frequency components. The Brindabella result is substantially larger than the Livingstone result indicating that the separation between the coarse and fine scale structure is much greater in the Brindabella site than at Livingstone. This is consistent with the generally steep smooth slopes observed at the Brindabella site, with fine-scale variations being relatively minor compared to the slopes. The negative Rinker result indicates that the high frequency components are sufficiently powerful enough to induce slope reversals, but its small magnitude shows that the slope reversals are relatively minor. This is probably a consequence of the glacial character of the Rinker DEM, with poorly organised drainage and larger variation in the size and shape of first-order catchments.

The above analysis is based on some unrealistic assumptions concerning the perfect coherence of phases in the two halves of the spectrum, but it supports the earlier suggestion that there is a connection between the wavelength at which the spectrum becomes steeper and the length scale at which the landscape is divided into hillslopes. If this connection is indeed valid then the average hillslope length should be half the critical wavelength \( \lambda_c = 1/f_c \).

Using this interpretation of the inflection in the spectrum, the spectra of the three sites can be used to indicate the grid resolution required to capture first order catchments. If the spectrum is accurate for wavelengths longer than the critical wavelength, then the DEM has sufficient resolution to capture first order catchments, although it may not capture the details within those catchments. At the Livingstone site, the 20 m resolution DEM accurately captures the spectrum down to the critical 250 m wavelength. The 40 m DEM captures most, but not all, of the structure down to 250 m, while the 80 m DEM clearly misses much of this structure. Twenty metres is therefore a good choice of resolution for this site for modelling processes at the hillslope scale; forty metres would still be useful but less accurate, but eighty metres would not be appropriate. These results parallel the findings of Gessler (1996) for soil-landscape associations. Similarly the 1:25 000 scale source data appears to be just fine enough to capture the shapes of first order catchments and coarser data would not be appropriate. The same conclusions can be made for the Brindabella DEM, but for Rinker Lake the analysis shows that the size of the first order catchments (if the concept is meaningful in that landscape) is not well defined so the same conclusions do not apply.

3.6 Implications for Scaling

The power spectral density can be used directly to predict the effect of scale on slope. The variance of slope of a surface is given by the second moment of the power
Figure 3.27: Square root of slope variance as a function of highest frequency included for the three study sites.

The spectrum (Brown and Scholz, 1985):

\[
\text{var(slope)} = \int P(f) f^2 df
\]  

(3.19)

When \( P(f) \) follows a power law this integral converges as long as the exponent of the power law is less than \(-3\). Steeper (more negative) spectral slopes result in convergence to a finite variance of slope, while shallower (less negative) spectral slopes result in divergence to infinite slope variance. This qualitative difference in behaviour has been noted in the previous chapter where the threshold for fractal behaviour, with its concomitant divergence of slope in Equation (2.6), corresponds to a spectral slope of \(-3\).

Using the same model for the continuous power spectrum as presented in the previous section, the variance of slope can be shown to be:

\[
\text{var(slope)} = k_1 \left( \frac{f_c^{2b_1+3} - f_{min}^{2b_1+3}}{2b_1 + 3} + f_c^{2(b_1-b_2)} \frac{f_{max}^{2b_2+3} - f_c^{2b_2+3}}{2b_2 + 3} \right) \quad b_1, b_2 \neq -3/2
\]  

(3.20)

where, as before, \( f_{min}, f_{max} \) and \( f_c \) are the lowest, highest and critical frequencies and \( b_1 \) and \( b_2 \) are the exponents of the amplitude spectrum so the corresponding exponents of the power spectrum are \( 2b_1 \) and \( 2b_2 \). This equation does not require the coherence of phases assumed in the previous section, so its results can be interpreted.
3.6. IMPLICATIONS FOR SCALING

quantitatively.

Using $f_{\text{max}}$, the highest frequency, as a variable controlling the level of detail in the representation this equation describes the variance of slope as a function of resolution. Figure 3.27 shows the square root of slope variance as a function of highest frequency, computed using Equation (3.20) for the three sites.

The slope variance converges to a constant value at scales finer than about twice the critical frequencies. This indicates that including information at finer scales than this does not expand the distribution of slopes. This is not to say that no additional information is included, but that the additional variation in the surface is not large enough to change the frequency distribution of slope. At all three sites it is apparent that little change in slope variance can be expected at scales finer than about 100 m.

This result can be compared with the findings of Moore et al. (1993a) where the cumulative probability distribution of slope was determined for a range of DEM resolutions, using the Brindabella and Livingstone — there called Wagga — sites and one other site. The parameters of fitted two-parameter Weibull distributions were plotted against resolution. The scale parameter of the Weibull distribution, which characterises the variance of the slope, was found to be approximately constant for resolutions finer than 100 m and to decrease at coarser resolutions at a rate related to the fractal dimension. In a similar manner, the different shapes of the curves in the low-frequency (broad scale) portions of Figure 3.27 are due to the different spectral exponents in the broad scale portions of the power spectra. Brindabella, which has a spectral exponent more negative than $-3$, has a slope variance curve that is converging to a value (curving downwards) even before the critical frequency is reached. Rinker, which has a spectral exponent more positive than $-3$, is diverging (curving upwards) before the critical frequency is approached, while Livingstone (exponent $\simeq -3$) appears to be slowly converging with a very slight downwards curve.

A similar relationship can be derived for total curvature (using the fourth moment of the power spectral density) but not for the more useful directional curvatures such as plan and profile curvature because they are sensitive to the shape of the surface, and the power spectrum is not sufficient to characterise shape. Contributing area and specific catchment area likewise cannot be characterised using power spectral density because it is non-local and is sensitive to convergence and divergence along flow paths. Information on such surface organisation is contained in both the magnitude and phase components of the Fourier transform, but phase is not included in the power spectrum and cannot readily be analysed to reveal this information. This structural information is not represented in the frequency domain of the power spectrum: its focus is purely on scale.

This analysis has demonstrated that spectral analysis techniques can be used to determine the scaling properties of at least one topographic parameter, slope,
3.7. LIMITATIONS OF SPECTRAL ANALYSIS

relatively easily and with considerably less effort than analysing a number of different resolution DEMs. Similar analysis of other topographic parameters would require a different decomposition that is sensitive to the organisation of the landscape surface.

3.7 Limitations of Spectral Analysis

In spite of the ability of spectral analysis to extract useful information about changes in surface morphology at various scales, there are some significant limitations on the use of spectral analysis for characterising topography. These fall under the headings of stationarity, sample dependence and the oscillatory nature of the basis functions.

3.7.1 Stationarity

The issue of stationarity has already arisen in the analysis and interpretation of the spectra. The Fourier transform assumes stationarity of the signal — the mean, variance and higher order moments should all be independent of location — but topography is manifestly non-stationary, and in fact the spatial variation of statistical parameters is of great interest. The stationarity condition arises because the Fourier transform uses a single sinusoidal function to describe all the variation at a particular wavelength, which means that the fine scale detail at one location is assumed to be identical to the fine scale detail at a distant location. This property of Fourier transforms also prevents localising features within a sample (a finite DEM) since the analysis treats the sample as one unit and produces results which describe the whole sample.

As noted earlier, power law relationships between spectral power and spatial frequency may arise spuriously because of non-stationarity of the surface. Determining whether an observed power law relationship is due simply to non-stationarity or is a more intrinsic property of the nature of the surface is very difficult. The approach taken to assess stationarity earlier in this chapter could conceivably be extended to finer subdivisions of the data, but the spectra become rapidly less reliable as the sample size decreases preventing effective comparisons.

3.7.2 Sample Dependence

When different samples are taken from a DEM, the spectrum commonly varies from sample to sample; this follows naturally from the problem of non-stationarity. It is usual in spectral analysis to expect that increasing the amount of data should produce a more reliable spectrum, but the non-stationarity of topography means that every expansion in the data set being analysed (potentially) brings new structures into the sample and changes the spectrum. The spectrum would not (in general) tend to
converge to a stable form but would continue evolving as the sample region included larger areas. Does the spectrum computed from a given sample properly represent the structure of the surface or is it an uninterpretable combination of several different spectra from regions with differing characteristics? The validity of the spectrum could be assured if there were a priori grounds for deciding that the region covered by the sample was homogeneous. In some cases it may be possible to argue this but in general the homogeneity of an area is one of the questions to be answered by the analysis, not presumed by it.

3.7.3 Oscillatory Basis Functions

The third difficulty with spectral analysis is that it uses oscillatory waveforms (sine and cosine functions) as the basis functions into which the sample data is decomposed. If the basis functions are not a good representation of the fundamental shapes occurring in the landscape then good localisation in scale (wavelength) is not possible. While profiles from landscapes sometimes appear periodic, surfaces are usually more irregular in their orientations and structures so periodic functions do not represent the surface well (with a few exceptions, where the geomorphic processes give rise to regularly spaced features in the same orientation, such as dunes). For this reason, there is little point extending to two dimensions the analysis of the previous section: the association of low frequencies with broad scale features and high frequencies with fine scale features is not as close as the analysis requires, because the shape of the Fourier basis functions does not match the shape of the landforms.

Non-sinusoidal shapes of a given size produce harmonics at finer scales (shorter wavelengths) which can overwhelm the contribution of smaller-amplitude features at those wavelengths. This also implies that a single feature in the landscape (if such a thing exists) is represented in the Fourier transform by a substantial number of sinusoidal components, and conversely a single component in the Fourier transform contains contributions from a number of surface features. An extreme example of this would be a cliff — a genuine discontinuity in the data set — which would introduce substantial high frequency energy into the spectrum that may mask the spectral characteristics of fine-scale structures elsewhere in the landscape.

3.8 Summary

Fourier spectral analysis has been used here to

- study the scaling properties of topographic surfaces;
- examine the source of the commonly observed change in spectral slope at 200 to 500 m wavelengths;
3.8. SUMMARY

- link the wavelength at which spectral slope becomes more negative to hillslope length; and

- derive the scaling behaviour of topographic slope.

Although the power law exponents in the power spectra usually fall outside the range allowable under the fractal model, the identification of power law relationships still indicate the existence of scaling relationship that might be related to geomorphic processes. The power law relationship permits the analytical evaluation of the relationship between slope variance and scale.

The analysis of DEMs of different scale source data and resolutions demonstrated that the change in spectral slope commonly observed at wavelengths of 200 to 500 m has a topographic component, as well as a contribution from cartographic generalisation and interpolation. This analysis establishes that the assumption of a single-power law relationship across all scales used in studies such as Polidori et al. (1991) is unwarranted. The modelling of the effects of mapping and interpolation on the spectrum is imperfect. Further study of the phenomena using a wider range of source data scales and different landforms would help to refine the models of the mapping and interpolation filters.

The connection between hillslope length and the wavelength at which spectral slope becomes more negative is supported (but not proven) by the analysis of the simple model of topographic profiles. This connection is physically reasonable and consistent with measurements of drainage density of the surfaces. It is also consistent with the effects of grid resolution on soil-landscape associations reported by Gessler (1996).

The use of spectral analysis is limited by non-stationarity of the topography and the use of oscillatory basis functions. These limitations prevent confident characterisation of the surface using the power spectrum. The existence of non-stationarity in the surface reduces the confidence in the measured power law exponent, as it may be an aggregate of slightly different spectral exponents in different locations as suggested by Figure 3.25. The reliance on oscillatory basis functions hampers the interpretation of the power spectrum in terms of geomorphic features.
Wavelet analysis is a method of decomposing a signal into a set of oscillatory basis functions at different frequencies in a similar way to the Fourier transform. Unlike the Fourier transform, the wavelet transform uses localised functions — wavelets — rather than infinitely repeating sine and cosine functions. This gives the wavelet transform two significant advantages for topographic analysis: the ability to analyse non-stationary signals and the ability to provide spatially variable results.

The wavelet transform uses a single basis function that is localised in both position and frequency and can be translated and dilated to cover the entire position-frequency plane. This contrasts with the Fourier transform which uses many basis functions each at a different frequency and with no position localisation. The effect of using dilation (or stretching) of the wavelet to account for different scales is that high frequencies are analysed over a small distance while low frequencies are analysed over much larger distances.

The discussion here will be very brief and will touch on only a few of the important properties of wavelets. More thorough but still simple reviews of wavelets from various perspectives can be found in Rioul and Vetterli (1991), Chui (1992b) and Jawerth and Sweldens (1994). More detailed treatments can be found in Daubechies (1992) and Chui (1992a).
4.1 Theory

The wavelet transform is, like the Fourier transform, the product of the signal with the basis function

\[ \mathcal{W}(v(x)) = T(x, a) = |a|^{-1/2} \int_{-\infty}^{\infty} \psi \left( \frac{u - x}{a} \right) v(u) du \]  \hspace{1cm} (4.1)

where \( x \) is location, \( a \) is scale, \( \psi \) is the wavelet basis function and \( v \) is the signal to be analysed (Daubechies, 1992, for example). \( \mathcal{W} \) denotes the wavelet transform operator. Scale is approximately equivalent to wavelength or the reciprocal of frequency in spectral analysis. In contrast to the Fourier transform, the wavelet transform of a signal is a function of location.

The wavelet transform has an inverse, known as the “resolution of identity” formula:

\[ \mathcal{W}^{-1}(T(x, a)) = f(x) = C_\psi^{-1} \int_{-\infty}^{\infty} |a|^{-1/2} \int_{-\infty}^{\infty} \psi \left( \frac{u - x}{a} \right) T(u, a) du \hspace{1cm} (4.2) \]

where \( C_\psi \) is a constant that depends only on the wavelet:

\[ C_\psi = 2\pi \int_{-\infty}^{\infty} |\Psi(f)|^2 |f|^{-1} df \]  \hspace{1cm} (4.3)

\( \Psi(f) \) is the Fourier transform of \( \psi(x) \). For this integral to converge, \( \Psi(f) \) must tend to 0 as \( f \) tends to 0, implying that the mean of \( \psi(x) \) must be 0 (Daubechies, 1992).

The discrete wavelet transform of finite sampled data is:

\[ T_{m,n} = a_0^{-m/2} \sum_{j=1}^{n} v(x_j) \psi \left( \frac{x_j - x_n}{a_0^{-m}} \right) \]  \hspace{1cm} (4.4)

where \( m \) is the step in location \( (x) \) and \( n \) is the step in scale \( (a) \).

4.1.1 Choice of Wavelet

In both the continuous and discrete cases the wavelet \( \psi \) must have a zero average value for \( C_\psi \) to be finite and for the inverse wavelet transform to exist. For good analysis performance \( \psi \) is typically chosen to be well localised in both position and scale (frequency). In other respects there is considerable freedom and scope for trading off some characteristics for others.

Some choices of the wavelet \( \psi \) lead to a set of orthonormal bases in a similar way to the Fourier transform. In these cases, the coefficients of the discrete wavelet transform can be used to reconstruct the original signal using the original wavelet (orthogonal wavelet bases) or using another related wavelet (bi-orthogonal bases).
4.2. MULTIRESOLUTION ANALYSIS

![Diagram of multiresolution analysis]

Figure 4.1: A schematic diagram of multiresolution analysis. A function $f$ is passed through two filters, a low pass filter $H$ producing the low-frequency component $u_1$ and a high pass filter $G$ producing the high-frequency component $v_1$. The low frequency component is further processed by the same pair of filters (with their frequencies halved) producing another low and high frequency pair, and so on. Reconstruction is the same process in reverse: the final low and high frequency components $u_n$ and $v_n$ are recombined to produce $u_{n-1}$ which is then combined with $v_{n-1}$ and so on until at the final stage $f$ is reconstructed.

Most other choices of wavelet allow only approximate rather than perfect reconstruction. Another desirable property of wavelets is symmetry: analysis with a strongly asymmetric wavelet would give different results if the data were reversed, which is undesirable. For efficiency in computing the transform, compact support (a finite region over which the wavelet is non-zero) is attractive. Unfortunately some of these properties are incompatible. For example, it can be shown that there are no symmetric, compactly supported, orthonormal wavelets apart from one simple but unattractive exception, the square Haar wavelet (Daubechies, 1992, chapter 8). All other compactly supported, orthonormal wavelets are asymmetric, and no other symmetric, orthonormal wavelets are compactly supported. If reconstruction is less important than analysis then orthonormality can be sacrificed to allow symmetry and compact support.

4.2 Multiresolution Analysis

Multiresolution analysis is the iterative decomposition of a signal $f$ into a low-frequency part $u$ and high-frequency part $v$ using low and high frequency filters $H$ and $G$ respectively. The first low-frequency part $u_1$ can be further decomposed into two parts $u_2$ and $v_2$ and so on to form a cascade of components as shown in Figure 4.1. Since each component at the first level contains half the frequency space of the original signal, they can be subsampled by a factor of two without loss of information (this results in aliasing in the high-frequency case but because there is no low-frequency content there is no loss of information and the signal can still be recovered). The second level of decomposition, which decomposes only the $u_1$ component, must therefore only process half the number of samples as the first level. If there are $N = 2^k$ samples in the original data set, this halving of the number of samples can continue until only one sample remains in $u_k$ whereupon the decomposition is complete. This progressive
sub-sampling has two effects: firstly, the number of coefficients produced by the analysis is the same as the original number of data points; secondly because the number of data points halves at each level the total number of operations is proportional $N$ (Riou and Vetterli, 1991). This efficient transform is called the fast wavelet transform or FWT by analogy with the fast Fourier transform, and is closely related to techniques independently developed in signal processing known as sub-band coding.

It would appear that perfect separation of the frequency space is required for this scheme to work properly; however if overlapping low-pass and high-pass filters are matched in a particular way the original signal can still be reconstructed perfectly. It turns out that “matched in a particular way” corresponds exactly to the use of orthonormal wavelets as the filters. Multi-resolution analysis is therefore intimately connected with orthonormal wavelets.

This multiresolution decomposition can be performed in two dimensions, in which case the low-frequency $u$ part can be sub-sampled to 1/4 the size of the original data, and three high-frequency components $v_h$, $v_v$ and $v_d$ corresponding to horizontal, vertical and diagonal components.

### 4.3 An Example of Wavelet Analysis

The complex Morlet wavelet (Daubechies, 1992) provides analysis capabilities similar to the Fourier transform: it is based on sine and cosine functions and because it is complex valued it produces familiar magnitude and phase results. The complex Morlet wavelet is defined as:

$$\psi(t) = e^{-t^2/\alpha^2} \left( e^{i\pi t} - e^{-\pi^2 \alpha^2/4} \right)$$

(4.5)

where $\alpha$ is a parameter specifying the width of the Gaussian envelope of the wavelet. The exponential factor outside the parentheses provides the Gaussian envelope, the first term inside the parentheses is a complex exponential equivalent to the sum of
a sine and a cosine function, and the second term is a constant that ensures a zero mean value for the wavelet. For this study $\alpha$ was set to 2. Figure 4.2 shows the real (cosine) and imaginary (sine) components of the wavelet.

The useful properties of wavelet analysis techniques are shown in Figure 4.3, which is an implementation of the discrete wavelet transform, Equation (4.4), using logarithmic steps on the scale axis and constant steps (independent of scale) on the location axis. At scale 0 the wavelet “frequency” takes its highest value of one cycle per two samples, the same as the Nyquist frequency in spectral analysis; the lowest “frequency” at scale 100 is one cycle per 23 sample intervals, or about four cycles over the length of the data set.

The signal being analysed is 100 samples of the sum of two sine functions, one with period 12.5 sample intervals and the other with period 6.67 sample intervals. To this smooth signal are added two impulses of amplitude 1 at the 50th and 60th samples. The sine functions have a well-defined frequency but no particular location; the impulses have a well-defined location but no particular frequency.

The wavelet transform detects both the continuous sine signals (broad peaks at a constant scale) and the abrupt impulses (peaks at constant locations). The magnitude surface (Figure 4.3(b, c)) shows peaks due to the sine waves that are relatively constant at all locations but are reasonably well localised on the scale axis. It also shows peaks — particularly at the fine scale — at the locations of the impulses. There are also fine-scale peaks at the ends of the data set where the wavelet transform responds to the discontinuity in the data or its slope.

The phase of the wavelet transform is shown in Figure 4.3(d). Each cycle from black through grey to white represents one cycle of the phase response. At the broad scales, the wavelet “tracks” the low frequency component of the signal while at finer scales it responds to the high-frequency sine wave. This periodic pattern is disturbed by the impulses at the finest scales.

### 4.4 Limitations of Wavelet Analysis

The wavelet transform of a one-dimensional signal as shown here produces a two-dimensional surface of wavelet coefficients. This representation is highly redundant, with many correlations between the individual wavelet coefficients $T(a, b)$ (Daubechies, 1992). When applied to a two-dimensional signal such as a surface or image the wavelet transform produces a two-dimensional magnitude and phase spectrum at every point on the surface; in total this is a four-dimensional result. This high-dimensional data set could be reduced by collapsing the two-dimensional magnitude spectrum to one dimension in the same way as was done for the Fourier transform, and the variation with orientation could be characterised to measure anisotropy if
Figure 4.3: A signal (a) consisting of two sine functions plus two impulses and its wavelet transform: the magnitude as a surface (b), magnitude as an image with large magnitudes giving lighter colours (c), and the phase as an image (d) (the horizontal axis is location and the vertical axis is scale with large scales at the top).
desired. A further reduction of data could be achieved if a function with few parameters was fitted to the one-dimensional spectrum. The spatial variation of these wavelet-spectral parameters could then be studied.

However, wavelet analysis still does not address the third problem of the Fourier transform method which is the use of oscillatory basis functions. Although there is a wide choice of basis functions, the wavelet transform requires the basis function to have zero mean which necessitates a degree of oscillation. A decomposition into components that more closely match the nature of real topographic features would provide greater analytical power and require less effort to deal with unwanted side-effects.
The analysis of scale to this point in the study has focussed on two linked concepts: scaling laws and characteristic scales. Scaling laws provide the basis for the scaling of representations of topographic forms and processes. Characteristic scales are those at which the scaling laws are observed to change: they are the scale boundaries that represent barriers to the scaling of form and process.

Characteristic scales can be identified with the lengths of particular features in the landscape. In Chapter 3 it was demonstrated how the abrupt increase in spectral slope could be connected with hillslope length. This link between characteristic scale and structure gives a concrete meaning to scale in topographic surfaces.

A characteristic scale detected by Fourier analysis indicates the presence throughout the study area of many features (hillslopes in this case) at that particular length. This constancy of hillslope length is linked with a uniform drainage density. In heterogeneous landscapes the variety of hillslope lengths would prevent the clear identification of a single characteristic scale, in spite of the existence of features such as hillslopes or first-order catchments each with a well-defined size.

5.1 Scale and Structure

The central argument of this thesis is that if scale is to be used quantitatively in studies of landscapes it must be connected explicitly with topographic features. A general-purpose analysis method may identify a characteristic scale or a scaling relationship, but without an explicit link between the analytic method and topographic forms such results are difficult to interpret.
5.2. REPRESENTATION AND ANALYSIS

The fractal model fails not only because it insists on particular ranges of scaling exponents, but also because it ignores any organisation of the surface into hillslopes and catchments, plains and mountain ranges. Fourier and wavelet analysis relax the limitations on scaling exponents but still ignore the particular structure of topographic surfaces. A new method of analysing landscapes is called for, one that recognises and takes advantage of the organisation of the landscape into recognisable features at a range of scales, in the same way that ANUDEM (Hutchinson, 1988; Hutchinson, 1989) takes advantage of the organisation of the landscape to sensibly interpolate a surface from contours. Effective characterisation of scale in topography will remain out of reach as long as we rely on analysis tools that cannot distinguish up from down, valleys from ridges, and peaks from depressions.

5.2 Representation and Analysis

An ideal representation of topography would not only permit high fidelity representations with minimal data volume but would also contain all the interesting information about the landscape. An ideal analysis tool would provide not only accurate information about the landscape, but also a complete and unique characterisation of the landscape so that the surface could be reconstructed from the results of the analysis. This convergence of analysis and representation would lead to a single tool with powerful analysis capabilities well tuned to the particular characteristics of topography as well as an efficient and accurate representation.

The existing digital representations of landscapes such as grids encapsulate very little information about landform, so extensive analysis is required to derive that information. A Fourier transform of the surface provides some structural information and also permits reconstruction (provided the phase information is not discarded) but further processing is required before useful results emerge, as shown in Chapter 3.

Triangulated irregular networks (TINs) have been seen as an attractive representation of terrain (Weibel and Heller, 1993; Hutchinson and Gallant, 1997) because they can capture the significant points in the landscape: “significant” here means important for capturing the shape of the landscape. TINs and other locally-adaptive representations such as quadtrees have the very attractive capability of including more points in places where the landscape is more complex. However TINs are not an analysis tool and, like grids, do not directly contain useful information about the surface.

Pike (1988) discusses the geometric signature, a method for characterising landforms based on various measures of terrain shape such as the mean and variance of slope and its distribution, mean and variance of slope (profile) curvature, slope between reversals and relative relief. Part of Pike’s vision for such a measure is that, with the right choice of additional measures and techniques, “a geometric signature
could be complete enough to reconstruct the original piece of topography from which it came”.

5.3 Representation using Topographic Features

The idea that the quality of information contained in a terrain representation depends on the type of representation is not new. Several authors have suggested methods of describing landforms using component features of various types.

Speight (1968) describes a method of identifying, from a contour map, land elements such as crests, hillslopes, foot slopes and water courses using a set of rules relating to surface slope, contour and profile curvature and specific catchment area. These land elements can be aggregated into land systems, providing a description of the landscape at two hierarchical levels. Unfortunately this method suffers from a degree of arbitrariness in the selection of thresholds for slope, curvatures and specific catchment area. These thresholds would probably require adjustment in different landscapes. This system is also difficult to replicate as an automated algorithm.

A method of representing landscapes that recognises both scaling and characteristic scales is described by Clarke (1988), based on the principle that “Modeling topography depends greatly upon the separation of landforms into their component parts”. Clarke separates topography into a deterministic broad-scale component described using a few Fourier components, and a stochastic self-similar component determined from the original landscape. A final local smoothing operator is used to fill sinks; Clarke links this conceptually to diffusive erosion processes but without any attempt to match any real erosion process.

Connelly (1972) discusses information theory in the context of terrain representations, noting that the coding of landforms, “the creation of a numerical ‘language’...which accurately represents terrain”, is of much more interest to geomorphologists than the simple encoding of elevation. A description of terrain can be considered as a coded message, and careful choice of the code can lead to efficient representations. The compactness of the representation is limited by the information content of the terrain itself, that information which must be transmitted to differentiate this particular landscape from all others. Connelly continues the message analogy:

Textual messages are analysed for information content by subdividing them into their component parts, letters, numbers and punctuation. To analyse landforms in a similar fashion, they too must be subdivided into a set of component features. These component features must be so simple that they can be easily coded and yet they must be descriptive so that, when re-assembled, they accurately reconstruct the landform (Connelly, 1972).
This is an exciting concept. Unfortunately, Connelly then continues: "Let us, then, consider elevation to be made up of a large number of spot elevations." This is analogous to representing text as black and white spots, completely ignoring the higher-level information content of the message. One interesting point raised during Connelly's analysis of the information content of elevation arrays (grid DEMs) is that the vast majority of the possible states of a regular array do not represent anything recognisable as topography, so some very strong limits can be placed on the plausible combinations of values in such an array. These limits correspond to the constraints imposed by geomorphic processes.

Blöschl and Sivapalan (1995) conclude their review of scale issues of hydrological modelling by noting that "organisation in catchments is one of the keys to 'enlightened scaling'. New and innovative techniques for quantifying organisation are needed."

The idea that a good representation of a complex object should use the properties of component features appears in contexts other than topography. For example, Cipra (1997) reports on attempts to describe the differences between magnetic resonance images of brains using image deformations, with the constraint that the deformations used should be biologically or physically meaningful. If successful, this technique would allow the interpretation of the differences between two images in terms connected with brain morphology and function. It uses a cascade of deformations from coarse scale to fine scale, each matched to known geometric or biological sources of variation. Another example is the use of adaptive optics in astronomical telescopes to counteract the effects of atmospheric turbulence (Davis, 1997). Here the problem of controlling the mirror shape is solved using a basis set of mirror deformations that best control each type of atmospheric disturbance at its own characteristic frequency (temporal scale). The common thread in these techniques is the use of knowledge about the morphology (in a general sense) of the system in question to develop an information-rich representation. The recurrence of multiple scales in both these examples is noteworthy.

5.4 The Characteristics of a Good Representation

Although scattered through the literature, there appears to be a recurring theme in geomorphology that a representation of topography using topographic features is a worthy goal. Such a representation potentially offers direct measures of geomorphically significant parameters, high efficiency (low redundancy), adaptivity and, if approached in the right manner, innate scalability.

The remainder of the thesis will apply the idea of a representation of terrain that directly links scale and structure, but before commencing it is worth considering the desirable characteristics of a multi-scale feature-based representation of topography.
1. Each feature should be a recognisable landform element (in the general sense). Candidate features might include hills, catchments, plains, fans and rivers.

2. Each feature should have a well-defined scale, which implies some limit on the complexity of its shape.

3. Each feature should have a well-defined shape, but with some flexibility to allow for variations in detail while retaining the same overall form.

4. The parameters describing each feature should be directly interpretable in terms of surface shape or geomorphic processes.

5. The representation should be unique, so that every surface has only one possible representation, or at least one representation should be clearly better (in a quantifiable way) than other possible representations.

In the following chapter a scheme is presented that meets some of these goals and illustrates the use of a feature-based representation for studying scaling properties.
6 A Multi-Scale Feature-Based Representation

6.1 The Feature

The terrain representation proposed here is a superposition of features of a pre-defined shape at various scales. The elevation at any point is the sum of the elevations of the features at that point. A surface can be constructed by introducing broad-scale features first and refining the surface by adding finer features onto the broader features. To avoid introducing discontinuities in the surface the edges of the features must blend smoothly to zero.

The feature chosen to develop the method has an elliptical plan form and a smooth polynomial profile form as shown in Figure 6.1. This is a simple feature with sufficient flexibility to represent topographic surfaces. The parameters of this feature are its location \((x, y)\), length or scale \(L\), width \(w\), orientation \(\theta\) and height \(h\): note that height can be negative, and width is expressed as a multiple of length and takes values between 0 and 1. Other features with similar characteristics but different details could also be used. The polynomial profile is:

\[
F_0(r) = \begin{cases} 
(1 - 4r^2)^3 & |r| < 1/2 \\
0 & \text{otherwise}
\end{cases}
\]  

(6.1)
where $r$ is the distance from the centre of the feature scaled to take account of the elliptical shape. The full equation for the feature can be developed as a series of coordinate transformations starting with a unit diameter circle centred at the origin:

$$F_1(x, y) = F_0 \left( (x^2 + y^2)^{1/2} \right)$$

This circle is then scaled by $L$ in the $x$ direction and $Lw$ in the $y$ direction to form an ellipse:

$$F_2(x, y, L, w) = F_1 \left( \frac{x}{L}, \frac{y}{Lw} \right)$$

Rotation by $\theta$ is then introduced:

$$F(x, y, L, w, \theta) = F_2(p, q, L, w) \quad (6.2)$$

where $p = x \cos \theta - y \sin \theta$

$q = -x \sin \theta + y \cos \theta$

This feature satisfies requirements 2 and 3 of the list of Section 5.4: it has a well-defined scale and shape. It partly satisfies requirements 1 (being a recognisable feature) and 4 (having directly interpretable parameters). The degree to which this model satisfies these requirements will be discussed as part of the assessment of surface decompositions in Section 6.5.4.

To demonstrate that this feature is capable of representing landforms, Figure 6.2 shows an idealised small catchment synthesised using four of these features. The parameters of the features are listed in Table 6.1. The first, broadest scale feature is
narrow (width = 0.2) and oriented east-west across the patch. The second and third features form hilltops and ridges running approximately north-south on either side of the valley. The fourth feature is negative and forms the valley catchment in the slope on the side of the hill. More features could be added to the surface to represent additional detail.

Constructing an artificial surface using these features is simple, but the more interesting problem of decomposing an existing surface into features is more difficult. Many approaches are possible ranging from global techniques to adjust a large number of features simultaneously to match the entire surface, or iterative schemes that fit the broadest- or finest-scale features first. The technique used here is an iterative algorithm called positive wavelet analysis.

6.2 Positive Wavelet Analysis of Profiles

Positive wavelet analysis (Watson and Jones, 1993) is similar to wavelet analysis in that it uses a single function (or basis) which is translated and dilated, but it employs a positive pulse as its basis instead of an oscillatory function; the feature of Figure 6.1 is a candidate positive wavelet. Because the positive pulse does not have zero mean it is not possible to reconstruct the original data from the wavelet coefficients as per the normal wavelet decomposition (Section 4.1). However, by using a process called correlation detection, individual features in the form of translated and dilated (scaled) copies of the basis function can be extracted from sample data.

To demonstrate the correlation detection algorithm, the technique will be used to decompose a one-dimensional profile of the surface. The unit positive wavelet in one dimension consists of the profile of Equation (6.1) scaled by $L$ and offset by location...
6.2. POSITIVE WAVELET ANALYSIS OF PROFILES

\[ x_0: \]
\[ F_{1D}(x, L, x_0) = F_0(\frac{x - x_0}{L}) \]  \hspace{1cm} (6.3)

The signal or function to be analysed will be denoted \( g(x) \). In its one-dimensional form, the correlation detection algorithm analyses \( g(x) \) to find the location and scale of the positive wavelet giving the largest correlation value:

\[ T(x, L) = L^{-1/2} \int_{-\infty}^{\infty} g(t) F_0 \left( \frac{x - t}{L} \right) dt \]  \hspace{1cm} (6.4)

which is the wavelet transform of the function \( g(x) \) using the wavelet \( F_0(\cdot) \), located at \( x \) and scaled by \( L \). This correlation integral gives a large value where the wavelet \( F_0(\cdot) \) and the profile \( g(x) \) are most similar in shape.

Correlation detection of a single feature (in this case corresponding to the wavelet itself) is illustrated in Figure 6.3. The correlation surface \( T(x, L) \) has a single maximum at the location and scale of the signal feature.

The location, scale and amplitude of the \( i \)th detected wavelet are denoted \( x_i, L_i \) and \( m_i \), respectively, with:

\[ m_i = C L_i^{-1/2} T(x_i, L_i) \]  \hspace{1cm} (6.5)

Note that for a constant correlation \( T \), the amplitude \( m \) is inversely proportional to scale \( L \). \( C \) is a normalising constant dependent only on the wavelet shape:

\[ C = \frac{1}{\int_{-\infty}^{\infty} F_0^2(x) dx} \]  \hspace{1cm} (6.6)

This normalisation results in the detected amplitude of 1.0 when the profile \( g(x) \) contains a wavelet of amplitude 1.0, and for the function of Equation (6.3) \( C = 3003/1024 \).

The detected wavelet \( m_i F(\frac{x - x_i}{L_i}) \) can be subtracted from the original function \( g(x) \) to leave a residual profile and the process repeated to find more features. This iteration finds the largest magnitude features first (which in topographic data tend to also be at broad scales) followed by progressively smaller features. This leads to a relatively sparse representation of the sample being analysed provided the shape of the positive wavelet is a good match to the shape of the features in the sample. It is likely that some of the fine scale features will be corrections for errors in the coarse scale components rather than real features in the data and these can be eliminated by global minimisation of error after detection of the features (Watson and Jones, 1993).

End effects in Equation (6.4) cause substantial distortion of the correlation surface at the data boundaries. There are several possible treatments to overcome this:

- constant extension – the terminal data points are extended outside the boundary
Figure 6.3: Correlation surface for a single feature of unit amplitude and scale (width) of 400. (a) the feature; (b) contours of the correlation surface; (c) a perspective view of the correlation surface.
6.2. **POSITIVE WAVELET ANALYSIS OF PROFILES**

![Graphs](image)

Figure 6.4: Illustration of the types of data extension used to reduce end effects. (a) original data, (b) constant extension, (c) periodic extension, (d) reflection and (e) inverted reflection.

as far as required.

- periodic extension — the data is replicated in its original order so the last data point is followed by the first. This usually results in a discontinuity at the data boundary.

- reflection — the data is replicated in reverse order so the last data point is followed by the second-from-last and so on. This results in continuity across the boundary but with discontinuous slope.

- reflection with inversion — as for reflection, but the data is inverted and offset to maintain continuity in both value and slope.

Figure 6.4 illustrates these different types of extension. Tests showed that the least distortion was obtained at broad scales using periodic extension and at fine scales using reflection with inversion, but periodic extension produced the most distortion at fine scales and reflection with inversion the most distortion at broad scales. Simple reflection was chosen as a compromise that was reasonably consistent across all scales.

As in spectral analysis, the mean is removed before analysis; if it were not removed, the first feature found would be required to account for the mean value which requires a feature much larger than the size of the data set. Ideally, there would be additional data beyond the finite data set to characterise the shape of the surrounding surface.
at this coarse scale but this is not the case in practice. Trends, however, are not removed, neither is any windowing required. Apart from the end effects this analysis technique is not affected by non-stationarity, and trends will be accounted for by the largest-scale features detected; the data set does contain sufficient information to determine the size of these features.

The correlation integral (6.4) is evaluated on a regular grid in the \((x, L)\) plane to determine local maxima and minima, with linear steps in \(x\) (at every point in the profile) and logarithmic steps in \(L\). The coordinates of each local extremum is refined after detection by a local maximisation algorithm. The algorithm used is an adaptation of the Levenberg-Marquardt non-linear maximisation method (Marquardt, 1963; Press et al., 1989) which uses first and second derivatives to rapidly locate extreme points; the continuity of derivatives of \(F_0(x)\) helps maintain stability of the maximisation algorithm.

In wavelet parlance scale \(L\) is a quantity with units of length, so is well defined: a large value of \(L\) corresponds to coarse scales or large lengths. For the remainder of this chapter, "scale" will be used in this sense as well as in the more generic sense, so the phrases "large scale" and "increasing scale" should be understood to mean "at large length scales" and "at increasingly large length scales" respectively. The terms "broad scale" and "fine scale" retain their usual meanings.

6.3 Profile Results

Figure 6.5(a) shows the profile being analysed and the first three components detected using the iterative procedure. The close fit of the positive wavelet to the large central feature is an encouraging indication that the wavelet form chosen is reasonable. As noted previously, the largest features in the profile are detected first. The correlation surface from which the first three features were detected is shown in Figure 6.6(a). The minima and maxima are numbered to correspond with the features in Figure 6.5(a). Figure 6.5(b) shows the next three components detected after those of Figure 6.5(a) are removed, and Figure 6.6(b) shows the corresponding correlation surface. The first six components together are shown in Figure 6.5(c). The overall shape of the profile is quite well represented by these six features. Figure 6.5(d) shows the fit using twenty components. As these plots show, the method progressively adds more "bumps" to the fitted profile to bring it closer to the data. The correlation detection process does not seek in the first instance to minimise the difference between the data and the positive wavelet representation but maximises the similarity in shape between the signal and the wavelet. The new wavelet is only subtracted from the data after its parameters are determined, so that the next wavelet can be detected. A total of 75 wavelets (225 parameters) were used to decompose the profile (which itself contains 291 points). The amplitude of the 75th component was less than 1.5m, and there is