of the conventional attribute half-time are remarkably similar in many parts of Fig. 13. We should not be surprised. Half-time is the

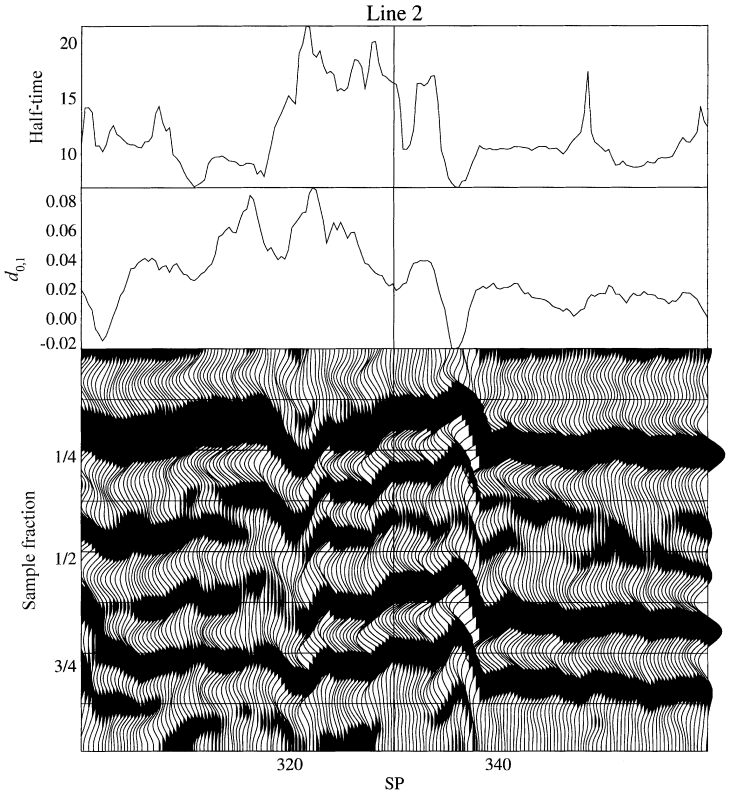

Figure 13. Traces from Line 2 in Fig. 12, together with the two attributes, half-time and $d_{0,1}$.

proportion of the trace segment length required to build-up half the total energy in the trace segment: it is a low-frequency measure. Similarly, $d_{0,1}$ represents a very low-frequency octave band, and is hence also a low-frequency measure.

The results shown in Figs 11 and 13 illustrate the flexibility of the DWT coefficients as attributes: they can be used to give local information on trace differences, or can be used in a similar way to conventional attributes as a measure of a whole trace segment.

Conclusions

The discrete wavelet transform can parsimoniously represent localized changes in sequences such as segments of seismic traces. Certain coefficients or functions of coefficients from the transform might thus capture the effect of changes and can be used to cluster or discriminate among traces. We have demonstrated the application of the transform to derive attributes from small segments of seismic traces, which can be used for trace characterization. Further work is required to understand fully the features which the transform captures in various combinations of the coefficients.

We have shown that the choice of the type of wavelet filter coefficients needs consideration, as do uncertainties in picking events and treatment of truncations, resampling, etc. However, as the algorithm is refined and improved (as was the fast Fourier transform), the discrete wavelet transform should prove increasingly useful.

Acknowledgements

This work was funded by the OSO/NERC Hydrocarbon Reservoirs Link Programme, with additional support from British Gas Exploration and Production, EECal, Enterprise Oil plc, Geco Prakla, OMV (UK) Ltd and Scott Pickford plc.

We are grateful to the editor and three referees for helpful comments which improved the presentation and content of the paper.

References

Bois P. 1980. Autoregressive pattern recognition applied to the delimitation of oil and gas reservoirs. *Geophysical Prospecting* **28**, 572–591.

Bois P. 1982. Some comments on the application of pattern recognition to oil and gas exploration. *Geoexploration* **20**, 147–159.

Bradshaw G.A. and Spies T.A. 1992. Characterizing canopy gap structures in forests using wavelet analysis. *Journal of Ecology* **80**, 205–215.

Cleveland W.S. and McGill M.E. 1988. *Dynamic Graphics for Statistics*. Wadsworth.

Daubechies I. 1992. Ten lectures on wavelets. CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM.

Dumay J. and Fournier F. 1988. Multivariate statistical analyses applied to seismic facies recognition. *Geophysics* **53**, 1151–1159.

Friedman J.H. 1987. Exploratory projection pursuit. *Journal of the American Statistical Association* **82**, 249–266.

Goupillaud P., Grossmann A. and Morlet J. 1984. Cycle-octave and related transforms in seismic signal analysis. *Geoexploration* **23**, 85–102.

Hagen D.C. 1982. The application of principal components analysis to seismic data sets. *Geoexploration* **20**, 93–111.

Jones M.C. and Sibson R. 1987. What is projection pursuit?. *Journal of the Royal Statistical Society* **A150**, 1–36.

Lendzionowski V., Walden A.T. and White R.E. 1990. Seismic character mapping over reservoir intervals. *Geophysical Prospecting* **38**, 951–969.

Morlet J., Arens G., Fourgeau E. and Giard D. 1982a. Wave propagation and sampling theory—Part I: complex signal and scattering in multilayered media. *Geophysics* **47**, 203–221.

Morlet J., Arens G., Fourgeau E. and Giard D. 1982b. Wave propagation and sampling theory—Part II: sampling theory and complex waves. *Geophysics* **47**, 222–236.

Olmo G., Lo Presti L. and Spagnolini U. 1994. Wavelet transform in geophysical signal processing—an application to velocity analysis. 56th EAEG meeting, Vienna, Extended Abstract HO 14.

Pike C.J. 1994. Analysis of high resolution marine seismic data using the wavelet transform. In: *Wavelets in Geophysics* (eds E. Foufoula-Georgiou and P. Kumar), pp. 183-211. Academic Press, Inc.

Simoncelli E.P., Freeman W.T., Adelson E.H. and Heeger D.J. 1992. Shiftable multiscale transforms. *IEEE Transactions on Information Theory* **38**, 587–607.

Strang G. 1993. Wavelet transforms versus Fourier transforms. *Bulletin of the American Mathematical Society* **28**, 288–305.

Walden A.T. 1992. Clustering of attributes by projection pursuit for reservoir characterization. In: *Automated Pattern Analysis in Petroleum Exploration* (eds I. Palaz and S.K. Sengupta), pp. 173-199. Springer-Verlag, Inc.

Walden A.T. 1994. Spatial clustering: using simple summaries of seismic data to find the edge of an oil-field. *Applied Statistics* **43**, 385–398.

Wang Y. 1995. Jump and sharp cusp detection by wavelets. *Biometrika* **82**, 385–397.