
Chapter 2

Modelling Rough Surfaces

2.1 Introduction

The goal of this thesis is to develop a classifier that is suited to the discrimination of rough surfaces, and which is robust to changes in illuminant tilt. It is the belief of the author that this goal can best be achieved by gaining an understanding of, and an ability to model, the physical, as well as the computational aspects of this problem. This chapter considers the first data structure (or signal) in the surface to symbol chain, the surface topography. In subsequent chapters we shall consider the transformations of this signal that occur at higher levels in the process of recognition.

The aim of this chapter is to develop a physically based model of the surface. The first objective is to arrive at a method for characterising textured surfaces and to define the limitations of this technique. Having found a means of description, the second objective is to adopt a group of models which are defined in terms of that method. These models will then form the basis of subsequent analysis of the recognition process.

To attain these two objectives, this chapter is structured in the following way. The first section consists of a literature survey of methods for the description of rough surfaces. The method adopted, the power spectrum, forms a complete description of a surface in conjunction with the phase spectrum. The next section develops a stochastic model for the phase spectra, which includes an *ad hoc* definition of the textures that lie within the scope of this thesis; this section also describes a simple test to determine whether a surface complies with the phase model's definition. Having adopted the power spectrum as the surface descriptor, we next consider rough surfaces models defined in terms of the power spectrum. Three models are adopted: a linear roll-off (or fractal)

model; a fractal model modified to have a flat spectrum below a parameterised cut-off frequency; and a model that varies the cut-off frequency with polar angle to introduce surface directionality. These models are discussed with reference to experimental findings reported in the literature.

2.2 Rough Surface Description

2.2.1 A Brief Review of Possible Sources of Models

Several fields of research have been surveyed to consider descriptive techniques and to construct a suitable model. One source considered was visual texture analysis, where investigators have used a wide variety of techniques such as Markov modelling [Chellappa85], ARMA modelling [Kayshap84] and Gabor filtering [Jain91]. Whereas early approaches in texture analysis were based purely on discrimination, more recently, techniques which attempt to model textures have come to the fore. Furthermore, most techniques are inherently two dimensional and explicitly consider anisotropy. There are, however, two drawbacks to the use of texture analysis techniques. Firstly, most are designed to segment an image and are consequently optimised for the localisation of a texture, rather than its accurate representation. The second and more serious difficulty stems from the fact that these techniques have been developed for visual textures. Applying them to an application outside their original context has the consequence that reported parameter values cannot be used to model surfaces, so any surface model would lack experimental verification in this important respect. Texture analysis techniques will therefore not be considered as a means of either surface description or modelling.

Another potential area of interest is terrain modelling, e.g. [Austin94]; recently interest in modelling land surfaces has increased, largely motivated by the development of fractals and microcomputers. Because the subject is comparatively new, there is little literature and theoretical framework compared to some of the more established areas such as scattering. The physical basis of the application, and the overlap in descriptive methods, does however, make this a possible source of realistic modelling data.

The texture of surfaces is of interest in many engineering applications. The interaction of light with surfaces which have been “optically finished”, i.e. manufactured to an accuracy comparable with the wavelength of light, is of considerable interest in

scattering theory. Surfaces manufactured to less exacting standards are also studied, usually with reference to their frictional properties in the field of tribology¹. The combination of theoretical background and the emphasis on actual physical measurements makes these fields the most attractive areas to survey, and this chapter will be restricted to these fields.

Both scattering theory and tribology use surface description as a stepping stone towards subsequent analysis of their respective properties of interest. Despite this common interest in surface description, the fields are distinct and often use different parameters. In general, the descriptive techniques favoured in scattering theory, e.g. the Power Spectral Density (PSD) and rms roughness, are more suited to this research than their counterparts, e.g. the Autocorrelation Function (ACF) and centre line average, used in tribology. On the other hand, tribological surfaces exist on a scale that is much closer to, and which may actually overlap with, that at which a vision system is likely to operate. It follows that surfaces of this scale are more likely to resemble those which a vision system will encounter and are therefore more reliable sources of model parameters. Fortunately, there is a large degree of overlap in the subjects and we choose to treat both simultaneously.

A description of a surface may be made on several levels: a single parameter may be sufficient to characterise a surface for some purposes; in other cases a much greater degree of description is required. The organisation of this section follows an increase in the sophistication of the description and its ability to define a surface; the flow of the section is in consequence approximately chronological. We have resolved three distinct levels of description. The first level seeks to characterise some property of the surface with a single parameter such as the root mean square (rms) height or gradient. On the second level, a statistical model, in the form of a histogram, is applied to the variation of height or gradient. This is a natural extension to rms values, i.e. the standard deviation of surface height and slope, but in another respect it represents a paradigm shift since it allows the use of surface models, such as Beckmann and Spizzichino's height model [Beckmann63] or Torrance and Sparrow's slope model [Torrance67]. At the third level in our taxonomy are those techniques which incorporate spatial interaction, such as the PSD and the ACF. Given certain assumptions discussed elsewhere in this chapter, these form the basis for models of a wide range of surface textures.

¹ Tribology is the study of friction, lubrication and wear between moving surfaces.

2.2.2 Single Parameter Description

The most basic form of profile description requires the use of only one parameter. The 1930's saw new emphasis placed on profile measurement and a large number of measures were developed around this period [Parsons, p.268]. These measures vary in the degree to which they are specialised to a particular application and in their mathematical tractability. We will consider the two most common and general measures of roughness, *rms roughness* σ_s and the *Centre Line Average* (or *average roughness*) R_{cla} , in addition we will discuss the much less popular *rms slope* parameter.

σ_s (2.2.2a) and R_{cla} (2.2.2b) have, broadly speaking, been used in different fields; σ_s roughness has tended to be used to describe optical finish, whereas R_{cla} is more commonly associated with machined surfaces [Bennett89 p.39]. Despite this, even in the field of machining, most theoreticians prefer to use σ_s , since it allows the use of statistical random process techniques (these will be discussed later in this chapter). It is worth noting that for surfaces which have height distributions conforming to the normal distribution, there is an approximate relationship between the two parameters (2.2.2c). In this thesis, however, the preferred parameter will be *rms roughness* due to its compatibility with statistical analyses.

$$\sigma_s = \sqrt{\frac{1}{n} \sum_{x=0}^n [s(x) - \bar{s}(x)]^2} \quad (2.2.2a)$$

where $s(x)$ represents the height of the surface at a point x along the profile.

$$R_{cla} = \frac{1}{n} \sum_{x=0}^n |s(x)| \quad (2.2.2b)$$

$$R_{cla} \approx 0.8 \sigma_s \quad (2.2.2c)$$

The *rms slope* parameter, m_{rms} (2.2.2d), is much less commonly used than either of the other measurements. Bennett states that the measured *rms slope* may vary by a factor of 50 depending on the profilometer and the separation of sampling points [Bennett89 p.40]. Whereas *rms roughness* is commonly used in physical scattering

models, such as Beckmann and Spizzichino [Beckmann63], *rms slope* has been used in geometric optics models such as Torrance and Sparrow [Torrance67]. Despite the problems associated with this parameter, it is useful in the context of this thesis. The same texture analysis technique may be applied to images ranging from electron microscopy to satellite imagery; absolute scale is in many cases relatively unimportant. The slope of a facet is more important than its height, and the *rms slope* provides a parameter particularly suited to the purposes of this research.

$$m_{rms} = \sqrt{\frac{1}{n} \sum_{x=0}^n [s'(x) - \overline{s'(x)}]^2} \quad (2.2.2d)$$

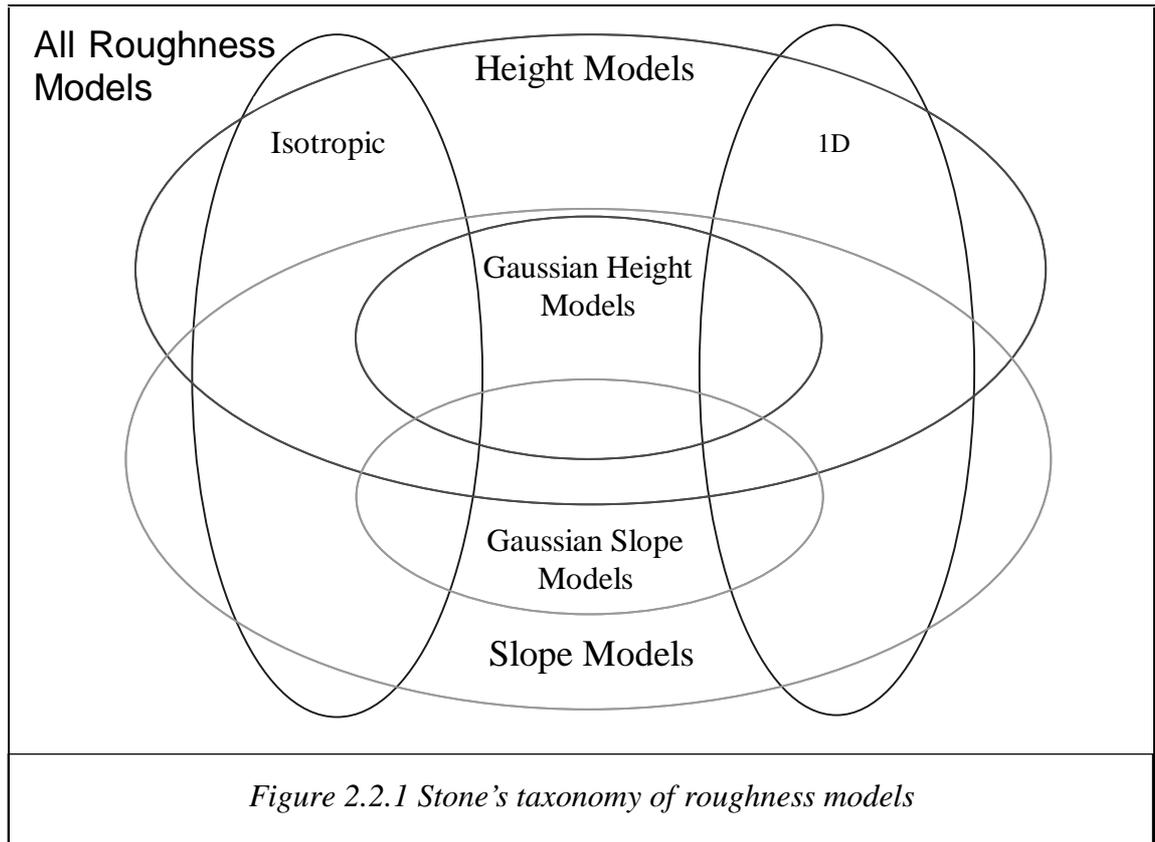
where $s'(x)$ is the derivative of the function $s(x)$ at point x .

In practice, the values of both the *rms roughness* and *slope* depend on the measuring instrument. The length of the sample, the area resolution and the sampling rate all affect the parameters—this being especially true of the *rms slope*. Strictly speaking, neither the *rms slope* nor the *rms roughness* are defined for non-bandlimited surfaces.

2.2.3 Histogram Description

In retrospect, it now seems a natural progression from considering the *rms roughness*, i.e. the standard deviation of the heights of a surface, to the adoption of a statistical model of height distributions, though this development does not seem to have followed quickly. We believe that use of histogram description represents a significant shift in the way surfaces were thought about since it not only implies a new degree of discrimination between surfaces but also allows a certain degree of modelling.

The histogram does give an insight into the nature and history of a surface. Surfaces which are the result of a large number of random events tend to have a Gaussian distribution. Wear, grinding and abrasion tend to wear down and deform summits while leaving valleys largely unaffected, which is reflected in the histogram taking on a negative skew [Bennett89 pp.42-44]. Some turning and milling operations and the presence of relatively large particulate matter on an otherwise smooth surface will result in the histogram taking on a positive skew. It is worth noting, however, that use of skew, and of higher moments of the distribution in general, is subject to artefacts in the sampling process.



The histogram has been used to model surfaces: Beckmann uses a normal distribution of heights in his scattering model, while Torrance uses a normal distribution of slopes for his geometric model of specular reflection from rough surfaces. Stone [Stone94] develops a taxonomy of models used for reflectance modelling in machine vision, shown in *Figure 2.2.1*. Stone defines two important classes, *isotropic* and *one dimensional* shown as the ellipses with vertical major axes. However, the major division is between models defined in terms of either slope or height. Important subclasses of each are defined by the Gaussian constraint. Since differentiation is a linear operation, Gaussian height models have Gaussian slope distributions, though unless integrability is enforced, the converse may not be true. While first order statistical surface models have proved successful in these areas, their lack of spatial information means that they are clearly inadequate for the representation of texture.

The statistical parameters of the height distribution have been further exploited in conjunction with some spatial information. The two point height probability distribution, analogous to the co-occurrence matrices used in texture analysis, allows the description of correlation and structure within the profile. The related technique of Markov chains, again with related applications in texture analysis, has also been applied to profile description

[Ogilvy p.29]. However, these techniques have not been widely used in surface description.

2.2.4 PSD and ACF

The forms of profile description discussed above do not take into account any spatial correlation within the surface. This is a profound limit to their ability to characterise surfaces to any satisfactory degree. At the next level of description the surface profile is treated as a realisation of a random process. The tools with which Wiener laid the foundations of random process theory: the autocorrelation function (ACF) and the power spectral density (PSD) are now applied to the surface profile. The autocorrelation function is the most popular approach in production engineering, while the PSD is more commonly associated with scattering theory, due to the surface spectrum's close relationship to the scattered intensity of light from the surface.

Assume the surface profile is wide sense stationary and has zero mean. The autocovariance function (ACVF) (2.2.4a) is the average product of the heights of two points on a profile separated by a distance t , known as lag. The autocorrelation function (2.2.4b) is the normalised (by surface variance) form of the ACVF².

$$c(t) = E[s(x)s(x+t)] \quad (2.2.4a)$$

$$r_c(t) = E\left[\frac{s(x)s(x+t)}{\sigma_s^2} \right] \quad (2.2.4b)$$

where t is lag.

In *Figure 2.2.2* we plot the ACF for various surfaces. The ACF for a white noise process approximates an impulse at zero lag; the ripple texture, shown in *Figure 2.2.3a*, has a gradually decreasing periodic ACF, whereas that of the fractal surface (*Figure 2.2.3b*) falls monotonically as lag increases. The form of the decay for real surfaces has been the subject of much controversy and several models have been proposed; the most common models are the exponential (2.2.4c) and the Gaussian (2.2.4d). The exponential form appears to be the better fit to the experimental data, however, it does suffer from a discontinuity at zero lag; Ogilvy and Foster [Ogilvy89a] show that an exponential decay is equivalent to a linear roll-off for at least part of the power spectrum. In practice, many

² These definitions are based on those used in scattering theory texts, [Ogilvy] and [Bennett], signal processing texts, e.g. [Therrien] define the autocorrelation and autocovariance functions as:

surfaces have ACF which do not conform to either form. Consider the ACF of the ripple surface (the image of which is shown in *Figure 2.2.3a*), this exhibits zero crossings and a large positive peak, suggesting that the surface contains a strong periodicity at that lag.

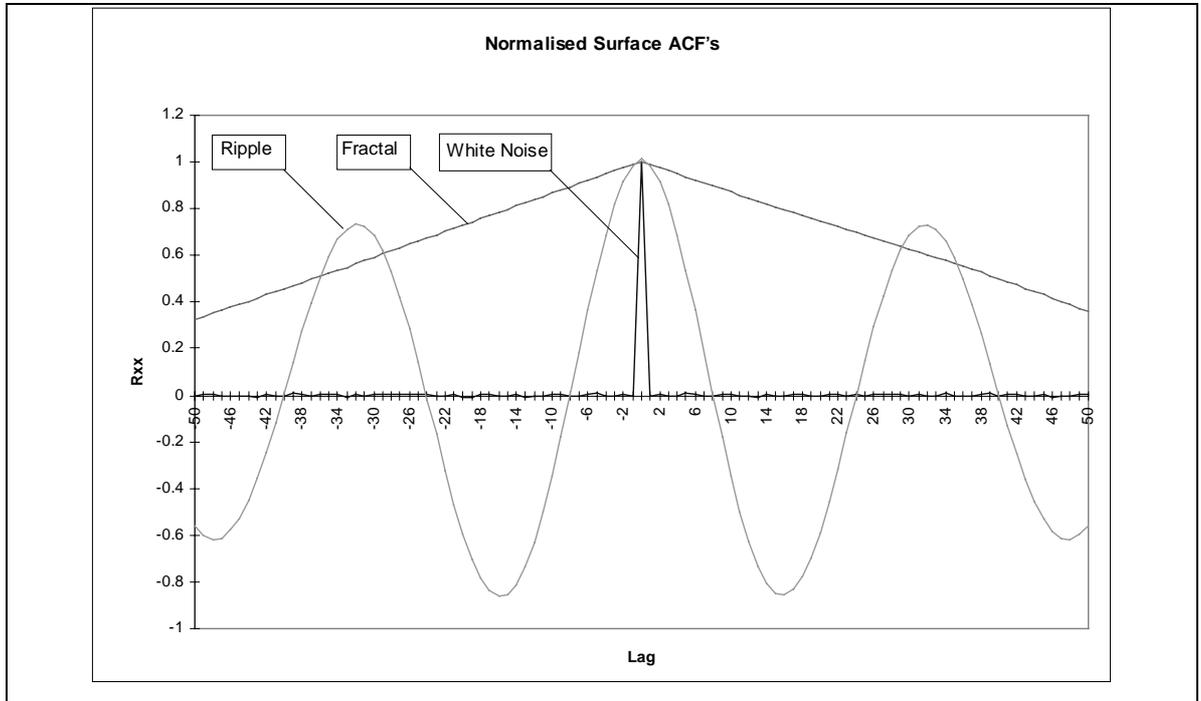


Figure 2.2.2 The 1D ACF for various surfaces

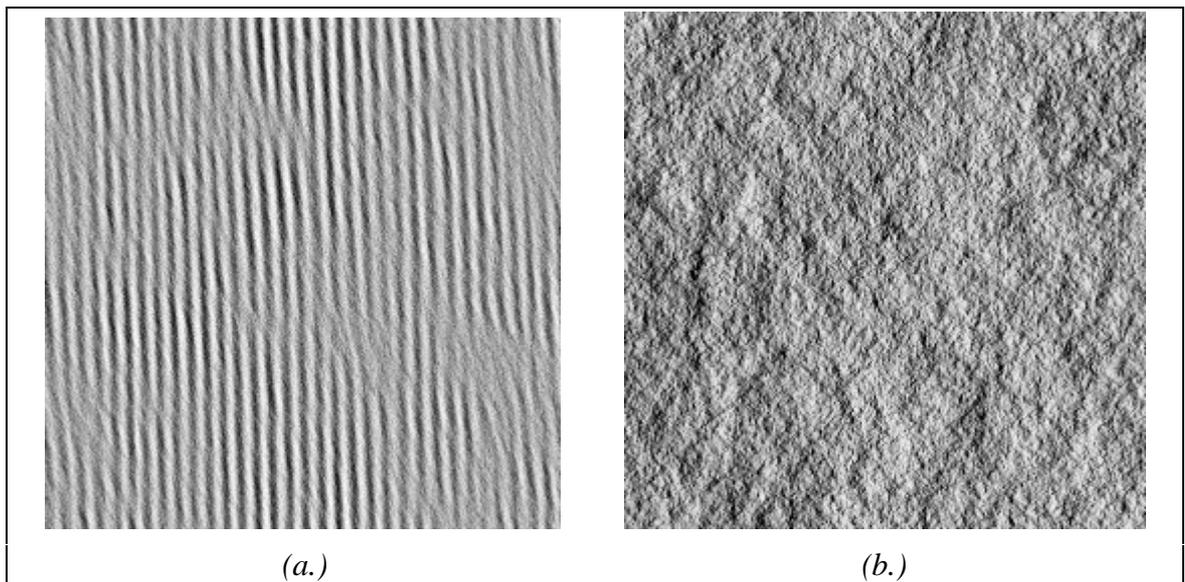


Figure 2.2.3 Synthetic ripple(a) and fractal (b) textures.

$$r(t) = E[s(x)s(x+t)] \text{ and } c(t) = E[(s(x) - \overline{s(x)})(s(x+t) - \overline{s(x)})] \text{ where } \overline{s(x)} = E[s(x)].$$

$$r(t) = \exp\left(\frac{-|t|}{\lambda_0}\right) \quad (2.2.4c)$$

$$r(t) = \exp\left(\frac{-t^2}{\lambda_0^2}\right) \quad (2.2.4d)$$

where λ_0 is the correlation distance.

The power spectral density is equivalent to the ACF and forms a Fourier transform pair with the ACVF—it is also easily related to the *rms roughness* (2.2.4e) and slope (2.2.4f) of a bandlimited surface. Equations (2.2.4e) and (2.2.4f) also illustrate that these parameters are not defined for non-bandlimited fractal surfaces.

$$\sigma_s^2 = \int_{-\infty}^{\infty} S_{1d}(\omega) d\omega \quad (2.2.4e)$$

$$m_{rms}^2 = \int_{-\infty}^{\infty} \omega^2 S_{1d}(\omega) d\omega \quad (2.2.4f)$$

where $S_{1d}(\omega)$ is the power spectrum of the surface profile.

While the rms roughness is an isotropic parameter, the rms slope will vary with direction for an anisotropic surface. In this thesis we will assume that the dominant directionality of the surface is aligned with the image axes, and the directionality of the slope distribution may be parameterised using the rms slopes of profiles taken in the direction of the x and y axes, p_{rms} and q_{rms} respectively. The two dimensional equivalents of (2.2.4f) are shown in (2.2.4g) and (2.2.4h).

$$p_{rms} = \int_{v=-\infty}^{\infty} \int_{u=-\infty}^{\infty} u^2 S(u, v) du. dv \quad (2.2.4g)$$

$$q_{rms} = \int_{v=-\infty}^{\infty} \int_{u=-\infty}^{\infty} v^2 S(u, v) du. dv \quad (2.2.4h)$$

where u is the frequency in the x-direction

and v is the frequency in the y-direction

2.2.5 Relationship between Profile and Surface Spectra

Most work in tribology and scattering deals with one dimensional profiles, whereas we are primarily interested in two dimensional surfaces. The generalisation from one to two dimensions is not trivial and we give a brief description of the process based on two

papers [Nayak71] and [Church83]. We begin with the calculation from a surface PSD. Consider a sinusoidal surface (shown in *Figure 2.2.4b*) with PSD:

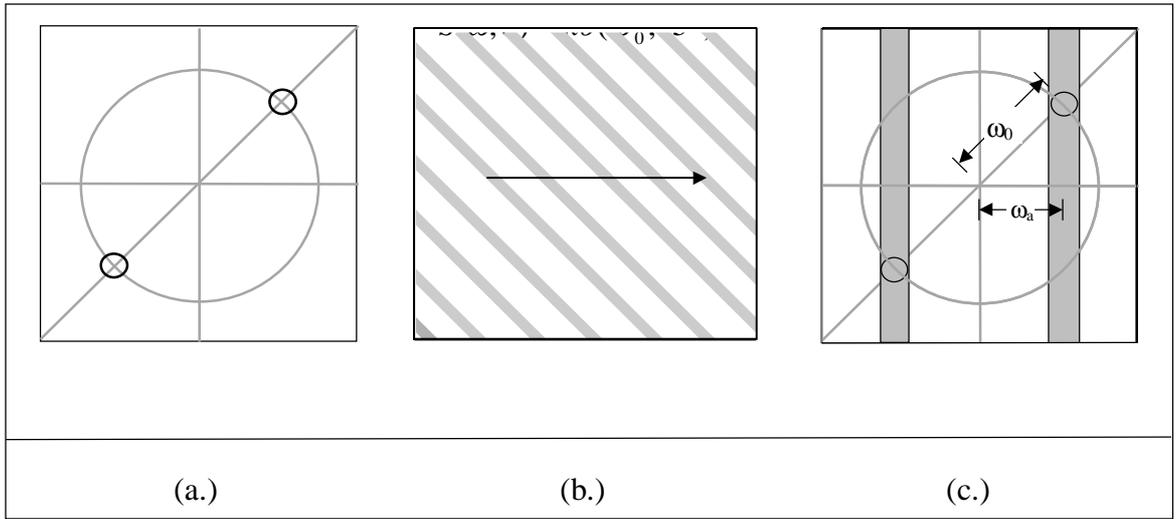


Figure 2.2.4 Calculation of profile spectrum of 2D spectrum.

Now, consider a profile taken in the direction of the x-axis, *Figure 2.2.4b*, this will be sinusoidal with frequency ω_a , such that $\omega_a < \omega$. More generally, the power spectrum of a profile at a given frequency will be the sum of the two dimensional PSD at that frequency and direction and the projections of the two dimensional PSD at all directions and frequencies that project onto this component *Figure 2.2.4c*. Again assuming the profile is in the direction of the x-axis, mathematically this can be stated as:

$$S_{1d}(u) = \int_{-\infty}^{\infty} S(u, v) dv \quad (2.2.5a)$$

Consider now a two dimensional surface, such that the two dimensional power spectral density is dependent only on radial frequency ω i.e. the surface is isotropic.

let
$$\omega = \sqrt{u^2 + v^2} \quad (2.2.5b)$$

$$S_{1d}(\gamma) = 2 \int_{\omega=\gamma}^{\infty} S_{2d}(\omega) \cdot \frac{\omega}{(\omega^2 - \gamma^2)} d\omega \quad (2.2.5c)$$

where γ is the frequency parameter in the one dimensional power spectrum.

If we wish to calculate the two dimensional signal given a profile PSD, we must either assume extreme anisotropy or isotropy. Let us consider the latter case.

Given a one dimensional profile with spectrum $S_{1d}(\gamma)$, we may calculate the PSD of the two dimensional spectrum $S_{2d}(\omega)$, along one direction using:

$$S_{2d}(\omega) = -\frac{1}{\pi} \int_{\gamma=\omega}^{\infty} \frac{d}{d\gamma} S_{1d}(\gamma) \cdot \frac{1}{(\gamma^2 - \omega^2)} d\gamma \quad (2.2.5d)$$

2.2.6 Summary

In this section we have surveyed methods of surface description used in tribology and scattering theory. Three levels of description were discussed. It was argued that single parameter description is not sufficient for our purposes, though the rms slope parameter does measure a surface characteristic which is relevant. The surface height histogram is usually assumed to be unimodal Gaussian, in which case it is closely related to the rms parameters. The histogram, however, like the single parameters, lacks any description of the spatial interaction of points on the surface and therefore does not contain sufficient information for texture description. Second order statistics, i.e. the PSD and the ACF, which incorporate the spatial relationship of surface heights were then considered. The relationship of the 1D and 2D forms of the power spectra was discussed, and an analytical expression from the literature for the isotropic case was quoted.

The evidence considered in this section suggests that the power spectrum should be adopted as the principle means of description. However, the rms slope was also found to be a useful surface parameter and will be used throughout this thesis. The surface height histogram, and consequently the slope histogram, will be assumed to be Gaussian.

2.3 An Admissibility Criterion Based on Phase

In the previous section it was stated that the most sophisticated level of surface description commonly in use in tribology and scattering is the PSD (or equivalently the ACF). It follows that surfaces with identical power spectra are considered to be identical for the purposes of the analysis. This thesis is concerned with distinguishing between surfaces that have dissimilar appearance under identical conditions. The sufficiency of the PSD-only approach must therefore be considered in this context. Following from our PSD-only approach we must impose certain bounds on the phase spectra of the surfaces which we will consider in this thesis.

Our next objective is to determine for what type of surface the condition holds. Unfortunately, we do not possess a data set of surface topography. We do however have a series of textured images each of which form a two dimensional data set sharing many of the textured characteristics of the surface. By applying a test of the phase condition to this data set, we will identify those characteristics which either include or exclude a particular data set from the scope of this thesis.

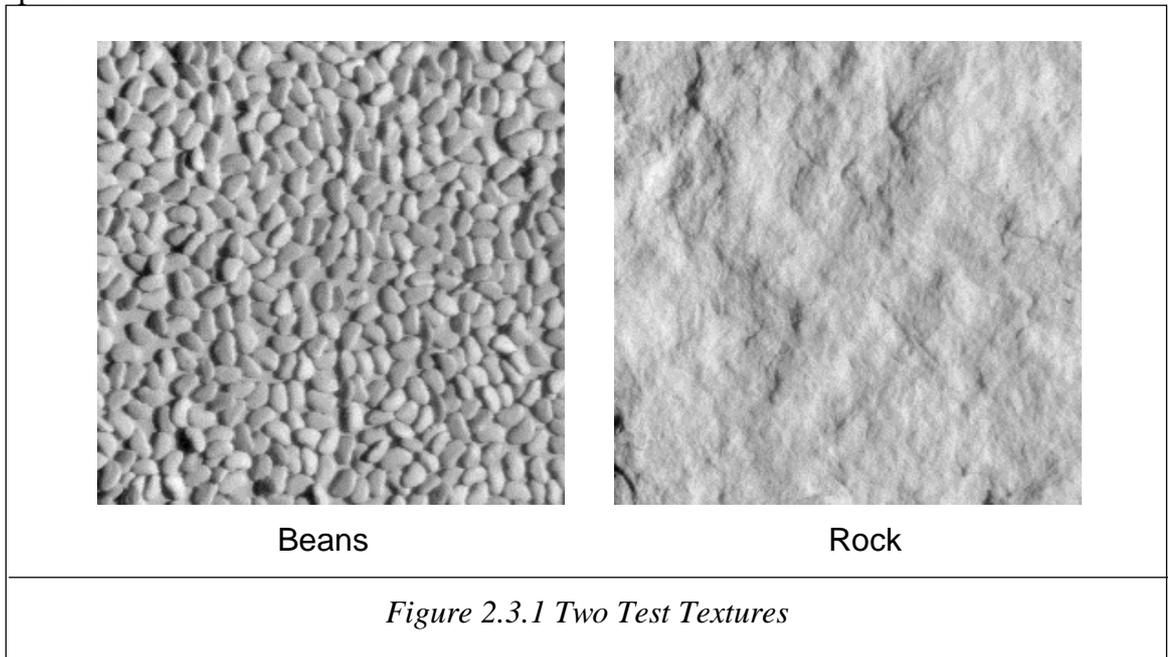
We now consider some general requirements of the phase condition. Firstly, the model must be sufficiently specific to allow discrimination, yet general enough to allow the grouping of similar textures. A further requirement is that the assumptions of the model should easily be met for synthesis. Finally, the degree to which the model is appropriate for a particular texture must be measurable. With these requirements in mind, next we consider the nature of a possible model.

2.3.1 A Phase Condition

It is well established that images with the same PSD may be easily discriminable: both the power and phase spectra are required for complete reconstruction of the image. This applies equally well to the surface. If we discriminate between surfaces solely on the basis of their PSD we cannot *guarantee* a correspondence with any reasonable visual discrimination. One approach to circumvent this problem is to adopt a model of the phase spectra and to restrict our research to textures which satisfy the model's assumptions. If the model is appropriate, we will be able to extract the PSD of a texture and use it, together with a random realisation of the phase component of the model to generate a texture which can be reasonably said to belong to the same class as the original texture.

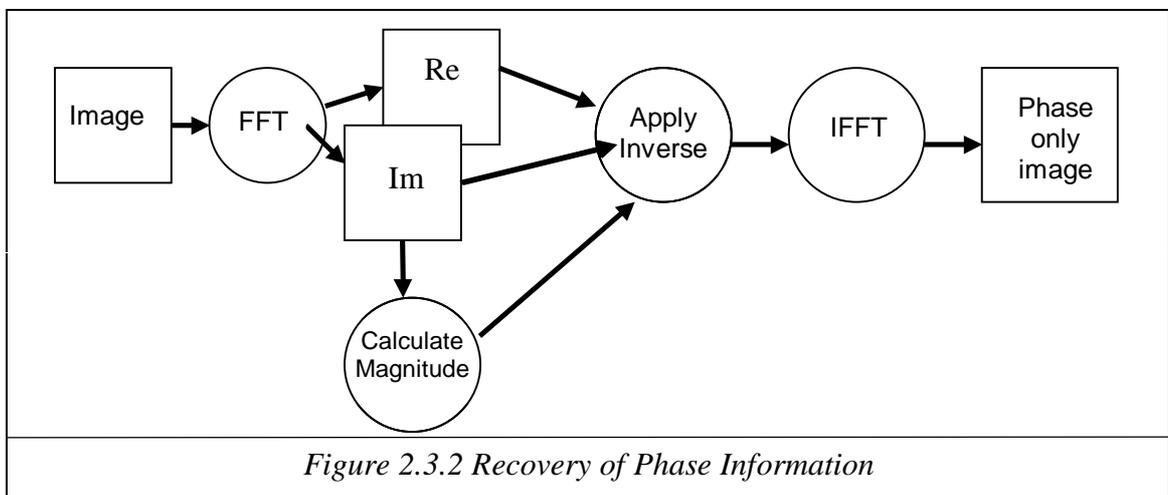
We adopt the concept of a maximum entropy phase spectrum, i.e. the phase spectrum should contain no discriminatory information and all characteristics of the texture should be encoded in the power spectrum. Clarke [Clarke92] has shown that the appearance of deterministic textures, i.e. those containing a large degree of structure such as Beans (*Figure 2.3.1*) is more sensitive to changes in phase than that of unstructured textures such as Rock. It is reasonable to conclude that unstructured textures are more effectively defined in terms of a PSD-only model than their structured counterparts. This leads us to predict that the maximum entropy requirement will be fulfilled by unstructured

textures; the next step is to formalise this either directly, or indirectly, in terms of the phase spectra.



2.3.2 A Simple Test for the Condition

Given the difficulty in interpreting phase information, it would clearly be advantageous to express information held in the phase spectra in another form in order to assess its effect. If we inverse filter the Fourier transform of the original texture with the resultant of its real and imaginary parts, we obtain a scalar field with a uniform PSD. Any structure in the field will be due to the phase spectrum. An unstructured texture will produce an uncorrelated random field, whereas a structured texture will have a structured field. If we can differentiate between these two cases we can discover how well a texture is described by the PSD-only model.



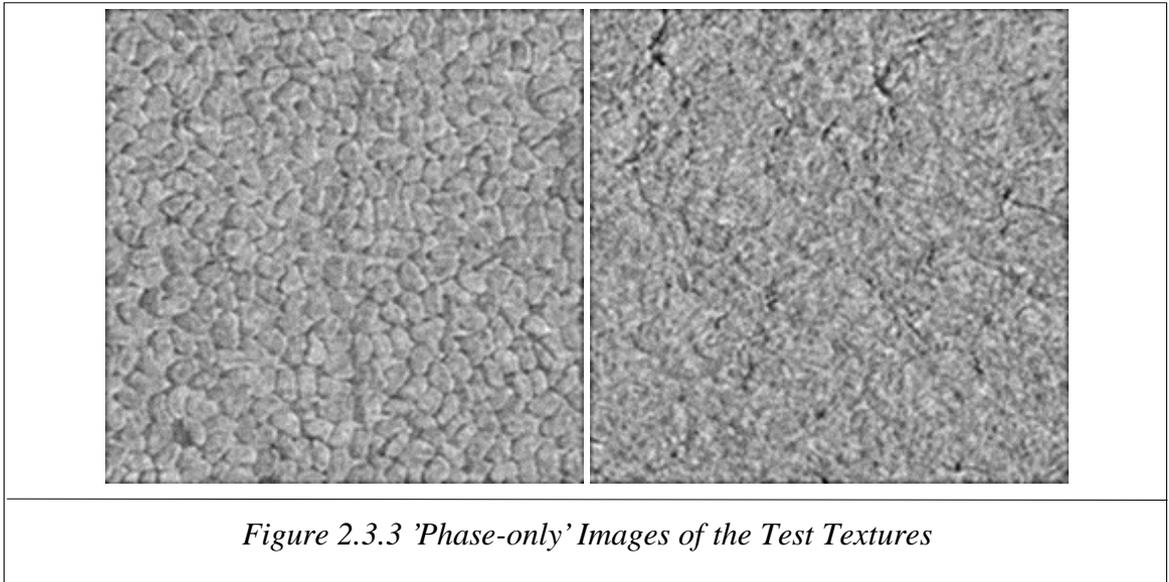
We must now deal with the problem of detecting texture with a uniform power spectral density. Clearly frequency domain models will be inappropriate; we must therefore use probabilistic means. The second order statistics of a spatially uncorrelated random field are completely specified by the first order statistics. If the second order statistics predicted from the mean and standard deviation statistics differ significantly from the measured statistics we may infer that the field is correlated and the texture does not conform to our criteria.

We use the χ^2 goodness of fit test to ascertain whether the measured second order statistics are drawn from the same distribution as those predicted from the first order statistics. We use the following framework:

define i as the intensity of a point (x,y) and g as the intensity at point $(x+\delta x,y+\delta y)$.

For an uncorrelated field the values of i and g are independent and the joint probability function of i and g , $p(i/g)$, is equal to the product of the probabilities of each of the individual probability functions, $p(i)$ and $p(g)$, i.e :

$$p(i/g)=p(i).p(g) \quad \text{for all } \delta x,\delta y \text{ except } \delta x=\delta y=0$$



since i and g share the same uniform probability function:

$$p(i/g)=p(i)^2=p(g)^2 \quad \text{for all } \delta x,\delta y \text{ except } \delta x=\delta y=0 \quad (2.3.2a)$$

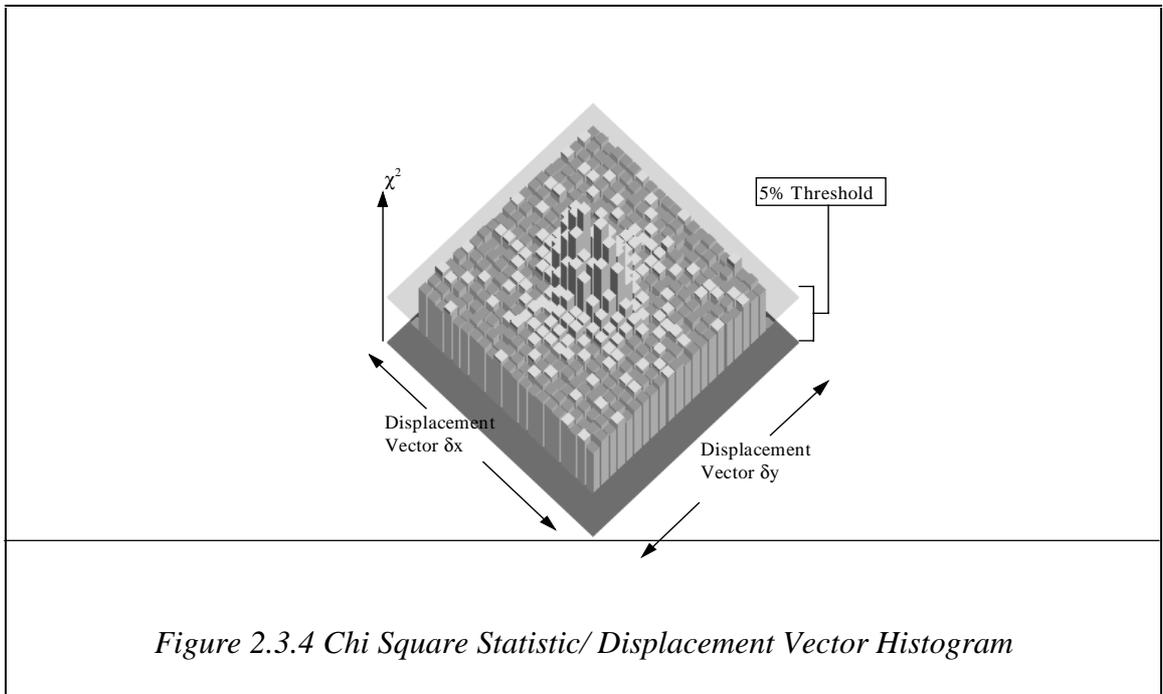
The observed histogram will have $f(i,g)$ elements in each bin (i,g) . For a random field with n elements, the expected number of elements in each bin, by (2.3.2a), will be:

$$E(i,g)= np(i/g)=n.p(i)^2$$

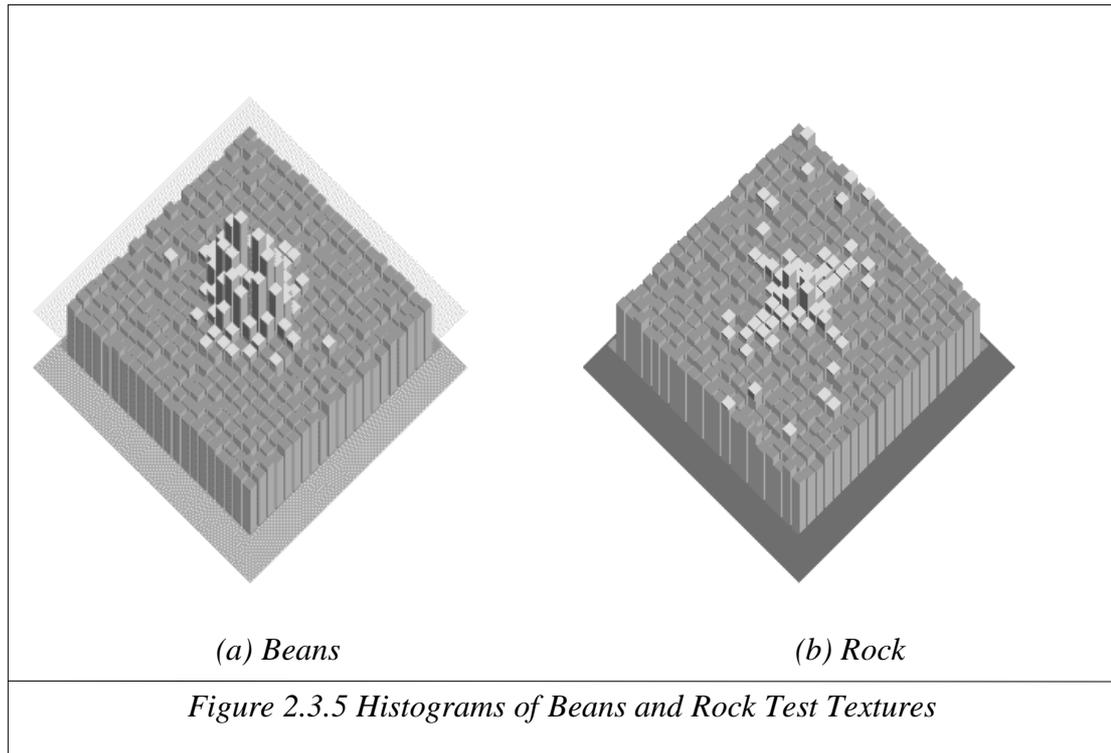
The form of the χ^2 test for this application is shown below:

$$\chi^2 = \sum \frac{[f(i,g) - E^2(i,j)]^2}{E^2(i,j)}$$

We adopt the following procedure: the whitened image is requantised to 32 levels and the histogram equalised; allowing all bins to be filled to a level that will avoid the problems associated with the Chi statistic when it is applied to empty or nearly empty bins. The Chi square goodness of fit criteria is then applied to assess the similarity of the function $f(i,g)$ with $E^2(i)$. For a histogram of these dimensions and a 5% level of significance, the Chi square test gives a threshold of 1098. A plot of the statistic for each displacement vector is given in a 3D histogram (*Figure 2.3.4*). Columns exceeding the 5% threshold are shaded more lightly than those that fall below the level.



In *Figure 2.3.5* the histograms for the Beans and Rock textures are plotted. The first point we note is that in both cases the highest values are clustered around the zero displacement vector. This indicates that the spatial interactions of pixels in the whitened textures is highly localised, suggesting that both the phase-only textures exhibit the Markov property. Beyond this, comparison of the graphs shows that while both textures do exceed the 5% threshold, the structured texture does so in a much more marked fashion.



The test is now applied to a wider range of textures, these are shown in Appendix B. Since it is not practical to plot such a wide range of displacements for all the textures, we take advantage of the localisation property observed earlier and define a second order neighbourhood composed of the eight closest displacements. From this neighbourhood we extract two figures: the mean and the maximum value. The results are tabulated in *Table 2.3.1* and shown in *Figure 2.3.6*.

If we consider the statistic we note that the textures do fall into two distinct categories: those with statistics less than 1500 and those with values greater than 2000. The latter group consists of stones, chips and beans, i.e. those textures which we intuitively identified as being structured. It is also worth noting that the anaglypta textures, which consist of a regularly repeated primitive, have among the lowest Chi statistics.

Texture	Mean	Max		Texture	Mean	Max
Isoroc	1033	1055		Maze	975.5	1044
Pitted	938	965		Ripple	1025	1052
Radial	994	1023		Lip	993	1048
Rock	1081	1179		Stripl	1122	1173
Slab	1063	1109		Beans	2103	2537
Slate	983	1023		Chips	2186	2341
Striate	1057	1137		Rock (Chantler)	1304	1589
Twins	1023	1100		Stones	3380	4555

Table 2.3.1 Chi Statistics

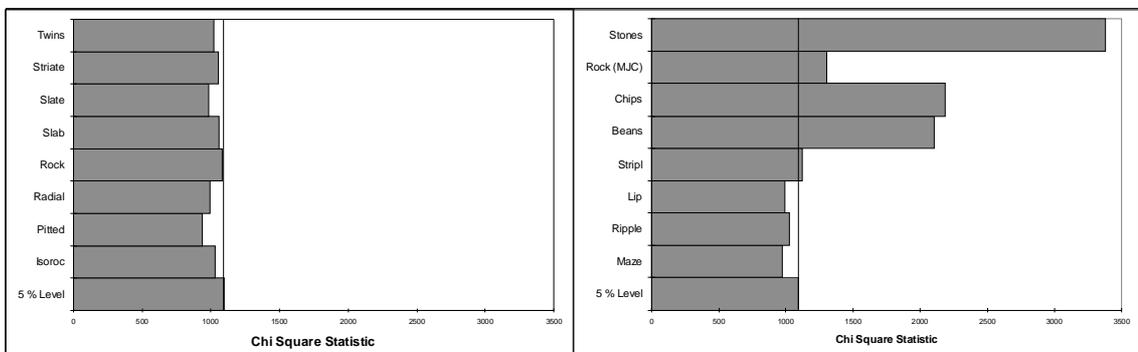


Figure 2.3.6 Histogram of Chi Square Statistic for various textures



Figure 2.3.7 Phase rich textures.

It is clear that the textures Stones, Chips and Beans do contain important phase information, though the other textures also contain a certain degree of structure which is independent of PSD. The key characteristic which differentiates phase-rich textures is the presence of step changes in the value of the data set. In the image data set this occurs due

to cast shadows or the clear delimiting of a texture primitive. In terms of the surface, a step change would manifest itself as a rapid change in height, or 'cliff' topography. This seems to be rare in rough surfaces which have their origin in physical, rather than biological or synthetic processes. The maximum entropy phase condition is therefore compatible with the chosen data set.

By defining random phase textures, we can restrict the scope of this thesis to a class of textures which we can confidently discriminate on the basis of the PSD alone. In the next section we consider models that are expressed in terms of the power spectrum; our interpretation of these models implicitly assumes the phase spectra to be realisations of an uncorrelated random process.

2.4 Models of Surface Roughness

In the previous sections of this chapter we considered some popular techniques for describing engineering surfaces, culminating in the power spectrum. We now look at some models which have been developed from these methods. Two aspects of the model: the *power roll-off* and the *surface directionality* are considered. In the literature these aspects are rarely treated in the same paper: most roll-off models are one dimensional, and directional models are generally stated in terms of *m parameters*, rather than explicitly in terms of the spectrum.

2.4.1 Modelling Roll-Off

The models are generally stated in one of three ways: PSD, ACF or fractal techniques. In this thesis, for the purposes of consistency we will state each model in terms of its PSD. As we would expect, the models do have characteristics in common—an inverse power law is a common feature of most for at least part of the frequency range.

The first model we consider was proposed by Sayles and Thomas in their paper in Nature [Sayles78]. The paper showed that a large number of surfaces, ranging over eight decades of surface roll-off had power spectra exhibiting an inverse square power law with increasing wavelength. The thesis of the paper was that a wide range of surfaces may be modelled with an expression of the form:

2.4 Models of Surface Roughness

In the previous sections of this chapter we considered some popular techniques for describing engineering surfaces, culminating in the power spectrum. We now look at some models which have been developed from these methods. Two aspects of the model: the *power roll-off* and the *surface directionality* are considered. In the literature these aspects are rarely treated in the same paper: most roll-off models are one dimensional, and directional models are generally stated in terms of *m parameters*, rather than explicitly in terms of the spectrum.

2.4.1 Modelling Roll-Off

The models are generally stated in one of three ways: PSD, ACF or fractal techniques. In this thesis, for the purposes of consistency we will state each model in terms of its PSD. As we would expect, the models do have characteristics in common—an inverse power law is a common feature of most for at least part of the frequency range.

The first model we consider was proposed by Sayles and Thomas in their paper in Nature [Sayles78]. The paper showed that a large number of surfaces, ranging over eight decades of surface roll-off had power spectrums exhibiting an inverse square power law with increasing wavelength. The thesis of the paper was that a wide range of surfaces may be modelled with an expression of the form:

$$S_{1d}(\omega) = \frac{2\pi k_t}{\omega^2} \quad (2.4.1a)$$

It follows that these surfaces can be characterised by the factor k_t , known as *topothesy*. In retrospect, the similarity to fractals is obvious and while Sayles and Taylor do not mention it in their paper, Berry and Hannay do relate Sayles' work to that of Mandelbrot in their reply [Berry78]. Berry goes on to describe Sayles' treatment of data, as 'procrustean' and Sayles himself states that many machined surfaces will have their longer wavelengths suppressed by processing. This remark would seem to characterise many of the subsequent modifications to the linear roll-off model.

A more empirical approach to spectral description adopted by several authors is to split the spectrum into two regions of different fractal dimension. Both [Hasegawa93] and [Oden92] fit two lines: with low roll-off at low frequencies and high roll-off at high frequencies. Imre et al. [Imre93], split the spectrum into three regions, microscopic, fractal and macroscopic. The fractal region has dimension 2.4, while the micro and macroscopic regions have no roll-off. A similar, though more elegant, model was developed by Underwood and Banerji in their paper dealing with fractured steel samples. Instead of a linear fractal curve, they report a reversed sigmoidal curve [Underwood86].

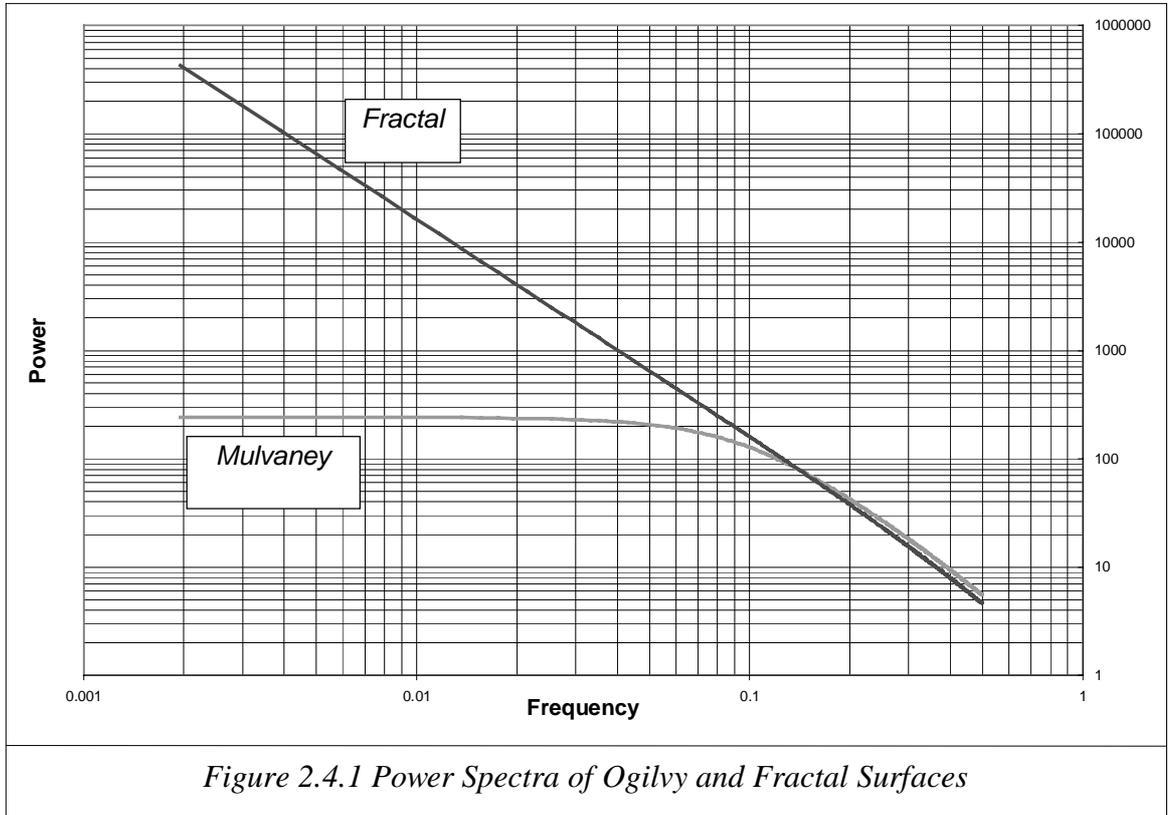
Echoing Sayles' reservations on the power of low frequency components for machined surfaces, Mulvaney et al. have developed an alternative model from experimental data (2.4.1b) and goes onto develop an estimate for surface variance (2.4.1c) [Mulvaney89]. Interestingly the PSD of the exponential correlation function mentioned in section 2.2.4 is of a similar form (2.4.1d) [Ogilvy89a].

$$S_{1d}(\omega) = \frac{k_t}{1 + \left(\frac{\omega}{\omega_c}\right)^2} \quad (2.4.1b)$$

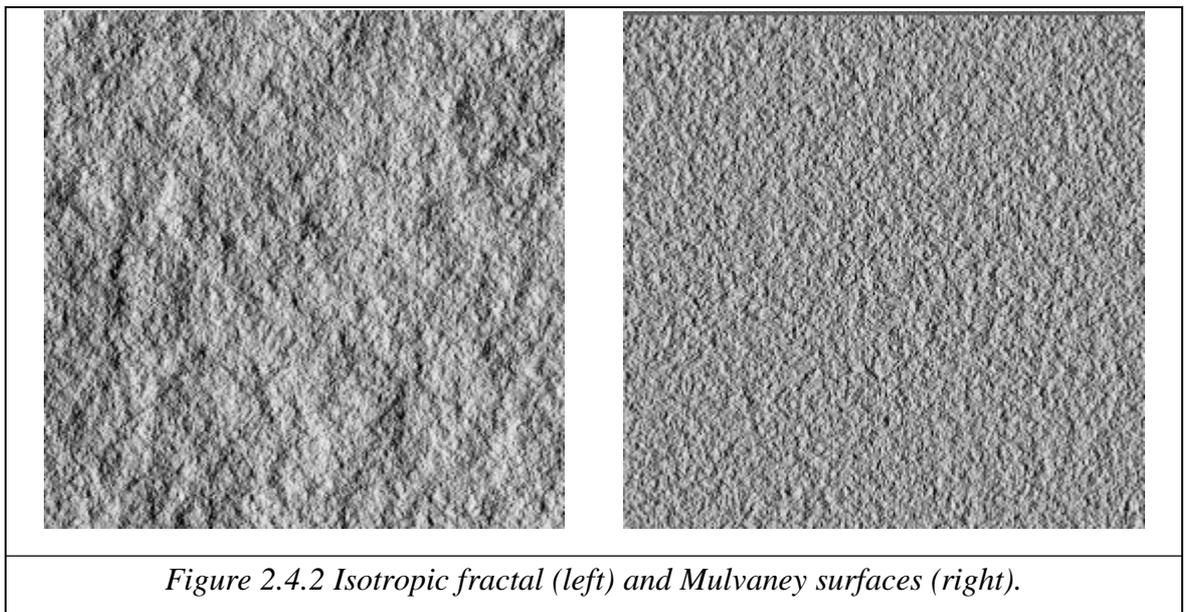
$$\sigma_s^2 = \frac{\pi k_t \omega_c}{2} \quad (2.4.1c)$$

$$S_{1d}(\omega) = \left(\frac{2}{\lambda_0 \sqrt{2\pi}}\right) \frac{1}{\left(\frac{1}{\lambda_0}\right)^2 + \omega^2} \quad (2.4.1d)$$

where λ_0 is the correlation length, i.e. the distance from the origin it takes the ACF to fall to e^{-1} of its value at the origin. The fractal and Mulvaney power spectra is shown in *Figure 2.4.1*.



The illuminated fractal and Mulvaney surfaces are shown in *Figure 2.4.2*, it is interesting to note that, while both are realistic, the Mulvaney surface resembles a rough surface which has undergone a degree of physical processing.



Both the fractal and Mulvaney models are stated in terms of profile spectra, using equation 2.2.5d we can predict the spectra of isotropic surfaces conforming to these models. Substituting the fractal model into equation 2.2.5d leads to:

$$s_{2d}(\omega) = \frac{-1}{\pi} \int_{\gamma=\omega}^{\gamma=\infty} \frac{d}{d\gamma} \left(\frac{k_t}{\gamma^2} \right) \frac{1}{(\gamma^2 - \omega^2)^{\frac{1}{2}}} d\gamma$$

$$s_{2d}(\omega) = \frac{-1}{\pi} \int_{\gamma=\omega}^{\gamma=\infty} \frac{k_t}{\gamma^3 (\gamma^2 - \omega^2)^{\frac{1}{2}}} d\gamma$$

$$s_{2d}(\omega) = \frac{\pi k_t}{2\omega^3}$$

A fractal surface profile with roll-off $\beta_{1d}=2.0$ corresponds to a fractal surface with power roll-off $\beta_{2d}=3.0$. We may repeat the process for the Mulvaney profile spectrum :

let $\alpha = \omega_c^{-2}$

and $s_{1d}(\omega) = \frac{k_t}{\alpha\gamma^2 + 1}$

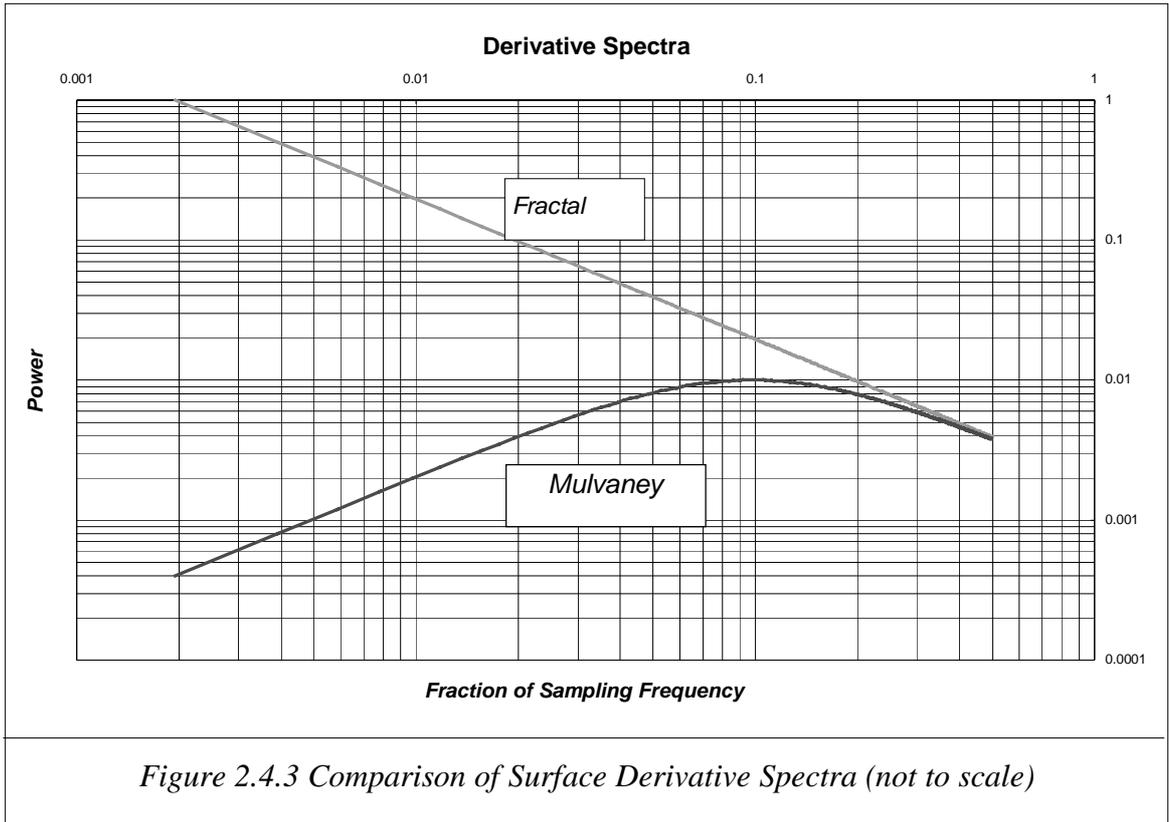
$$s_{2d}(\omega) = \int_{\gamma=\omega}^{\gamma=\infty} \frac{k_t}{\gamma^2 (\alpha\gamma^2 + 1)^2 (\gamma^2 - \omega^2)^{\frac{1}{2}}} d\gamma$$

$$s_{2d}(\omega) = \frac{\pi k_t}{4\sqrt{\alpha}} \frac{\omega\sqrt{\alpha + \omega^{-2}}}{(\alpha\omega^2 + 1)^2}$$

$$s_{2d}(\omega) = \frac{\pi}{4\sqrt{\alpha}} (\alpha\omega^2 + 1)^{-\frac{3}{2}}$$

For ω small, this corresponds to a white noise spectrum, and for ω large to fractal roll-off with $\beta=3.0$.

Since this thesis is concerned with the surface image, which is related to the surface facet slopes, we differentiate the two surface models (in any direction) and the plot the power in *Figure 2.4.3*. Both models have a $1/\omega$ characteristic at high frequencies. This does not converge to zero as frequency tends to infinity. Consequently, without bandlimiting the slope variance will tend to infinity. This serves to underline the dependency of the m_{rms} parameter on sampling frequency. Since the model is used in discrete form we assume the surface is bandlimited above the Nyquist frequency.



At low frequencies, the power of the derivatives of the Mylvaney model varies quadratically with frequency and therefore tends to zero as frequency approaches zero. The amount of derivative power present at wavelengths greater than the sampling window size will depend on the frequency of the breakpoint. For the fractal model the $1/\omega$ relationship still holds at low frequencies. The derivatives are therefore non-stationary, with unstable statistical properties.

The $1/\omega$ characteristic is both very common in practice and difficult to accommodate in signal theory. However Keshner states that for a $1/\omega$ process, "If the time over which the process is observed is short compared with the time elapsed since the process began then the exact ACF can be approximated." He goes on to say, "the PSD (of a ω^{-1} process) is stationary except for the steady state value which depends logarithmically on the time elapsed since the process was started" [Keshner82]. For ease of analysis, we will assume that the surface spectra are bandlimited for frequencies less than the fundamental.

2.4.2 Modelling Directionality

In this section we will examine the most common model of directionality and consider some frequently made assumptions. We will also survey the literature to find what degree of directionality occurs in real surfaces. Finally, we will consider the question of whether fractal dimension is rotation invariant for anisotropic surfaces.

We begin by considering the Longuet-Higgins (LH) description [Longuet-Higgins57]. This was developed to describe the sea surface and forms the foundations of statistical geometry. Nayak [Nayak71] was the first to apply the LH approach to engineering surfaces and therefore provided the motivation for the use of random process techniques on surfaces. We have neglected the LH description, since it is first order i.e. does not incorporate any spatial information and is ultimately used to develop a number of parameters, such as summit density, which are of no direct relevance to this work. However, the first stage of the LH model does provide a useful basis for discussion of surface directionality.

If we assume that the direction of the surface grain is aligned with one of the axes, then we may use the matrix described by Bush et al. [Bush79].

$$\begin{matrix} m_{00} & 0 & 0 & -m_{20} & -m_{02} \\ 0 & m_{20} & 0 & 0 & 0 \\ 0 & 0 & m_{02} & 0 & 0 \\ -m_{20} & 0 & 0 & m_{40} & 0 \\ -m_{02} & 0 & 0 & 0 & m_{04} \end{matrix} \quad (2.4.2a)$$

where

$$m_{fg} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u^f v^g S(u, v) du, dv \quad (2.4.2b)$$

From equation (2.4.2b) we can see that m_{00} is independent of direction. We do however note that the one dimensional profiles in the Hasegawa study do show considerable variation in surface variance [Hasegawa93]. We presume this is due to limited data length and/or detrending removing significant low-shifted frequencies.

From (2.4.2a) we infer that surfaces are assumed to contain one dominant directionality. While this is probably the case for the majority of surfaces, we note that at least one of our real, manmade, test surfaces exhibit two well resolved directionalities.

In the previous section we noted that most surface roughness models have a power law roll-off over at least part of their spectrum. We now consider whether fractal dimension, and therefore the rate of roll-off, is rotation invariant for anisotropic surfaces. Hall et al. develop the theorem that, “*the fractal dimensions along the transects of a stationary stochastic surface are all identical, save that in one special direction.*” He further states that “*the majority of processed surfaces we have examined appear to have identical fractal dimensions, even though they are markedly anisotropic in other characteristics*” [Hall95]. In his study of an anisotropic thin-film rigid disk, Majumdar and Tien found that the x spectrum consisted of two regions with $\beta=-1.94$ and -1.42 , the y spectrum had $\beta=-2.51$, though the low frequency spectrum oscillated and did not follow any power-law. In the most comprehensive study of this area, Hasegawa et al. measured fractal dimension for several sample types at various orientations [Hasegawa93]. The roll-offs are shown in *Table 2.4.1*, where Hasegawa divided the spectrum into two regions; we have denoted these regions by (a) and (b). The results shown do not appear to give a definitive answer to the question of the rotation dependency of fractal dimension.

Angle	0°	30°	60°	90°	120°
<i>Lapped Surface</i>	1.56	1.54	1.56	1.69	1.69
<i>Electric Discharge</i> (a)	1.57	1.87	1.58	1.56	1.57
(b)	1.66	1.80	1.86	1.65	1.74
<i>Grinding1</i> (a)	1.54	1.55	1.55	1.52	1.59
(b)	1.77	1.79	1.73	1.78	1.73
<i>Grinding2</i> (a)	1.67	1.72	1.68	1.73	1.70
(b)	1.68	1.65	1.64	1.72	1.70

Table 2.4.1 Estimated Fractal dimensions in various directions, (from Hasegawa93)

In fact we find this model inadequate in one respect; it does not allow for a smooth transition between isotropic and highly directional textures. Setting the correlation lengths in the orthogonal directions to equal values results in a cross-like power spectrum, *Figure 2.4.5b*, rather than an isotropic spectrum. However, this model is well suited to extremely anisotropic textures and we shall use this model to produce a series of exemplar textures.

2.4.3 Assumptions of the Models

We make several assumptions in connection with the models we have adopted. We assume that a surface can be reconstructed from its gradient field i.e. the gradient field must be integrable; if a surface s is a potential field to the gradient field \mathbf{S} then \mathbf{S} is conservative, i.e. an integral of the gradient field around any path is equal to zero, Eq 2.4.3.

$$\text{Curl } \mathbf{S} = 0 \quad (2.4.3)$$

As a consequence of (2.4.3), the first and second partial derivatives of s are continuous and the mixed derivatives equal [Thomas p.1065].

A stricter assumption is that the surface is bandlimited. The low pass component of this assumption allows a discrete form of the surface model to be used, and in consequence the surface model can be differentiated an infinite number of times. A further consequence of this is that the *rms slope* of a surface will be finite. The high pass component of the assumption implies that the surface is stationary. It follows that the surface can be completely predicted from a realisation of the phase spectrum and the model without recourse to initial conditions. A further consequence is that the surface statistics are stable. Within the bandlimited region we assume the surface spectrum conforms to either the fractal, Mulvaney or directional Ogilvy models.

We assume the surface phase spectrum satisfies the maximum entropy condition, i.e. there is no discriminatory information held in the phase spectra and classification must be carried out purely on the power spectra. A consequence of this assumption and the central limit theorem is that the height distribution, and consequently the slope distribution will be Gaussian, allowing a more tractable analysis in subsequent chapters. We make the further assumption in this thesis that the slope distribution will have a standard deviation not exceeding 0.5.

2.4.4 Summary of the Models

We have identified three complementary models:

1. a fractal model,
2. the Mulvaney model, and
3. the Ogilvy model.

Surfaces (1) and (2) are isotropic and differ in the low frequencies. At high frequencies, the Mulvaney model exhibits behaviour identical to the fractal model, i.e. a power law roll-off, however, at low frequencies the Mulvaney spectrum is white. By varying with direction the frequency at which the transition between the two behaviours occur, Ogilvy produces our third model. The fractal and Ogilvy surfaces will be used extensively in this thesis as exemplars, though the Mulvaney model will be used to a lesser extent.

2.5 Summary of Surface Description and Modelling

The topography of a rough surface is modelled as the scalar field $s(x,y)$ which acts as a potential field to the (conservative) vector field $S(x,y)$. The probability function associated with each of the scalar fields of $S(x,y)$ as well as the height field itself are assumed to be Gaussian. The standard deviations of these distributions, p_{rms} and q_{rms} form useful surface parameters and will be used extensively in this thesis.

The surface power spectra $S(u,v)$ is the principle means of surface description—it is assumed to conform to one of three prototypical forms:

- (1) a fractal form, this follows an inverse power law throughout its range,
- (2) the Mulvaney form, exhibiting roll-off only above a cut-off frequency, and
- (3) the Ogilvy form, which varies the frequency at which cut-off occurs with orientation to introduce surface directionality.

The phase spectrum of the surface is assumed to contribute no discriminatory information, i.e. surfaces with the same surface model and model parameters, but different phase spectra, will be considered to be of the same type. In fact, this is only a reasonable assumption for a limited subset of textures. Intuitively, this subset is characterised by the absence of step changes and a lack of structure in the data set.

2.5.1 Relevance of this Chapter to subsequent chapters

This chapter describes the first stage in the classification process. In consequence, many of the assumptions and decisions made in this chapter have consequences throughout the rest of the thesis. In this sub-section we show the relevance of these decisions to later chapters.

Throughout this thesis we model the topography of a rough surface with a discrete function $s(x,y)$ with an associated derivative vector field $S(x,y)$. The distribution of heights and gradients is assumed to be Gaussian. In Chapters 3, 4 and 6 we model imaging and classification as a linear process, consequently the assumption of normality is maintained throughout these sections. The slope distributions are parameterised by the rms slope parameters (m_{rms}, p_{rms} and q_{rms}). Unlike the rms roughness, the slope parameters are directional, and consequently provide a simple indication as to the directionality of the surface. In Chapter 3 we will see that the adequacy of the linearity assumption for the rendering process is a function of slope. Slope parameters therefore represent an important surface characteristic and are used in Chapters 3, 7 and 8. As a consequence of starting from a height, rather than a derivative, model we assume the derivative fields are conservative and integrable. This is relevant in Chapters 7 and 8 where we consider reconstruction of the surface from derivative estimates.

The scalar field $s(x,y)$ is described using the power spectrum, $S(u,v)$. Use of the PSD allows integration of the surface model with the illumination, noise and filtering models of Chapters 3,4 and 6 respectively. The power spectrum is assumed to be bandlimited. The low pass component of this assumption allows a discrete form of the surface model to be used, consequently the *rms slope* of a surface will be finite. The high pass component implies that the surface is stationary and the surface statistics are stable.

We assume the surface phase is random, i.e. there is no discriminatory information held in the phase spectra, and classification must be carried out solely on the power spectra. This will affect the classifier design in Chapter 5. This assumption in conjunction with the central limit theorem supports the assumption that the height distribution, and consequently the slope distribution will be Gaussian.

This chapter has also justified the adoption of three surface models defined in terms of the power spectrum:

- (1.) an isotropic and fractal model—this obeys a power law roll-off within the observed bandwidth.
- (2.) Mulvaney's model is also isotropic and can be seen as a generalisation of the fractal case, with a flat response below a cut-off frequency.
- (3.) Ogilvy's model, varies the frequency at which the transition between white noise and fractal roll-off occurs with angle to produce a directional surface.

The spectral characteristics of these models has influenced the design of the classifier developed in Chapter 5. The fractal and Ogilvy models are used throughout this thesis as test cases for isotropic and anistropic surfaces respectively. The requirement for a series of isotropic surfaces in Chapter 6, is partially fulfilled using the Mulvaney model.

	3 Image Formation	4. Imaging	5. A classifier	6. The Effect of tilt on the classifier	7 Compensation Schemes	8 A Model - based scheme
s(x,y)						
normality						
m _{rms}						
curl \underline{S} =0						
PSD						
Phase						
Fractal						
Ogilvy						
Mulvaney						

Figure 2.5.1 Assumptions developed in this chapter and their use in later chapters. (Shading indicates use of assumption in chapter)

2.6 Conclusions

The aim of this chapter was to define a system of surface modelling, whereby different realisations of the same model, with identical parameters, could be classified as belonging to the same class of surface on the basis of their visual appearance. After considering several fields it was decided to consult the related fields of tribology and scattering in order to find an established methodology of surface description and modelling. Several levels of description were discussed, before a second order technique, the PSD, was adopted for the description of rough surfaces.

The description is in terms of the power spectrum. However, surfaces may have the same PSD but radically different visual appearance due to the influence of their phase spectra. We circumvent this problem by adopting a maximum entropy condition for the textures' phase spectra: a texture's phase spectrum may have any realisation so long as it doesn't carry any information. A simple test was developed to assess whether this condition holds for a given texture. Only textures which fulfilled the criterion will be considered in this report.

The models considered are defined in terms of the description adopted: the power spectrum. Two characteristics of the spectrum were considered: the roll-off of power with increasing frequency and directionality. All the models display a power law roll-off of power with frequency for at least part of their frequency range. While the fractal model shows this characteristic throughout the frequency range, Mulvaney and Ogilvy models have a limiting process at low frequencies where the spectrum is approximately white. The Ogilvy model, models directionality by varying the frequency at which the transition from white to fractal behaviour occurs with direction.