Stochastic Analysis of Surface Roughness
A Fractal-based Approach

Steven John Davies

July 1998

A thesis submitted for the degree of Doctor of Philosophy of
The Australian National University.

Declaration

Unless otherwise specified in the text, this thesis describes my own work, supervised by Professor P. G. Hall.

S. J. Davies

Acknowledgements

To my wife, Robyn, for her much needed support and encouragement.

To my parents, John and Jacqueline, who made many sacrifices so that I would receive the best possible education.

To my supervisors and mentors, Peter Hall, Nick Fisher & Graham Constantine. They provided the problem, guided me through my attempts at tackling it, and carefully scrutinised the end result.

To my friend, Dan Lunn, for his constant badgerings. I hope now that we can change the subject.

To my friends and colleagues at C.S.I.R.O. and the A.N.U. who withstood my interruptions, listened and gave help.

Abstract

The need to characterise surface roughness arises in many branches of science and engineering, in areas as diverse as soil science, polymer chemistry and manufacturing. The ability to capture information about surface properties with a few statistical indicators can provide the basis for control and monitoring of industrial processes, and the indicators themselves may be important covariates in scientific investigations.

In the past, the extent of an analysis of roughness has been limited largely to calculation of some simple roughness parameters. Although these parameters are appealing by virtue of their ease of calculation, they do not perform well in terms of characterising surfaces and, from a statistical point of view, there are many problems associated with their use.

The two main aims of this thesis are to provide better parameters for characterising surface roughness, and to provide statistical tools to aid scientists and engineers in making inferences. In developing the methodology, special care has been taken to make the methods as practicable as possible.

The main parameter chosen to characterise surface roughness is fractal dimension. It has the appealing property that, when using it to judge whether one surface is rougher than another, the decision will, in many circumstances, agree with the subjective ordering made by a trained observer. This is a property the parameters in current use do not enjoy.

Fractal dimension is itself limited in terms of its ability to characterise surface roughness. It is possible for two surfaces to have similar fractal dimensions, yet one may appear much rougher than the other. This is usually due to the difference in amplitude between the peaks and troughs of two surfaces being large. Fractal

dimension, being a scale-invariant measure, does not take account of such a difference. So a parameter called topothesy, which is scale-sensitive, is introduced to complement fractal dimension.

Many methods have been proposed for estimating fractal dimension from data, but little attention has been given to understanding how the estimators perform. This thesis includes theoretical and numerical analyses of the performance of estimators of fractal dimension and topothesy in a number of different situations.

The analyses provide a guide to how well estimators will perform in general, but do not quantify the precision of an estimate calculated from a given data set. So, this thesis also proposes methods to calculate standard errors of fractal dimension and topothesy, thereby providing an objective way of making comparisons between surfaces. Methods for carrying out hypothesis tests are also discussed.

Roughness characterisation methods are also developed for two-dimensional data. In the past, most of the data sets used to gauge surface roughness were one-dimensional profiles, collected along a straight path across a surface. As a result of advances in both the technology for measurement equipment and the storage capacity of computers, genuine two-dimensional surface data sets are now quite common. This raises the important question of anisotropy: how do the roughness properties of the surface vary with orientation? This question is thoroughly explored and appropriate methods devised.

Computational efficiency has been a prime consideration in practical aspects of this study.

Contents

Abstract iv

1 Introduction 1

1.1 Motivation 2

1.2 Examples of applications 3

1.3 Visualising the data 8

1.4 Statistical modelling 10

2 Characterising surface roughness 15

2.1 Review of existing methods 17

2.2 What is roughness? 19

2.3 Fractal methods 21

2.3.1 Fractal dimension and fractal index 21

2.3.2 Point estimation 23

2.3.3 Topothesy 28

2.4 Comparison of existing and fractal methods 30

3 The variogram 34

3.1 Definition 36

3.2 Estimation 37

3.3 Advantages of variogram over autocovariance 41

3.4 Exploratory analysis 42

3.5 Properties 49

3.5.1 One-dimensional case 50

3.5.2 Two-dimensional case 53

4 One-dimensional transects **57**

4.1 Point estimation 59

4.1.1 Theoretical performance of estimators 59

4.1.2 Numerical issues 61

4.1.3 Choice of k 63

4.1.4 Possible improvements 64

4.2 Quantifying the error 65

4.2.1 Plug-in method 66

4.2.2 Bootstrap method 67

4.3 Fitting a variogram model 69

4.3.1 Model validation 73

4.4 Effects of measurement error 78

4.4.1 Effects of smoothing the variogram 79

4.4.2 Effects of zeroing the smoothed variogram 80

4.5 Analysis of the roller data 83

5 Isotropic surfaces **87**

5.1 Extension from one dimension 88

5.2 Box counting 96

5.3 Application to soil surfaces 100

6 Anisotropic surfaces **104**

6.1 General anisotropy 105

6.2 Weak anisotropy 111

6.3 A test for isotropy 113

6.3.1 Comparing fractal dimension between surfaces 115

6.4 Analysis of the polymer surfaces 116

Bibliography **121**

Chapter 1

Introduction

Fractal analysis of surface roughness is more than a mathematical curiosity. It has many applications in science and engineering, where it is used to aid understanding of the physical processes occurring at the interface between objects or between an object and its environment. In this chapter we describe examples of such applications from three different disciplines. The examples are elaborated throughout the thesis, providing the motivation for the methods developed and illustrations of how the methodology can be applied.

As in most areas of statistical practice, visualisation has an important role to play. We look at conventional ways of visualising surfaces in three dimensions: contour plots, wireframe perspective plots and height encoded images. These are arguably good devices for discerning the features of relatively smooth surfaces, but they lack clarity and intepretability when used to visualise rough surfaces. One natural way to view such data is to artificially reconstruct its appearance to the human eye. This is achieved by computer-generated renderings of the surfaces.

A thorough statistical analysis must necessarily be based on assumptions set out clearly in mathematical terms. We describe the stochastic framework that we shall use to model the surface data. We also detail other major assumptions made about the data used throughout the thesis, and provide physical and statistical justification for them.

1.1 Motivation

The scientific interest in surface roughness is extensive, as can be judged from the vast literature on the subject. As well as journals solely devoted to the topic, such as *The Journal of Tribology*, there are many articles in a variety of fields.

A large portion of the literature is concerned with physical effects of surface roughness, usually in the setting of particular applications. For example, Thomas & Atkinson (1997) consider the 'Ammonium uptake by coral reef – effects of water velocity and surface roughness on mass transfer'; and Amis (1996) addresses 'The effect of surface roughness in fibroblast adhesion in vitro'. Some of the many physical properties studied are absorption, magnetisation, adhesion, impedance, flow, growth, friction, conductivity, reflectance and heat transfer.

Another portion of the literature concentrates on the effects of processes, both natural and manufactured, on surface roughness. Two recent examples are McCarrol & Nesje (1996), who study the use of 'Rock surface roughness as an indicator of degree of rock surface weathering'; and Hassan (1997), who studies 'The effects of ball- and roller-burnishing on the surface roughness and hardness of non-ferrous metals'.

The remainder of the literature is concerned with the measurement and characterisation of surface roughness. Measurement is usually achieved through contact methods such as stylus profilometry, or non-contact methods using optics or even acoustics (Swart *et al.*, 1996). An overview of the various characterisation methods is given in Chapter 2.

The development of statistical methods for analysing surfaces has typically been driven by the technologies used to record data. A relatively old but still rather common approach to measurement is stylus profilometry, in which a fine stylus is drawn across a surface and the current generated by its oscillations is taken as a measure of height. The total electrical charge resulting from these oscillations is proportional to the mean absolute deviation of surface height, not the mean squared deviation. Therefore, L_1 measures of variation can be more attractive than L_2 measures, quite apart from any statistical advantages that either might have. More

recent technologies allow extensive and detailed surface height data to be recorded, and so permit more sophisticated, flexible approaches to data analysis. The new advances include refinements to the performance of stylus profilometers, and optical profilometry based, for example, on scanning with a laser, or with white light from an optical fibre.

However, all profilometer data are recorded from what are essentially line transects of the surface (albeit with a non-infinitesimal width representing the diameter of the stylus or light beam) and so are still rather restrictive. New technologies such as scanning electron microscopes allow genuinely two-dimensional data to be gathered, typically in the form of digital images. Such data offer exciting opportunities for addressing issues of anisotropy and spatial variation in a way that is difficult even for extensive profilometer data. This thesis is motivated by these possibilities. We suggest new methods for analysing surface data, taking advantage of the new technologies and allowing questions of spatial variability to be addressed.

1.2 Examples of applications

The following examples have motivated much of the work within the thesis and are used throughout to illustrate how the methodology could be applied.

Roller profile In the manufacture of rolled products such as sheet metal and paper, the surface roughness of the roller is crucial. If the roller is too smooth, it may slip or skid, causing tears in the product. On the other hand, if the roller is too rough this adversely affects the quality of the rolled product, for example by causing perforations. Therefore it is important for the surface roughness of a roller to lie within predetermined control limits. In order to prevent the manufacturing faults mentioned, the rollers are periodically inspected for wear measured in terms of adverse changes to the surface roughness.

The data we have are from a polished metal roller used for rolling sheet metal. They were obtained using a standard commercial stylus profilometer, and consist of 1150 equally spaced heights (above a datum level) along a 4.5mm section of the

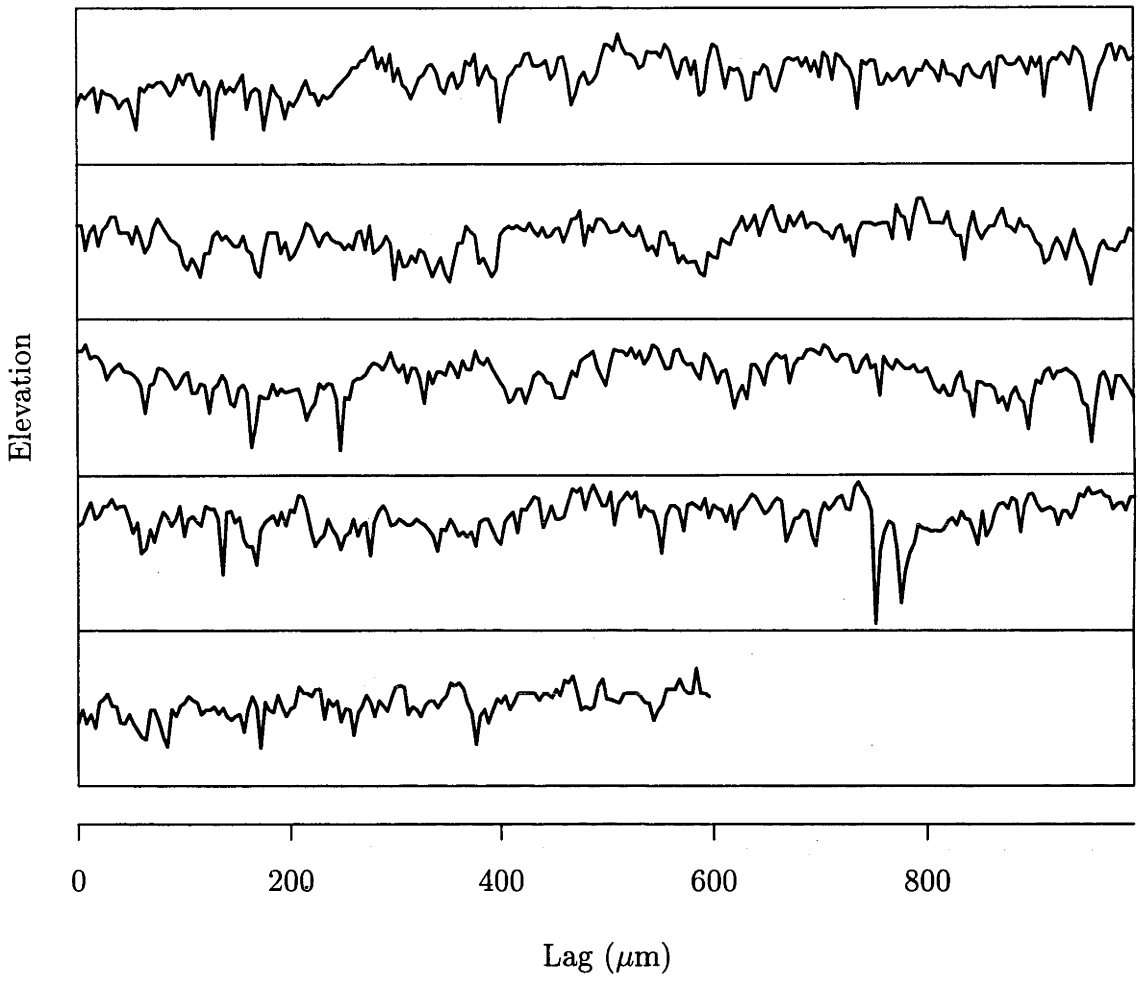
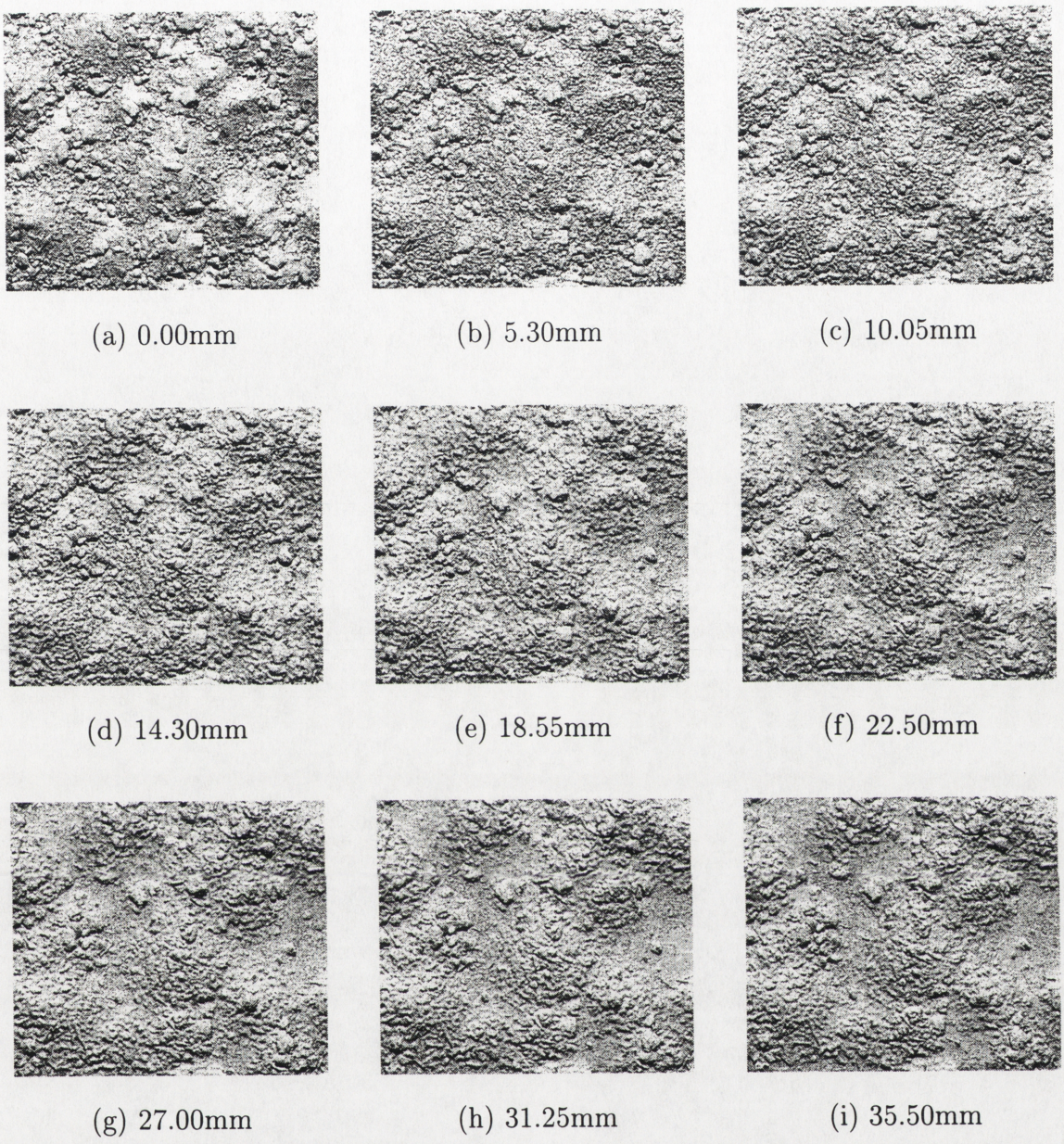


Figure 1.1: The roller profile data broken up into 5 segments with consecutive segments plotted down the page. The scale at the bottom applies both horizontally and vertically.

roller. The stylus had a nominal tip radius of 4μm. The electrical signal giving the profile height, is recorded continuously but is digitised at discrete intervals, approximately 4μm each, by the instrument used to produce the data.

The data are shown in figure 1.1 and a roughness analysis is performed in Chapter 4.

Soil surfaces In Soil Science, considerable attention is being placed on understanding the interaction between land management and environmental conditions, as an aid to developing sustainable and productive agricultural systems. To do this,



(a) 0.00mm (b) 5.30mm (c) 10.05mm

(d) 14.30mm (e) 18.55mm (f) 22.50mm

(g) 27.00mm (h) 31.25mm (i) 35.50mm

Figure 1.2: Rendered images of the soil surface, dry and after eight successive periods of rainfall. The cumulative amount of rainfall is indicated in the caption for each panel.

scientists must understand many of the physical processes that take place, such as infiltration of water, runoff of water, erosion by water and wind, gas exchange, evaporation and heat flux. All of these processes occur at the interface between the atmosphere and soil, namely at the soil surface; and the roughness of the soil surface

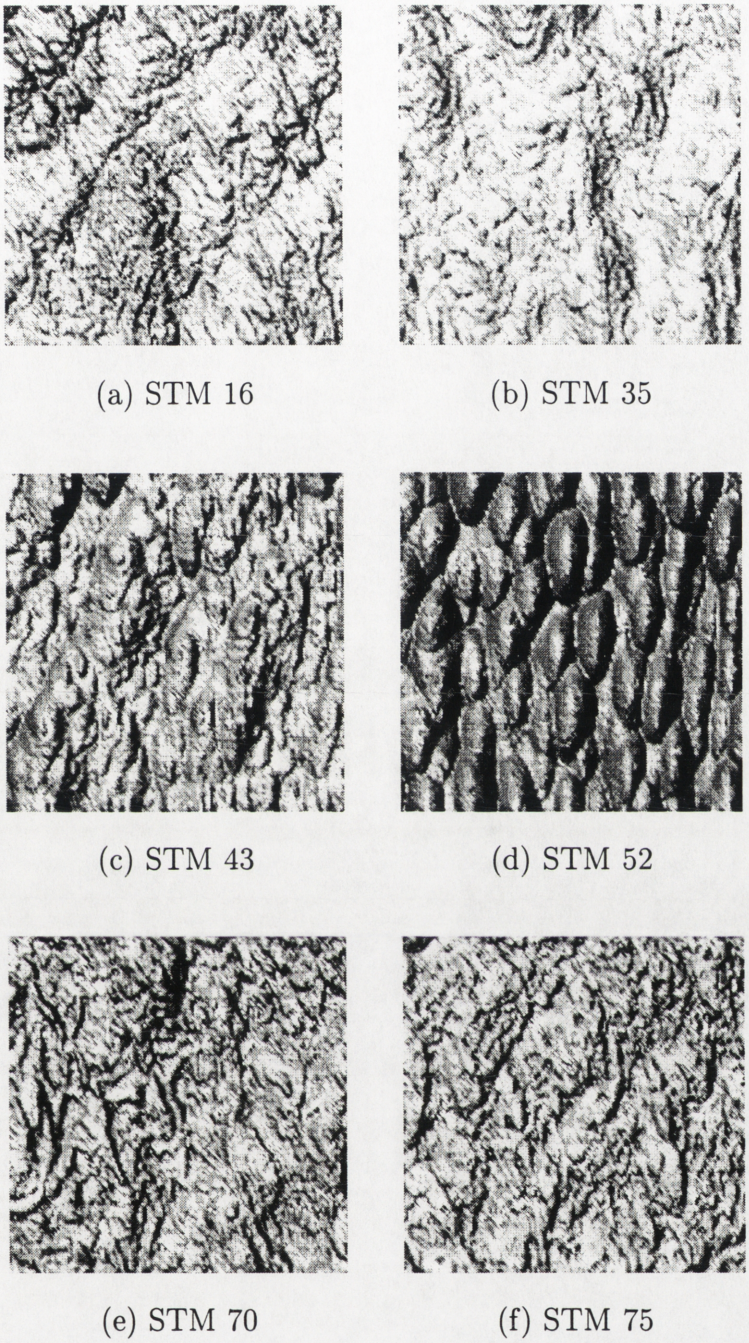
is an important factor in determining the effects of such processes. Contributing to the complexity of the problem is the dynamic nature of the surface as it varies with time due to these and other processes, for example rain and cultivation.

The data we have come from a particular study to assess the availability of water for plant growth after rain. Rainfall itself changes the surface topography through the impact of raindrops and the transportation of particles with the movement of water over the surface. Therefore it is necessary to understand the effect of rainfall on the soil surface as well as the effect of surface topography on the behaviour of water. The water that infiltrates the surface becomes available for plant growth; the remainder either evaporates or runs off. To infiltrate the surface the water needs to be stationary, collected in depressions on the surface, or travelling at low velocity. The surface roughness contributes to the number and size of such depressions and the speed at which water can flow.

A single surface measuring 600mm $\times$ 500mm was constructed in the laboratory with dry soil, a sandy loam sampled from Cowra, in South-Eastern Australia. Under the soil was a layer of free-draining coarse sand. The central 512mm $\times$ 450mm area of the surface was scanned by a laser device. Height averages over circular regions approximately 0.5mm in diameter were taken at 1mm intervals along parallel lines 1mm apart. The soil surface was then subjected to simulated rainfall consisting of 2.7mm drops falling vertically 13 meters and achieving 97 percent terminal velocity. The rainfall intensity was 90 mm per hour, which would correspond to a heavy storm in the region from which the soil was sampled. Rain was allowed to fall until ponding occurred. (Ponding is the phenomenon of free water appearing on the surface due to rainfall exceeding infiltration.) The surface was then scanned again. The process was repeated 8 times in all, with a night between each rainfall, yielding a time series of 9 images of the soil surface.

The soil surface data are rendered in figure 1.2 and a roughness analysis is given in Chapter 5.

Polymer surfaces The aim in the third application was to identify whether two alternative manufacturing processes could produce polymers with similar surface



(a) STM 16 (b) STM 35

(c) STM 43 (d) STM 52

(e) STM 70 (f) STM 75

Figure 1.3: Rendered images of the six polymer surfaces.

roughness properties.

The ultimate use for the polymers was to be as a plastic wrap for food, an application where roughness properties are important in promoting the longevity of the food. Decay is hastened by the presence of microorganisms, and smoother

wraps offer less opportunity for such organisms to adhere. Another contribution of statistical analysis is to compare surfaces in terms of smoothness.

The data set comprises three pairs of samples of the polymers, one sample from each pair produced by one manufacturing process and the other sample by an alternative process. Both samples in a pair were produced with similar input parameters to the process. The input parameters were different between pairs.

The surface elevations of each polymer sample were measured by coating it with a layer of gold just a few molecules thick, and analysing the coated plastic using a scanning, tunnelling electron microscope. The data from this source were recorded on a 128×128 grid, in the form of measurements that were height averages very close to grid point centres. For these data, pixels were 40 nanometres square.

The polymer surfaces data are rendered in panels of figure 1.3 and a roughness analysis is given in Chapter 6.

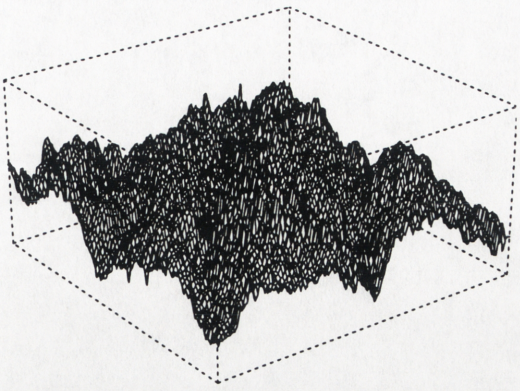
1.3 Visualising the data

Visualisation is an important tool in many statistical applications. For the surface scientist, visual characterisation plays an important and sometimes crucial role in topography analysis. Stout *et al.* (1993, page 197) state that "it is generally accepted that a full intuitive appreciation of a surface can only be achieved by 3-D visualisation methods and that 2-D profiles are inadequate for qualitative assessment."

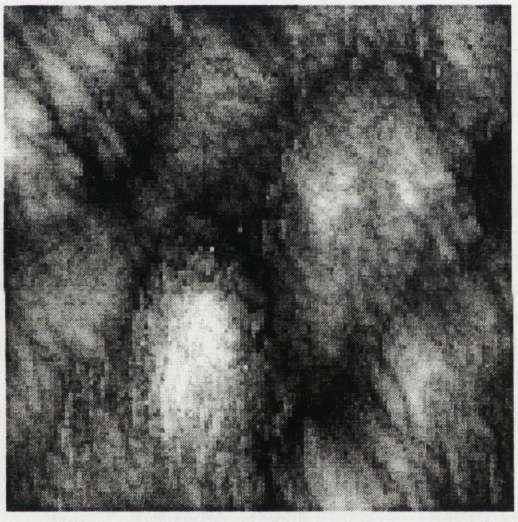
Profile data Nevertheless, if the only data available are profiles, then it is still of value to get an appreciation of the nature of the data by a simple graph. When analysing the surface roughness of profiles using a simple scatterplot of heights against position, features to which one might pay particular attention are the irregularity of the sample path and the scale of its oscillations. To do this consistently, equal scalings should be applied to both the heights and the positions.

Figure 1.1 is a scatterplot of the heights of the metal roller data. The heights have been joined together by straight line segments to help the viewer discern the

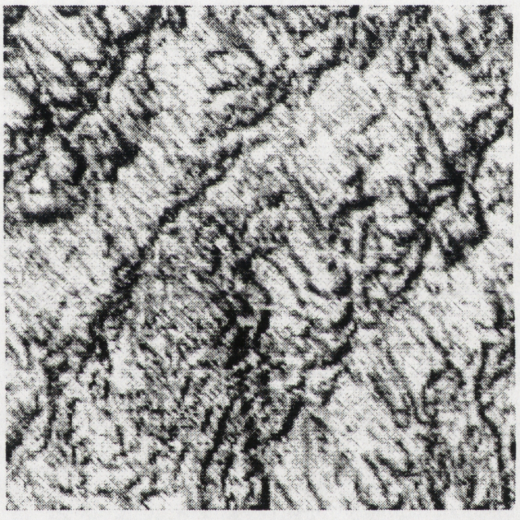
(a) contour plot



(b) perspective plot



(c) height encoded image



(d) rendered image

Figure 1.4: Four different visualisations of the first polymer surface data

sequence of points. The original surface probably followed a far more irregular path between successive points.

Gridded data There are several common ways of viewing the 3-D surfaces in 2 dimensions, including contour plots, perspective/wireframe plots and colour/greyscale height coded images. These are arguably good graphical devices for looking at smooth surfaces, such as purely mathematical functions, but for rough surfaces the

graphs end up a mess of ink in which it is hard to discern features.

One relatively recent way of viewing raw surface height data is to construct an image of the surface similar to how it would appear to the human eye. The eye remains the most powerful and versatile processor, particularly when complicated and ill-defined but patterned data are involved. A "rendered" image is created by calculating how much light from a particular source would be reflected to the viewer in a particular location. Figure 1.4 is a comparison between a contour plot, a perspective plot, an image plot and a computer rendered scene of the surface. Notice, in panel (d), that we can see evidence of directional dependence in the texture, something which may not be seen in the other graphs except by a trained observer.

The sequence of rendered images for the soil surfaces is shown in Figure 1.2. With careful inspection, it is possible to see how the surface is changing over time with successive amounts of rainfall. The higher areas are gradually being eroded, whereas deposits are being formed in the lower areas. This is easier to see if the individual 'snapshots' are animated to form a short movie sequence and played in time, allowing the viewer to switch between scenes both backwards and forwards.

Figure 1.3 shows the surfaces of the six plastic sheets. The fourth sample appears smoother and more regular than the others, although its fractal analysis was not markedly atypical; in fact its fractal properties are on a much finer scale than the "bubbles".

1.4 Statistical modelling

Throughout the thesis, we think of surface data as realisations from a single-valued two-dimensional random field $X(\boldsymbol{t})$ in the case of surface data, and from a one-dimensional random process $X(r)$ in the case of profile data. Thus, the data sets may be considered as functions from one- or two-dimensional grids to the set of real numbers $x : \mathcal{G} \to \mathbb{R}$ — $\mathcal{G}$ is either a one-dimensional equally spaced grid of points on the real line for profile data or a two-dimensional equally spaced grid of points

on the real plane for surface data. Therefore the data from a single surface/profile, $x(\cdot)$, is really one datum from an underlying random field $X(\cdot)$.

Since it is hard to carry out any statistical analysis from a single datum, we need to make some assumptions concerning the properties of these processes in order to obtain an effective increase in the sample size. In doing so, we should also ensure that any assumptions made are reasonable, both statistically and physically. By 'reasonable' we mean that they need not necessarily be true, but that they are close enough to the truth that any departure will have a negligible effect on the assumption-based methods.

The first assumption we make is that of *intrinsic stationarity*, which is defined for a random field (process) through first differences:

$$\mathrm{E}\left[X(\boldsymbol{s}+\boldsymbol{t}) - X(\boldsymbol{s})\right] = 0,$$
$$\mathrm{var}\left[X(\boldsymbol{s}+\boldsymbol{t}) - X(\boldsymbol{s})\right] = v(\boldsymbol{t}).$$

That is, the random field has constant expectation, and the variance of the difference of the random field at two locations depends solely on the relative displacement between the two locations and not on their positions. The function $v(\boldsymbol{t})$ is known as the *variogram* and is studied in detail in chapter 3.

An earlier form of intrinsic stationarity is the *intrinsic hypothesis* of Matheron (1971). The intrinsic hypothesis differed in that it allowed for a linear drift in the process $X(\cdot)$.

The first part of the assumption, namely, that the random field has constant expectation, is not as unreasonable as might first appear considering the nature of the data. In the case of the practical examples outlined in section 1.2, the data are often taken on a much finer scale than the whole surface, so any change in the trend over the whole surface would be slight over the range of the data. The second part of the assumption requires that the process that produced the surface, again at the scale on which the data were collected, behaves in a uniform manner.

Another assumption that might reasonably be made in some circumstances, but not always, is that the process that determined the surface topography does not favour any particular direction. This might be a valid assumption for the soil surface

data, where the soil surfaces were artificially constructed, but not for the polymer data, where the polymer is manufactured by an extrusion process: the polymer is drawn out in one direction. More formally, this assumption is that of *isotropy* and is a tightening of the intrinsic stationarity assumption where we now require that the variance of the difference of the random field at two locations depend solely on the distance, rather than displacement, between the two locations, and not on their relative orientation:

$$\text{var}\,[X(\boldsymbol{s}+\boldsymbol{t}) - X(\boldsymbol{s})] = v(\|\boldsymbol{t}\|).$$

The function $v(\|\boldsymbol{t}\|)$ is now one-dimensional, as it would be if the random field were itself one-dimensional. Accordingly we call the distance $\|\boldsymbol{t}\|$ the *lag*, to follow one-dimensional time series terminology. Methods for characterising surface roughness from two-dimensional data assuming isotropy, and from one-dimensional profile data, are considered and applied to the soil surface data in Chapter 5.

If a random field is not isotropic then we say it is *anisotropic*. Methods for deducing whether a surface is either isotropic or anisotropic, and methods for characterising the surface roughness for anisotropic surfaces, are given and applied to the polymer data in Chapter 6.

When deriving performance properties, it is sometimes necessary to assume that the surfaces may be modelled by Gaussian processes. Indeed, Thomas (1982, page 9) writes:

> Many formative processes, particularly those carried out in controlled conditions in research laboratories, can produce textures with height distributions that are accurately Gaussian. Their Gaussian nature is not an artifact: it arises as the natural result of a well-known statistical property, and will occur whenever a texture is created as the cumulative result of a large number of randomly located events.

However, the methods developed are also applicable to a wide range of non-Gaussian surfaces: see for example Hall & Roy (1994), who show that similar performance results may be obtained for suitably well-behaved functions of Gaussian surfaces.

In order to compare the asymptotic properties of one-dimensional and two-dimensional estimators, the properties will be expressed in terms of overall sample size, N, the number of points in the domain of $x(\cdot)$. Lower case n will be used for the number of unique points in the one-dimensional projections of the domain of $x(\cdot)$. So for one-dimensional data, $N = n$, and for two-dimensional data, $N = n^2$.

Self-similar and self-affine Self-similarity is a common assumption made about profile data in order to model the irregular behaviour of the data.

There are a number of conflicting definitions of self-similarity, due to their different origins in mathematics and statistics. Here, we employ the definition from Taqqu (1988) for a one-dimensional process: a process $X(t)$ is *self-similar* with parameter H if $X(at)$ and $a^H X(t)$ have identical finite-dimensional distributions for all scalar values of $a > 0$.

For an intrinsically stationary process X, self-similarity implies that $v(t) = c|t|^\alpha$ where $\alpha = 2H$. To see this, observe that

$$\begin{aligned} v(as) &= \text{var}[X(as + at) - X(at)] \\ &= \text{var}\{a^H[X(s + t) - X(t)]\} \\ &= a^{2H} v(s). \end{aligned}$$

To obtain the desired form for the variogram, substitute $a = |s|^{-1}$ and $c = v(1)$.

Self-similarity is sometimes taken to be the case where $c = 1$, and the term *self-affine* used for the more general case when c is allowed to vary. In that case, a self-similar process may be specified by the single parameter, H.

In principle, the definition of self-similarity of a one-dimensional process can be applied in two dimensions, by replacing the one-dimensional argument t by a two-dimensional vector $\boldsymbol{t}$. It should be noted that this is more restrictive than the condition that every one-dimensional transect of $X(\boldsymbol{t})$ be self-similar. The former restricts H to a single value, independent of orientation of $\boldsymbol{t}$, whereas the latter allows H to vary with orientation.

Fractional Brownian motion If $X(t)$ is intrinsically stationary, self-similar and *Gaussian*, then it can be modelled by a scaled *fractional Brownian motion*, $\lambda Z(T+t)$, where

$$Z(0) = 0, \quad \mathrm{E}\, Z(t) = 0, \quad \text{and} \quad \mathrm{E}\, Z(t)^2 = |t|^{2H}.$$

Modelling profiles using fractional Brownian motion makes an implicit assumption of self-similarity. From experience, we believe that in many cases such an assumption is unwarranted. As Thomas (1982) warns,

> The theoretician and the instrument designer start with some postulated mathematical model and are developing numerical descriptions of surface texture and theories of surface interaction. Both approaches are invaluable. But there is an obvious danger that they may diverge; in particular, theories may be developed which contain mathematically attractive assumptions not generally valid for real surfaces, and parameters may become commonly measured and specified merely because their determination is instrumentally convenient.

As $H \to 1$, or equivalently as $\alpha \to 2$, the dependence between increments over large lags becomes so great that the fractional Brownian motion tends to a straight line. None of the data sets illustrated in section 1.2 seems to exhibit this attribute.

It is sometimes thought that the assumption of self-similarity is necessary for developing fractal methods. This is in fact not the case, and has unfortunately led to some earlier methods that lack consistency when the assumption is not valid. In practical terms, this can be almost all of the time. None of the methods developed in this thesis relies on the assumption of self-similarity.

Chapter 2

Characterising surface roughness

We explore existing methods for the characterisation of surface roughness, paying particular attention to their limitations. Most of these methods are for one-dimensional data and do not extend readily to higher dimensions. Some of the methods concentrate on a related attribute such as the area about a trend line, and fail to distinguish between surfaces with different roughness characteristics but similar area.

We distinguish two qualitatively different components of roughness: erraticism *and* scale. *Most of the existing methods implicitly combine these two components, thereby losing important information. At this point we introduce fractal dimension and show how, as a scale-independent measure, it quantifies the first component, that of erraticism.*

We model the behaviour of the variogram in the vicinity of the origin as an approximation to a power law. This law can be defined directly in terms of fractal dimension, but we define it in terms of an intermediate quantity, fractal index. *Fractal index is useful in that, like fractal dimension, it measures erraticism, but it has the added advantage of being independent of the dimension of the data. We derive the simple linear relation between fractal index and fractal dimension that allows fractal dimension to be estimated.*

Associated with this power law is a scale factor. This serves as a measure of the scale component of roughness. To be consistent with similar terminology from Sayles & Thomas (1978b), this scale factor will be referred to as topothesy.

Finally, we describe a way of comparing the "overall" roughness between two surfaces. This yields a partial ordering of surfaces in which either (a) one of two surfaces is rougher than the other, or (b) the two surfaces are incomparable. The method is based on the dominance of one variogram over another, and incorporates comparisons of both fractal index and topothesy.

2.1 Review of existing methods

There are numerous measures extant for quantifying surface roughness from profile data. Thomas (1982, pages 86–87) provides a table of definitions and origins for 23 common measures. This is by no means an exhaustive list.

Although most of the existing roughness measures have been developed for one-dimensional data, some measures exist for two-dimensional gridded data; see for example Stout *et al.* (1993). These include extensions of the one-dimensional parameters, where this is possible, as well as measures specifically designed for two dimensions.

Many measures are application–specific, being defined in terms of how the surface may be expected to behave functionally in a given environment. One such example is the 'bearing length', or 'bearing area', which seeks to quantify the proportion of the surface that would support a flat object being forced upon it, by gravity for instance.

Some parameters are easily recognisable as statistical moment estimators, *e.g.* skewness and kurtosis. However, the usual sampling assumption underlying the analysis of these estimators, that of independence, has been ignored, severely affecting the validity of such approaches in the presence of correlation.

Nearly all measures of roughness can be shown to be measures of scale, in that a rescaling of the data is reflected as a similar rescaling to the measures. Thus, for two surfaces similar in all other respects, these measures provide an adequate means of differentiating the surfaces by size.

Amongst the most commonly used roughness measures (possibly due to their appearance in measurement standards) are: average roughness, root-mean-square roughness, and ten-point height. Define the *mean level* as $\bar{x} = n^{-1} \sum x_i$. Then the *average roughness* is calculated as,

$$R_a = \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}|,$$

the *root-mean-square roughness* by,

$$R_q = \left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{\frac{1}{2}},$$

and if $x_{[1]} \leq \cdots \leq x_{[n]}$ denote the order statistics of the x_i's, the *ten-point height* is the quantity,

$$R_z = \frac{1}{5}\sum_{i=1}^{5}(x_{[n+1-i]} - x_{[i]}),$$

for $n \geq 10$.

Caution must be used when using any of these measures to compare surfaces. Indeed, many practitioners advise that to use these measures to compare surfaces, the surfaces should be recorded over the same interval at the same sampling rate by the same machine. Two of these imperatives relate to the sources of error.

Firstly, there is the accuracy of the measurement equipment used to obtain the profile data from the surface. Different machines that employ different measurement techniques, such as stylus and optical profilometers, cannot measure the exact elevation of a surface at an exact point. Instead, they perform height averages in the vicinity of the point. In addition, each type of machine gives a different type of average.

The second source of error, which is due to increasing the sample size by increasing the range over which the profile is measured, can be attributed to bias. If, for the moment, we assume that X has constant finite variance and a covariance function $\gamma(\cdot)$, then the expected value of R_q^2 is

$$\mathrm{E}(R_q^2) \;=\; \gamma(0) - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma(i-j) \;=\; \left(1-\frac{1}{n}\right)\gamma(0) - \frac{2}{n}\sum_{i=1}^{n-1}\left(1-\frac{i}{n}\right)\gamma(i)$$

(*cf.* 3.3). For finite samples the bias term can be quite considerable, especially if $\gamma(\cdot)$ decays slowly as would be the case for a smooth surface. Therefore, profiles measured over different ranges but at the same sampling rate may have considerably different biases that would dominate any comparison. When the assumption of a

constant finite variance for X is dropped, the problem is exacerbated. In this case, the expected R_q^2 is a weighted average of the variogram over the range observed:

$$\mathrm{E}(R_q^2) = \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} v(i-j) = \frac{1}{n}\sum_{i=1}^{n-1}\left(1-\frac{i}{n}\right)v(i). \tag{2.1}$$

So, depending on the form of the variogram, samples taken over wider ranges are measuring different quantities, and are therefore incomparable.

These problems are common to most existing scale–dependent roughness measures, including the other two quantities defined above, R_a and R_z.

To date, there do not seem to be any methods for gauging the accuracy of point estimates in any of the existing methods. We shall address this shortcoming in later chapters.

2.2 What is roughness?

Despite the plethora of roughness measures, there does not seem to be a generally agreed definition, either mathematically or in natural language, of what roughness is. As we saw in the previous section, it has generally been the case that a particular algorithm has been applied to data to obtain a measure, and the algorithm itself has been used as a definition of roughness, *e.g.* root-mean-square roughness.

There has been little discussion of what the result of these algorithms actually measures. Rather, there is usually an implicit assumption that the value achieved will be different for different data sets, and a hope that the implied ordering of data sets, from smooth to rough, will somehow agree with a person's perception of roughness.

This last property, of a measure ordering surfaces according to roughness in agreement with an experienced observer, is an appealing one. It would allow the observer to gain an appreciation of a surface without necessarily having to see it. It would also allow an analysis that is usually subjective to be more objective. This is particularly advantageous when there are many surfaces to analyse and an automatic method of analysis is sought. Such applications arise in process monitoring

in manufacturing, where it is necessary to regularly inspect component surfaces for wear. Sadly, the existing roughness measures do not have these properties and have proven unreliable in this application.

Erraticism One of the appealing properties of fractal dimension as a quantifier of roughness is that it captures the degree of erratic behaviour of sample paths. Adler (1981, Chapter 8) coined the term *erraticism* for this property when applied to the sample path of a stochastic process. For one-dimensional processes, it refers to the irregularity of oscillations as opposed to the scale of oscillations. This idea of the degree of irregularity or randomness of sample paths ties in well with human perceptions of roughness.

However, a measure of erraticism alone does not fully capture the whole concept of roughness. Imagine a profile that took the form of a sine wave. Its path is very regular and smooth, but if we were to dramatically increase the frequency and amplitude of the sine wave, then the path might appear rough in some sense.

Although it may be argued that increasing the frequency somehow increases the erraticism, if we were to look at the path on a smaller scale, *i.e.* by looking at it under magnification, then we may revert to our original characterisation of the sample as smooth.

Perhaps, then, *scale-independence* might be a desirable property for a measure of erraticism to possess. True scale-independence can only be achieved if the stochastic process is self-similar. As we discussed in Section 1.4, this property does not seem to apply in many practical situations. Therefore, we shall require that measures of erraticism be *vertically scale-independent* functionals, in the sense that any vertical rescaling of the process leaves the measure of erraticism unaffected.

To date, fractal dimension and closely related methods appear to be the only such quantifiers of erraticism. None of the traditional roughness measures manages to treat erraticism at all adequately.

Scale Nearly all existing measures of roughness provide a measure of scale, in that they respond monotonically to a change in scale of the data, either horizontally or

vertically. If the data are magnified, then these measures increase.

Indeed, most traditional measures of roughness are linear functionals, for which increases in the data are reflected by the same increases in the measure. Since this aids comparison of such measurement between surfaces, we note this as a desirable property for measures of scale.

If two surfaces have similar erraticism, a measure of scale would seem appropriate to differentiate the two surfaces in terms of roughness, the surface that varies more being the rougher. In this way, measures of erraticism and scale complement each other.

2.3 Fractal methods

Undoubtedly, the name most commonly associated with the development of fractals and fractal dimension, and their application in the physical sciences, is that of Benoit Mandelbrot. The interested reader may consult Mandelbrot (1975, 1977, 1982, 1985) and Mandelbrot *et al.* (1984) for an appreciation of early works. Since then, there has been considerable investigation of the mathematical theory of fractals (Barnsley 1988 and Tricot 1993), and of the practical applicability of fractals in a wide variety of areas ranging from urban and landscape planning (*e.g.* Milne, 1991a,b) to oceanography and meteorology (*e.g.* Jain, 1986; Morrison & Srokosz, 1993), defence science (*e.g.* Lo *et al.*, 1993), and the study of musical scores (*e.g.* Hsü & Hsü, 1990; Lewin, 1991).

In the remainder of this thesis we shall confine our attention to fractal methods as they apply to the problem of characterising surface roughness.

2.3.1 Fractal dimension and fractal index

Fractal dimension There is a variety of different definitions of fractal dimension, including box-counting or Minkowski-Bouligand dimension (*e.g.* Barnsley 1988, Chapter 5; Taylor & Taylor 1991, Chapter 5), and Hausdorff dimension (*e.g.* Adler, 1981, Chapter 8). In general, a suitably defined deterministic curve or surface can

have different fractal dimensions, depending on the definition used, or have well-defined dimension according to one approach but not another. However, for the Gaussian-based models that we have in mind, all the common definitions are applicable, and all produce the same numerical value for dimension.

The practical estimation of fractal dimension is typically based on estimation of a quantity that is sometimes called *fractal index*, that describes the way in which the variogram of the underlying stochastic process varies near the origin. (The term 'fractal index' is sometimes used to describe the rate of decay of the covariance function for arbitrarily large lags, rather than small lags. The duality between behaviour at infinity and the origin in the case of perfectly self-similar processes has caused the terminology to be used for behaviour at both extremes.) Quite apart from theoretical considerations, practical limitations often restrict estimation to moment properties of stochastic processes, so it is particularly helpful to be able to characterise relatively complex properties of sample paths in terms of properties of low-order moments, in particular of covariance. Thus, many estimators of fractal dimension are based on an estimate of fractal index, and use a simple formula relating the two to produce an estimate of dimension. Estimators based on the variogram and periodogram are of this type.

Fractal index The *fractal index* of an intrinsically stationary process with variogram $v(\cdot)$ may be defined as the common value, α, of the real numbers

$$\begin{aligned} a &= \sup\{\xi > 0 : v(\boldsymbol{t}) = O(\|\boldsymbol{t}\|^{\xi}) \text{ as } \boldsymbol{t} \to \boldsymbol{0}\} \\ b &= \inf\{\xi > 0 : \|\boldsymbol{t}\|^{\xi} = O[v(\boldsymbol{t})] \text{ as } \boldsymbol{t} \to \boldsymbol{0}\}, \end{aligned}$$

provided they do have a common value. In practice, it is usual to assume a relatively simple model for the variogram, that ensures equality of a and b. For example, in the case of isotropy it might be supposed that

$$v(\boldsymbol{t}) = c\|\boldsymbol{t}\|^{\alpha} + o(\|\boldsymbol{t}\|^{\alpha}) \tag{2.2}$$

as $\boldsymbol{t} \to \boldsymbol{0}$. The value of α always lies between 0 and 2, and (for a Gaussian process or field) equals 2 if X is differentiable.

Since the fractal index is so pervasive in this thesis, we reserve the Greek letter α to symbolise it, and the accented $\hat{\alpha}$ to denote an estimator for it.

Relation between dimension and index If the fractal index α is well-defined, and if the process X is sufficiently closely related to (see below) a Gaussian field, then with probability 1 the realizations of X have fractal dimension

$$D = d + 1 - \tfrac{1}{2}\alpha; \tag{2.3}$$

see for example Adler (1981, Chapter 8), Sayles & Thomas (1978b) and Thomas & Thomas (1988). Note particularly that $d \leq D < d + 1$. Processes X for which D is larger have rougher–looking realizations, at least on a sufficiently fine scale.

With regard to the phrase 'sufficiently closely related to', Hall & Roy (1994) showed that for (2.3) to hold it is not necessary that X be Gaussian. Indeed for (2.3) to hold, it is sufficient that X be a smooth function of a sequence of Gaussian random fields, with the fields having well-defined fractal dimension and the function satisfying a suitable moment condition. This implies that variogram-based methods for fractal dimension estimation are not restricted by a Gaussian assumption, but are appropriate to a much wider class of random field models.

IMPORTANT NOTE: Because of the negative relationship between fractal dimension and fractal index in (2.3), it is imperative that the distinction between the two be emphasised. Fractal dimension increases with an increase in the erraticism of a curve, and is therefore a measure of increasing *roughness*. In contrast, fractal index decreases with a similar increase in erraticism, and so is a measure of increasing *smoothness*. Throughout this thesis, it will be necessary to switch attention from fractal dimension to fractal index and back again.

2.3.2 Point estimation

Over the past 20 years, increasing interest in fractal dimension has led to the invention of many dimension estimators. Some estimators are more general in nature

and apply to a wide range of data types, *e.g.* to random sample paths within a plane, such as the coastline of Britain. Because of their general nature, these methods often perform more poorly than specific methods tailored to special data types. Nevertheless, they do provide a means of dimension estimation in cases where there might not otherwise be one.

In the present problem, we choose to impose structure on the sample paths, specifically, that they are continuous functions. Consequently, we shall confine attention to those dimension estimators specifically designed for such data sets. We now give descriptions of three of the most common approaches to dimension estimation for continuous random functions.

These methods are described in the one-dimensional setting, since most of the work on the estimators is for one-dimensional data, the most common case. It is often beneficial to gain an understanding in one dimension before progressing to two or more dimensions. The two-dimensional setting will be discussed in detail in Chapters 5 and 6.

Variogram method

This method relies on the one-dimensional version of (2.2):

$$v(t) = c|t|^{\alpha} + o(|t|^{\alpha}) \qquad \text{as } t \to 0. \tag{2.4}$$

Once this assumption has been made, $\log v(t)$ should be an approximately linear function of $\log |t|$, for small t:

$$\log v(t) = \log c + \alpha|t| + o(1). \tag{2.5}$$

So, given an estimator $\hat{v}(\cdot)$ for the variogram, we may construct a plug-in estimator of α as the slope of the linear regression of $\log \hat{v}(l/n)$ on $x_l = \log l$ for the first k values of l, *i.e.*

$$\hat{\alpha} = \left\{ \sum_{l=1}^{k} (x_l - \bar{x}) \log \hat{v}(l/n) \right\} \Big/ \left\{ \sum_{l=1}^{k} (x_l - \bar{x})^2 \right\}, \tag{2.6}$$

where $\bar{x} = k^{-1} \sum x_l$. Then using (2.3), fractal dimension is estimated by $\hat{D} = 2 - \hat{\alpha}/2$.

Practical considerations, such as choice of k and the statistical properties of $\hat{\alpha}$, and therefore of $\hat{D}$, will be discussed in Chapter 4.

Box counting

Because it is simple to understand and easy to implement, box counting is a popular method for dimension estimation. The basic box counting method involves partitioning the plane into equal–sized squares, or boxes. The box covering is taken to be the set of boxes that intercept the curve. Fractal dimension is calculated from the limit of the ratio of either the logarithm of the box–covering area, or the number of boxes in the covering, to the box width, as box width tends to zero. See Barnsley (1988) for details.

In practical terms, the smallest box width is determined at least partly by the fineness of discretisation of the data, making it difficult to calculate fractal dimension from its formal definition. Consequently, the slope of a log–log regression, similar to that employed in the variogram method, is used. However, Hall & Wood (1993) show that this method is prone to bias of the order of $(\log n)^{-1}$. Instead they provide an alternative log–log regression method. Firstly, to imitate collecting the data at coarser scales, the data are sub-sampled at regular intervals: the greater the interval the coarser the scale. The box–covering area is then calculated for each sub-sample, and the log of the area is regressed against the log of the width of the sub-sampling interval.

Following their notation, define

$$\mathcal{B}(i,l) = \{(i-1)lm/n, [(i-1)m+1]l/n, \ldots, ilm/n\} \qquad (1 \le i \le q_l, 1 \le l \le k),$$

where q_l denotes the integer part of $(n-1)/lm$. Here, l denotes the level of discretisation, m the width of a block, and $\mathcal{B}(i,l)$ the i^{th} block of indices. The approximate area of the box–covering for the i^{th} block is

$$A_{il} = \epsilon_l(U_{il} - L_{il}),$$

where

$$\epsilon = lm/n, \quad U_{il} = \max_{j \in \mathcal{B}(i,l)} X(j/n), \quad \text{and} \quad L_{il} = \min_{j \in \mathcal{B}(i,l)} X(j/n).$$

The individual box covering areas are summed over blocks to obtain the total box–covering area,

$$A(l) = \sum_i A_{il} = \epsilon_l \sum_i (U_{il} - L_{il}).$$

An estimate, $\hat{D}$, of fractal dimension may be obtained from the slope of a log–log regression,

$$2 - \hat{D} = \left\{ \sum_{l=1}^{k} (x_l - \bar{x}) \log A(l) \right\} \Big/ \left\{ \sum_{l=1}^{k} (x_l - \bar{x})^2 \right\}, \tag{2.7}$$

where $x_l = \log l$ and $\bar{x} = k^{-1} \sum_{l=1}^{k} x_l$.

The similarity between this box counting method and the variogram method above is not restricted to the use of log–log regression. When $X(\cdot)$ is Gaussian and m is taken as 1, the two methods coincide.

In this case, U_{il} and L_{il} are, respectively, the maximum and minimum of just two values, $X[(i-1)l/n]$ and $X(il/n)$. Therefore,

$$(U_{il} - L_{il}) = |X(il/n) - X[(i-1)l/n]|$$

is the absolute difference of $X(\cdot)$ over a lag of l/n. For a Gaussian process this absolute difference has an expected value of $2\pi^{-1/2} v (l/n)^{1/2}$.

Thus, the expected box–covering area is

$$\begin{aligned} \mathrm{E}\{A(l)\} &= \epsilon_l \sum_{i=1}^{q_l} \mathrm{E}\,|X(il/n) - X[(i-1)l/n]| \\ &= (l/n) q_l (2/\pi)^{1/2} v (l/n)^{1/2} \\ &= (2/\pi)^{1/2} v (l/n)^{1/2}. \end{aligned}$$

If we use $(\pi/2)A(l)^2$ as the estimate of the variogram in (2.6) then we get (2.7) exactly.

In this instance, the equivalence of the two methods allows us to compare the two methods. The implicit variogram estimate in the box–counting method is the

average of just n/l summands chosen from a possible $n - l$. All other estimators, including the empirical variogram introduced in the next chapter, use $n - l$. Consequently, these estimators perform better, although their rates of convergence are of a similar order.

The fact that the box counting method is a special case of the variogram method, which performs more poorly by a constant factor, explains why we choose to concentrate on variogram–based methods rather than box–counting methods in this thesis.

Spectral methods

Fractal dimension can also be estimated using spectral methods based on the periodogram (see Dubuc *et al.*, 1989). These methods make use of the relation between fractal dimension and fractal index (2.3) and assume second-order moment properties similar to (2.4). Although spectral methods are popular for estimating fractal dimension, Constantine & Hall (1994) note,

> Consistent estimation of the spectral density, $f(\cdot)$, requires statistical smoothing, and a second level of smoothing is needed to estimate α (and hence D) from $f(\cdot)$. This complicates both the theoretical and practical sides of the procedure. By way of contrast, $v(\cdot)$ may be estimated directly and unbiasedly from observations of the stochastic process X on a grid; smoothing is not required at this level.

Nevertheless, Chan *et al.* (1995) propose a method, based on a continuous version of the periodogram, which avoids the complications in analysis introduced by statistical smoothing as noted above. Their method follows.

For convenience and without loss of generality, the domain on which X is observed is taken to be $(-1, 1)$. Define

$$A(\omega) = \int_{-1}^{1} X(t)\cos(\omega t)dt, \qquad B(\omega) = \int_{-1}^{1} X(t)\sin(\omega t)dt$$

$$I(\omega) = A(\omega)^2 + B(\omega)^2, \qquad \text{and} \qquad J(\omega) = A(\omega)^2.$$

Here, $I(\omega)$ is the full continuum periodogram and $J(\omega)$ is a semiperiodogram.

Despite their similar construction, $\mathrm{E}\{A(\omega)^2\}$ and $\mathrm{E}\{B(\omega)^2\}$ have quite different asymptotic behaviour:

$$\mathrm{E}\{A(\omega)^2\} \sim C_\alpha\, \omega^{-(\alpha+1)}$$

whereas

$$\mathrm{E}\{B(\omega)^2\} \sim \begin{cases} C_\alpha\, \omega^{-(\alpha+1)}, & \text{if } 0 < \alpha < 1, \\ C_1\, \omega^{-2}, & \text{if } \alpha = 1, \\ C_2\, \omega^{-2}, & \text{if } 1 < \alpha < 2, \end{cases}$$

as $\omega \to \infty$ through integer multiples of 2π, where C_1 and C_2 are constants and C_α is nonzero and depends on α. From this it can be seen that, for $\alpha > 1$, α can not be estimated directly from $I(\omega)$. Therefore estimation of α, and hence D, is based solely on the semiperiodogram $J(\omega)$, and is again achieved by performing a log-log linear regression, *viz.*

$$\hat{\alpha} = -\left\{\sum_{j=1}^{k}(x_j - \bar{x})\, \log \hat{A}(\omega_j)^2\right\} \Big/ \left\{\sum_{j=1}^{k}(x_j - \bar{x})^2\right\} - 1,$$

where $w_j = 2\pi j$, $x_j = \log \omega_j$, $\bar{x} = k^{-1}\sum x_j$ and $\hat{A}(\omega)$ is an estimator for $A(\omega)$. Then D is estimated by $\hat{D} = 2 - \hat{\alpha}/2$, as for the variogram method.

We do not intend to explore spectral methods any further in this thesis. Instead we have chosen to concentrate on characterising surface roughness using parameters, including fractal dimension, derived from the variogram.

2.3.3 Topothesy

Sayles & Thomas (1978b) coined the term *topothesy* for a parameter that describes the scale of oscillations of a surface. They also stated that this parameter "uniquely defines the statistical geometry of the random components of an isotropic surface". However, Berry & Hannay (1978) showed that in reaching this conclusion Sayles and Thomas had effectively taken the fractal index to be 1 for each surface under

study. This value was obtained from data pooled from all of the example surfaces under investigation. Individual estimates for the fractal index for each surface were calculated in Sayles & Thomas (1978a), and indeed a wide range of values, covering almost the whole possible range, was obtained. Topothesy may not define the statistical geometry uniquely; however, it does provide a measure of the scale of surface oscillations complementary to the measure of erraticism offered by fractal dimension.

There seems to be no agreed formal definition of topothesy; note the different definitions in Sayles & Thomas (1978b), Berry (1979), and Thomas & Thomas (1988). In particular, there exist non-equivalent definitions in the frequency and spatial domains. Therefore we propose an alternative, more convenient, definition based on the multiplicative factor c in (2.2). Since topothesy is to be a measure of scale, it is not unreasonable to require that the topothesy of $\lambda X(t)$ be λ times the topothesy of $X(t)$, *i.e.* a linear functional of $X(t)$.

An expression that fits this requirement is simply $c^{1/2}$, where c is the constant of proportionality in (2.2). Note that Davies & Hall (1999) define topothesy as c. The definition of $c^{1/2}$ is preferred here for its linearity property.

We can estimate c as the intercept of the linear regression of $\log v(l/n)$ on x_l:

$$\hat{c} = \exp\left\{ k^{-1} \sum_{l=1}^{k} \log \hat{v}(l/n) - \hat{\alpha}\bar{x} \right\}. \tag{2.8}$$

Then the estimated topothesy is just $\hat{c}^{1/2}$.

For anisotropic two-dimensional data, c will be a function of orientation rather than a constant. This will be examined in detail in chapter 6.

Unlike fractal index, topothesy has physical dimensions. A simple dimensional analysis using (2.2) yields physical dimensions in metres (m) for topothesy of $\mathrm{m}^{1-\alpha/2}$, or equivalently m^{D-d} — that is, metres to the power of the difference between fractal dimension and Euclidean dimension. In practice, it is probably easier to interpret if the unit for length is the same as the original unit of measurement, *e.g.* $(\mu\mathrm{m})^{D-d}$.

Bearing this in mind, the two roughness parameters, fractal dimension and topothesy, could be combined in a simple expression to characterise the roughness

of a profile: for example, 'the roller has a fractal roughness of 54.6 $(\mu\text{m})^{0.255}$.'

2.4 Comparison of existing and fractal methods

A numerical study was carried out to show how fractal methods perform in comparison with existing methods.

The three traditional roughness parameters R_a, R_q and R_z defined in Section 2.1 and the two fractal estimators $\hat{D}$ and $\hat{c}$ defined using the variogram method were calculated for a collection of simulated surface profiles. In calculating $\hat{D}$ and $\hat{c}$, the one-dimensional empirical variogram (*cf.* 3.2) was used as an estimator of $v(\cdot)$, and k was taken as 2.

Each simulated profile was a realisation of a stationary Gaussian random process with the stable exponential covariance model

$$\gamma(t) = \gamma(0)\ \exp(-\lambda|t|^{\alpha}),$$

for its covariance function. Here, $\gamma(0)$, the variance of the process, was set equal to the value 0.04 for all of the simulations. The simulations were performed using the algorithms in chapter 7.

Seven sets of 1000 profiles were simulated, each with different values for the pair (λ, α). The α's took the values 0.25(0.25)1.75. The corresponding values for λ were calculated so that the range of dependence within each process was approximately half the range of the data. To achieve this, λ was calculated so that $\gamma(0.5)/\gamma(0) = e^{-3} \approx 0.05$, which implies that $\lambda = 3 \times 2^{\alpha}$.

Typical profiles for each pair of parameter values (λ, α) are graphed in Figure 2.1.

Figure 2.2 contains 5 panels of boxplots, one panel for each of the roughness parameters discussed so far. Within each panel there are seven boxplots corresponding to the seven values of α. Each boxplot depicts the distribution of measures of 1000 simulated profiles.

Of the three traditional roughness parameters, the ten-point height R_z provides the best discrimination between the simulated profiles. However, its superiority is

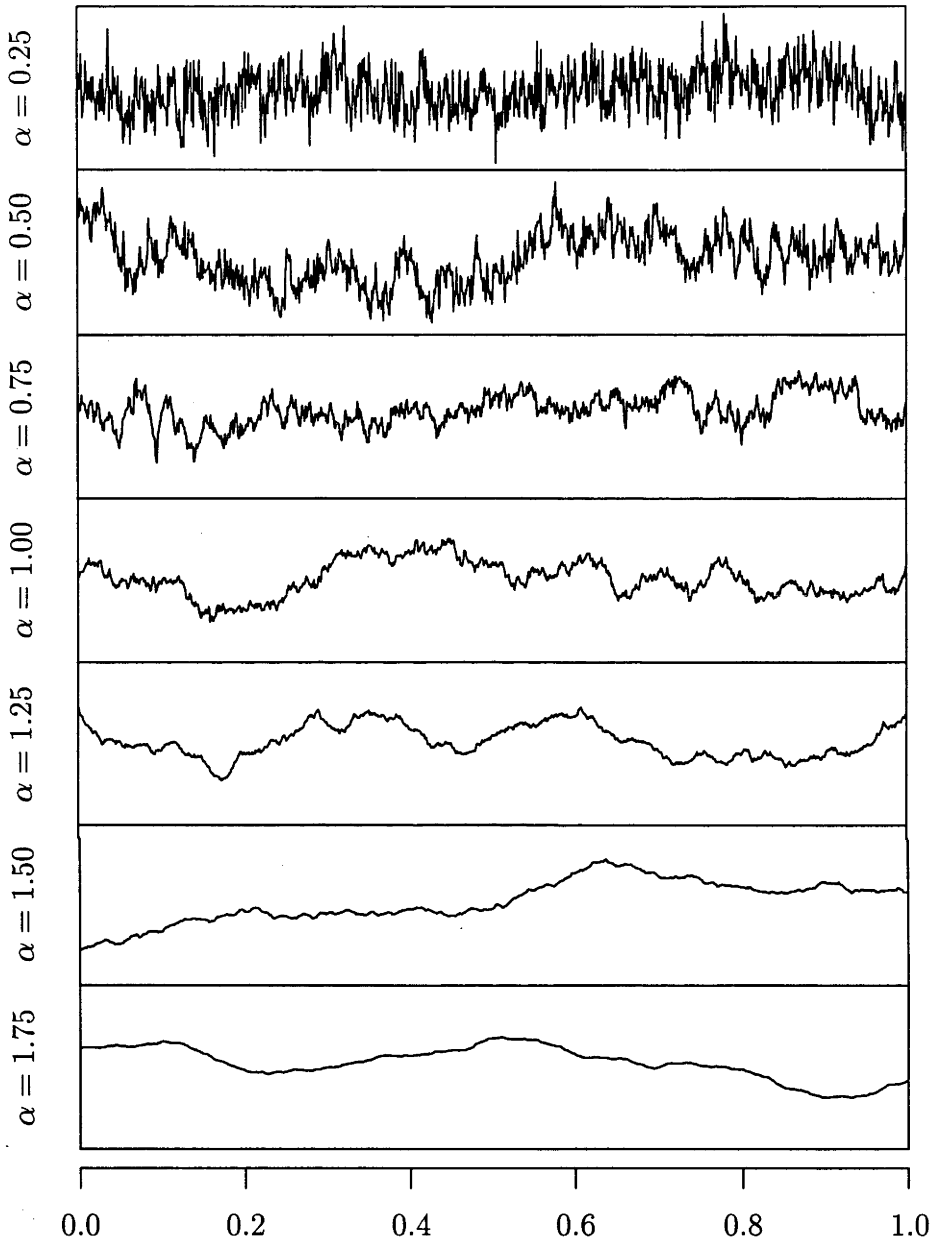


Figure 2.1: Simulated profile data sets. Each profile is a realisation of a Gaussian random process with a stable exponential covariance model with fractal index α. The standard deviation of each surface was 0.2 and the decay constant c was calculated so that $\gamma(0.5)/\gamma(0) = 0.05$. The vertical scales for each profile are the same as for the horizontal scale.

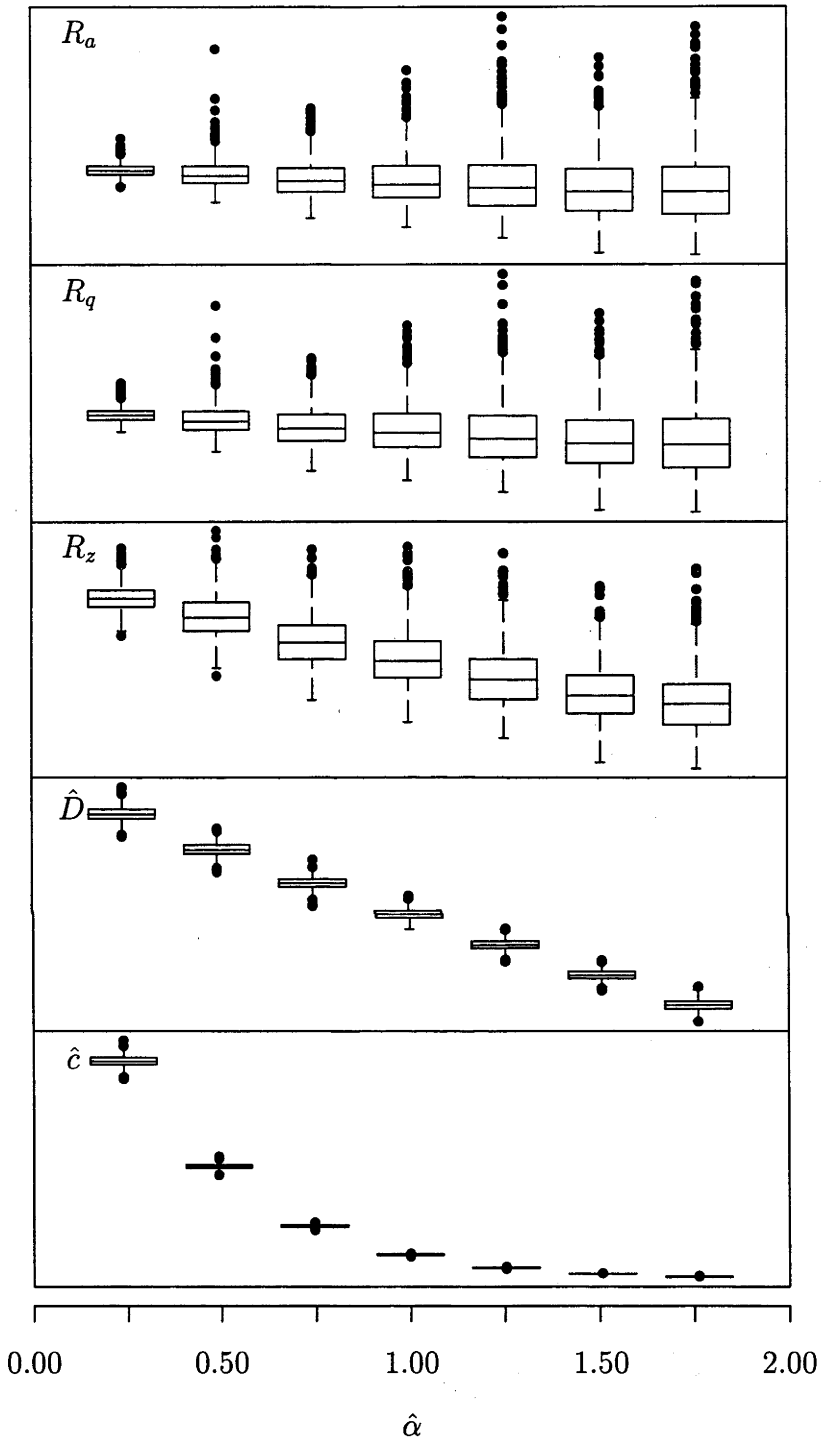


Figure 2.2: Boxplots of roughness area R_a, root-mean-square roughness R_q, ten-point height R_z, estimated fractal dimension $\hat{D}$ and estimated topothesy $\hat{c}$ for 1000 profile data sets, simulated with 7 different values for α; a typical profile for each value of α is shown in Figure 2.1.

only marginal when compared with the dramatic differences between boxplots for estimated fractal dimension and estimated topothesy.

In this instance, the difference in nature between profiles from the different models is exhibited in both the fractal dimension and the topothesy. Since topothesy is a linear functional, we could quite easily exaggerate the vertical scale of profiles from different models by certain factors so that their estimated topothesies were similar. It is comforting to know that this would have no effect on the estimate of fractal dimension, since it is scale-independent.

If, on the other hand, profiles from a particular application were found to have similar fractal dimensions, topothesy could still be used to determine which might be considered roughest.

Chapter 3

The variogram

All of the methods for characterising and comparing surface roughness developed in this thesis are based on the variogram. Because the variogram plays such an important rôle in these methods, it is fitting that a chapter be devoted to its study, which addresses computational and stochastic properties of an appropriate estimator for it.

There are a number of different estimators of the variogram; however for our purposes the naive empirical variogram *will suffice. Its limitations are mitigated by the typical size of data sets in surface problems, and its benefits include mathematical tractability, ease of computation (for which we provide computationally efficient algorithms), and its familiarity to various scientific and engineering groups, although, in some cases, not as an estimator of the variogram.*

The variogram and the auto-covariance function of a surface are similar in that roughness properties that we wish to estimate can be derived from estimators of either. However, there are a number of advantages in using the variogram over the auto-covariance in this instance: it exists for a larger class of problems, it excludes an unnecessary parameter, and its naive estimator is unbiased.

We look at a few ways of displaying the variogram for one- and two-dimensional data. Some of these graphs can provide useful inferences about the roughness char-

acteristics of a surface, such as isotropy or a dominant roughness direction.

The statistical properties of the methods depend heavily on properties of the empirical variogram as an estimator. We derive its moment properties and asymptotic distributional behaviour in terms of the two roughness parameters, fractal index and topothesy.

3.1 Definition

The variogram was defined in section 1.4 for an intrinsically stationary random process. More generally, the variogram for a random process $X(\boldsymbol{t})$ may be defined as a function of two arguments, location $\boldsymbol{t}$ and displacement $\boldsymbol{h}$, by

$$v(\boldsymbol{t}, \boldsymbol{h}) = \text{var}\,[X(\boldsymbol{t} + \boldsymbol{h}) - X(\boldsymbol{t})].$$

From this definition, the following observation about $v(\boldsymbol{t}, \boldsymbol{h})$ may be deduced:

$$v(\boldsymbol{t}, \boldsymbol{h}) = v(\boldsymbol{t} + \boldsymbol{h}, -\boldsymbol{h}). \tag{3.1}$$

If $X(\boldsymbol{t})$ is intrinsically stationary then $v(\boldsymbol{t}, \boldsymbol{h})$ is independent of location and depends only on displacement. In this case we write it as $v(\boldsymbol{h})$.

The simple observation (3.1) then becomes $v(\boldsymbol{h}) = v(-\boldsymbol{h})$, that is, symmetry about the origin. This implies that when calculating or estimating $v(t)$ from one-dimensional data, one need only calculate $v(t)$ for nonnegative values of t, and similarly when estimating $v(\boldsymbol{t})$ from two-dimensional data, one need only be concerned with $\boldsymbol{t}$ in a half-plane.

A necessary condition for the validity of variograms, in the sense that associated sample paths are real-valued, is that of *conditional non-positive definiteness*, that is, for any finite number of spatial locations $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_m$ and real numbers $c_1, \ldots, c_m$ satisfying $\sum_{i=1}^{m} c_i = 0$,

$$\sum_{i=1}^{m} \sum_{j=1}^{m} c_i c_j v(\boldsymbol{t}_i - \boldsymbol{t}_j) \leq 0.$$

In chapter 6 we explore the implications of this important property for fractal dimension and topothesy in the anisotropic two-dimensional setting.

Many methodologies based upon the variogram, including the best-known application of kriging, require that variogram estimators also be conditionally non-positive definite. However, most numerical estimators fail in this respect. To overcome this problem, the numerical estimators are used to fit various variogram models that are conditionally non-positive definite, and then these fitted models are used. This is the case in chapter 7, where we need to simulate surfaces with similar appearances.

3.2 Estimation

One of the simplest estimators of the variogram is the *empirical variogram.* Given a realisation of an intrinsically stationary random process $X : \mathcal{G} \to \mathbb{R}$, the empirical variogram is defined as

$$\hat{v}(\boldsymbol{h}) = |C(\boldsymbol{h})|^{-1} \sum_{\boldsymbol{t} \in C(\boldsymbol{h})} [X(\boldsymbol{t}+\boldsymbol{h}) - X(\boldsymbol{t})]^2 \tag{3.2}$$

where

$$C(\boldsymbol{h}) = \{\boldsymbol{t} \in \mathcal{G} : \boldsymbol{t}+\boldsymbol{h} \in \mathcal{G}\},$$

and $|C(\boldsymbol{h})| \geq 1$. For one-dimensional data, $|C(h)| = n - n|h|$.

When dealing with gridded data, the domain of $\hat{v}(\cdot)$ is itself a grid, centred at the origin. Thus, we can use perspective plots and similar graphical methods to explore its features.

The empirical variogram is also called the *sample variogram* in statistics, or the *structure function* in other sciences. Interestingly, the structure function is already used in some areas of application to estimate fractal dimension, although the reasons for its use lack a theoretical basis.

There are other methods of estimating the variogram; see for example Cressie (1991). Most are concerned with robustifying the estimate against domination by a few summands in (3.2). When dealing with smaller non-gridded data sets (often the case in geostatistics) this is an important practical consideration. However, surface data sets are gridded and often large, implying that the number of summands $|C(\boldsymbol{h})|$ that go into estimating $v(\boldsymbol{h})$ is large. The differences between different variogram estimators for such data are negligible. So, for mathematical convenience, we shall use the naive empirical variogram as our variogram estimator, although the methods developed are appropriate for other estimators too.

We shall see that point estimation of fractal dimension and other roughness parameters requires the variogram to be estimated at only a relatively small number of displacements. However, other methods presented, including the model fitting for simulation of section 4.3, require the variogram to be estimated over a large portion

of its domain, if not its whole domain. For this reason we provide a more efficient holistic algorithm for its computation.

One-dimensional algorithm

In the one-dimensional case,

$$\begin{aligned}\hat{v}(h) &= (n-|h|)^{-1}\sum_{t=1}^{n-|h|}[X(t+|h|)-X(t)]^2 \\ &= (n-|h|)^{-1}(S_{1,|h|}-2C_{|h|}+S_{2,|h|})\end{aligned}$$

where

$$S_{1,j}=\sum_{u=1}^{j}X_u^2, \qquad S_{2,j}=\sum_{u=j}^{n}X_u^2, \qquad \text{and} \qquad C_j=\sum_{u=1}^{n-j}X_uX_{u+j},$$

for $0 \le j < n$. Here, $\{S_{1,j}\}$ and $\{S_{2,j}\}$'s are sequences of partial sums and, assuming that they are calculated in an efficient manner and intermediate results are kept, will require $O(n)$ operations.

The quantity C_j is recognisable as the discrete convolution of $\{X_j\}$ with the reverse of itself. Using the Fast Fourier Transform (FFT) to effect the convolution in the Fourier domain, we may calculate all of the C_j's in $O(n \log n)$ time.

To avoid edge effects, embed the sequence of X_i's in a sequence of zeroes of overall length $2n$: for $0 \le j < 2n$ put

$$Y_j = \begin{cases} X_{j+1}, & \text{if } 0 \le j < n \\ 0, & \text{otherwise.} \end{cases}$$

It should be noted that $\{X_j\}$ need only be extended to length $2n-1$, but we choose to extend it to a length of $2n$, since algorithms for implementing the FFT perform better if the sequence length is a highly composite number.

Then, if $\{F_l\}$ is the FFT of $\{Y_j\}$,

$$F_l = \sum_{j=0}^{2n-1} Y_j e^{-2\pi ijl/n}.$$

Now, $\{C_j\}$ is the circular convolution of $\{Y_j\}$ with its reverse:

$$C_j = \frac{1}{2n} \sum_{l=0}^{2n-1} |F_l|^2 e^{2\pi i l j/n}.$$

The computational complexity of a sequential algorithm is the maximum complexity of its component stages. Hence, the one-dimensional empirical variogram may be computed in $O(n \log n)$ operations, compared with $O(n^2)$ for direct computation.

Two-dimensional algorithm

Although this is a direct extension of the one-dimensional algorithm, there is a significant difference: $\hat{v}(\cdot)$ needs a half-plane rather than just positive values. The following method calculates $\hat{v}(\boldsymbol{h})$ for all displacements $\boldsymbol{h}$ in $\{1-n, \cdots, n-1\}^2$.

Because indices of arrays are usually nonnegative, we introduce the following notation to switch between the domain of $\hat{v}(\cdot)$ and the indices of a two-dimensional array. Variables annotated with a prime, such as j', range over $1-n, \ldots, n$ and will be used for the domain of $\hat{v}(\cdot)$, whereas unannotated variables, such as j, range over $0, \ldots, 2n-1$ and will be used to index arrays. Furthermore, the implicit relationship between a variable and its annotated counterpart is given by

$$j' = \begin{cases} j, & \text{if } 0 \le j \le n \\ j - 2n, & \text{if } n < j < 2n, \end{cases} \qquad j = \begin{cases} j', & \text{if } 0 \le j' \le n \\ j' + 2n, & \text{if } -n < j' < 0. \end{cases}$$

From (3.2),

$$\hat{v}(j', k') = [(n - |j'|)(n - |k'|)]^{-1}(S_{1,(jk)} - 2C_{jk} + S_{2,(jk)})$$

— note the absence of primes on the subscripts — where

$$S_{1,(jk)} = \sum_u \sum_v X_{uv}^2, \qquad S_{2,(jk)} = \sum_u \sum_v X_{u+j',v+k'}^2$$

and

$$C_{jk} = \sum_u \sum_v X_{uv} X_{u+j',v+k'},$$

and the ranges for the summations are defined by $1 \leq u, u+j' \leq n$ and $1 \leq v, v+k' \leq n$.

By analogy with the one-dimensional algorithm, embed the $n \times n$ array of data $\{X_{uv}\}$ into the $2n \times 2n$ array $\{Y_{jk}\}$,

$$Y_{jk} = \begin{cases} X_{j+1,k+1} & \text{if } 0 \leq j,k \leq n \\ 0 & \text{otherwise.} \end{cases}$$

Perform the two-dimensional FFT on $\{Y_{jk}\}$:

$$F_{lm} = \sum_{j=0}^{2n-1} \sum_{k=0}^{2n-1} Y_{jk} e^{-2\pi i(jl+km)/n}.$$

The two-dimensional FFT is just the sequential iteration of the one-dimensional FFT, n times in both directions. It requires $O(n^2 \log n)$ operations.

Note that C_{jk} is the two-dimensional convolution of $\{Y_{jk}\}$ with the reversal, in both directions, of itself:

$$C_{jk} = \tfrac{1}{(2n)^2} \sum_{l=0}^{2n-1} \sum_{m=0}^{2n-1} |F_{lm}|^2 e^{2\pi i(jl+km)/n}.$$

We may obtain $S_{1,(jk)}$ and $S_{2,(jk)}$ from the four intermediate quantities,

$$p_{1,qr} = \sum_{u=1}^{q} \sum_{v=1}^{r} X_{uv}^2, \qquad p_{2,qr} = \sum_{u=1}^{q} \sum_{v=r}^{n} X_{uv}^2,$$
$$p_{3,qr} = \sum_{u=q}^{n} \sum_{v=1}^{r} X_{uv}^2, \qquad p_{4,qr} = \sum_{u=q}^{n} \sum_{v=r}^{n} X_{uv}^2,$$

where q and r range over $1, \ldots, n$. Then,

$$S_{1,(jk)} = S_{2,(2n-j,2n-k)} = \begin{cases} p_{4,jk} & \text{if } j,k < n \\ p_{3,(j,k-n-1)} & \text{if } j < n < k \\ p_{2,(j-n-1,k)} & \text{if } k < n < j \\ p_{1,(j-n-1,k-n-1)} & \text{if } n < j,k \\ 0 & \text{if either } j = n \text{ or } k = n. \end{cases}$$

To compute them efficiently, observe that $\{p_{1,qr}\}$, $\{p_{2,qr}\}$, $\{p_{3,qr}\}$ and $\{p_{4,qr}\}$ are all arrays of partial sums. If care is taken each will require only $O(n^2)$ operations.

Accumulating the computational order of the various stages we can see that the overall algorithm requires $O(n^2 \log n)$ operations, compared with $O(n^4)$ for direct evaluation. Thus, if N is the total size of a data set, n for one-dimensional data and n^2 for two-dimensional data, then both algorithms are $O(N \log N)$.

3.3 Advantages of variogram over autocovariance

If X has a well-defined covariance function then it is possible to construct plug-in estimators of fractal dimension and topothesy from the covariance function of X, since

$$v(\boldsymbol{t}) = 2[\gamma(0) - \gamma(\boldsymbol{t})].$$

However, there are several compelling reasons to employ variogram methods.

Firstly, the variogram exists for a wider range of processes. We model profiles and surfaces by intrinsically stationary and continuous random processes. Continuity ensures that the variogram exists, but for the processes to have a covariance function we also require that they have finite variance. This may not necessarily be a physically valid assumption.

As an example, consider the case of fractional Brownian motion. It is intrinsically stationary and continuous and has a valid variogram that depends only on displacement. However, it does *not* have a covariance function that depends on displacement only: either the fBm is tied at one end, in which case the covariance is non-stationary, or the variance is infinite.

Secondly, if a stationary process has a covariance function, then it is difficult, if not impossible, to construct an unbiased estimator without making extra assumptions. For example, the naive estimator $\hat{\gamma}(h)$ for the covariance function of a one-dimensional process, similar to the empirical variogram estimator for the variogram,

will be heavily biased even at the origin:

$$\begin{aligned}
\mathrm{E}\,\hat{\gamma}(0) &= \mathrm{E}\,n^{-1}\sum_i (X_i - \overline{X})^2 \\
&= \mathrm{E}\,n^{-1}\sum_i \left[(X_i - \mu) - (\overline{X} - \mu)\right]^2 \\
&= \mathrm{E}\,n^{-1}\sum_i \left[(X_i - \mu)^2 - 2(X_i - \mu)(\overline{X} - \mu) + (\overline{X} - \mu)^2\right] \\
&= n^{-1}\sum_i \mathrm{E}(X_i - \mu)^2 - n^{-2}\sum_i\sum_j \mathrm{E}(X_i - \mu)(X_j - \mu) \\
&= \gamma(0) - n^{-2}\sum_i\sum_j \gamma(i - j).
\end{aligned}$$

Furthermore, if the process exhibits long–range dependence, where $\gamma(h)$ does not decrease to zero sufficiently quickly, then the estimator will not even be consistent.

These two problems are both concerned with the variance of a random process: either it is infinite or it is difficult to estimate. Our last reason for preferring the variogram is that the variance of a process is irrelevant when estimating the main roughness parameters, fractal dimension and topothesy. This is because the variance is a "wider range" parameter, in that its effects may be gauged by observing the process at larger displacements, whereas fractal dimension and topothesy operate more locally.

3.4 Exploratory analysis

One-dimensional data The simplest way to explore features of bivariate data is to plot their untransformed values using a scatterplot. Indeed, for variogram estimators this has often been the only graphical exploration carried out and even then, averages of binned values are sometimes all that is actually displayed.

The most common use of such a plot is to help the experienced observer choose a suitable model for the variogram to be fitted to the data and then used in subsequent analysis. By 'suitable' we mean that the model is a valid variogram and has characteristics or features similar to the empirical variogram.

We shall use some of the standard geostatistical terminology. The *sill*, if it exists,

is the fixed value to which the variogram converges as $|t| \to \infty$, and the *range* is the lag at which the variogram can be considered to be "close" to the sill. The *nugget effect* is the size of any discontinuity of the variogram at the origin; its existence is due either to measurement error or to variation in the process at a finer scale than that measured. Another feature is the behaviour of the variogram near the origin: is it concave, linear or convex? In roughness terms, this last property has implications for the value of the fractal index α; the variogram will be concave for $\alpha < 1$, linear for $\alpha = 1$ and convex for $\alpha > 1$.

Figure 3.1 shows a plot of a large portion of the empirical variogram against lag for the roller data described in section 1.2. Both positive and negative lags are

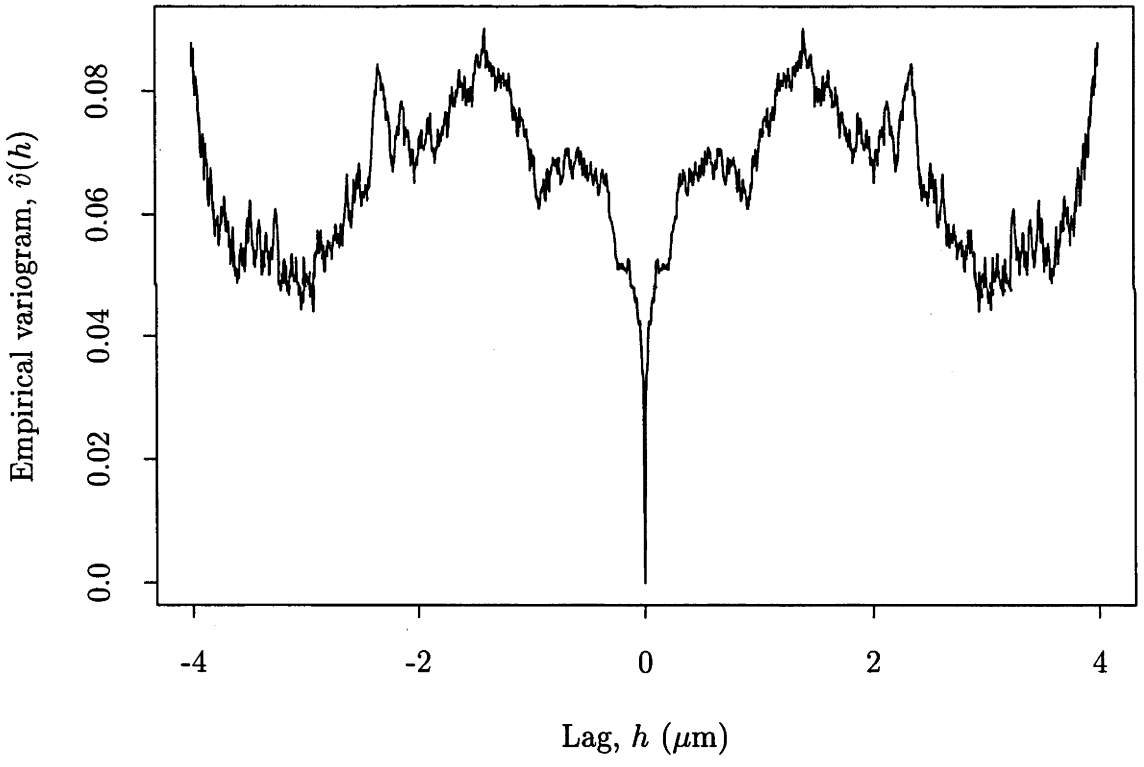


Figure 3.1: Plot of the empirical variogram calculated from the roller data. The graph is highly concave near the origin. This suggests a value of $\alpha < 1$, which in turn indicates a fractal dimension for the roller surface. This rapid decline also makes it difficult to discern whether there is a nugget effect in the underlying variogram or not.

included, so that the behaviour at the origin is easier to ascertain. The variogram appears concave at the origin, implying a value for α less than 1, which indicates a fairly rough surface. Because of the rapid decline of the variogram near the origin, it is difficult to infer a nugget effect or not.

An alternative to the straightforward scatterplot of variogram estimator against lag is a log-log plot of the variables. We have already seen in section 2.3.2 that this transforms the behaviour of the variogram near the origin from a power law to a linear relationship. Comparing the slopes of lines may be easier than comparing the shapes of curves.

The slopes of these lines correspond to the amount of erraticism in the process: the steeper the line, the smoother the surface. Note that the gradients of the lines are bounded by 2, corresponding to an angle of approximately 63°. Human perception is better at differentiating lines up to an angle of 45° Cleveland (1993), and 63° is not a slope with which most people can identify. For these reasons it may be preferable to plot half of the log of variogram against the log of the lag, or $\log \hat{v}(t)^{1/2}$. This puts the variogram onto the same scale as lag. Lines will then have slopes of $\alpha/2$, ranging from a lower limit of 0, which implies a completely erratic (discontinuous) process, to an upper limit of 1 for a smooth (differentiable) process.

Taking logarithms also has the advantage that it allows us to look at the whole variogram at once, while at the same time magnifying the area of greatest interest, near the origin.

So far we have assumed that the data X_i are *exact* measurements from a continuous process. In practice, the X_i's are likely to be subject to error from a number of sources, for example due to the precision of the measurement instrument or to local averaging of the surface. This was indeed the case for the roller surface data and the soil surface data, where the data were discretised height averages. So, variogram estimators calculated from X may exhibit a nugget effect.

If we include an error term in X by setting X equal to a continuous process Y plus independent error η, then an update of model (2.4) becomes

$$v(t) = \nu[1 - \delta(t)] + c|t|^{\alpha} + o(|t|^{\alpha}) \qquad \text{as } t \to 0,$$

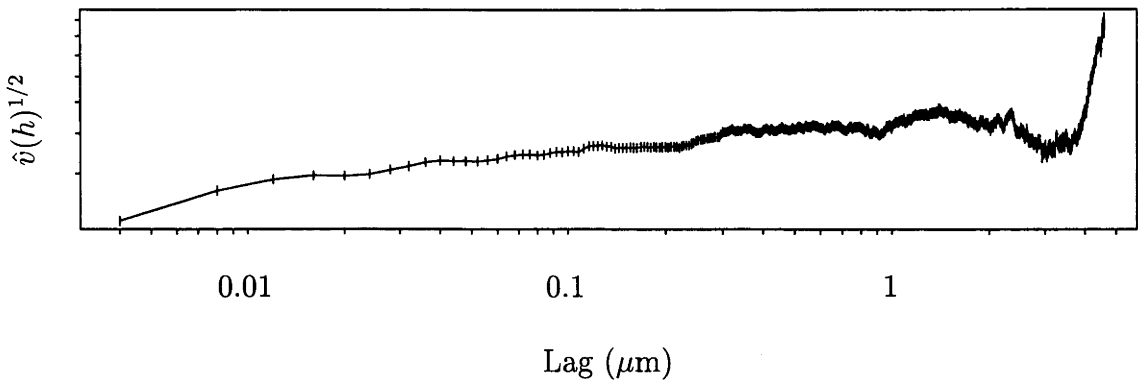


Figure 3.2: Plot on log scales of the empirical variogram versus the lag for the roller data. The slope over the first few lags is 0.31 corresponding to a fractal index $\alpha = 0.61$ and a fractal dimension $D = 1.69$.

where $\nu = 2 \operatorname{var} \eta$ is the nugget effect and δ is the Dirac delta function.

Assuming that, on the measurement scale, the measurement error does not dominate the approximate power law, *i.e.* $\nu < c|t|^{\alpha}$, then

$$\log v(t) \approx \log c + \alpha \log |t| + \log[1 + \frac{\nu}{c|t|^{\alpha}}].$$

The last term is positive and decreases with increasing $|t|$, so that a high nugget effect will manifest itself as an increasing slope near the origin.

This behaviour is contrary to that of the examples discussed in section 1.2. All of their slopes tend to decrease rather than increase with increasing lag. Therefore, we conclude that the measurement error, and hence the nugget effect, is negligible for these examples. However, it is a consideration that should be borne in mind.

Figure 3.2 is a plot of half the log of the empirical variogram against log lag for the roller surface data. The slope over the first two lags is 0.31, corresponding to an estimated fractal dimension of 1.69.

Two-dimensional data Two-dimensional surface data allow us to address the issue of anisotropy and how it affects the roughness properties developed in chapter 2. Before we can explore this relationship, we need to assess whether a surface is indeed anisotropic. To explore this graphically we draw a contour plot of the

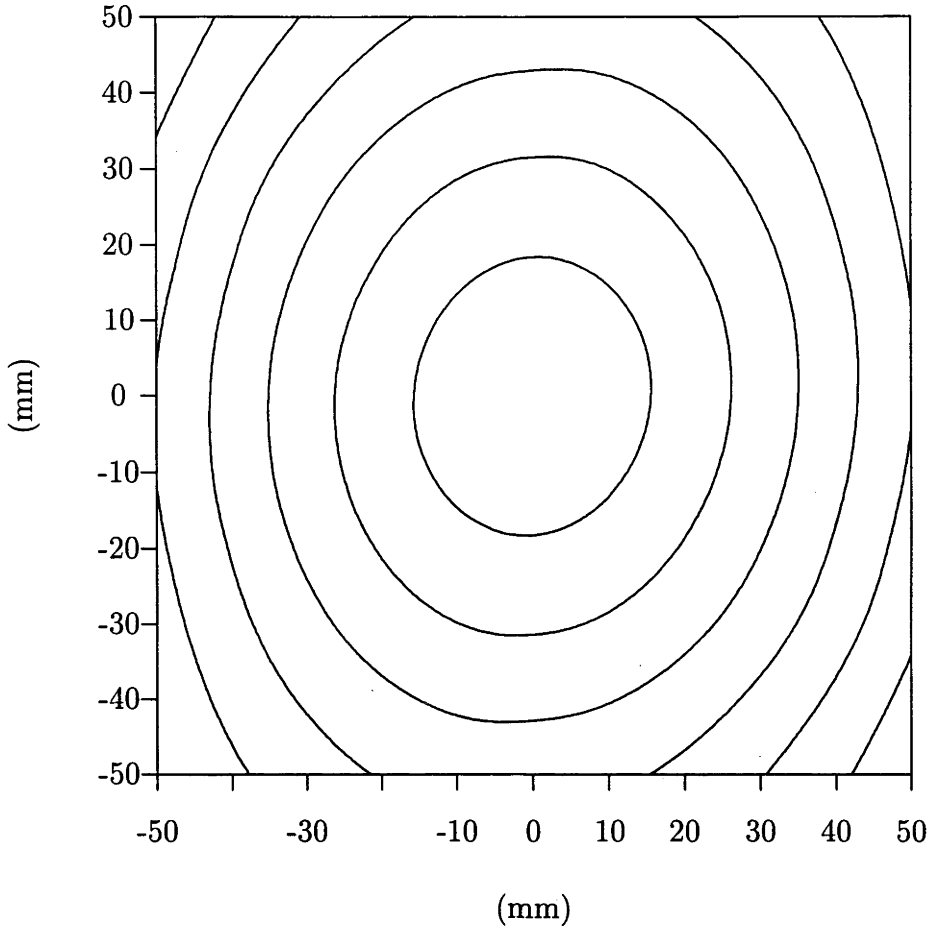


Figure 3.3: Contour plot of the two-dimensional empirical variogram for the soil surface after the maximum amount of simulated rainfall (35.50mm). The individual contours are close to circular in shape indicating the soil surface is close to isotropic. This is as expected since the the soil surface was constructed in the laboratory using a method which did not favour any particular direction.

variogram. A contour plot is more appropriate in this situation than it was for the raw surface data, since the variogram is generally a smooth function and so contours will be easier to interpret. Anisotropy of the data will be manifested by noncircular contours. Because our roughness properties are determined in a local sense, we need only draw contours in the neighbourhood of the origin.

Figure 3.3 depicts contours of the empirical variogram for the wettest soil surface. Although the contours are not perfectly circular, there is not substantial evidence of

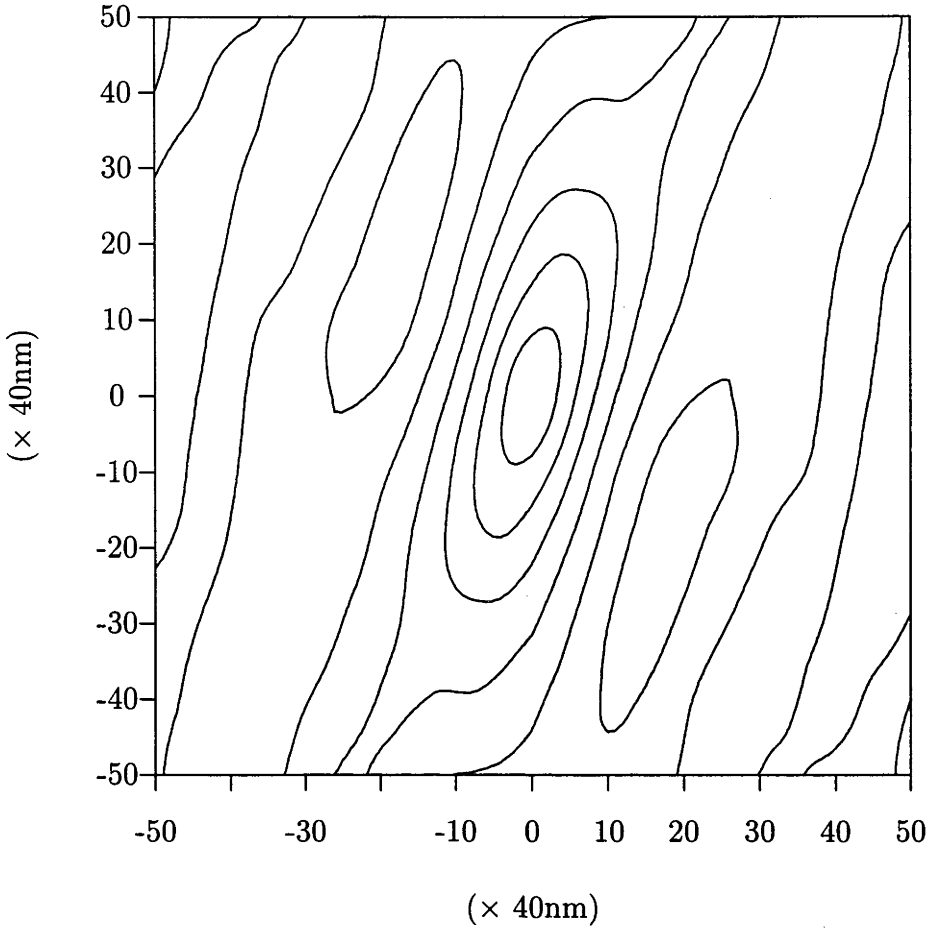


Figure 3.4: Contour plot of the two-dimensional empirical variogram for the fourth polymer data set. The anisotropy of this data set is clear; the contours near the origin are far from circular.

anisotropy. Figure 3.4 depicts the contours of the empirical variogram for the fourth polymer surface. In comparison with Figure 3.3, its contours exhibit a high degree of elongation suggesting a far greater degree of anisotropy than the soil data. Chapter 6 contains a formal statistical test which corroborates this graphical evidence.

Once the existence of anisotropy is decided, we can turn our attention to the effects of anisotropy on roughness properties. More specifically, we wish to investigate how erraticism and scale vary with orientation. To explore this, we use a simple but informative extension of the one-dimensional log-log plot, by combining the one-dimensional log-log plots for each distinct orientation into a single graph.

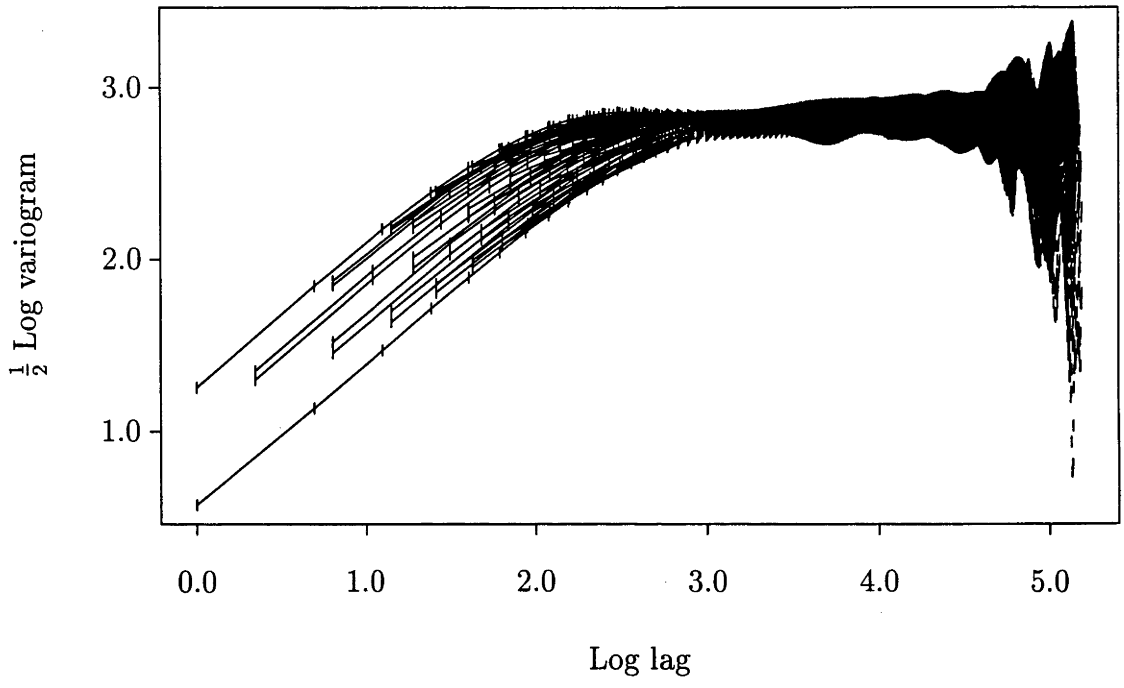


Figure 3.5: Plot of the logarithm of empirical variogram versus the logarithm of lag for the fourth data set. Points originating from displacements with common orientation are joined by lines. The anisotropy inferred in Figure 3.4 is clearly shown here by the difference in value of the variogram in different orientations. However, the form of this anisotropy seems to be limited to the intercepts of the one-dimensional variograms, all having a common slope. This phenomenon is explored in detail in chapter 6.

An indication of how the fractal index and topothesy vary can then be gauged by the relative slopes and intercepts of lines joining points with small lag.

Figure 3.5 is a plot of half the log of the empirical variogram against log lag for the fourth polymer data set. The most striking feature of this graph is that the lines between points at lower lags are parallel, indicating a common fractal dimension and varying topothesy of line transects with different orientations. This in turn implies that the anisotropy of this surface can be explained solely in terms of topothesy. This phenomenon is explored in detail in chapter 6, where it is shown that the effective independence of fractal dimension and direction is a consequence of the

conditional non-positivity constraint on the variogram.

The slopes of these lines are close to 45°, indicating that the surface is rather smooth; this is in agreement with a visual assessment of the rendered surface in panel (d) of Figure 1.3.

One other point to note from Figure 3.5 is the abundance of larger lags at which the empirical variogram is approximately 2.75, providing strong evidence that the variogram has a sill.

3.5 Properties

An important feature of surface roughness data sets is that they are usually very large. Accordingly, we shall require asymptotic distributional properties of the variogram in subsequent chapters. In this section we will describe properties of the variogram estimator in terms of the roughness characteristics of the underlying surface.

It has been shown, as noted in Cressie (1991, page 71), that the limit distribution of the empirical variogram is Gaussian, provided certain mixing conditions are satisfied. We formalise these mixing conditions in terms of the roughness characteristic fractal index, for one- and two-dimensional gridded data. We also give an expression for the asymptotic distribution when these mixing conditions are not satisfied.

Cressie (1991, page 72) states,

> Intuitively, the rate of convergence of these estimators to their limiting distribution will be slower when the correlations between the data are stronger.

We formalise this statement, again in terms of the fractal index, and give infill asymptotic convergence rates for both one- and two-dimensional data.

3.5.1 One-dimensional case

Before we state moment and distributional properties of $\hat{v}$, we give mild assumptions which ensure they are true. We shall require that

$$v(t) = c|t|^{\alpha}\{1 + o(1)\} \tag{3.3}$$

as $t \to 0$, where $c > 0$ and $0 < \alpha < 2$; and that

$$\sup_{\epsilon<|t|<1} |v''(t)| < \infty \qquad \text{for all } \epsilon \in (0,1) \tag{3.4}$$

and

$$v''(t) = c\alpha(\alpha-1)|t|^{\alpha-2}\{1 + o(1)\} \qquad \text{as } t \to 0. \tag{3.5}$$

Condition (3.5) is a second derivative version of the asymptotic power law assumption (3.3).

Bias The empirical variogram is unbiased:

$$\begin{aligned} \mathrm{E}\,\hat{v}(h) &= |C(h)|^{-1} \sum_{t \in C(h)} \mathrm{E}[X(t+h) - X(t)]^2 \\ &= |C(h)|^{-1} \sum_{t \in C(h)} \mathrm{var}\,[X(t+h) - X(t)] \\ &= v(h). \end{aligned}$$

Variance The convergence rate of $\hat{v}(h)$ is determined in a curious manner by the amount of dependence in the process X: the rate of convergence increases with increasing dependence until a critical point, after which it is no longer influenced by dependence. This is a known property of dependent processes and has usually been studied in terms of either mixing conditions or m-dependence, both measures of dependence (*e.g.* Rosenblatt, 1961; Davis & Borgman, 1982).

Rather than using mixing conditions, we state the rates of convergence in terms of α, since this is more relevant to the present application. Also, recall that α provides a measure of dependence.

Formally, if $h = l/n$ then the variance of $\hat{v}(h)$ is $O[\lambda(n)^2]$, where

$$\lambda(n) = \begin{cases} n^{-1/2-\alpha} & \text{if } 0 < \alpha < \frac{3}{2} \\ n^{-2}(\log n)^{1/2} & \text{if } \alpha = \frac{3}{2} \\ n^{-2} & \text{if } \frac{3}{2} < \alpha \leq 2, \end{cases}$$

as $n \to \infty$; see Constantine & Hall (1994) and Matthews (1998).

Note that here we are interested in the behaviour of the empirical variogram near the origin, as measurements are taken on a progressively finer grid. To this end we study the performance of $\hat{v}(h)$ as h decreases. This explains why the convergence rates appear to increase with increasing correlation within X, counter to our previous claims. To illustrate that the claims hold, note that, proportionally, $\text{var}[\hat{v}(h)/v(h)] = O[n^{2\alpha}\lambda(n)^2]$ as $n \to \infty$. These convergence rates do decrease with increasing dependence within X.

Distribution The convergence rate for $\hat{v}(h)$ is exact, in the sense that $\lambda(n) \times [\hat{v}(l/n) - v(l/n)]$ has a nondegenerate limiting distribution. The form of this distribution also depends on α. For values of α not exceeding the critical value of $\frac{3}{2}$, the limiting distribution is Normal. Above $\frac{3}{2}$, the limiting distribution takes the form of a Rosenblatt distribution; these distributions were introduced by Taqqu (1988) and include the special case considered by Rosenblatt (1961). To show that the limiting distribution is a Rosenblatt distribution, we use the method of cumulants.

Firstly, notice that the empirical variogram estimator at lag h can be written as a quadratic form,

$$\hat{v}(h) = (n - nh)^{-1}\boldsymbol{\delta}_h' I_{n-nh}\boldsymbol{\delta}_h,$$

where $\boldsymbol{\delta_h}$ is the $(n - nh)$-vector of increments whose i^{th} component is $\delta^{(i)} = X(i + h) - X(i)$. I_m is the $m \times m$ identity matrix.

Since $X(\cdot)$ is Gaussian and intrinsically stationary, $\boldsymbol{\delta}_h$ is also Gaussian and has a covariance matrix V_h whose $(i, j)^{\text{th}}$ element is $\tau(i - j, h)$ where

$$\tau(u, h) = \tfrac{1}{2}v(u + h) + \tfrac{1}{2}v(u - h) - v(u). \tag{3.6}$$

Now, $\tau(\cdot, h)$ is a second difference of the variogram and so behaves very much like the second derivative of $v(\cdot)$ away from the origin, as will be seen later.

From formulae giving the characteristic function of a quadratic form in Gaussian variables (*e.g.* Johnson & Kotz, 1970, chapter 29), the characteristic function $\chi_n(t; h)$ of the empirical variogram is

$$\chi_n(t; h) = \prod_{j=1}^{n-nh} [1 - 2it(n - nh)^{-1}\lambda_j]^{-1/2},$$

where the λ_j's are the eigenvalues of the covariance matrix V_h. From the characteristic function we can obtain the individual cumulants,

$$\begin{aligned} \text{cum}_k\{\hat{v}(h)\} =& 2^{k-1}(k-1)!(n-nh)^{-k}\sum_{j=1}^{n}\lambda_j^s \\ =& 2^{k-1}(k-1)!(n-nh)^{-k}\,\text{tr}(V_h^k). \end{aligned} \tag{3.7}$$

The last equality uses the identity that the sum of k^{th} powers of the eigenvalues of a matrix is equal to the trace of that matrix raised to the k^{th} power.

Note that V_h is a Toeplitz matrix whose $(i, j)^{\text{th}}$ element is $\tau(i - j, h)$. For such a matrix, the trace of its k^{th} power is given by

$$\text{tr}(V_h^k) = \sum_{i_1=1}^{n-nh} \cdots \sum_{i_k=1}^{n-nh} \tau(i_1 - i_2, h) \ldots \tau(i_k - i_1, h).$$

Using Euler's form for the remainder term in Taylor's theorem, we can rewrite (3.6) as

$$2\tau(u, h) = h^2 \int_0^1 (1-t)\{v''(u+ht) + v''(u-ht)\}dt. \tag{3.8}$$

In view of (3.4) and (3.5), there exist constants $\epsilon_1, \epsilon_2 > 0$, and $0 < C_1(\epsilon_1, \epsilon_2) \leq C_2 < \infty$, such that

$$C_1 \leq \frac{\int_0^1 (1-t)v''(u+ht)dt}{\{\max(|u|, |h|)\}^{\alpha-2}}$$

for all $|u| \leq \epsilon_1$ and $|h| \leq \epsilon_2$, and

$$\frac{\int_0^1 (1-t)v''(u+ht)dt}{\{\max(|u|, |h|)\}^{\alpha-2}} \leq C_2$$

for all $|u| \leq 1$ and $|h| \leq 1$. From these bounds, certain integral approximations to series, and the fact that $3/2 < \alpha < 2$, we may prove that

$$n^{-k} \operatorname{tr}(V_h^k) = h^{2k} J_k(v) + o(h^{2k}), \tag{3.9}$$

where

$$J_k(v) = \int_0^1 \cdots \int_0^1 \left[\prod_{j=1}^{k} v''(x_j - x_{j+1}) \right] dx_1 \ldots dx_k,$$

as $n \to \infty$.

The fact that

$$C_3 \leq |v''(t)|/|t|^{\alpha-2} \leq C_4$$

for $|t| \leq C_5$, and $C_5 > 0$, implies that the integral in the definition of $J_k(v)$ is well-defined.

Combining (3.7) and (3.9) we see that, if $h = l/n$, then

$$\operatorname{cum}_k[n^2 \hat{v}(h)] \to 2^{k-1}(k-1)! l^{2k} J_k(v),$$

the cumulants of a Rosenblatt distribution (see Taqqu, 1988).

3.5.2 Two-dimensional case

The results for the two-dimensional empirical variogram are similar to those for the one-dimensional case. Since validation of the results requires little extra demonstration, we confine ourselves to studying a special case.

In order to state the two-dimensional results clearly, we introduce the following notation. Denote the second-order partial derivatives of v by $v_{pq}(u_1, u_2) = \partial^2 v(u_1, u_2)/\partial u_p \partial u_q$, and let ϕ_1 and ϕ_2 be the direction cosines of the unit vector $\boldsymbol{\phi}$ in the direction ϕ, that is $\boldsymbol{\phi} = (\phi_1, \phi_2)' = (\cos\phi, \sin\phi)'$.

First we require assumptions analogous to (3.3), (3.4) and (3.5):

$$v(\boldsymbol{t}) = c(\arg \boldsymbol{t}) \|\boldsymbol{t}\|^{\alpha} \{1 + o(1)\} \tag{3.10}$$

as $\boldsymbol{t} \to \mathbf{0}$, where $c(\cdot)$ is a positive periodic function, $0 < \alpha < 2$,

$$\sup_{\epsilon < \|\boldsymbol{t}\| < 1} |v_{pq}(\boldsymbol{t})| < \infty \qquad \text{for all } \epsilon \in (0,1) \tag{3.11}$$

and

$$v_{pq}(\boldsymbol{t}) = c_{pq}(\arg \boldsymbol{t}) \|\boldsymbol{t}\|^{\alpha-2}\{1 + o(1)\} \quad \text{as } \|\boldsymbol{t}\| \to 0. \tag{3.12}$$

Here $\arg \boldsymbol{t}$ denotes the angle made by the 2-vector $\boldsymbol{t}$ to a fixed direction in the plane. The quantities $c_{pq}(\theta)$ are periodic functions depending on α, $c(\theta)$, $c'(\theta)$ and $c''(\theta)$, given by

$$\begin{aligned}
c_{11}(\theta) &= \alpha[(\alpha-1)\theta_1^2 + \theta_2^2]\, c(\theta) - 2(\alpha-1)\theta_1\theta_2 c'(\theta) + \theta_2^2 c''(\theta), \\
c_{12}(\theta) = c_{21}(\theta) &= \alpha(\alpha-2)\theta_1\theta_2\, c(\theta) + (\alpha-1)(\theta_1^2 - \theta_2^2)\, c'(\theta) - \theta_1\theta_2\, c''(\theta), \\
c_{22}(\theta) &= \alpha[\theta_1^2 + (\alpha-1)\theta_2^2]\, c(\theta) - 2(\alpha-1)\theta_1\theta_2 c'(\theta) + \theta_1^2 c''(\theta).
\end{aligned}$$

For this we require that $c(\theta), c'(\theta)$ and $c''(\theta)$ exist, are finite, and do not vanish simultaneously.

Variance The main difference between the one- and two-dimensional results is the critical point at which behaviour changes. For two-dimensional data from a process satisfying assumptions (3.10), (3.11) and (3.12), if $\boldsymbol{h} = n^{-1}\boldsymbol{l}$ then the variance of $\hat{v}(\boldsymbol{h})$ is $O(\lambda_{2D}(n)^2)$, where

$$\lambda_{2D}(n) = \begin{cases} n^{-1-\alpha} & \text{if } 0 < \alpha < 1 \\ n^{-2}(\log n)^{1/2} & \text{if } \alpha = 1 \\ n^{-2} & \text{if } 1 < \alpha \le 2, \end{cases}$$

as $n \to \infty$; see Davies & Hall (1999).

Distribution Similar to the one-dimensional case, if conditions (3.11) and (3.12) are satisfied then, for $1 < \alpha < 2$, the asymptotic distribution of $n^2[\hat{v}(n^{-1}\boldsymbol{l}) - v(n^{-1}\boldsymbol{l})]$ is a Rosenblatt distribution, whose cumulants are $2^{k-1}(k-1)!\|\boldsymbol{l}\|^{2k} J_k(v, \arg \boldsymbol{l})$, where

$$J_k(v,\phi) = \int_{I^2} \cdots \int_{I^2} \prod_{j=1}^{k} \sum_{p=1}^{2} \sum_{q=1}^{2} \phi_p \phi_q v_{pq}(\boldsymbol{x}_j - \boldsymbol{x}_{j+1})\, d\boldsymbol{x}_1 \ldots d\boldsymbol{x}_k. \tag{3.13}$$

Ruled surface A *ruled surface*, X_θ, is an extension of a one-dimensional process to two dimensions by 'drawing out' the process in a specified direction. This is formulated mathematically by

$$X_\theta(\boldsymbol{t}) = X(\boldsymbol{t}'\boldsymbol{\theta}),$$

where $\boldsymbol{t}'\boldsymbol{\theta}$ denotes the scalar product of $\boldsymbol{t}$ and $\boldsymbol{\theta}$. Note that $\boldsymbol{\theta}$ is perpendicular to the direction of drawing.

For such a surface the variogram obeys a similar law, $v(\boldsymbol{t}) = v(\boldsymbol{t}'\boldsymbol{\theta})$, allowing its derivatives to be expressed in terms of a one-dimensional variogram:

$$v_{11}(\boldsymbol{t}) = \theta_1^2 v''(\boldsymbol{t}'\boldsymbol{\theta}), \quad v_{12}(\boldsymbol{t}) = v_{12}(\boldsymbol{t}) = \theta_1\theta_2 v''(\boldsymbol{t}'\boldsymbol{\theta}) \quad \text{and} \quad v_{22}(\boldsymbol{t}) = \theta_2^2 v''(\boldsymbol{t}'\boldsymbol{\theta}).$$

Although these derivatives do not satisfy (3.11) and (3.12), if we assume that v'' obeys the one-dimensional assumptions (3.4) and (3.5), the cumulants of $\hat{v}(\boldsymbol{h})$ will behave similarly to those in the one-dimensional case, although the value for $J_k(v, \phi)$ will be slightly different.

To see this, note that

$$\sum\sum \phi_p\phi_q v_{pq}(\boldsymbol{t}) = \sum\sum \phi_p\phi_q\theta_p\theta_q v''(\boldsymbol{t}'\boldsymbol{\theta}) = (\boldsymbol{\theta}'\boldsymbol{\phi})^2 v''(\boldsymbol{t}'\boldsymbol{\theta}).$$

Therefore, by rotating the $\boldsymbol{x}_j$'s so that their first component is in the direction of θ in (3.13), we get

$$J_k(v, \phi) = (\boldsymbol{\theta}'\boldsymbol{\phi})^{2k} \int_0^{\theta_1+\theta_2} \cdots \int_0^{\theta_1+\theta_2} \prod_{j=1}^{k} L(y_j) v''(y_j - y_{j+1}) dy_1 \ldots dy_{j+1}$$

where L is linear. Comparing this to its one-dimensional form $J_k(v)$, we see that the two are indeed well-defined for the same range of α.

Thus, the empirical variogram for data from a ruled surface has similar distributional properties to the one-dimensional estimator. In particular, the critical point for α is $\frac{3}{2}$.

This is not altogether surprising, since measurements of a ruled surface taken over a two-dimensional grid are essentially not providing any extra information. In

practical terms, for those applications where the surface can be modelled as the superposition of a dominant ruled surface and a finer rough surface, it may be more cost-effective to collect one-dimensional data across and along the direction of the rulings. Applications for which this is the case are not as rare as might be expected. Some processes do generate surfaces with a dominant direction, for example milled engineering surfaces, extruded materials and ploughed fields. The latter example is particularly relevant to the soil application detailed in section 1.2, but on a larger scale.

Comparison between dimensions The main difference between the one- and two-dimensional empirical variograms is the critical point above which their convergence rates progressively deteriorate, $\frac{3}{2}$ and 1, respectively. Below these points, the estimators behave commensurately, in that $\text{var}\{\hat{v}(\cdot)/v(\cdot)\} = O(N^{-1})$ where N is the total sample size. It is interesting to note that the analyses of all the sample surfaces described in section 1.2 lead to estimates of α exceeding the two-dimensional critical point, and mostly below the one-dimensional critical point. So for surfaces that may be assumed isotropic for physical reasons, extensive one-dimensional data may provide more information than two-dimensional data, at least in terms of variogram-based analyses.

Chapter 4

Analysing roughness from one-dimensional transects

The estimators of fractal index and topothesy for one-dimensional data, introduced in chapter 2, quantify the erraticism and scale of a particular transect. Before the estimators can be put into practice, it is important to stress that the error inherent in their estimation is due to the finite, as opposed to infinitesimal, nature of the data. To gauge how the error depends on sample size, we present asymptotic convergence results for the estimators, showing that they are consistent and showing how their rates of convergence depend on the true values of the roughness parameters.

The remainder of the chapter concentrates on practical aspects of roughness analysis using the estimators of fractal index and topothesy. For instance, we consider the choice of the number of regression points included in the estimation, showing that relatively few points are required to obtain good estimates.

The asymptotic convergence results confirm that the errors in estimating the fractal index and topothesy decrease with increasing sample size, but they do not quantify the amount of error for particular finite samples. We give two model–based methods for estimating these errors, providing a practical means of assessing the precision of fractal index and topothesy estimates.

Both these methods require smooth estimates of the variogram, that can be obtained by fitting "local" models to the empirical variogram. However, there are two significant problems in applying existing fitting procedures: the highly dependent nature of the empirical variogram, and the generality of the fitted model. Singly or combined, these problems lead to volatile parameter estimates, or to no estimates at all, due to failure in convergence. To overcome these problems we provide a heuristic fitting procedure tailored to the model we wish to fit. To supplement that procedure we give a bootstrap method for model validation.

In many cases, the surface elevation data are not exact: no measurement device is perfect. The effects of a particular kind of measurement error, that due to local averaging, are shown to have a dramatic effect on the roughness analysis. We explore modifications to the basic estimators that ameliorate the effects of measurement error.

The methods developed are also applied to roller data.

4.1 Point estimation

For one-dimensional surfaces, roughness is characterised by fractal dimension D and topothesy $c^{1/2}$, as discussed in Chapter 2. Estimates were derived from $\hat{\alpha}$ and $\log \hat{c}$, the slope and intercept of a simple linear regression on the log scale.

The defining equations for $\hat{\alpha}$ and $\log \hat{c}$, (2.6) and (2.8), may be rewritten as simple linear sums,

$$\hat{\alpha} = \sum_{l=1}^{k} a_l \log \hat{v}(l/n) \qquad \text{and} \qquad \log \hat{c} = \sum_{l=1}^{k} b_l \log \hat{v}(l/n), \tag{4.1}$$

where $a_l = (x_l - \bar{x}) / \sum (x_l - \bar{x})^2$, $b_l = k^{-1} - a_l \bar{x}$, $x_l = \log(l/n)$ and $\bar{x} = k^{-1} \sum x_l$. Note that $\sum a_l = 0$, $\sum a_l \log(l/n) = 1$, $\sum b_l = 1$ and $\sum b_l \log(l/n) = 0$. Rewriting the equations in this way reduces the complexity of equations that follow.

4.1.1 Theoretical performance of estimators

The theoretical properties of $\hat{\alpha}$ and $\log \hat{c}$ have been discussed by Constantine & Hall (1994) and Matthews (1998). For completeness, we state their results and show how they may be obtained from properties of the empirical variogram discussed in section 3.5.

There are two main sources of error in $\hat{\alpha}$: a systematic error due to the departure of the underlying variogram from a power law, and a random error due to the stochastic nature of the variogram estimator. To formulate this mathematically, we may write

$$\hat{\alpha} = \alpha + E_1 + E_2,$$

where

$$E_1 = \sum a_l \log \frac{v(l/n)}{c|l/n|^{\alpha}} \qquad \text{and} \qquad E_2 = \sum a_l \log \frac{\hat{v}(l/n)}{v(l/n)},$$

are the systematic and random errors, respectively.

In E_1, the form of the ratio of $v(t)$ to $c|t|^{\alpha}$ depends on the $o(|t|^{\alpha})$ term in (2.4). In order to study the contribution of this term to E_1, and hence to the error in $\hat{\alpha}$,

we refine (2.4) to model $o(|t|^\alpha)$ more carefully. Assume that

$$v(t) = c|t|^\alpha + d|t|^{\alpha+\beta} + o(|t|^{\alpha+\beta}),$$

or equivalently, that

$$v(t) = c|t|^\alpha\{1 + d|t|^\beta + o(|t|^\beta)\},$$

for $d \neq 0$ and $\beta > 0$, as $t \to 0$. Then

$$E_1 = C_1 n^{-\beta}\{1 + o(1)\}$$

as $n \to \infty$, where $C_1 = d\sum a_l l^\beta$.

The behaviour of E_2 is determined by the error about the mean of $\hat{v}(h)$ at small lags. To see this, note that

$$\log\frac{\hat{v}(h)}{v(h)} = \log\left[1 + \frac{\hat{v}(h) - v(h)}{v(h)}\right].$$

Thus, from the stochastic properties of the empirical variogram developed in section 3.5,

$$E_2 = n^\alpha \lambda(n) R\{1 + o_p(1)\} \tag{4.2}$$

as $n \to \infty$. Here, R is a random variable which is independent of n, unbiased and with a non-degenerate distribution that is Gaussian for $0 < \alpha \leq 3/2$ and non-Gaussian for $3/2 < \alpha < 2$. The form of the non-Gaussian distribution is related to the Rosenblatt distribution, since R is a linear combination of a small number of Rosenblatt-distributed variables.

Which of E_1 and E_2 dominates the error of $\hat{\alpha}$ depends on the values of α and β. For values of α below the critical point of $3/2$, E_1 dominates when $\beta < 1/2$, and for values of α above the critical point, E_1 dominates when $\beta < 2 - \alpha$. At all other times E_2 is, asymptotically, the dominant source of error. As an example, consider a Gaussian process with the stable exponential covariance function, $\gamma(t) = \exp(-c|t|^\alpha)$. Here, $\beta = \alpha$ and E_1 dominates only when $\alpha < 1/2$.

The rates of convergence of estimators of α and c lessen as α gets close to 2 because the amount of information contained in oscillations of X decreases as the

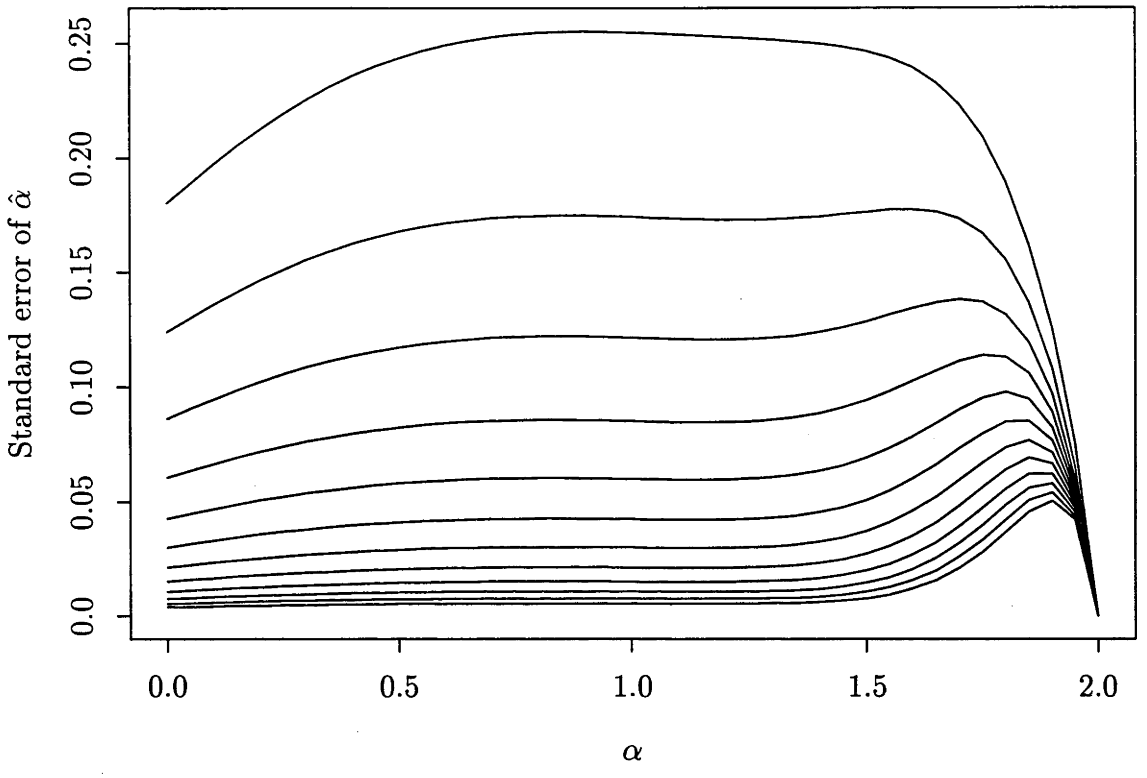


Figure 4.1: Graphs of the parameter α and corresponding standard error $s_{\hat{\alpha}}$ of its estimate $\hat{\alpha}$, based on the simulation study; $k = 4$ points were used to fit the regression line. Each curve represents a different sample size n ranging from 2^5 at the top to 2^{16} at the bottom, in powers of 2. The rates of convergence for $\alpha > 3/2$ appear less than for $\alpha < 3/2$.

process becomes smoother. In the extreme case when $\alpha = 2$, there is insufficient information in a record of X on a finite interval to estimate c consistently. As α increases to 2, one needs to examine successively higher-order properties in order to obtain good performance; see Kent & Wood (1997).

4.1.2 Numerical issues

The convergence problems as α approaches 2 are observable in a simulation study. Figure 4.1 depicts standard errors of $\hat{\alpha}$ for transects from fBm (fractional Brownian motion) models with different values for α. Each curve represents a different sample size, n, with n ranging from 2^5 to 2^{16} in powers of 2.

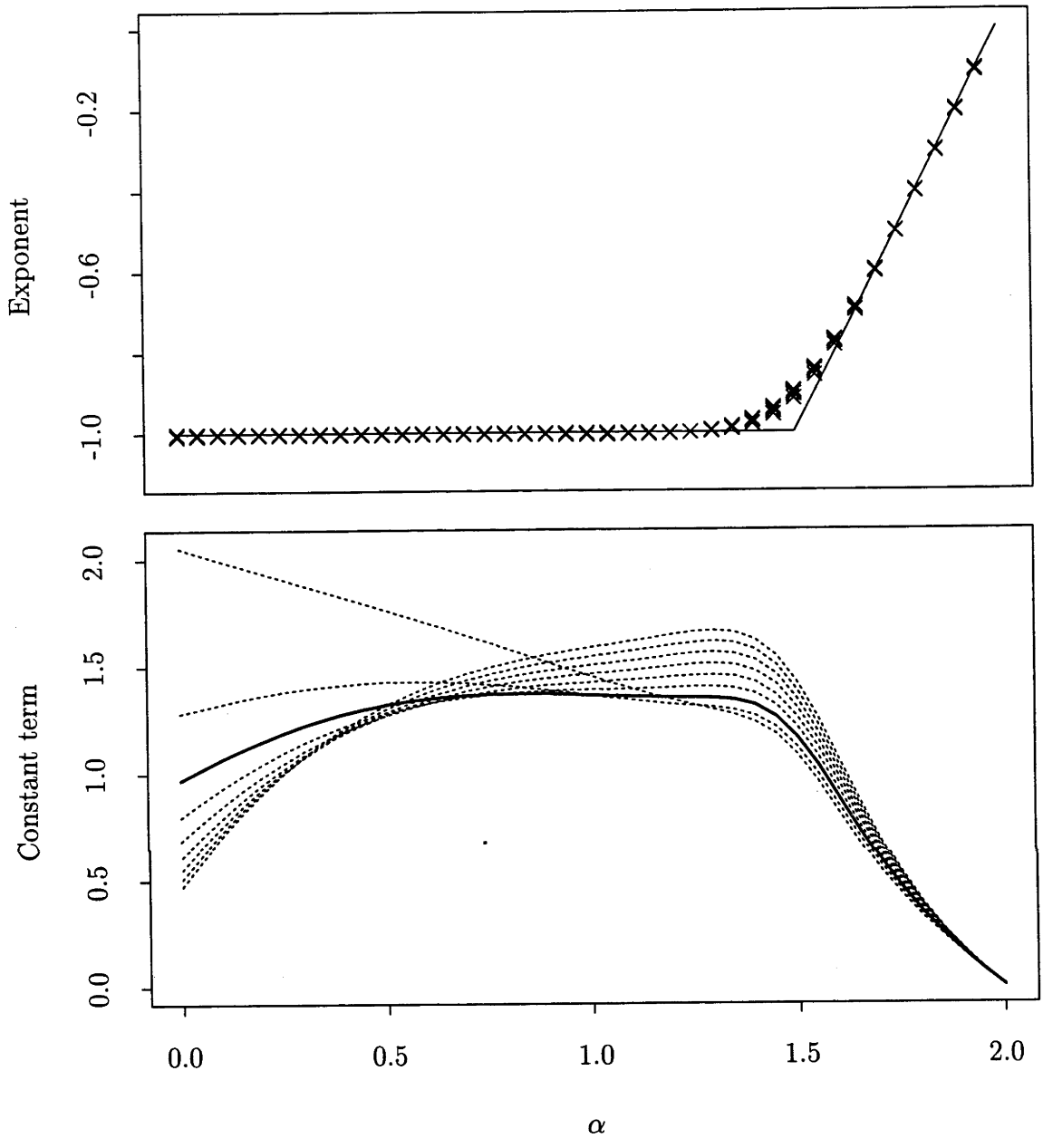


Figure 4.2: (a) Scatterplot of exponents of n from estimated convergence rates of the standard error of $\hat{\alpha}$, calculated using $k = 2, \ldots, 10$, for values of $\alpha = 0(0.05)2$. The line denotes the theoretical convergence rates. (b) Graph of constant term in the convergence of the standard error of $\hat{\alpha}$. Each dotted line represents the number of regression points $k = 2, \ldots, 10$ used in estimating $\hat{\alpha}$. The solid line represents a global optimum value of 4 for k. The number of points used in the regression k does not affect the asymptotic convergence rates of $\hat{\alpha}$, but, for finite samples, a value of $k = 4$ appears optimal for fBm.

The convergence of the standard error to 0 as α approaches 2 is a consequence of the underlying fBm model. For an fBm process, the correlation between increments tends to 1 as α approaches 2, and sample paths tend towards a straight line. For straight lines, $\hat{\alpha} = \alpha$, and there is no error. This is an artifact of the fBm model. In general, for Gaussian processes with an underlying stable exponential covariance function, the standard error of $\hat{\alpha}$ does not converge to 0 as α approaches 2; rather, each has a limit that depends on specific parameter values.

Notice how the convergence rate of the standard error lessens for α above 1.5, supporting the theoretical results obtained above. To quantify this, experimental convergence rates were obtained by simple linear regressions of the logarithm of standard error against the logarithm of sample size n, for different values of α; see for example panel (a) of Figure 5.1. The 12 points in each linear regression were calculated for values of n ranging from 32 to 65536 in powers of 2, and the variogram model used was the fBm model. The slope of a regression line then acts as an estimator of the exponent of the sample size at which the standard error converges. For example, a slope of -0.5 implies a convergence rate of $n^{-1/2}$. Exponents were estimated for a range of values of α, and are shown in panel (a) of Figure 4.2. For *all* values of α there are a number of estimates of convergence exponent. These relate to the number of design points $k = 2, \ldots, 10$, in the regression, used in determining the standard errors of $\hat{\alpha}$, showing that choice of k does not affect convergence. The solid line in the graph is the theoretical convergence rate.

4.1.3 Choice of k

In addition to saving computational labour, there are two reasons for choosing k relatively small. Firstly, it minimises bias from the $o(1)$ term in equation (2.5). Secondly, estimates of the variogram exhibit high correlation between values at neighbouring lags, and using more lags in the regression does not improve the accuracy of estimates as much as in the case of data with lower correlation. Indeed, if the number of points in the regression increases with sample size, both bias and variance will increase; so, even in an asymptotic sense, as $n \to \infty$, the optimal k is

bounded. This has been shown theoretically by Constantine & Hall (1994).

As mentioned above, panel (a) of Figure 4.2 shows experimental convergence rates for the standard error of $\hat{\alpha}$, and indicates that the choice of k is not important with respect to convergence; however, we can seek to minimise the constant term of the standard error by choosing k wisely. Approximate values for this constant term may be obtained from the intercept of the regression lines used to determine the approximate convergence rate. Panel (b) of Figure 4.2 contains 9 curves, corresponding to $k = 2, \ldots, 10$, for this constant term against α. The value of k for which the average value of the constant term is minimised is 4 — the solid line — providing a global optimum in the limiting case of fBm. Note that, for $\alpha > 1$, a slight improvement in the constant term of the variance may be achieved by taking k less than 4.

4.1.4 Possible improvements

The estimators $\hat{\alpha}$ and $\log \hat{c}$ minimise a least squares criterion. This is the usual fitting method for a model with independent Gaussian errors. Although there is numerical evidence that the logarithm of the empirical variogram is approximately Gaussian for a range of α (Baczkowski & Mardia, 1987), the errors about the mean are highly correlated. Indeed, as α approaches 2, the correlation between the values of $\hat{v}(\cdot)$ at different lags tends to 1.

Since the errors are correlated, an improvement in the performance of $\hat{\alpha}$ and $\hat{c}$ might be obtained by taking these correlations into account in the fitting criterion. For example, it may be better to minimise a *generalised* least squares criterion: a quadratic form in the residuals using the inverse of their covariance matrix. This would require calculation, or prior knowledge, of the covariance matrix of the variogram at different lags. Calculation of the covariance matrix is not a simple task, and errors are compounded when the inverse of the matrix is taken. The problem is exacerbated by the high correlations between lags.

In an extensive study, Kent & Wood (1997) compare ordinary least squares and generalised least squares estimates of fractal index based on increments of the

underlying process X, of which the present method is a special case. Although some improvement may be gained using the Kent-Wood generalised least squares estimators, they do not perform consistently better than the ordinary least squares estimators; indeed they may perform more poorly.

A simpler alternative to generalised least squares is to minimise a *weighted* least squares criterion. This is similar to generalised least squares but ignores off-diagonal elements in the covariance matrix. As such, it does not take correlations into account, but re-weights individual contributions in inverse proportion to their variances.

Cressie (1985) explores variogram model fitting by weighted least squares. He shows that, for one-dimensional equally-spaced data, minimising

$$\sum_h |C(h)| \left\{ \frac{\hat{v}(h)}{v(h;\boldsymbol{\omega})} - 1 \right\}^2 \tag{4.3}$$

is a good approximation to weighted least squares, where $\boldsymbol{\omega}$ is the vector of model parameters.

When $\boldsymbol{\omega}$ is chosen such that $v(h;\boldsymbol{\omega})$ is close to $\hat{v}(h)$,

$$\frac{\hat{v}(h)}{v(h;\boldsymbol{\omega})} - 1 \approx \log \hat{v}(h) - \log v(h;\boldsymbol{\omega}).$$

Also, for large n and small values of h, $|C(h)|$ is virtually independent of h. Thus, for the present application, by minimising a least squares criterion on the log scale, we are in fact approximating (4.3) and hence approximating weighted least squares.

4.2 Quantifying the error

In the previous section, we derived asymptotic expressions for the bias and variance of the estimators. This allowed us to show that the estimators are consistent, but did not provide methods for estimating the error for finite samples.

In the present section, we propose two numerical methods for quantifying the errors. The first is a plug-in approach based on asymptotic approximations of the variance of $\hat{\alpha}$, and the second employs the parametric bootstrap. Both provide estimates of standard error; however, the bootstrap method may be modified to provide

confidence intervals. This is particularly relevant when the asymptotic distribution of $\hat{\alpha}$ is non-Gaussian, *i.e.* for α above the critical point 3/2.

The extra utility of the bootstrap method comes at a computational cost. Since it uses Monte Carlo simulations it is computationally expensive, whereas the plug-in method is relatively quick to implement. Therefore, the choice of which method to use may depend on sample size.

Both the plug-in and bootstrap methods require the underlying process to be Gaussian. However, for the bootstrap method, this restriction is due to the availability of appropriate modelling and simulation methods for other distributional families. Such methods are not considered in the thesis.

4.2.1 Plug-in method

The variance of $\hat{\alpha}$ can be written in terms of the variances and covariances of the logarithm of the empirical variogram at small lags:

$$\mathrm{var}(\hat{\alpha}) = \sum_{l_1}\sum_{l_2} a_{l_1} a_{l_2} \,\mathrm{cov}\{\log \hat{v}(l_1/n), \log \hat{v}(l_2/n)\}.$$

If we have estimates $s(h_1, h_2)$ of the covariances, $\mathrm{cov}\{\log \hat{v}(h_1), \log \hat{v}(h_2)\}$, then the plug-in estimate of the variance of $\hat{\alpha}$ is

$$s^2_{\hat{\alpha}} = \sum_{l_1}\sum_{l_2} a_{l_1} a_{l_2} s(l_1/n, l_2/n). \tag{4.4}$$

Now, the covariance between the empirical variogram at two lags on the log scale can be approximated by rescaled covariances of the untransformed empirical variogram:

$$\mathrm{cov}\{\log \hat{v}(h_i), \log \hat{v}(h_j)\} \approx \frac{\mathrm{cov}\{\hat{v}(h_i), \hat{v}(h_j)\}}{v(h_i)v(h_j)}. \tag{4.5}$$

Let $T(v; t, h_1, h_2) = v(t + h_1 - h_2) - v(t + h_1) - v(t - h_2) + v(t)$. Then, for a Gaussian random field, the covariances of the empirical variogram estimates are

$$\mathrm{cov}\{\hat{v}(h_1), \hat{v}(h_2)\} = \tfrac{1}{2}|C(h_1)|^{-1}|C(h_2)|^{-1} \sum_{t_1}\sum_{t_2} T(v; t_1/n - t_2/n, h_1, h_2)^2. \tag{4.6}$$

If $v(h)$ were known, $\mathrm{cov}\{\hat{v}(h_1), \hat{v}(h_2)\}$ could be calculated directly and used in (4.5) to obtain $s(h_1, h_2)$. Since $v(\cdot)$ is not known, an estimator for it such as $\hat{v}(\cdot)$ could be used instead. However, the high variability of $\hat{v}(\cdot)$ at larger lags would dominate the sums in (4.6), especially when squared. Instead, a suitable model $\tilde{v}(\cdot)$ is fitted to $\hat{v}(\cdot)$ using the methods described in section 4.3, and substituted for $v(\cdot)$ in (4.6) to obtain an estimate of $\mathrm{cov}\{\hat{v}(h_1), \hat{v}(h_2)\}$.

Thus, after some manipulation to reduce computational effort, our estimates $s(h_1, h_2)$ of the covariances $\mathrm{cov}\{\log \hat{v}(h_1), \log \hat{v}(h_2)\}$ are given by

$$s(h_1, h_2) = [2n^2(1-h_1)(1-h_2)\tilde{v}(h_1)\tilde{v}(h_2)]^{-1} \sum_{u=1-(n-nh_1)}^{n-nh_2-1} N(u/n)T(\tilde{v}; u/n, h_1, h_2)^2,$$

where $N(t) = n \times \min(1-h_1+t, 1-h_1, 1-h_2, 1-h_2-t)$. Then the estimate of the variance of $\hat{\alpha}$ is the quadratic form, $s^2_{\hat{\alpha}}$, defined in (4.4).

The variance of $\log \hat{c}$ can be calculated in a similar manner by substituting b_l for a_l in (4.4). To obtain an estimate for the variance of $\hat{c}$ we need to rescale. Thus,

$$s^2_{\hat{c}} = \hat{c}^2 \sum_{l_1} \sum_{l_2} b_{l_1} b_{l_2} s(l_1/n, l_2/n)$$

is our estimate for the variance of $\hat{c}$.

Note that the difference in behaviour of $\hat{\alpha}$ for α above and below the critical point, is attributable to the contributions of the summands in (4.6). When $\alpha < 3/2$, the main contribution to the sums comes when $t_1 - t_2$ is small. This implies that the choice of the variogram model to fit to the empirical variogram is not crucial, since they all have a similar behaviour near the origin. Nevertheless, when $\alpha > 3/2$ the influence of summands for larger values of $t_1 - t_2$ increases, and they may contribute as highly as for smaller values of $t_1 - t_2$. So, when $\alpha > 3/2$ the choice for the model $\tilde{v}(\cdot)$ is more important.

4.2.2 Bootstrap method

Confidence intervals for $\hat{\alpha}$ and $\hat{c}$ may also be obtained using the parametric bootstrap, starting with an appropriate model for the covariance of the process that

produced the data. The procedure is as follows.

Step 1: Fit a valid variogram model to the empirical variogram using the methods described in section 4.3. The model used should permit the full range of allowable values for α and c. Denote the fitted variogram by $\tilde{v}(\cdot)$.

Step 2: Simulate a number, B say, of realisations of Gaussian random fields with the fitted model as their variogram. Denote a typical realisation by X_b^*.

Step 3: For all simulated processes X_b^*, calculate their respective empirical variograms $\hat{v}_b^*(\cdot)$ and use (4.1) to obtain B pairs of estimates, $(\hat{\alpha}_b^*, \hat{c}_b^*)$, for fractal index and topothesy.

Standard errors

Step 4: Calculate the sample standard deviations of the resampled parameter values $\{\hat{\alpha}_b^*\}$ and $\{\hat{c}_b^*\}$ to obtain estimated standard errors for the point estimates, $\hat{\alpha}$ and $\hat{c}$.

Confidence intervals

Step 5: Use equations similar to those in (4.1), substituting the fitted variogram model $\tilde{v}(u)$ from Step 1 for $\hat{v}(u)$, to obtain expected values (α^*, c^*) for $(\hat{\alpha}_b^*, \hat{c}_b^*)$ calculated from the simulated data sets. The difference between (α^*, c^*) and the parameter estimates $(\hat{\alpha}, \hat{c})$ from the original data set can be used to correct $(\hat{\alpha}_b^*, \hat{c}_b^*)$ for bias, yielding $(\tilde{\alpha}_b^*, \tilde{c}_b^*)$.

Step 6: Let $\tilde{\alpha}_{[1]}^* \leq \cdots \leq \tilde{\alpha}_{[B]}^*$ and $\tilde{c}_{[1]}^* \leq \cdots \leq \tilde{c}_{[B]}^*$ be ordered sets of values taken from the $\tilde{\alpha}_b^*$'s and the $\tilde{c}_b^*$'s respectively. Marginal β-level confidence intervals for α and c are then given by $(\tilde{\alpha}_{[\frac{1}{2}\beta B]}^*, \tilde{\alpha}_{[(1-\frac{1}{2}\beta)B]}^*)$ and $(\tilde{c}_{[\frac{1}{2}\beta B]}^*, \tilde{c}_{[(1-\frac{1}{2}\beta)B]}^*)$, respectively.

Because of its general nature, application of the bootstrap method is not confined to fractal index and topothesy: it can also be used to estimate standard errors or approximate confidence intervals for any of the traditional roughness measures discussed in section 2.1.

4.3 Fitting a variogram model

Both the plug-in and bootstrap methods for estimating standard errors require an estimate of the variogram. In the case of the bootstrap procedure, in which the variogram estimate is used to simulate similar data, the variogram estimate must be a valid variogram. For the plug-in method, the variance of $\hat{\alpha}$ requires the aggregation of successive squared second-differences of the variogram. Since the empirical variogram is not guaranteed to be a valid variogram, a valid variogram estimate is needed. This is obtained by fitting a suitable variogram model to the empirical variogram.

For a variogram model to suit our purposes, we require that it be general enough to cover the full range of data sets to which the methodology may be applied, not just to a limited number of special cases. However, many of the existing variogram models in current use fail this criterion, because they allow only a small (finite) number of values for the rate of expansion near the origin, α. The exception is the power law model,

$$v(t; c, \alpha) = c|t|^{\alpha}, \tag{4.7}$$

which for a Gaussian process implies fBm.

This is a useful model as it is exactly the approximation upon which our methods are based. However, as mentioned earlier, as α approaches 2, the correlation across lags becomes so great that sample paths of such processes will tend to a straight line. If it is our aim to make the methodology cover a wider class of data sets, for example to include smoother but non-monotonically varying transects, then a better model would be one that allows the correlation to decay over larger lags but retains the behaviour near the origin. Such a model is obtained from the *stable exponential* covariance function, $\gamma(t; \sigma, \lambda, \alpha) = \sigma^2 \exp(-\lambda|t|^{\alpha})$, which implies the following model for the variogram:

$$v(t; \sigma, \lambda, \alpha) = 2\sigma^2[1 - \exp(-\lambda|t|^{\alpha})]. \tag{4.8}$$

Here $2\sigma^2$ represents the variogram sill.

Notice that if the product $2\sigma^2\lambda$ remains constant while σ increases indefinitely, then in the limit the stable exponential variogram (4.8) is the power law variogram (4.7) with $c = 2\sigma^2\lambda$. So the power law model can be treated as an extreme case of the stable exponential model. In practice then, (4.8) may be used as the initial model and if, after fitting the model, the estimate of σ is an order of magnitude greater than $\hat{c}$, then (4.7) may be an adequate model over the range of the data.

As noted at the end of 4.2.1, for $\alpha < 3/2$ the variance properties of $\hat{\alpha}$ and $\hat{c}$ depend mostly on the behaviour of the variogram near the origin. Since (4.7) and (4.8) behave similarly in this neighbourhood, choice of which model to fit will provide only a second-order improvement in estimation of variance. Therefore, in this case, (4.7) may be preferable for computational ease and efficiency.

For $\alpha > 3/2$, it is important to provide a reasonable model for the underlying variogram over the whole range of the data. Nevertheless, particular emphasis should be placed on fitting the variogram near the origin, since this is the area of main concern as well as that of least variability in the empirical variogram. Such emphasis may be achieved by appropriate weighting in the fitting criterion. Choosing optimal weights is difficult to achieve due to an implied recursion: the variance of the empirical variogram at a specific lag has a major contribution to the corresponding weight for that lag, and this variance is precisely the quantity we are trying to estimate using the weights. To some extent, this problem can be ameliorated by fitting the model on a log scale, as noted at the end of section 4.1.4, and by using as weights the numbers of increments used to obtain the empirical variogram at different lags.

However, there is another problem in fitting (4.8). That model is no longer easily transformed into a linear model, as is (4.7), and hence simple linear methods may not be used to obtain parameter estimates directly. Indeed, the model (4.8) is so highly non-linear, that even non-linear minimisation algorithms have convergence problems.

Heuristic fitting procedure To overcome the problems outlined above, and as an alternative to other minimisation procedures, a heuristic iterative procedure tai-

lored to fitting (4.8) was employed. The procedure involves estimating the model parameters separately and then iteratively adjusting them until a "reasonable" fit is obtained to the empirical variogram. In the rest of this chapter, the fitted variogram models will be denoted by $\tilde{v}$ and model parameter estimates will also be accented by the tilde: $\tilde{\alpha}, \tilde{\lambda}$ and $\tilde{\sigma}$.

The most difficult parameter to estimate consistently is σ^2, the variance (if it exists) of the surface. This was highlighted in chapter 2 where the performance of R_q, an estimator of σ, was compared to that of $\hat{D}$ (or equivalently $\hat{\alpha}$) and $\hat{c}$. Its poor performance may be attributed to the relative lack of relevant information in a transect of data. If it is assumed that correlation decays sufficiently fast, then only the differences of the process over a smaller number of large lags may provide consistent information for estimating σ. In contrast, a larger number of increments over small lags can be used to obtain estimates for α and c. To compound the problem, R_q is a weighted average of the empirical variogram in which smaller lags are given a higher weighting than larger lags; see (2.1) on page 19.

The estimator we propose for σ^2 is the median of the empirical variogram over its range. This provides some adjustment for the incorrect weighting and is more robust against the wild fluctuations of the empirical variogram. It is also equivariant under monotone transformations of the empirical variogram, such as taking the logarithm. So we take

$$\tilde{\sigma}^2 = \tfrac{1}{2}\,\mathrm{median}_t\{\hat{v}(t)\}.$$

as the estimator of surface variance.

The initial estimates for α and λ in (4.8) are obtained from the asymptotic expansion of $v(t;\alpha,\lambda,\sigma)$:

$$v(t;\alpha,\lambda,\sigma) = 2\lambda\sigma^2|t|^\alpha + O(|t|^{2\alpha}),$$

as $t \to 0$. Thus,

$$\tilde{\alpha} = \hat{\alpha} \qquad \text{and} \qquad \tilde{\lambda} = \hat{c}/(2\tilde{\sigma}^2)$$

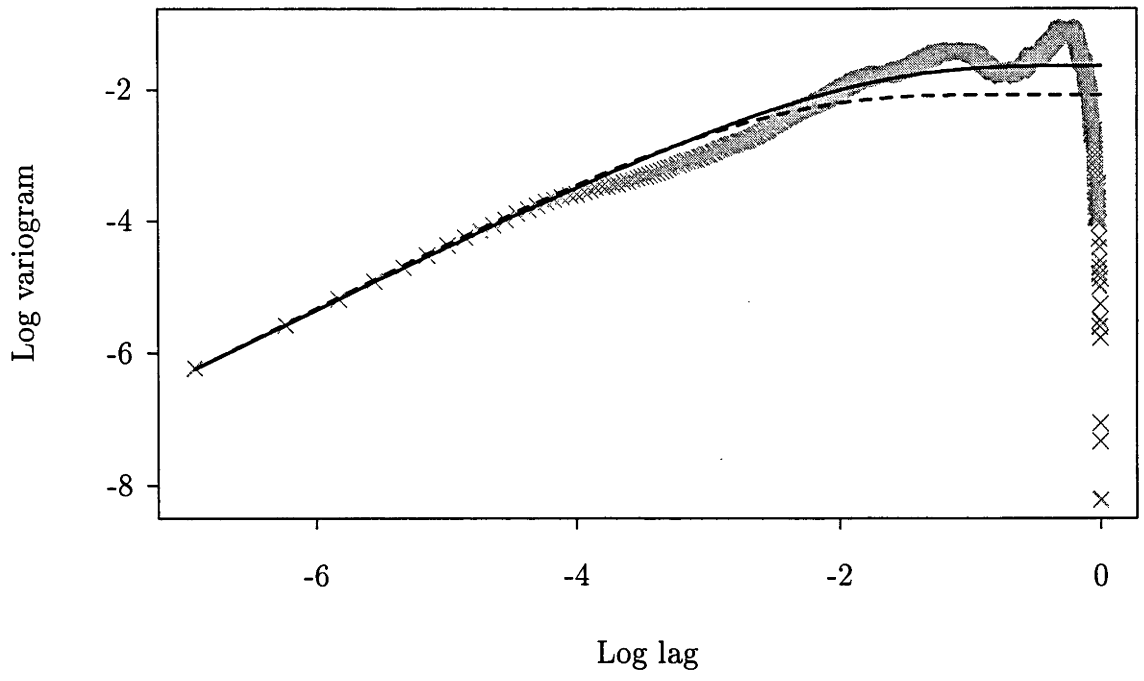


Figure 4.3: Fit of the stable exponential variogram model, using the heuristic fitting procedure described in the text, to the empirical variogram (marked with $\times$'s) of simulated data. The solid line depicts the fitted model and the dashed line represents the underlying variogram from which the data were simulated. The heuristic method fits closely in the most important region, near the origin, while also capturing the decay in correlation at higher lags.

These parameter estimates define a curve that is asymptotically equal to $\hat{c}|t|^{\hat{\alpha}}$. However, if the discretisation of the data is not fine enough, it may depart appreciably from the empirical variogram because of the size of the bias inherent in $\hat{\alpha}$ and $\hat{c}$. Therefore, an adjustment to parameter estimates is required so that expected values of the estimates of $\hat{\alpha}$ and $\hat{c}$, from data generated from the $\tilde{v}(\cdot)$, are close to themselves.

This is done using the following iterative procedure. Firstly estimate the expected values of $\hat{\alpha}$ and $\hat{c}$ from $\tilde{v}(\cdot)$ by

$$\hat{\alpha}_e = \sum a_l \log v(l/n; \tilde{\sigma}, \tilde{\lambda}, \tilde{\alpha}) \qquad \text{and} \qquad \log \hat{c}_e = \sum b_l \log v(l/n; \tilde{\sigma}, \tilde{\lambda}, \tilde{\alpha}).$$

Then update $\tilde{\alpha}$ and $\tilde{\lambda}$ by assigning

$$\tilde{\alpha} := \tilde{\alpha} + \hat{\alpha} - \hat{\alpha}_e \qquad \text{and} \qquad \tilde{\lambda} := \tilde{\lambda}\frac{\hat{c}}{\hat{c}_e}.$$

[The symbol ":=" is used for assignment to differentiate it from the notion of equality.] These two bias-correcting steps are repeated until $\hat{\alpha} - \hat{\alpha}_e$ is small. Usually only a few iterations are required.

Figure 4.3 is a log-log plot of the empirical variogram, calculated from a simulated data set with a sample size of 1024. The solid curve represents the fitted model after three iterations of the fitting procedure. The dashed curve is the actual variogram from which the data were simulated: $v(\cdot\,; 1, 16, 2^{-1})$.

4.3.1 Model validation

To supplement the fitting procedure, a method for model validation is described below. In practical situations, this will allow us to make inferences about the fitted model before using the model in further analysis. As an example, consider the situation outlined above in which the relative values for parameter estimates of c, λ and σ^2, obtained from fitting the stable exponential model, imply that the power model is more appropriate. Some model validation may be used to assess this. Indeed, since the high correlation between lags has been largely ignored in fitting the model, the actual fitted model may also be unrealistic. Validation may help here, too.

The method of model validation we propose is graphical. It seeks to provide a visual check that a proposed model is not unreasonable for the data. This is achieved by calculating the range of variogram estimators that might be expected from the model, and then using extreme values of these to define a set of simultaneous intervals, or a confidence set. The actual variogram estimate obtained from the original data can then be compared with these bands. If it crosses the bands then the model would appear to be implausible; see Figure 4.4.

Since the bands are simultaneous intervals of the variogram estimator over all possible lags, the procedure requires a simultaneous multivariate test. One method

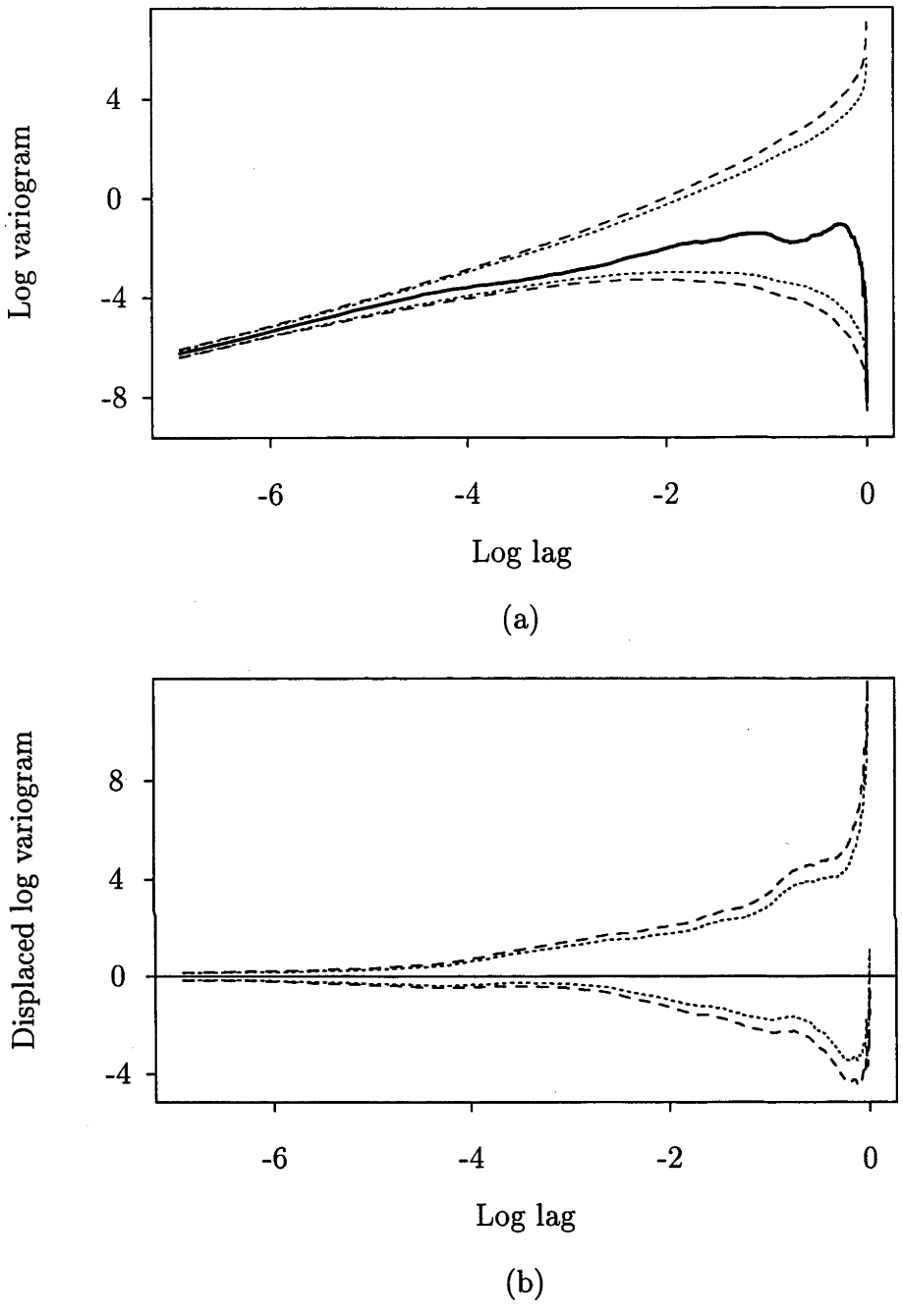


Figure 4.4: Critical regions at the 95% level (dotted line) and 99% level (dashed line) to test whether a power model is reasonable for the simulated data in Figure 4.3. In the top panel, it is difficult to see whether the empirical variogram crosses either of the lower critical boundaries because of the steepness of the curve at large lags. The problem is removed in the bottom panel by subtracting the empirical variogram from all curves. Then if any of the boundaries crosses the horizontal line at 0, the departure of the data from the model is (highly) significant.

of constructing such a test is by using approximations to produce Bonferroni-style intervals. The problem with this is that it neglects the large amount of correlation across lags. As an alternative, the proposed test reduces the simultaneous multivariate test to a univariate test based on an order statistic of the rescaled residuals of the empirical variogram about its mean.

Formally, the model validation utilises the test of the null hypothesis,

$$H_0 : X \sim \text{IGRP}(\tilde{v}) \qquad \text{against} \qquad H_1 : X \nsim \text{IGRP}(\tilde{v}),$$

where $\text{IGRP}(\tilde{v})$ is an intrinsic stationary Gaussian random process with variogram $\tilde{v}(\cdot)$. Although the emphasis is on testing whether $\tilde{v}(\cdot)$ is an appropriate underlying model for the data, differences in other assumptions, such as Gaussianity or intrinsic stationarity, may also cause rejection. However, the amount of averaging implicit in the empirical variogram will dampen the sensitivity to such departures.

Define an acceptance region Ω_β by

$$\mathbb{P}\left[(\hat{v}_0[t_1], \ldots, \hat{v}_0[t_m])' \in \Omega_\beta\right] < 1 - \beta,$$

or equivalently, on the log scale and for a different Ω_β, by

$$\mathbb{P}\left[(\log \hat{v}_0[t_1], \ldots, \log \hat{v}_0[t_m])' \in \Omega_\beta\right] < 1 - \beta, \tag{4.9}$$

where $\hat{v}_0(t)$ is an estimator for the variogram under H_0. As it stands, (4.9) does not define Ω_β uniquely. Given a certain alternative hypothesis, it may be possible to derive a unique Ω_β to obtain a most powerful test against that alternative. However, for the practical purpose of being able to visualise the region Ω_β graphically, Ω_β is constrained in shape to be the product of simultaneous univariate intervals: $\Omega_\beta = (\omega_\beta^L[t_1], \omega_\beta^U[t_1]) \times \cdots \times (\omega_\beta^L[t_m], \omega_\beta^U[t_m])$, where $\omega_\beta^L(t)$ and $\omega_\beta^U(t)$ are the lower and upper limits of the interval at t. Thus, (4.9) becomes

$$\mathbb{P}\left[\omega_\beta^L(t) < \log \hat{v}_0(t) < \omega_\beta^U(t), \text{for } t = t_1, \ldots, t_m\right] < 1 - \beta,$$

and, if the intervals are centred about $\text{E}\{\log \hat{v}_0(\cdot)\}$, it becomes

$$\mathbb{P}\left[|\log \hat{v}_0(t) - \text{E}\{\log \hat{v}_0(t)\}| < \omega_\beta(t), \text{for } t = t_1, \ldots, t_m\right] < 1 - \beta. \tag{4.10}$$

Note that $\omega.(t)$ is a function whose form is still to be chosen. Since the test is a simultaneous multivariate test, one way to establish a form for ω_β is to construct equi-probable Bonferroni-style intervals by combining the m separate univariate regions defined by

$$\mathbb{P}\left[|\log \hat{v}_0(t) - \mathrm{E}\{\log \hat{v}_0(t)\}| < \omega_\beta(t)\right] < 1 - m^{-1}\beta$$

for $t = t_1, \ldots t_m,$. Thus, all other things being equal, each interval has an equal chance of failing to cover the empirical variogram. Unfortunately, because of their conservative nature Bonferroni intervals are too conservative for correlated data, providing better approximation for independent data. Indeed, the approximation also deteriorates as the number of simultaneous variables increases.

However, the property of equalising the sensitivity of the overall test across lags is a good one, and may be used in (4.10) by rescaling deviations from the expected value by the standard error, to obtain

$$\omega_\beta(t) = \zeta_\beta \,\mathrm{var}\{\log \hat{v}_0(t)\}^{1/2}.$$

In this form, the simultaneous multivariate test is reduced to a single-valued test, that of finding the value for ζ_β for which

$$\mathbb{P}\left[z < \zeta_\beta\right] < 1 - \beta,$$

where z is the random variable defined by

$$z = \max_{t=t_1 \ldots t_m} \frac{|\log \hat{v}_0(t) - \mathrm{E}\{\log \hat{v}_0(t)\}|}{\mathrm{var}\{\log \hat{v}_0(t)\}^{1/2}}.$$

The distribution of z is complex, depending very much on the correlation across lags of the variogram estimator and other underlying properties. So the following Monte Carlo method, similar to the bootstrap method for estimating variances in section 4.2.2, is employed to estimate ζ_β.

Step 1: Simulate a number, B say, of realisations of Gaussian random fields with $\tilde{v}(\cdot)$ as their variogram. Identify each as X_b^*.

Step 2: For all simulated processes X_b^*, calculate their respective empirical variograms $\hat{v}_b^*(\cdot)$.

Step 3: Estimate $\mathrm{E}\{\log \hat{v}_0(t)\}$ and $\mathrm{var}\{\log \hat{v}_0(t)\}$ by

$$m(t) = \frac{1}{B}\sum_{b=1}^{B} \log \hat{v}_b^*(t) \qquad \text{and} \qquad s(t)^2 = \frac{1}{B-1}\sum_{b=1}^{B} [\log \hat{v}_b^*(t) - m(t)]^2,$$

respectively, for $t = t_1, \ldots, t_m$.

Step 4: Put

$$z_b^* = \max_{t=t_1,\ldots,t_m} \frac{|\log \hat{v}_b^*(t) - m(t)|}{s(t)}$$

Step 5: Order the z_b^*'s: $z_{[1]}^* \leq \cdots \leq z_{[B]}^*$. Then $\hat{\zeta}_\beta = z_{[(1-\beta)B]}^*$ and

$$\Omega_\beta = \prod_{t=t_1,\ldots,t_m} (m(t) - \hat{\zeta}_\beta s(t),\ m(t) + \hat{\zeta}_\beta s[t])$$

is a nominal $100(1-\beta)\%$ acceptance region for the variogram estimator. If the variogram estimator at any lag falls outside the corresponding interval for that lag then the null model is rejected at the $100(1-\beta)\%$ level.

It should be stressed that the above method does not use the bootstrap. In a bootstrap test it is necessary to approximate the null distribution. Here the null distribution is known, and in principle there need be no systematic difference between the nominal significance level and the actual significance level. However, there will be a difference between nominal and actual significance levels, depending on the number of Monte Carlo simulations, B. For $B = 1000$, the difference is small.

The test was applied to the simulated data shown in Figure 4.3, to see whether the power law model, obtained by using $\hat{\alpha}$ and $\hat{c}$ calculated from the data as model parameters, was an appropriate underlying model for the data. We performed 1000 simulations in the Monte Carlo procedure and resulting critical bands were calculated at the 95% and 99% levels. These are shown in panel (a) of Figure 4.4 along with the empirical variogram from the simulated data shown in Figure 4.3. Due to the steepness of the curves at high lags, it is difficult to ascertain whether the

empirical variogram crosses either of the critical boundaries. To overcome this problem panel (b) shows all curves with the empirical variogram subtracted; thus the horizontal line represents the empirical variogram. Since the 95% boundary crosses the horizontal line, the null power model is rejected as being unsuitable.

4.4 Effects of measurement error

In practice, virtually all measurements of a continuous process are smoothed to some extent during recording. A common theoretical approximation to the smoothing is that the measurement at each location is obtained by averaging uniformly over a region about that location. This is indeed the case for optical profilometers. For other instruments, the average is often taken with respect to a *point spread function* whose shape more closely resembles a circularly symmetric bivariate Normal density than it does a uniform density over a square. The shape of the point spread function is sometimes known quite accurately.

The observed, degraded signal may be modelled as

$$Y(t) = \int X(t-s)H(s)ds,$$

where $H(\cdot)$ is the point spread function. It is assumed that $\int H(s)ds = 1$, so that it preserves average elevation. $H(\cdot)$ represents a spatial average over two dimensions, but we shall concentrate on the one-dimensional case to identify the main effects, as the two-dimensional case is similar.

Let $v_X(\cdot)$ and $v_Y(\cdot)$ be the respective variograms of X and Y. Then

$$v_Y(t) = \int v_X(t-s)K(s)ds - \int v_X(s)K(s)ds, \tag{4.11}$$

where $K(t) = \int H(s-t)H(s)ds$ is the convolution of $H(\cdot)$ with the reflection of itself. Hence $\int K(s)ds = 1$, $K(\cdot)$ is symmetric, and if $H(\cdot)$ is compactly supported then so is $K(\cdot)$.

Note that the variogram of Y is not only a smoothed version of the variogram of X, but that it is also shifted so that the property $v_Y(0) = 0$ is maintained. We shall now examine the two effects separately.

4.4.1 Effects of smoothing the variogram

Firstly, consider the first integral in (4.11). By Taylor expanding $v_X(t-s)$ about $v_X(t)$ inside the integral it may be shown that, for t outside the support of $K(\cdot)$,

$$\int v_X(t-s)K(s)ds = v_X(t) + R_1(t)$$

where $|R_1(t)| < \frac{1}{2}\kappa_2 v_X''(t)$, and $\kappa_2 = \int s^2 K(s)ds$. Derivatives of odd order disappear since $K(\cdot)$ is symmetric. The bound on $R_1(t)$ assumes that $v_X''(t)$ is decreasing.

For t inside the support of $K(\cdot)$ the effects of smoothing are overwhelming in that any analysis of $v_Y(t)$ as t approaches 0 will relate to properties of $K(t)$ rather than $v_X(t)$. So we recommend that the variogram be ignored for lags at least up to the width of the support of $H(\cdot)$. We shall assume for the rest of this section that t is outside the support of $K(\cdot)$.

If we are to make this assumption and look at the behaviour of the smoothing as $n \to \infty$ then we need to make the corresponding assumption that the support for $K(\cdot)$ is decreasing commensurately. In practice, it is of little value to make measurements on progressively finer grids below the support of the point spread function. So we set $K_n(s) = nK(ns)$, which also preserves the property $\int K_n(s)ds = 1$, and look at the asymptotic behaviour of $\int v_X(j/n - s)K_n(s)ds$.

Assuming the approximate power laws (2.4) and (3.5) for $v_X(t)$ and $v_X''(t)$ respectively,

$$\int v_X(j/n - s)K_n(s)ds = n^{-\alpha}c|j|^\alpha \left\{1 + \frac{R_1(j)}{c|j|^\alpha} + o(1)\right\},$$

as $n \to \infty$. The ratio of $R_1(j)$ to $c|j|^\alpha$ is less in magnitude than $\kappa_2\alpha(\alpha-1)|j|^{-2}/2$, which decreases rapidly with increasing j.

EXAMPLE 4.1 Consider the extreme case of a uniform point spread function. Here, Y is the average of X over a finite support; that is, $H(t) = 1/\zeta$ for $|t| < \zeta/2$ and $H(t) = 0$ elsewhere. Then $K(t) = \zeta^{-1}(1-\zeta^{-1}|t|)$ for $|t| < \zeta$ and $K(t) = 0$ elsewhere.

The quotient of the smoothed $v_X(t)$ to $v_X(t)$ itself can be obtained directly at multiples of ζ:

$$\frac{\int_{-\zeta}^{\zeta} v(j\zeta - s)K(s)ds}{v(j\zeta)} = \frac{(j-1)^{\alpha+2} - 2j^{\alpha+2} + (j+1)^{\alpha+2}}{(\alpha+1)(\alpha+2)j^\alpha}. \qquad (4.12)$$

By Taylor-expanding the numerator on the right-hand side of (4.12), we may obtain the following bound for the remainder:

$$\frac{R_1(j\zeta)}{v(j\zeta)} < \frac{\alpha(\alpha-1)}{12} j^{-2} < \frac{1}{6} j^{-2}.$$

Thus, the contribution due to smoothing is at most one-sixth when the variogram is calculated at the width of the point spread function, and rapidly decreases thereafter. To cater for these effects, an extra term may be added to the fitted model, although this has the disadvantage of making the model non-linear. Alternatively, since the contribution is less than 0.05 at twice the support, it may be better to continue to use the linear model but to ignore variogram estimates below this point.

This example was for a uniform point spread function. For other point spread functions that are more concentrated at the centre of their support, the contribution of the remainder term due to smoothing will be less.

4.4.2 Effects of zeroing the smoothed variogram

When calculating the variogram of Y, a more important factor than variogram smoothing is the contribution of the shift in (4.11) by a constant term, $C = \int v_X(s)K(s)ds$, since this term does not diminish with increasing t. Therefore we need to take its effects into account either by updating our fitted model to include the constant term, or by modifying our approach to mitigate its effects. We explore both avenues here.

Non-linear least squares One updated model containing the constant term has the form $v(t) \sim c|t|^\alpha - C$, which is no longer easily transformed into a linear model by taking logarithms. However, it is still preferable to fit this model on the log scale because, as noted in chapter 3, taking logarithms provides some variance stabilisation. Thus, to fit this model to the empirical variogram we can use a general algorithm to minimise the least squares criterion,

$$\sum_{l=1}^{k} \{\log \hat{v}_Y(h_l) - [\log c + \alpha \log |h_l| + \log(1 - C/c|h_l|^\alpha)]\}^2.$$

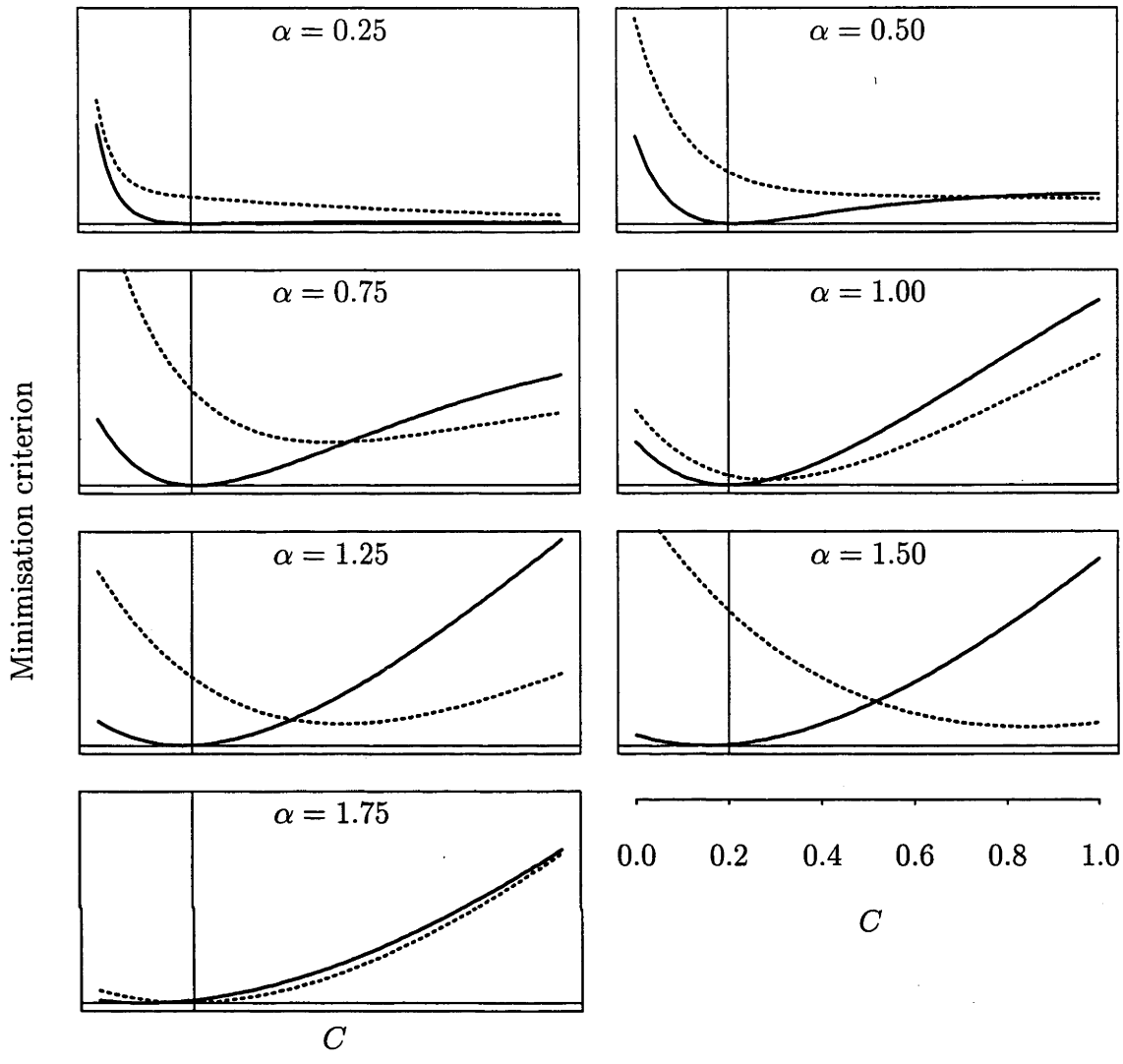


Figure 4.5: Graphs showing least squares criteria for fitting the best line to the logarithm of *offset* empirical variogram estimates against logarithm of lag, as described in the text. The minimum of each curve provides an estimate of the correction required to reduce the measurement effects on the variogram of smoothing the data. The curves in each panel are the criteria calculated for processes that have been uniformly locally averaged, thereby reproducing the effects of measurement error. Each panel is the result of calculations for processes from different underlying fBm models with the values for α as indicated. The solid line in each panel represents the expected criterion and the dotted line represents the criterion as calculated from a single realisation of a locally averaged process. The horizontal line is at 0 and the vertical line is at the actual offset caused by the averaging.

A similar criterion, but one that preserves some linearity, is

$$\sum_{l=1}^{k}[\log(\hat{v}_Y(h_l) + C) - (\log c + \alpha \log|h_l|)]^2. \tag{4.13}$$

For a given C, (4.13) is minimised when α and $\log c$ take on the values

$$\hat{\alpha} = \sum_{j=1}^{k} a_j \log(\hat{v}_Y(h_j) + C) \qquad \text{and} \qquad \log \hat{c} = \sum_{j=1}^{k} b_j \log(\hat{v}_Y(h_j) + C),$$

respectively. After substituting these for α and c in (4.13), performing some algebraic manipulation, and noting that $\sum a_j = 0$ and $\sum b_j = 1$, (4.13) becomes

$$\sum_{l=1}^{k}\left\{\sum_{j=1}^{k}(b_j + a_j \log|h_l|)\log\frac{\hat{v}_Y(h_l) + C}{\hat{v}_Y(h_j) + C}\right\}^2,$$

thus transforming the minimisation from a search over three variables to one over a single variable, C.

In trying to minimise this function, one problem with the formulation becomes apparent. As C becomes large, the criterion approaches its theoretical minimum of 0. This general decline towards 0 may remove any local minimum. The solid lines in Figure 4.5 depict the minimisation criterion for the deterministic case, in which $v_X(t) = |t|^\alpha$ for $\alpha = 0.25(0.25)1.75$. The dotted line in each panel represents the minimisation criterion applied to the empirical variogram $\hat{v}_Y(\cdot)$ as calculated from a realisation of a process with underlying variogram $v_Y(\cdot)$. Here, $v_Y(t)$ is calculated from $v_X(t) = |t|^\alpha$ using (4.11) and assuming measurement error in the form of uniform local averaging. Notice how the local minima are completely removed for $\alpha = 0.25$ and 0.50. Notice also that the location of actual minima can be dramatically different from their true values.

Taking differences Alternatively, the effects of shifting the variogram by a constant may be removed by taking first-order differences of the variogram. An approximate model for these differences can then be obtained by Taylor-expanding the difference of $v[(j+1)/n] - v(j/n)$ about j, to yield

$$v_Y[(j+1)/n] - v_Y(j/n) = n^{-\alpha} c\alpha |j|^{\alpha-1}[1 + R_2(j)]$$

where $|R_2(j)| < |(\alpha-1)j^{-1}/2|$. In fitting models to the differences, the contribution from $R_2(j)$ is too large to ignore, as it stands. However, the contribution may be reduced by expanding the same difference about $(j+\frac{1}{2})/n$ instead of j/n. Then,

$$v_Y[(j+1)/n] - v_Y(j/n) = n^{-\alpha}c\alpha|j+\tfrac{1}{2}|^{\alpha-1}[1+R_3(j+\tfrac{1}{2})]$$

where $|R_3(j+\frac{1}{2})| < |(\alpha-1)(\alpha-2)j^{-2}/4!|$.

This approximate power law leads to the following linear estimators of α and c:

$$\hat{\alpha} = 1 + \left[\sum_{l=1}^{k-1}(x_{l+1/2}-\bar{x})\log\{\hat{v}[(l+1)/n]-\hat{v}(l/n)\}\right] \Big/ \left[\sum_{l=1}^{k-1}(x_{l+1/2}-\bar{x})^2\right] \tag{4.14}$$

and

$$\hat{c} = n\hat{\alpha}^{-1}\exp\left\{(k-1)^{-1}\sum_{l=1}^{k-1}\log\{\hat{v}[(l+1)/n]-\hat{v}(l/n)\} - (\hat{\alpha}-1)\bar{x}\right\}, \tag{4.15}$$

where $x_l = \log(l/n)$ and $\bar{x} = (k-1)^{-1}\sum x_{l+1/2}$.

A major difference between these estimators and those introduced at the beginning of the chapter is the instability of the logarithm of a difference when applied to random variables such as $\hat{v}[(l+1)/n]$ and $\hat{v}(l/n)$. Indeed, the stochastic error in either of these can give rise to a negative difference, despite their strong positive correlation. So, in practical terms, the estimators in (4.14) and (4.15) will require very little stochastic error, and hence very large sample sizes, to be effective.

4.5 Analysis of the roller data

The discussion in section 4.4 on measurement error is very pertinent to the roller data since the stylus used to record the data had a nominal width equal to the width of digitisation. Although the stylus does not perform local averaging as an optical profilometer would, its shape and other factors such as load will cause smoothing of the profile and therefore measurement error similar to that discussed. Following the recommendations of section 4.4 the empirical variogram widths up to the width of

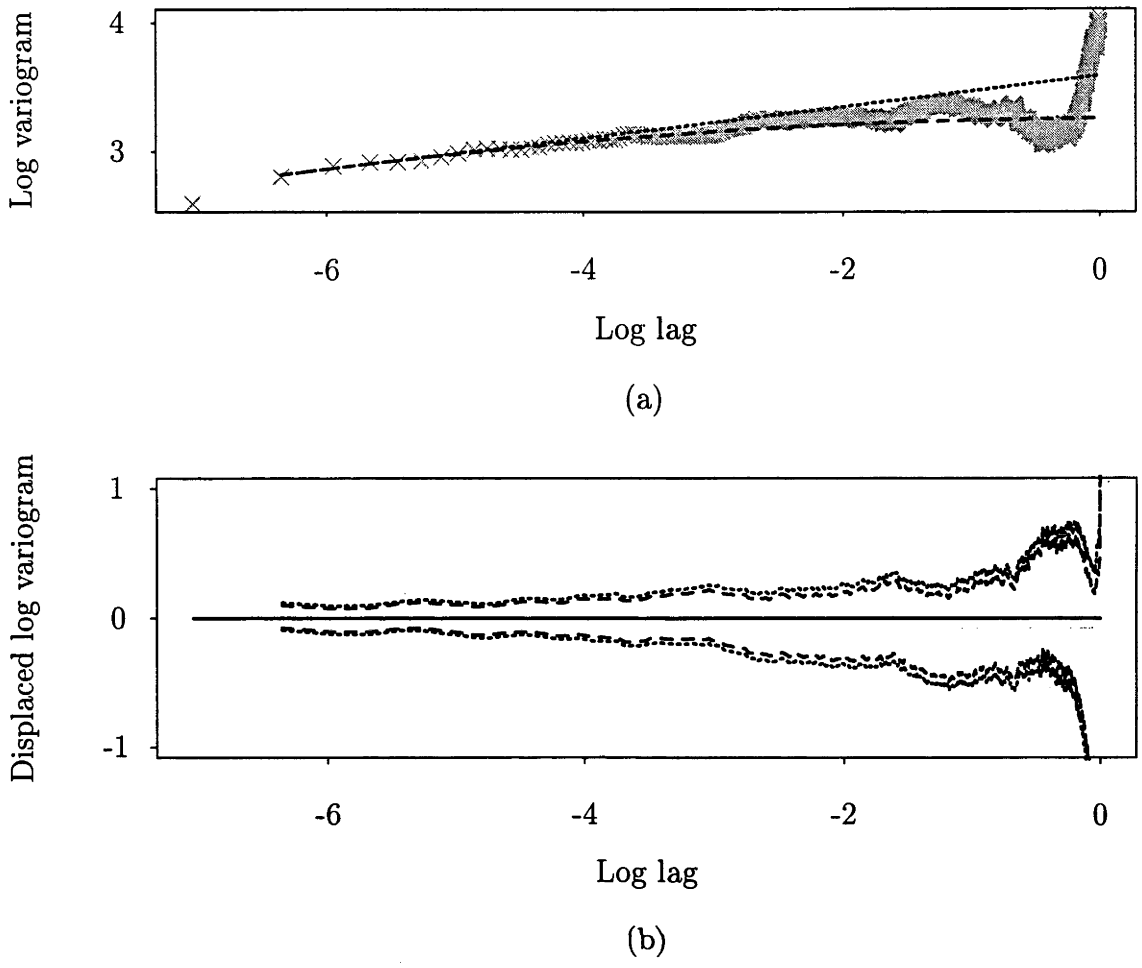


Figure 4.6: (a) The results of fitting the power model (dotted line) and the stable exponential model (dashed line) to the empirical variogram of the roller data excluding the initial lag, using the methods described in the text. (b) Implied critical region bands at the 95% (dashed line) and 99% (dotted line) levels to assess whether the fitted power model in (a) is a reasonable underlying model for the roller data. The bands have been translated to remove the empirical variogram, to aid interpretability. Since the bands do not cross the solid horizontal zero-line, there is not enough evidence to reject the fitted power model.

the stylus will be excluded from the analysis; for the roller data, this width is the first lag.

Thus, $\hat{\alpha}$ and $\hat{c}$ were calculated similarly to (4.1), but with l ranging from 2 to $k+1$, where k was taken to be 4, the value indicated in section 4.1.3. Point estimates

for $\hat{\alpha}$ and $\hat{c}$ are given in Table 4.1. The resultant estimated fractal dimension $\hat{D}$ is 1.88, a value close to the upper limit of 2, suggesting that this profile of the surface is extremely erratic.

	$\hat{\alpha}$	$\hat{c}$
	0.242	0.131
Plug-in s.e.	(0.046)	(0.036)
Bootstrap s.e.	(0.049)	(0.040)

Table 4.1: Estimated fractal index and topothesy for the roller data, with estimates of standard error from the plug-in and bootstrap methods. The corresponding estimate of fractal dimension $\hat{D}$ is 1.88, indicating a highly erratic profile.

$\tilde{\alpha}$	$\tilde{\lambda}$	$\tilde{\sigma}$
0.242	0.131	0.123

Table 4.2: Estimated model parameters from fitting the stable exponential variogram model to the empirical variogram of the roller data.

The power model for the variogram implied by these estimates is shown as the dotted line in panel (a) of Figure 4.6. It can be seen that this model agrees with the data over a larger set of lags than those used in estimation. Of course, this is partly because of the correlation across lags, but the effects of this correlation are smaller for the small values of α that the data appear to exhibit. So, the agreement may also suggest that the power model is a fairly good approximation over a wider range of scales than the digitisation level. Indeed, the model validation procedure of section 4.3.1 was used to test whether the fitted power model was a reasonable underlying model for the data. The critical bands, displaced by the empirical variogram of the roller data, are shown in panel (b) of Figure 4.6. Since

they do not intersect with the zero-line, it is not reasonable to reject the power model on this basis.

In order to obtain estimated standard errors for $\hat{\alpha}$ and $\hat{c}$, the stable exponential model for the variogram was fitted, yielding estimates of the model parameters given in Table 4.2. The resultant fit is shown as the dashed curve in panel (a) of Figure 4.6. This model was used to obtain standard errors for parameter estimates using both the plug-in and bootstrap methods described in the previous section. These standard errors are given in Table 4.1.

Chapter 5

Isotropic surfaces

This chapter deals with estimation of fractal dimension and topothesy measures that characterise roughness of two-dimensional isotropic surfaces. The models and methods presented are a direct extension of those for one-dimensional profile data described in chapter 4.

Although the one-dimensional methods translate directly to the higher dimensional case, the performances of estimators differ. It is shown that the critical value of fractal index at which the asymptotic behaviour of estimators changes is 1 for the two-dimensional isotropic case, whereas it was 3/2 for the one-dimensional case. The critical point is the value below which estimators are root-n consistent and asymptotically Gaussian. These properties are derived for the isotropic estimators, and numerical evidence is given to support these results.

Methods for estimating standard errors of estimators, and for constructing confidence intervals for parameters, are given. A box-counting method for estimating the fractal dimension of two-dimensional data is also derived, and a comparison is made with the variogram method.

The methods are used to analyse the soil surface data described in chapter 1.

5.1 Extension from one dimension

The simplest way to extend the methods for one-dimensional data to two dimensions is to assume that the underlying two-dimensional process is *isotropic*. In this case, the question of how the variograms of differently oriented transects might vary is ignored.

The validity of the assumption of isotropy can be checked using numerical or graphical evidence, for example the almost-circular contours of the two-dimensional empirical variogram for the wettest soil surface as shown in figure 3.3. Alternatively, there may be physical considerations that justify it. For example, the soil surfaces were artificially constructed and the experiment was designed not to favour any specific direction, nor did the artificial rainfall that was used in a controlled laboratory environment. Nonetheless, in the absence of *a priori* physical justification, it may be prudent to conduct some confirmatory analysis of isotropy to strengthen inference made purely from graphs of the empirical variogram or parameter estimates. A hypothesis test for isotropy against the alternative of weak anisotropy is given in chapter 6. (It is similar to the Monte Carlo test presented in section 4.3.1.) A process is said to be *weakly anisotropic* if there exists a linear transformation of its domain that will make the transformed process isotropic.

Suppose that the two-dimensional process X is isotropic. Then its underlying variogram at displacement $\boldsymbol{t}$ depends only on lag and not on orientation:

$$v(\boldsymbol{t}) = v(\|\boldsymbol{t}\|), \qquad \text{for all } \boldsymbol{t} \in \mathbb{R}^2. \tag{5.1}$$

Here, notation is abused by using $v(\cdot)$ to represent both two-dimensional and one-dimensional variograms, each being distinguished by its argument and context. Thus, in (5.1), the expression on the l.h.s. is the value of a two-dimensional variogram at displacement $\boldsymbol{t}$, and the expression of the r.h.s. $v(r)$ is the value of a one-dimensional variogram, $v(h)$, at lag $h = \|\boldsymbol{t}\|$.

From (5.1), the assumption concerning the behaviour of the variogram near the origin, corresponding to the one-dimensional counterpart (2.2), is that

$$v(\boldsymbol{t}) = c\|\boldsymbol{t}\|^\alpha + o(\|\boldsymbol{t}\|^\alpha), \qquad \text{as } \boldsymbol{t} \to \mathbf{0}, \tag{5.2}$$

whence,

$$\log v(\boldsymbol{t}) = \log c + \alpha \log \|\boldsymbol{t}\| + o(1) \qquad \text{as } \boldsymbol{t} \to \mathbf{0}. \tag{5.3}$$

This ensures that the fractal index α is well-defined; see the discussion on fractal methods in section 2.3. If the process X is sufficiently related to a Gaussian field (Hall & Roy, 1994), then the realisations of X have fractal dimension

$$D = 3 - \tfrac{1}{2}\alpha.$$

Estimators for D are obtained by plugging in an estimator for α and, as with the one-dimensional methodology, estimators for α and c may be obtained from the slope and intercept of a linear regression of the logarithm of estimated variogram against the logarithm of lag.

Since the variogram is independent of orientation then we need only provide an estimate for the one-dimensional $v(\|\boldsymbol{t}\|)$. The appropriate naive estimator averages the squared interpoint difference of a process over all pairs of points at a lag $h = \|\boldsymbol{t}\|$ apart. So, if $C_I(h)$ is the set of pairs of grid locations lag h apart,

$$C_I(h) = \{(\boldsymbol{r}, \boldsymbol{s}) \in \mathcal{G} \times \mathcal{G} : \|\boldsymbol{s} - \boldsymbol{r}\| = h\},$$

then the *isotropic empirical variogram* is defined by

$$\hat{v}_I(h) = |C_I(h)|^{-1} \sum_{(\boldsymbol{r},\boldsymbol{s}) \in C_I(h)} \{x(\boldsymbol{s}) - x(\boldsymbol{r})\}^2,$$

and is an estimator of $v(\|\boldsymbol{t}\|)$.

Let $u_0 \le u_1 \le \cdots \le u_N$ be the ordered sequence of lags for which $\hat{v}_I(\cdot)$ is calculable, *i.e.* the u_l's are those values for which $|C(u_l)| \ge 1$. For convenience, the sequence starts at u_0 so that $u_0 = 0$. In the one-dimensional case, u_l was an integer multiple of u_1. Then α and c may be estimated by

$$\hat{\alpha}_I = \sum_{l=1}^{k} a_l \log \hat{v}_I(u_l) \qquad \text{and} \qquad \log \hat{c}_I = \sum_{l=1}^{k} b_l \log \hat{v}_I(u_l), \tag{5.4}$$

where $a_l = (\log u_l - \overline{\log u}) / \sum (\log u_l - \overline{\log u})^2$, $b_l = k^{-1} - a_l \overline{\log u}$ and $\overline{\log u} = k^{-1} \sum \log u_l$. Note that $\sum a_l = 0, \sum a_l \log u_l = 1, \sum b_l = 1$ and $\sum b_l \log u_l = 0$, as in the one-dimensional case.

Theoretical properties Although the estimators $\hat{\alpha}_I$ and $\log \hat{c}_I$ are expressed in a form similar to their one-dimensional equivalents $\hat{\alpha}$ and $\log \hat{c}$, their theoretical performance differs significantly. The main difference is a change in the critical value of α at which convergence slows and asymptotic distributional properties become non–Gaussian. Whereas this value was 3/2 for the one-dimensional case, we shall show that it is 1 for the isotropic two-dimensional case.

In order to obtain expressions for the asymptotic behaviour of $\hat{\alpha}_I$ and $\hat{c}_I$, it is necessary to be more explicit about the form of the $o(\|\boldsymbol{t}\|)$ term in (5.2). Thus, assume that

$$\log v(\boldsymbol{t}) = \log c + \alpha \log \|\boldsymbol{t}\| \{1 + d\|\boldsymbol{t}\|^{\beta} + o(\|\boldsymbol{t}\|^{\beta})\} \qquad \text{as } \boldsymbol{t} \to \boldsymbol{0}. \tag{5.5}$$

Then for two-dimensional data taken on an $n \times n$ grid covering a fixed area, the errors in $\hat{\alpha}_I$ and $\hat{c}_I$ depend on the fineness of the grid in the following way:

$$\hat{\alpha}_I - \alpha = (Rn^{\alpha}\lambda_{2D}(n) + Cn^{-\beta})\{1 + o_p(1)\} \tag{5.6}$$

and

$$\hat{c}_I - c = c(Rn^{\alpha}\lambda_{2D}(n) + Cn^{-\beta}) \log n\{1 + o_p(1)\} \tag{5.7}$$

as $n \to \infty$, where C is a constant and R is a random variable that is Gaussian for $0 < \alpha \leq 1$ and non-Gaussian for $1 < \alpha < 2$. Recall that $\lambda_{2D}(n)$ also depends on α:

$$\lambda_{2D}(n) = \begin{cases} n^{-1-\alpha} & \text{if } 0 < \alpha < 1 \\ n^{-2}(\log n)^{1/2} & \text{if } \alpha = 1 \\ n^{-2} & \text{if } 1 < \alpha \leq 2. \end{cases}$$

PROOF From the definition of $\hat{\alpha}_I$ at (5.4), $\hat{\alpha}_I$ may be broken into two components, one systematic and one random:

$$\hat{\alpha}_I = \sum a_l \log v(u_l) + \sum a_l \log\{1 + \{\hat{v}_I(u_l) - v(u_l)\}/v(u_l)\}.$$

Under assumption (5.5), this becomes

$$\hat{\alpha}_I = c\sum a_l + \alpha \sum a_l \log u_l + d\sum a_l |u_l|^{\beta}\{1 + o(1)\} + \sum a_l \log\{1 + \{\hat{v}_I(u_l) - v(u_l)\}/v(u_l)\},$$

as $u_l \to 0$. Now the first two summands are 0 and α, respectively, since $\sum a_l = 0$ and $\sum a_l \log u_l = 1$. Thus,

$$\hat{\alpha}_I = \alpha + d\sum a_l |u_l|^\beta \{1 + o(1)\} + \sum a_l \log\{1 + \{\hat{v}_I(u_l) - v(u_l)\}/v(u_l)\} \qquad (5.8)$$

as $u_l \to 0$.

For data collected on an $n \times n$ grid covering a fixed area, the u_l's are the ordered values of $\|\boldsymbol{t}\|$, for those values of $\boldsymbol{t}$ at which $\hat{v}(\boldsymbol{t})$ is calculable. Put $\boldsymbol{w} = n\boldsymbol{t}$ and $w_l = nu_l$. This implies that there is a $\boldsymbol{w}$ for which $\|\boldsymbol{w}\| = w_l$.

In chapter 3 it was shown that $\{\hat{v}(n^{-1}\boldsymbol{w}) - v(n^{-1}\boldsymbol{w})\} = R_{\boldsymbol{w}}\lambda_{2D}(n)\{1 + o_p(1)\}$ as $n \to \infty$, where $R_{\boldsymbol{w}}$ is a random variable, independent of n, whose distribution depends on the underlying variogram $v(\cdot)$.

Now $\hat{v}_I(u_l)$ is the average of the small finite number of $\hat{v}(\boldsymbol{t})$'s for which $\|\boldsymbol{t}\| = u_l$:

$$\hat{v}_I(u_l) = |C_I(u_l)|^{-1} \sum_{\|\boldsymbol{t}\|=u_l} |C(\boldsymbol{t})|\hat{v}(\boldsymbol{t}).$$

So,

$$\begin{aligned}\hat{v}_I(u_l) - v(u_l) =& |C_I(w_l/n)|^{-1} \sum_{\|\boldsymbol{w}\|=w_l} |C(n^{-1}\boldsymbol{w})| R_{\boldsymbol{w}}\lambda_{2D}(n)\{1 + o_p(1)\} \\ =& R_l\lambda_{2D}(n)\{1 + o_p(1)\},\end{aligned}$$

as $n \to \infty$, where R_l is the weighted combination of the correlated $R_{\boldsymbol{w}}$'s.

Updating (5.8) gives,

$$\hat{\alpha}_I = \alpha + d\sum a_l |w_l/n|^\beta \{1 + o(1)\} + \sum a_l \log\left[1 + \frac{R_l\lambda_{2D}(n)\{1 + o_p(1)\}}{c|w_l/n|^\alpha\{1 + o(1)\}}\right]$$

as $n \to \infty$, which, after Taylor–expanding the logarithm about 1, collecting n's and subtracting α from both sides, yields the desired result for $\hat{\alpha}$:

$$\hat{\alpha}_I - \alpha = Cn^{-\beta}\{1 + o(1)\} + Rn^\alpha\lambda_{2D}(n)\{1 + o_p(1)\}$$

as $n \to \infty$, where $C = d\sum a_l w_l^\beta$ and $R = c^{-1}\sum a_l R_l w_l^{-\alpha}$.

A similar result is obtained for $\log\hat{c} - \log c$, following the same path but substituting b_l for a_l. Whereas the a_l's are independent of n, the b_l's introduce a $\log n$ term, *viz.*

$$b_l = k^{-1} - a_l k^{-1}\sum \log u_l = k^{-1} - a_l k^{-1}\sum \log(w_l/n) = a_l \log n + O(1)$$

as $n \to \infty$. Thus,

$$\log \hat{c} - \log c = \{Cn^{-\beta} + Rn^{\alpha}\lambda_{2D}(n)\} \log n\{1 + o_p(1)\} \tag{5.9}$$

as $n \to \infty$, for the same R and C.

Now,

$$\hat{c} - c = c\{\exp(\log \hat{c} - \log c) - 1\}$$

which, after making the substitution implied by (5.9) and Taylor-expanding, yields the desired result for $\hat{c}$. □

The amount of data on the $n \times n$ grid is of order n^2, and so a convergence rate of $\hat{\alpha} - \alpha = O_p(n^{-1})$ corresponds to "root-N consistency" in more classical problems. If $\beta \geq 1$, this rate is attained in (5.6) if and only if $0 < \alpha < 1$. Similar behaviour is observed in the case of inference about fractal index in one-dimensional processes (see Constantine & Hall 1994, and chapter 4), except that the dividing line between "root-N consistency" and a slower convergence rate occurs at $\alpha = 3/2$, not $\alpha = 1$.

In both contexts, these respective values of α also represent the dividing line between circumstances where $\hat{\alpha}$ is asymptotically Normally distributed and those where it is not, with Normality occurring for values of α less than the critical one. This is of some practical significance, since many of the real bivariate data sets that we have encountered are well-approximated by processes having α between 1 and 3/2.

Numerical properties The rates of convergence of estimators of α and c deteriorate as α gets closer to 2, because the amount of information contained in oscillations of X decreases as the process becomes smoother. In the extreme case when $\alpha = 2$, there is insufficient information in a record of X on a finite interval to estimate c consistently. As α increases to 2, one needs to examine successively higher-order properties in order to obtain similar performance; see Kent & Wood (1997).

The problems as α approaches 2 are observable in a simulation study, such as that summarised in Figure 5.1. Panel (a) depicts the logarithm of the standard

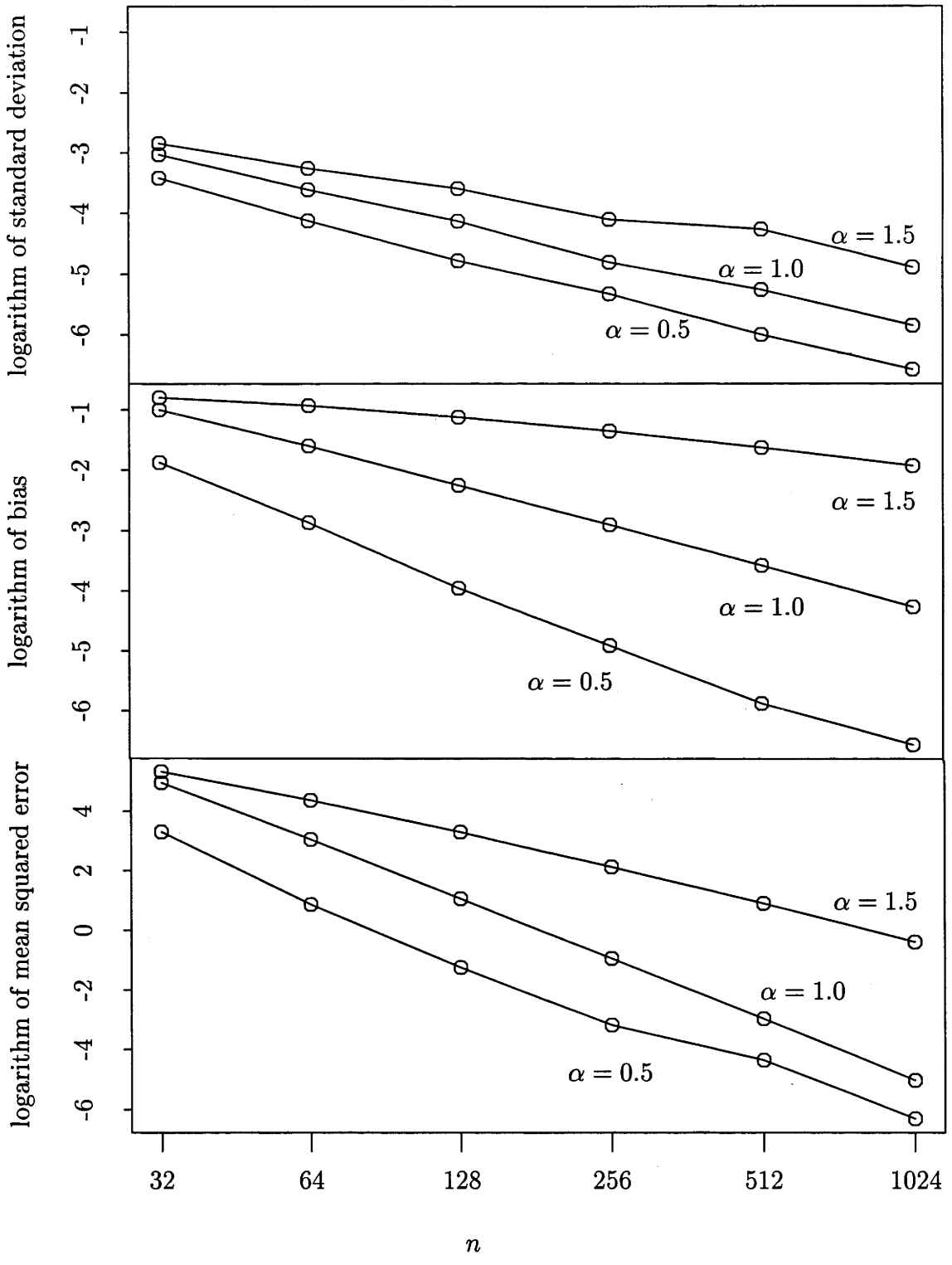


Figure 5.1: Plots of logged standard error, bias and mean squared error as a function of the logarithm of sample size, for a two-dimensional Gaussian random field with covariance $\gamma(t) = \exp(-8\|t\|^{\alpha})$ and for different values of α.

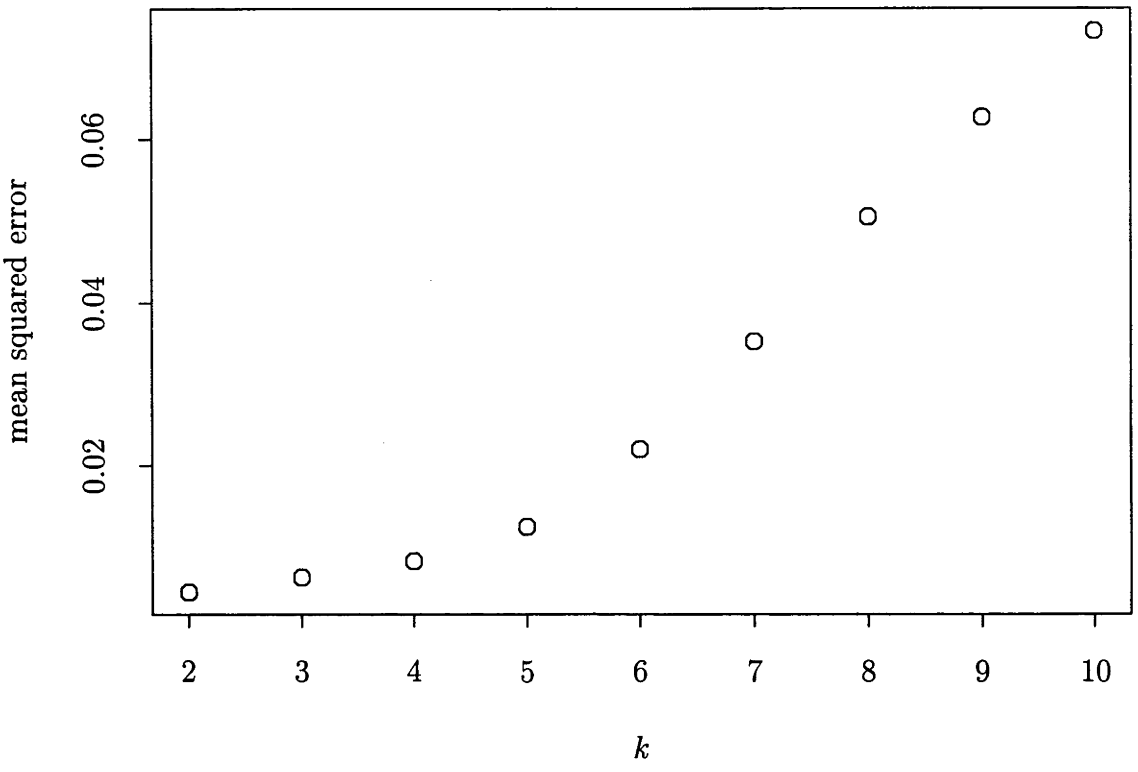


Figure 5.2: Plot of mean squared error of $\hat{\alpha}$ against the number of points, k, used to compute it by simple linear regression. The data were from a Gaussian process with covariance $\gamma(t) = \exp(-8|t|^{1.25})$

error of $\hat{\alpha}$, panel (b) the logarithm of the bias in $\hat{\alpha}$, and panel (c) the logarithm of the mean squared error of $\hat{\alpha}$, plotted against $\log n$ for $\alpha = 0.5$, 1.0 and 1.5. Data were generated from a two-dimensional Gaussian random field with covariance $\gamma(t) = \exp(-16\|t\|^{\alpha})$, in the range $t \in [0,1]^2$. We took $k = 4$ and u to equal n^{-1} multiplied by one of $1, 2^{1/2}, 2, 2^{3/2}$. For respective values of n the curves were based on $4(1024/n)^2$ simulations. For $\alpha > 1$, performance deteriorates as α increases, owing to an increase in variance; while for $\alpha < 1$, performance improves with increasing α, due to a decrease in bias.

Choice of k In addition to saving computational labour, there are two reasons for choosing k relatively small. Firstly, it minimises bias from the $o(1)$ term in equation (5.3). Secondly, estimates of the variogram exhibit high correlation between values

at neighbouring lags, and using more lags in the regression does not improve the accuracy of estimates as much as in the case of data with lower correlation. Indeed, taking the number of points in the regression to diverge with increasing sample size will inflate both bias and variance, and so even in an asymptotic sense, as $n \to \infty$, the optimal k is bounded. This has been shown theoretically by Constantine & Hall (1994).

To provide a numerical illustration, Figure 5.2 depicts a plot of mean squared error of $\hat{\alpha}$ against k, in the case of a Gaussian process on the interval $[0, 1]$ with covariance $\gamma(t) = \exp(-16|t|^{\alpha})$ and $\alpha = 1.25$. Clearly, mean squared error performance is optimised at a relatively small value, $k = 2$. Similar results are obtained in the bivariate case.

Quantifying the error Estimates of the variances of $\hat{\alpha}_I$ and $\hat{c}_I$ may be obtained from the parametric bootstrap procedure described in section 4.2.2, or from the following plug-in formulae:

$$s^2_{\hat{\alpha}_I} = \sum_{l_1=1}^{k}\sum_{l_2=1}^{k} a_{l_1}a_{l_2}s_I(u_{l_1}, u_{l_2}) \quad \text{and} \quad s^2_{\hat{c}_I} = \hat{c}_I^2 \sum_{l_1=1}^{k}\sum_{l_2=1}^{k} b_{l_1}b_{l_2}s_I(u_{l_1}, u_{l_2}),$$

where $s_I(u_1, u_2)$ is an estimator for $\mathrm{cov}\{\log \hat{v}_I(u_1), \log \hat{v}_I(u_2)\}$.

Here, $s_I(u_1, u_2)$ is constructed in a similar fashion to its one-dimensional equivalent $s(h_1, h_2)$ as defined in section 4.2.1. However, since $\hat{v}_I(h)$ is an average of a number of $\hat{v}(\boldsymbol{t})$'s and since these are calculated from data across a two-dimensional grid, its formulation is slightly more complex. Put

$$S(\boldsymbol{h}_1, \boldsymbol{h}_2) = \{\boldsymbol{t}_1 - \boldsymbol{t}_2 : \boldsymbol{t}_1 \in C(\boldsymbol{h}_1) \text{ and } \boldsymbol{t}_2 \in C(\boldsymbol{h}_2)\}$$

and, for $\boldsymbol{t} \in S(\boldsymbol{h}_1, \boldsymbol{h}_2)$, let $N(\boldsymbol{t})$ be the number of pairs $(\boldsymbol{t}_1, \boldsymbol{t}_2)$ in $C(\boldsymbol{h}_1) \times C(\boldsymbol{h}_2)$ such that $\boldsymbol{t}_1 - \boldsymbol{t}_2 = \boldsymbol{t}$, and let

$$T_I(v; \boldsymbol{t}, \boldsymbol{h}_1 - \boldsymbol{h}_2) = v(\|\boldsymbol{t} + \boldsymbol{h}_1 - \boldsymbol{h}_2\|) - v(\|\boldsymbol{t} + \boldsymbol{h}_1\|) - v(\|\boldsymbol{t} - \boldsymbol{h}_2\|) + v(\|\boldsymbol{t}\|).$$

Then,

$$s_I(u_1, u_2) =$$
$$\{2\tilde{v}(u_1)\tilde{v}(u_2)|C_I(u_1)||C_I(u_2)|\}^{-1} \sum_{\|\boldsymbol{h}_1\|=u_1} \sum_{\|\boldsymbol{h}_2\|=u_2} \sum_{\boldsymbol{t}\in S(\boldsymbol{h}_1,\boldsymbol{h}_2)} N(\boldsymbol{t}) T_I(\tilde{v}; \boldsymbol{t}, \boldsymbol{h}_1, \boldsymbol{h}_2)^2,$$

where $\tilde{v}(\cdot)$ is an estimator for the underlying one-dimensional variogram obtained by fitting a suitable valid model to $\hat{v}_I(\cdot)$. The one-dimensional fitting procedure described in section 4.3 can be used to fit the model.

The bootstrap method for estimating variances may also be used to obtain confidence intervals for $\hat{\alpha}_I$ and $\hat{c}_I$.

5.2 Box counting

Existing box counting methods for estimating fractal dimension are defined for subsets of Euclidean space with any number of dimensions. As such they are general in nature and, for the surface data, they do not take advantage of continuity of the underlying process nor the regular fashion in which the data are recorded.

As well as taking advantage of the structure of the data, the modified box counting method for one-dimensional data of Hall & Wood (1993), as described in section 2.3.2, improves on the performance of existing methods. The box counting described below is a direct extension of this to two-dimensional data.

The two-dimensional analogue of this method is to conduct linear regression of the logarithm of box covering *volume* against the logarithm of box width, for those data pairs with small box width. The fractal dimension is then calculated in an analogous way, as a linear function of the slope of the log-log regression.

The two-dimensional box-covering is a direct extension of the one-dimensional box-covering described in section 2.3.2. If $\mathcal{B}(i, l)$ is as defined in section 2.3.2 then denote its two-dimensional counterpart by

$$\mathcal{B}(\boldsymbol{i}, l) = \mathcal{B}(i_1, l) \times \mathcal{B}(i_2, l) \qquad (1 \le i_1, i_2 \le q_l, 1 \le l \le k),$$

where $\boldsymbol{i} = (i_1, i_2)'$. Recall q_l denotes the integer part of $(n-1)/lm$. Here, l denotes the level of discretisation and m the width of a block.

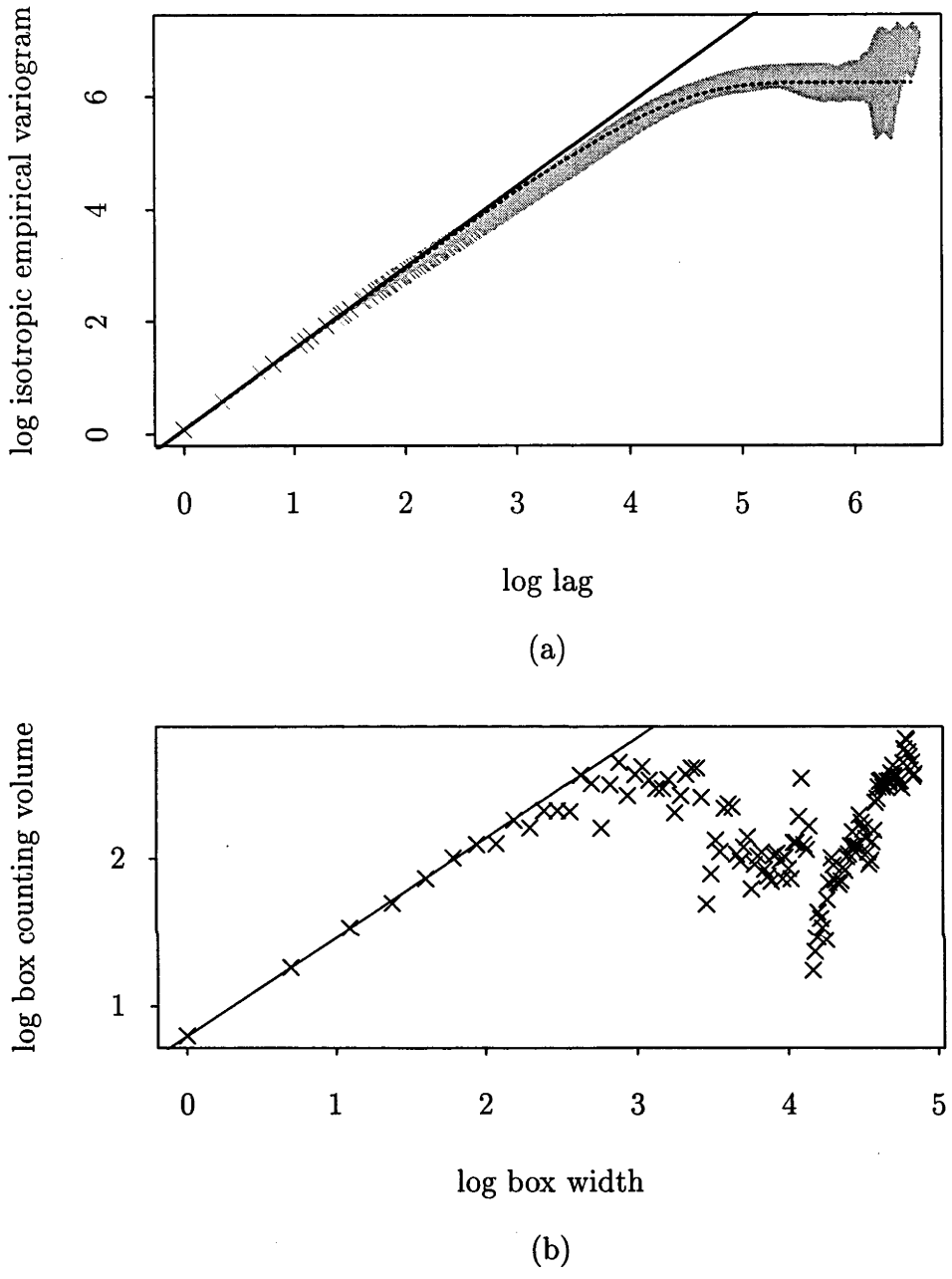


Figure 5.3: Panel (a) is a scatterplot of the log of the isotropic empirical variogram, $\hat{v}_I$, against log lag for the dry soil surface data shown in panel (a) of Figure 1.2. The straight line represents a least-squares fit through the points with lags: $1, 2^{1/2}, 2, 2^{3/2}$. The dotted curve represents a heuristic fit of the stable exponential variogram model to all of the points. Panel (b) is a scatterplot of the log of the box-covering volume, $V(l)$, against the log of minimum box-width (discretisation level l) for the same dry soil surface data set. The straight line represents a least-squares fit through the points with box-width: $1, 2, 3, 4$.

The approximate volume of the box–covering for the i^{th} block is

$$V_{il} = \epsilon_l^2 (U_{il} - L_{il}),$$

where

$$\epsilon = lm/n, \quad U_{il} = \max_{j \in \mathcal{B}(i,l)} X(j/n), \quad \text{and} \quad L_{il} = \min_{j \in \mathcal{B}(i,l)} X(j/n).$$

The individual box–covering volumes are summed over blocks to obtain the total box–covering volume,

$$V(l) = \sum V_{il} = \epsilon_l^2 \sum (U_{il} - L_{il}).$$

Using the *Box Counting Theorem* (Barnsley, 1988, page 176), a box counting estimator, $\hat{D}_{BC}$, for fractal dimension may be obtained from the slope of a log-log regression

$$3 - \hat{D}_{BC} = \left\{ \sum_{l=1}^{k} (x_l - \bar{x}) \log V(l) \right\} \Big/ \left\{ \sum_{l=1}^{k} (x_l - \bar{x})^2 \right\},$$

where $x_l = \log l$ and $\bar{x} = \sum_{l=1}^{k} x_l$.

The variogram and box counting methods for estimating fractal dimension are compared in Figure 5.3. Panel (a) depicts the variogram method and shows a scatterplot of the pairs $(\log u_l, \log \hat{v}_I(u_l))$ for the first soil surface data set, *i.e.* before any rainfall. The solid straight line represents the ordinary least-squares fit to the four points with $u_l = 1, 2^{1/2}, 2, 2^{3/2}$. The estimate of fractal dimension obtained from the slope of the regression line is 2.35. Estimates of fractal dimensions for the other soil surface data sets are given in Table 5.1.

Panel (b) of Figure 5.3 depicts the box counting method and shows a scatterplot of the pairs $[\log l, \log V(l)]$, also for the soil surface before rainfall. The solid straight line represents the ordinary least squares fit to the four points with $l = 1, 2, 3, 4$. The estimate of fractal dimension is 2.35.

There appears to be less structure and more variability in the scatterplot of panel (b) than in the scatterplot of panel (a). This is because the variogram estimator is an average of $O(n^2)$ terms, whereas the box-covering volume estimator is

Rainfall	Fractal Dimension		Fractal Index			Topothesy		
(mm)	$\hat{D}$	$(l_{\hat{D}}, u_{\hat{D}})$	$\hat{\alpha}_I$	$(s_{\hat{\alpha}_I})$	$[s_{\hat{\alpha}_I}]$	$\hat{c}_I$	$(s_{\hat{c}_I})$	$[s_{\hat{c}_I}]$
0.00	2.35	(2.34, 2.36)	1.30	(0.01)	[0.01]	2.02	(0.02)	[0.02]
5.30	2.45	(2.44, 2.46)	1.10	(0.01)	[0.01]	2.89	(0.02)	[0.02]
10.05	2.31	(2.29, 2.32)	1.39	(0.01)	[0.01]	1.80	(0.02)	[0.02]
14.30	2.31	(2.30, 2.32)	1.39	(0.01)	[0.01]	1.56	(0.02)	[0.02]
18.55	2.36	(2.35, 2.37)	1.28	(0.01)	[0.01]	1.57	(0.02)	[0.02]
22.50	2.27	(2.26, 2.28)	1.46	(0.01)	[0.01]	1.16	(0.02)	[0.02]
27.00	2.37	(2.36, 2.38)	1.26	(0.01)	[0.01]	1.41	(0.02)	[0.02]
31.25	2.27	(2.26, 2.28)	1.46	(0.01)	[0.01]	1.08	(0.02)	[0.02]
35.50	2.28	(2.26, 2.29)	1.45	(0.01)	[0.01]	1.08	(0.02)	[0.02]

Table 5.1: Results of analysing the soil-surface data after successive amounts of rainfall. For each data set, the table shows point estimates and 95% confidence intervals for fractal dimension, and point estimates & standard errors for fractal index and topothesy. The standard errors shown in parentheses were calculated using the plug-in method and the standard errors shown in brackets were calculated using the bootstrap method.

an average of $O(n)$ terms. The terms in both estimators are similar in construction and therefore contain similar information.

Nevertheless, the estimates of fractal dimension obtained from both methods agree to two decimal places. Since they both yield similar values, it may be advantageous to use the box-counting method for large data sets, since it requires less time to compute. However, for precision the variogram method is recommended.

5.3 Application to soil surfaces

Computer-generated scene renderings of the raw soil-surface data are shown in Figure 1.2. (The data sets and the experiments from which they were collected are described in section 1.2.) On first inspection, it is difficult to see any significant changes in the surface after successive amounts of rainfall.

However, there are two features in the sequence that may be noted after close scrutiny. The first concerns the difference between the dry soil sample, panel (a) of Figure 1.2, and the soil after 5.30mm of rain, panel (b). It might be said that the surface of the latter exhibits a rougher texture than that of the former. The second feature concerns the sequence of soil surfaces, from directly after the initial period of rainfall until the wettest soil surface, panels (b) through (i) of Figure 1.2. In this sequence, the lower-lying regions of the soil surface appear to become progressively smoother, whereas the higher regions are becoming slightly more granular in appearance.

These two features are expected, and correspond to the initial "spattering effect" of raindrops on a dry soil surface, and to the transportation of smaller particles from high regions to lower regions with the flow of rain over the surface, respectively.

Point estimates of fractal dimension and topothesy were obtained for each data set using the methods of section 5.1. These estimates are shown in Table 5.1, along with confidence intervals for fractal dimension and estimates of standard error for fractal index and topothesy.

The confidence intervals for fractal dimension and topothesy are also depicted in Figure 5.4, each panel showing how the fractal dimension and topothesy vary with successive amounts of rainfall. The estimated fractal dimension varies relatively little over time, whereas estimated topothesy decreases steadily after an initial increase. This would suggest that the fractal dimension of the soil surface is relatively unaffected by rainfall, and it is topothesy that explains the two features of initial "spattering" and later erosion and deposition of finer particles.

Care must be taken when comparing the topothesies of surfaces with different fractal dimensions as it is not clear how to interpret any differences. However, if

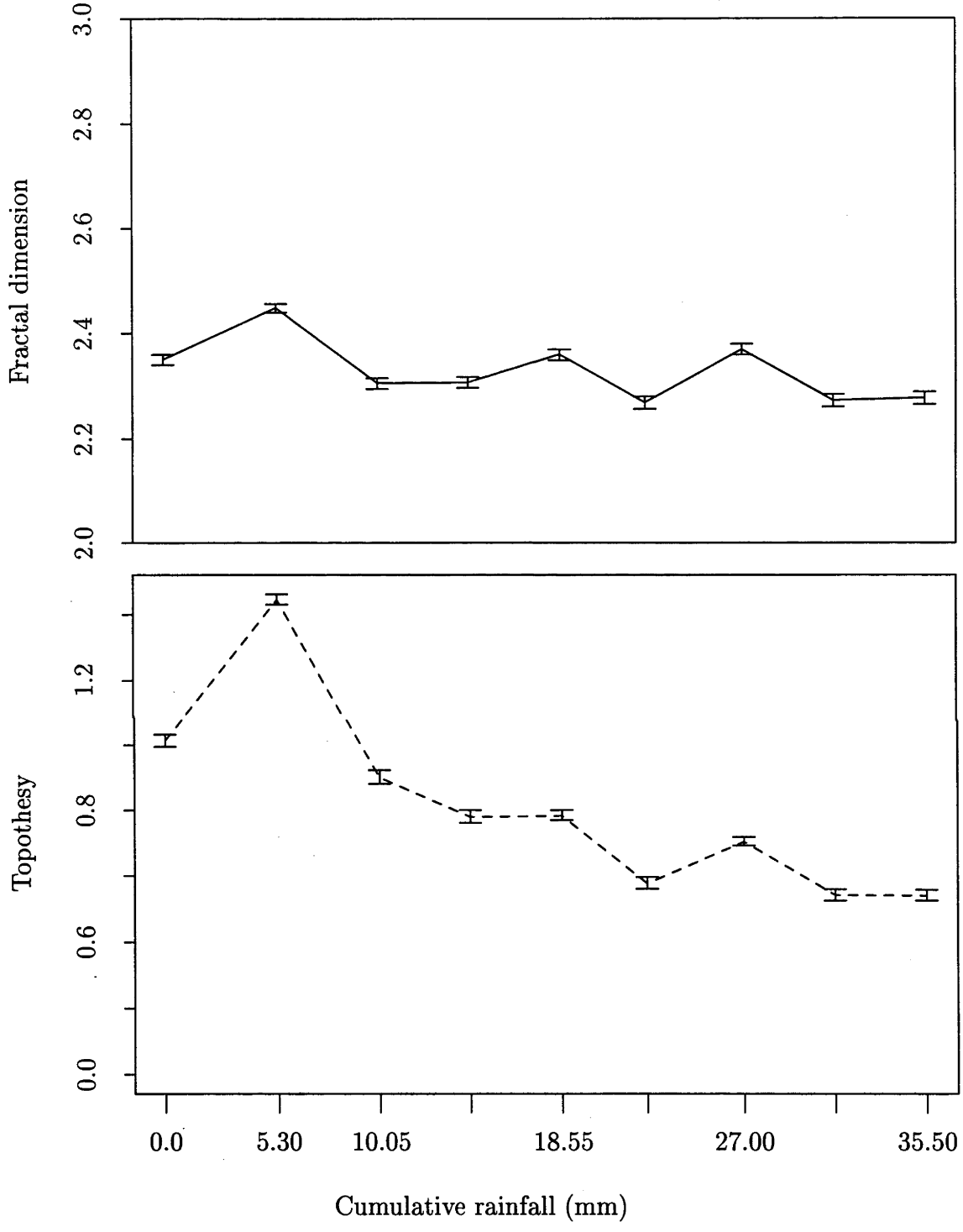


Figure 5.4: Graph of 95% confidence intervals for fractal dimensions (joined by a solid line) and topothesies (joined by a broken line) versus cumulative rainfall for the soil surface data.

Rainfall	$\tilde{\alpha}$	$\tilde{\lambda}$	$\tilde{\sigma}$
0.00	1.30	10.27	18.02
5.30	1.11	4.49	17.67
10.05	1.39	17.06	17.52
14.30	1.39	14.82	17.32
18.55	1.28	7.92	17.07
22.50	1.47	19.02	16.71
27.00	1.26	6.79	16.40
31.25	1.46	18.22	16.15
35.50	1.45	17.43	16.06

Table 5.2: Estimated model parameters obtained from fitting the stable exponential variogram model to the isotropic empirical variogram using the heuristic fitting procedure of section 4.3.

the fractal dimensions are similar, inferences based on comparing topothesies are arguably justified. In such cases, it is preferable to estimate a common fractal dimension with individual topothesies simultaneously for the surfaces.

Since fractal dimension estimates vary relatively little for the soil surface data, it might be supposed that fractal dimension remains constant. The least squares estimators of fractal index and topothesy of section 5.1 were modified to obtain estimators for a common fractal index and different topothesies for the soil surface data sets. The common fractal index was 1.34, corresponding to a fractal dimension of 2.33. The values for topothesy were similar to those from the individual fractal analyses of each data set, as tabulated in Table 5.1. Indeed, a plot of those topothesies

estimated assuming a common fractal dimension is very similar to that in Figure 5.4. This provides some support that topothesy exhibits the features observed in the raw data.

The standard errors and confidence interval estimates in Table 5.1 were obtained using the plug-in and bootstrap methods described in section 5.1. Both methods required valid variogram models for the underlying process. Parameters estimates from fitting the stable exponential variogram model to each of the soil data sets are given in Table 5.2. The fitting procedure used was the heuristic method described in section 4.3. The fitted model for variogram of the dry surface data is depicted by the dotted curve in panel (a) of Figure 5.3.

Chapter 6

Anisotropic surfaces

This chapter is concerned with more general surfaces for which it may not be reasonable to make the assumption of isotropy due to physical considerations, nor to assume isotropy given empirical evidence. The variogram models from the previous chapter are extended to cater for these surfaces, by allowing fractal index and topothesy to vary with orientation.

It is shown that the amount by which the fractal index and hence fractal dimension of line transects can vary is restricted. In fact it can take at most two values. The restrictions on topothesy are not so great and it generally take the form of a positive periodic function.

Methods for estimating the fractal index and the topothesy function for general anisotropy are given. Also, attention is paid to a special form of anisotropy that appears common in data sets in the literature. For this form of anisotropy, a model is developed and parameterised in a way that allows existing concepts to be quantified.

This special model is used as part of a bootstrap hypothesis test for isotropy. Since the bootstrap test methods are quite general, it is shown how to construct a two-surface test for common fractal dimension.

All the methods developed are applied to the polymer data described in chapter 1.

6.1 General anisotropy

To date there has been little attention paid to characterising surface roughness in terms of anisotropy. Most methods have involved describing roughness with one or two global parameters. This has been due largely to the nature of the measurement devices used to obtain surface elevation data. Devices such as stylus and optical profilometers only record data along a single transect of the surface. In order to satisfactorily address the question of anisotropy using data from these devices, it would be necessary to measure the same surface many times with the same device, while accurately recording the relative positions and orientations of successive transects. Instead, the question of anisotropy has largely been ignored. Only recently, with the advent of more modern measurement devices such as the laser scanning device used to measure the soil surface and the scanning tunnelling microscope used to record the polymer surfaces, have genuinely two-dimensional data been gathered. With it, attention is starting to concentrate on how to characterise the anisotropy of surface roughness.

So far, methods for both one-dimensional data and two-dimensional isotropic data are based on properties of a univariate variogram. However for anisotropic data, the variogram is a function of two variables, lag and orientation. This has implications for the two roughness parameters, fractal dimension and topothesy. In the univariate cases these parameters took on a single value but, in the bivariate anisotropic case, the parameters may also vary with orientation. Simply extending the model for the univariate variogram (2.4) to two dimensions leads to the following model:

$$v(\boldsymbol{t}) = c(\arg \boldsymbol{t})\|\boldsymbol{t}\|^{\alpha(\arg \boldsymbol{t})} + o(\|\boldsymbol{t}\|^{\alpha(\arg \boldsymbol{t})}), \quad \text{as } \boldsymbol{t} \to \mathbf{0}. \tag{6.1}$$

Here, fractal index and topothesy are functions of orientation.

Although the form of these functions is not specified in (6.1), it is reasonable to suspect that properties of the variogram affect the possible forms these functions can take. This is indeed the case as shown by the following theorem, which shows that the extent to which the fractal index of line transects can vary is limited. The

theorem asserts that for any three orientations, the two lowest fractal indices must be identical. Hence, if fractal dimension can be expressed in terms of fractal index by the canonical formula $D = 2-\alpha/2$, in the case $d = 1$, then the two highest values of fractal dimensions must be equal. In simpler terms this says "if the intrinsically stationary stochastic surface X satisfies the usual one-dimensional scaling laws then the fractal dimensions of its line transect processes are the same in all directions except possibly one, in which fractal dimension may be less than in all others."

THEOREM 1 *Let $v(\cdot)$ denote a valid variogram function in the plane with the property that, for any three different orientations $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3$ (denoting unit vectors such that $\boldsymbol{\theta}_i \neq \pm\boldsymbol{\theta}_j$ for $i \neq j$), the quantities*

$$a_i = \sup\{\alpha > 0 : v(u\boldsymbol{\theta}_i) = O(|u|^\alpha) \text{ as } u \to 0\} \qquad \text{and}$$
$$b_i = \inf\{\alpha > 0 : |u|^\alpha = O[v(u\boldsymbol{\theta}_i)] \text{ as } u \to 0\}$$

satisfy $a_i = b_i$ $(= \alpha_i$, say). Order the $\boldsymbol{\theta}_i$'s so that $\alpha_1 \leq \alpha_2 \leq \alpha_3$. Then $\alpha_1 = \alpha_2$.

PROOF Since $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3$ are distinct two-dimensional vectors, there exist non-zero scalars r_1, r_2, r_3 such that $r_1\boldsymbol{\theta}_1 + r_2\boldsymbol{\theta}_2 + r_3\boldsymbol{\theta}_3 = \mathbf{0}$. Construct a triangle whose vertices are

$$\boldsymbol{t}_1 = \mathbf{0}, \quad \boldsymbol{t}_2 = r_1\boldsymbol{\theta}_1 \quad \text{and} \quad \boldsymbol{t}_3 = r_1\boldsymbol{\theta}_1 + r_2\boldsymbol{\theta}_2.$$

Choose any positive real numbers $c_1, c_2 > 0$ and put $c_3 = -(c_1+c_2)$ so that $\sum c_i = 0$. Then, because variograms are necessarily conditional non-positive definite,

$$0 \geq \sum\sum c_i c_j\, v(\boldsymbol{t}_i - \boldsymbol{t}_j) = 2c_1c_2v(r\, r_1\boldsymbol{\theta}_1) + 2c_2c_3v(r\, r_2\boldsymbol{\theta}_2) + 2c_3c_1v(r\, r_3\boldsymbol{\theta}_3).$$

Now suppose that, contrary to the claim of the theorem, $\alpha_1 < \alpha_2 \leq \alpha_3$. Divide both sides of the inequality above by $v(r\, r_1\boldsymbol{\theta}_1)$, and let $r \to 0$, obtaining $0 \geq 2c_1c_2$, which contradicts the assumption that $c_1c_2 > 0$. □

EXAMPLE 6.1 In section 3.5 the mathematical concept of a *ruled surface* was introduced; a ruled surface, X_θ, is an extension of a one-dimensional process to two

dimensions by 'drawing out' the process in a particular direction. From its definition, the construction of a ruled surface is not unlike the manufacturing processes of extrusion in the case of plastic surfaces or milling in the case of metal surfaces. It is conceivable therefore that such surfaces may exhibit a special orientation where fractal dimension is different from all others.

A suitable refinement of (6.2) for the underlying variogram of a ruled surface is then

$$v(\boldsymbol{t}) = c\,|\cos(\arg \boldsymbol{t} - \psi)|^{\alpha}\|\boldsymbol{t}\|^{\alpha} + o(\|\boldsymbol{t}\|^{\alpha}), \quad \text{as } \boldsymbol{t} \to \mathbf{0}.$$

Empirical evidence of the assertion of the theorem may be observed in Figure 6.1. For each of 24 unit vectors $\boldsymbol{\theta}$, Figure 6.1 (a) illustrates least-squares fits of straight lines through pairs $(\log r, \log \hat{v}(r\boldsymbol{\theta}))$ with varying r, for the polymer surface data in panel (c) of Figure 1.3. The vectors $\boldsymbol{\theta}$ were approximately equally spaced around the circle, subject to the tangents of their orientations being rational. Although the intercepts vary considerably the slopes of the lines are similar, suggesting that the topothesies vary significantly with orientation but that the fractal dimensions do not. For the sake of comparison, Figure 6.1 (b) shows analogous lines for the surface depicted in panel (d) of Figure 1.3. This surface is smoother but the evidence of local self-affineness is similar. Formal testing will be addressed in section 6.3, where bootstrap methods for these data are discussed.

A result related to the Theorem, that fractal dimension is the same in "almost all" directions with respect to Lebesgue measure, is given in the non-stationary case by Marstrand (1954), Falconer (1985, Chapter 6) and Mattila (1985). The more specific result in the Theorem, that fractal dimension is the same in all directions except possibly one, is of significantly greater physical interest because most manufactured surfaces are produced in a manner that ascribes special importance to a particular "manufacturing axis". Thus, it is feasible that an orientation of special fractal character might exist. Engineering surfaces that have been milled, ground, face-tuned or bored, often have surfaces that closely resemble the mathematical ideal of a ruled surface; see Stout *et al.* (1993) for examples of such surfaces. After processing, they are not unlike the superposition of a ruled surface and a stationary,

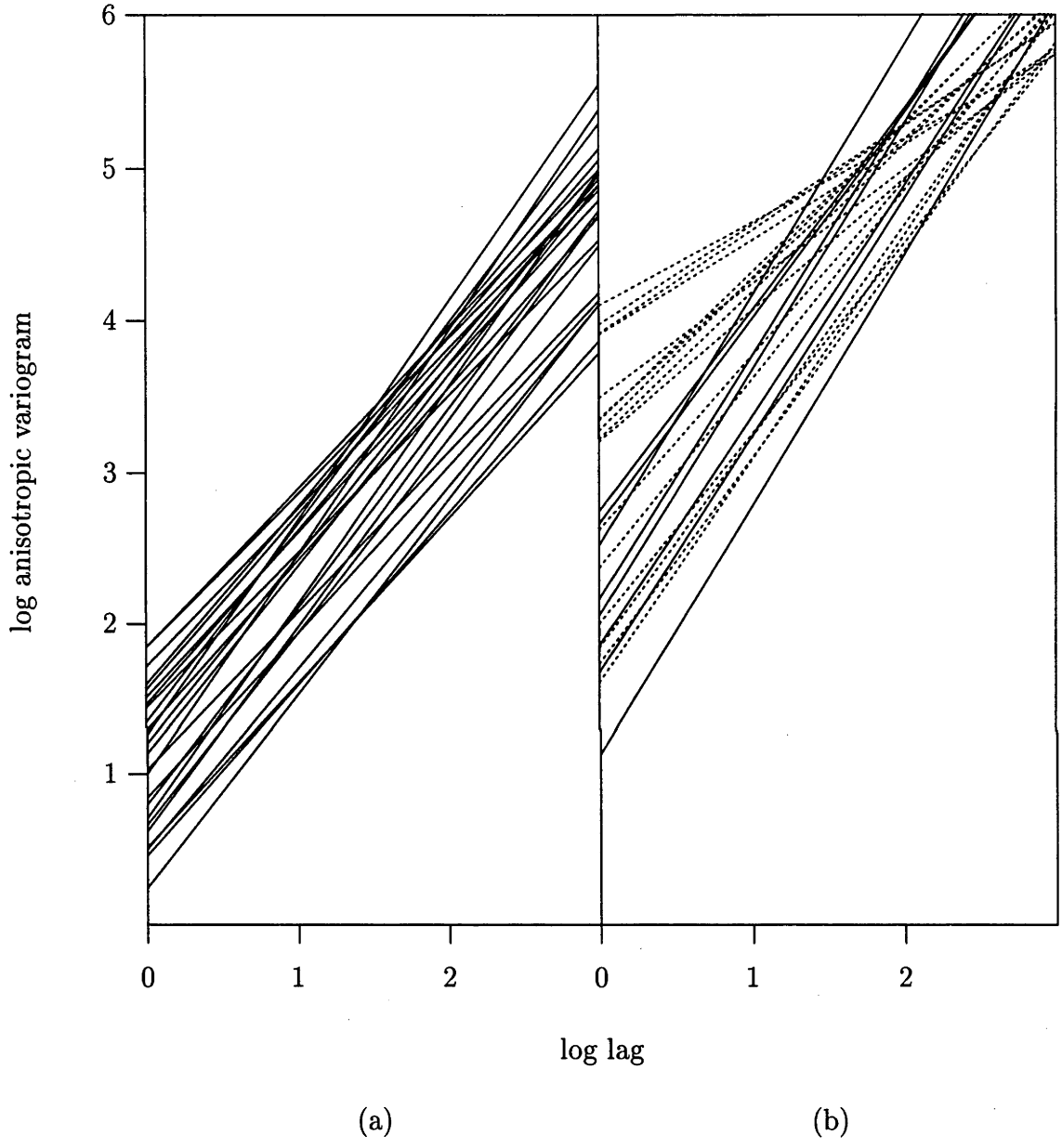


Figure 6.1: Least-squares fits to the scatterplots of the log of the empirical variogram, $\hat{v}(\cdot)$, against log lag for line transects in different orientations. Panel (a) is for data depicted in panel (c) of Figure 1.3, while panel (b) is for data in panel (d) of Figure 1.3. In the latter, broken lines represent fits based on larger lags.

stochastic, isotropic surface.

For cases where there exists an exceptional direction corresponding to lower fractal dimension, we may interpret the Theorem by saying that the roughness of a strongly directional surface will be experienced in any direction which is not orthogonal to that of the highest-frequency oscillations. One example of this phenomenon is that of a stationary process which may be decomposed into one component that is a ruled surface and another that is perhaps more directionally homogeneous (in any event, not a ruled surface with the same axis as the first one) but strictly smoother than the ruled surface. The unique direction in which the fractal dimension is less than that of the overall surface is orthogonal to the oscillations of the ruled surface. See also Hall & Davies (1995).

These direction invariance properties may be regarded as the basis for physical measurement of surface roughness based on stylus or optical profiling. If one's goal is estimation of fractal dimension, as distinct from scale, then it is often not essential to be meticulous about the orientation of a profile. Nevertheless, even if the dimension of line transects is the same in all directions, it can be statistically beneficial to take measurements in the direction in which the scale of fluctuations is greatest.

It should be stressed that even in cases where the fractal dimension of line transect samples can assume different values in different directions, the fractal dimension of the surface $\{X(t), t \in \mathbb{R}^2\}$ "as a whole" is usually well-defined and equals 1 plus the higher of the two fractal dimensions of line transect sections: see for example Federer (1969, Section 3.10).

Implications of the theorem The dependence of scale on orientation is more complex, but may be appreciated from a version of (2.4) without assuming isotropy. If $d = 2$ then (2.4) should generally be replaced by

$$v(\boldsymbol{t}) = c(\arg \boldsymbol{t})\|\boldsymbol{t}\|^\alpha + o(\|\boldsymbol{t}\|^\alpha), \qquad \text{as } \boldsymbol{t} \to \boldsymbol{0}, \tag{6.2}$$

where $c(\arg \boldsymbol{t})$ is a continuous, periodic and nonnegative function, nondegenerate in the case of anisotropy; and $0 < \alpha \leq 2$ is a constant.

The function $c(\cdot)$ is strictly positive except possibly for a single special value of

$\arg \boldsymbol{t}$, ψ say, where $c(\arg \boldsymbol{t})$ vanishes. Fractal index and fractal dimension for line transect samples in this direction are determined by the form of the remainder term $o(\|\boldsymbol{t}\|^\alpha)$ when $\arg \boldsymbol{t} = \psi$. In all other directions, they equal α and $2 - \frac{1}{2}\alpha$, respectively. Note particularly that in the case of (6.2), c is a function — the topothesy function.

Point estimation For data recorded on a bounded two-dimensional grid, there are only a finite number of possible orientations at which the variogram may be estimated. These orientations are those of lines passing through two or more grid points. Consequently, there are only a finite number of orientations at which naive empirical estimates of $c(\cdot)$ may be calculated. Since concern is still primarily with the behaviour of the variogram in the vicinity of the origin, the number of orientations at which $c(\cdot)$ is directly estimable is smaller. In constructing estimators for α and $c(\cdot)$, only those values of the variogram at displacements with lags not greater than R will be used.

Let $\theta_1, \ldots, \theta_\nu$ denote the ν distinct orientations that represent lines through points on the grid at most a distance R apart. For each orientation θ_i, let u_{ij}, for $1 \le j \le k_i$, be the k_i ordered positive values not exceeding R for which $\hat{v}(u_{ij}\boldsymbol{\theta}_i)$ is calculable. Recall the convention $\arg \boldsymbol{\theta} = \theta$. Put $c_i = c(\theta_i)$.

Estimators for α and c_i may be obtained by minimising a least–squares criterion, to fit

$$\log \hat{v}(\boldsymbol{t}) = \log c(\arg \boldsymbol{t}) + \alpha \log \|\boldsymbol{t}\|,$$

over the set of displacements for which $\|\boldsymbol{t}\| < R$. These least–squares estimators are given by

$$\hat{\alpha} = \sum_{i=1}^{\nu} \sum_{j=1}^{k_i} (\log r_{ij} - \overline{\log r_{i\cdot}}) \log \hat{v}(r_j \boldsymbol{\theta}_i) \Big/ \sum_{i=1}^{\nu} \sum_{j=1}^{k_i} (\log r_{ij} - \overline{\log r_{i\cdot}})^2 , \tag{6.3}$$

and

$$\log \hat{c}_i = \frac{1}{k_i} \sum_{j=1}^{k_i} \log \hat{v}(r_j \boldsymbol{\theta}_i) - \hat{\alpha}\, \overline{\log r_{i\cdot}} \tag{6.4}$$

where $\overline{\log r_{i\cdot}} = k_i^{-1} \sum \log r_{ij}$.

To obtain estimates of $c(\phi)$ for general ϕ, let $\hat{c}(\phi)$ be the result of passing a local linear smoother (Fan, 1993; Hastie & Loader, 1993) through the pairs of points $(\theta_i, \hat{c}_i)$, using a compactly supported kernel K and interpreting both ϕ and θ_i as if they took values on the interval $(-\pi/2, \pi/2]$ wrapped around a circle.

Moment and distributional properties of $\hat{\alpha}$ and $\hat{c}_i$ are very similar to those of $\hat{\alpha}_I$ and $\hat{c}_I$, as described in chapter 5 for isotropic two-dimensional data. Also, estimates of the standard errors of $\hat{\alpha}$ and $\hat{c}_i$ may be obtained similarly to those described in section 5.1.

6.2 Weak anisotropy

One particular form of anisotropy that appears relevant, at least to the polymer surface data, is that which causes the underlying variogram to have elliptical contours. This is often called *weak anisotropy* in the context of surface metrology and *geometrical anisotropy* in the context of geostatistics.

Geometrical anisotropy is usually given a more formal definition. It refers to the situation in which there exists a linear transformation that, when applied to the argument of a geometrically anisotropic variogram, corrects for the anisotropy. Let $v(\boldsymbol{t})$ be a geometrically anisotropic variogram and M be a correcting linear transformation. Then

$$v(\boldsymbol{t}) = v(||M\boldsymbol{t}||),$$

where, abusing notation, $v(t)$ is a valid univariate variogram.

This definition is not limited to two dimensions, and in geostatistics is often applied to problems in three or more dimensions. However, for the problem of characterising surface roughness, it will only be used in the two-dimensional setting. Hence we prefer to use the terminology of *weak anisotropy*.

One advantage of using this definition is that properties of the elliptical contours can be used to characterise the surface. For instance, the common orientation of the ellipses is that in which the topothesy function attains its minimum, and the eccentricity of the ellipses is related to the degree to which a surface is anisotropic.

These are often important factors when considering the manufacture or application of surfaces.

A refinement of assumption (6.2) for the behaviour of the variogram in the neighbourhood of the origin for a weakly anisotropic surface is

$$v(\boldsymbol{t}) = c_0 \{1 + s\cos 2(\arg \boldsymbol{t} - \psi)\}^{\alpha/2} \|\boldsymbol{t}\|^{\alpha} + o(\|\boldsymbol{t}\|^{\alpha}), \qquad \text{as } \boldsymbol{t} \to \mathbf{0}. \tag{6.5}$$

This leads to the following model for the logarithm of the variogram at small lags:

$$\log v(\boldsymbol{t}) = \log c_0 + \frac{\alpha}{2} \log \{1 + s\cos 2(\arg \boldsymbol{t} - \psi)\} + \alpha \log \|\boldsymbol{t}\|. \tag{6.6}$$

These equations have been parameterised in a manner that makes them easily interpretable. The constant c_0 provides an "average" measure of scale or *average topothesy*, and is comparable to the topothesy from the isotropic model. The parameter s lies between 0 and 1 and provides a measure of the *strength of anisotropy* of a surface. A value of 0 for s would result in the topothesy function degenerating to a constant, and hence imply the expansion (2.4) of the variogram of an isotropic process, *i.e.* no anisotropy in a local sense. If s were nonzero then ψ, which we might call the *lay*, in agreement with its use in surface metrology, would be the orientation at which the topothesy function attained its maximum value.

To see how (6.5) was obtained, first formulate parametric equations for concentric elliptical paths centred at the origin. If the eccentricity of these ellipses is λ and their major axes are oriented in the direction ψ then suitable equations are

$$\|\boldsymbol{t}\| \cos(\arg \boldsymbol{t} - \psi) = r \cos \theta, \tag{6.7}$$

$$\|\boldsymbol{t}\| \sin(\arg \boldsymbol{t} - \psi) = r\,\lambda \sin \theta. \tag{6.8}$$

Here eccentricity is defined to be the ratio of the furthest point on the ellipse from its centre to the closest point on the ellipse from its centre, and hence is bounded below by 1.

For any given family of concentric ellipses specified by the pair (ψ, λ), the equations allow the determination of the "radius" of a particular ellipse given any point

on that ellipse. This can be seen from the following identity, achieved by summing the squares of (6.7) and (6.8) after appropriate scaling,

$$r^2 = \|\boldsymbol{t}\|^2 \left(\cos^2(\arg \boldsymbol{t} - \psi) + \frac{\sin^2(\arg \boldsymbol{t} - \psi)}{\lambda^2} \right). \tag{6.9}$$

Suppose that $v(\cdot)$ has concentric elliptical contours. Then the value of the variogram at points on the same ellipse are all equal. The furthest points on concentric ellipses from their centres all lie in the same direction and so the variogram along this direction is a valid one-dimensional variogram. This gives rise to

$$v(\boldsymbol{t}) = v(r).$$

For $v(r)$ satisfying the one-dimensional model (2.4) and r obtained from (6.9),

$$v(\boldsymbol{t}) = c \left\{ \cos^2(\arg \boldsymbol{t} - \psi) + \frac{\sin^2(\arg \boldsymbol{t} - \psi)}{\lambda^2} \right\}^{\frac{\alpha}{2}} \|\boldsymbol{t}\|^\alpha + o(\|\boldsymbol{t}\|^\alpha),$$

which, after applying some trigonometric identities and reparametrising, becomes

$$v(\boldsymbol{t}) = c_0 \left(1 + s \cos 2(\arg \boldsymbol{t} - \psi)\right)^{\frac{\alpha}{2}} \|\boldsymbol{t}\|^\alpha + o(\|\boldsymbol{t}\|^\alpha),$$

where $c_0 = c(\sqrt{\lambda^2 + 1}/2\lambda)^\alpha$ and $s = (\lambda^2 - 1)/(\lambda^2 + 1)$. Thus, the strength of anisotropy s is determined solely by the eccentricity of the elliptical contours λ.

From (6.5), the topothesy function is given by

$$c(\arg \boldsymbol{t}) = c_0 \left\{1 + s \cos 2(\arg \boldsymbol{t} - \psi)\right\}^{\frac{\alpha}{2}}.$$

6.3 A test for isotropy

The problem of detecting anisotropy and assessing its degree has received little attention in the literature. Indeed, in geostatistics, where it has long been recognised that it is important to ensure that anisotropy present in the data is detected, most texts use graphical methods for detecting anisotropy.

A notable attempt to address the gap was made by Baczkowski & Mardia (1990). They developed a variogram–based test for symmetry about the diagonal of a rectangular grid. For mathematical convenience they assumed a doubly geometric process

model for the data which, apart from failing to be radially isotropic, is too restrictive for the present application. However, they showed that this test for symmetry may be used in a limited number of situations to test for isotropy.

The method for testing for isotropy suggested here is based on the model (6.5) for weakly anisotropy processes. The value of the model parameter s determines whether the model is of a weakly anisotropic process or an isotropic process. So a test for isotropy against the specific alternative of weak anisotropy can be carried out by testing the value of the strength of anisotropy s. Formally, we wish to test

$$H_0 : s = 0 \qquad \text{against} \qquad H_1 : s > 0.$$

The obvious suitable test statistic is then $\hat{s}$, the estimate of s obtained from (6.6). Unfortunately, simple distributional properties of $\hat{s}$ are difficult, if not impossible, to obtain due to the complicated correlation structure inherent in the empirical variogram, the iterative nature of the fitting procedure, and the dependence on other parameter values. Monte Carlo methods, similar to those described in section 4.3.1 for model validation, again provide a practical solution to estimating the null distribution.

Step 1: Choose a maximum lag R which defines a suitable neighbourhood of the origin. Fit the weakly anisotropic model (6.6) using those pairs $(\boldsymbol{t}, \hat{v}(\boldsymbol{t}))$ for which $\|\boldsymbol{t}\| < R$, to obtain estimates $\hat{\alpha}, \hat{c}_0, \hat{s}$ and $\hat{\psi}$ of model parameters. The estimate $\hat{s}$ will be used as the test statistic.

Step 2: Fit a suitable null (isotropic) model for the variogram, either the power law model (4.7) or the stable exponential model (4.8), using all pairs $(\|\boldsymbol{t}\|, \hat{v}(\boldsymbol{t}))$ for which $\hat{v}(\boldsymbol{t})$ is calculable. This can be done using the heuristic fitting procedure suggested in section 4.3. The resultant fitted model, $\tilde{v}(\cdot)$ say, will be used to generate the null distribution of $\hat{s}$.

NOTE: if the power law model (4.7) is considered suitable then the parameter estimates $\hat{\alpha}$ and $\hat{c_0}$, calculated in STEP 1, offer practical estimates of parameters for the purposes of simulation.

Step 3: Simulate B realisations of Gaussian random fields with $\tilde{v}(\cdot)$ as their variogram. Identify each as X_b^*.

Step 4: For all simulated processes X_b^*, calculate their respective empirical variograms $\hat{v}_b^*(\boldsymbol{t})$ and fit (6.6) using those pairs $(\boldsymbol{t}, \hat{v}_b^*(\boldsymbol{t}))$ for which $\|\boldsymbol{t}\| < R$, to obtain model parameter estimates, $\tilde{\alpha}_b^*, \tilde{c}_{0b}^*, \tilde{s}_b^*$ and $\tilde{\psi}_b^*$.

Step 5: Order the $\tilde{s}_b^*$'s: $\tilde{s}_{[1]}^* \leq \cdots \leq \tilde{s}_{[B]}^*$. The null hypothesis is rejected at the $100(1-\beta)\%$ level if the test statistic $\tilde{s}$ exceeds $\tilde{s}_{[(1-\beta)B]}^*$.

6.3.1 Comparing fractal dimension between surfaces

Most of the steps involved in the above algorithm are general; only a few are specific to the test of isotropy. A modification of the test above, producing a two–surface test for common fractal dimension, is given below.

Step 1: Fit a variogram model to the data. The model used should permit the whole range of allowable values for each of the anisotropic roughness parameters.

For the parameter under investigation, "pool" the parameter estimates obtained from the two surfaces, and substitute the pooled estimate for the values in the models from which we shall resample. For example, to compare the fractal dimensions of two anisotropic surfaces, one might take the average of their estimated fractal indices and substitute it for the original fitted values in the respective fitted models, leaving all other parameter estimates the same.

Step 2: From the modified models for each surface, simulate a number, B say, of realisations of Gaussian random fields. These simulations will then be conducted under the null hypothesis that the surfaces have similar underlying parameter values.

Step 3: For each of the $2B$ simulated surfaces, calculate their respective empirical variograms $\hat{v}_b^*(\cdot)$ and fit model (6.6), to obtain B sets of parameter estimates from each model.

Step 4: For the parameter in question, calculate the absolute difference of its estimated values from every combination of pairs of the simulated surfaces, one from either model, to obtain a bootstrap distribution. Then compare the absolute difference of the parameter's estimated values for the two real surfaces. Compare this statistic with the bootstrap statistics. The proportion of bootstrap statistics that exceed the actual statistic yields an approximate p-value.

6.4 Analysis of the polymer surfaces

Computer–generated scene renderings of the six polymer surfaces are shown in Figure 1.3. Based on these pictures, an initial subjective ranking of the surfaces might nominate the surface in panel (d) as the smoothest, the surface in panel (c) as the next smoothest and those in panels (a), (b), (e) & (f) as being equally the roughest. It might also be noted that in panel (d) there is evidence of directional anisotropy rectangular to the edges of the data. If the page is rotated through 90°, all panels show evidence of similar directional anisotropy. These features are confirmed later using the methods developed in the chapter.

Contour plots of the sample variogram for the polymer surface data were generally of elliptical shape near the origin, providing more graphical evidence that the surfaces are anisotropic; see for example Figure 3.4. One possible reason for the anisotropy might be traced back to the directional nature of the extrusion process used to manufacture the polymer. This is a topic that surface scientists might explore if their aim was to produce isotropic surfaces as well as surfaces with the right roughness characteristics. Our interest here is in characterising the roughness, accounting for the anisotropy in our models, and seeking to quantify the anisotropy.

To this end, estimates $\hat{\alpha}$ and $\hat{c}_i$ were calculated from (6.3) and (6.4) respectively for each data set, using those θ_i's for which there existed $r_{ij} \leq 5$. Figure 6.2 illustrates a scatterplot of pairs $(\theta_i, \hat{c}_i)$ for the third polymer surface, panel (c) of Figure 1.3. The LOESS statistical package (Cleveland & Devlin, 1988; Cleveland &

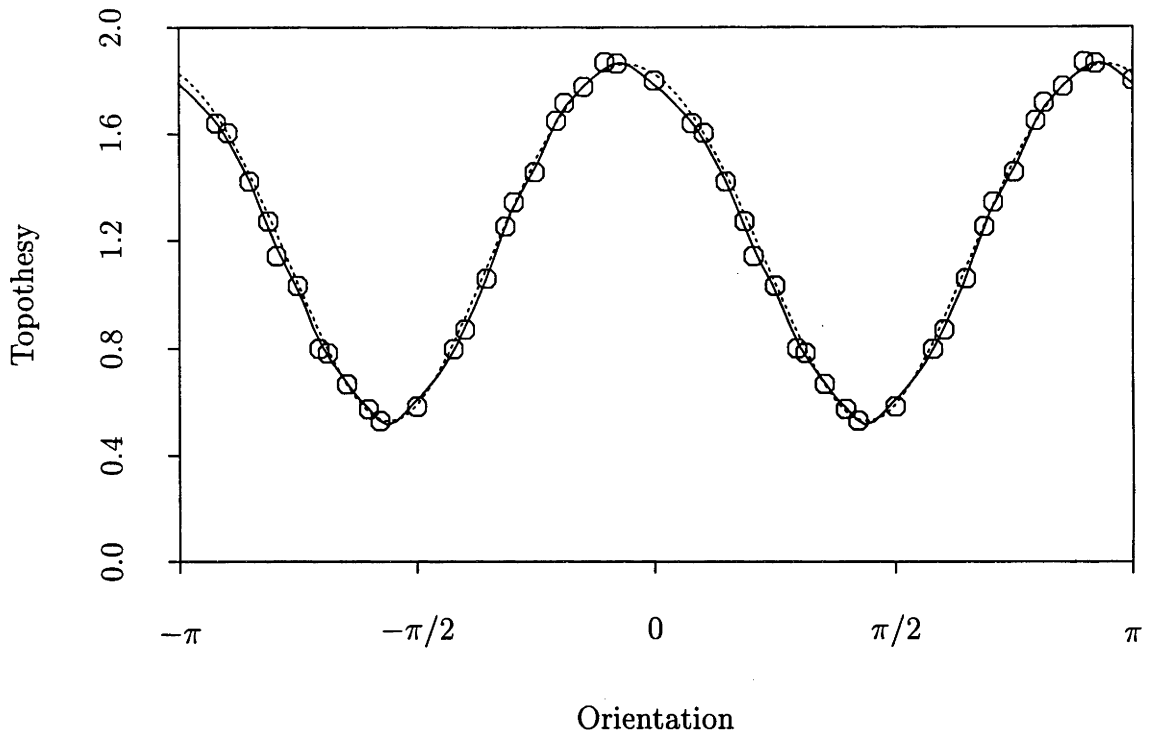


Figure 6.2: Estimates of the topothesy functions for the third plastic food wrap data set. The points are plotted at $(\theta_i, \widehat{c}_i)$. The solid line is a local linear smoother fitted to these points. The dashed line is the parametric estimate of topothesy under the assumption of weak anisotropy.

Grosse, 1991), employing a tricubic kernel and an average bandwidth of $\pi/10$, was used to produce the local linear smooth depicted by the solid line . (The amount of smoothing needed to produce graphs such as Figure 6.2 is usually easy to determine by eye, since the points on the curve are themselves estimates and so have relatively little noise.)

Fractal indices $\hat{\alpha}$ were used to compute estimators of fractal dimension by substituting into the formula $\hat{D} = 3 - \frac{1}{2}\hat{\alpha}$. Their values are listed in the row labelled "Fractal dimension (i)" in Table 6.1 .

Because the contours of the empirical variogram were elliptical in nature for each data set, the weak anisotropy model (6.5) was fitted, using ordinary least-squares, to the log of the empirical variogram in identical neighbourhoods of the origin.

Data set	(a)	(b)	(c)	(d)	(e)	(f)
Fractal dimension (i)	2.50	2.49	2.30	2.19	2.47	2.59
	(0.010)	(0.012)	(0.014)	(0.011)	(0.012)	(0.009)
Fractal dimension (ii)	2.50	2.48	2.31	2.19	2.46	2.58
	(0.012)	(0.015)	(0.016)	(0.012)	(0.015)	(0.012)
Average topothesy, c_0	1.01	0.63	1.28	4.11	1.30	1.36
	(0.014)	(0.012)	(0.043)	(0.194)	(0.025)	(0.019)
Strength of anisotropy, s	0.41	0.67	0.72	0.69	0.60	0.59
	(0.036)	(0.025)	(0.023)	(0.025)	(0.030)	(0.025)
Lay, ψ	0.12	0.12	-0.18	-0.04	0.17	0.09
	(0.052)	(0.030)	(0.026)	(0.029)	(0.035)	(0.031)

Table 6.1: The estimated fractal dimensions, average topothesies, strengths of anisotropy, and lays (in radians), for the six polymer surface data sets. The figures in parentheses are bootstrap estimates of the associated standard errors.

The corresponding estimators of fractal dimension ("Fractal dimension (ii)", in the second row), average topothesy c_0, strength of anisotropy s and lay ψ are listed in Table 6.1, along with bootstrap estimates of error in parentheses.

The fitted topothesy function implied by the total fitted model, in the case of the third polymer data set, is depicted by the broken line in Figure 6.2. This is to be compared with the scatterplot and the local linear smooth. Note that both curves differ from a pure sinusoid in that the curvature of the crests is less than that of the troughs.

The results in Table 6.1 suggest that the fourth data set has lower fractal dimen-

Data sets	statistic	P-value
(a) & (b)	0.016	0.637
(c) & (d)	0.133	0.000
(e) & (f)	0.125	0.000

Table 6.2: Test statistics and associated P-values for parametric bootstrap tests for common fractal dimension between successive pairs of the six plastic food wrap data sets.

sion than the others and, in partial compensation, higher average topothesy. This surface may show evidence of two fractal dimensions at different lags, or scales: see panel (b) of Figure 6.1, where the lines corresponding to larger lags (the broken lines) have lesser slope. However, this is probably due, at least in part, to greater error in estimation of the variogram, mainly owing to the greater bias at greater lags.

From a material science point of view, the estimates of fractal dimension provide criteria for ranking polymer surfaces in order of roughness. Furthermore, the resulting rankings agree with subjective assessment. This provides important information for determining the manufacturing process that produces the polymer with the most desirable properties.

As mentioned in section 6.1, the polymer surfaces sets were manufactured by two slightly different processes, those depicted in panels (a), (c) & (e) of Figure 1.3 by one process and those shown in panels (b), (d) & (f) by the other. Successive pairs of data sets had similar input parameters for each process. One of the aims of the experiment was to determine whether the roughness properties of surfaces manufactured by the first process were different from those for the second process, for a range of input parameter values.

Table 6.2 contains results from bootstrap hypothesis tests for common fractal

dimension. The inference is that the difference in fractal dimension for the first pair of surfaces is possibly caused by stochastic error, whereas differences for the other pairs are most likely not due to stochastic error.

Bibliography

Adler, R. J. (1981). *The Geometry of Random Fields.* New York: Wiley.

Amis, A. (1996). The effect of surface roughness on fibroblast adhesion in vitro – annotation. *Injury – International Journal of the Care of the Injured 27*, 43–43.

Baczkowski, A. J. and K. V. Mardia (1987). Approximate lognormality of the sample semi-variogram under a Gaussian process. *Communications in Statistics B, (Simulation and Computation) 16*, 571–585.

Baczkowski, A. J. and K. V. Mardia (1990). A test of spatial symmetry with general application. *Communications in Statistics A, (Theory and Methods) 19*, 555–572.

Barnsley, M. (1988). *Fractals Everywhere.* New York: Academic Press.

Berry, M. V. (1979). Diffractals. *Journal of Physics A 12*, 781.

Berry, M. V. and J. H. Hannay (1978). Topography of random surfaces. *Nature 273*, 573.

Chan, G., P. Hall, and D. S. Poskitt (1995). Periodogram-based estimators of fractal properties. *Annals of Statistics 23*, 1684–1711.

Cleveland, W. (1993). *Visualizing Data.* New Jersey: Hobart Press.

Cleveland, W. and S. Devlin (1988). Locally-weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association 83*, 596–610.

Cleveland, W. and E. Grosse (1991). Computational methods for local regression. *Statistics and Computing 1*, 47–62.

Constantine, A. G. and P. Hall (1994). Characterising surface smoothness via estimation of effective fractal dimension. *Journal of the Royal Statistical Society, Series B, (Methodological) 56*, 97–113.

Cressie, N. (1985). Fitting variogram models by weighted least squares. *Journal of the International Association of Mathematical Geology 17*, 693–702.

Cressie, N. (1991). *Statistics for Spatial Data.* New York: John Wiley.

Davies, S. and P. Hall (1999). Fractal analysis of surface roughness by using spatial data. *Journal of the Royal Statistical Society, Series B, (Methodological) 61*, 3–37.

Davis, B. and L. Borgman (1982). A note on the asymptotic distribution of the sample variogram. *Journal of the International Association for Mathematical Geology 14*, 189–193.

Dubuc, B., J. F. Quiniou, C. Roques-Carmes, C. Tricot, and S. W. Zucker (1989). Evaluating the fractal dimension of profiles. *Phys. Rev. Ser. A 39*, 1500–1512.

Falconer, K. (1985). *The Geometry of Fractal Sets.* Cambridge, U.K.: Cambridge University Press.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Annals of Statistics 21*, 196–216.

Federer, H. (1969). *Geometric Measure Theory.* Berlin: Springer.

Hall, P. and S. Davies (1995). On direction-invariance of fractal dimension on a surface. *Applied Physics, Series A 60*, 271–274.

Hall, P. and R. Roy (1994). On the relationship between fractal dimension and fractal index for stationary stochastic processes. *Annals of Applied Probability 4*, 241–253.

Hall, P. and A. Wood (1993). On the performance of box-counting estimators of fractal dimension. *Biometrika 80*, 246–252.

Hassan, A. (1997). The effects of ball- and roller-burnishing on the surface roughness and hardness of non-ferrous metals. *Journal of Materials Processing Technology 72*, 385–391.

Hastie, T. and C. Loader (1993). Local regression: automatic kernel carpentry. *Statistical Science 8*, 120–143.

Hsü, K. J. and A. J. Hsü (1990). Fractal geometry of music. *Proceedings of the National Academy of Science 87*, 938–941.

Jain, P. S. (1986). Fractal dimensions of clouds around Madras. *Mausam 40*, 311–316.

Johnson, N. and S. Kotz (1970). *Distributions in Statistics*, Volume 3. Continuous Univariate Distributions, Part 2 of *Houghton–Mifflin Series in Statistics.* Boston: John Wiley & Sons.

Kent, J. T. and A. T. A. Wood (1997). Estimating the fractal dimensions of a locally self-similar Gaussian process using increments. *Journal of the Royal Statistical Society, Series B, (Methodological) 59*, 679–699.

Lewin, R. (1991). The fractal structure of music. *New Scientist 130*, 15.

Lo, T., H. Leung, J. Litva, and S. Haykin (1993). Fractal characterisation of sea-scattered signals and detection of sea surface targets. *IEE Proceedings, Series F 140*, 243–250.

Mandelbrot, B. B. (1975). *Les Objets Fractals.* Paris: Flammarion.

Mandelbrot, B. B. (1977). *Fractals: Form, Chance and Dimension.* New York: Freeman.

Mandelbrot, B. B. (1982). *The Fractal Geometry of Nature.* New York: Freeman.

Mandelbrot, B. B. (1985). Self-affine fractals and fractal dimension. *Physica Scripta 32*, 257–260.

Mandelbrot, B. B., D. E. Passoja, and A. J. Paullay (1984). Fractal character of fracture surfaces of metals. *Nature 308*, 721–722.

Marstrand, J. (1954). Some fundamental geometrical properties of plane sets of fractal dimensions. *Proceedins of the London Mathematical Society 4*, 257–302.

Matheron, G. (1971). *The Theory of Regionalized Variables and its Applications.* Fontainebleau: Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau.

Matthews, D. (1998). *On Fractal Dimension and its Application.* Ph. D. thesis, Australian National University, Canberra.

Mattila, P. (1985). On the Hausdorff dimension and capacities of intersections of sets in n-space. *Acta Mathematica Scientia 152*, 77–105.

McCarrol, D. and A. Nesje (1996). Rock surface roughness as an indicator of degree of rock surface weathering. *Earth Surface Processes and Landforms 21*, 963–977.

Milne, B. T. (1991a). Lessons from applying fractal models to landscape patterns. In M. G. Turner and R. H. Gardner (Eds.), *Quantitative Methods in Landscape Ecology*, pp. 199–235. New York: Springer.

Milne, B. T. (1991b). The utility of fractal geometry in landscape design. *Landscape and Urban Planning 21*, 81–90.

Morrison, A. I. and M. A. Srokosz (1993). Estimating the fractal dimension of the sea surface: a first attempt. *Annals Geophysicae 11*, 648–658.

Rosenblatt, M. (1961). Independence and dependence. In J. Neyman (Ed.), *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, pp. 411–433. University of California Press.

Sayles, R. S. and T. R. Thomas (1978a). Sayles and Thomas reply. *Nature 273*, 573. Reply to Berry & Hannay (1978).

Sayles, R. S. and T. R. Thomas (1978b). Surface topography as a nonstationary random process. *Nature 271*, 431–434.

Stout, K. J., P. J. Sullivan, W. P. Dong, E. Mainsah, N. Luo, T. Mathia, and H. Zahouani (1993). *The Development of Methods for the Characterisation of Roughness in Three Dimensions.* Commission of the European Communities.

Swart, P., B. Lacquet, and C. Blom (1996). An acoustic sensor system for determination of macroscopic surface roughness. *IEEE Transactions on Instrumentation and Measurement 45*, 879–884.

Taqqu, M. (1988). Self-similar processes. In S. Kotz and N. Johnson (Eds.), *Encyclopedia of Statistical Sciences*, Volume 8, pp. 352–357. New York: Wiley.

Taylor, C. C. and S. J. Taylor (1991). Estimating the dimension of a fractal. *Journal of the Royal Statistical Society, Series B, (Methodological) 53*, 353–364.

Thomas, F. and M. Atkinson (1997). Ammonium uptake by coral reefs – effects of water velocity and surface roughness on mass transfer. *Limnology and Oceanography 42*, 81–88.

Thomas, T. R. (1982). *Rough Surfaces.* Harlow: Longman.

Thomas, T. R. and A. P. Thomas (1988). Fractals and engineering surface roughness. *Surface Topography 1*, 143–152.

Tricot, C. (1993). *Curves and Fractal Dimension.* New York: Springer.