THE DETECTION OF CHANGE

IN SPATIAL PROCESSES

WITH

ENVIRONMENTAL APPLICATIONS

ELAINE B. MARTIN B.Sc. B.Sc.

A Dissertation Submitted To The

University of Glasgow

For The Degree of Doctor of Philosophy

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

ABSTRACT

Ever since Halley (1686) superimposed onto a map of land forms, the direction of trade winds and monsoons between and near the tropics and attempted to assign them a physical cause, homo-sapiens has attempted to develop procedures which quantify the level of change in a spatial process, or assess the relationship between associated spatially measured variables.

Most spatial data, whether it be originally point, linear or areal in nature, can be converted by a suitable procedure into a continuous form and plotted as an isarithmic map i.e. points of equal height are joined. Once in that form it may be regarded as a statistical surface in which height varies over area in much the same way as the terrain varies on topographic maps. Particularly in environmental statistics, the underlying shape of the surface is unknown, and hence the use of non-parametric techniques is wholly appropriate. For most applications, the location of data points is beyond the control of the map-maker hence the analyst must cope with irregularly spaced data points. A variety of possible techniques for describing a surface are given in chapter two, with attention focusing on the methodology surrounding kernel density estimation.

Once a surface has been produced to describe a set of data, a decision concerning the number of contours and how they should be selected has to be taken. When comparing two sets of data, it is imperative that the contours selected are chosen using the same criteria. A data based procedure is developed in chapter three which ensures comparability of the surfaces and hence spurious conclusions are not reached as a result of inconsistencies between surfaces. Contained within this chapter is a discussion of issues which relate to other aspects of how a contour should be drawn to minimise the potential for inaccuracies in the surface fitting methodology.

Chapter four focuses on a whole wealth of techniques which are currently available for comparing surfaces. These range from the simplest method of overlaying two maps and visually comparing them to more involved techniques which require intensive numerical computation. It is the formalisation of the former of these techniques which forms the basis of the methodology developed in the following two chapters to discern whether change/association has materialised between variables.

One means of quantifying change between two surfaces, represented as a contoured surface, is in terms of the transformation which would be required for the two surfaces to be matched. Mathematically, transformations are described in terms of rotation, translation and scalar change.

Chapter five provides a geometrical interpretation of the three transformations in terms of area, perimeter, orientation and the centre of gravity of the contour of interest and their associated properties. Although grid resolution is fundamentally a secondary level of smoothing, this aspect of surface fitting has generally been ignored. However to ensure consistency across surfaces, it is necessary to decide firstly, whether data sets of different sizes should be depicted using different mesh resolutions and secondly, how fine a resolution provides optimal results, both in terms of execution time and inherent surface variability. This aspect is examined with particular reference to the geometric descriptors used to quantify change.

The question of random noise contained within a measurement process has been ignored in the analysis to this point. However in practice, some form of noise will always be contained within a process. Quantifying the level of noise attributable to a process can prove difficult since the scientist may be over optimistic in his evaluation of the noise level. In developing a suitable set of test statistics, four situations were examined, firstly when no noise was present and then for three levels of noise, the upper bounds of which were 5, 15 and 25%.

Based on these statistics, a series of hypothesis tests were developed to look at the question of change for individual contour levels i.e. local analysis, or alternatively for a whole surface by combining the statistics and effectively performing a multivariate test. A number of problems are associated with the methodology. These difficulties are discussed and various remedial measures are proposed.

The theoretical derivation of the test statistic, both in the absence and presence of random noise, has proved mathematically to be extremely complex, with a number of stringent assumptions required to enable the theoretical distribution to be derived. A major simulation study was subsequently undertaken to develop the empirical probability distribution function for the various statistics defining change for the four levels of noise. Also for each of the statistics, the resultant power of the test was examined.

The remaining chapter explicitly examines two case studies and how the methodology developed in the preceding two chapters may be implemented. The first example cited raises the question, 'Has a seasonal temperature change resulted during the fifty year span, 1930 to 1980, within the contiguous United States of America?' The data base was provided by the United States Historical Climatology Network (HCN) Serial Temperature and Precipitation Data, Quinlan et al (1987).

The second problem examines whether there is an association between background radiation levels, within three regions of the south-west England, and the location of various forms of leukaemia or whether case location is a product of the population distribution. Differences between this example and the previous illustration materialise in terms of the spatial resolution of the data; the leukaemia data are defined as punctual data points and are extremely sparse; the population distribution is defined as areal regions; with the radiation data being of a more continuous format. The methodology developed required modification, but aside of this a preliminary set of conclusions were reached.

ACKNOWLEDGEMENTS

The author would like to express her sincere gratitude to her supervisor, Dr. E.M. Scott, for her encouragement and guidance throughout the period of her research. In addition, the author would like to thank the members of the statistics department at Glasgow University for their support and friendship during this time.

The author is indebted to Dr. D. Sanderson for the many stimulating discussions, the opportunity to partake in the collection of aerial survey radiation data and finally, for allowing her to make use of previously collected radiation data.

The author would also like to thank Dr. F. Alexander and the Leukaemia Research Fund for providing the case/control data for the leukaemia study.

This research has been supported by the Science and Engineering Research Council.

Finally, the author would like to express her heartfelt gratitude to her family for their tolerance and understanding throughout the years.

Table of Contents

List of Figures

Chapter 5 The Charaterisation of Spatial Change

Chapter 6 An Hypothesis Testing Approach to Surface Comparison

List of Tables

Chapter 1 Spatial Data Analysis - Its Application

Chapter 2 Surface Mapping

Chapter 3 Surface Representation

Chapter 4 Different Methods of Surface Comparison

Chapter 7 :- The Application Of The Methodology To Two Environmental Case Studies

CHAPTER 1

SPATIAL DATA ANALYSIS - ITS APPLICATION

§1.1 Introduction :- Statistics and Spatial Data

The first manifestation of statistics for spatial data appears to have arisen in the form of data maps, Halley (1686) superimposed onto a map of land forms the direction of trade winds and monsoons between and near the tropics, and attempted to assign them a physical cause.

Spatial models appeared much later. Student (1907) examined the distribution of particles throughout a liquid and instead of analysing their spatial positions, he aggregated the data into counts of particles per unit area and found they followed a Poisson distribution. Fisher (1935) during the 1920's and 1930's established the principles of randomisation, blocking and replication. Although these control for bias and randomisation neutralises the effect of spatial correlation, they do not neutralise the spatial correlation at spatial scales larger or smaller than the plot dimension.

More recently nearest neighbour techniques for analysing agricultural field trials have attempted to take spatial dependence into account by using residuals from neighbouring plots as covariates or by differencing, Besag and Kempton (1986).

In areas such as geology, ecology and environmental science, it is not often possible (or appropriate) to randomise, block and replicate the data. There is therefore a need for new statistical methods and approaches that address new questions arising from old and new techniques. Many of the resulting problems such as resource assessment, environmental monitoring and medical imaging are spatial in nature.

The development of statistical procedures to quantify the level of change in an environmental process or assess the relationship between associated spatially measured variables is an area of considerable statistical and practical interest. Much of the analysis of spatially collected data is concerned with the identification and explanation of spatial structure, and with the analysis and explanation of possible links between two spatial processes e.g. humans and the environment.

The form of a spatial analysis undertaken is in part conditioned by the dimensionality of the data, Unwin (1981) recognised four separate classes; point, line, area and surface. Within each category techniques both descriptive and analytical exist, table 1.1 collates some of these methods.

Points	Lines
mean centre/standard distance	random walk
standard deviational ellipse	vectors
gradient analysis	graph theory :-
nearest neighbour analysis	nodality
variance/mean ratio test	connectivity
quadrant analysis	flow analysis
space clustering	dispersion

Area	Surfaces
location quotients	isolines
space clustering	trend surface analysis :-
poisson probability	power series polynomials
hierarchial clustering	fourier series

Table 1.1 :- Spatial analysis techniques

Methodological problems stem from the fact that spatial patterns are frequently intricate and complex, and to date many investigators have grossly simplified these patterns.

An additional problem of applying statistics to spatial data concerns the problem of scale. Often spatial patterns or variable associations show striking differences at different scales of analysis. It can be argued that scale should be used creatively, Cleek (1979). For one thing, replication of findings at different scales tends to confirm hypothesis. Furthermore, White (1972) stated that one can try to identify the scale at which a certain process is most effective: this in itself may provide clues as to how the process works. As in all forms of analysis the reliability and validity of the data will potentially limit the value of sophisticated statistical models and techniques. Methods which account for random error in a spatial variable may help to extend the feasibility and power of the technique.

In different areas of application, various specialised difficulties will arise, each separate analysis must take account of these localised difficulties e.g. in medical geography, the spatial position of an individual at the time of diagnosis is taken to be the current place of residence. For chronic diseases such as cancer, which have a long latent period, the place of diagnosis may have little aetological significance.

All the problems discussed are confined to individual sets of spatial data. Effectively performing a comparison of spatial variables increases the number of potential pitfalls. The development of a method which invokes only automatic procedures will reduce the possibility of attaining spurious conclusions regarding associations/differences.

The first step in developing such a framework is to convert the data, of whatever dimensionality, to a continuous form, by one of the available interpolation techniques. Once in this format it may then be presented as a contoured surface, the question of contour selection and fitting is an essential stage of the analysis since incompatible selection may result in inaccurate statements.

Once both these stages have been successfully implemented, the question of comparison may be examined. Intuitively, the most appealing means of comparison is to superimpose the surfaces and describe the change in terms of various geometric properties of the contour. Based on this notion, a series of statistical tests were constructed to look at change/association between variables.

§1.2 Surface Mapping

The definition of a map is given to be a representation of (part of) the earth's surface, showing physical and political features. In mathematics, a mapping is described as a translation from one vector space to another. For the practitioner, the first vector space is the real multidimensional world and the second is the face of a sheet of paper on which the map is drawn. From this definition, a map and surface are one and the same since a surface is described as having length and breadth but no thickness. The paper accordingly restricts a map to a three dimensional vector space where the length and breadth represent the spatial co-ordinates and the z-variable, the phenomenon being mapped, is defined in terms of lines or symbols.

Representing spatial information in terms of surfaces/maps has four advantages:-

1. They provide a synoptic expression of the information represented.

2. They enable a number of important spatial properties that were not initially measured to be isolated e.g. orientation, shape and relative location. Maps/surfaces therefore contain spatial structure.

3. They are models of the real world.

4. Finally, they are communication devices which may be used to express ideas or generate hypothesis.

However, maps/surfaces contain three theoretical difficulties as models of the real world:-

1. They are static and cannot be drawn to include a time dimension.

2. They are the result of a spatial process which theoretically may be defined as:-

$$z_i = f(x_i, y_i)$$

The z-values above result from a deterministic process but in practice they are the result of a stochastic process in which chance variation, ε_i, is unquantifiable:-

$$z_i = f(x_i, y_i) + \varepsilon_i$$

3. Maps/surfaces are constrained by humans and may be used to display information inadequately or alternatively the data used in their construction may be of poor quality.

The level of measurement of a spatially distributed variable is a basic control on the choice of map type, method of analysis and ultimately on the nature of the inferences that can be drawn from a study of a variable's statistical structure.

Stevens (1946) identified four basic levels; nominal, ordinal, interval and ratio. The properties of which are discussed by Siegel (1956). Nominal data serve merely as symbols and cannot be manipulated mathematically, the simplest form are dichotomous variables e.g. yes/no. Statistical and cartographic operations are limited, but a number of primary operations may be executed. The second form of data, ordinal, assumes relationships are both transitive and asymmetric. Statistically the remaining two levels, ratio and interval, may be treated in a similar manner since both allow the distance between categories to be defined for equivalents units of measurement.

Apart from the hierarchical property of level of data, measurement has the characteristic of dimensionality. Non-dimensional variables play an important role where comparisons between sets of numbers are to be performed. Dimensional analysis is more restrictive in practice but is possibly more valuable in applied work.

In theory there is almost no limit to the choice of units of measurement and dimensions of data which may fall into any one of the four categories. However locational data is restricted to one of four types of geographical data; point, line, area or surface. Table 1.2 summarises the various types of maps which may be produced using such information, Unwin (1981).

	Dimension			
Level	Point	Line	Area	Surface
Nominal	Dot maps	Network map	Coloured area map	Freely coloured map
Ordinal	Symbol map	Ordered network map	Ordered coloured map	Ordered chorochromatic map
Interval and Ratio	Graduated symbol map	Flow map	Choropleth map	Contour type map

Table 1.2 : Various combinations of map types

Nowadays a surface is a starting point of an analysis. The next stage is to summarise and describe the distribution prior to examining whether or not a recognisable pattern exists. For each of the twelve possible map forms, various types of analysis can be undertaken. Unwin (1981) examines some of these procedures. Dimension is the strongest single influence on the type of statistical inference which may be undertaken.

A further feature which determines the type of analysis which is appropriate relates to whether or not the data is isotropic or anisotropic. The former describes a data set in which the properties are the same in any direction whilst for the later, these vary with direction. Analysis of an anisotropic surface is more complex since the direction of variability is not always known.

Particularly in environmental statistics, the underlying shape of the surface is unknown hence the use of non-parametric techniques is wholly appropriate. For most applications, the location of data points is beyond the control of the map-maker and does not correspond to a geometric pattern. In these cases, the analyst must cope with irregularly spaced data points. The mapped surface need not necessarily pass through the points exactly hence the irregular location of points is not a complication. Other techniques necessarily pass through all the points. A more detailed discussion of possible techniques for describing a surface are given in chapter two with attention focusing on the methodology surrounding kernel density estimation.

Density estimation provides a natural and intuitive means for describing the data, and the technique does not differ vastly from some of the more traditional approaches for representing spatial data such as nearest-neighbour and kriging. It is this technique which has been used in the development of surfaces for subsequent analysis.

§1.3 Surface Representation

Viewing locational data as a statistical surface Robinson (1961) suggested that much of morphometric analysis may be applicable to the 'topography' of isarithmic surfaces. Although isarithmic maps are three-dimensional in the sense that a distribution varies continuously over an area, they are commonly referred to as single component maps since only one component, the z-variable, is being mapped. The most common method of showing variation in the z-dimension is through the use of isolines or perspective plots. A wide variety of terms have been used to describe surfaces of this form, the most

general being, isoline and isarithm. Table 1.3 summarises some of the more specialised terminology.

Isoline	Quantity represented	Isoline	Quantity represented
Contour	Equal heights	Isotherms	Equal temperatures
Isovels	Equal speed	Isoheyts	Equal rainfall
Isonoets	Equal intelligence	Isohels	Equal sunshine
Isotachs	Equal wind speed	Isobar	Equal pressure
Isophers	Equal freight rate	Isochrone	Equal time/distance

Table 1.3 :- Various specialised forms of isolines/isarithms.

Clearly such illustrative procedures are based on the assumption that the distribution is continuous. Warntz (1959) demonstrated that considerable theoretical and operational gains were to be had by regarding such variations as continuous.

Two types of isolines exist, the first refers to the more conventional form where the surface exists at all points. The second variant describes maps in which the isolines display spatial variation in derived quantities, that are themselves related to some area e.g. population density. These are more commonly referred to as isopleths. Problems associated with such maps are, unlike isolines, the spaces between the isopleths do not have any values related to them. Secondly the form of the isoline is dependent on :-

1. The shape of the areal units.

2. The location of the control points

In the 1950's a number of authors advocated their use particularly Mackay (1951), Schmid and MacCannell (1955), and Porter (1958) whilst more recently Nordbeck and Rystedt (1970) and Tobler and Lau (1970) have also supported their application. There are however situations where it is prudent to retain the data in their discontinuous form. The non-overlapping cells are mapped as choropleth maps with appropriate values assigned to each cell. Analysis should then proceed by area-based methods.

Figure 1.1 summarises the relationship between measured point data, networks and contours in surface mapping.

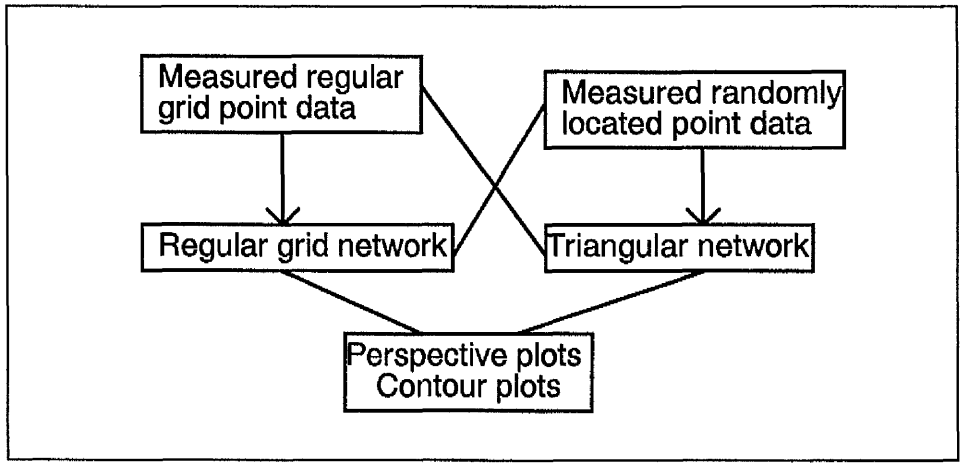


Figure 1.1 :- Relationship between point data and plot types.

A perspective plot takes a point, P, in three-dimensional space and transforms it to a point in two-dimensions, Q, its two-dimensional image. The general isometric projection is the simplest type. To evaluate Q from P, P is multiplied by:-

$$\begin{pmatrix} \cos\alpha & -\cos\beta & 0 \\ \sin\alpha & \sin\beta & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

The standard isometric projection sets $\alpha = \beta = 30°$.

Once an isarithmic surface has been produced to describe a set of data, a decision concerning the number of contours and how they should be selected has to be taken. When comparing two sets of data, it is imperative that the contours selected are chosen using the same criteria. A data based procedure is developed in chapter three which ensures comparability of the surfaces of interest, and hence avoids spurious conclusions being reached, as a result of inconsistencies between surfaces. Contained within this chapter are other issues which relate to how a contour should be drawn to minimise the potential for inaccuracies in the surface fitting methodology.

§1.4 Surface Comparison

A great deal of interest has focused on the detection of common or differing features between two or more distinct variables in various fields of application e.g. image processing, geology, medical geography etc. Invariably we are attempting to make some objective comments concerning the existence of an 'areal ' association between variables, or alternatively, trying to discern the type of change which may have resulted due to some, possibly unknown, underlying process.

However caution should be displayed when interpreting the findings. Where a similarity is seen to exist between two variables, a causal link may not necessarily hold. Similar forms can be generated by markedly different processes and only seldom will any single analysis be able to discriminate between them. Although this observation suggests that this mode of analysis is of little value, in many situations it is an important step to identifying the possible processes which may be responsible for the change.

Surface comparison is further complicated by the surfaces themselves, like must be compared with like, or the association may be a by-product of the technique and not the processes hence the emphasis on surface construction in chapter three.

The number of types of comparisons which can be made is enormous. Restricting ourselves to the four basic types of data, Table 1.4 shows that there are at least ten different types of comparisons.

DISTRIBUTION A	DISTRIBUTION B			
	point	line	area	surface
point	1	2	3	4
line		5	6	7
area			8	9
surface				10

Table 1.4 :- Types of comparison involving the four basic data types.

Comparisons between similar distributions is the most obvious e.g. area distribution of woodland and soil type (8). Nevertheless comparisons of distributions of differing spatial

dimensions may be undertaken e.g. point location of leukaemia sufferers and the spatial dispersion of a specific radionuclide (4) but potentially different types of analysis will be required. (This problem is dealt with more fully in Chapter 7).

Before attempting to compare surfaces, the analyst must first define their objectives and determine what is to be compared. Also, consideration must be given to the difference between a single-valued comparison function, that gives an overall comparison, and a spatial comparison or function which produces a map of areal goodness-of-fit that permits a two-dimensional similarity interpretation. Chapter four focuses on a whole wealth of techniques, currently available. These range from the simplest method of overlaying two maps and visually comparing them, to more involved techniques which require intensive numerical computation. It is the formalisation of the former of these techniques, which forms the basis of the methodology developed in chapters five and six to discern whether change/association has materialised between variables, or over time.

One means of quantifying change between two surfaces, represented as contoured surfaces, is in terms of the effective transformation which would be required for the two surfaces to be matched. The most usual mathematical definition of transformation is in terms of quantities relating to rotation, translation and scalar changes.

This approach has previously been used to collate two sets of images i.e. image registration. In both medical and satellite imagery, these transformation measures are generally considered to be nuisance parameters and the data is accordingly transformed to remove these factors. The technique for isolating these change parameters is based on the movement of easily identifiable points on a surface, landmark data. These points are not anticipated to deviate in locality over time or between variables. Once the two surfaces have been 'matched', change is invariably described in a subjective rather than quantitative manner, Coombes et al. (1991).

In environmental applications, two problems are associated with such an approach, firstly it is unusual to have data of this format, since in field studies it is extremely difficult to return to the same site where the initial measurement was recorded after time, t, primarily because of practical difficulties and secondly, the type of transformation causing the change is of considerable importance to the earth scientist since it may give them a handle on the underlying process contributing to the change.

Both the validity and the properties of the four measures area, perimeter, orientation and centre of gravity are assessed in chapter five as to their feasibility for forming the basis of measures used to quantify the three transformations, rotation, translation and scalar change. The question of grid resolution is examined specifically in terms of these descriptors. This aspect of surface fitting acts as a secondary level of smoothing, the finer the mesh, the more smooth the resultant picture. In terms of a comparison it is essential that compatibility across surfaces is achieved. First, we must decide whether mesh resolution should vary with data set size and secondly what an optimal resolution is in terms of time and resultant surface variability.

The question of random noise contained within a measurement process has been ignored in the analysis to this point, in practice, some form of noise will always be contained within a process. Quantifying the level of noise attributable to a process can prove difficult since the scientist may be over optimistic in their evaluation of the noise level. In developing a suitable set of test statistics, four situations were examined, firstly when no noise was present and then for three levels of noise, the upper bounds of which were 5, 15 and 25%.

Based on these statistics, a series of hypothesis tests were developed, both to assess change at a local level i.e. change between individual contour levels, and on a global scale. A number of problems are discussed and various remedial measures are proposed.

The theoretical derivation of the test statistic, both in the absence and presence of random noise, has proved mathematically, extremely complex. A number of stringent assumptions would be required to enable the theoretical distribution to be derived. A major simulation study was subsequently undertaken to develop the empirical probability distribution function for each of the various statistics defining change for the four levels of noise. Also for each of the statistics, the resultant power of the test was examined.

The remaining chapter explicitly examines two data sets. They illustrate how the methodology developed in the preceding chapters may be transposed to real problems and the problems which materialise when implementing the technique.

The first data set deals with potentially the simplest situation, the comparison of two sets of similar measurements recorded at the same locality. The example is concerned with the topical subject of climate change and the notion of global warming. It examines the

question of whether there has been a change in the seasonal temperature during the 50 year period 1930 to 1980 within the contiguous United States of America. The data base was provided by the United States Historical Climatology Network (HCN) Serial Temperature and Precipitation Data, Quinlan et al (1987).

The second example deals with a more complex situation in which the spatial resolution of the variables of interest differ, the questions of interest are therefore formulated in a different way and the methodology was modified accordingly. The problem examines whether there is an association between the level of background radiation, within three separate regions of the south-west of England, and the location of various forms of leukaemia, or whether the cases are purely a product of the population distribution. Differences between this example and the previous illustration, materialise primarily in terms of the composition of the data; the leukaemia data is defined in terms of punctual data points and is extremely sparse in quantity, the population distribution is defined in terms of areal regions i.e. civil parishes and finally, the radiation data is of a more continuous format.

The latter example highlights the difficulties surrounding analysis of data varying in spatial resolution and quantity, but it also emphasis the potential applicability of the method to problems where the spatial resolution does differ between variables. Although the sparsity of the leukaemia data serves as a restriction to any formal analysis, alternative quantitative methods and subjective techniques founded on the methodology were implemented to answer the question of interest.

CHAPTER 2

SURFACE MAPPING

§2.1 Interpolation Methods

Most locational data, whether it be originally point, linear or areal in nature, can be converted into a continuous form and plotted as a contour map by one of the many available interpolation methods. Once in that form, it may be regarded as a statistical surface in which height varies over area, in much the same way, as terrain varies on a topographic map.

If an investigator has control over the location of data points, they would be arranged in a regular lattice or grid with uniform spacing between points. In this case interpolation could be accomplished by fitting a hyperbolic paraboloid to every four data points by double linear interpolation, Switzer et al. (1964), fitting a polynomial to, up to 25 surrounding data points, by Newton's divided difference formula, Steffensen (1927), Berezin and Zhidkov (1965), or bicubic spline interpolation, de Boor (1962).

For most applications, the location of data points is beyond the control of the map maker and does not correspond to a geometric pattern. In these cases, the analysist must cope with irregularly spaced data points. Sometimes the surface to be mapped need not fit the data points exactly, i.e. the mapped surface need not attain the value of the dependent variable specified at the data point. Subsequently the irregular location of data points will not present a serious complication. An example of such is a trend surface (i.e. a polynomial of specified order) which can be fitted to the data by regression, Krumbein (1959).

For the surface to fit the data points exactly, Bengtsson and Nordbeck (1964) grouped data points into triangles or quadrilaterals. Shepard (1968 a,b) developed an interpolation algorithm, SYMAP, which produced a smooth, continuously differentiable surface which passed through all data points, the method being based on the idea of moving averages.

Table 2.1, although not exhaustive, lists three groups of interpolation techniques with examples of each type:-

Non-parametric density estimation	Trend surfaces	Distance related
Naive estimator	Polynomial	Moving averages
Kernel density estimation	Fourier series	Moving median
Nearest neighbour (balloon density)	Spatial filtering	Varying Quantile Method
Adaptive kernel method	Linear programming	Kriging
Maximum penalised method		Newton's statistical prediction technique
Orthogonal Series estimator		The polygonal method
Reflection and Replication Techniques		Thiessen polygon
Transformation technique		

Table 2.1 :- Summary of various interpolation techniques.

A resumé of four of the above techniques is presented focusing on the theory but also incorporating a number of practical illustrations :-

1. Moving averages
2. Kriging
3. Polynomial trend surface
4. Kernel density estimation

The first three methods are more commonly associated with geological mapping, however more recently the ideas of density estimation has been used to produce surfaces for analysing geographical data, especially in terms of disease mapping and clustering, Bithell (1990).

§2.2 Moving Averages

In the interpolation or girding of spatial data, one of the simplest methods is that of moving averages. A 'true value', z_p, at point P is estimated by taking the average of all surrounding points within a certain limiting radius. This average, z_p, is a linear estimator with weights 1/n on the n points lying within the radius and zero on all other observations. Clarke (1979) showed the resulting estimate of z_p, does not necessarily minimise the sum of squared deviation where no attempt has been made to optimise these weights.

The reason why this estimator is poor is because the value at the centre of the search circle is estimated from observations that lie between the two extremes, i.e. from observations which lie close to the limiting distance and also, from those points which lie close to the centre of the circle, all the points are equally weighted. A modification of this approach that corrects for this failing is to allocate weights, dependent on their distances, from the critical point P, by using a monotone decreasing function e.g. an inverse weighting function i.e. :-

$$z_p = \frac{\sum_{i=1}^{n}\left(z_i / \overline{p}_i\right)}{\sum_{i=1}^{n}\left(1 / \overline{p}_i\right)}$$

where z_i = value of the i^{th} dependent variable to be mapped
$\overline{p}_i$ = distance from P to data point i

If the point P is very near data point i, then $\overline{p}_i$ is small, the further away we move, the larger $\overline{p}_i$ becomes, so the weight accordingly decreases.

Therefore as the distance $\overline{p}_i \rightarrow 0$ the value at P is the limit of:-

$$\frac{\left(z_i / \overline{p}_i\right)}{\left(1 / \overline{p}_i\right)} = z_i$$

Thus for points near data point i, the computed value is near to value z_i.

It may be shown that by computing such a value z_p for every point P over the map area, a 'smooth' surface is obtained. This method has several shortcomings :-

1. As the number of data points becomes large this method for calculating z_p becomes long and inefficient.

2. Direction from data points to the point P is not considered giving interpolated values that are implausible in some situations. Under this simple method, the computed value at P depends only on the distances to the data points 1,2,.....n but not on their location relative to P. For example, assuming $\bar{p}_1$ and $\bar{p}_2$ are held fixed in length, the following configuration of data points give identical values at P:-

+	+	+		+	+	+
1	P	2		P	2	1

3. In practice we can alter the weighting. In the above example of inverse distance weighting, use was made of the actual distance $\bar{p}_i$, more generally, we could weight using $\bar{p}_i^{\,b}$, in which the exponent b is set to some value other than unity. Values greater than this will decrease the relative effect of distant points whilst, for values less than one, the importance of distant points will be increased. Many commercial computer programs use this method with b=2, giving an inverse distance squared weighting. With inverse square weighting, the surface is level at every data point (i.e. the derivative is zero) regardless of the location of that data point or its data value.

A number of embellishments were made to this technique to compensate for the above deficiencies, Shepherd (1968 a,b).

This idea is extended further in the description of kriging which takes into account both the variation in the measurements and distances between the measurement points

§2.3 Kriging

Many geological surfaces, both real and conceptual, can be regarded as regionalised variables which have properties intermediate between a truly random variable and one completely deterministic. Typically, regionalised variables are functions describing natural phenomena that have geographic distributions such as the elevation of the ground surface, or changes in grade within an ore body. Unlike random variables, regionalised variables have continuity from point to point but the changes in the variable are so complex that they cannot be described by any tractable deterministic function.

Even though a regionalised variable is spatially continuous, it is not usually possible to know its value everywhere. Instead its values are known only through samples taken at specific locations.

Geostatistics was originally developed by Georges Matheron of the Centre de Morphologic Mathematique in Fontainebleau, France. It involves estimating the form of a regionalised variable in one, two and three dimensions. One of the basic statistical measures of geostatistics is the semi-variance which is used to express the ratio of change of a regionalised variable along a specific orientation i.e. the semi-variance is a measure of the degree of spatial dependence between samples along a specific support.

§2.3.1 Semi-variogram

For simplicity assuming the samples are point measurements and the support is regular, the semi-variance, γ_h, is defined as:-

$$\gamma_h = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} \left(z(x_i) - z(x_i + h) \right)^2$$

where

$z(x_i)$ = measurement of a regionalised variable taken at location i.

$z(x_i + h)$ = measurement taken h intervals away.

$n(h)$ = number of pairs of samples at distance h apart.

If we evaluate the semi-variances for different values of h, the results may be plotted in the form of a semi-variogram. For small h, the points tend to be similar and the semi-

variance will be small, increasing h, the points being compared are less and less closely related and their distances become larger, resulting in a large value of γ_h. At some distance the points being compared are so far apart that they are unrelated and their squared difference becomes equal in magnitude to the variance around the average value. The semi-variogram consequently develops a flat region, the sill. The distance at which the sill is attained is termed the span or range. Figure 2.1 illustrates a hypothetical semi-variogram. The nugget effect i.e. the value of the semi-variogram at h=0, corresponds to the mean squared sampling and assaying error.

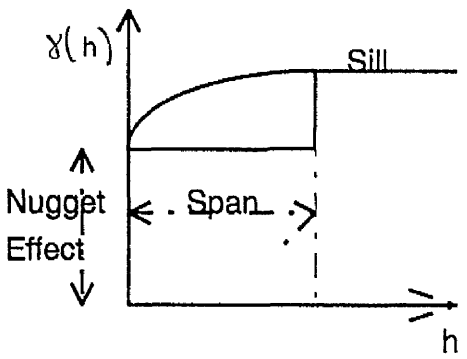


Figure 2.1 :- Example of an hypothetical semi-variogram.

In principle the experimental semi-variogram is known only at discrete points representing distances h, whilst in practice the semi-variance may be required for any distance. For this reason the discrete experimental semi-variogram is modelled by a continuous function that can be evaluated for any desired distance.

Fitting a model equation to an experimental semi-variogram is a trial-and-error process, usually done by eye. Clark (1979) describes and gives examples of the manual process whilst Olea (1977) provides a program that computes a linear semi-variogram having the same slope at the origin as the experimental semi-variogram, he illustrates the technique on a set of geological data.

If the regionalised variable has been sampled at a sufficient density relative to the range, there will be no significant difference between estimates assuming a linear semi-variogram model and other semi-variogram models.

When the form of the semi-variogram is known, it is possible to estimate the value of the surface at any unsampled location, this estimation procedure is termed kriging. The term kriging is derived from the name of D. G. Krige, a South African mining geologist and statistician, who first introduced the idea to avoid systematic overestimation of reserves in the field of mining.

Unlike convential contouring algorithms, this method provides a measure of the error or uncertainty of the contoured surface. Kriging uses the information from the semi-variogram to find an optimal set of weights which are then used to estimate the surface at unsampled locations.

§2.3.2 Various forms of kriging

Six types of kriging are :-

1. Punctual
2. Block
3. Lognormal
4. Disjunctive
5. Universal
6. Generalised covariances

Table 2.2 describes the situation where each of the forms is most applicable.

Distribution of data	Stationarity			
	Stationary	Simple drift	Local trends	Severe anistropy
Normal	Simple kriging (point or block)	Universal kriging	Generalised covariance	?
Simple (known distribution)	Lognormal kriging	?	?	?
Complex	Disjunctive kriging	?	?	?

Table 2.2 :- Situations in which the various types of kriging are applicable.

Punctual kriging is the simplest form of kriging. The data consists of measurements taken at dimensionless points and the estimates are made at other dimensionless points.

For simplicity, we will assume the variable to be mapped is statistically stationary i.e. free from drift, for a regionalised variable z(x) :-

$$E[z(x)] = m \qquad \text{constant}$$
$$E[z(x) - m][z(x') - m] = \text{cov}(x - x')$$

i.e. the covariance between two points x and x' does not depend separately on x and x', but only on the vector $(x\text{-}x')$.

The value at an unsampled location may be estimated as a weighted average of the known observations, with the resultant value at point P given by :-

$$\hat{X}_p = \sum_{i=1}^{n} w_i x_i \qquad \text{where} \qquad \sum_{i=1}^{n} w_i = 1$$

The estimate $\hat{X}_p$ will differ from the true, unknown value X_p by an amount termed the estimate error, ε_p:-

$$\varepsilon_p = (X_p - \hat{X}_p)$$

There are an infinite number of possible weight combinations that can be selected, each gives a different estimate and estimation error, however only one combination produces a minimum estimation error, it is this unique combination of weights that kriging aims to find.

A complete derivation of the kriging equations is provided by Olea (1975) for punctual kriging, optimum values for the weights are found by solving a set of simultaneous equations which include values from a semi-variogram of the variable being estimated. The weights are evaluated such that the resulting estimates are unbiased and have minimum estimated variance.

For n data points we have (n+1) equations with only n unknowns. Since we have more equations than unknowns, the additional degree of freedom may be used to ensure the solution has the minimum possible estimation error. This is accomplished by including a Lagrange multiplier, λ. The complete set of equations have the following form in matrix notation:-

$$\begin{bmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \gamma(h_{13}) & \cdots & \gamma(h_{1n}) & 1 \\ \gamma(h_{21}) & \gamma(h_{22}) & \gamma(h_{23}) & \cdots & \gamma(h_{2n}) & 1 \\ \vdots & \vdots & \ddots & \cdots & \vdots & 1 \\ \gamma(h_{n1}) & \gamma(h_{n2}) & \gamma(h_{n3}) & & \gamma(h_{nn}) & \vdots \\ 1 & 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ \lambda \end{bmatrix} = \begin{bmatrix} \gamma(h_{1p}) \\ \gamma(h_{2p}) \\ \vdots \\ \gamma(h_{np}) \\ 1 \end{bmatrix} \qquad (1)$$

In general we solve the equation :-

$$[A][W]=[B]$$

for a vector of unknown coefficients [W]. The terms in matrix [A] and vector [B] are taken directly from the semi-variogram or the mathematical function describing it. Once the unknown weights have been determined, the variable at location P is given by:-

$$\hat{X}_p = \sum_{i=1}^{n} w_i x_i$$

For block kriging the same form of equations as given in (1) are valid, but the γ values are replaced by $\overline{\gamma}$. $\overline{\gamma}$ is the semi-variogram which expresses the variance between the observation and the block to be estimated.

Where the distribution of the variable of interest is not even approximately normal but strongly skewed, it is possible to fit a lognormal distribution. Krige (1978) considers this method of kriging. The data are first transformed using an appropriate transformation :-

$$y_i = \log(x_i + a)$$

where a is an arbitrary constant to optimise the fit to a normal distribution. These values are then used to compute the semi-variogram and generate the ordinary kriging estimates. Link and Koch (1975) have shown various problems are associated with this type of kriging.

Theoretically, the best possible estimator of x_p is some function of the data $x_1, x_2, \ldots, x_n$ and is defined to be the conditional expectation of x_p given the n observations. In order to obtain this estimator we need to know the precise distribution of the (n+1) variables. For the simplest case of a normal distribution and stationarity, the conditional expectation can be defined and is identical with the best linear estimator.

In disjunctive kriging, a 'best-fit' appropriate transformation is used, consisting of a set of Hermite polynomial functions, Journal and Huijbregts (1978). By using such transformations, it is possible to convert the data into a form which will approximate to a univariate normal distribution for the x_i values and bivariate normal distribution for every pair of values x_i, x_j. The class of possible functions is restricted to those which are linear combinations of univariate functions of the data. The estimator obtained is non-linear and many of the desirable properties of simple linear kriging are discarded e.g. unbiasedness.

A study of the sea-floor by Journal (1969) provoked Matheron's interest in the problems of non-stationarity and the development of universal kriging. The approach parallels that of trend surfaces where the phenomena is split into two components, a deterministic trend (or drift) plus a random error (or fluctuation), section 2.4. The difference between the two approaches is that the fluctuations used in geostatistics are not assumed to be independent, as in trend surfaces. Instead of assuming the drift is a constant, it is allowed to follow a trend which may be expressed in terms of a simple polynomial.

Although the idea is a step away from the rarely adhered to requirement of stationarity, major problems exist with this technique, Armstrong (1984). These are primarily associated with obtaining a valid estimation of the semi-variogram and secondly the indeterminacy of the drift.

Finally generalised covariances rely on the theoretical fact that lowest order differences are the strongest forms of stationarity. These differences are termed generalised

increments. The concept of generalised increments leads directly to that of generalised covariances, using this notion the kriging equations may be rewritten in terms of generalised covariances. The result of this is to relax the stationarity requirements but the cost is measured in terms of the complexity of the theory and the resultant computations.

Generally the ideas of geostatistics have advantages over other more conventional techniques for analysing spatial data. They have a sound theoretical basis, they allow some estimate of the quality of estimation procedures and finally they have some claim to statistical properties such as unbiasedness, linearity and minimum variance. However they require some fairly stringent assumptions to be made which in practice can rarely be adhered to.

Regionalised variable theory is not valid under conditions where its defined form of stationarity does not hold. However, although unproved, most users of the method suggest that departures from stationarity are not of practical significance since local stationarity may be assumed.

A second assumption is that of normality. The majority of data sets are not strictly normal, transforming them to achieve normality and then applying the kriging equations results in the estimate being sub-optimal, non-linear and biased.

Finally the major problem lies with the corner-stone of parametric geostatistics, the semi-variogram. Whatever data set is used the semi-variogram will often depart significantly from all the theoretical models, skilful interpretation is required to fit one or more models to the empirical curve.

§2.4 Trend Surface Analysis

Trend surface analysis was first drawn to the attention of earth scientists in the mid 1950's by Oldham and Sutherland (1955), Miller (1956), Krumbein (1956), Grant (1957) and Whitten (1957). These authors used the techniques for the analysis of gravity maps, stratigraphic maps, isopach maps and maps representing specific attributes of sedimentary and igneous rocks, respectively.

In the course of time, the number of applications has increased significantly and the method has been refined and generalised. More recent studies in which trend surface

analysis was used as a primary tool for arriving at conclusions include Anderson (1970), Bradley (1970) and Haining (1987).

A trend is any large scale systematic change that extends smoothly and predictably from one map-edge to the other. Examples of such systematic trends might be the dome of atmospheric pollution over a city, the steady rise in cirque floor levels noted across many mountain ranges and so on. The detection and separation of such trends is the object of trend surface analysis.

Once again thinking of a surface mapping as a scalar field:-

$$z_i = f(x_i, y_i)$$

A trend surface simply specifies a precise mathematical form for the function, f, and fits it to the observed data by least squares regression.

In practice no simple function will fit the observed data. Firstly, even when the underlying surface is simple, measurement errors will have been introduced into the observed data. Secondly, in practice it is exceedingly unlikely that only one trend-producing process will be in operation. It follows that as well as trend represented by a simple mathematical function of the x and y co-ordinates, there will also be local departure from this i.e. residuals:-

$$z_i = f(x_i, y_i) + \varepsilon_i$$

The main problem in trend surface analysis is to decide upon a particular function for the trend part of the function. Although there is an enormous range of possible functions, the simplest trend surface is an inclined plane i.e. a linear trend surface:-

$$z_i = \beta_{00} + \beta_{10} x_i + \beta_{01} y_i + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2) \quad (2)$$

For the general case of a polynomial of degree m:-

$$z_i = \sum_{j=0}^{p} \sum_{k=0}^{q} \beta_{jk} x_i^j y_i^k + \varepsilon_i \quad (3)$$

In specific applications, $p+q \leq m$ where m denotes the degree of the trend surface. A trend surface of degree m has $r = \frac{1}{2}(m+1)(m+2)$ coefficients β_{jk} These are calculable only if the number of observations satisfies the condition $n \geq m$. Where n=m a special type of surface is fitted with no residuals.

Returning to the simplest case of a linear trend surface, (2), the constants have a simple physical interpretation, the first β_{00} represents the height of the plane surface at the map origin where $(x_i, y_i) = (0,0)$. The second β_{10} is the surface slope in the x-direction and β_{01} gives its slope in the y-direction. For interpretation of quadratic and cubic coefficients see Cliff et al. (1975).

The theory of least squares estimation is used to fit a surface to the data i.e. the sum of the squares of the residuals at all data points is minimised for all possible surfaces of that degree, determining uniquely the coefficients, β_{jk}, in equation (3) i.e.

$$z(\underline{u}) = f(\underline{u})^T \beta + \varepsilon(\underline{u})$$

where $$S = f(\underline{u}) = \begin{bmatrix} 1 & x_1 & y_1 & x_1^2 & y_1^2 & \cdots & x_1^p y_1^q \\ 1 & x_2 & y_2 & x_2^2 & y_2^2 & \cdots & x_2^p y_2^q \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & y_n & x_n^2 & y_n^2 & \cdots & x_n^p y_n^q \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_{00} \\ \beta_{10} \\ \beta_{01} \\ \vdots \\ \beta_{pq} \end{bmatrix}$$

Thus $\underline{\beta}$ satisfies:-

$$\underline{\beta} = (\underline{S}^T \underline{S})^{-1} \underline{S}^T \underline{Z}. \qquad (4)$$

In general equation (4) is valid, however $(\underline{S}^T \underline{S})$ may in practice be singular hence its inverse does not exist. Various methods have been developed to circumvent this particular problem e.g. Whitten (1970) developed a method using orthogonal polynomials for irregularly spaced data.

Various degrees of polynomial for a trend surface have been cited in the literature e.g. Krumbein (1959) illustrated the use of both a linear and quadratic surface to describe the trend in the lithologic composition (classic ratio) of Pennsylvanian rocks in parts of Kansas and Oklahoma. Whitten (1970) employed octic surfaces in his description of the sub-surface elevation on top of the Devonian Dundee Limestone in a 900 square-mile area of Central Michigan. Coons et al. (1967) examined the Mid-Continent gravity high to assess the relation between basement and Paleoxoic structures using surfaces of up to a complex 15^{th} order polynomial.

The method of trend surfaces works quite well with surfaces that are of a simple form but it is very sensitive to data point distribution and aberrant values. Gray (1972) and Robinson (1972) illustrate the validity of such a statement, whilst Davies (1973) demonstrates the pattern of the data is unimportant as long as the area of interest is covered, the clustering of points within an area does not influence the final surface. A further problem is that the number of data points should be restricted in order that rather extensive interpolations are based on relatively few data.

A further cautionary note to the usage of this technique is that care should be taken in not relying too heavily on the statistical significance tests which have been developed to decide the degree of a trend surface (analysis of variance), or to check residuals for 'outliers' (t-tests), or on confidence belts which can be calculated for the trend surfaces. These statistical tests produce exact results only if the residuals are stochastically independent. In reality the residuals from a trend surface are frequently subject to significant spatial auto-correlations. In that situation, the test statistics based on the model of uncorrelated residuals may be severely biased, Agterberg (1964), Watson (1971).

Various other techniques have been advocated to eliminate this problem. Instead of postulating a trend-residual model, a trend-signal-noise model may be defined. The trend signal denotes a continuous random function with noise describing a stochastically independent component. Haining (1987) considers alternative ways of fitting trend-surface models which have been modified to include a random process model to accommodate spatial autocorrelation. Haining considered two cases, firstly where the order of the trend was assumed known and secondly when the order is unknown. For the former the method of parameter estimation used was the Cochrane-Orcutt time series procedure, Ripley (1981). In case of the order of trend being unknown, a first order

model is estimated and the residuals are tested for autocorrelation using the generalised Moran coefficient (GMC), Ripley (1981) for autocorrelation. If autocorrelation is present, an iterative scheme is performed to estimate the model parameters, otherwise they may be estimated using ordinary least squares. This procedure is repeated for a second order model. If none of the second order coefficients are significant or, the increase in R^2 is not significant, the first-order model is retained; otherwise the third-order model is estimated and the procedure repeated.

Trend surface analysis is probably most useful when trend and residuals can be linked to separate spatial processes which were operative on a regional and local scale.

§2.5 Density Estimation

Density estimation is possibly the most important topic in applied statistics since unless the density $f(\underline{x})$ is known its characteristics must be inferred from a sample $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$. Two possible approaches to density estimation are parametric and non-parametric. The former assumes the data are drawn from one of the known parametric family of distributions. The density, f, underlying the data is then estimated by finding estimates of the unknown parameters from the data and substituting these estimates into the formula for the function, f.

The second approach, non-parametric, is less rigid in its assumptions concerning the distribution of the observed data. The data determines an estimate of the probability density, f. Density estimation was first proposed by Fix and Hodges (1951) as a method for relaxing the rigid distributional assumptions inherent to discriminant analysis.

For the remainder of the section, attention will focus on the second form of density estimators, in particular those applicable to multivariate data since the ensuing work is spatial in nature. Additional problems arise when estimating a density function in a multi-dimensional setting. These primarily relate to regions of the sample space which are devoid of observations, this is referred to as the 'empty space phenomena' by Scott and Thompson (1983).

A variety of non-parametric probability density estimators have been proposed and studied since the pioneering work on kernel methods by Rosenblatt (1956) and Parzen

(1962). The types of estimators fall into four categories kernel, histogram-type, orthogonal series and splines, some overlap does exist. Only the first two forms of estimators will be discussed, Fryer (1977) reviews all four methods, not only with reference to their theoretical properties but also to their applicability to real life problems. Most density estimation research has dealt with the one-dimensional case and applications to relatively small data sets. Multivariate extensions of histograms, kernel and nearest neighbour estimators have been studied theoretically and are usually applied to bivariate data sets.

§2.5.1 Histogram based estimators

The histogram is extremely efficient computationally compared to kernel methods, but statistically it is quite inefficient. Relative sample sizes required for histograms to have errors comparable to kernel estimators increases rapidly with increasing sample size and dimension. Scott (1985a) studied the frequency polygon which is formed by the linear interpolation of adjacent mid-bin values of a histogram, statistically it has the same order of efficiency as the kernel estimator but still has the computational efficiency of the histogram. Problems exist with such an approach, the achievement of the same order of efficiency requires greater sampling density and secondly, the bins of the optimal frequency polygon are wider than those for the optimal histogram. With large bins, bin edge effects become more pronounced.

Scott (1985b) proposed two other variants, the average shifted histogram and a frequency polygon of the average shifted histogram. With the latter it can be shown to be functionally identical to a related interpolated kernel estimator of binned data.

§2.5.2 Kernel estimators

Moving away from the idea of mapping based on binned data and its related problems to the kernel method of density estimation. The definition of a kernel estimator as a sum of 'bumps' centred at the observation is easily generalised to the multivariate case. Rosenblatt's results for the naive estimator were extended to the bivariate case by Maniya (1961), with Cacoullos (1964,1966) obtaining the p-dimensional multivariate equivalent of Parzen's work, particularly with reference to the evaluation of optimal mean square error and other invariance properties.

§2.5.3 Fixed kernel estimators

The simplest form of the multivariate kernel density estimator is the fixed kernel. Let $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ be an independent, identically distributed random sample from a multivariate distribution with density f. The kernel density estimate of f based on these data is:-

$$\hat{f}(\underline{x}) = n^{-1}h^{-d}\sum_{i=1}^{n} K\left\{h^{-1}(\underline{x}-\underline{x}_i)\right\}$$

where $K\{\ \}$ is the kernel function defined for d-dimensional $\underline{x}$ satisfying :-

$$\int_{R^d} K(\underline{x})d\underline{x} = 1$$

and h is the smoothing parameter, assumed to tend to zero as n tends to infinity.

For mathematical reasons, the kernel selected is a radially symmetric, unimodal probability function. Within this family of functions, it has been demonstrated that in terms of efficiency, the choice of kernel is not critical. The selection should be conditioned by the intended use to which the methodology is being applied. Various examples of kernel functions are:-

1. Epanechnikov Kernel

$$K_e(\underline{x}) = \begin{cases} \frac{1}{2}c_d^{-1}(d+2)(1-\underline{x}^T\underline{x}) & \underline{x}^T\underline{x} < 1 \\ 0 & \text{otherwise} \end{cases}$$

where c_d = volume of the unit d-dimensional sphere

2. Standard multivariate normal density function

$$K(\underline{x}) = (2\pi)^{-d/2}\exp\left[-\tfrac{1}{2}\underline{x}^T\underline{x}\right]$$

3. $$K_2(\underline{x}) = \begin{cases} 3\pi^{-1}(1-\underline{x}^T\underline{x})^2 & \underline{x}^T\underline{x} < 1 \\ 0 & \text{otherwise} \end{cases}$$

4. $$K_3(\underline{x}) = \begin{cases} 4\pi^{-1}\left(1-\underline{x}^T\underline{x}\right)^3 & \underline{x}^T\underline{x} < 1 \\ 0 & \text{otherwise} \end{cases}$$

The last two apply solely to the two-dimensional situation.

§2.5.3.1 Smoothing parameter selection

The smoothing parameter, or alternatively the window width, is important for its role in determining the resultant shape of the density estimate.

Figure 2.2 displays pictorially the consequences of varying the smoothing parameter. The estimates are based on a set of fifty data points simulated from a bivariate normal distribution using the NAG (1984) random number generator subroutine, G05EAF.

Figure 2.2(a) displays the optimal surface as determined by one of the automatic procedures, least squares cross-validation, described later. Selecting too small a value for the smoothing parameter results in a more spiked representation of the surface 2.2(b), by selecting a larger value of smoothing parameter than deemed 'optimal' by one of the automatic procedures, results in the density estimate being oversmoothed and hence causing any detail to be obliterated, figure 2.2(c).

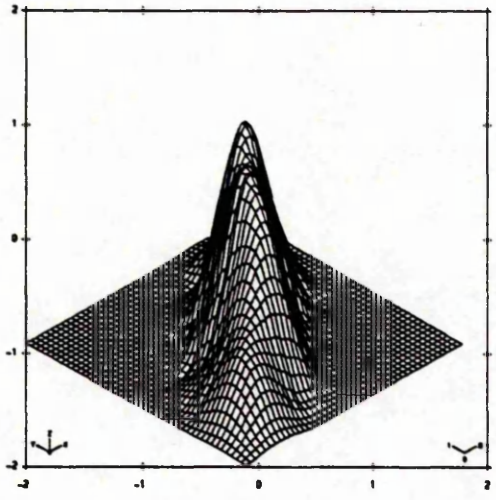


Figure 2.2 (a) :- Optimal procedure for selection of h.

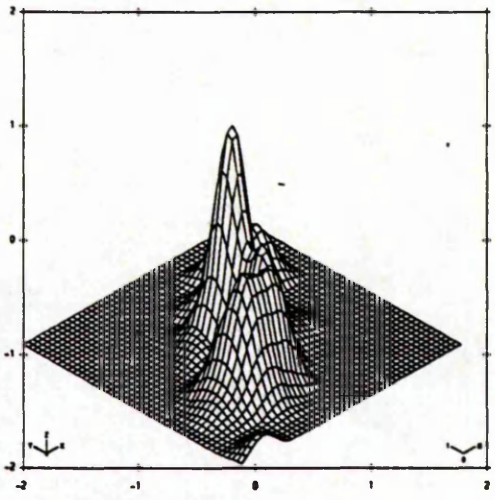


Figure 2.2 (b) :- Under-estimation of h.

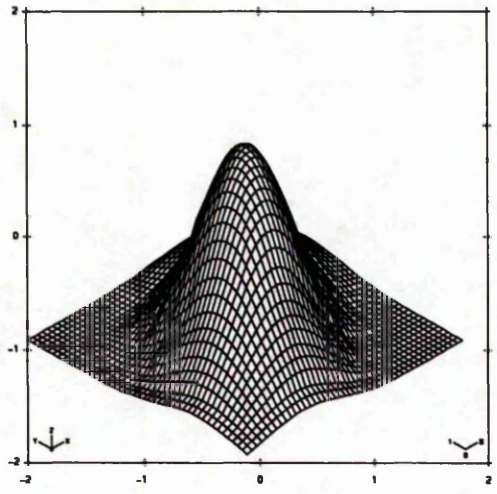


Figure 2.2 (c) :- Over-estimation of h.

Figure 2.2 :- Surfaces produced for varying levels of smoothing parameter.

A variety of techniques have been developed to overcome the subjectivity of selecting h. Where the kernel is purely for illustrative purposes, selecting the window width in this manner is fairly standard. Five techniques for selecting h automatically or semi-automatically are suggested in the literature:-

1. Likelihood cross-validation
2. Least squares cross-validation
3. Test-graph method
4. Bootstrap method
5. Selecting h assuming the data arises from a standard distribution

The first two methods use the concept of cross-validation. The idea of cross-validation was developed by Mosteller and Wallace (1963) for use in discriminant analysis. In the context of smoothing parameter selection, the cross-validation criterion is the sub-division of a sample size, n, into a 'construction' subsample, size (n-1), and a validation subsample, size 1, in all n possible ways. Stone (1974) provides a detailed description of cross-validation in various areas of application.

§2.5.3.1.1 Likelihood cross-validation

This method as its name suggests uses the ideas of likelihood to assess the adequacy of fit for different smoothing parameters.

A score function CV(h) is evaluated based on the log-likelihood averaged over each choice of omitted $\underline{x}_i$. The value of smoothing parameter for which the function is optimised is selected as the 'optimal' value:-

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n} \log \hat{f}_{-i}(\underline{x}_i)$$

Scott and Factor (1981) adjudged the behaviour of CV(h) to be unduly sensitive to outliers and as a result tended to oversmooth the function. The problem is potentially more worrisome with multivariate data, due to the difficulties of detecting outliers.

One of the requirements of a statistical procedure is that it should be consistent i.e. good estimates of the quantity of interest should be obtained if very large samples are available. If one of the tails is monotonic and dies off at an exponential rate, the use of likelihood cross-validation leads to inconsistent estimates of the density, Schuster and Gregory (1981).

Although simple to implement and computationally not excessive in terms of time, the method is not ideally suited for use in real data sets

§2.5.3.1.2 Least squares cross-validation

The idea suggested by Rudemo (1982) and Bowman (1984) is based on minimising a suitably well behaved loss function, the integrated squared error:-

$$I.S.E. = \int_{R^d} \left\{\hat{f} - f\right\}^2 d\underline{x} \quad = \quad \int_{R^d} \hat{f}^2 - 2\int f\hat{f} + \int f^2$$

$$R\left(\hat{f}\right) = \int_{R^d} \hat{f}^2 - 2\int \hat{f} f$$

The basic principle of least-squares cross-validation is to construct an estimate of $R\left(\hat{f}\right)$ from the data themselves and minimise this estimator over h to give the choice of window width. Silverman (1986) demonstrates how a score function may be evaluated from the data to allow this value of h to be calculated. Defining

$$\hat{f}_{-i}(\underline{x}) = (n-1)^{-1}h^{-d}\sum_{\substack{j=1 \\ i\neq j}}^{n} K\left\{h^{-1}(\underline{x}-\underline{x}_j)\right\}$$

$$\text{Let} \quad M_0(h) = \int \hat{f}^2 - 2n^{-1}\sum_{i=1}^{n}\hat{f}_{-i}(\underline{x}_i) \qquad (4)$$

M_0 depends on the data. The idea of least squares cross-validation is to minimise the score M_0 over h. Equation 4 is not in a form suitable for computation. Letting $K^{(2)}$ be the convolution of the kernel with itself. It may be shown that:-

$$\int \hat{f}(\underline{x})^2 d\underline{x} \;=\; n^{-2}h^{-d}\sum_i\sum_j K^{(2)}\left\{h^{-1}(\underline{x}_i-\underline{x}_j)\right\}$$

and

$$n^{-1}\sum_i \hat{f}_{-i}(\underline{x}_i) = n^{-1}(n-1)^{-1}\sum_i\sum_j h^{-d}\hat{K}\{h^{-1}(\underline{x}_i-\underline{x}_j\} - (n-1)^{-1}h^{-d}K(\underline{0})$$

The resultant score function is :-

$$M_1(h) = n^{-2}h^{-d}\sum_i\sum_j K^*\left\{h^{-1}(\underline{x}_i-\underline{x}_j)\right\} + 2n^{-1}h^{-d}K(\underline{0})$$

where $K^*(\underline{t}) = K^2(\underline{t}) - 2K(\underline{t})$

$K^2(\underline{t})$ is the convolution of the kernel with itself.

Stone (1984) provides a strong large sample justification of least squares cross-validation. Stone's theorem states that the score function $M_1(h)$ says asymptotically, just

as much about the optimal smoothing parameter, from the integrated squared error point of view, as if we actually knew the underlying density, f.

§2.5.3.1.3 Test Graph Method

The third of the methods may be described as semi-automatic. For a set of independent, identically distributed observations from a d-dimensional density, f, $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$, the estimator function is given by :-

$$f_n(\underline{x}) = \sum_{j=1}^{n} n^{-1} h^{-d} K\left\{h^{-1}(\underline{x} - \underline{x}_j)\right\}$$

The test graph method is then:-

$$\nabla^2 f_n(\underline{x}) = \sum_{j=1}^{n} n^{-1} h^{-d-2} \nabla^2 K\left\{h^{-1}(\underline{x} - \underline{x}_j)\right\}$$

The window width is chosen to give the best estimate of the density such that the random fluctuations, in the second derivative, ∇^2, of the estimate, will be asymptotically of maximum size $\pm\delta \sup|f''|$, where δ is a calculable constant and the value of $\sup|\nabla^2 f|$ can be estimated from the test graph.

The subjectivity of the method arises since test graphs are drawn of the second derivative of the estimator for various band widths, to assess which gives the 'right size' of fluctuations. The problem with this technique is the difficulty in assessing the graphs. In two-dimensions, contour plots may be used, but in higher dimensions difficulties of presentation are encountered. The second problem is the time required to produce each graph. Finally to achieve reasonably accurate results, a fairly sizeable data set is recommended.

Despite these problems, Silverman (1978) demonstrated the acceptability of the results in general. In conclusion it is felt that this method is most useful as an aid to checking a pre-selected window width.

§2.5.3.1.4 Bootstrap method

The selection of bandwidths for density estimation based on the bootstrap, once again involves the idea of minimising the integrated squared error over various values of the smoothing parameter, h, Faraway and Jhun (1990).

The integrated mean squared error may be decomposed into a bias and variance component. Direct application of the bootstrap fails since it is incapable of estimating the bias term in this context. To overcome this problem, an initial estimate of the density, with the bandwidth chosen by one of the other methods, is calculated and resampling is subsequently done from this. The bias term is then calculable.

A by-product of this technique is the construction of a confidence band for the smoothing parameter.

Overall this method is shown to perform comparatively better than cross-validation in terms of the integrated mean squared error. The bandwidths are generally larger than for cross-validation resulting in smaller variation. The negative factor of this approach is the time taken to evaluate the smoothing parameter.

§2.5.3.1.5 Estimating 'h' from a standard distribution

The final method for selecting bandwidths to be discussed is to assume the underlying density is standard e.g. multivariate normal. Based on the ideas of bias and variance the window width can be shown to be :-

$$h_{opt} = A(K)n^{-1/(d+4)} \qquad (5)$$

where $A(K) = \left[\delta\beta\alpha^{-2}\left\{\int\left(\nabla^2\theta\right)^2\right\}^{-1}\right]^{1/(d+4)}$ depends on the kernel and θ is the d-variate normal density. For specific kernels A(K) has been calculated, Silverman (1986).

The window width may then be evaluated directly from (5) if the Fukunaga estimate (1972) is used :-

$$\hat{f}(\underline{x}) = \frac{(\det S)^{-1/2}}{nh^d} \sum_{i=1}^{n} K\left\{h^{-2}(\underline{x}-\underline{x}_i)^T S^{-1}(\underline{x}-\underline{x}_i)\right\}$$

where $K(\underline{x}^T\underline{x}) = K(\underline{x})$

S = covariance matrix

If the kernel is radially symmetric and the data untransformed, a single scale parameter, σ, for the data is evaluated from:-

$$\sigma^2 = d^{-1}\sum_i s_{ii}$$

and the value σh_{opt} is the resultant window width.

This method gives a 'quick and dirty' estimate for the smoothing parameter simplifying the task of finding the optimal value for the other techniques.

§2.5.3.1.6 Summary

The importance of selecting the smoothing parameter was summarised in figure 2.2. However a whole wealth of techniques are available for selecting h. Biasing the conclusions towards the requirements of the proceeding work, eliminates three of the methods, firstly the test graph method is not fully automated, secondly estimating the smoothing parameter from a standard distribution requires the assumption that the data arose from a standard distribution, the use of a non-parametric mode of analysis was selected to avoid such a restriction. Finally the use of likelihood cross-validation is restricted by its sensitivity to outliers, which is especially problematical for multivariate data.

Finally, of the remaining two methods, the bootstrap technique requires the smoothing parameter to be selected initially by some existing technique. The overall reduction in integrated mean squared error over that of the least squares cross-validation method does not necessarily warrant the additional increase in computational time. In light of this, least squares cross-validation was used for smoothing parameter selection.

§2.5.4 Alternative Kernel Estimators

Although the kernel estimator can be shown to exhibit a series of desirable properties, in terms of practicalities the main drawback of the method is its inability to deal satisfactorily with the tails of the distribution, without oversmoothing the main part. Two possible adaptive approaches are the nearest neighbour and adaptive kernel methods.

§2.5.4.1 Nearest neighbour method

This approach described by Loftsgaarden and Quesenberry (1965) has also been termed the balloon density by Tukey and Tukey (1981). The estimator in d-dimensions is defined in the following manner:-

$$\hat{f}(\underline{t}) = \frac{k/n}{v_k(\underline{t})} = \frac{k/n}{c_d r_k(\underline{t})^d}$$

where $r_k(\underline{t})$ - Euclidean distance from $\underline{t}$ to the k^{th} nearest data point

$v_k(\underline{t})$ - d-dimensional volume of the d-dimensional sphere of radius $r_k(\underline{t})$

c_d - volume of the unit sphere in d-dimensions

For a sample size, n, one expects $nf(\underline{t})v_k(\underline{t})$ observations to lie within the sphere radius $r_k(\underline{t})$, centre $\underline{t}$.

The overall estimates obtained by the nearest neighbour approach are not satisfactory for a number of reasons:-

1. They are prone to local noise.
2. The tails of the estimator are unduly heavy.
3. It is discontinuous and its integral over all space is infinite.

As a technique it is not the most suitable method due to the above limitations. The adaptive kernel method overcomes these difficulties while still being adaptive to the local density.

§2.5.4.2 Adaptive kernel estimators

This method offers a combination of the desirable smoothness properties of the Parzen-type estimators with the data-adaptive characteristics of the k-nearest neighbour approach.

A kernel is placed at each of the observed data points but the window width is allowed to vary according to the underlying density. Intuitively the smoothing parameter $h(\underline{x}_i)$ should vary inversely with the true density, f, i.e. in regions of low density a broader kernel should be used and vice-versa.

Theoretical work of Abramson (1982) showed that $h(\underline{x}_i) \propto f^{-1/2}(\underline{x}_i)$ is a good choice. The density $f^{-1/2}(\underline{x}_i)$ is unknown but may be replaced with an appropriate estimate. This estimate makes use of one of the alternative methods, the choice of which is not crucial, since the adaptive method is insensitive to the fine detail of the pilot study.

Breiman et al. (1977) describe the technique and compare it to the Parzen estimator. Overall, they found it was superior to the best Parzen estimate.

Silverman (1986) feels the technique has advantages over the ordinary fixed kernel in some situations, but each case should be treated separately.

§2.6 Summary

Within this chapter four methods of surface fitting have been discussed. Each of the techniques requires some unknown parameter to be defined:-

1. Moving average - weighting method.
2. Trend surface - order of polynomial.
3. Kriging - form of semi-variogram.

4. Kernel density estimation - smoothing parameter, h.

The unknown parameters are all basically controls on the smoothness of the resultant surface. Of the four techniques, only the fixed kernel may be defined to be fully automatic. Currently no automatic selection procedure is available for selecting the 'optimal' order of polynomial or weighting method. In terms of kriging, the model for the semi-variogram requires considerable skill to fit it to the empirical data.

Potential drawbacks of the fixed kernel as mentioned previously relates to their inability to deal with long tailed distributions. However this apart, it was decided that since least human intervention was required it was the most suitable for fitting a surface. However any of the other methods are feasible and the ensuing work is wholly applicable using the other techniques.

CHAPTER 3

SURFACE REPRESENTATION

§3.1 Contour Selection

Before proceeding to investigate the various aspects of contour selection and fitting, it is necessary to describe the two types of question which may arise when comparing surfaces since they condition the form of contour selection. The most common form of questions arising are:-

1. Are the underlying statistical distributions of the surfaces similar?

2. For a specific level of contour, does the spatial pattern differ between surfaces?

A subtle difference exists between the selection procedures for the two questions, the latter does not require a formal selection method since the levels will be the same for each surface. In terms of the former question, the two surfaces may not necessarily describe the same variable, therefore a procedure which ensures contour selection is consistent is essential to avoid reaching spurious conclusions. Motivation for a quasi-automatic technique which allows the spacing and the number of class intervals to be selected was prompted by the first of these questions. Later within the chapter the question of contour accuracy is examined with its role in influencing various geometric properties of a contour being discussed.

§3.2 Contour Notation

Before proceeding to examine specific topics relating to the description of a surface in terms of a set of contours, various definitions and notation are introduced.

A closed contour is defined to be a closed plane figure bounded by three or more straight line segments that terminate in pairs at the same number of vertices, and do not intersect other than at their vertices. The sum of the interior angles is $(n-2)*180^0$, where n is the number of sides; the sum of the exterior angles is always 360^0. The polygon is

approximated by a set of N points, the first is arbitrary but the sequence is delineated in an anticlockwise direction, figure 3.1 (a).

An open contour, is a contour which has two or more of its ordinates defined to be end-points on the boundary of the region. For analysis, if the region is defined a-priori, an open contour is considered equivalent to a closed contour, figure 3.1 (b), although the conditions of a closed contour are not necessarily satisfied.

A disjoint contour is one which comprises more than one element, these may be either open or closed, figure 3.1 (c).

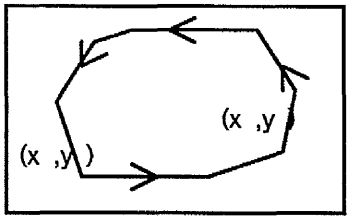


Figure 3.1 (a) Closed contour.

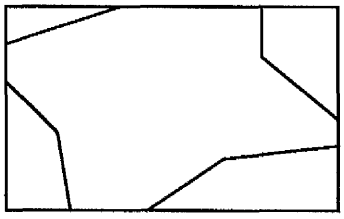


Figure 3.1 (b) open contour

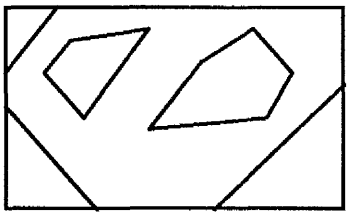


Figure 3.1 (c) Disjoint contour

Figure 3.1 :- Definition of various contour forms.

A summary of the contour notation used subsequently is given below :-

C_{ij} contour j, in surface i, $i = 1, N_S$, $j = 1, N_C$

N_S total number of surfaces

N_C total number of contours

N total number of point describing contour j.

where $C_{ij} = (x_{ij1}, y_{ij1}), (x_{ij2}, y_{ij2}), \ldots, (x_{ijN}, y_{ijN})$

x_{ijk} k^{th} x-ordinate of contour j, in surface i

y_{ijk} k^{th} y-ordinate of contour j, in surface i

§3.3 Number of Class Intervals

There is no definitive answer to the number of contours which might be used for constructing a map. A finely contoured map based on a few control points gives an impression of accuracy unwarranted by the primary information, conversely it is wasteful to use very few contours when in practice much is known about the inflexion points in the surface under investigation.

Dobson (1973) stated that in absolute terms the human eye can confidently identify only a small number of shadings or symbols i.e. between four and ten. This view was supported by Monmonier (1973), Robinson and Sale (1969) and Jenks and Caspell (1971).

One guide as to the choice of the number of contours was given by Brooks and Carruthers (1953) who advocated the number of classes in a histogram should satisfy:-

$$\text{number of classes} \leq 5*\log_{10}(\text{number of observations})$$

In terms of a contour map based on n observations, the surface should not consist of more than $(5\log_{10}(n)-1)$ contours, table 3.1.

Number of observations	50	100	150	200	500
Maximum number of contours	7	9	9	10	12

Table 3.1 :- Brooks and Carruthers (1953) contour selection procedure.

§3.4 Selection of Class Intervals

A second question pertaining to the display of spatial data in terms of isolines concerns the selection of class intervals. Numerous class interval selection methods have been proposed, Evans (1976) categorised these on the basis of four groups:-

1. Exogenous
2. Arbitrary
3. Idiographic
4. Serial

The first of these groups, exogenous, selects the intervals on some meaningful baseline level which relates to, but is not derived from the data to be mapped e.g. a critical population density threshold. As its name suggests, those methods categorised under the arbitrary heading are founded on a series of numbers of no particular significance. Usage of this method is indefensible and should never be implemented.

§3.4.1 Idiographic group

More specific to the details of the data set to be mapped, the idiographic group may be divided into three sub-groups:-

§3.4.1.1 Natural break methods

Two groups fall under this heading multi-modal and multi-step. Intervals for the former are based on 'natural breaks' in regions where the frequency is low. Jenks and Coulson (1963) found that clear natural break classes do not occur and subjective judgements vary greatly between people. It is usually possible to find apparent breaks but these are often the result of small sample size and their significance should not be exaggerated.

In a similar way the multi-step methodology divides the cumulative distribution into a series of 'treads' and 'rises'. The same set of problems exist as for multi-modal plus double the number of contours are required.

Neither of these methods should be utilised in quantitative mapping unless significant multimodality has been demonstrated statistically.

§.3.4.1.2 Percentiles

The percentile system of subdividing the frequency distribution is very useful in ensuring equal representation of classes. It is invariant to scalar, rotational and translational changes. The main disadvantage is that class intervals vary irregularly in

different parts of the measurement scale and between maps of the same variable, for different areas or times. Furthermore a percentile based map provides no information on the underlying frequency distribution. It is therefore incumbent that a histogram is positioned alongside the map with the class boundaries accordingly marked. This facilitates interpretation in relation to the measurement scale but Schultz (1961) still advocates very careful reading of the scale. In terms of map comparisons it is not advisable unless the underlying statistical distribution is uniform.

§3.4.1.3 Nested means class limits

Two classes are formed about the mean with each class then being subdivided at its own mean, Scriptor (1970). This method balances the desirable properties of equal numbers per class against that of equal class widths. Class intervals are narrow in the modal parts of the frequency distribution and broad in the tails. Extreme values influence the positions of the means of various orders so that the less closely spaced the values are in a given magnitude range, the broader the class.

For a rectangular frequency distribution, nested means approximate the equal interval or percentile solutions. For a normal distribution, it will approximate a standard deviation based method and finally for a 'j' shaped distribution, a geometric progression, see section 3.4.2.2 and section 3.4.2.4, respectively.

Nested means provide a generally applicable, replicable yet highly inflexible interval system. They do not form a numerical series independent of the data and do not allow numbers of classes other than 2^m, where m is an integer.

§3.4.2 Serial class

The serial class encompasses all methods with limits which are mathematically derived and fixed in relation to some statistical descriptor e.g. mean, median, range or standard deviation. Once again there are a number of techniques specific to this category, many of which are variants of a theme.

§3.4.2.1 Normal percentiles

This method of standardised class intervals based on the normal probability distribution was developed by Armstrong (1969). For normal data this method is preferable to true percentiles in being uninfluenced by minor details of the frequency distribution. Evans et al. (1975) mention the desirability of transforming the data to achieve normality and then applying this method. The same set of problems arise with the implementation of this method as for true percentiles.

§3.4.2.2 Standard deviation

The class width is defined as a proportion, K_S, of the standard deviation. Class intervals are centred on the mean, which is a class mid-point if the number of classes is odd and a class boundary if the number is even; the highest and lowest classes are necessarily open-ended.

This method is particularly suited to data sets which display approximate normality or are fairly symmetric with a pronounced mode near the mean. The internal intervals are equal and for skewed unimodal distributions a transformation may be applied to achieve approximate normality

Use of a standard deviation basis does not imply a class interval of one standard deviation which is too coarse if more than four contours are employed. Selecting a proportion too large results in the tail classes being under-utilised, alternatively too small a value reverts the method to that of a percentile based solution.

§3.4.2.3 Equal arithmetic intervals

Contour maps constructed by dividing the range into the desired number of levels fall under the serial group heading. The contours are selected as for topographic maps, but this is rarely feasible for statistical maps due to the structure of the data set and the possible existence of outliers.

An extreme example is cited in Jenks and Coulson (1963). They studied an area in Central Kansas where the population ranged from 1.6 to 103.4 persons per square mile.

On a map with seven equal intervals only four classes were represented with over 90% of the map lying within a single band.

§3.4.2.4 Geometric progression

This method has been developed to deal with 'j' shaped distributions. The geometric progression is set to pivot about some measure of central tendency e.g. median. The median is then made the geometric mid-point of the class, if an odd number of classes is desired, or the boundary between two central classes if an even number are required.

The sequence of class widths are delineated as:-

$$a \quad ax \quad ax^2 \quad ax^3 \quad \ldots\ldots\ldots \quad ax^{(N-1)}$$

where a = size of first term
x = base of the geometric progression
N = number of classes

The major drawback of this technique is its severe limitations of use. The base, x, requires trial and error to determine its ideal value for each data set, hence data sets contoured using this method are not easily automated.

§3.4.3 Illustration of four contouring techniques

Figures 3.2 to 3.5 provide illustrative examples of four of the contouring methods described above. The surface, based on 3432 points, describes the distribution of potassium 40 in the south-west of England around the towns of Yeovil, Taunton and Illminster.

Table 3.2 summarises the contour levels for the four methods of contouring. The original data set, although not reproduced, is characterised by fairly long tails, taking logs of the data draws the tails in slightly.

Contouring technique	Contour Level 1	2	3	4	5	6	7
Equal range	32.85	65.70	98.55	131.4	164.2	197.1	229.9
Nested means	33.05	59.02	82.02	94.88	106.7	129.4	177.8
Percentile	39.96	63.33	85.67	95.40	103.1	111.0	146.2
Standard deviation	7.11	36.37	65.24	94.88	124.1	153.4	182.6

Table 3.2 :- Contour levels for the various methods.

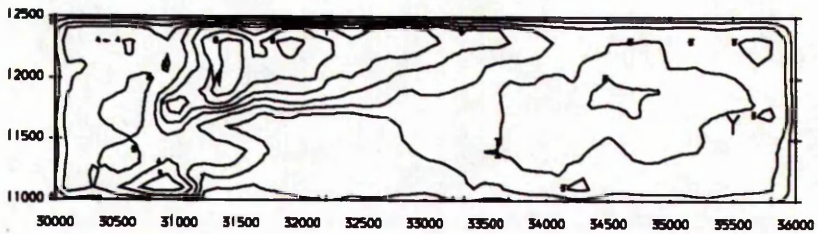


Figure 3.2 :- Surface contoured using the equal range technique.

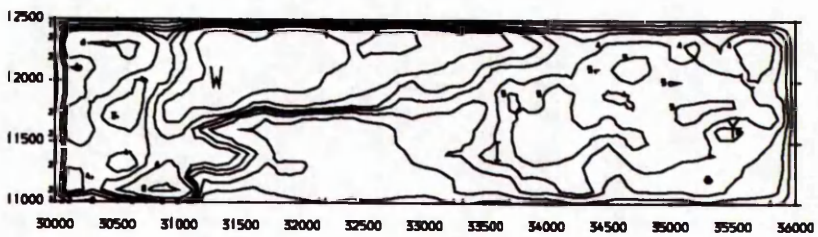


Figure 3.3 :- Surface contoured using the nested-means methodology.

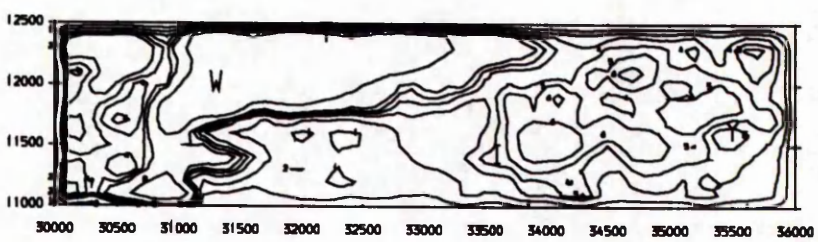


Figure 3.4 :- Surface contoured using the percentile methodology.

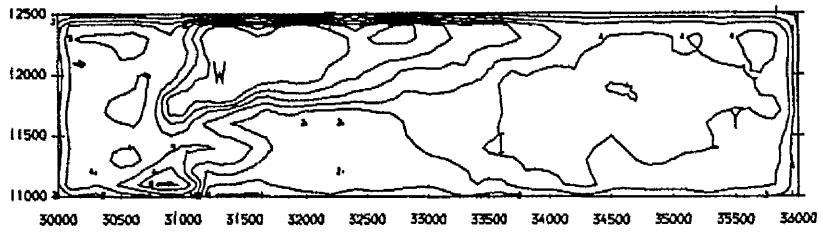


Figure 3.5 :- Surface contoured using the standard deviation technique.

The contour levels for percentile and nested means methods are similar, this was reflected in the resultant surfaces. The problem as mentioned previously with the nested means method is the restriction on the number of classes selected whilst for the percentile technique comparability between surfaces is not necessarily assured.

For this example, the equal range method and standard deviation techniques produce similar results. The major drawback of the former of these methods concerns the distortion of results if an outlier is present. The levels selected may then be severely biased towards one of the extremes. Furthermore the equal range method does not guarantee comparability of surfaces. The standard deviation method is a more robust method which works equally well for data which is originally near normal in distribution, or alternatively, which requires to be transformed to achieve normality. One drawback of the standard deviation and equal range methods is that for distributions with long tails, one or both of the extreme levels may not be represented in the final surface due to a lack of information in these regions. This problem does not affect the comparison of surfaces since comparability is achieved between the remaining levels.

This problem apart, it was decided that in view of the versatility of the standard deviation method and the knowledge that comparability is assured between surfaces of approximately normal data, contour selection should be based on this method.

A word of caution at this point is that many commercially available contouring routines use the equal range or percentile technique for automatic contouring. Comparisons using this form of contour production should be treated with caution.

§3.5 A Semi-Automatic Contour Selection Procedure

A simulation study was carried out to investigate whether it was possible to create an empirical rule which allowed the number of levels, and the standard deviation proportion to be selected on a consistent, and automatic basis, rather than from examining the resultant distribution for various permutations of the two variables, until a satisfactory solution was derived. The study was based on how well the spatial distribution of various data sets were reproduced for different numbers of contours and standard deviation proportions.

Initially a sequence of 50, 100 and 200 points were simulated from a standard normal distribution using the random number generator from within Minitab, release 7.1. The response of the standard normal distribution was examined for a range of standard deviation proportions, K_S, 0.3 - 1.4, and number of classes, N_C, 4 to 10. The response of each data set was gauged in terms of a series of histograms, figures 3.6 - 3.12, constructed from the percentage number of points contained within each band for the various proportions, and how well the known underlying distribution was reproduced. If the resultant histogram satisfied the following criteria the combination of the two factors was deemed satisfactory to represent the spatial distribution of the data:-

1. No excessive pooling in the open-ended classes.

2. No under-utilisation of the open-ended classes.

3. The resultant histogram mirrored the underlying distribution of the data i.e. normal.

From these diagrams, an ad-hoc selection rule, similar to that developed by Brooks and Carruthers (1953) was formulated to initially select the desired number of contours for the size of the data set. A second rule was constructed to select an appropriate value for the standard deviation procedure based on the number of contour levels selected.

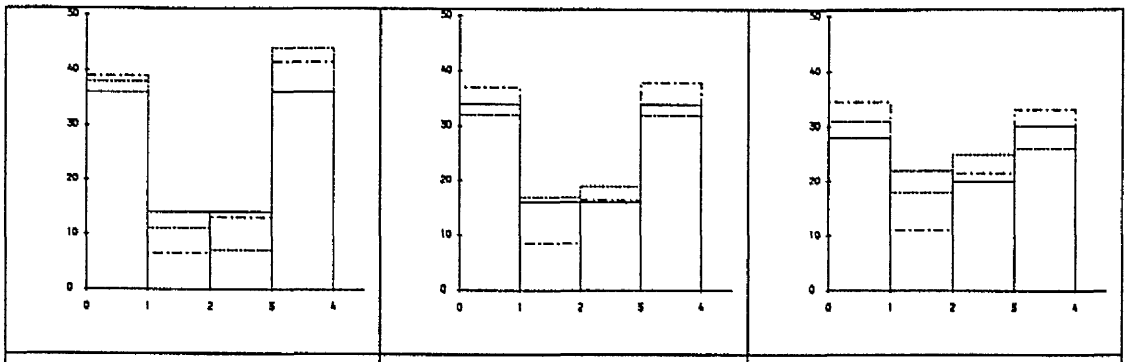


Figure 3.6 (a) :- $K_S = 0.3$ | Figure 3.6 (b) :- $K_S = 0.4$ | Figure 3.6 (c) :- $K_S = 0.5$

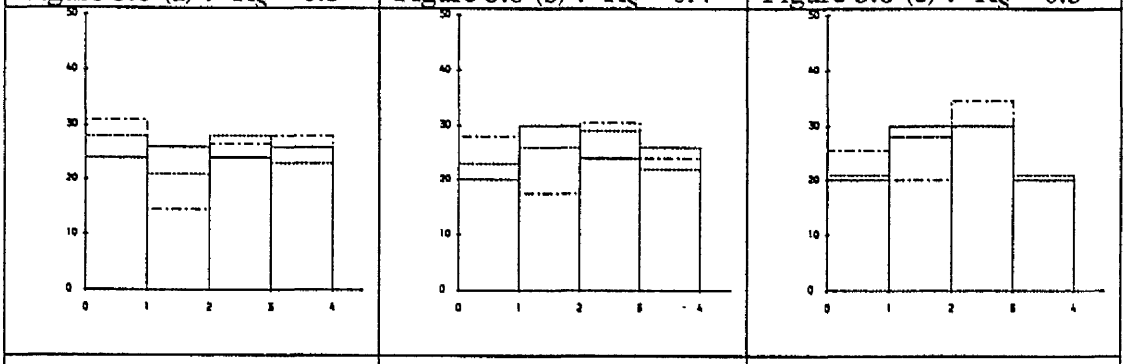


Figure 3.6 (d) :- $K_S = 0.6$ | Figure 3.6 (e) :- $K_S = 0.7$ | Figure 3.6 (f) :- $K_S = 0.8$

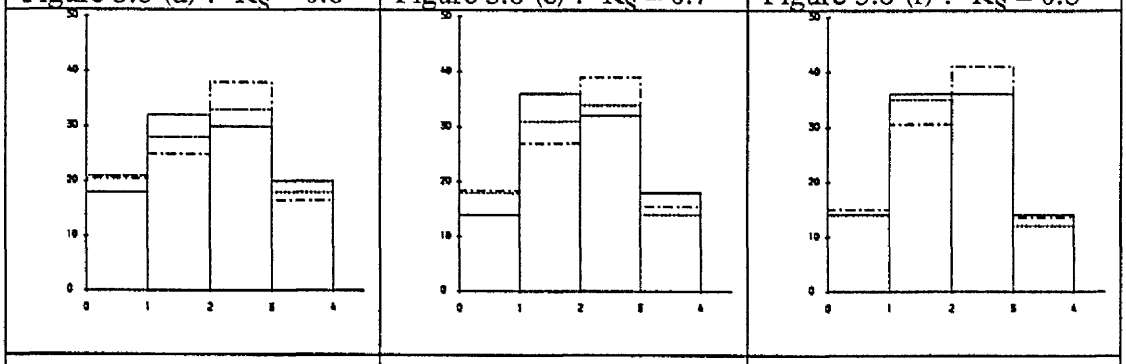


Figure 3.6 (g) :- $K_S = 0.9$ | Figure 3.6 (h) :- $K_S = 1.0$ | Figure 3.6 (i) :- $K_S = 1.1$

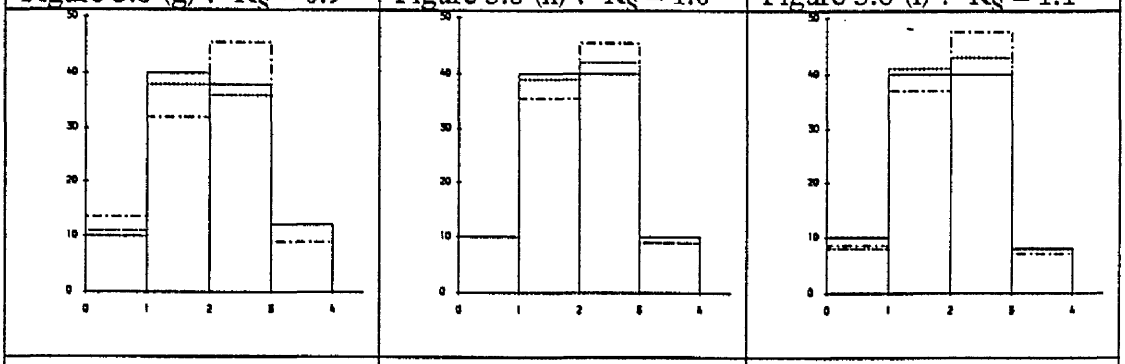


Figure 3.6 (j) :- $K_S = 1.2$ | Figure 3.6 (k) :- $K_S = 1.3$ | Figure 3.6 (l) :- $K_S = 1.4$

Key :- ———— 50 points ..... 100 points —·—·—·— 150 points

Figure 3.6 :- Histogram of the % number of points for four contour levels.

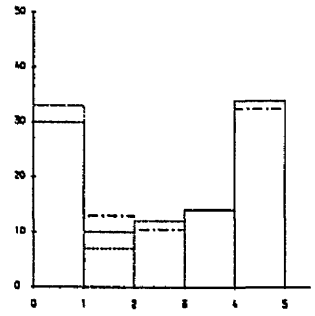


Figure 3.7 (a) :- $K_S = 0.3$

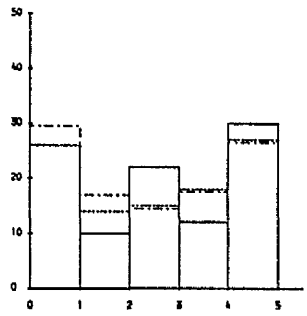


Figure 3.7 (b) :- $K_S = 0.4$

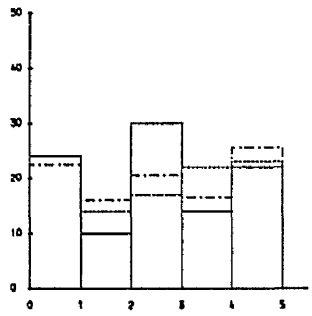


Figure 3.7 (c) :- $K_S = 0.5$

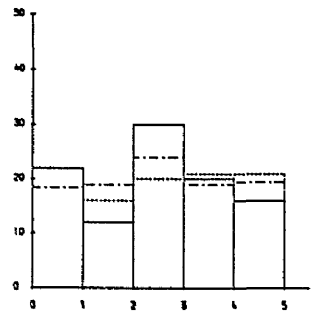


Figure 3.7 (d) :- $K_S = 0.6$

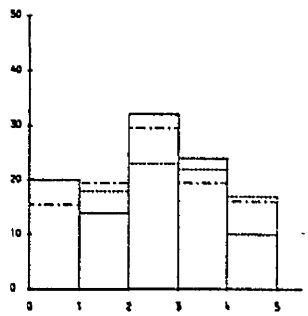


Figure 3.7 (e) :- $K_S = 0.7$

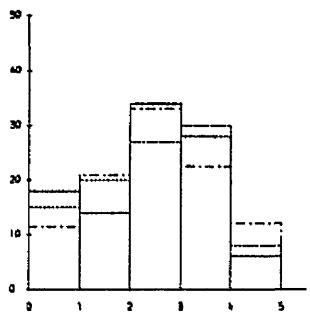


Figure 3.7 (f) :- $K_S = 0.8$

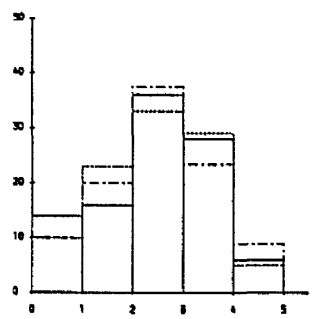


Figure 3.7 (g) :- $K_S = 0.9$

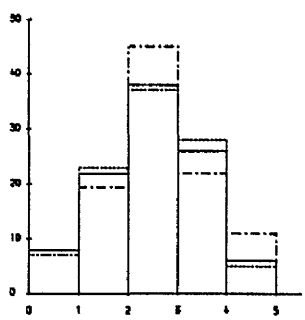


Figure 3.7 (h) :- $K_S = 1.0$

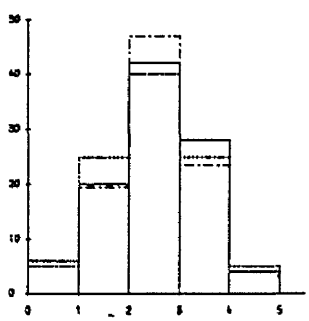


Figure 3.7 (i) :- $K_S = 1.1$

Key :- ——— 50 points ····· 100 points —·—·—·— 150 points

Figure 3.7 :- Histogram of the % number of points for five contour levels.

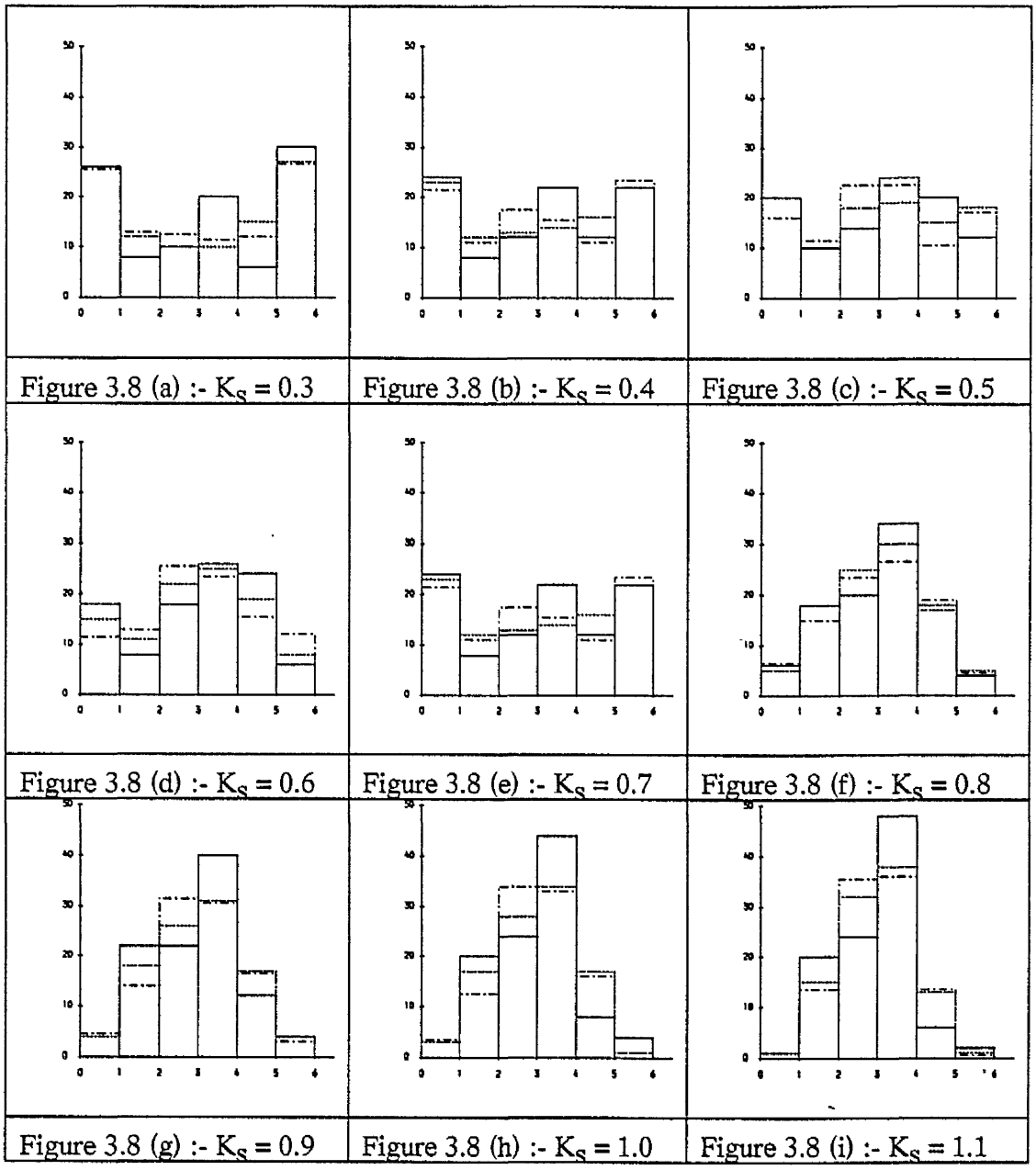


Figure 3.8 (a) :- $K_S = 0.3$ | Figure 3.8 (b) :- $K_S = 0.4$ | Figure 3.8 (c) :- $K_S = 0.5$

Figure 3.8 (d) :- $K_S = 0.6$ | Figure 3.8 (e) :- $K_S = 0.7$ | Figure 3.8 (f) :- $K_S = 0.8$

Figure 3.8 (g) :- $K_S = 0.9$ | Figure 3.8 (h) :- $K_S = 1.0$ | Figure 3.8 (i) :- $K_S = 1.1$

Key :- ———— 50 points ..... 100 points – – – – – 150 points

Figure 3.8 :- Histogram of the % number of points for six contour levels.

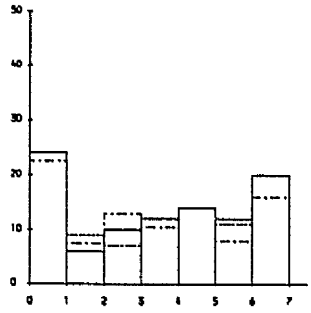


Figure 3.9 (a) :- $K_S = 0.3$

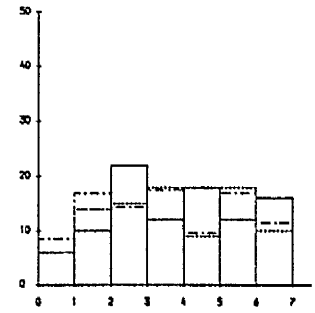


Figure 3.9 (b) :- $K_S = 0.4$

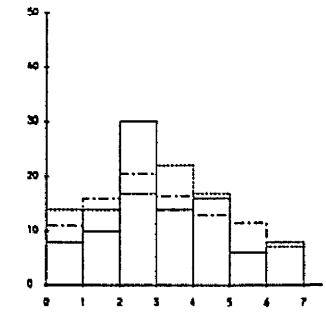


Figure 3.9 (c) :- $K_S = 0.5$

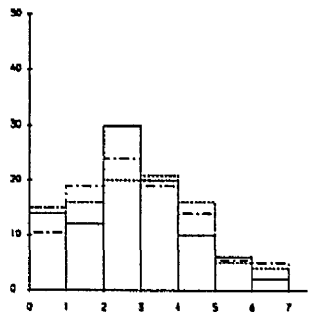


Figure 3.9 (d) :- $K_S = 0.6$

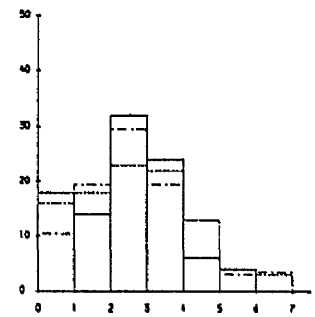


Figure 3.9 (e) :- $K_S = 0.7$

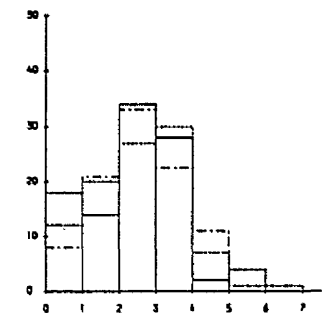


Figure 3.9 (f) :- $K_S = 0.8$

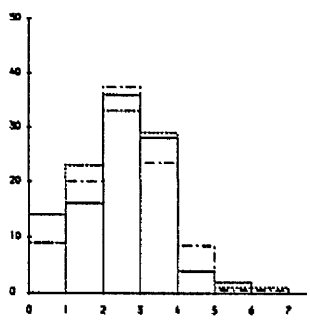


Figure 3.9 (d) :- $K_S = 0.6$

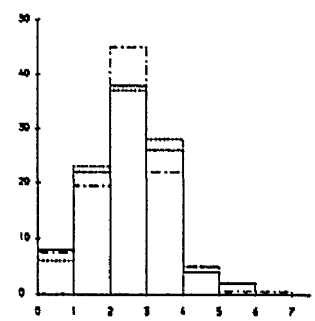


Figure 3.9 (e) :- $K_S = 0.7$

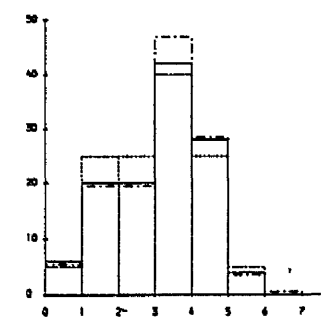


Figure 3.9 (f) :- $K_S = 0.8$

Key :- ———— 50 points ..... 100 points —·—·—·— 150 points

Figure 3.9 :- Histogram of the % number of points for seven contour levels.

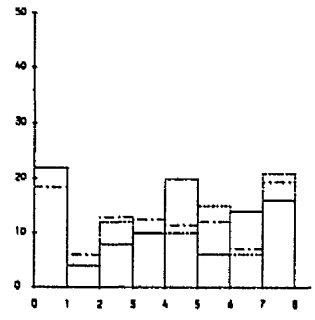


Figure 3.10 (a) :- $K_s = 0.3$

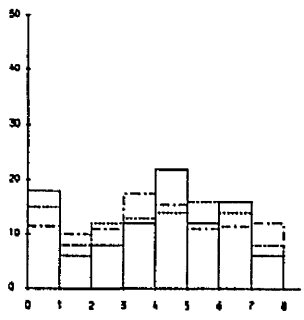


Figure 3.10 (b) :- $K_s = 0.4$

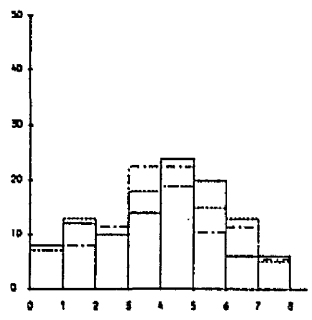


Figure 3.10 (c) :- $K_s = 0.5$

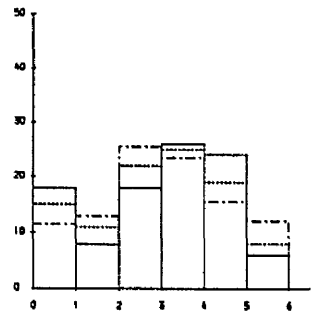


Figure 3.10 (d) :- $K_s = 0.6$

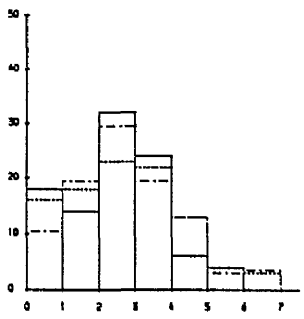


Figure 3.10 (e) :- $K_s = 0.7$

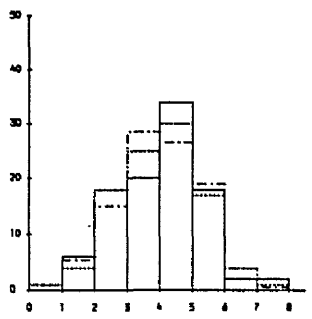


Figure 3.10 (f) :- $K_s = 0.8$

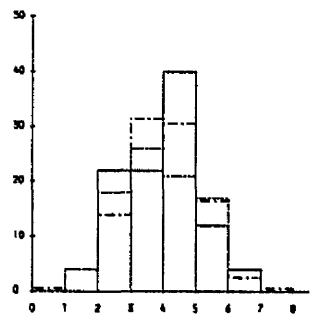


Figure 3.10 (g) :- $K_s = 0.9$

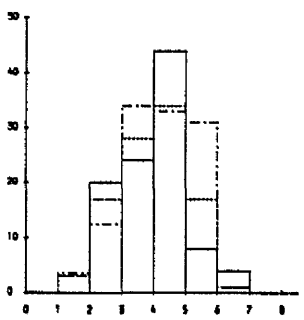


Figure 3.10 (h) :- $K_s = 1.0$

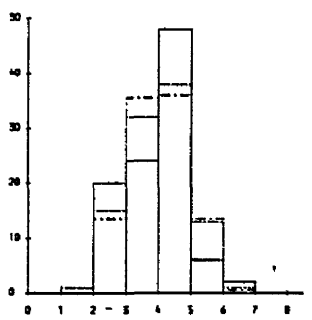


Figure 3.10 (i) :- $K_s = 1.1$

Key :- ———— 50 points ..... 100 points -·-·-·-·- 150 points

Figure 3.10 :- Histogram of the % number of points for eight contour levels.

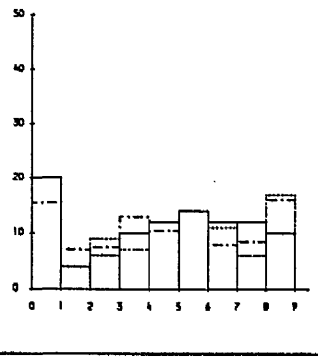


Figure 3.11 (a) :- $K_S = 0.3$

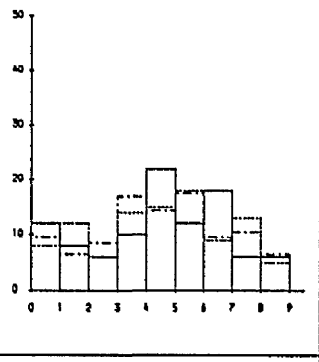


Figure 3.11 (b) :- $K_S = 0.4$

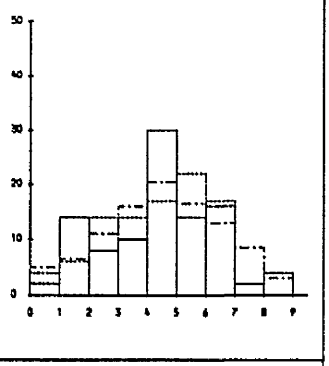


Figure 3.11 (c) :- $K_S = 0.5$

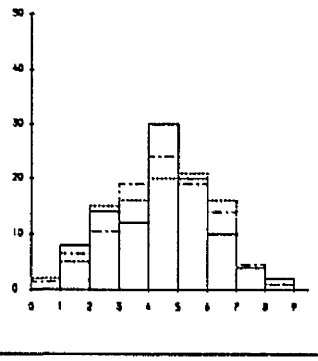


Figure 3.11 (d) :- $K_S = 0.6$

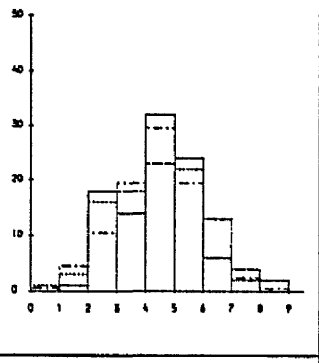


Figure 3.11 (e) :- $K_S = 0.7$

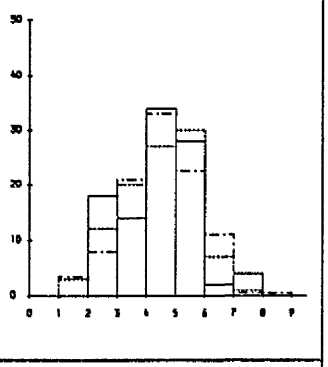


Figure 3.11 (f) :- $K_S = 0.8$

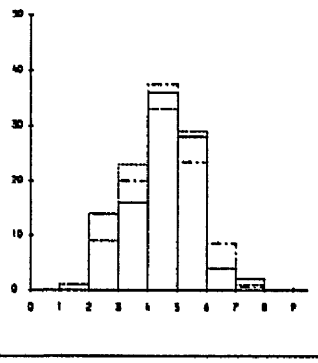


Figure 3.11 (g) :- $K_S = 0.9$

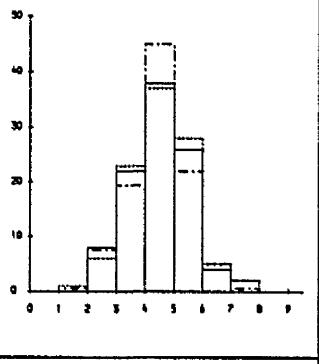


Figure 3.11 (h) :- $K_S = 1.0$

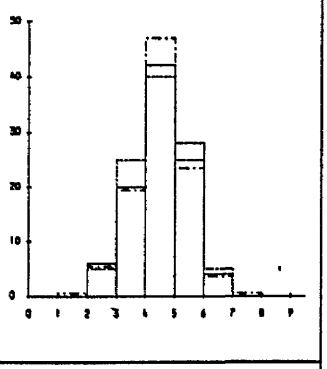


Figure 3.11 (i) :- $K_S = 1.1$

Key :- ———— 50 points ..... 100 points -·-·-·-·- 150 points

Figure 3.11 :- Histogram of the % number of points for nine contour levels.

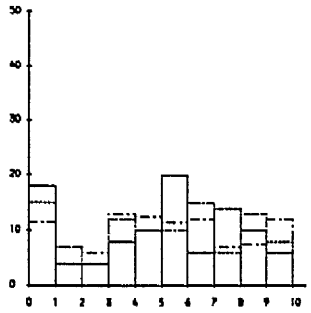


Figure 3.12 (a) :- $K_S = 0.3$

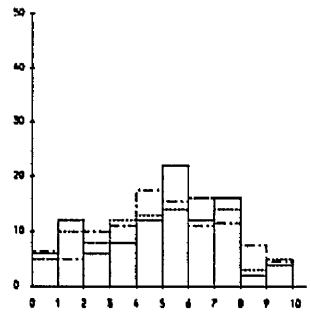


Figure 3.12 (b) :- $K_S = 0.4$

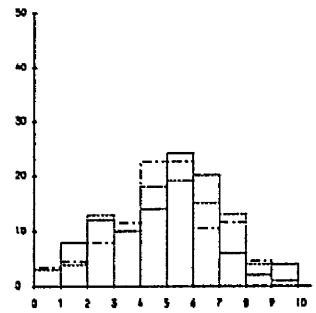


Figure 3.12 (c) :- $K_S = 0.5$

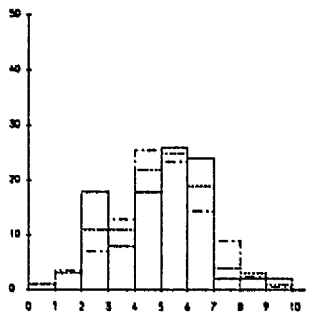


Figure 3.12 (d) :- $K_S = 0.6$

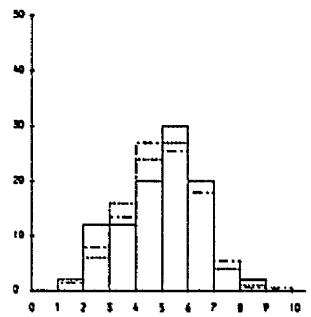


Figure 3.12 (e) :- $K_S = 0.7$

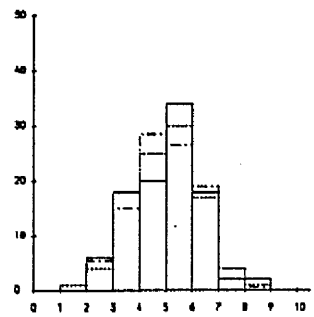


Figure 3.12 (f) :- $K_S = 0.8$

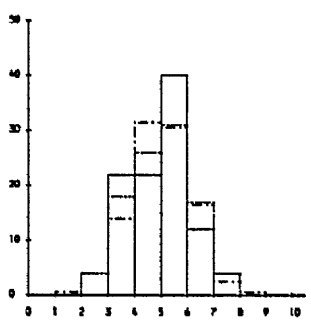


Figure 3.12 (g) :- $K_S = 0.9$

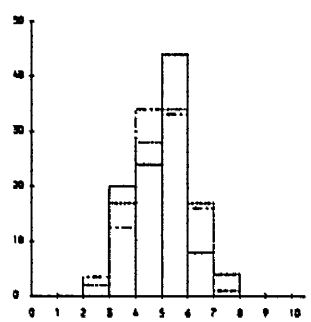


Figure 3.12 (h) :- $K_S = 1.0$

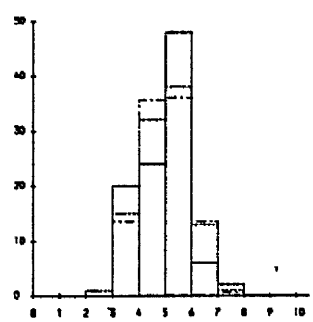


Figure 3.12 (i) :- $K_S = 1.1$

Key :- ———— 50 points ..... 100 points -·-·-·-·- 150 points

Figure 3.12 :- Histogram of the % number of points for ten contour levels.

The result of the study was the formulation of the following two very simple rules :-

1. $\{\log_e(N) \le C_c \le 2\log_e(N)\}$

2. $\left\{\frac{1}{\log_e(C_c)} \le K_s \le \frac{1}{\log_e(C_c/2)}\right\}$

where C_c = number of contour classes
K_s = standard deviation proportion
N = number of data points

These are displayed graphically in figure 3.13.

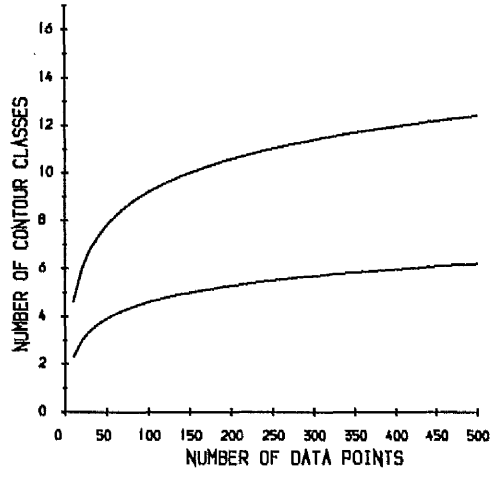


Figure 3.13 (a) :- Selection of number of contour classes.

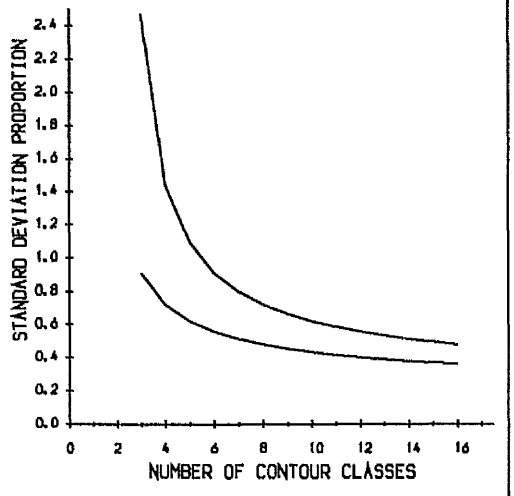


Figure 3.13 (b) :- Selection of standard deviation proportion.

Figure 3.13 :- Graphical representation of the two rules for selecting number of contours and standard deviation proportion.

The above rules were validated on three different data sets, a N(4,10), a Un(7,4) and a Ga(1,0.4), solely for those values of standard deviation proportion and number of contours, determined from the rules. Table 3.3 summarises the relevant combinations of

interest e.g. for six contours, standard deviation proportions of 0.6-0.9 were examined in steps of 0.1 for 50, 100 and 200 points.

Number of contours		4	5	6	7	8	9	10
Standard deviation proportion (K_s)		0.8-1.4	0.7-1.0	0.6-0.9	0.6-0.7	0.5-0.7	0.5-0.6	0.5-0.6
Number of data points	50	yes	yes	yes	yes	no	no	no
	100	no	yes	yes	yes	yes	yes	no
	200	no	no	yes	yes	yes	yes	yes

Table 3.3 :- Summary of the standard deviation proportions and number of contours.

From the histograms of the results, figures 3.14 - 3.20, a well defined rule of thumb materialised as to how the standard deviation proportion should be selected from within an interval. The greater the tendency for the original data to emulate an uniform distribution the closer to the lower bound K_S should be selected; the more peaked the data set becomes, the nearer to the upper bound the ideal value lies. Applying these directives ensures the three criteria mentioned earlier are satisfied.

The rules are not binding and values lying outwith the recommended ranges should still produce satisfactory results. The main advantage of using the above empirical rules is that for data sets differing in size, a number of classes appropriate to all the data sets may be selected. Secondly, the rules are easily implemented with values being defined a-priori, this dispenses with the previous time absorbing trial-and-error method .

No matter what means of contouring interval selection is used, the same technique should be implemented for all surfaces and any parameters required as input defined on the same criteria where a comparison is to be undertaken.

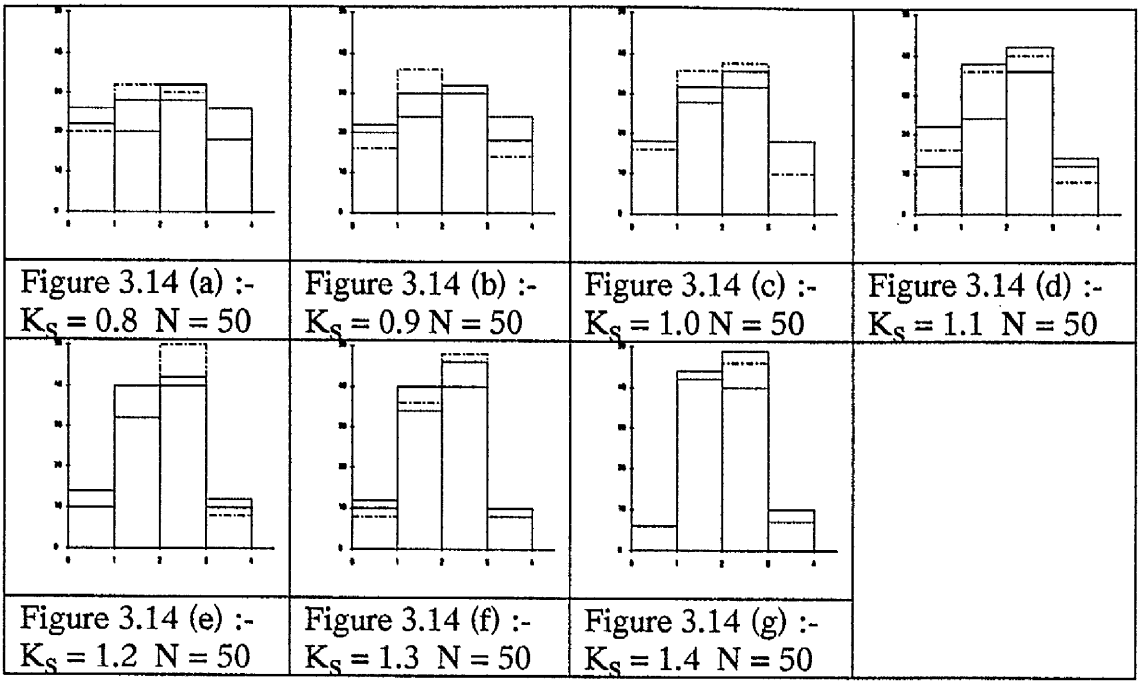


Figure 3.14 (a) :- $K_S = 0.8$ N = 50

Figure 3.14 (b) :- $K_S = 0.9$ N = 50

Figure 3.14 (c) :- $K_S = 1.0$ N = 50

Figure 3.14 (d) :- $K_S = 1.1$ N = 50

Figure 3.14 (e) :- $K_S = 1.2$ N = 50

Figure 3.14 (f) :- $K_S = 1.3$ N = 50

Figure 3.14 (g) :- $K_S = 1.4$ N = 50

Figure 3.14 :- Histograms describing the results for four contours and the relevant standard deviation proportions.

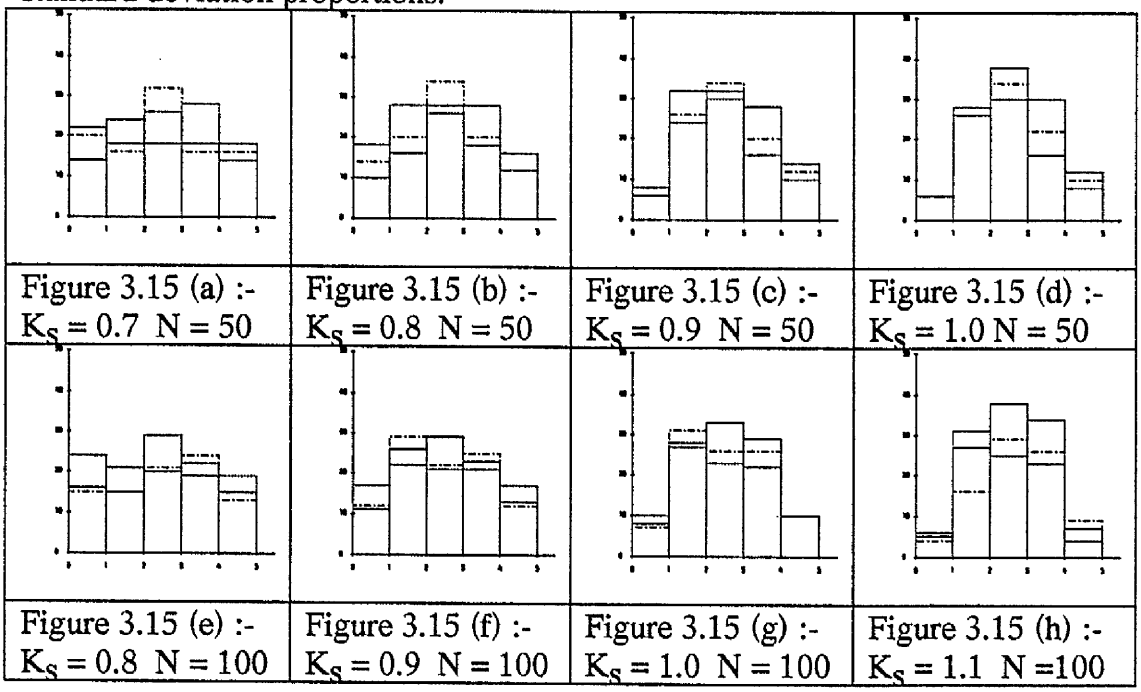


Figure 3.15 (a) :- $K_S = 0.7$ N = 50

Figure 3.15 (b) :- $K_S = 0.8$ N = 50

Figure 3.15 (c) :- $K_S = 0.9$ N = 50

Figure 3.15 (d) :- $K_S = 1.0$ N = 50

Figure 3.15 (e) :- $K_S = 0.8$ N = 100

Figure 3.15 (f) :- $K_S = 0.9$ N = 100

Figure 3.15 (g) :- $K_S = 1.0$ N = 100

Figure 3.15 (h) :- $K_S = 1.1$ N =100

Key :- ——— N(4,10) ..... Un(7,4) – – – – – Ga(1,0.4)

Figure 3.15 :- Histograms describing the results for five contours and the relevant standard deviation proportions.

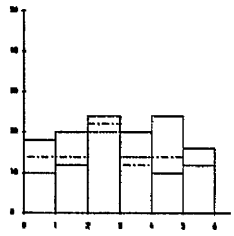


Figure 3.16 (a) :- $K_S = 0.6$ N = 50

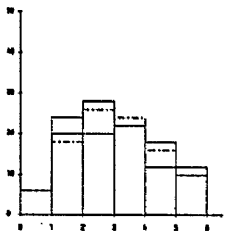


Figure 3.16 (b) :- $K_S = 0.7$ N = 50

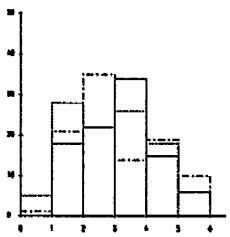


Figure 3.16 (c) :- $K_S = 0.8$ N = 50

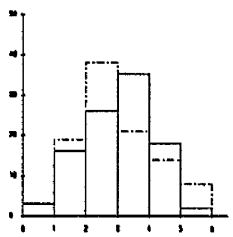


Figure 3.16 (d) :- $K_S = 0.9$ N= 50

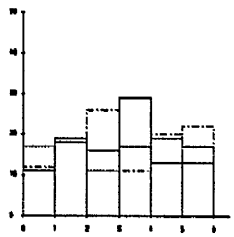


Figure 3.16 (e) :- $K_S = 0.6$ N = 100

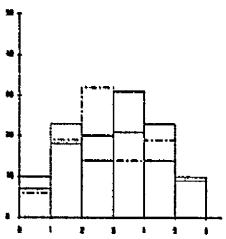


Figure 3.16 (f) :- $K_S = 0.7$ N = 100

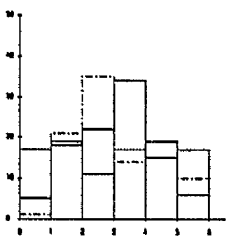


Figure 3.16 (g) :- $K_S = 0.8$ N = 100

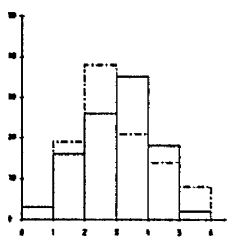


Figure 3.16 (h) :- $K_S = 0.9$ N = 100

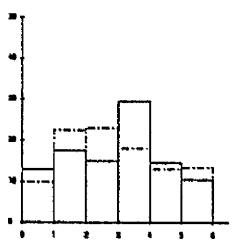


Figure 3.16 (i) :- $K_S = 0.6$ N= 200

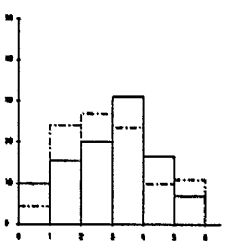


Figure 3.16 (j) :- $K_S = 0.7$ N = 200

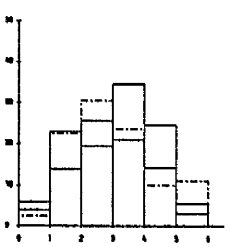


Figure 3.16 (k) :- $K_S = 0.8$ N = 200

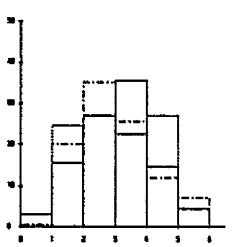


Figure 3.16 (l) :- $K_S = 0.9$ N= 200

Key :- ——— N(4,10) ..... Un(7,4) —·—·—·— Ga(1,0.4)

Figure 3.16 :- Histograms describing the results for six contours and the relevant standard deviation proportions.

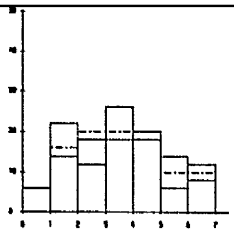


Figure 3.17 (a) :-
$K_S = 0.6$ N = 50

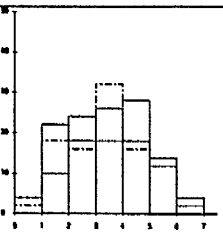


Figure 3.17 (b) :-
$K_S = 0.7$ N = 50

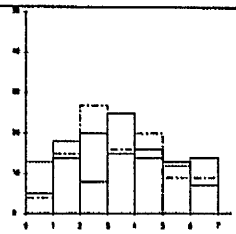


Figure 3.17 (c) :-
$K_S = 0.6$ N = 100

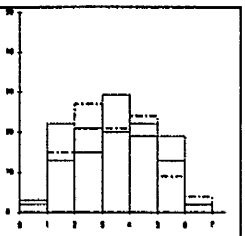


Figure 3.17 (d) :-
$K_S = 0.7$ N = 100

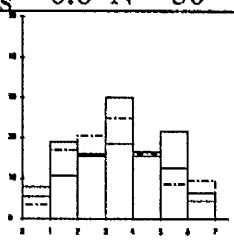


Figure 3.17 (e) :-
$K_S = 0.6$ N = 200

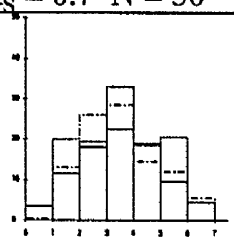


Figure 3.17 (f) :-
$K_S = 0.7$ N = 200

Figure 3.17 :- Histograms describing the results for seven contours and the relevant standard deviation proportions.

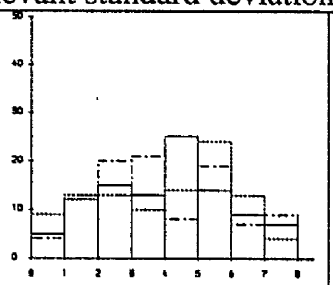


Figure 3.18 (a) :-
$K_S = 0.5$ N = 100

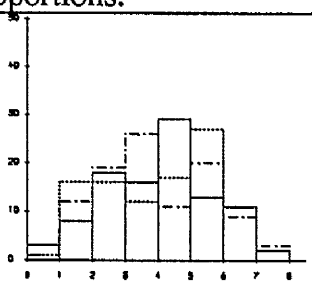


Figure 3.18 (b) :-
$K_S = 0.6$ N = 100

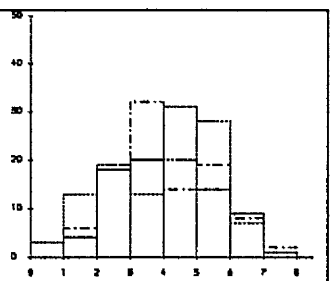


Figure 3.18 (c) :-
$K_S = 0.7$ N = 100

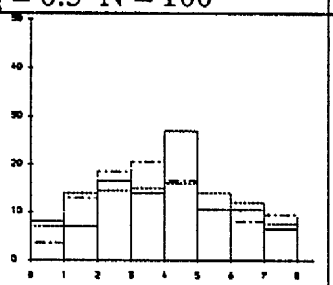


Figure 3.18 (d) :-
$K_S = 0.5$ N = 200

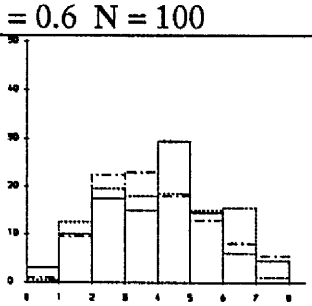


Figure 3.18 (e) :-
$K_S = 0.6$ N = 200

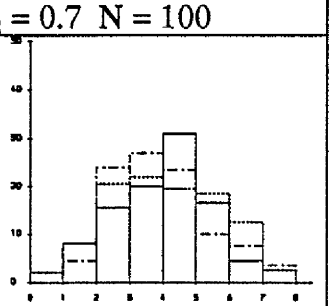


Figure 3.18 (f) :-
$K_S = 0.7$ N = 200

Key :- ——— N(4,10) ····· Un(7,4) –·–·–·– Ga(1,0.4)

Figure 3.18 :- Histograms describing the results for eight contours and the relevant standard deviation proportions.

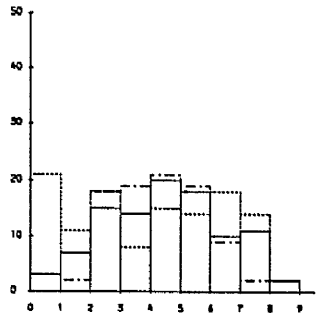


Figure 3.19 (a) :- $K_S = 0.5$
Number of points = 100

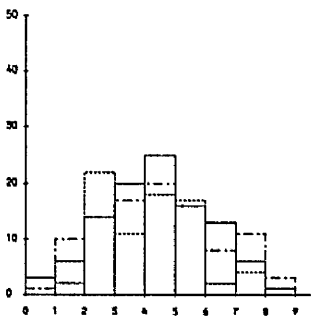


Figure 3.19 (b) :- $K_S = 0.6$
Number of points = 100

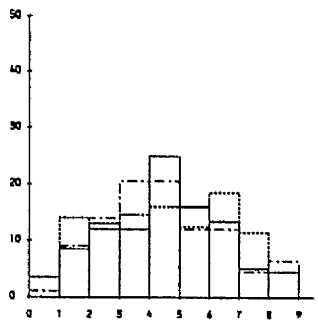


Figure 3.19 (c) :- $K_S = 0.5$
Number of points = 200

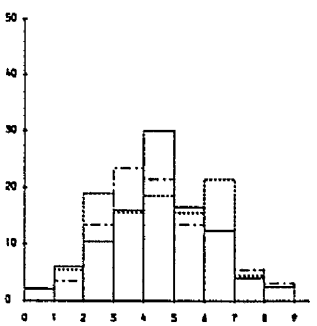


Figure 3.19 (d) :- $K_S = 0.6$
Number of points = 200

Key :- ——— N(4,10) ····· Un(7,4) —·—·—·— Ga(1,0.4)

Figure 3.19 :- Histograms describing the results for nine contours and the relevant standard deviation proportions.

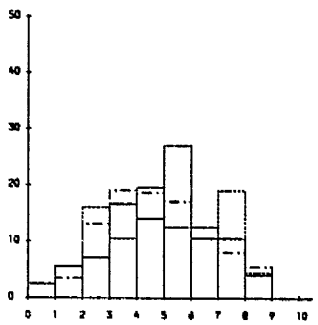


Figure 3.20 (a) :- $K_S = 0.5$
Number of points = 200

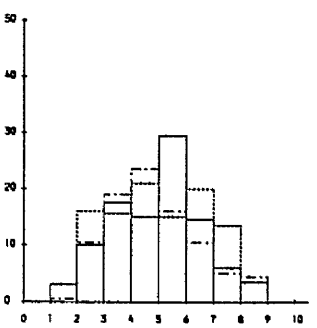


Figure 3.20 (b) :- $K_S = 0.5$
Number of points = 200

Key :- ——— N(4,10) ····· Un(7,4) —·—·—·— Ga(1,0.4)

Figure 3.20 :- Histograms describing the results for ten contours and the relevant standard deviation proportions.

§3.6 Locating and Tracing The Contour

The next step in producing a pictorial representation of the underlying spatial distribution is to physically draw the contours using the a-priori criteria. The form of the data determines the most plausible method for contouring. Routines which contour data on a rectangular grid are the simplest and many algorithms exist. Alternative methods use the ideas of triangulation.

§3.6.1 Contouring by gridding

Gridding is the process of determining values of the surface at a set of locations that are arranged in a regular pattern which completely covers the mapped area. In general, values of the surface are not known at these uniformly spaced points and so they must be estimated from the irregularly located control points where the values are known. The locations where estimates are made are referred to as 'grid points' or 'grid nodes'. The methods discussed in sections 2.3 to 2.5 enable a regular grid to be evaluated e.g. kernel density estimation, trend surface analysis, kriging.

A contour of a given height may be produced in two ways. First each rectangle of the grid is examined in turn and the sections of the contour within that rectangle drawn. Alternatively, once part of a contour has been located within a grid rectangle, the rest of the contour may be traced through the whole grid. Further contours are then sought and when found, they too are traced through the grid.

Broadly speaking, a contour of height h, crosses one of the grid lines (lines between two adjacent grid points, ht(A), ht(B) if:-

$$ht(A) \leq h \leq ht(B)$$

The exact point of intersection of the contour with the gridline must be calculated. If the grid points are sufficiently close that a contour does not cross the line joining them more than once, then a good approximation to the point of intersection may be made using inverse linear interpolation. Having found a starting point, the next point must be found and in a similar manner further points, so the contour may be traced through the grid.

In the 1960's Dayhoff (1963) and Cottafava and le Moli (1969) developed various contouring algorithms. The embryonic ideas of Cottafava and le Moli were brought to maturity in conjunction with those of Dayhoff by Heap and Pink (1969).

Heap and Pink realised that all contours must either cross the boundary or a horizontal grid line hence it was only necessary to record the intersections of the contour with horizontal lines i.e.

$$ht(A) < h \leq ht(B) \qquad (1)$$

where A is immediately to the left of B is sufficient

All such interval lines are marked during a preliminary scan. The grid points on the boundary are then examined to see whether they satisfy (1). Each time one is found, the contour is traced through the grid. This accounts for all open contours. When all open contours are drawn, the closed contours are sought. A search is first made for a marked grid line, when found it serves as the starting point, the contour is traced removing the appropriate marks, until the starting point is reached. This process is repeated until no marks remain.

Two forms of degeneracy are present in contouring, the first which is not applicable in the case of Heap's and Pink's routine, refers to the case where the grid point has the same height as the contour of interest. Cottafava and le Moli (1969) encountered this problem but by 'virtually' altering the value of the height at the grid point by a small amount this problem is eradicated. Rothwell (1971) and Crane (1972) adopted a similar approach.

Secondly if we consider a rectangle where the height bears the relationship to the contour height as shown, figure 3.21(a). Given entry on a particular side, all three sides appear to be the exit side. This suggests that another contour passes through the rectangle hence three solutions are plausible, figure 3.21 (b) - 3.21 (d) :-

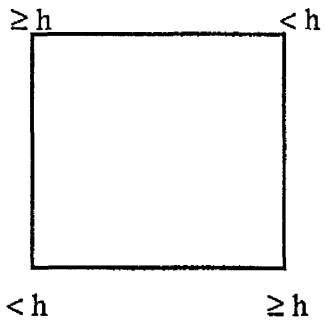


Figure 3.21 (a) :- Illustration of a situation where a degeneracy may arise.

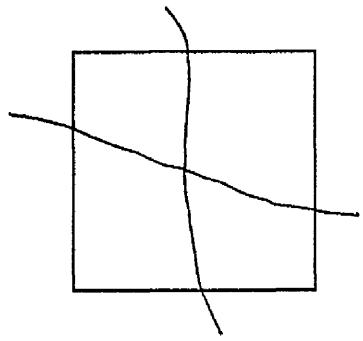


Figure 3.21 (b) :- Solution A.

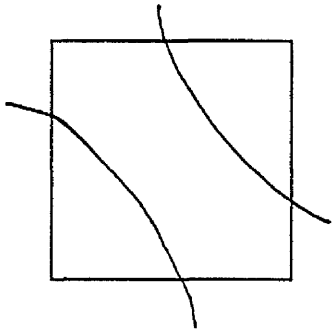


Figure 3.21 (c) :- Solution B.

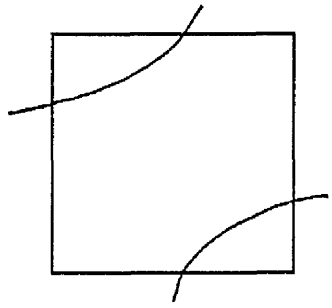


Figure 3.21 (d) :- Solution C.

Figure 3.21 :- Illustration of contour degeneracy.

Heap and Pink used the idea of Dayhoff to circumvent this problem. They approximated the height at the centre of the rectangle by averaging the height at the four grid points. The rectangle was then divided into four triangles which cannot be degenerate.

Cottafava and le Moli (1969) argue the situation only occurs rarely and select either solutions B or C depending on how the rectangle is encountered. Rothwell (1971) adopts a more complex solution involving directions.

§3.6.2 Contouring By Triangulation

This method avoids the necessity for initially interpolating the grid points onto a regular grid. Control points are assumed to be located without any particular regularity, these are initially connected by straight lines. This forms a mesh of triangles that covers the

surface. By interpolating down the sides of the triangles, locations can be found where the elevation is a constant, specified value.

It is apparent that if the control points are connected in a different manner, a different set of triangular plates will be defined and a different sequence could result in conspicuously different-appearing contour lines. A set of unique, 'optimal' triangles are therefore required. Gold et al. (1977) suggested possible solutions, the individual triangles should be as near equilateral as possible, they should have the minimum possible height or alternatively the longest leg of each triangle should be the shortest possible.

This problem impeded the development of triangulation based contouring algorithms since an iterative process was required to obtain the optimal configuration. Using a network referred to as Delaunay triangulation Gold et al. (1977) and McCullagh and Ross (1980), developed an algorithm which produced almost optimal networks on the first pass. Delaunay triangles are defined uniquely for a given set of data points. The triangles formed being as nearly equiangular as possible with the longest sides of the triangles being as short as possible.

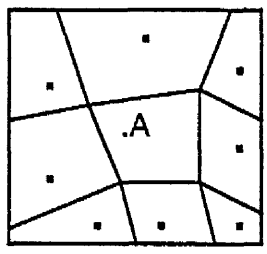


Figure 3.22 (a) :- Illustration of Thiessen polygon.

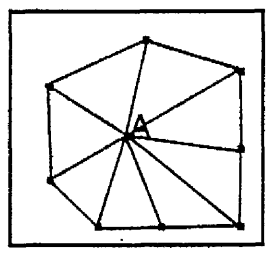


Figure 3.22 (b) :- Delauney triangulation network.

Figure 3.22 :- Illustration of contouring by triangulation.

The idea behind a Delaunay triangular network is as follows; in a field of scattered points, each point is surrounded by an irregular polygon such that every location within a polygon is closer to the enclosed part than it is to any other, figure 3.22 (a). Conversely, every location outside a specific polygon is closer to some other point than it is to the point within the polygon. This is the most compact division of space. These are commonly referred to as Thiessen, Dirichlet or Voronoi polygons. Immediately

surrounding the Thiessen polygon enclosing a specific point A, are other Thiessen polygons, each of which encloses a single point. If these points are connected by a straight line, the result is a Delaunay triangular network, figure 3.22 (b).

The assumption that the triangles represent tilted flat plates is a very crude approximation of a surface. A better approximation can be achieved using curved or bent triangular plates, particularly if these can be made to join smoothly across the edges of the triangles. Tipper (1979) described how the three-dimensional equivalent of the spline function may be used to ensure that abrupt changes in direction where contour lines cross from one plate to another are avoided. McCullagh (1983) uses a tricubic polynomial to avoid this problem, the details are given by Birkhoff and Mansfield (1974).

The finished map, contoured by triangulation is very similar to a map produced using a gridding algorithm. The main difference arises where the surface slope changes abruptly, the triangulation method makes the surface look sharper than it is in reality. Along the margins of maps this problem may be corrected by the judicious insertion of pseudo-points, but these cannot be inserted into the body of the map.

§3.7 Alternative Contouring Techniques

Other forms of contouring are available. Powell (1973) divided each rectangle into eight triangles and approximated the surface by piecewise quadratic functions. A contour line may then be approximated as a sequence of pieces of conic sections which can be drawn easily, since conic sections have a convenient parametric form.

McLain (1974) describes a method for arbitrary data points, but advocates a two-stage process, first interpolating heights to a regular mesh and then contouring the rectangular mesh. The method suggested was that of bicubic splines. Once the intersection of a contour with a grid line has been located, the contour is traced through the rectangle by a series of short steps in one of eight directions, N, NE, E, SE, etc. The direction of the next step is selected from one of three directions which depend on the previous step i.e. one in the same direction and two at 45 degrees to either side of the previous step. The step selected is that which is closest in value to the contour height.

Many other techniques are available, but the methodology of Heap and Pink (1969) is simple to understand and has stood the test of time. Coding has been made available,

Heap and Pink (1969), and after minor modifications and conversion to Fortran 77, this method was used to produce all contours.

A further advantage of implementing this sub-routine and not using one of the commercially available packages for producing contours e.g. Ghost 80, UNIRAS, S-Plus etc. was the extra feature that contour co-ordinates were accessible to the user. This is not a feature of the majority of graphical packages.

§3.8 Accuracy of Statistical Maps

Within the previous sections, methodology was developed for selecting the number and bounds of class intervals and a number of techniques were described as to how a contour may be drawn. However a question fundamental to the construction of a statistical map is how accurate is the resultant surface.

The question of accuracy arises when an isarithmic map has been constructed from data which do not cover the entire domain of the mapped area, these data points are subject to observational and/or location errors and finally to errors generated by the method of interpolation between punctual sample points. In such instances the map is referred to as an estimated map as opposed to the underlying error free map.

Accuracy is then a measure of the displacement between the estimated and true map, the latter which is rarely, if ever, available. Various attempts have been made within the literature to quantify the accuracy of maps.

Blumenstock (1953) discussed the problem of assessing reliability of isarithmic maps for meteorological maps but his analysis is of wider significance for all maps of geographical phenomena. The method he initiated was to first take the basic data and estimate the magnitude of all the sources of unreliability. The data was then corrected for bias and thirdly, the standard error was computed in terms of the observational and sampling error. The final step required the determination of the chance that any one particular plotted value will be in error by x units, solely on the basis of the unreliability of the data. This final step enabled an estimate to be obtained as to how many of the plotted values lie outside their correct isarithmic intervals.

The method is constrained in its usefulness since it is reliant on the person implementing the technique to define levels of reliability of their data and secondly it assumes the errors are normally distributed.

Switzer (1975) examined the question of error induced as a result of interpolating between punctual sample data. He considered a map to be a partition of a domain, R, into k sub-domains, k being the number of colours used in a map. The sub-domains are denoted $R_1, R_2,...R_k$ in the true map and are denoted $\hat{R}_1, \hat{R}_2, \ldots, \hat{R}_K$ in the true map. The n data points underlying the estimated map are at locations $s_1, s_2, s_3, \ldots, s_n$, which may or may not be at the centres of basic sampling cells, $S_1, S_2, S_3, \ldots, S_n$. The estimated map is constructed by assigning a cell to the i^{th} sub-domain $\hat{R}_i$ if the data point inside that cell is observed to fall in R_i, the n basic sampling cells being a relatively fine partition of the domain, R.

As an index of precision, Switzer examined the discrepancy between the true and estimated maps as measured by the mismatch areas:-

$$\text{i.e.} \qquad L_{ij} = \mu\left(R_i \cap \hat{R}_j \right) \qquad i \neq j$$

$$\mu = \text{area}$$

Where L_{ij}, the precision index, can be expressed in terms of Lebesgue integrals:-

$$L_{ij} = \sum_{h=1}^{n} \delta_j(s_h) \int_{S_h} \delta_i(s) d\mu(s) \qquad (6)$$

where L_{ij} = area of that region which belongs to true subdomain i but is represented as subdomain i on the estimated map.

δ_i = indicator function for the true subdomain i

S_h = sampling cells

s_h = data points

Problems with such a technique relate to the difficulty of modelling the form of the location error, differences in the index will materialise depending on the model selected, e.g. spherical, square, hexagonal, etc.

Other techniques although developed for choropleth maps may be extended to contour maps, Jenks and Caspall (1971). Problems with attempting to evaluate an accuracy index are that the indices themselves require the estimation of various quantities, or alternatively, certain assumptions require to be satisfied for the index to be applicable, so far no method has been developed to circumvent these problems.

Mackay (1953) showed problems in accuracy may not necessarily originate from the original data but from the locational pattern of the recording points or 'control' points. Where choropleth maps of discrete counties are concerned, the framework of control points is fixed; where isarithmic maps are being developed from point observations, the framework problem may be more complex. The implementation of Heap and Pink's triangulation method for contouring takes account of this possible error source.

Hsu and Robinson (1970) and Morrison (1971) illustrate that both sample size and the method of areal sub-division affect the quality of data and hence of isopleth maps.

A further source of error is attributable to how the points on a contour are joined. They may be joined as straight lines or by a curve fitting algorithm. Although the latter produces aesthetically pleasing contours, they are not necessarily accurate. Too often the contours produced by this method reflect the curve fitting algorithm used rather than the data being contoured, especially if local curve fitting algorithms are used. As a consequence, when a curve fitting algorithm is used there is no guarantee that contours of different heights will not cross, the situation is consequently totally unacceptable.

In practice to avoid this problem straight lines should be used to join the points on a contour. Not only does it have the advantage of simplicity of implementation but it also gives the viewer of the resultant map a good appreciation of how coarse or fine the map really is.

The utilisation of straight lines to join points of equal heights is of less consequence when the grid over which we interpolate is extremely fine. Neither the shape of the contour nor various geometric attributes associated with it, such as area, perimeter, will be unduly affected. For a coarse grid differences will arise in the measures and possibly shape. A very simple example illustrates this point.

For a simple construct, the circle, the question of what kind of accuracy is achievable with x-data points was examined. A circle of unit radius, centre the origin was depicted using 4, 8, 16, 32 and 64 data points. Table 3.4 evaluates the accuracy of the result by taking the ratio of the grid value to the true value for area and perimeter.

Number of points	Angle subtended between successive data points	Ratio of grid area to true area	Ratio of grid perimeter to true perimeter
4	90	0.6366	0.90031
8	45	0.9003	0.97449
16	22.5	0.9745	0.99358
32	11.25	0.9936	0.99839
64	5.625	0.9984	0.99960

Table 3.4 :- Accuracy of evaluation of a unit circle for x-data points.

The results of the table show that changing the number of points to specify the area of interest has less effect on perimeter than area. For both measures 99% accuracy was achieved using 32 points, specifying additional points served only to increase the computational time for minimal gain in accuracy. By increasing the number of points a more aesthetically satisfactory picture results.

The question of grid resolution is of considerable interest since it may inadvertently have an effect on the outcome of the analysis. This is examined in detail in chapter 4.

In reality, it is extremely difficult to place any sort of numerical bounds on the accuracy of the resultant statistical map since the true surface is never known.

§3.9 Summary

The construction of a contour plot from a set of spatially referenced data follows a sequence of clearly defined steps:-

1. If the data are irregularly spaced, an interpolation technique which enables a regular grid to be estimated is implemented. The methodology used was that of kernel density estimation with smoothing parameter selection undertaken using least-squares cross validation.

2. Select number and bounds of class intervals using the following empirical rules and by studying the histogram of the data to decide whether the variables should be selected closer to the upper/lower bound:-

1. $\{\log_e(N) \leq C_c \leq 2\log_e(N)\}$

2. $\left\{\frac{1}{\log_e(C_c)} \leq K_s \leq \frac{1}{\log_e(C_c/2)}\right\}$

where C_c = number of contour classes
K_s = standard deviation proportion
N = number of data points

3. Using Heap's and Pink's contouring methodology construct the contour plot using straight lines to join the contour nodes.

By implementing the three steps described on a consistent basis, the error attributable to the procedure should be of the same order for all surfaces.

CHAPTER 4

DIFFERENT METHODS OF SURFACE COMPARISON

§4.1 :- Introduction

Many methods are available for effecting a comparison between two or more spatially referenced data sets, these range from the very simple to the more complex and computationally intensive approaches. The form of the analysis depends primarily on four factors:-

1. The manner in which the data is presented e.g. point, line, areal or surface.
2. Whether variables to be compared relate to the same measurement.
3. Whether the comparison is inter-regional.
4. Whether the observations are recorded at the same locality.

A simple way of categorising the various methods available for performing a comparison is to group them according to their mode of interpretation:-

1. Subjective comparison.
2. Comparison technique for a global analysis.
3. Comparison technique relevant to a local analysis.

§4.2 Subjective Analysis

The simplest and oldest method of comparison is to overlay the maps of interest and describe how their distributions differ.

Where information has been recorded on variables at the same locality i.e. a repeated measures format, an isopach/residual map may be constructed to describe the spatial arrangement of the differences between the surfaces.

Other subjective approaches include plotting dependent versus independent variables and making a visual assessment of the strength of the relationship. McGlashen (1972) conducted a survey in Central Africa of fifty five diseases and twenty environmental

factors that might be associated with the diseases. Data came from patient records at 84 hospitals. From visual examination of these disease and factor plots, McGlashen was subsequently able to perform a contingency table analyses. For example, he compared the number of annual cases of diabetes mellitus with whether or not cassava was the staple food eaten by hospital patients. In a similar manner Prentice et al. (1991) subjectively assessed the goodness of fit between observed and simulated isopall maps i.e. maps of equal pollen count, by plotting the observed levels of pollen count against those values simulated from a response surface model.

These three techniques of subjective comparison, superposition of maps, isopach maps and dependent versus independent variables are illustrated in figure 4.1. They examine the relationship between the dispersion of two measures of ambient radioactivity, beta and gamma radionuclides, within the south-west of England. The majority of external radiation dose to the population originates from radionuclides in two of the three natural radioactive decay series i.e. ^{238}U and ^{232}Th. The decay of these sources of gamma radiation causes the emission of both alpha and beta particles into the atmosphere. It is therefore anticipated that a strong association will exist between the two radiation variables.

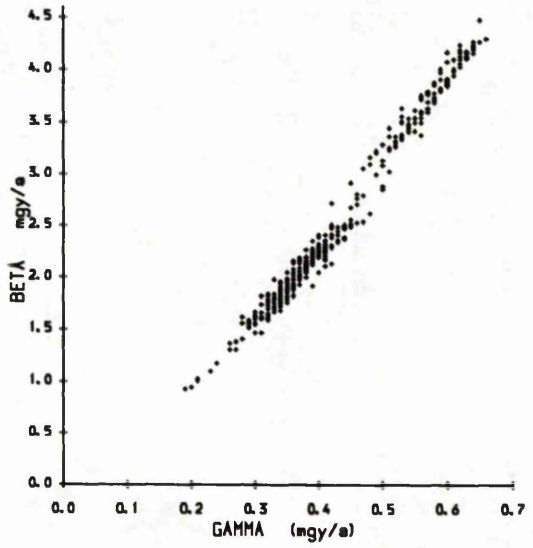
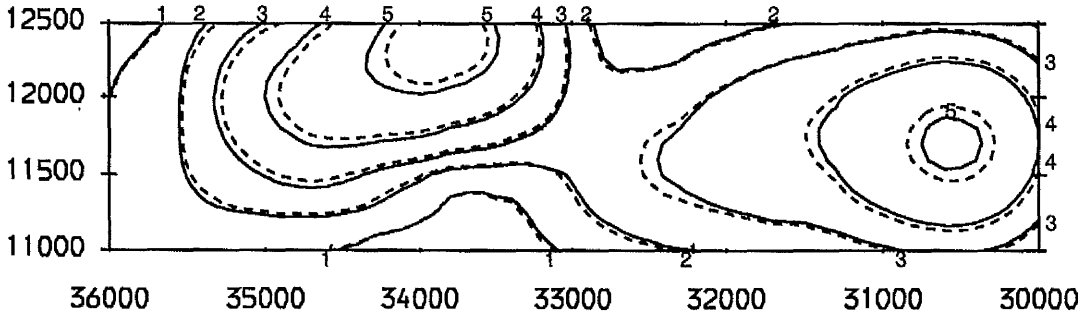


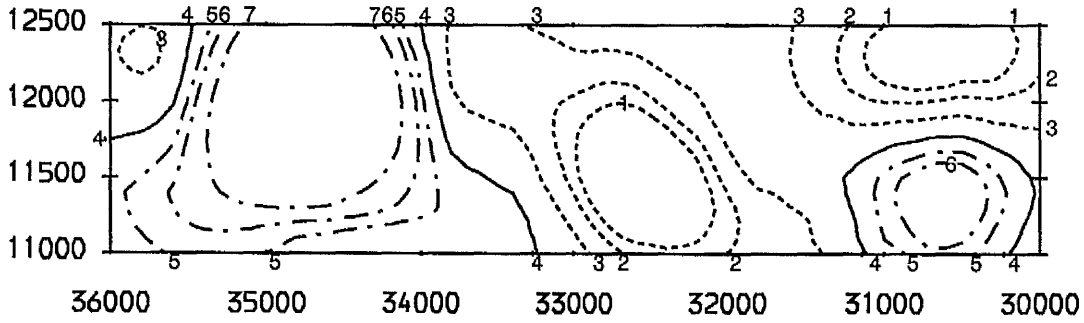
Figure 4.1(a) :- Plot of Gamma (mgy/a) versus Beta (mgy/a) for the south-west of England.



KEY

	Beta (mgy/a)		Gamma (mgy/a)	
1	5% percentile	2	25% percentile	
3	50% percentile	4	75% percentile	
5	95% percentile			

Figure 4.1(b) :- Spatial distribution of beta and gamma radiation variables within the south-west of England.



KEY

— — —	1	-1.5	—·— 5	0.5
— — —	2	-1.0	—·— 6	1.0
— — —	3	-0.5	—·— 7	1.5
———	4	0.0		

Figure 4.1(c) :- Residual/isopach map of the two radiation variables, beta and gamma, in the south-west of England.

Figure 4.1. :- Various forms of subjective analysis.

Subjectively, all three methods suggest the existence of a strong areal dependence between the two radiation fields, the information imparted by each diagram is different. The first diagram ignores the spatial dimension of the data. A simple graphical presentation of the relationship between the beta value at station i and the gamma value at station i is illustrated in figure 4.1 (a). This plot is only applicable when the measurements are recorded at the same locality, thus restricting its usefulness in presenting comparative information.

The remaining two methods incorporate the spatial dimension of the data. Figure 4.1 (b) illustrates the superposition of the two surfaces representing the spatial distribution of the variables. A strong spatial association is apparent between the two variables, as indicated from the preliminary discussion. Where differences do materialise between the variables, this mode of presentation enables those areas of greatest change/similarity to be both located and also, a description given as to the structure of the change. One method of describing change is in terms of various mathematical descriptors e.g. change in orientation, location and/or size between a set of comparable contours. This idea forms the basis of the methodology developed in the following two chapters to describe and quantify change. For this example, differences between the two data sets are minimal, it primarily arises in terms of the size of the contours; contours relating to the gamma radiation field extend over a smaller area than those depicting beta radiation, particularly in the west, the converse is true in the east.

The specialised nature of the data i.e. observations recorded at the same locality, enables a residual map to be constructed. Before the residual map could be produced, it was first necessary to normalise the two data sets. Interpretation of a residual map of this form is not as intuitive to non-statisticians, however it localises those areas of greatest change in both a positive and negative direction. In the west, the beta values tend to be higher than those of the gamma radiation field. In the east, with the exception of the south-east corner, the reverse is true.

The overlaying of the surfaces is the most versatile of the methods since it firstly, incorporates the spatial dimension of the data and secondly, variables recorded at different locations may be examined using this method, this is not plausible for the other two methods.

A fundamental drawback to all the above methods lies in their subjectivity. Confronted with the same two maps, not everyone will agree that an areal association exists, or will assess the degree of association as the same.

§4.3 Global Analysis

In terms of an overall analysis, a number of numerical measures have been proposed which attempt to eliminate the uncertainty attributable to subjective techniques.

§4.3.1 Lorenz curves

The Lorenz (1905) curve is a diagrammatic tool which allows for the visual and quantitative comparison of the cumulative relationship between two variables. The Lorenz curve may be defined mathematically as the curve whose ordinate and abscissa are Φ and F, respectively, such that:-

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

$$\Phi(x) = \frac{1}{\upsilon}\int_{-\infty}^{x} xf(x)dx$$

where υ is the population mean of x. Convexity to the F-axis is a necessary condition for all Lorenz curves.

The commonest equality measure is the Gini (1913-1914) coefficient which is a direct function of the Lorenz curve. It can be shown, Kendall and Stuart (1958), that the Gini coefficient is equal to twice the area between the line $F = \Phi$ and hence the Lorenz curve is defined to be :-

$$G_{coeff} = \frac{1}{2\upsilon}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|x-y|dF(x)dF(y)$$

A Lorenz curve is an area-by-area plot of the ratios of the two variables made in order to indicate similarities of the distribution. The calculations involved for two variables z_1, z_2 are:-

1. evaluate $\frac{z_{1j}}{z_{2j}} = z_{rj}$ where z_{ij} = variable i for region j

z_{rj} = ratio for region j

i = 1,2 j = 1,,..,N

2. rank z_{rj}, smallest proportion given rank one.

3. evaluate $\frac{z_{ij}}{\sum_{j=1}^{N} z_{ij}} \times 100\% = sz_{ij}$

where sz_{ij} = standardised variable i for region j

4. Maintaining z_{rj}, the sz_{ij} are accumulated, $\%z_{ij}$, i.e. cumulative percentages for variable i, region j

5. plot $\%z_{1j}$ against $\%z_{2j}$

Where the distributions are proportionally identical in each area, the resultant plot is a straight line through the origin with slope one, with completely separate distributions resulting in a line which would follow the z_2 axis. In a real world situation the curve will lie between the two extremes.

The difference between the plotted curve and the theoretical optimal is a measure of the dissimilarity of the distributions. Various indices are available for quantifying the level of dissimilarity :-

1. $D_m = \max\left|\%z_{1j} - \%z_{2j}\right|$

2. $D_1 = \frac{1}{2}\sum_j \left|\%z_{1j} - \%z_{2j}\right|$

3. $D_f = \frac{\text{area between the curves}}{\text{area below diagonal}}$

The latter index appears in the literature in a variety of guises and has been given different names according to the nature of the two variables. In population studies where z_1 is the population and z_2 the land area, it is termed the index of population concentration. For the same z_2, but z_1, the land area under a particular use, it is described as the coefficient of areal localisation.

An example of a Lorenz curve is given in figure 4.2. The curve examines the relationship between the area of civil parish districts within the south-west of England, around the towns of Illminster and Yeovil, and the population contained within each of these parishes.

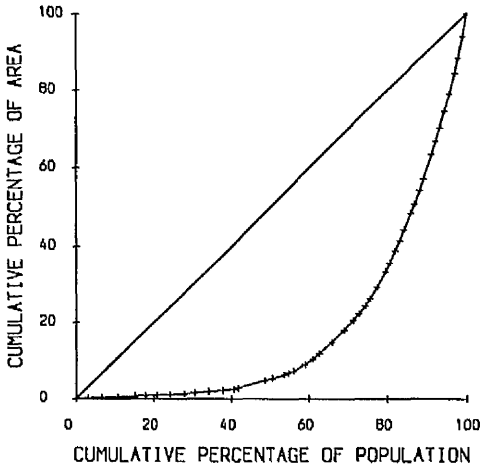


Figure 4.2 :- Lorenz curve for civil parish data within the south-west of England.

	D_m	D_l	D_f
Dissimilarity results	51.10	51.10	71.04

Table 4.1 :- Dissimilarity levels for the Lorenz curve

Table 4.1 reports the results for the various measures of dissimilarity. D_f is numerically greater than the other two but it makes use of all the available information; in all three cases the level of dissimilarity indicates that the distributions do not correspond very closely. Generally small civil parishes are associated with high populations and vice-versa. The small parishes tend to be located in urban areas with the geographically much larger parishes within rural areas.

The Lorenz curve is a useful graphical and numerical tool for making comparisons, but it suffers from a number of disadvantages:-

1. Its use is constrained to choropleth maps.
2. It should not be used with negative numbers or density ratios.
3. It will not discriminate between different arrangements of the areal units.

§4.3.2 Coefficient of areal correspondence

Minnick (1964) constructed a cartostatistical method for determining areal correspondence that was simple, meaningful and clearly interpretable. In terms of basic algebra:-

$$C_a = \frac{A \cap B}{A \cup B}$$

$$= \frac{\text{area over which phenomena are located together}}{\text{total area covered by the two phenomena}}$$

Completely separate distributions give a value of zero whilst for exactly coincident distributions, a value of one is reported for the coefficient of areal correspondence, C_a.

Returning to the example cited on the relationship between the two radiation fields, section 4.2, a value for the coefficient of areal correspondence was calculated to be 0.974, when the two fields were compared at their respective median values. Interpretation of this value confirms analytically what had been expressed in terms of our subjective beliefs i.e. the two radiation variables display a strong association.

An equivalent measure, the proportion of total area of the observed map categorised correctly by a simulated map, PCC, was used by Webb et al. (1987) to quantify the level of association between observed and simulated fossil-pollen maps. They acknowledged that the significance of measures of association is difficult to assess analytically since the map patterns are spatially highly-correlated, Cliff and Ord (1981). In an attempt to compensate for this drawback, an empirical reference distribution for each of the measures was computed for all possible comparisons among the observed maps for all pollen types and times, but those comparisons within each type at different times were excluded. This reference distribution gave an indication of how large PCC could be when

no association between maps was expected. For a specific comparison of an observed and simulated map, a large value for one of the measures relative to the values in the reference distribution, indicates a stronger association between the simulated and observed map than that expected for the comparison of two observed maps for different pollen types.

A further modification and development of the method was introduced by Court (1970). For the simplest case of one isopleth, representing the median value of the distribution, areally weighted, the two surfaces were superimposed and the resultant diagram described four regions, table 4.2 :-

		REGION A	
		above median	below median
REGION B	above median	(+ , +)	(- , +)
	below median	(+ , -)	(- , -)

Table 4.2 :- A 2×2 resemblance matrix.

For such a 2×2 resemblance matrix, the coefficient of medial correlation q_m is defined to be:-

$$q_m = \frac{\sum(\text{like area}) - \sum(\text{unlike area})}{\sum \text{area}}$$

This coefficient has three advantages over C_a, the coefficient of areal correspondence, first its limits are -1 to +1, with a perfect negative areal correspondence, giving a value of -1. Secondly, the method may be extended to deal with an even number of classes. According to Court (1970) the sampling distribution of this coefficient is roughly normal hence statistical significance may be tested, although care should be taken due to the presence of spatial auto-correlation.

Hugg (1979) used this method to compare the geographic distribution of work disability and poverty status for persons aged 18-64 using the fifty states of the United States as units of observations.

§4.3.3 Correlation coefficient

An approach suggested by Robinson and Bryson (1957) and Mirchink and Bukhartsev (1959) was the use of Pearson's product moment correlation coefficient:-

$$r = \frac{\sum_{i=1}^{n} (z_{1i} - \bar{z}_1)(z_{2i} - \bar{z}_2)}{n s_{z1} s_{z2}}$$

$$\text{where} \quad s_{zj} = \text{sqrt}\left[\frac{\sum (z_{ji} - \bar{z}_j)^2}{n}\right]$$

This coefficient is equally applicable for interval or ratio scattered data. Pyle (1973) compared census tract maps of measles incidence in Akron, Ohio for 1970-1971 against various demographic and socio-economic variables. Ecological correlation was also used by Gesler et al. (1980) to compare maps of community characteristics to disease reporting and hospital use by census tract in Central Harlem Health District, New York City.

A non-parametric form of the correlation coefficient, Goodman and Kruskal's gamma, Goodman and Kruskal (1954), was also used by Webb et al. (1987) to assess the association between the pollen-fossil maps. The level of agreement or disagreement between the observed and simulated fossil-pollen maps was assessed by mapping the results on a large spatial scale. The authors agreed that although some spatial smoothing was introduced and the pollen counts were recorded for only four levels, the resultant effect was desirable since it reduced palynological noise, i.e. small scale spatial and temporal variability and secondly it reduced the scale discrepancy between the coarse spatial scale of the model and the generally finer scale of the data. An ordered contingency table based on these levels was constructed using the interpolated values from the contouring step. The coefficient values were calculated:-

$$\gamma = \frac{p - q}{p + q} \qquad 0 \leq \gamma \leq 1$$

where p = number of concordant pairs of observations
q = number of discordant pairs of observations

This coefficient has the advantage of having a direct probabilistic interpretation, it is the difference in probability of like rather than unlike orders for the two variables when two individuals are selected at random. γ takes the value one when the data are concentrated in the upper-left to lower-right diagonal and the value zero in the case of independence, however zero need not imply independence. This coefficient has similar drawbacks/limitations in its use and interpretation as the PCC described earlier. A procedure for formal testing also requires a reference distribution.

Tests of statistical significance may be performed to test whether the correlation coefficient is significant, these are fundamentally incorrect since various assumptions inherent to such a test cannot be satisfied in such a situation, i.e. data points are rarely independently distributed, and distributions to be compared will invariably demonstrate significant spatial autocorrelation. The correlation coefficient as a measure of similarity is appropriate as it stands but further inferential analysis should be avoided.

In many situations, subtle changes in the distribution of a variable will occur over a time span, a correlation analysis will potentially report a high level of correspondence, but in reality it imparts no information on the complexities of these changes.

§4.3.4 Comparison analysis of trend maps

Merriam and Sneath (1966) developed a very simple procedure which allowed trend surfaces to be compared. Data points which represent a surface are not independent hence they believed it may be more prudent to estimate the similarity between surfaces in terms of the coefficients of the trend.

A model for surface i, assuming a linear trend is given by :-

$$z_{ij} = \beta_{00} + \beta_{10}x_{ij} + \beta_{01}y_{ij} + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N(0, \sigma^2)$$

where z_{ij} = value of dependent variable at point j, for surface i.
(x_{ij}, y_{ij}) = spatial location of dependent variable j, for surface i.

The values of the coefficients of the fitted surface may then be used as mathematical descriptors of the observed surface. In practice the base is excluded. The similarity between the surfaces may be expressed in one of two ways :-

1. A correlation coefficient

$$r = \frac{\text{cov}\beta_{ik}}{\sqrt{\text{var}\beta_i \, \text{var}\beta_k}}$$

where

$\text{cov}\beta_{ik}$ = covariance between the coefficients for surface i and k.

$\text{var}\beta_i$ = variance of the coefficients for surface i.

Instead of comparing observations x_1 and x_2, the coefficients (β's) of the equations for trend surface i and trend surface k are compared.

2. The taxonomic distance

$$d_{ik} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(\beta_{ji} - \beta_{jk})^2}$$

where $j = 1, s$ where s = order of the trend surface

i, k = surface of interest

This measure is the square root of the mean of the squared differences between equivalent coefficients for the same order surfaces to be compared. The taxonomic distance is no more efficient than the correlation coefficient, it may be easier to interpret since it is always positive and is not constrained to values less than one. Identical trend surfaces give taxonomic distances of zero, with increasingly dissimilar trend surfaces having increasingly greater taxonomic distances between them.

For trend surfaces of order s, the higher order polynomial terms will generally take small values, to use the values as they stand amounts to estimating the variance of the differences corrected for height. In order to ensure all the terms make an equal contribution to the overall similarity measure, the coefficients are normalised prior to

calculating either of the measures. The implementation of this method requires that the surfaces to be compared are of the same order.

Classification or comparison of regional trend surfaces by grouping on the basis of the calculated coefficients of the polynomial terms is restricted by two factors. First, the trend surface parameters are not invariant under certain changes of scale and orientation of the co-ordinate system and secondly, the estimators are still correlated. Miesch and Conner (1967) showed that shifting the origin of the co-ordinate system changed both the values of the estimated coefficients and the percentage explanation of individual terms, the overall percentage explanation by terms for a given order remains unchanged. The absolute values of the regression coefficients, will occasionally be subject to extreme fluctuations in the higher powers as a result of machine rounding and truncation. Variation will also arise within the data due to measurement error.

These factors all serve to restrict the applicability of this method of comparison to those situations where a comparison is to be effected between quadrat systems of similar size and shape. An example of its implementation is given in chapter 7 where the surface temperature for two years are compared for the contiguous United States of America.

§4.3.5 Difference maps

The idea of an isopach/residual map has been extended and formalised to situations where the observations have not necessarily been recorded at the same locality. In such cases an intermediate grid is required to be evaluated before the difference between the two matrices of grid values is taken, the resultant matrix is then contoured.

The most complicated scenario is where interest is in two different variables each of which has been recorded at differing locations. Davies (1973) believes that to directly compare two maps under these conditions, one must be expressed in terms of units of the other or alternatively, both converted to standardised unitless forms.

Expressing one variable in terms of the other allows the user to perceive where the mapped variable is 'greater/smaller' than predicted on the basis of the other variable. The ideas of least squares regression, Seber (1977), enables the implementation of this procedure. For the vector of grid values for variables X and Y, we compute;

$$\hat{Y}_i = \beta_0 + \beta_1 X_i$$

Typically linear regression is used, but low order polynomial regression is equally applicable. The result of the above is a vector of predicted values of $\hat{Y}$. The $\hat{Y}$'s are based solely on values of the second variable X, hence a residual surface $(Y_i - \hat{Y}_i)$ may be regarded as a map of the differences between X and Y. No statistical assessment of the regression of Y on X is possible where the two variables were not measured at the same control points, since the regression is based on the estimates of X and Y at the grid points. The difference/residual surface may be displayed in terms of positive and negative residuals.

Problems with the methodology exist:-

1. The estimates $\hat{Y}$ account for only a part of the variation in Y. Unless the correlation of X and Y is 'high' serious errors will be introduced by the substitution of the estimate $\hat{Y}$ for X.

2. It may not be possible to decide which variable should be used as the estimator. If the correlation between X and Y is 'high' the two regression lines will nearly coincide. If the correlation is not pronounced, the two lines may produce radically different results, Mills (1955). In practice, it may be argued that in the absence of a high correlation, it is pointless to compare the two by a predictive model.

The problems inherent to difference maps based on estimated or predicted variables can be avoided if the original two maps are converted to a standardised format. After the data on each map have been standardised, they may be contoured in the conventional manner. However the contour values will be in units of standard deviation above or below the mean. The resultant contour map is liable to contain ambiguous areas, i.e. an expected positive difference can result from subtracting a low positive area from a high positive area, however it is also achievable by subtracting a large negative area from a low positive area. Similar ambiguous cases result for negative differences.

§4.3.6 Pattern of differences

Cliff (1970) pioneered a method based on the analysis of the pattern of differences between two maps. The criteria for executing such a technique was that the variables of interest were measured in the same units and related to the same areas. Cliff then constructed a three-colour map, coloured according to the relation between the variables. The fundamental proposition in Cliff's method was that if the two maps did not differ significantly, in a spatial sense, then the distribution of the three colours will not be significantly different from random. On the other hand, any spatial pattern in the colours is indicative of some unknown spatial process. The test statistic is a simple joins-count approach:-

$$z = \frac{\text{observed number of joins - expected number given by a random process}}{\text{standard deviation of the expected values}}$$

Expressions for the expected number and standard deviation of joins of both the same color and different colors can be found in Cliff's (1970) work.

§4.4 Local Analysis

The final category of surface comparison techniques relate to those methods which fall under the heading of local analysis. The term local analysis refers to those techniques where specific aspects of the spatial distribution are examined, in terms of a contoured map, this may be the 75^{th} percentile, for example.

One feature of many of these methods is that invariably it is convenient to regard the underlying qualitative variation as a multi-colour pattern or, where appropriate as a two-colour black-white pattern. In general the techniques in this group are computationally more intensive and specialised.

§4.4.1 Complexity index

Complex spatial geologic patterns may be regarded as realisations of random processes, Dacey (1964), Watson (1971), Matern (1960). The estimated parameters of such processes serve as convenient summary characteristics of the observed spatial patterns

e.g. patchiness and prevalence and provide a basis for their classification and comparison, Switzer (1973).

These ideas formed the basis of an approach which attempts to describe the level of pattern complexity of a map. By complexity, Switzer meant the spatial scale of variation. A pattern that has a self-contained area is less complex than the same proportion of area distributed in many scattered smaller areas. This notion of complexity as a scale measurement may also be viewed as a measure of patchiness of the pattern. One intuitive index cited was :-

$$X = \frac{\text{total length of boundaries}}{(\text{area of region})^{1/2}}$$

The larger the value of X, the more complex the pattern. X is also invariant to the choice of measurement unit. We shall need some convention on how to measure boundary length, which is assumed to be finite, section 3.8.

It can be shown, Matern (1960), that if the pattern is regarded as a realisation of a random process, then :-

$$\text{mean of } X = \frac{1}{2}\pi(\text{area of region})^{1/2} Q'(0)$$

$Q'(0)$ = derivative of $Q(d)$ at $d=0$

$Q(d)$ = probability that two colours distance d apart are of different colours.

Because we are concerned here about the estimation of 'pattern properties', from discretely spaced data, we may wish to know how the complexity parameter, X, might be estimated from a square grid, say. Basically we require to estimate Q'(0). Switzer proposes one method for evaluating Q'(0) for a square grid. By altering the sample space and shape of the grid and the sampling density, the complexity index, X, will be changed.

§4.4.2 Image registration

In image-analysis where the analysis of two or more images of the same scene is to be undertaken, registration is required. Image registration is the process of determining the position of corresponding points in two images of the same scene. If the difference between the images is any combination of translation, rotation and scaling then by determining the positions of a minimum of two corresponding points, control points, in the image, the images may be registered.

Extraction of control points is an almost impossible task. If it is possible to find straight lines within the image, the intersection of the lines produce control points, Stockman et al. (1982), alternatively the image can be segmented and closed-boundary regions defined within the image, the centre of gravity of these regions will then produce control points. Goshtasby et al (1986) used the idea of centres of gravity of closed boundary regions as control points. Various segmentation techniques are available, the main objective of these is to produce a desirable number of closed boundary regions within the image.

A point pattern matching technique is required to establish correspondence between control points in an image. One method is to match point patterns by a clustering approach, Stockman et al (1982). Using the clustering technique, matching is carried out between all possible pairs of points in the two sets. When matching point pairs, the translational, rotational and scaling differences between them are determined and a point entered into a parameter space showing the parameter values. Correct matches tend to make a cluster whilst mismatches randomly fill the parameter space. The parameter values corresponding to the most dense cluster are used to map one set to another and determine the correspondence between the two sets of points.

By knowing corresponding control points in the images, corresponding regions may be identified. For comparison of two regions in image analysis, it is desirable to refine the regions to become as similar as possible. If two corresponding regions are more similar, it is anticipated their centres of gravity correspond to each other more closely. Region similarity is then obtained by comparing the shapes of the segmented regions. Since the images have translational, rotational and scaling differences, the shape measures must be invariant to these transformations.

Some of the techniques suggested by Goshtasby et al. for defining shape similarity are Fourier descriptors, shape signatures, centroidal profiles, invariant moments and shape matrices. In terms of the latter, the shape is transformed into a binary matrix by polar quantization of the shape. The zeros and ones in the matrix show points that belong to the outside and inside of the shape respectively. The dimensions of the matrix that determine the quantization steps are determined by the user. The fewer the number of ones in the obtained matrix, the more similar the shapes. By choosing differing dimensions the results will not necessarily exhibit robustness.

In reality we are not always interested in removing the scaling difference, rotational and translational changes since it is these factors which give us a handle on the possible sources of the underlying processes which are responsible for the change.

§4.4.3. Image restoration

A further set of spatial comparison techniques contained within the image processing literature have arisen through the assessment of image processing algorithms for the restoration of images. A number of numerical measures are available for quantifying the discrepancy or 'distance' between two images:-

1. The distance between two grey scale images x and y, measured by the root mean squared difference between corresponding pixel values is given by :-

$$\left[\frac{1}{N}\sum_{t\in T}(x(t) - y(t))^2\right]^{\frac{1}{2}}$$

where x(t) denotes the brightness value of image x at pixel t. This is an example of an L^2 metric. It has many mathematical advantages and is the basis of the optimal linear (Wiener) filtering theory, Hamming (1983). A further example using distance between corresponding pixel values is that of:-

$$\frac{1}{N}\sum_{t\in T}|x(t) - y(t)|$$

2. In classification problems, where the pixel values are class labels, image distances can be measured by the pixel disagreement rate i.e. the proportion of pixels given conflicting class labels:-

$$\frac{1}{N}\text{number}\{t \in T:\ x(t) \neq y(t)\}$$

These two measures involve a pixel by pixel comparison, although widely used, these measures are generally recognised as unsatisfactory, Besag (1986), since they ignore the spatial context and are inadequate in expressing human perceptions of similarity.

Baddeley (1987) has suggested a modification which attempts to combine the ideas of the L^2 metric with those of the Hausdorff metric. The Hausdorff distance between two sets of pixels $X, Y \subseteq T$ is:-

$$H(X,Y) = \max\left\{ \sup_{t \in X} d(t,Y) \ ,\ \sup_{t \in Y} d(t,X) \right\}$$

where d(t,X) is the shortest distance from a pixel $t \in T$ to a subset $X \subseteq T$

$$d(t,X) = \inf_{s \in X} d_t(s,t)$$

i.e. H(X,Y) is the largest distance from a point in one set to the nearest point in the other set.

A major drawback of the Hausdorff metric is its susceptibility to outliers and its lack of robustness. The metric is equal to the distance from one white pixel, say, to the remaining white pixels hence the alteration to the value of a single pixel can markedly affect the value of the metric.

For $\lambda > 0$, the λ metric between images x,y, defined by Baddeley, is denoted as :-

$$\Delta_\lambda(x,y) = \sup_{t \in T} \max\left\{\delta_{\lambda,x,y}(t); \delta_{\lambda,y,x}(t)\right\}$$

where for each $t \in T$

$$\delta_{\lambda,x,y}(t) = \inf\{a > 0 : \exists\, s \in T, d_T(s,t) < a\lambda, d_v(x(t),y(s)) < a\}$$

In other words, two images are closer than a units in the metric if, for every pixel in the x-image, there is a pixel in the y-image less than $a\lambda$ units away, with a brightness value differing by less than a and vice versa. Intuitively λ is the 'rate of exchange' between errors in pixel brightness and errors in pixel distance.

The L^p metrics compare the image brightness values of x and y at the same pixel position, t, i.e. a 'vertical' comparison whilst the Generalised Hausdorff metric basically performs a lateral comparison since H(X,Y) equals the maximum distance from a pixel in one image to the nearest pixel with the same value in the other image. Finally the λ-metric equals the maximum height of a rectangle needed to touch the graph of y from any point on the graph of x and vice-versa and is effectively a trade-off between vertical and lateral comparisons.

§4.4.3 Shape change

Scientists and philosophers in many disciplines have long recognised the potential for gaining insight into spatial processes by studying spatial form. Study of the distinctive shape of a distribution occupied distinguished biologists like Thomson (1917), who saw in changing forms clues to biological growth and evolution. Scholars in all disciplines who have attempted to formalise process-form arguments have found it necessary to devise adequate methods for describing shape.

If shape is stable through time the effect of the conflicting forces has been resolved and an equilibrium state has been achieved. However continued growth or atrophy is indicative that such a state has not yet been obtained; the forces operative on the object still remain unresolved. If one considers processes to be a collective term for all the unresolved forces continuing to shape the object in question, studies of form and process are legitimate.

Generally shape is defined to be the set of properties possessed by any closed figure of at least two-dimensions which has a planar representation, and which possesses precise boundaries i.e. form with size removed.

The statistical analysis of shape data has a vast range of applications in biology, archaeology, geography and chemistry for example. Two main classes of shape analysis exist, those relating to outline data and secondly, those methods based on landmark data.

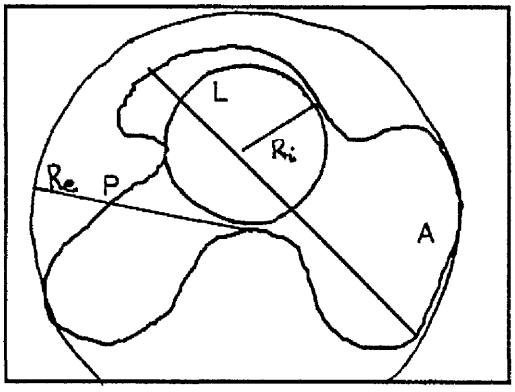
§4.4.4.1 Outline data

Shape factors are very simple to derive computationally and require only basic information about the shape. In trying to measure shape, two questions must be asked:-

1. What characteristics do we measure?
2. How do we combine them into an effective index?

In addressing the first, we assume that the shape of each spatially discrete area is being studied separately. The attributes most commonly measured are area, perimeter, major axis, radii of internal and external enclosing circles, figure 4.3. In order of increasing dimensional scale these are :-

1. Points within the closed figure e.g. centre of gravity.
2. Lines within the closed figure e.g. perimeter.
3. Area of the closed figure.



KEY

A	Area
P	Perimeter
L	Longest axis
R_i	Radii of internal enclosing circle.
R_e	Radii of external enclosing circle.

Figure 4.3:- Commonly measured shape descriptors.

For all three situations, many different 'shapes' will share similar numerical values. The number of indices based on these primary measures is very large, Boots and Lamoureaux (1972). Table 4.3 describes some of the more commonly cited measures for describing shape, particularly in geographic applications.

Measure	Mathematical formulation	Reference
Elongation ratio	L/L' L'- length minor axis	Werrity (1969)
Form ratio	A/L^2	Horton (1932)
Circularity ratio	$(4A)P^2$	Miller (1953)
Compactness ratio	$(2\sqrt{\pi A})/P$ A/A' A' - area of smallest enclosing circle $1.273A/L^2$	Richardson (1961) Cole (1964) Gibbs (1961)
Ellipticity index	$L/2\{A/[\pi(L/2)]\}$	Stoddart (1965)
Radial shape index	$\sum_{i=1}^{n}\left\lvert(100d_i/\sum_{i=1}^{n}d_i)-(100/n)\right\rvert$ d_i = radial distance from a point to the circumference of the circle.	Boyce and Clarke (1964)

Table 4.3. :- Elementary measures for measuring the shape of geographic areas.

An interesting shape measure was described by Young et al. (1974), to circumvent some of the potential problems associated with the above indices i.e. lack of robustness and to a lesser extent, the possibility that some of the properties may be altered in the transition from analysis of shapes of a continuous form to analysis on a discrete grid.

The measure described by Young et al. was based on the notion of bending energy. They suggested a two-dimensional outline made out of a homogeneous material, if allowed to adopt its 'free form', would assume the shape of a circle since a circle minimises stored energy. More convoluted outlines require additional work in the form of bending energy. The curvature calculation is simple. The shape is divided into n small regions and the curvature K_n is defined to be :-

The total 'bending energy' is given by $\sum_{i=1}^{n} K_i^2$ over the whole region. The measure is invariant to position and rotation but is affected by size as well as shape differences.

Other techniques based on the definition of the curvature at sample points on an outline, or within a surface, have been used to compare biological forms. One method is to describe the continuous curvature at sample points on the outline. The outline can then be considered to be a continuous curve. Implicit in this function, describing the curve, are the actual Euclidean locations of any sample points that are required, (each sample point can be described by its tangent angle and arc-length from an arbitrary start point).

The local curvature at any point on the outline of a shape can be calculated from the 'chain-code' directions from pixel to pixel, of an outline. By repeated averaging of adjacent curvatures, a smooth graph can be drawn. Graphs of different forms may then be compared.

Bookstein has suggested it is possible to compare forms by sampling the tangent angle function, arc-length and tangent angle at landmarks and analysing these values by a multivariate statistical technique.

A mathematical technique used to describe the entire facial surface shape and the changes occurring in the face has been developed based on a classification system inaugurated by Besel and Jain (1988). By decomposing the surface into fundamental shape patches, an objective, quantitative and qualitative description of the face can be produced. Each surface point on the face is classified as belonging to one of eight surface types by computing values of the Gaussian and mean curvature, Coombes et al (1991). Gaussian curvature is a measure of the curvature at a point on the space surface, given as a ratio of the discriminant of the two fundamental forms of the surface, the first describes the metric and the arc length and the second, defines the direction cosine of the normal to the surface. Points on the surface may then be classified as flat, elliptic, parabolic or hyperbolic. The mean curvature is the sum of the principal curvatures. These two curvatures are independent and both are needed to describe a surface unambiguously.

The advantage of this method of describing the surface is that it is independent of orientation, rotation and displacement. Thus the description of the face will be the same from any viewpoint.

In order to produce a classification, the signs of the Gaussian and mean curvatures are used. These are computed by passing a local neighbourhood operator over a depth map. The points on the face may then be colour-coded according to the surface type to which they belong and a 'surface type' image may be produced. As all data has some random variation, it is necessary to set thresholds on the curvatures, below which the surface is classified as flat.

For the clinician, the advantage of such a method lies in the fact that the facial surface is described quantitatively but at the same time, retains a reasonable amount of descriptiveness, allowing rapid appreciation of the differences between surfaces.

§4.4.4.2 Landmark data

Morphometrics, the study of geometrical form of organisms, combines themes from biology, geometry and statistics. Data for morphometric studies usually include geometric locations of landmarks i.e. points that correspond biologically from form to form.

There are two general methods for characterising landmarks, firstly those which locate landmarks by the juxtaposition of different identifiable structures e.g. in fish, the 'anterior fin base' and 'posterior fin base' delimit the body outline. Secondly, those located using geometric properties e.g. the tip of a tooth may be taken at the point where the curvature of the edge is greatest. Landmark locations may be augmented by information about the curving of external or internal boundaries between landmarks.

Two approaches to the investigation of group differences/associations in size and/or shape or between size change and/or shape change are :-

1. Multivariate morphometrics.
2. Deformation analysis.

In multivariate morphometrics, configurations of landmarks are measured one at a time in collections of 'morphometric variables'. Some of these measures are size variables e.g. distances between landmarks, whilst other variables have a value independent of geometric scale, such as ratios of distances or other shape variables, as well as functional transforms of these ratios.

Generally, Bookstein (1986) concluded that although 'size' and 'shape' are verbally orthogonal, computationally and conceptually they are inextricably entangled.

The second morphometric tradition concentrated on the theme of deformation. This was introduced into descriptive biology by Thompson (1961) under the heading of 'Cartesian Transformation'. A deformation is a mapping which takes neighbouring points to neighbouring points and which alters lengths of short segments by factors which never get too large or too small. The notion is an informal version of what mathematicians call a diffeomorphism: e.g. the reals and the interval $(0,\infty)$ are diffeomorphically equivalent, since the diffeomorphism

$$f: \Re \rightarrow (0,\infty) : f(x) = e^x$$
$$g: (0,\infty) \rightarrow \Re : g(x) = \log x$$

i.e. a one-to-one transformation which, along with its inverse has a derivative at every point of a region and its image.

Most techniques for the geometric study of mappings, choose to model the configuration of landmarks by a map from some algebraic simple family, and then interpret either the coefficients of the fitted map or else, its distributed 'error of fit'.

Sneath (1967) expresses the Cartesian co-ordinates of the landmarks of one form by a cubic bivariate polynomial in the x and y ordinates of the same landmarks in another form. The resultant coefficients are impossible to interpret directly. Sneath's purpose was instead to summarise them in a single net measure of dissimilarity between forms.

In general these analysis suffer from the dependence on landmarks being readily and unambiguously identifiable. On gently curving surfaces such as the human back or face this is not the case, hence it serves as a constraint on the accuracy of the technique.

Many papers have since suggested methods for the analyses of shape in two-dimensions and almost all have been based on the movement of homologous landmarks, Bookstein (1978, 1984 a,b,1986), Siegel and Benson (1982).

A whole wealth of other techniques for describing shape and hence allowing comparisons to be effected are summarised in O'Higgens and Johnson (1985) and include medial transforms Blum (1973), and fourier analysis, Erlich et al. (1983).

§4.5 Summary

The advancement of technology has enabled spatial comparisons to move in leaps and bounds over the past two decades. Prior to this time, techniques available were simplistic and the numerical measures defined were invariably difficult to interpret and information was ignored.

Major strides have since taken place especially within the field of image analysis with research initially focusing on the development of good restoration algorithms. A by-product has been an attempt to develop statistics which express the deviance from the 'true' picture and hence allow the performance of algorithms to be assessed and compared. Ideas based on distance metrics have been the main contributors in this area.

Within the field of medical imagery, particularly reconstructive surgery, interest has focused on change through surgery or growth. The use of landmarks has been paramount in this field but the question of selection of such points has raised a number of questions as to the suitability and robustness of this approach. Other methods such as the idea of curvature have been suggested to describe a surface, however for comparison a subjective-based ordering mechanism is used to discriminate between images.

Much work has been done in the field of shape analysis to describe surfaces as diverse as drainage basins, central business districts and human faces. These measures have been used to assess similarities/differences between the variables of interest. The major feature common to all these methods is that they are invariant to translational, rotational and scaling differences. In the field of environmental sciences this aspect of change potentially enables the scientist to diagnose the process which is responsible for the change. The next two chapters will concentrate on the development of test statistics

based on these three modes of transformation to examine the question of change or similarity.

CHAPTER 5

THE CHARACTERISATION OF SPATIAL CHANGE

§5.1 Introduction

Within this chapter, the framework of the methodology used to describe change between two spatial processes represented as contoured surfaces, is developed. The contoured surface being produced as described in chapter 3.9. Although the procedure is explicitly formulated in terms of expressing change, it is equally applicable for investigating associations between spatial processes.

Subjectively change may be described by evaluating those features of the contoured surface which the eye perceives as having changed when the two surfaces are superimposed. These changes will invariably be summarised in terms of one or all of the three transformations:-

1. Scalar
2. Translation
3. Rotation

Figure 5.1 illustrates five simulated examples of two contours where various transformations have been imposed. The first shows two identical contours i.e. the shape, size and spatial location are coincident. Figure 5.1(b) is an example of two contours relating in spatial locality but differing in size. A change in the orientation of one of the contours has occurred due to some external process in figure 5.1(c). The penultimate diagram describes the situation where contour A has been displaced with respect to contour B. Finally figure 5.1(e), illustrates the most complicated scenario where translational, rotational, scalar and shape change have all resulted. The value of undertaking a statistical procedure to assess change of this magnitude is questionable due the primary differences which exist between the two surfaces.

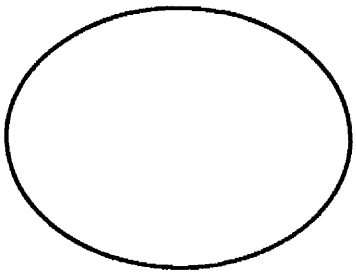


Figure 51(a) :- No change.

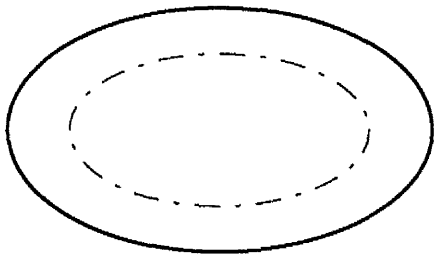


Figure 5.1(b) :- Scalar change.

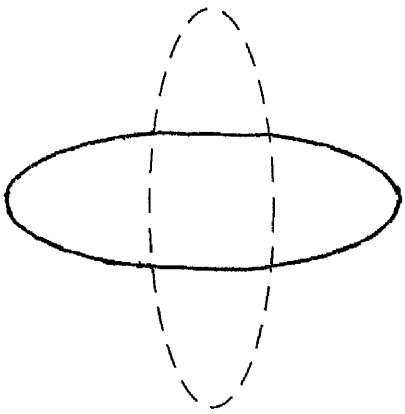


Figure 5.1(c) :- Angular change.

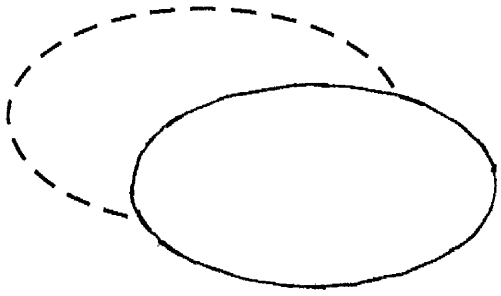


Figure 5.1(d) :- Location shift.

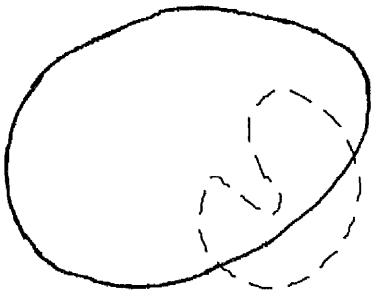


Figure 5.1 (e) :- Complex change.

KEY

SURFACE A

SURFACE B

Figure 5.1 :- Illustration of the three transformations.

Three stages were required for the formalisation of a set of test statistics which would allow the various transformations to be quantified in terms of potential change statistics :-

1. The definition of various geometric properties of a contour i.e. area, perimeter, centre of gravity and orientation. A simulation study was performed to examine how the variability of these measures was affected by changing the fineness of the underlying mesh.

2. Based on these quantities, potential statistics were examined to assess which enabled the three transformations, scalar change, rotation and translation to be quantified satisfactorily. A whole series of plausible test statistics can be listed, ranging from those involving ratios or differences to those which are specific to a particular area of application. The method of assessment for each of the proposed test statistics was based on the following criteria :-

1. Comparability of results.
2. Bounds of $-\infty$ and $+\infty$ are unattainable in practice.
3. Robustness, i.e. ordering of the surfaces is irrelevant.

The first and third conditions are of particular relevance in constructing a statistical distribution since the resultant distribution should be globally applicable.

3. Those measures deemed to describe most satisfactorily the three transformations from part 2 were then formalised. These provided the basis of the test statistics for describing change and their distributions are derived in chapter 6.

§5.2. Various Contour Descriptors For Characterising Transformation Parameters

§5.2.1 Scale

The most common form of transformation liable to be found in the environment is scalar change i.e. expansion/shrinkage of a process. Examples include the decay of a radionuclide and the deforestation of an area.

The simplest criteria for describing scale is in terms of the relative size of the contour. Two measures which describe the specific size of an object are area and perimeter. These two measures quantify different aspects of size; area refers to the part of a two-dimensional surface enclosed within a specified boundary or geometric measure, whilst perimeter relates to the length of a 'curve' enclosing a region of a space. For simple surfaces, such as a circle or rectangle, the two measures are related, figure 5.2.

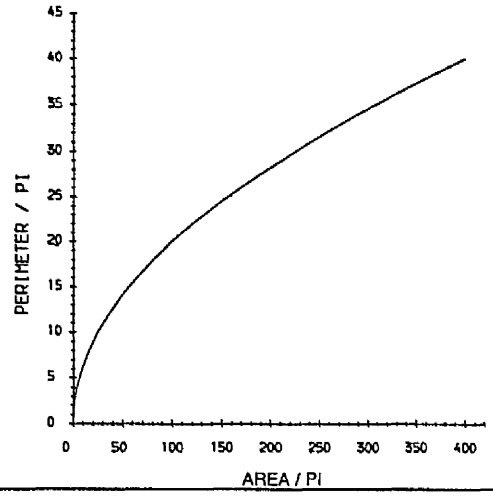


Figure 5.2 (a) Circle

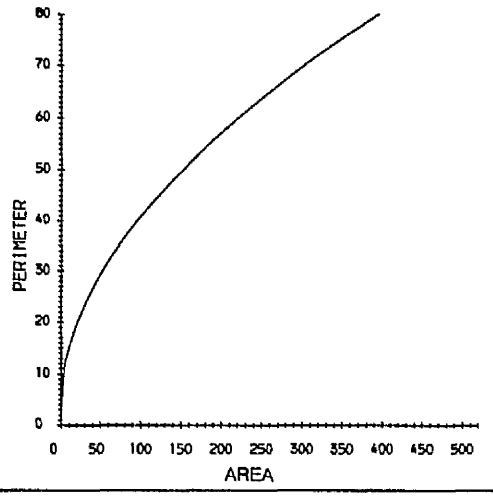


Figure 5.2 (b) Rectangle

Figure 5.2 :- Relationship between area and perimeter for simple constructs.

In practice surfaces will be more complex hence a simple relationship as illustrated above is unlikely to arise, although the assumption of independence may potentially still be violated. This question is investigated further in chapter 6 in terms of a global approach for detecting change.

§5.2.1.1 Area

The evaluation of an areal quantity is simply achieved.

Let A_{ij} = area of contour j, in surface i

where $A_{ij} = \iint_{C_{ij}} dxdy$ C_{ij} - region of interest

The analytical evaluation of the region of interest will invariably prove to be impossible to achieve apart from for simple regions. A numerical procedure was therefore required for quantifying area.

The area of any polygon represented as a vertex list may be calculated by summing the areas of the trapezia under each side, down to the axis. The direction of the sides must be taken into account, so that sides on the bottom of the polygon are subtracted from the total, figure 5.3. Care must be taken to ensure the polygon is stored in an anti-clockwise direction, if stored in a clockwise direction, the absolute value should be taken.

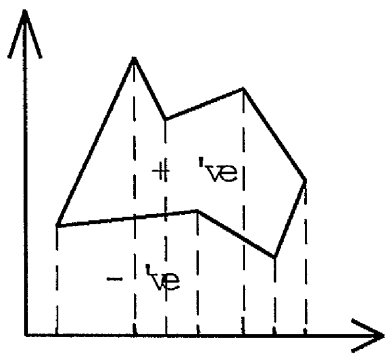


Figure 5.3 :- Evaluation of the area of a polygon based on the trapezia rule.

$$A_{ij} = \frac{1}{2}\left[\left(x_{ij1}y_{ij2} + x_{ij2}y_{ij3} + \ldots + x_{ij(N-1)}y_{ijN} + x_{ijN}y_{ij1}\right) - \left(x_{ij2}y_{ij1} + x_{ij3}y_{ij2} + \ldots + x_{ijN}y_{ij(N-1)} + x_{ij1}y_{ijN}\right)\right]$$

$$= \frac{1}{2}\left\{\left[\sum_{k=1}^{N-1}\left\{\left(x_{ijk}y_{ij(k+1)}\right) - \left(x_{ij(k+1)}y_{ijk}\right)\right\}\right] + \left[\left(x_{ijN}y_{ij1} - x_{ij1}y_{ijN}\right)\right]\right\}$$

The major problem with this approach is that if the polygon is sited a long way from the x-axis, the area of the trapezia will be much larger than the area of the polygon and accuracy will be lost. Temporarily making one vertex the origin will avoid this problem.

§5.2.1.2 Perimeter

A similar formulation for perimeter is possible.

Let P_{ij} = perimeter of contour j, in surface i,

where $$P_{ij} = \sum_{k=1}^{N-1}\left\{\left(x_{ijk} - x_{ij(k+1)}\right)^2 + \left(y_{ijk} - y_{ij(k+1)}\right)^2\right\}^{1/2} + \left\{\left(x_{ijN} - x_{ij1}\right)^2 + \left(y_{ijN} - y_{ij1}\right)^2\right\}^{1/2}$$

§5.2.2 Orientation

One possible definition to be used in assessing rotation is the orientation of a contour i.e the angle subtended by the major axis and the x'-axis. The origin being defined by the ordinates of the contour's centre of gravity, $C(XCG_{ij},YCG_{ij})$. The principal axis, PP', is delineated by the line which describes the maximum distance between two points on the contour's boundary, which lie on a straight line, pass through the centre of gravity and is wholly contained within the bounds of the contour, figure 5.4

The theoretical derivation of the angle of orientation is simply a function of the distances PC and AC. Empirically it is not so straightforward. The crucial step is to identify the principal axis. To simplify the calculations the contour of interest should first be translated so that the centroid becomes (0,0). The orientation of the figure will accordingly be preserved.

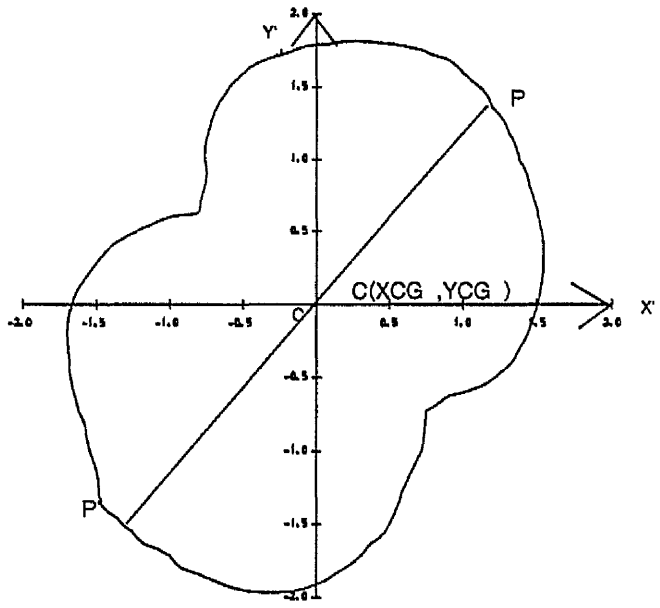


Figure 5.4 :- Definition of the angle of orientation.

One method for locating the principal axis is :-

1. Ascertain whether the mesh points, A, B where A≠B, comprising part of the contour boundary are in diagonally opposite quadrants. Simply checking the following truth statement enables verification of the above statement:-

$$\left[\left(\mathrm{sn}(x_{1jk}).\mathrm{NE}.\mathrm{sn}(x_{2jk})\right).\mathrm{AND}.\left(\mathrm{sn}(y_{1jk}).\mathrm{NE}.\mathrm{sn}(y_{2jk})\right)\right] \Rightarrow \mathrm{TRUTH} \qquad \text{where sn = sign}$$

2. If step 1 is true, then the angles subtended by points A and B with the x-axis are calculated, θ_{ijA}, θ_{ijB}, respectively, where:-

$$\theta_{ijS} = \mathrm{modulus}\left\{\tan^{-1}\left(\frac{y_{ijS}}{x_{ijS}}\right)\right\} \qquad S = A, B.$$

3. If $\theta_{ijA} = \theta_{ijB}$, then the two points lie at the extremes of a straight line which passes through the origin and the distance d_{ijS} may be calculated.

4. Steps 1 to 3 are repeated for S=1,N where N = Number of interpolated contour points.

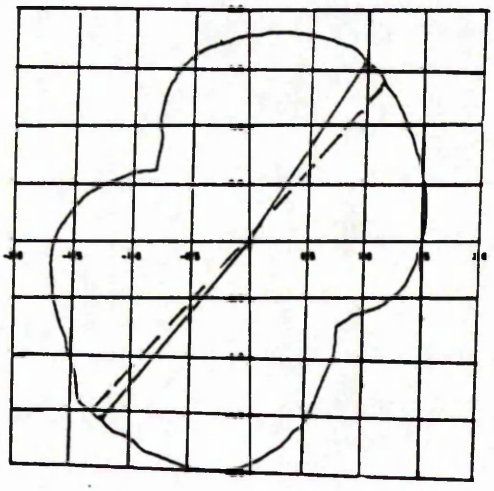
5. The angle of orientation is the angle which subtends the maximal distance between two end points:-

$$OR_j = \left\{\theta_{ijS} / \max(d_{ijS})\right\}$$

A number of problems arise with this technique. Theoretically the test $\theta_{ijA} = \theta_{ijB}$ is valid, but in practice is unworkable. Two reasons for this are:-

1. Inaccuracies due to rounding errors. When working in real space it is seldom possible to evaluate two numbers which coincide to more than two or three decimal points.

2. Theoretically a closed contour is defined to be continuous, in practice the ordinates of a contour's boundary are only known at discrete points, the mesh intersections. The finer the mesh, the more valid the assumption of continuity. This feature complicates the location of the principal axis, figure 5.5. The 'theoretical true' principal axis may not necessarily be described by the defined mesh points, hence physically it will be impossible to locate the true principal axis.



– – – – True principal axis

——— Empirical principal axis

Figure 5.5 :- Problem concerning the theoretical and empirical definition of principal axis.

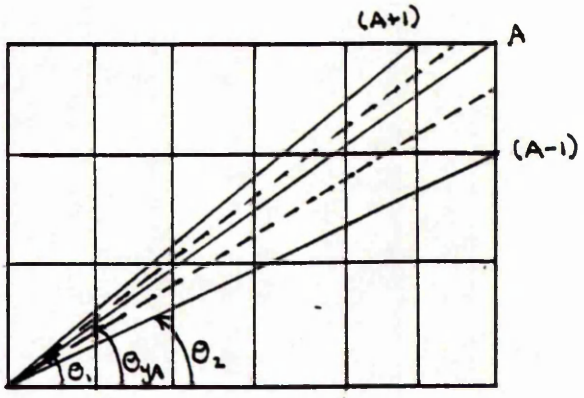


Figure 5.6. Definition of orientation bounds.

The points which lie closest to the 'true' empirical principal axis may not always be strictly diametrically opposite. The simplest method to account for these discrepancies, due to the discretisation of the contour's boundary, is to define bounds for the angle of orientation, figure 5.6 :-

$$\left[\theta_{ijA} - \frac{(\theta_{ijA} - \vartheta_2)}{2}\right] \leq \theta_{ijS} \leq \left[\theta_{ijA} + \frac{\vartheta_1 - \theta_{ijA}}{2}\right]$$

where $$\vartheta_1 = \text{mod ulus}\left\{\tan^{-1}\left(\frac{y_{ij(A+1)}}{x_{ij(A+1)}}\right)\right\}$$

$$\vartheta_2 = \text{mod ulus}\left\{\tan^{-1}\left(\frac{y_{ij(A-1)}}{x_{ij(A-1)}}\right)\right\}$$

By defining the angle of interest in terms of a closed interval, we are theoretically eliminating the problem of discretisation by defining the contour boundary in terms of 'continuous' segments.

Inherent within the definition of the principal axis was the assumption that the axis was contained wholly within the bounds of the contour. The simplest way of verifying this assumption is not violated is to perform the ray test i.e. only one value of d_{ijS} is calculated for each mesh point; where more than one value results, the ray test will be negative and the combination of points do not define the major axis. Computationally, time is increased since it is not possible to stop the procedure once a diametrically point is located, (N-S) calculations have to be performed, for S=1, N.

A final problem is where the contour of interest is almost cyclic in definition. Definition of the angle of orientation will then present problems, unless the cyclic property is recognised at the subjective stage.

§5.2.3. Centre of gravity

One possible measure which may form the basis of translational movement is that of the displacement of the centre of gravity.

Analytically the centre of gravity may be found by:-

$$XCG_{ij} = \frac{1}{M}\iint_{C_{ij}} xf(x,y)dxdy$$

$$YCG_{ij} = \frac{1}{M}\iint_{C_{ij}} yf(x,y)dxdy$$

where $M = \iint_{C_{ij}} f(x,y)dxdy$

C_{ij} - the region of interest.
XCG_{ij} - x-ordinate of the centre of gravity
YCG_{ij} - y-ordinate of the centre of gravity

As for area, it is seldom feasible to define the region C_{ij} mathematically, however a simple numerical procedure will enable the critical point to be located.

$$XCG_{ij} = \frac{1}{3}\left\{\frac{XV}{A_{ij}} + x_{ijN}\right\}$$

$$YCG_{ij} = \frac{1}{3}\left\{\frac{YV}{A_{ij}} + y_{ijN}\right\}$$

where

$$XV = \sum_{k=2}^{N}\left[(x_{ijN}-x_{ijk})(y_{ijk}-y_{ijN})+(x_{ijk}-x_{ijN})(y_{ijk}-y_{ijN})\right]\left\{(x_{ijk}+x_{ij(k-1)})\right\}$$

$$YV = \sum_{k=2}^{N}\left[(x_{ijN}-x_{ijk})(y_{ijk}-y_{ijN})+(x_{ijk}-x_{ijN})(y_{ijk}-y_{ijN})\right]\left\{(y_{ijk}+y_{ij(k-1)})\right\}$$

A_{ij} = area for contour j in surface i (§5.2.1.1)

§5.3 Variability Of The Contour Descriptors

Section 3.8 mentioned the importance of mesh resolution in influencing the results for area and perimeter for a very simple example. In this section, the level of variation

attributable to different levels of grid resolution, in each of the four primary statistics is examined. The reason for such an investigation is that for comparative work it is essential the level of noise resulting from the surface fitting procedure is minimised, and where data sets differ in size, the noise contributions should be of a similar order. A simulation study was undertaken to examine the level of grid resolution which both minimised the error due to the surface fitting technique and secondly did not increase the computational time for minimal gain in accuracy.

§5.3.1 Simulation study

A study was performed on three sizes of data set 50, 100 and 150 points, these were believed to be representative of sizes likely to be encountered in environmental applications. Each set of observations were generated by simulating a bivariate normal distribution, using the NAG (1984) subroutine G05EAF:-

$$f(\underline{x}) = (2\pi)^{-1}|\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu})^T \Sigma^{-1}(\underline{x}-\underline{\mu})\right\}$$

where $\mu = (0,0)^T$

$$\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

The fixed kernel method of density estimation was used for surface evaluation, the kernel function was defined to be that of the standard multivariate normal density function and the smoothing parameter was evaluated using least squares cross-validation.

The measure used to define the levels of variability introduced into the contour descriptors as a result of the mesh resolution was the coefficient of variation :-

$$Cv = \left(\frac{\text{standard deviation}}{\text{mean}}\right) \times 100\%$$

The coefficient of variation is dimensionless hence it enables the relative variability between the data sets to be compared.

The effect of mesh resolution on the contour descriptors was investigated by examining the effect of changing the number of grid points over which the raw data was

interpolated to produce the surface. Based on a square grid, nine sizes of grids were investigated; (22^2), (32^2), (39^2), (45^2), (50^2), (55^2), (63^2), (67^2) and (71^2). The theoretical distribution of each of the primary measures was unknown, hence the standard bootstrap, Efron (1979), was used to evaluate the coefficient of variation. The bootstrap is an easily implemented although computationally intensive device which allows the sampling distribution of some statistic to be estimated; this parametric statistic frequently being a parameter estimate.

Suppose for a sample of N independent observations $x_1, x_2, x_3, \cdots, x_N$, we are required to estimate a parameter $\hat{\theta}$. The underlying distribution, F, of the observations is unknown but it is assumed to be both unimodal and well behaved. Let F_n be the empirical probability distribution of F, having mass 1/N at each observed x_i (i = 1,.....,N). A random sample, herewith known as a bootstrap sample is drawn with replacement from F_n;

$$x_1^*, x_2^*, \ldots, x_N^* \sim F_n \qquad (1)$$

and the characteristic of interest $\hat{\theta}^*$ is calculated;

$$\text{where} \quad \hat{\theta}^* = \hat{\theta}(x_1^*, x_2^*, \ldots, x_N^*) \qquad (2)$$

Steps (1) and (2) are repeated S times to give:-

$$\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_S^*$$

The mean and standard deviation of this sample can then be evaluated;

$$\hat{SD} = \left\{ \frac{1}{S-1} \left\{ \sum_{s=1}^{S} \left(\hat{\theta}_s^* - \bar{\hat{\theta}}^* \right) \right\}^2 \right\}^{1/2}$$

$$\text{where} \qquad \bar{\hat{\theta}}^* = \frac{1}{S} \sum_{s=1}^{S} \hat{\theta}_s^*$$

The more re-samples performed the closer the parameter of interest tends to its true value. In reality the process has to be terminated at some stage, hence a compromise

between time and accuracy has to be reached. The number of replicates performed for this study was 250 since initial results indicated reasonable stability at this level.

The simulation study took the following form:-

1. Simulate data set $x_1, x_2, x_3, \cdots, x_N$ from a standard bivariate normal distribution.

2. Generate bootstrap sample $\underline{x}_1^*, \underline{x}_2^*, \ldots, \underline{x}_N^*$.

3. Calculate the smoothing parameter using least squares cross validation.

4. Construct surface from bootstrap sample.

5. Evaluate area, perimeter, centre of gravity and orientation for the contour depicting the upper quartile.

6. Repeat steps (2) to (5) 250 times.

The usage of the standard bootstrap in step 2 resulted in the contour breaking down into subsidiary components for some combinations of data. The problem was traced to one of the bootstraps peculiar properties. Every value within a bootstrap sample is drawn from the original data with replacement, some values will be repeated several times causing the sample to become over-discretised. In addition the surface constructed using the standard bootstrap appeared to deviate from the original data, especially for 50 data points.

The simplest remedial action was to replace the standard bootstrap by its smoothed counterpart, samples generated using this technique do not possess this property. Instead of resampling with replacement from the original observations, $x_1, x_2, x_3, \cdots, x_N$, a non-parametric estimate $\hat{f}_h$ of the underlying density is evaluated and resampling is performed from $\hat{f}_h$, the smoothed version. Effectively each point in the new sample is a perturbation of the original data point randomly selected as before i.e. the bootstrap sample is obtained by independently sampling the distribution $\hat{f}_h$ with density:-

$$\hat{f}_h(\underline{t}) = \frac{1}{nh}\sum_{i=1}^{N} K\left\{\frac{1}{h_{bs}}(\underline{t}-\underline{x}_i)\right\}$$

The function K is assumed to be a bounded density function symmetric about zero and with unit variance. The density is a kernel estimator of density, f, of the population F, with smoothing parameter h_{bs}.

The following procedure may then be used for evaluation of a smoothed bootstrap:-

1. Choose I uniformly with replacement from i=1,...,N.

2. Generate $\underline{\varepsilon}$ to have probability density function K.

3. Let $\underline{Y} = \underline{X}_I + h\underline{\varepsilon}$

If realisations $\underline{Y}$ are required to have the same first and second moment properties as those of the observed sample then, by using the following transformation which scales the kernel to have the same variance matrix as the data, this requirement will be satisfied :-

$$\underline{Y} = \overline{\underline{X}} + \frac{\left(\underline{X} - \overline{\underline{X}} + h_{bs}\varepsilon\right)}{\left(1+h_{bs}^2\right)^{1/2}}$$

Hall et al (1989) demonstrated that the benefits to be derived from smoothing will diminish with increasing sample size, as $N \to \infty$, $h_{bs} \to 0$. This theoretical statement confirms the behaviour witnessed in terms of the over discretisation problem which was most noticeable for 50 data points. For 150 points, this problem was less serious.

The effect of using the smoothed version of the bootstrap as opposed to the standard form is expressed diagramatically in figure 5.7. Once again a set of 50 data points were simulated from a bivariate normal distribution, the resultant density being expressed in figure 5.7(a). Figure 5.7(b) is an example of a resample using the standard bootstrap, whilst the third diagram illustrates the smoothed bootstrap.

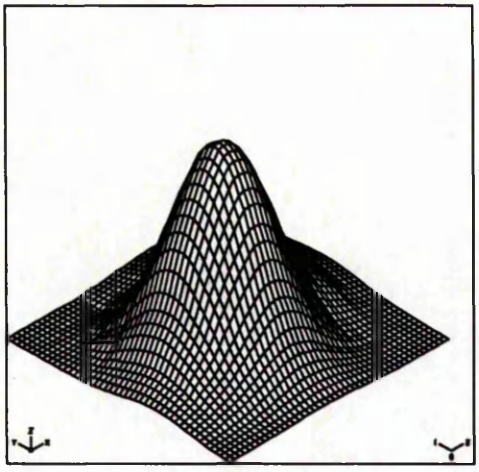


Figure 5.7(a) :- Original data.

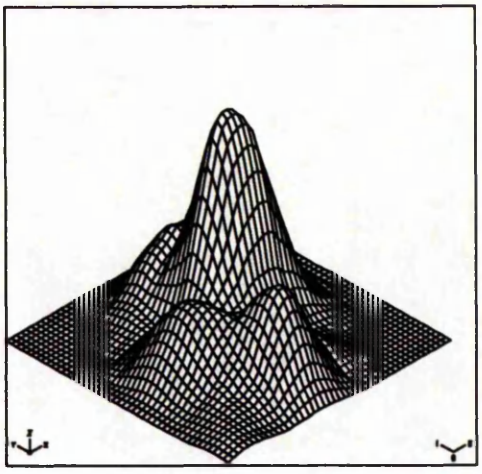


Figure 5.7(b) :- Standard bootstrap.

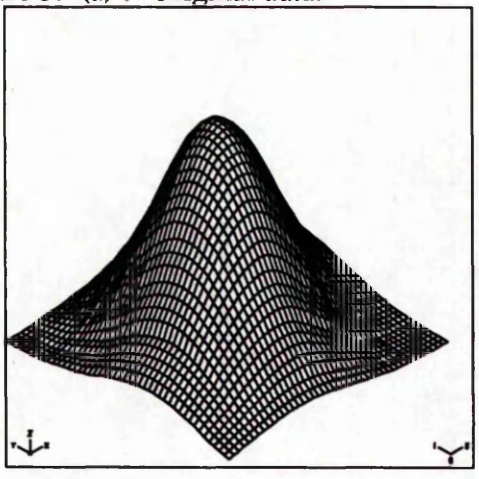


Figure 5.7 (c) :- Smoothed bootstrap.

Figure 5.7 :- Illustration of the use of the bootstrap.

The question of what value of bootstrap smoothing parameter should be used is a research topic in its own right. Within the simulation study a series of values were examined. The sequence for the study was as before but with step two now reading:-

2. Generate smoothed bootstrap data set with smoothing parameter, h_{bs}.

An additional step was added at the end to examine the question of the smoothing parameter for the smoothed bootstrap.

7. Repeat steps (2) to (6) for the range of h_{bs} under consideration.

The range of bootstrap smoothing parameters considered was determined by simulating 200 bivariate normal distributions and evaluating the smoothing parameters for each separate distribution. Based on the evidence of Hall et al (1989), the range for the smoothed bootstrap parameter was taken to be slightly wider than that of the smoothing parameter for the original data. Increments of 0.1 were examined for the smoothed bootstrap parameter. Table 5.1 collates these results.

	Smoothing parameter for the original data		Bootstrap smoothing parameter	
No. of points	h (lower)	h (upper)	hbs (lower)	hbs (upper)
50	0.20	0.65	0.15	0.80
100	0.21	0.57	0.15	0.8
150	0.22	0.55	0.15	0.8

Table 5.1 :- Range of smoothed bootstrap parameters investigated.

§5.3.2 Simulation results

§5.3.2.1. Area

Increasing the size of the data set has negligible effect on the coefficient of variability at each level of grid resolution. However across levels a trend is evident; increasing the grid fineness, the level of variability decreases, figure 5.8. Table 5.2 summarises the results from figure 5.8 for the mean level of the original smoothing parameter, calculated from the original 200 simulations used to produce the ranges for the smoothing parameter in table 5.1.

No. of points	Grid Size									
	22^2	32^2	39^2	45^2	50^2	55^2	59^2	63^2	67^2	71^2
	Coefficient of variation									
50	1.340	0.742	0.561	0.461	0.398	0.340	0.300	0.250	0.278	0.235
100	1.340	0.801	0.616	0.465	0.398	0.324	0.313	0.275	0.250	0.241
150	1.279	0.789	0.545	0.465	0.385	0.335	0.328	0.283	0.275	0.220

Table 5.2 :- Results for coefficient of variation for area.

The table indicates the presence of a strong exponential trend with a tailing off of variability occurring for grid levels of 63 and upwards. In terms of related data sets size, this corresponds to 80, 40 and 26 times the original sizes for 50, 100 and 150 data points, respectively. In real terms variation of the order of approximately 0.3% is introduced when interpolating over a grid of 4000 mesh points.

For 50 data points, a bootstrap smoothing parameter of less than 0.2 produced fairly unstable results, this behaviour was explained by the findings of Hall et al. (1989) who advocated a suitable value for the smoothed bootstrap should be calculated using the original data set, the lower bound for 50 data points, from our simulations, table 5.1, being 0.2 hence the instability. For values greater than 0.2, a remarkable degree of consistency was established across all values of h_{bs}.

If interest is solely in area, then a mesh of 4000 points and a value for h_{bs} selected equal to that calculated for the original data will produce a high degree of accord across surfaces. Variability of the 0.3% will be introduced into any ensuing analysis. This will invariably be masked by the other forms of error.

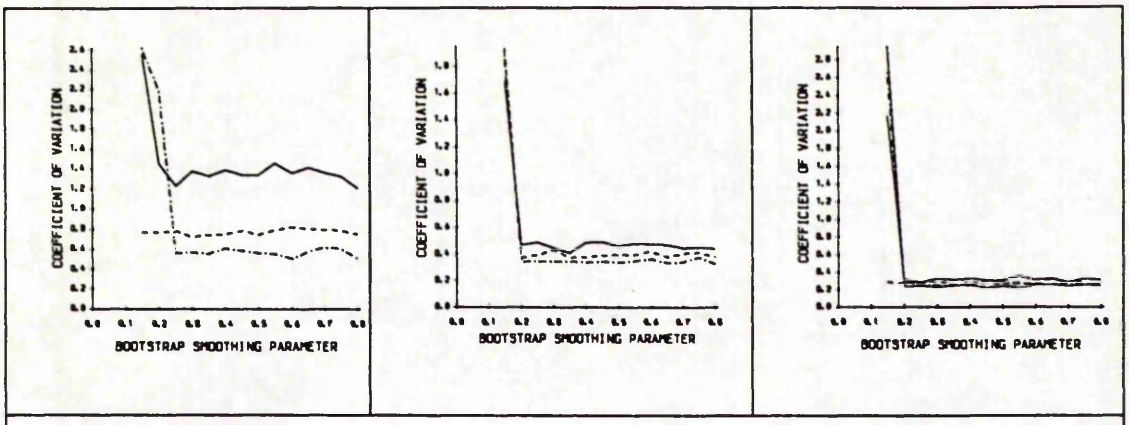


Figure 5.8 (a) :- Coefficient of variation for data set comprising 50 points.

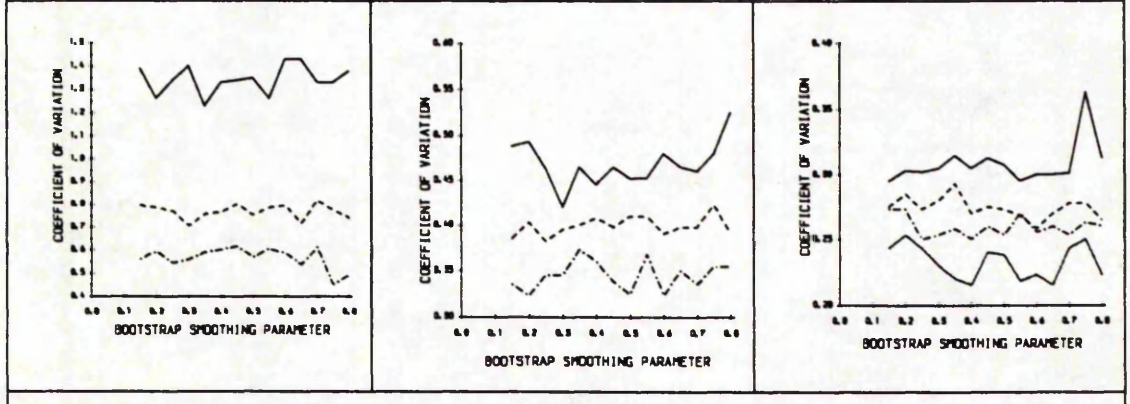


Figure 5.8 (b) :- Coefficient of variation for data set comprising 100 points.

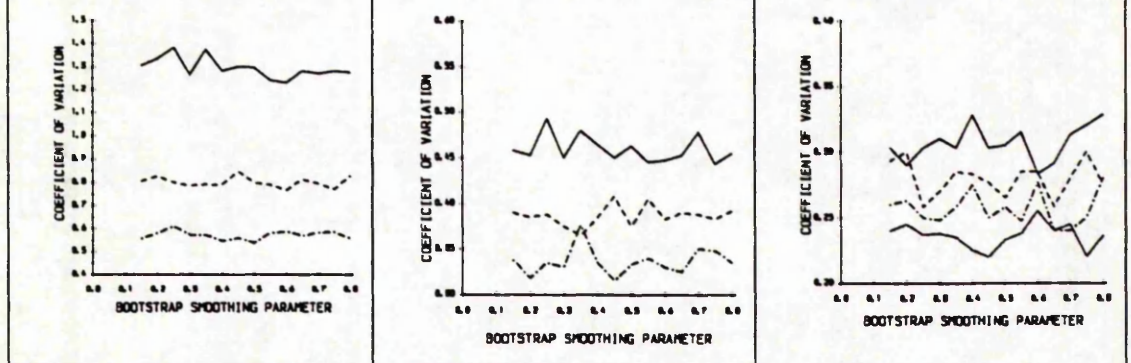


Figure 5.8 (c) :- Coefficient of variation for data set comprising 150 points.

KEY	KEY	KEY
——— 22 × 22	——— 45 × 45	——— 59 × 59
– – – – 32 × 32	– – – – 50 × 50	– – – – 63 × 63
·-·-·-·-· 39× 39	·-·-·-·-· 55 × 55	·-·-·-·-· 67 × 67
		·-·-·-·-· 71 × 71

Figure 5.8 :- Area results for coefficient of variation.

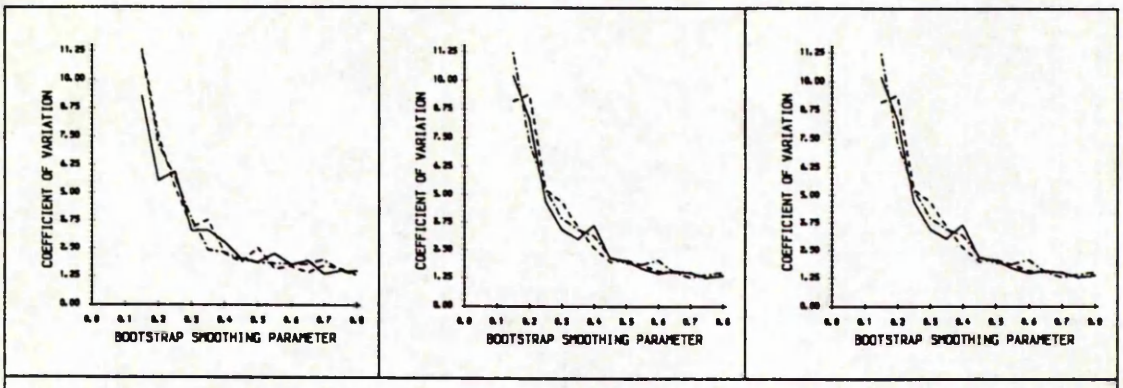


Figure 5.9 (a) :- Coefficient of variation for data sets comprising 50 points.

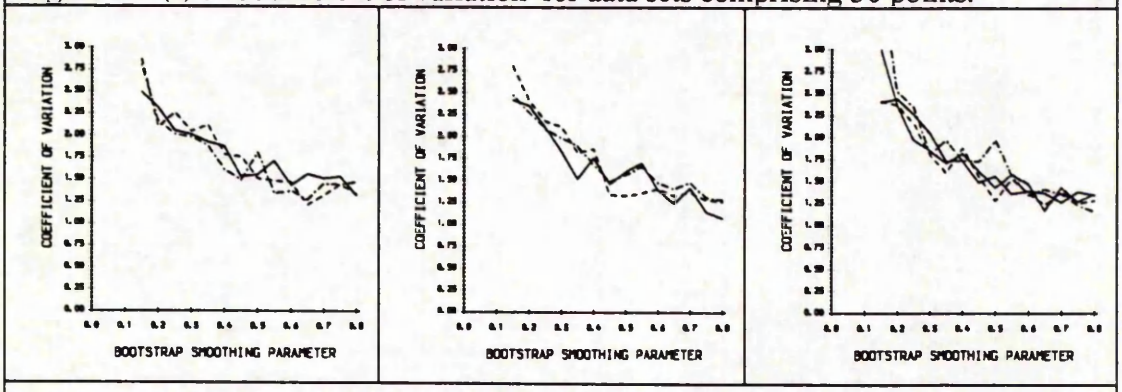


Figure 5.9 (b) :- Coefficient of variation for data sets comprising 100 points.

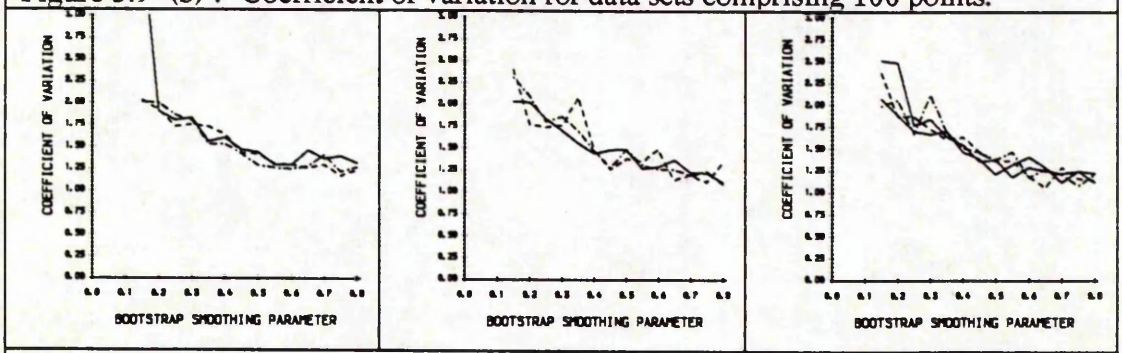


Figure 5.9 (c) :- Coefficient of variation for data sets comprising 150 points.

KEY	KEY	KEY
——— 22 × 22	——— 45 × 45	——— 59 × 59
– – – – 32 × 32	– – – – 50 × 50	– – – – 63 × 63
·-·-·-·-· 39× 39	·-·-·-·-· 55 × 55	·-·-·-·-· 67 × 67
		——·——·—— 71 × 71

Figure 5.9 :- Perimeter results for coefficient of variation

§5.3.2.2 Perimeter

The pattern of the levels of variability due to the surface fitting regime for perimeter differ to those of area, figure 5.9. In terms of the original data set size, a considerable difference is expressed between the results for 50, and 100 and 150 points. Table 5.3 reports the results for the mean value of the original smoothing parameter:-

No.of points	Grid Size									
	22^2	32^2	39^2	45^2	50^2	55^2	59^2	63^2	67^2	71^2
	Coefficient of variation									
50	1.87	2.54	1.84	2.02	2.05	1.90	1.56	1.78	2.42	2.53
100	1.53	1.76	1.49	1.46	1.33	1.36	1.49	1.56	1.73	1.61
150	1.60	1.64	1.53	1.44	1.47	1.32	1.47	1.64	1.51	1.64

Table 5.3 :- Results for the coefficient of variation for perimeter.

Firstly, across grid resolutions differences of 20% and upwards are reported between the results for 50 data points and the other two data sets. Secondly, a trend emerges in terms of the smoothed bootstrap parameter, the results do appear to stabilise from the apparent mean value of the smoothing parameter upwards. This levelling off of variability is masked by fluctuations in the results, particularily for the coarser grids. Increasing the number of simulations may smooth out of some of these fluctuatiuons.

A final feature of these results is the consistency of the results for all the values of mesh resolution considered, the range being slightly tighter for 150 points than for 100 points. In terms of selection of grid size, the variability in the perimeter results indicates that the effect on perimeter need not be considered, as long as the value of bootstrap smoothing parameter selected, lies close to the value of smoothing parameter determined from the original data.

In terms of the overall variability, approximately 1.5% variation is seen in the results, five times that for area. The reason for this five-fold increase may be explained by considering fractals.

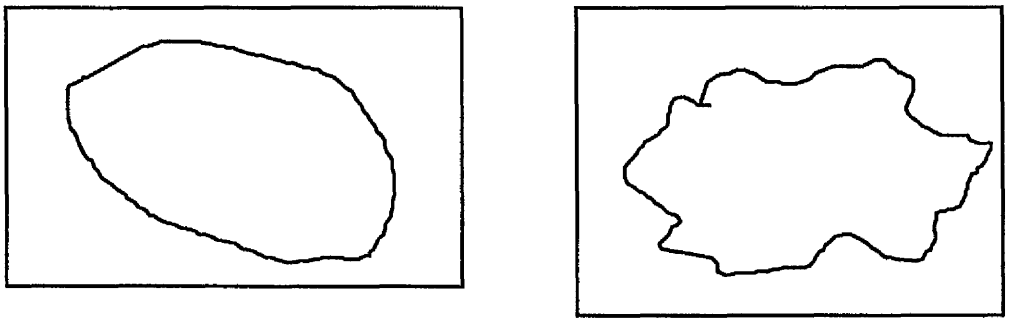


Figure 5.10 :- Variation in perimeter for a given area.

Figure 5.10 displays two shapes which have the same resultant area but whose perimeters differ considerably. These two shapes may easily be the result of two simulation runs. It is this feature of fractals which contributes to the differences in levels of variability for area and perimeter.

§5.3.2.3 Orientation

For the remaining geometric measurements, orientation and centroid displacement, the variability introduced into the results due to the surface fitting procedure was examined for each of the grid resolutions, but the various levels of bootstrap smoothing parameter were not examined. Computationally this procedure was extremely intensive. Utilising the results for the earlier measures, area and perimeter, the evidence strongly supported the case for choosing the smoothing parameter, h_{bs}, to take the same value as evaluated for the original data.

Figure 5.11 illustrates the results for orientation. Orientation appears to react in a similar manner to that of perimeter. Changes in grid resolution have a minimal effect on the level of variability introduced into the system for 100 and 150 points and for grids above 32×32 for 50 points. Increasing the sample size sees a marked reduction in the coefficient of variation between data sets of 100 and 150 points and 50 points.

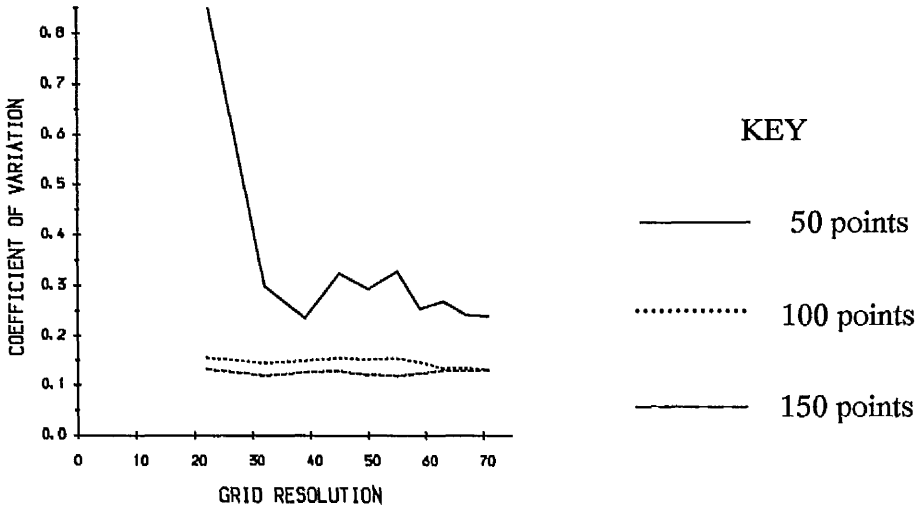


Figure 5.11 :- Coefficient of variation results for orientation.

One possible explanation is that for larger data sets, the contour shape is more clearly defined hence less variation will be introduced. Approximately half of the variability is seen in 100 points compared to that for 50 points, with a further reduction of a sixth for 150 points.

Although not reported, increasing the correlation from 0.5 to 0.9 resulted in a marked decrease in the coefficient of variation, particularly for 50 points. This raises the question of whether orientation should be used as a descriptor particularily for small data sets where orientation may be ill-defined. Generally there will be little change in the angular direction but where it does occur, the two contours to be compared will possibly be analogous in shape hence changes will be easily identifiable. There will be situations where the definition of orientation is questionable, as mentioned previously e.g. cyclic contours.

For orientation to be a satisfactory descriptor, the analysis of large data sets is required to ensure variation introduced as a result of the surface fitting technique is minimised, so that it can be separated from change due to some unknown underlying process.

§5.3.2.4 Centroid Displacement

The effect of mesh resolution on centroid displacement is shown in figure 5.12, effectively it is similar to that of area. Changes in grid resolution affect the level of variability, whilst changes in the size of the data set has no real effect.

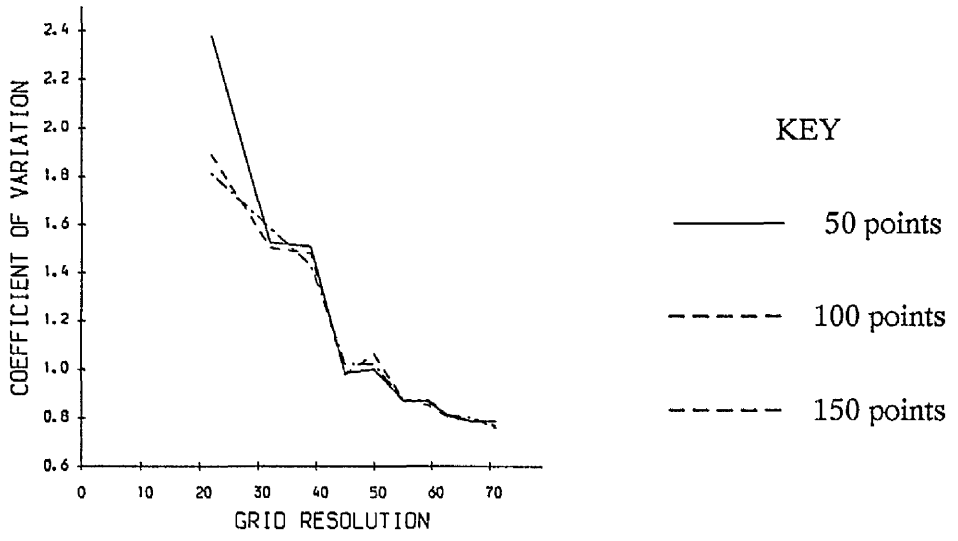


Figure 5.12 :- Coefficient of variation results for centroid displacement.

As for orientation, considerable variation is seen in the results. Potentially this may be a result of the variability introduced from evaluating the centre of gravity, since for the evaluation of both orientation and centroid displacement it is required to be evaluated.

§5.3.2.5 Smoothing Parameter

Although not strictly a geometric measurement of interest, it was decided for completeness to examine the level of variability introduced into the selection of the smoothing parameter when the bootstrap smoothing parameter and data set size were varied, the results are presented in table 5.4 and figure 5.13.

	50 points	100 points	150 points
Coefficient of Variation	0.100	0.087	0.072

Table 5.4 :- Coefficient of variation in the estimated smoothing parameter, h.

§5.3.2.4 Centroid Displacement

The effect of mesh resolution on centroid displacement is shown in figure 5.12, effectively it is similar to that of area. Changes in grid resolution affect the level of variability, whilst changes in the size of the data set has no real effect.

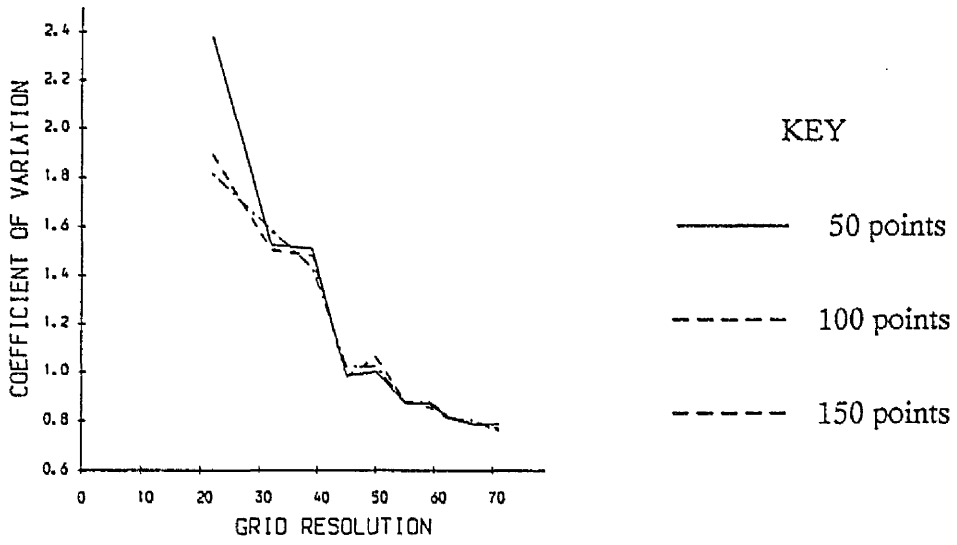


Figure 5.12 :- Coefficient of variation results for centroid displacement.

As for orientation, considerable variation is seen in the results. Potentially this may be a result of the variability introduced from evaluating the centre of gravity, since for the evaluation of both orientation and centroid displacement it is required to be evaluated.

§5.3.2.5 Smoothing Parameter

Although not strictly a geometric measurement of interest, it was decided for completeness to examine the level of variability introduced into the selection of the smoothing parameter when the bootstrap smoothing parameter and data set size were varied, the results are presented in table 5.4 and figure 5.13.

	50 points	100 points	150 points
Coefficient of Variation	0.100	0.087	0.072

Table 5.4 :- Coefficient of variation in the estimated smoothing parameter, h.

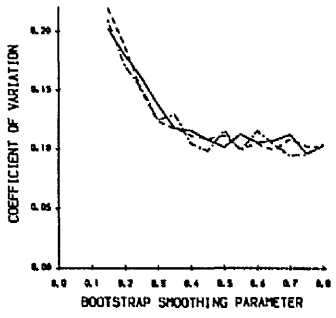


Figure 5.13 (a) :- 50 points

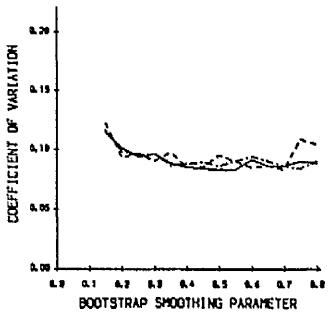


Figure 5.13 (b) :- 100 points

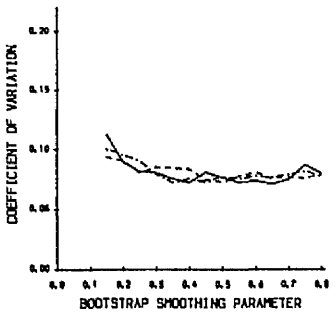


Figure 5.13 (c) :- 150 points

KEY :_ Mesh Resolution ——— 22^2 —·—·—·— 45^2 – – – – – – 63^2

Figure 5.13 :- Results for coefficient of variation for the smoothing parameter.

When evaluating a surface, it is the original localities of the points which are of importance for selecting the smoothing parameter, when using least squares cross-validation, hence the superposition of the results for different levels of mesh resolution. In terms of different values for the smoothed bootstrap, a similar pattern emerges to that of before. For the range of values which were defined for the original 200 simulations, the results were extremely consistent, particularily for the lower quartile upwards. This explains the sharp increase in the coefficient of variation for 50 points for values of the smoothed bootstrap less than 0.47, the lower quartile, and 0.39 for 100 points and 0.35 for 150 points.

§5.3.2.6 Summary

Returning to the original very simple example, section 3.8, which illustrated the importance of grid resolution, the results obtained in terms of the strong dependence of the areal results on grid size and the minor role it plays in influencing perimeter variability was confirmed.

The size of the data set was of crucial importance in terms of variability introduced into the perimeter results. In practice comparability of data set of 100 points and upwards of varying or equal size poses no problems. However for smaller data sets, comparisons between these and large data sets will display a lack of concordiality in variability due to the surface fitting mechanism. It is physically impossible to reduce the level of

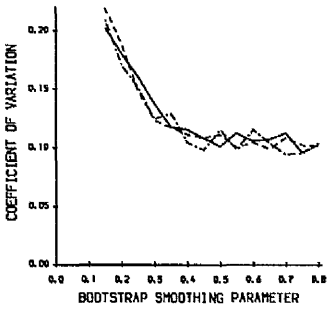


Figure 5.13 (a) :- 50 points

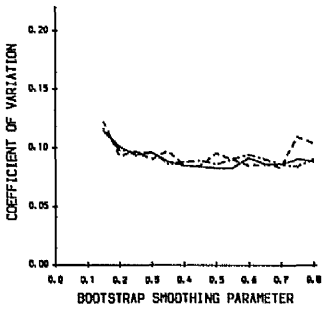


Figure 5.13 (b) :- 100 points

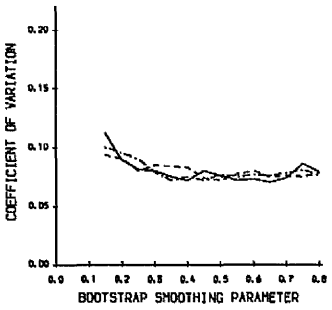


Figure 5.13 (c) :- 150 points

KEY :_ Mesh Resolution ——— 22^2 —·—·—·— 45^2 – – – – – – 63^2

Figure 5.13 :- Results for coefficient of variation for the smoothing parameter.

When evaluating a surface, it is the original localities of the points which are of importance for selecting the smoothing parameter, when using least squares cross-validation, hence the superposition of the results for different levels of mesh resolution. In terms of different values for the smoothed bootstrap, a similar pattern emerges to that of before. For the range of values which were defined for the original 200 simulations, the results were extremely consistent, particularily for the lower quartile upwards. This explains the sharp increase in the coefficient of variation for 50 points for values of the smoothed bootstrap less than 0.47, the lower quartile, and 0.39 for 100 points and 0.35 for 150 points.

§5.3.2.6 Summary

Returning to the original very simple example, section 3.8, which illustrated the importance of grid resolution, the results obtained in terms of the strong dependence of the areal results on grid size and the minor role it plays in influencing perimeter variability was confirmed.

The size of the data set was of crucial importance in terms of variability introduced into the perimeter results. In practice comparability of data set of 100 points and upwards of varying or equal size poses no problems. However for smaller data sets, comparisons between these and large data sets will display a lack of concordiality in variability due to the surface fitting mechanism. It is physically impossible to reduce the level of

variability for 50 points to that for 100 points. Realistically we are talking of variation of the level 1% to 2% in the results, this will generally be of lesser importance than variability due to physical measurements.

Returning to the areal measurement, it is this one measure which controls the level of grid over which to interpolate. A grid comprising 4000 points appears to minimise the levels of variability and hence ensure comparability.

For orientation and centroid displacement, the bootstrap smoothing parameter was selected to be equal to that for the original data. Orientation responded in a similar manner to that of perimeter i.e. data set size was the most influential factor in determining the level of variation due to the surface fitting technique whilst for centroid displacement, grid resolution was the controlling factor.

The other aspect which emerged during the examination of grid resolution concerned the choice of smoothing parameter when executing the smoothed bootstrap. An interval for the range of smoothing parameters calculated using least square cross-validation was constructed by a simulation study. By selecting any value which was contained within the interval which has as its lower bound the lower quartile, consistency was achieved.

In practice, a simulation study may be performed on the data sets of interest and the mean value taken for the smoothing parameter for the smoothed bootstrap. Computationally this is not excessive, unless the data set of interest is very large. Selecting the same value as for the original data set will be satisfactory, unless a high degree of accuracy is necessitated.

§5.4 Simple Statistics For Describing Change

The previous section examined the effect of grid resolution on various geometric descriptors and secondly, the level of bootstrap smoothing parameter required to minimise variability from the surface fitting procedure to ensure comparability was achieved across surfaces, generated from data sets of possibly differing size. We now move on to consider the comparison of two surfaces, constructed using the methods described in chapter 3.8.

variability for 50 points to that for 100 points. Realistically we are talking of variation of the level 1% to 2% in the results, this will generally be of lesser importance than variability due to physical measurements.

Returning to the areal measurement, it is this one measure which controls the level of grid over which to interpolate. A grid comprising 4000 points appears to minimise the levels of variability and hence ensure comparability.

For orientation and centroid displacement, the bootstrap smoothing parameter was selected to be equal to that for the original data. Orientation responded in a similar manner to that of perimeter i.e. data set size was the most influential factor in determining the level of variation due to the surface fitting technique whilst for centroid displacement, grid resolution was the controlling factor.

The other aspect which emerged during the examination of grid resolution concerned the choice of smoothing parameter when executing the smoothed bootstrap. An interval for the range of smoothing parameters calculated using least square cross-validation was constructed by a simulation study. By selecting any value which was contained within the interval which has as its lower bound the lower quartile, consistency was achieved.

In practice, a simulation study may be performed on the data sets of interest and the mean value taken for the smoothing parameter for the smoothed bootstrap. Computationally this is not excessive, unless the data set of interest is very large. Selecting the same value as for the original data set will be satisfactory, unless a high degree of accuracy is necessitated.

§5.4 Simple Statistics For Describing Change

The previous section examined the effect of grid resolution on various geometric descriptors and secondly, the level of bootstrap smoothing parameter required to minimise variability from the surface fitting procedure to ensure comparability was achieved across surfaces, generated from data sets of possibly differing size. We now move on to consider the comparison of two surfaces, constructed using the methods described in chapter 3.8.

§5.4.1 Scalar comparators

Three of the more intuitive comparators based on area, perimeter and area/perimeter combined are described in table 5.5.

	Area	Perimeter	Area/perimeter
Ratio	$\frac{A_{1j}}{A_{2j}}$	$\frac{P_{1j}}{P_{2j}}$	$\frac{A_{1j}/P_{1j}}{A_{2j}/P_{2j}}$
Difference	$A_{1j} - A_{2j}$	$P_{1j} - P_{2j}$	$\frac{A_{1j}P_{2j} - A_{2j}P_{1j}}{A_{2j}}$
Proportion	$\frac{A_{1j}}{A_{1j} + A_{2j}}$	$\frac{P_{1j}}{P_{1j} + P_{2j}}$	$\frac{A_{1j}(P_{1j} + P_{2j})}{P_{1j}(A_{1j} + A_{2j})}$

Figure 5.5 :- Operators based on area and perimeter for performing a scalar comparison.

Similar comments are applicable to area, perimeter and area/perimeter, only those properties relating to area will be discussed in detail.

§5.4.1.1 Ratio

The first of the statistics was that of the ratio of the two areas of interest. In terms of a good comparator, ratio concurs with the three definitions :-

1. For differing areal sizes, the results are scalar independent.

2. $\forall\ C_{ij}$, such that $A_{ij} > 0$, where C_{ij} is the region of interest, then

$$A_R = \begin{cases} 1 & \text{if } A_{1j} = A_{2j} \\ \rightarrow +\infty & \text{if } A_{1j} >> A_{2j} \\ \rightarrow 0 & \text{if } A_{1j} << A_{2j} \end{cases} \qquad \text{where} \quad A_R = \frac{A_{1j}}{A_{2j}}$$

From the definition of A_{ij} 0 and ∞ are unattainable in practice.

§5.4.1 Scalar comparators

Three of the more intuitive comparators based on area, perimeter and area/perimeter combined are described in table 5.5.

	Area	Perimeter	Area/perimeter
Ratio	$\frac{A_{1j}}{A_{2j}}$	$\frac{P_{1j}}{P_{2j}}$	$\frac{A_{1j}/P_{1j}}{A_{2j}/P_{2j}}$
Difference	$A_{1j} - A_{2j}$	$P_{1j} - P_{2j}$	$\frac{A_{1j}P_{2j} - A_{2j}P_{1j}}{A_{2j}}$
Proportion	$\frac{A_{1j}}{A_{1j} + A_{2j}}$	$\frac{P_{1j}}{P_{1j} + P_{2j}}$	$\frac{A_{1j}(P_{1j} + P_{2j})}{P_{1j}(A_{1j} + A_{2j})}$

Figure 5.5 :- Operators based on area and perimeter for performing a scalar comparison.

Similar comments are applicable to area, perimeter and area/perimeter, only those properties relating to area will be discussed in detail.

§5.4.1.1 Ratio

The first of the statistics was that of the ratio of the two areas of interest. In terms of a good comparator, ratio concurs with the three definitions :-

1. For differing areal sizes, the results are scalar independent.

2. $\forall$ C_{ij}, such that $A_{ij} > 0$, where C_{ij} is the region of interest, then

$$A_R = \begin{cases} 1 & \text{if } A_{1j} = A_{2j} \\ \rightarrow +\infty & \text{if } A_{1j} >> A_{2j} \\ \rightarrow 0 & \text{if } A_{1j} << A_{2j} \end{cases} \qquad \text{where} \quad A_R = \frac{A_{1j}}{A_{2j}}$$

From the definition of A_{ij} 0 and ∞ are unattainable in practice.

3. $\frac{A_{1j}}{A_{2j}} = \frac{1}{A_{2j}/A_{1j}}$ $\Rightarrow$ ordering is irrelevant to the outcome, since the reciprocal enables the same result to be achieved, when the two surfaces are transposed.

§5.4.1.2 Differences

For the difference operator all three conditions are not satisfied.

1. A very simple example illustrates the failure of the first property :-

$A_1 = 200$ units $\qquad A_2 = 100$ units

$A_1 = 2$ units $\qquad A_2 = 1$ unit

In terms of areal differences, there is a lack of consistency between the sizes of the differences hence comparability is impaired i.e. it is scale dependent.

2. $\forall\ C_{ij}$ such that $A_{ij} > 0$

$$A_D = \begin{cases} 0 & \text{if } A_{1j} = A_{2j} \\ \rightarrow +\infty & \text{if } A_{1j} >> A_{2j} \\ \rightarrow -\infty & \text{if } A_{1j} << A_{2j} \end{cases} \qquad \text{where} \quad A_D = A_{1j} - A_{2j}$$

In reality no surface is infinitely large or small hence the bounds are unattainable.

3. $|A_{1j} - A_{2j}| = |A_{2j} - A_{1j}| \Rightarrow$ ordering of the surface is irrelevant, if the absolute value is taken.

§5.4.1.3 Proportion

The idea of using proportion to assess the level of change has its foundation in probability. The conditions are satisfied as follows :-

3. $\frac{A_{1j}}{A_{2j}} = \frac{1}{A_{2j}/A_{1j}} \Rightarrow$ ordering is irrelevant to the outcome, since the reciprocal enables the same result to be achieved, when the two surfaces are transposed.

§5.4.1.2 Differences

For the difference operator all three conditions are not satisfied.

1. A very simple example illustrates the failure of the first property :-

$A_1 = 200$ units $\qquad A_2 = 100$ units
$A_1 = 2$ units $\qquad A_2 = 1$ unit

In terms of areal differences, there is a lack of consistency between the sizes of the differences hence comparability is impaired i.e. it is scale dependent.

2. $\forall\ C_{ij}$ such that $A_{ij} > 0$

$$A_D = \begin{cases} 0 & \text{if } A_{1j} = A_{2j} \\ \rightarrow +\infty & \text{if } A_{1j} >> A_{2j} \\ \rightarrow -\infty & \text{if } A_{1j} << A_{2j} \end{cases} \qquad \text{where} \quad A_D = A_{1j} - A_{2j}$$

In reality no surface is infinitely large or small hence the bounds are unattainable.

3. $|A_{1j} - A_{2j}| = |A_{2j} - A_{1j}| \Rightarrow$ ordering of the surface is irrelevant, if the absolute value is taken.

§5.4.1.3 Proportion

The idea of using proportion to assess the level of change has its foundation in probability. The conditions are satisfied as follows :-

1. $\frac{A_{1j}}{A_{1j}+A_{2j}}$ satisfies the assumption of comparability.

2. $\forall\ C_{ij}$ such that $A_{ij} > 0$

$$A_P = \begin{cases} \frac{1}{2} & \text{if } A_{1j} = A_{2j} \\ \rightarrow 1 & \text{if } A_{1j} >> A_{2j} \\ \rightarrow 0 & \text{if } A_{1j} << A_{2j} \end{cases} \qquad \text{where} \quad Ap = \frac{A_{1j}}{A_{1j}+A_{2j}}$$

Although unattainable, the limits are more acceptable to field scientists.

3. $\frac{A_{1j}}{A_{1j}+A_{2j}} = 1 - \frac{A_{2j}}{A_{1j}+A_{2j}} \Rightarrow$ since the measure is based on a probabilistic definition, the problem of ordering is avoided by subtracting one from one of the measures.

§5.4.1.4 Scalar summary

Both ratio and proportion measures are applicable for describing scalar changes, since both satisfy the three conditions described in section 5.1. Returning to the two simple areal regions described earlier, the circle and the rectangle, it can be shown that there is a strong association between the two operators. By defining the original dimensions to be a and b and r for the rectangle and the circle respectively and ka, lb and kr for the dimension of the new surface, where k and l are real in definition, the relationship between the measures are given in table 5.6.

	Ratio	Proportion
Rectangle	$\frac{1}{kl}$	$\frac{1}{(1+kl)}$
Circle	$\frac{1}{k^2}$	$\frac{1}{(1+k^2)}$

Figure 5.6 :- Relationship between scalar comparators for two surfaces.

1. $\frac{A_{1j}}{A_{1j}+A_{2j}}$ satisfies the assumption of comparability.

2. $\forall\ C_{ij}$ such that $A_{ij} > 0$

$$A_P = \begin{cases} \frac{1}{2} & \text{if } A_{1j} = A_{2j} \\ \rightarrow 1 & \text{if } A_{1j} >> A_{2j} \\ \rightarrow 0 & \text{if } A_{1j} << A_{2j} \end{cases} \qquad \text{where} \quad Ap = \frac{A_{1j}}{A_{1j}+A_{2j}}$$

Although unattainable, the limits are more acceptable to field scientists.

3. $\frac{A_{1j}}{A_{1j}+A_{2j}} = 1 - \frac{A_{2j}}{A_{1j}+A_{2j}} \Rightarrow$ since the measure is based on a probabilistic definition, the problem of ordering is avoided by subtracting one from one of the measures.

§5.4.1.4 Scalar summary

Both ratio and proportion measures are applicable for describing scalar changes, since both satisfy the three conditions described in section 5.1. Returning to the two simple areal regions described earlier, the circle and the rectangle, it can be shown that there is a strong association between the two operators. By defining the original dimensions to be a and b and r for the rectangle and the circle respectively and ka, lb and kr for the dimension of the new surface, where k and 1 are real in definition, the relationship between the measures are given in table 5.6.

	Ratio	Proportion
Rectangle	$\frac{1}{kl}$	$\frac{1}{(1+kl)}$
Circle	$\frac{1}{k^2}$	$\frac{1}{(1+k^2)}$

Figure 5.6 :- Relationship between scalar comparators for two surfaces.

Realistically the numerical results for the ratio and proportion between two surfaces differ only by a constant. Although, the surfaces of interest will tend to be more complex in nature and associations between the operators may not be as straight forward as those displayed in table 5.6, it was decided only to investigate the test statistics relating to ratio. Similar comments apply to the test statistics when applied to the measure of perimeter.

§5.4.2 Rotation

The simplest means to describe the change in the orientation of a contour after a time period, t, or between two variables is in terms of an angular metric.

§5.4.2.1 Ratio

The metric describing angular change as a ratio is :-

$$OR_R = \frac{\theta_{1j}}{\theta_{2j}} \qquad \text{where} \quad \begin{cases} 0 \le \theta_{ij} \le 2\pi \\ 0 \le OR_R \le \infty \end{cases}$$

Although the form of the operator is dimensionless, the inclusion of infinity as one of its bounds is undesirable. Unlike previous situations this bound is attainable in practice i.e. when the contour relating to surface two is parallel to the x-axis. Although ordering of the surfaces does not affect the results since the reciprocal provides an equivalent result, when the two surfaces are transposed, in practice the concept of ratio is unworkable.

§5.4.2.2 Differences

The second form is that of differences :-

$$OR_D = \theta_{1j} - \theta_{2j} \qquad \text{where} \quad \begin{cases} -2\pi \le OR_D \le 2\pi & \text{non-symmetric contour} \\ -\pi \le OR_D \le \pi & \text{symmetric contour} \\ 0 \le \theta_{ij} \le 2\pi \end{cases}$$

The three properties relating to ordering, dimensionality and the bound conditions are all satisfied for angular differences.
Although unrealistic in practice, the bounds of the operator OR_D have been split according to whether the contour is symmetric.

For all work on angular change, measurements will be made in terms of radians as opposed to degrees. The major difficulty in orientation lies in the problems incurred in defining the principal axis, once these have been overcome the choice of operator is unquestionable i.e. differences.

§5.4.3 Translation Descriptors

The final transformation is that of translation. Two possible statistics will be examined :-

1. The distance between two uniquely identifiable points in the area of interest i.e. centroid displacement.

2. The area of overlap.

A shift in location is liable to arise in many different situations e.g. in the marine environment, the transfer of oil slicks, saline or debris by the prevailing wind and tidal conditions.

§5.4.3.1 Centroid displacement

Three possible means of evaluating displacement are :-

1. Standard displacement

2. Percentage displacement

3. Standard deviation displacement

Before evaluating displacement, rotating the contour of interest such that its major and minor axis lie parallel to the x and y axis respectively, simplifies the ensuing analysis. Displacement may be defined more simply for an isotropic surface i.e. one in which the

variation is approximately constant over the surface compared to an anisotropic surface, where variability is dependent on direction and therefore interest is in analysing the directions separately. The two directional components are defined to be in the x and y directions. Prior knowledge of a process may enable the true directions of variability to be determined and displacement analysed accordingly. Table 5.7 identifies the form for each of the descriptors based on the centre of gravity for each surface.

	Standard Displacement	Percentage Displacement	Standard Deviation Displacement
Isotropic Surface	D_{COG}	Undefined	SD_{COG}
Anisotropic Surface X- Direction	$XCG_{1i} - XCG_{2i}$	$\frac{XCG_{1j} - XCG_{2j}}{\text{major axis}}$	$\frac{XCG_{1j} - XCG_{2j}}{SD_{xj}}$
Y- Direction	$YCG_{1i} - YCG_{2i}$	$\frac{YCG_{1j} - YCG_{2j}}{\text{minor axis}}$	$\frac{YCG_{1j} - YCG_{2j}}{SD_{yj}}$

where $D_{COG} = \sqrt{(XCG_{1j} - XCG_{2j})^2 + (YCG_{1j} - YCG_{2j})^2}$

$$SD_{COG} = \frac{\sqrt{(XCG_{1j} - XCG_{2j})^2 + (YCG_{1j} - YCG_{2j})^2}}{SD_{xyj}}$$

SD_{xyj} = pooled standard deviation in x and y direction, for contour level j.
SD_{xj} = pooled standard deviation in x-direction, for contour level j.
SD_{yj} = pooled standard deviation in y direction, for contour level j.

Table 5.7 :- Descriptors for the various forms of centroid displacement.

§5.4.3.2 Standard displacement

The first of the comparators suggested was standard displacement. Advantages associated with this measure are :-

1. Ease of definition in x and y directions and if necessary x-y direction.

2. Intuitively, it is easily interpreted.

3. The ordering of the surfaces is not a criterion which has to be considered since by taking the absolute value the same result is obtained. The sign of the difference is an indicator as to the direction of the change, this is not an issue at this juncture.

4. The achievable bounds are contained within the limits of $-\infty$ and $+\infty$.

The major problem with such a metric is that the measure is dimension dependent. For some surfaces interest may be in small measures e.g. millimetres, whilst in other examples, particularly geographical applications, we may be working in terms of kilometres. This disadvantage unfortunately negates the usefulness of the metric as a comparator unless some form of standardisation is performed.

§5.4.3.3 Percentage Displacement

A number of difficulties are encountered when defining percentage displacement:-

1. For isotropic surfaces, the definition of a combined measure is not intuitive. This feature is not necessarily a major drawback, there are many situations where the analyst is more concerned with examining the separate components of location shift.

2. The condition of robustness is violated, the major and minor axis are defined to be expressions of one of the surfaces. Changing the baseline surface is liable to affect the percentage displacement, unless the two surfaces are coincident in shape. Where interest is in change across time periods then a natural ordering is imposed, however for associations between variables, an ordering is much harder to implement.

3. A final problem associated with percentage displacement is concerned with the more fundamental interpretation of the resultant value. A 50% shift in the x-direction may relate to two units, whilst a displacement of 50% in the y-direction may be of the order of ten units, depending on the size of the contour under question. In reality the two unit shift may not be capable of being detected over and above the inherent natural variability attributable to random noise in the spatial measurements, see chapter 6.

Linked to this problem is that of the size of the contour i.e. a displacement of 50%, say, in one of the lower order contour levels is much less likely to be achieved due to the bounds imposed by the region of interest than for a higher order level when the area covered will be correspondingly less.

Although subjectively, we may state contour A has been displaced by 50%, mathematically this presents a number of subtle complications. A way round this problem is to measure location shift as a proportion of the standard deviation of the contour.

§5.4.3.4 Standard deviation displacement

The last of the descriptors for translational change is standard deviation displacement. The idea behind this potential test statistic is to express the level of change as a proportion of the standard deviation of the data points defining the contour of interest. Invariably, we will be comparing two contours which may be matched in terms of shape or location, hence it is feasible to assume the standard deviation of the two contours are similar so that pooling the variances is a valid assumption. It is under this premise that the definition of standard deviation displacement is made. For an isotropic surface, the variances in the x and y directions are pooled satisfying the property of robustness.

Using this formulation, it is hoped that consistency can be achieved in terms of describing displacement between surfaces with differing dimensions. Further to the definition of a good test statistic, the metric described is both dimensionless and the achievement of infinite bounds is impossible, where the two surfaces exist.

§5.4.3.5 Overlap

The second translation statistic is the overlap function, k_j, between two contours for surfaces, i and i_1, say,at level, j, k_j, can be expressed as :-

$$k_j = C_{ij} \cap_{i \neq i_1} C_{i_1 j} \qquad \text{where} \quad j = 1, N_C$$

$$i, i_1 = 1, N_S \quad i \neq i_1$$

where N_C = number of contours

N_S = number of surfaces

Computationally, the method for evaluating k_j is as follows :-

1. Define those points of C_{ij} which are contained entirely within or make contact with $C_{i,j}$.

2. Repeat step 1, transcribing $C_{i,j}$ with C_{ij}.

3. Combine the two data sets.

4. Order the co-ordinates sequentially in an anti-clockwise direction.

5. Evaluate the overlap area.

Two alternative forms exist for describing overlap, the coefficient of areal association, C_a, and standardised overlap, C_s.

$$C_a = \frac{\text{area over which the two phenomena are located together}}{\text{total area covered by the two phenomena}}$$

$$= \frac{A_{ov}}{A_{1j} + A_{2j} - A_{ov}} \qquad 0 \leq C_a \leq 1$$

$$C_s = \frac{\text{area over which the two phenomena are located together}}{\text{baseline area}}$$

$$= \frac{A_{ov}}{A_{1j}} \qquad 0 \leq C_s \leq 1$$

Both descriptors range from zero to one with completely separate distributions giving a value of zero whilst exactly coincident surfaces are described by a value of unity.

§5.4.3.6 Standardised overlap

The applicability of standardised overlap is restricted to situations where a natural ordering may be imposed on the surfaces so that the definition of a baseline surface is not

arbitrary. The other two conditions of, comparability of units and finite bounds are both satisfied.

§5.4.3.7 Coefficient of areal correspondence

As for standardised overlap, the two properties concerning bounds and dimensions are satisfied.

The main restriction of this descriptor relates to the special case of surface A being contained entirely within surface B. The measure then reduces to the ratio of the areas of the two surfaces and as a consequence imparts no information on the level of shift which has been incurred.

This situation will not always arise and in instances where a shift is not wholly contained within one of the contours, the metric C_a, reveals as much information about the level of displacement as measures founded on centroid displacement and no prior assumptions concerning the variance of the contour need be stated.

§5.4.3.8 Translation summary

Translation may be described in terms of the displacement of the centre of gravity or alternatively by the amount of overlap. Of the three metrics described to analyse the question of translation, only one turns out to be practically viable, standard deviation displacement. Intuitively the concept of standard displacement is the most satisfactory but the property of dimension dependence nullifies its global applicability.

Percentage displacement relies heavily on the natural ordering of surfaces to enable the metric to be evaluated, this feature is not satisfactory especially where comparisons are between different variables e.g. beta and gamma radiation levels.

The usage of standardised overlap as a possible test statistic is restricted by its reliance on ordering, hence the coefficient of areal correspondence will be used for quantifying overlap.

§5.4.4 Summary

Development of any statistical procedures to assess the level of change between two spatial processes will invariably necessitate some form of subjective pre-analysis. By examining a simple overlay of the two surfaces, a judicious choice of comparators may be selected.

Although this may be believed to bias the results since it is effectively performing a comparison based on prior information, it may be stated that in all situations a test of translation, scaling and rotation will be performed, it is only the format of the test which is under consideration, all tests should produce equivalent results. The main question which arises concerns the duplication of the test for scale by using the coefficient of areal correspondence for testing for translation, where surface A is enclosed by surface B. In this instance standard deviation displacement is a more appropriate measure. In all other instances, the overlap metric is possibly more valid as prior assumptions concerning the variability of the contour need not be made.

In terms of rotation, the choice of comparator is unambiguous i.e. differences in orientation of the two contours of interest, whilst for scalar change, interest will focus on the ratio of both area and perimeter.

CHAPTER 6

AN HYPOTHESIS TESTING APPROACH TO SURFACE COMPARISON

§6.1 Introduction

Within the previous chapter attention focused on a possible approach to describing change. The motivation for this approach arose from the mathematical description of those features which man perceives when assessing subjectively whether change has occurred between surfaces represented as contours, which have been overlain. The factors considered were area, perimeter, centre of gravity and orientation. Various combinations of the above factors allowed different aspects of change to be quantified in terms of rotational, translational and scalar change.

In terms of a scalar change, ratios, proportions and differences of area, perimeter and area/ perimeter combined were discussed. Theoretically, the three measures, area, perimeter and area/perimeter, are related hence the selection of potential test-statistic is not clear-cut. From the ensuing work on mesh resolution, the behaviour of the area/perimeter measure mirrored that of perimeter, hence only area and perimeter formulations were used as the basis of the subsequent test statistics. The test statistics selected for quantifying scalar change were based on the ratio of area and perimeter. Differences between the areas or perimeters of comparable contours are inappropriate since they lack global applicability, whilst measures formulated from proportions were shown to differ from ratio, only by a constant.

For rotation, only one plausible descriptor was suggested, namely angular change i.e. the difference between the angles of orientation for the contours of interest. Finally for translation, overlap and centroid displacement were shown to be applicable in different situations. Overlap generally being the better all round descriptor since no prior assumptions are made concerning its usage, unlike centroid displacement where various assumptions concerning the equality of the variances of the contours are made. Where contour A is contained entirely within contour B overlap reduces to the ratio of the two areas, hence to avoid the problem of correlation between this measure and that of areal

ratio, centroid displacement should be used. Table 6.1 summarises the descriptors which will be used as the basis of the test statistics.

Transformation	(1)	(2)	(3)	(4)
Translation	SD_{COG}	$\frac{XCG_{1j} - XCG_{2j}}{SD_{Xj}}$	$\frac{YCG_{1j} - YCG_{2j}}{SD_{Yj}}$	overlap k_j
Scalar	$\frac{A_{1j}}{A_{2j}}$	$\frac{P_{1j}}{P_{2j}}$		
Rotation	$\theta_{1j} - \theta_{2j}$			

Table 6.1:- Summary of descriptors.

A series of hypothesis tests based on these descriptors are developed within this chapter to examine the question of change between spatial processes. Interest will mainly focus on a local form of analysis, by local we refer to specific contour levels.

It is well known that one of the main problems associated with environmental research concerns the large amount of variation encountered in the measurements, whether it be attributable to field measurements or alternatively, measurements taken in the laboratory. In some cases, it is recognised that the level of variation in a measurement may result in the available analytical capacity being exceeded. It also follows that if data are to be used in a modelling exercise, realistic estimates of overall error must be given if a situation is not to arise where the errors so compound one another that totally unrealistic results are obtained.

Initially in this chapter we examine the simplest situation possible, no random noise is present in the system. For this situation, the question 'Does contour A of surface 1, differ from that of contour A of surface 2, according to characteristic x?' is examined.

Investigation of this question takes the form of a series of hypothesis tests using the descriptors summarised in table 6.1 as the basis for the test statistics. The distribution of the test statistics under the null hypothesis will be developed analytically, by way of a

simulation study, since the theoretical derivation of the distribution of the characteristics of interest is extremely complex with a large number of assumptions being required.

Following on from this, the distribution of the test statistic under various alternative hypothesis is examined and from the results, a set of power curves are constructed.

The second part of the chapter will follow the same procedure but this time varying levels of random noise will be introduced into the data, the levels of noise having upper bounds of 5%, 15% and 25%. These three values being selected as representative of noise levels which may arise in practice. Attention will focus on the ability of various test statistics to distinguish between change due to some underlying process and that due to inherent random noise.

Finally the results will be drawn together and a global approach to the analysis of change will be proposed. Problems concerned with the implementation of both the local and global approaches will be discussed and the limitations of the technique described.

§6.2 Simulation Study

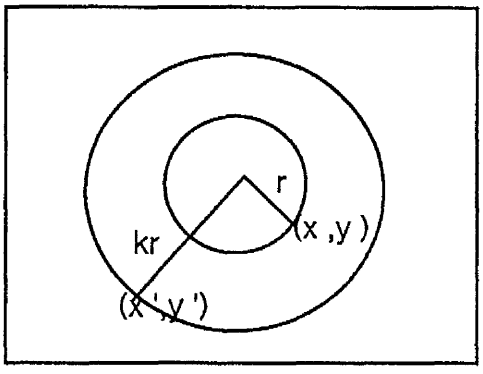
A major simulation study was undertaken to derive the empirical probability distribution function for the test statistics of interest, table 6.1. The null hypothesis, H0, takes the form of a simple hypothesis and effectively describes the situation of no change, Translating this to the parameters the null hypothesis are:-

1. Scalar H0 : $A_{1j} = A_{2j}$
2. Rotation H0 : $\theta_{1j} = \theta_{2j}$
3. Centroid displacement H0 : $C_d = 0$
4. Overlap H0 : area of overlap =100%

Before the power of the statistics under various alternative hypothesis could be assessed, it was necessary to introduce various levels of known change into the analysis. The form of the alternative hypothesis, H1, is composite and depending on the definition is either one or two-sided, this is discussed in detail in section 6.3.2.

§6.2.1 Scalar Change

For both area and perimeter, a similar procedure was implemented to introduce varying levels of change into the statistic of interest. The scale of a contour may be changed by multiplying each of the points by a constant, k, figure 6.1.



Key

k = constant ≥ 1

$x'_p = kx_p$

$y'_p = ky_p$

Figure 6.1 :- Scalar increase/decrease

When changing the scale of a pattern of points by a factor k, the points move k times as far away from the origin, unless the origin is (0,0). Although theoretically this is not undesirable, as it is the area and perimeter which are of interest, a number of practical considerations arise. Firstly, the contour of interest may be shifted outwith the region of investigation, resulting in a different contour being depicted as the i^{th} quartile, which would not necessarily be k times larger than the original contour as intended. The simplest means of avoiding this situation is to first translate the centre of the original data set to the origin (0,0) and then apply the transformation. The area/perimeter of the region should then be transformed by the corresponding factor. Running a test program to validate this procedure confirmed its appropriateness.

A further simplification implemented was to describe contours circular in shape since changes in areas and perimeters are simple to implement. This should not affect the resultant distribution under the null hypothesis which is shape independent.

§6.2.2 Orientation

In terms of the metric involving orientation change was introduced and controlled by the following process.

Suppose P is a point with co-ordinates (x_p, y_p) and that it is rotated about the origin of the co-ordinates through an angle θ. The resultant position of P, P* is given by:-

$$P^* = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_p \\ y_p \end{pmatrix}$$

When rotating a pattern of points through an angle θ, the rotation swings about the origin so that the pattern is moved elsewhere in space, as well as being rotated. This displacement may be avoided by initially translating the pattern so its centroid is at the origin (0,0) and then applying the rotation. As for the scalar metric, the assessment of the test statistic was carried out via a simulated example whose initial orientation was known.

§6.2.3 Centroid Displacement

A very simple transformation is required to achieve a known locational change:-

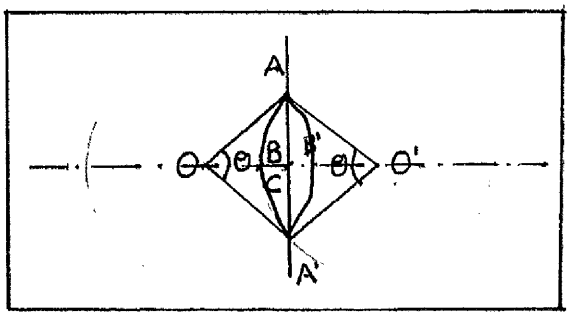
$$X^* = X + c\sigma_X \qquad (X,Y) \text{ original data point}$$
$$Y^* = Y + c\sigma_Y \qquad (X^*,Y^*) \text{ new data point}$$

where c = constant of proportion
σ_i = standard deviation of the original data set in direction i, i=x,y

The only minor modification required was to shift the region of interest by the same amount as the simulated data to ensure comparability of the contours of interest. By neglecting to make this translation, the shifted contour potentially failed to correspond in definition to the original contour unless coincident in size and shape.

§6.2.4 Overlap

Utilising the properties of a circle, the following method was devised to allow varying levels of change in overlap to be evaluated. Changes in overlap when dealing with two similar circles can be equated in terms of the difference between the sector area and the segment area, $A, figure 6.2.



KEY

$\angle AOA' = \angle AO'A' = \theta$ = overlap angle
$AOA'B' = AO'A'B = S_A$ = sector area
$AB'A'C = ABA'C = \$A$ = segment area
$ABA'B' = A_{ov}$ = overlap area

Figure 6.2 :- Definition of angles and areas used to evaluate overlap.

The above definitions may be described mathematically as follows :-

$$S_A = \frac{\theta r^2}{2}$$

$$\$A = S_A - \Delta AOA'$$

$$= \frac{\theta r^2}{2} - \frac{1}{2}\left(2r\sin\frac{\theta}{2} r\cos\frac{\theta}{2}\right)$$

$$= \frac{1}{2}r^2(\theta - \sin\theta)$$

$$A_{ov} = 2\$A$$

$$A_{ov} = r^2(\theta - \sin\theta)$$

From figure 6.2 the critical parameter for varying the overlap is the overlap angle, θ. For a specific level of overlap, the angle required to achieve this change for the coefficient of areal correspondance is as follows:-

$$C_a = \frac{A_{ov}}{A_{C_1} + A_{C_2} - A_{OV}}$$

$$= \frac{r^2(\theta - \sin\theta)}{\pi r^2 + \pi r^2 - r^2(\theta - \sin\theta)}$$

$$\frac{C_a}{1+C_a} = \frac{\theta - \sin\theta}{2\pi} \qquad \begin{matrix} 0 \le C_a \le 1 \\ 0 \le \theta \le 2\pi \end{matrix}$$

Figure 6.3 expresses the above relationship diagramatically.

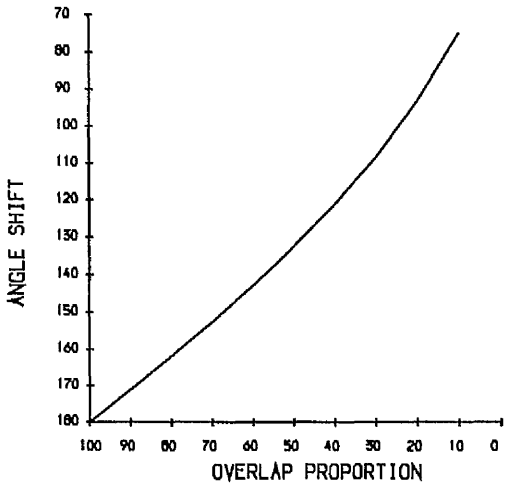


Figure 6.3 :- Values of angular displacement, θ, to achieve required level of overlap for the coefficient of areal correspondence.

A very simple angular transformation applied to the original co-ordinates allows the new set of co-ordinates of the displaced circle to be evaluated.

$$X_{new} = X_{old} + X_{shift}$$
$$Y_{new} = Y_{old}$$

$$\text{where } X_{shift} = 2\left\{r\cos\frac{\theta}{2}\right\}$$

An additional check was included in the program to ensure the radii of the two contours did not vary considerably and hence destroy the property of symmetry enjoyed when the two radii are equal and on which the whole argument for controlling overlap was founded.

§6.2.5 Summary

The investigation of the various test statistics postulated in the preceding sections was undertaken via a major simulation study. The form of the study was similar for the various measures with the two situations of no random noise and random noise being analysed separately.

The following sequence of steps were implemented to examine the effect of change on the various test statistics.

1. Simulate data points $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_N$, where N = 50, 100, 150, using the NAG subroutine G05EAF with

$$\mu = \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 1.0 & k \\ k & 1.0 \end{bmatrix}$$

$$\text{where } k = \begin{cases} 0.0 & \text{area} \\ 0.0 & \text{perimeter} \\ 0.9 & \text{orientation} \\ 0.5 & \text{centroid displacement} \\ 0.0 & \text{overlap} \end{cases}$$

2. Generate a smoothed bootstrap sample $\underline{x}_1^*, \underline{x}_2^*, \ldots, \underline{x}_N^*$, where the smoothing parameter is taken to be the value evaluated by least squares cross validation for the original data set $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_N$.

3. Apply the appropriate transformations to the bootstrap sample, table 6.2. Consider a point $P(x_p, y_p)$:-

	x-transformation	y-transformation	Conditions H_0	H_1
Area/ Perimeter	kx_p	ky_p	k=1	k>1
Orientation	$x_p \cos\theta - y_p \sin\theta$	$x_p \sin\theta + y_p \cos\theta$	$\theta=0$	$0\leq\theta\leq 2\pi$
Centroid Displacement	$x_p + c\sigma_x$	y_p	c=o	c>0
Overlap	$x_p + 2\left\{r\cos\frac{\theta}{2}\right\}$	y_p	$\theta=\pi$	$0\leq\theta\leq\pi$

Table 6.2 :- Transformations to point $P(x_p, y_p)$.

4. Calculate the smoothing parameter for the bootstrap sample using least squares cross-validation.

5. Construct the surface from the bootstrap sample using kernel density estimation and a grid resolution of 63×63.

6. Evaluate area ratio, perimeter ratio, location shift, orientation and overlap for the contour depicting the upper quartile.

7. Repeat steps (2)-(6) 250 times for the derivation of the empirical probability distribution function under the null hypothesis of no change. For the alternative hypothesis, 100 simulations were performed and the observed values for the appropriate test statistics calculated.

The reduction in the number of simulations under the alternative hypothesis was forced by the CPU time required to obtain the desired information. The resultant loss in accuracy was not critical since it was the general shape that was of interest, the power curves were possibly less smooth than might have been expected as a result.

§6.3 Simulation Results

§6.3.1 Distribution of the test statistic under the null hypothesis

The form of the null hypothesis for all the transformations is simple and in accordance with many applications is the hypothesis of no change.

§6.3.1.1 Area

The unsmoothed empirical probability distribution function for the test statistic relating to the ratio of two areas is given in figure 6.4.

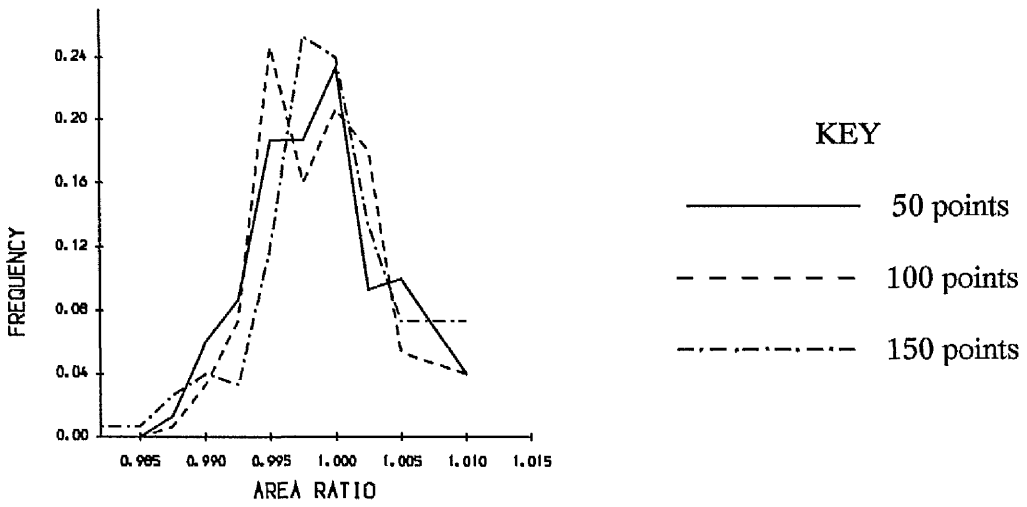


Figure 6.4 :- Empirical probability distribution for areal change under the null hypothesis of no change.

The most apparent feature is the consistency across the results for differing sizes of data set. This was anticipated from the earlier work concerning the investigation of the level of variability introduced as a result of the surface fitting procedure. In this case, data set size did not appear to unduly influence the variability of the results.

A numerical summary of the empirical probability distribution function in the form of the percentage points, q = 60, 80, 90, 97.5 and 99, is given in table 6.3, i.e.

$P(A_j > A_j(n;q)) = q/100$. The points tabulated correspond to area A_{1j} > area A_{2j} where the test statistic is defined to be $A_j = \frac{A_{1j}}{A_{2j}}$.

q	50 points	100 points	150 points
60	1.00316	1.00353	1.00385
80	1.00584	1.00485	1.00654
90	1.00691	1.00646	1.00839
95	1.00889	1.00803	1.00904
97.5	1.009815	1.01000	1.00979

Table 6.3 :- Percentage points for the empirical probability distribution function relating to areal change.

§6.3.1.2 Perimeter

Figure 6.5 illustrates the empirical probability distribution of the following test statistic for perimeter:-

$$P_j = \frac{P_{1j}}{P_{2j}} \qquad 0 < P_j < \infty$$

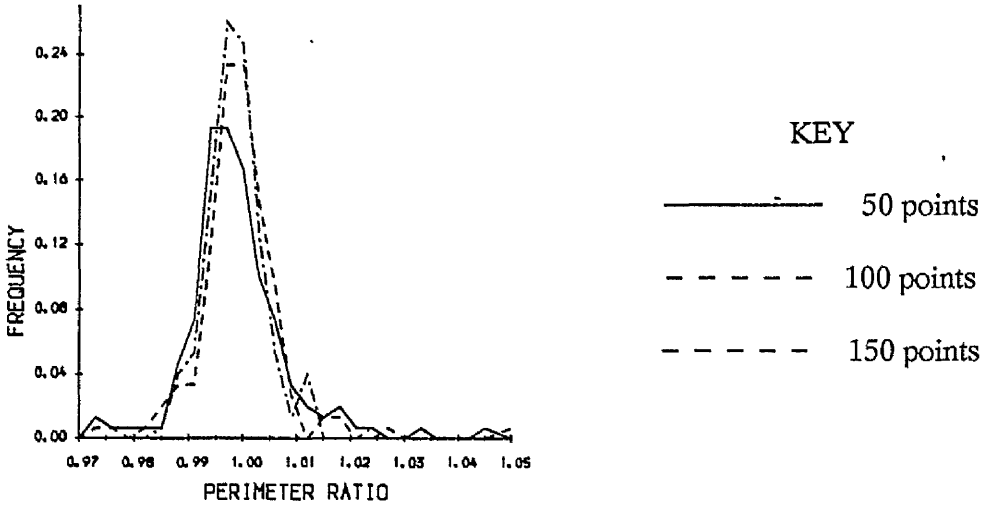


Figure 6.5 :- Empirical probability distribution for perimeter change under the null hypothesis of no change.

q	50 points	100 points	150 points
60	1.00551	1.00478	1.00398
80	1.01002	1.00704	1.00624
90	1.01623	1.00963	1.01138
95	1.02099	1.01607	1.01337
97.5	1.02929	1.01889	1.01412

Table 6.4 :- Percentage points for the empirical probability distribution function relating to perimeter change.

Table 6.4 summarises the numerical values obtained for the percentage points described for perimeter change based on the same criteria as for area i.e. perimeter P_{1j} > perimeter P_{2j}. Increasing the sample size sees a corresponding reduction in the range of the distribution function, with the tails of the distribution being shorter for the larger data sets. The shape of the distribution remains constant for all three data sets.

§6.3.1.3 Orientation

Under the null hypothesis of no change, the distribution of the test statistic is as described in figure 6.6..

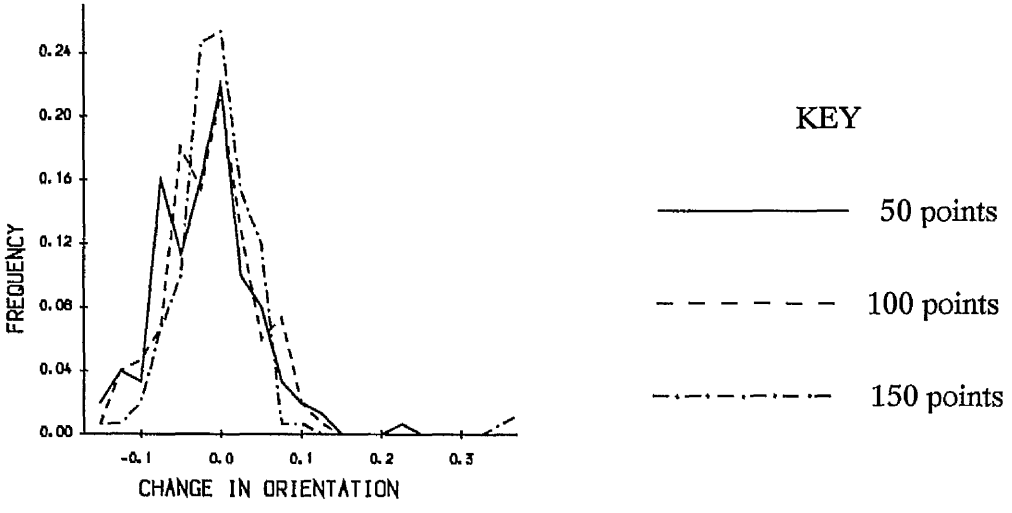


Figure 6.6 :- Empirical probability distribution for orientation under the null hypothesis of no change.

q	50 points	100 points	150 points
60	0.04786	0.03877	0.02690
80	0.06770	0.07449	0.04333
90	0.08780	0.09621	0.05465
95	0.10333	0.09966	0.06046
97.5	0.12561	0.10580	0.06777

Table 6.5 :- Percentage points for the empirical probability distribution function relating to the test statistic describing orientation.

The distributions for orientation are similar for various sizes of data set, however for 50 data points, an occasional more diverse result is recorded. From the plot, there appears to be minimal bias in the overall distribution with the mode centred approximately at zero. As for the earlier test statistics, the percentage points for the empirical probability distribution function are defined in table 6.5.

§6.3.1.4 Centroid Displacement

The empirical probability distribution function for centroid displacement is illustrated for the positive set of values, figure 6.7. We are only interested in the distance between two points, not the displacements, the direction vector being analysed in terms of the orientation of the contour. Although the value zero is plausible within the definition of centroid displacement, within the numerical accuracy of the procedure it is never achieved.

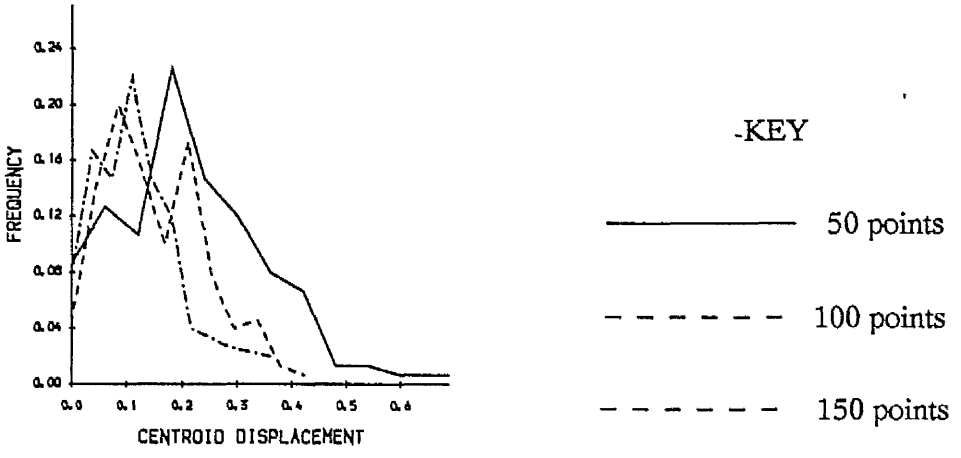


Figure 6.7 :- Empirical probability distribution for centroid displacement under the null hypothesis of no change.

The distribution illustrated is equally applicable to the x, y and x-y distances since the assumption of equality of variances has to be satisfied before the x and y distances can be combined.

A difference exists between the distributions for 50, and 100 and 150 points. For the former the results recorded are more diverse, with a shift of approximately 0.6 standard deviations being recorded under the null hypothesis, at a percentage level of 97.5 and the mode of the distribution occurring at 0.2 standard deviations, the maximum displacement for 100 and 150 points is approximately one third less with the distribution peaking at approximately 0.1 standard deviation. A numerical summary of the empirical probability distribution, in the form of a set of percentage points is given in table 6.6. The table being symmetric about zero.

q	50 points	100 points	150 points
60	0.35560	0.23950	0.19566
80	0.42299	0.29472	0.22753
90	0.45234	0.34955	0.27755
95	0.53601	0.35761	0.30000
97.5	0.60862	0.38940	0.34023

Table 6.7 :- Percentage points for the empirical probability distribution function describing centroid displacement.

§6.3.1.5 Overlap

The empirical probability distribution function of the test statistic under the null hypothesis of no change is described in figure 6.8, with the salient percentage points tabulated in table 6.7.

The results reported display greater diversity for 50 points, than for 100 and 150 points, suggesting that the measures of interest are more concisely defined for larger data sets. A lateral shift in the distributions away from the optimal value of one results with decreasing sample size.

q	50 points	100 points	150 points
60	0.87185	0.90219	0.91391
80	0.91106	0.92295	0.93814
90	0.94053	0.93507	0.95754
95	0.95742	0.95045	0.96398
97.5	0.96436	0.96339	0.97271

Table 6.7 :- Percentage points for the empirical probability distribution function relating to the test statistic describing overlap.

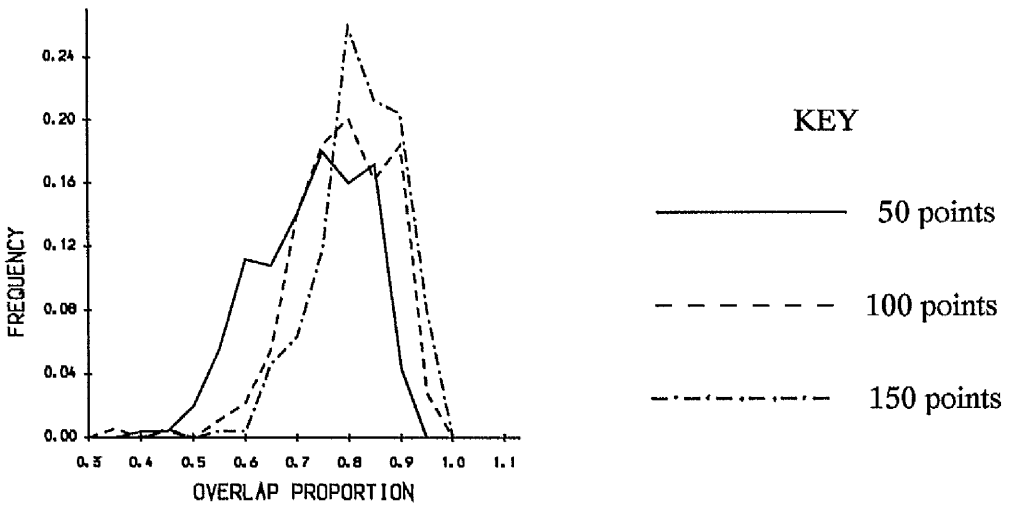


Figure 6.8 :- Empirical probability distribution for overlap under the null hypothesis of no change.

§6.3.2 Distribution under the alternative hypothesis

The alternative hypothesis, H1, can either be of a simple or composite form depending on the nature of the analysis. Typically it will be of a composite form and will refer to there either being change of any form i.e, a two-sided test or alternatively, the analyst may believe that a specific type of change is most probable hence a one-sided test is applicable.

The standard situations of both one and two-sided tests are relevant for test statistics pertaining to scalar and angular change. However for the test statistics describing translation i.e. overlap and centroid displacement, only one form of the alternative hypothesis is applicable i.e. one sided tests, since the statistics have one finite bound.

Once the alternative hypothesis have been stated, the relevant rejection regions, table 6.8 may be described.

H_0	H_1	Rejection region
$A_{1j} = A_{2j}$	$A_{1j} < A_{2j}$	$\left\{A_j : A_j < \frac{1}{A(N, \max(r_n) : q)}\right\}$
$A_{1j} = A_{2j}$	$A_{1j} > A_{2j}$	$\{A_j : A_j > A(N, \max(r_n) : q)\}$
$A_{1j} = A_{2j}$	$A_{1j} \neq A_{2j}$	$\left\{A_j : \max\left(A_j : \frac{1}{A_j}\right) > A(N, \max(r_n) : q)\right\}$
$OR_{1j} = OR_{2j}$	$OR_{1j} < OR_{2j}$	$\{OR_j : OR_j < -OR(N, \max(r_n); q)\}$
$OR_{1j} = OR_{2j}$	$OR_{1j} > OR_{2j}$	$\{OR_j : OR_j > OR(N, \max(r_n); q)\}$
$OR_{1j} = OR_{2j}$	$OR_{1j} \neq OR_{2j}$	$\{OR_j : \lvert OR_j \rvert \geq OR(N, \max(r_n); q)\}$
$P_{1j} = P_{2j}$	$P_{1j} < P_{2j}$	$\left\{P_j : P_j < \frac{1}{P(N, \max(r_n); q)}\right\}$
$P_{1j} = P_{2j}$	$P_{1j} > P_{2j}$	$\{P_j : P_j < P(N, \max(r_n); q)\}$
$P_{1j} = P_{2j}$	$P_{1j} \neq P_{2j}$	$\left\{P_j : \max\left(P_j, \frac{1}{P_j}\right) \geq P(N, \max(r_n); q)\right\}$
$OV_j = 1$	$OV_{1j} < OV_{2j}$	$\{OV_j : OV_j < OV(N, \max(r_n); q)\}$
$CD_j = 0$	$CD_{1j} \neq CD_{2j}$	$\{CD_j : \lvert CD_j \rvert > CD(N, \max(r_n); q)\}$

where N = number of points

$\max(r_n)$ = maximum level of random noise incorporated into the system

Table 6.8 :- Various forms of the alternative hypothesis.

By tabulating the numerical values of the empirical probability distribution function for the percentage points 60, 80, 90, 95 and 97.5, the critical values of the distribution for various significance levels are easily interpreted from tables 6.3 - 6.7.

§6.3.3 Power of the tests

The ability of the test statistic to correctly reject the null hypothesis for various values is described in terms of the power function of the test procedure.

Suitable choices for the alternative hypothesis were defined by initially testing a coarse grid of values and on the basis of these results, the critical regions were examined more closely. This was especially critical for assessing the test statistics based on area and perimeter.

Small changes in the area or perimeter ratios are seen to have a major effect on the resultant distribution. This extreme sensitivity of the test statistic may be shown to be a result of the underlying distribution of the test statistic.

Asymptotically, it may be assumed that both area and perimeter have an approximate normal distribution. On this premise and the assumption of independence between the two measures, Geary (1930) showed the ratio of two independent normal variables, $v = \frac{x}{y}$, has frequency function:-

$$f(v) = \frac{1}{\sqrt{2\pi}} \frac{\mu_y \sigma_x^2 + \mu_x \sigma_y^2 v}{\left(\sigma_x^2 + \sigma_y^2 v^2\right)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\frac{\left(\mu_x - \mu_y v\right)^2}{\sigma_x^2 + \sigma_y^2 v^2}\right\}$$

For constant variance and small changes in the mean of the two distributions i.e. small scalar changes, major shifts in the resultant distribution function occur for the ratio of the two normal distributions. Figure 6.9 illustrates the empirical distributions for the ratio of two normal distributions for 50 data points, whose variances are constant but means differ by 2 and 4% respectively. This illustration confirms the apparent sensitivity of the statistics to change. Using this information, an appropriate set of scalar changes were investigated.

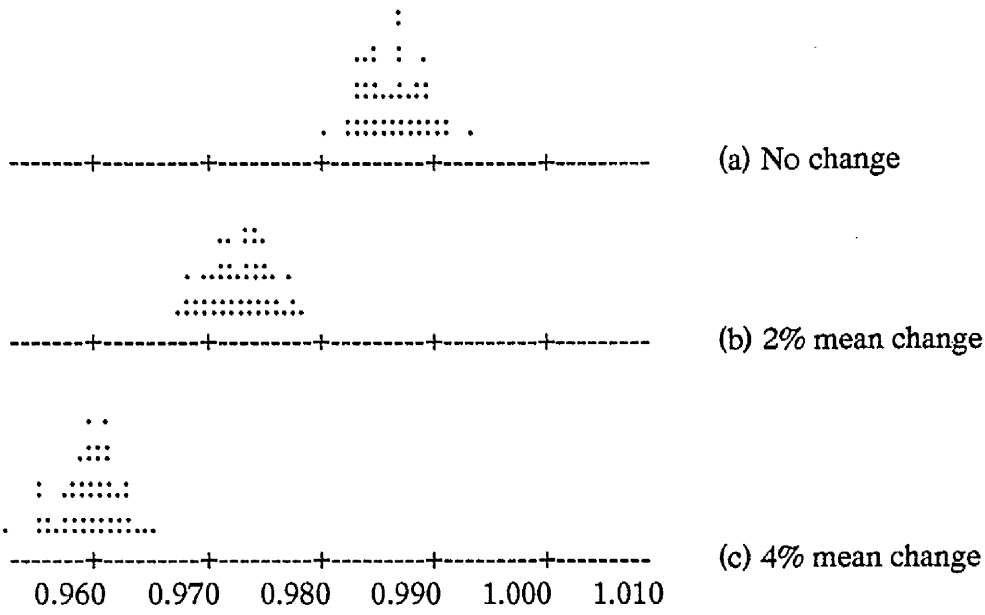


Figure 6.9 : Sensitivity of the scalar metrics.

For the remaining three statistics, the levels of change were constrained by their natural, numerical bounds, i.e. for orientation $(0,\pi]$, centroid displacement $(0,2\sigma]$ and overlap, (0,100%), where 0% is total separation and for 100%, contour A encompasses contour B. For each of these ranges an initial coarse set of values was investigated; for overlap, 90%, 70%, 50%, 30%. Once the critical zones were located, additional points were selected in that region.

§6.3.3.1 Area

Figure 6.10 describes the power curve for area. The areal constant, k, relates to areal change in the following manner:-

$$(1+(k\times 0.002))^2 \quad \text{where} \quad k>0$$

In real terms, the range of change examined was up to 2.4%.

An areal expansion/shrinkage of greater than 2%, for a 95% level of confidence, will be discerned to be attributable to some physical process and not simply a by-product of the surface fitting technique. The results for 50, 100 and 150 points are of a similar order.

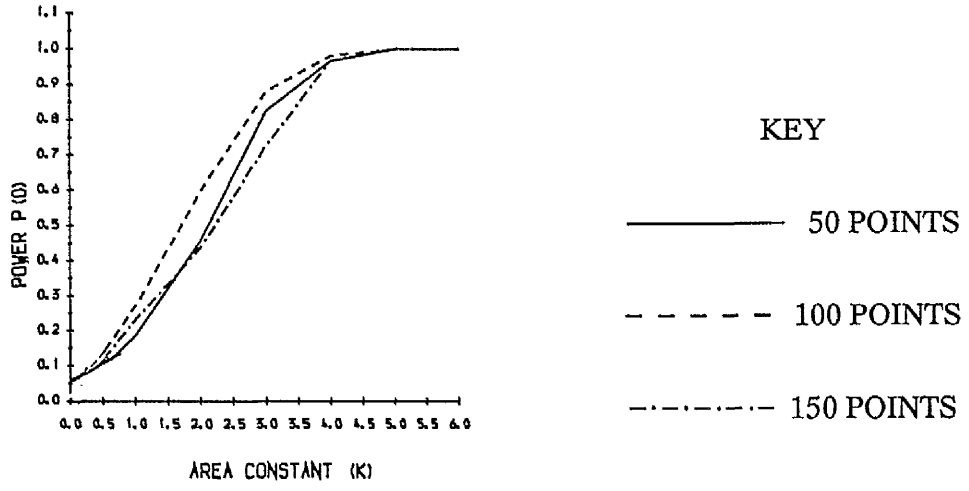


Figure 6.10 :- Power curve for area with no random noise.

§6.3.3.2 Perimeter

A similar procedure to that for area was performed. In this instance, k effectively introduces changes of the order 0 to 3% i.e.

$$(1+(k\times 0.002)) \qquad \text{where} \quad k>0$$

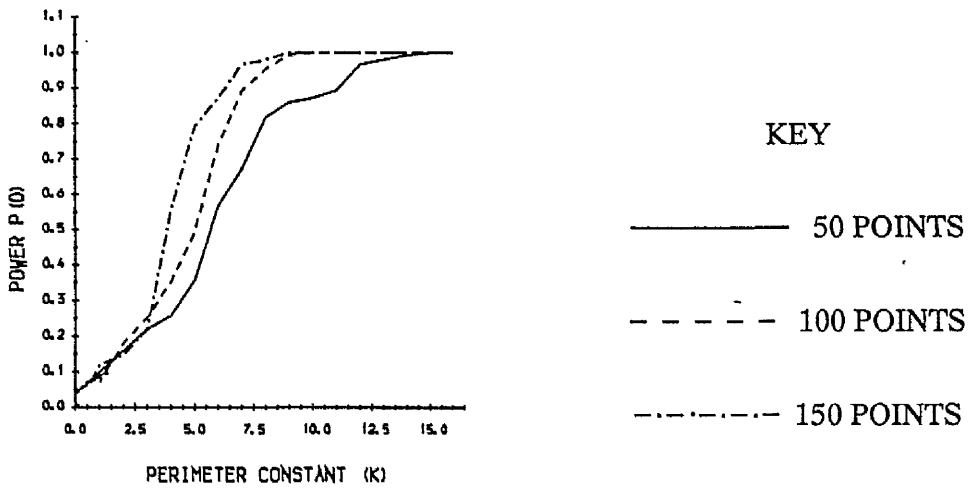


Figure 6.11 :- Power curve for perimeter with no random noise.

From the diagram of the power curves, figure 6.11, the results for 100 and 150 points are reasonably similar, however the power function, for 50 points, shows that greater

changes are required for change to be defined as statistically significant and not to be solely a facet of the surface fitting procedure.

§6.3.3.3 Orientation

Throughout the chapter, examination of the various test statistics has been based on the bivariate normal distribution. Depicting the distribution as a contour plot, the contours are theoretically symmetric about the major and minor axis when no random noise is present in the system. On this assumption the change in orientation should be assessed between $(0,\frac{\pi}{2})$ radians.

The response of the statistic for orientation is similar to that for area, angular change of between 20^0 and 25^0 is attributable to the surface fitting procedure, for a power of 0.05. Changes outside this range may be considered to be a result of external factors, figure 6.12

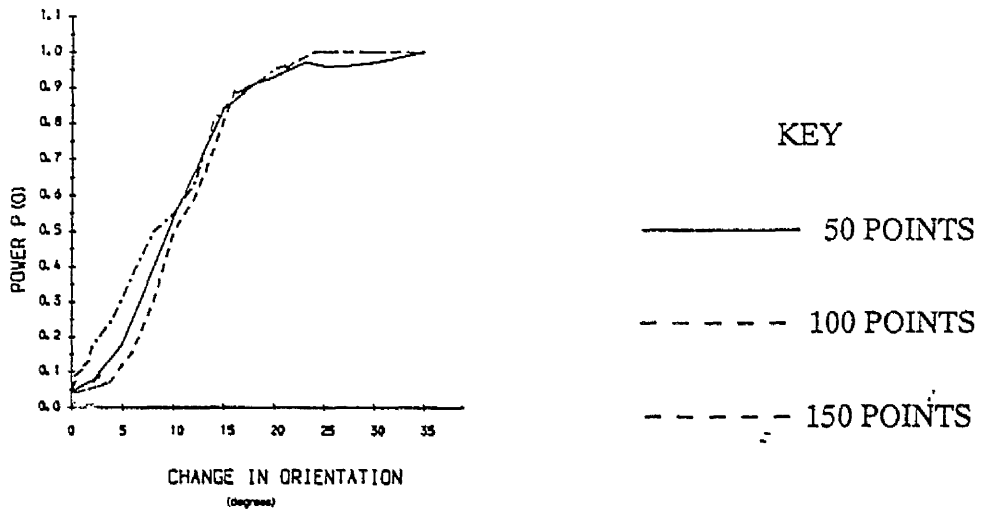


Figure 6.12 :- Power curve for the change in orientation with no random noise.

§6.3.3.4 Centroid Displacement

The resultant power curves are shown in figure 6.13. The measurement of centroid displacement along the x-axis is defined in terms of standard deviations i.e. k=1, describes a shift of one standard deviation.

In terms of the metric describing centroid displacement, it is apparent that the larger the data set, the more precise the results i.e. less variability is introduced due to the surface fitting methodology. A marked drop in the power for 50 data points from that recorded for 100 and 150 points is seen from figure 6.13.

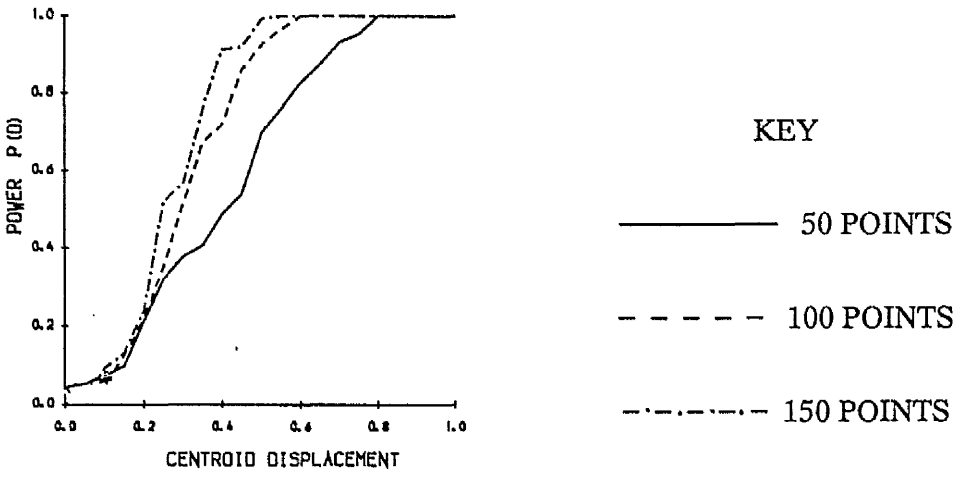


Figure 6.13:- Power curve for centroid displacement with no random noise.

§6.3.3.5 Overlap

For overlap, the power curves illustrated in figure 6.14, display differing behavioural patterns with 50 points having the weakest power, as expected.

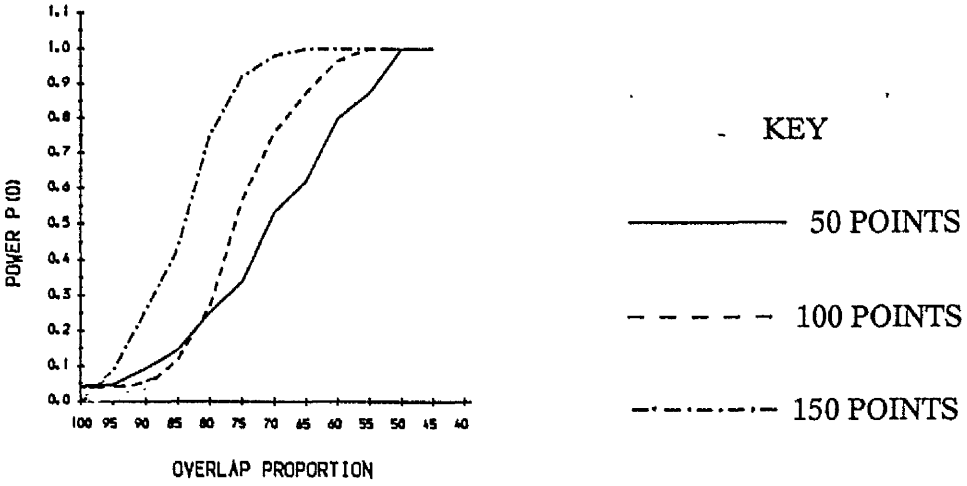


Figure 6.14 :- Power curve for overlap proportion with no random noise.

For the null hypothesis to be rejected, at the 95% confidence level, an overlap of the order of at most 50% is required for 50 data points and 60% and 75% for 100 and 150 points respectively.

It appears that problems may arise with the introduction of random noise, particularly for 50 points, resulting in situations where noise attributable to the surface fitting procedure will dominate the result, hence noise introduced from the surface fitting procedure will be inseparable from changes due to physical processes.

§6.3.4 Summary of no random noise

The initial part of the chapter has examined the simplest situation where no noise is present in the measurements. Even for this simple scenario, there are indications that the statistics describing orientation and overlap, may only be applicable where noise levels are firstly minimised and secondly, the original data set is as large as practically viable, both in terms of the economics and further, the experimental effort required.

Even where noise is absent and no ambiguity arises in defining the angle of orientation or the area of overlap, changes of the order of 20^0 and an overlap of 50% are required before we can be certain that change is a result of some underlying process, for 50 data points.

For the two scalar measures area and perimeter and a mesh network comprising approximately 4000 points, the level of change which signifies external processes are acting is surprisingly small e.g. for a contour of area 100 units, an increase to 102 units leads to the rejection of the null hypothesis for a significance level of 0.05. The introduction of noise into the system is liable to result in a much greater change being required before the null hypothesis is rejected.

Finally centroid displacement responds in a fairly stable manner with values of over 0.4 standard deviations being significant in the case of 100 and 150 points and 0.8 standard deviations for 50 points.

§6.4 Random Noise

Section 6.3 examined the simplest scenario where the measurements were not subject to random noise. In practice this is an unrealistic situation as measurements will always contain some form of inaccuracy.

This may be due to measurement error, bias or human error. Bias or persistent error is outwith the control of the statistician. The operator should recognise the symptoms when recording the measurements. Failure to do this will lead to an over/under estimation of results. The overall shape of the contour surface should be unaffected, other than be raised or lowered by a constant factor, metaphorically speaking. The worst scenario is the case where the operator recognises the occurrence of bias, recalibrates the instrument, continues making measurements but fails to correct the earlier results for bias. The surface is then a mis-match of results and worthless.

The only form of error in which we are interested is that due to measurement error, which is controlled by the accuracy of the instrument and human limitations, typically this form of error will be quantifiable, although over or under estimation may occasionally result. Typically in field measurements, the level of inaccuracy will be of a higher order than that pertaining to measurements recorded in the laboratory.

The noise levels which were believed to be most typical are those of 5, 15 and 25%. In this section, the analysis described in section 6.3 will be repeated but for the three levels of noise.

Two possible forms of noise may be introduced into a spatial system, either in terms of:

1. z-ordinate
2. x, y-ordinate.

Interest is specifically in the former of these error forms. A similar procedure to that described in section 6.2.5 was undertaken, step 1 was replaced by :-

1. Simulate data points $\underline{x}_{1R}, \underline{x}_{2R}, \ldots, \underline{x}_{NR}$ for N = 50, 100 and 150 points using the NAG subroutine G05EAF :-

$$\underline{x}_{iR} \sim N(\underline{0}, \Sigma) + N(0, \Sigma_1)$$

where $\Sigma = \begin{bmatrix} 1.0 & k \\ k & 1.0 \end{bmatrix}$ $\Sigma_1 = \begin{bmatrix} s & 0 \\ 0 & s \end{bmatrix}$

where $k = \begin{cases} 0.0 & \text{area} \\ 0.0 & \text{perimeter} \\ 0.9 & \text{orientation} \\ 0.5 & \text{centroid displacement} \\ 0.0 & \text{overlap} \end{cases}$ $s = \begin{cases} 0.05 \\ 0.15 \\ 0.25 \end{cases}$

§6.4.1 Distribution of the test statistic under the null hypothesis

As for the case of the null hypothesis where no random noise is present, the statement for the null hypothesis is the same i.e. no change.

§6.4.1.1 Area

The underlying empirical probability distribution function for the areal ratio is given in figure 6.15 with the relevant percentage points being tabulated in 6.9.

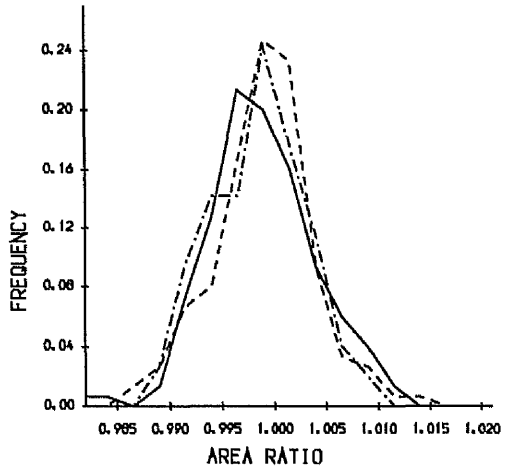


Figure 6.15 (a) :- 5% random noise

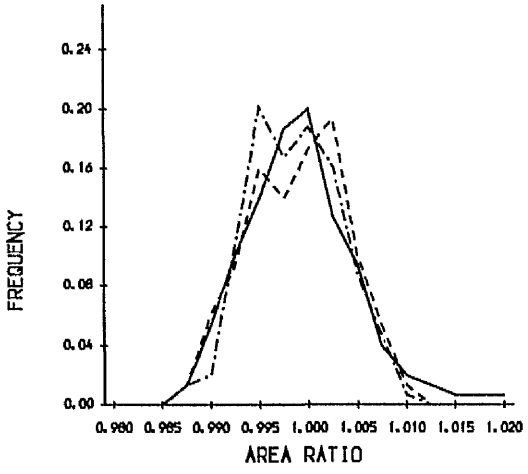
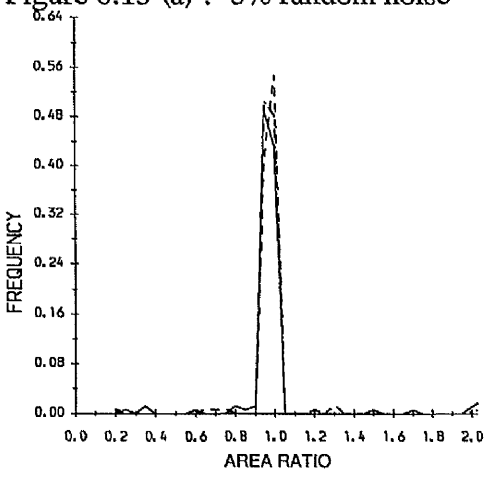


Figure 6.15 (b) :- 15% random noise



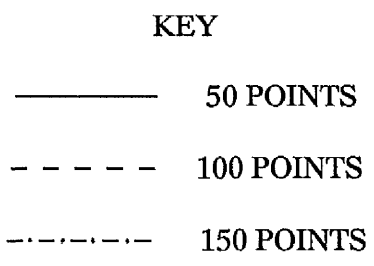


Figure 6.15 (c) :- 25% random noise

Figure 6.15 :- Empirical probability distribution for areal change under the null hypothesis of no change.

q	r_n	50 points	100 points	150 points
60	5%	1.00432	1.00361	1.00365
80	5%	1.00718	1.00554	1.00556
90	5%	1.00896	1.00781	1.00708
95	5%	1.01126	1.00945	1.00781
97.5	5%	1.01191	1.01177	1.00947

60	15%	1.00451	1.00451	1.00422
80	15%	1.00704	1.00660	1.00598
90	15%	1.00933	1.00835	1.00814
95	15%	1.01198	1.00923	1.00923
97.5	15%	1.01301	1.01006	1.01162
60	25%	1.00446	1.00452	1.00419
80	25%	1.00840	1.00742	1.00667
90	25%	1.01290	1.00972	1.00829
95	25%	1.58624	1.01202	1.01163
97.5	25%	2.88234	1.15785	1.06788

Table 6.9 : Percentage points for areal change under the null hypothesis.

A strong similarity is displayed between the results for 50, 100 and 150 points. Increasing the level of random noise, results in an elongation of the tails of the distribution. This is particularly prevalent for noise levels of the order 25%. For the case of 50 points, aberrant values are noted, the 97.5 percentage point for 50 points is almost double that for 95%. Increasing the number of simulations will possibly draw the value in slightly, but it will still be some distance away from the 95% point, since the distribution is naturally long tailed.

§6.4.1.2 Perimeter

The distribution under the null hypothesis in the presence of noise responds in a similar manner to that under the hypothesis of no change in the absence of no noise, figure 6.16.

The difference again materialises in the length of the tails of the distributions between data sets of differing sizes. Much closer accord is displayed between 100 and 150 points than for 50 points for 15% and 25% random noise. As for the case of areal change, we see that particularly for 50 data points, the distribution is elongated, this is accordingly reflected in the values of the percentage points, table 6.10. The results reported for 5% noise levels correspond closely to those evaluated for the case of no random noise in section 6.3.1.1. For 25% noise, the percentage points display a greater level of contrast between 100 and 150 points than for either 5% or 15%, in the tails of the distribution.

The introduction of noise into the system does not appear to bias the results, the mode of the distribution, for all permutations of size and noise, is still in the vicinity of one.

The shapes of the distributions remain constant for all permutations of noise level and data set size.

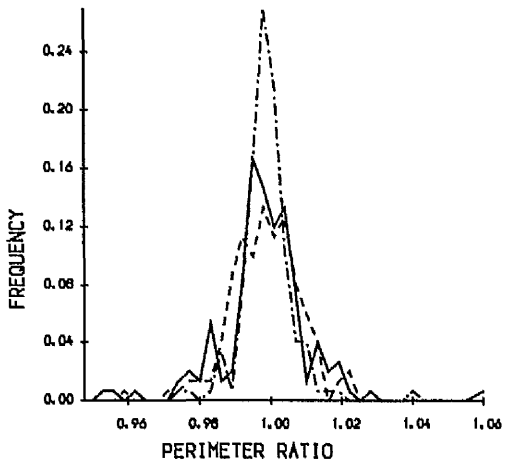


Figure 6.16 (a) :- 5% random noise

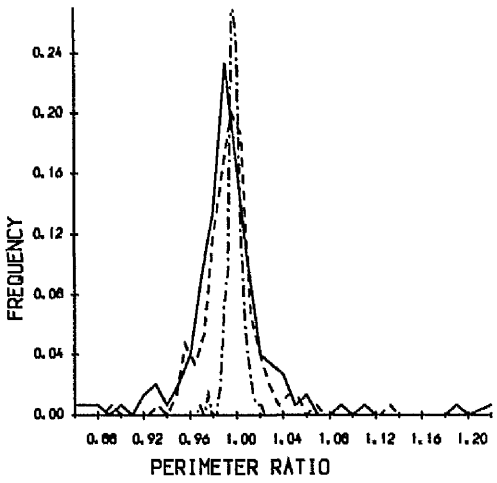
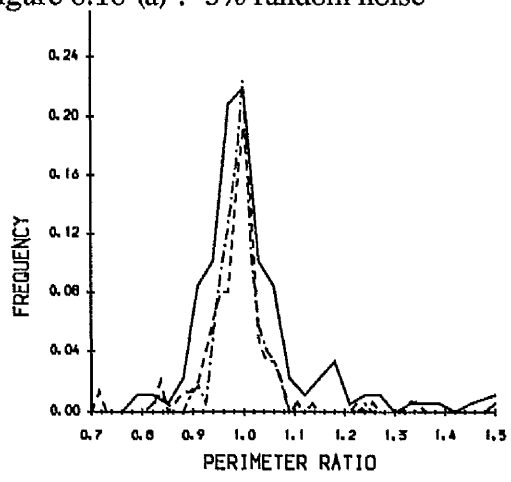


Figure 6.16 (b) :- 15% random noise



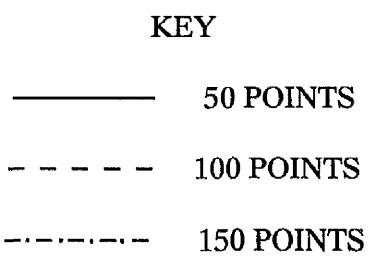


Figure 6.16 :- 25% random noise

Figure 6.16 :- Empirical probability distribution for the perimeter ratio under the null hypothesis of no change.

q	r_n	50 points	100 points	150 points
60	5%	1.00669	1.00789	1.00404
80	5%	1.01354	1.01191	1.00722
90	5%	1.01793	1.01563	1.01025
95	5%	1.02031	1.02303	1.01316
97.5	5%	1.02677	1.02663	1.01774
60	15%	1.01346	1.00972	1.00912
80	15%	1.03051	1.02123	1.01897
90	15%	1.04924	1.03320	1.02340
95	15%	1.07506	1.05203	1.02880
97.5	15%	1.15606	1.06293	1.05491
60	25%	1.06922	1.02781	1.02380
80	25%	1.17126	1.05261	1.05058
90	25%	1.24585	1.07865	1.06834
95	25%	1.58095	1.24007	1.08585
97.5	25%	1.65233	1.34295	1.17154

Table 6.10 : Percentage points for perimeter change under the null hypothesis.

§6.4.1.3 Orientation

The empirical probability distribution functions for orientation are presented in figure 6.17 with the analytical values given in table 6.11 for various percentage points.

Once again the main contrast between different sizes of data set is in the tails of the distribution. For increasing sizes of data set, the tails are terminated more sharply. This feature extends over all levels of noise. The mode of the distribution occurs at approximately zero for all the distributions, indicating that bias has not been introduced into the system.

The distribution of the test statistic for 25% noise is wider than for either of the other two levels and the tails are generally fatter.

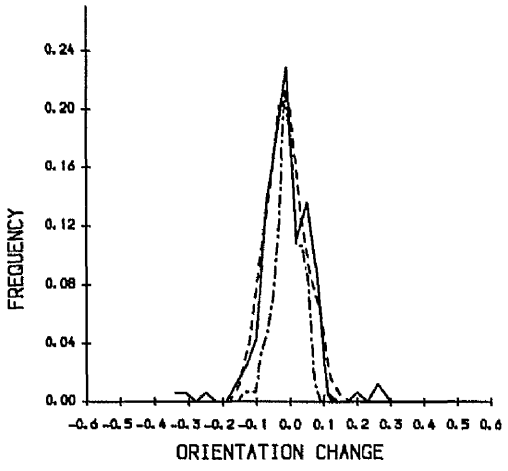


Figure 6.17 (a) :- 5% random noise

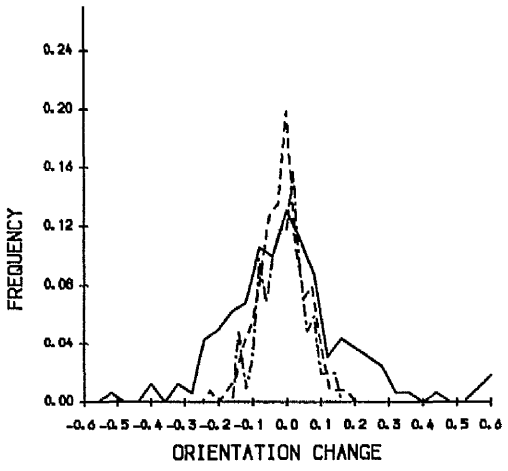
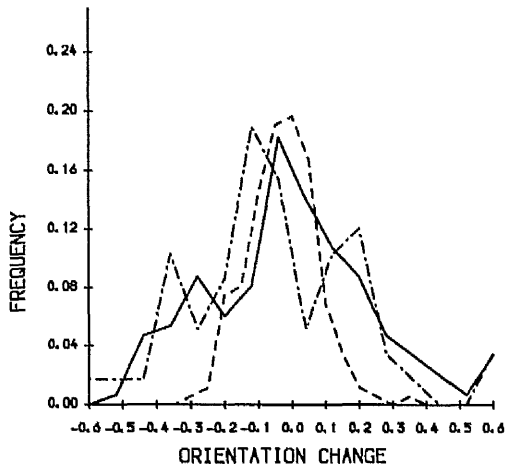


Figure 6.17 (b) :- 15% random noise



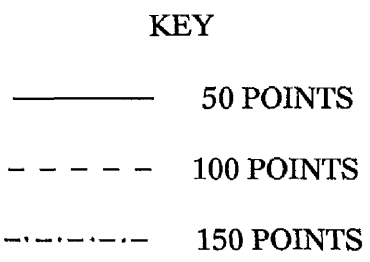


Figure 6.17(c) :- 25% random noise

Figure 6.17 :- Empirical probability distribution function for orientation change under the null hypothesis of no change.

q	r_n	50 points	100 points	150 points
60	5%	0.063614	0.055197	0.040060
80	5%	0.082298	0.078591	0.056560
90	5%	0.102002	0.097638	0.066350
95	5%	0.108241	0.117996	0.071623
97.5	5%	0.235798	0.124846	0.096091
60	15%	0.113325	0.051748	0.050417
80	15%	0.210964	0.833154	0.076049
90	15%	0.281352	0.105859	0.091681
95	15%	0.348310	0.119322	0.109681
97.5	15%	0.503490	0.155222	0.112509
60	25%	0.225310	0.059556	0.080961
80	25%	0.303834	0.093346	0.103648
90	25%	0.450519	0.145384	0.133834
95	25%	0.550824	0.169104	0.143449
97.5	25%	0.689743	0.205192	0.161665

Table 6.11 :- Percentage points for orientation change under the null hypothesis.

§6.4.1.4 Centroid Displacement

Once again for 5% and 15% random noise, a strong similarity in behaviour to that of the case for no random noise emerges. The results for 50 points being more diverse, figure 6.18.

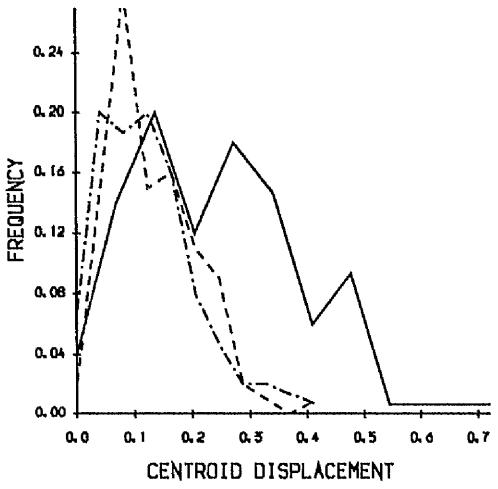


Figure 6.18 (a) :- 5% random noise

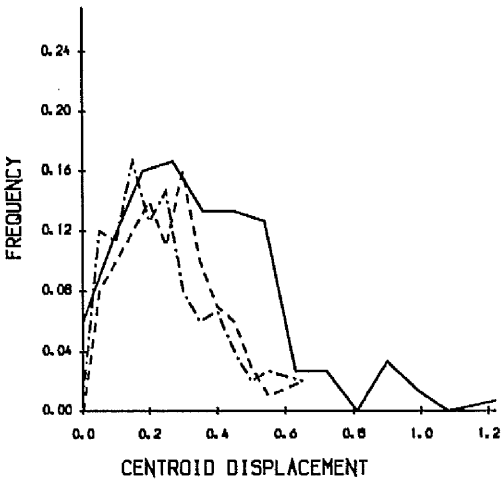
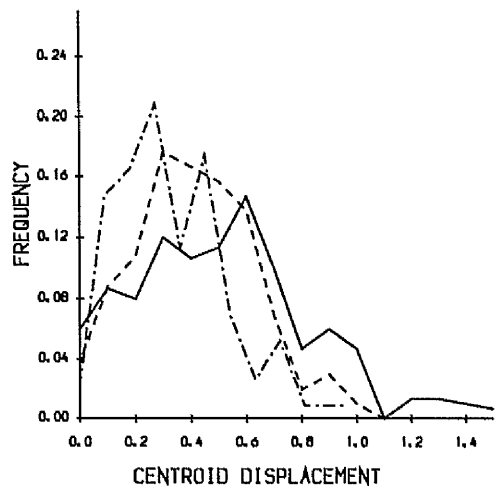


Figure 6.18 (b) :- 15% random noise



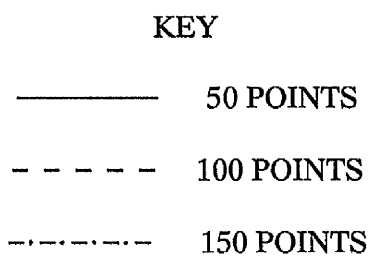


Figure 6.18 (c) :- 25% random noise

Figure 6.18 :- Empirical probability distribution for centroid displacement under the null hypothesis of no change.

By the time, the noise level attributable to the data is of the order of 25%, differences emerge between the three sizes of data set. The percentage points, table 6.12, confirm this behaviour. The tails are accordingly long for 50 data points. Although the distributions are long-tailed, the percentage points are contained within the limits of two standard deviations for all levels of random noise.

q	r_n	50 points	100 points	150 points
60	5%	0.38080	0.20772	0.20296
80	5%	0.48054	0.25640	0.25621
90	5%	0.51861	0.27862	0.30236
95	5%	0.53936	0.32125	0.34702
97.5	5%	0.61929	0.36598	0.39409
60	15%	0.35560	0.23949	0.19566
80	15%	0.42299	0.29472	0.22753
90	15%	0.45234	0.34955	0.27755
95	15%	0.53601	0.35761	0.30000
97.5	15%	0.60862	0.38940	0.34024
60	25%	0.78525	0.67000	0.52101
80	25%	0.94089	0.75300	0.62418
90	25%	1.02929	0.90167	0.73855
95	25%	1.20269	0.94069	0.79367
97.5	25%	1.33353	1.07630	0.85094

Table 6.12 :- Percentage points for centroid displacement under the null hypothesis.

§6.4.1.5 Overlap

In terms of the overlap function, the empirical probability distribution function is marginally different for each size of data set, figure 6.16. The saliant values being recorded in table 6.19.

The tails of the distribution are longer and fatter with increasing levels of noise. The mode of the distribution does not lie at 1.0 , the expected value, if no change has resulted. The underlying reason for this potential form of bias is because one is the upper bound of the distribution and with the introduction of noise, it would be anticipated that the two contours could be entirely contained within each other. For 5% noise and all three data sets, the mode lies at 75%, similarily for 100 and 150 points for 15% noise. For both 15% and 25% noise, for 50 data points, the mode is more difficult to select as the distribution is much flatter in shape, but it lies approximately at 40%.

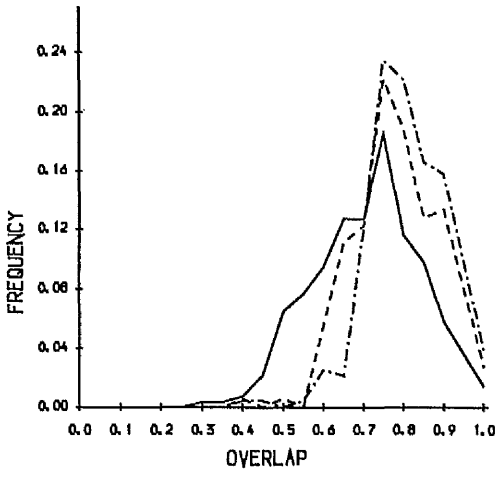


Figure 6.19 (a) :- 5% random noise

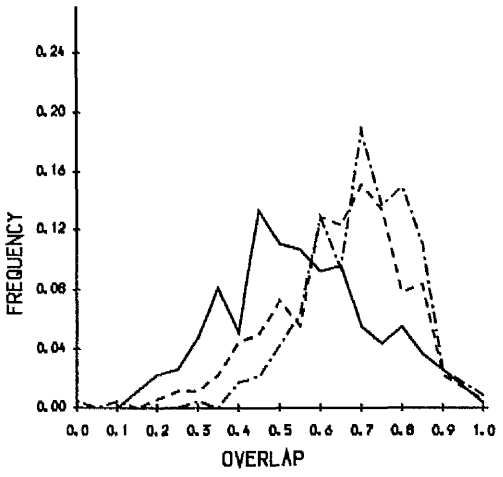
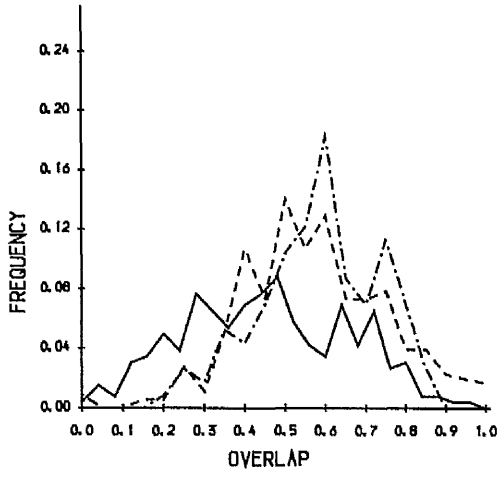


Figure 6.19 (b) :- 15% random noise



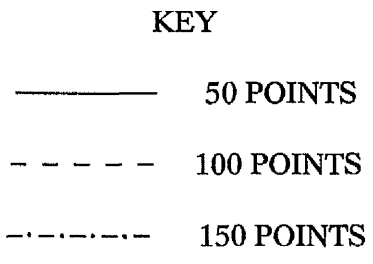


Figure 6.19 (a) :- 25% random noise

Figure 6.19 :- Empirical probability distribution for overlap under the null hypothesis of no change.

q	r_n	50 points	100 points	150 points
60	5%	0.833994	0.883402	0.898660
80	5%	0.887435	0.914999	0.922489
90	5%	0.921110	0.932646	0.940530
95	5%	0.938217	0.948872	0.956583
97.5	5%	0.950966	0.960851	0.962937
60	15%	0.714501	0.789122	0.826559
80	15%	0.821925	0.854482	0.869114
90	15%	0.869977	0.887272	0.886470
95	15%	0.905789	0.899752	0.912062
97.5	15%	0.924986	0.920836	0.925357
60	25%	0.672920	0.723764	0.732612
80	25%	0.742420	0.800562	0.787022
90	25%	0.798821	0.854832	0.818849
95	25%	0.834341	0.892620	0.843812
97.5	25%	0.862852	0.909092	0.863835

Table 6.13 :- Percentage points for overlap under the null hypothesis.

§6.4.2 Power Curves

§6.4.2.1 Area

For areal change, the power curves, figure 6.20, illustrate a strong similarity between the results for 100 and 150 points for 5% and 15% random noise, whilst for 25%, differences materialise between all three data sets. The results for 50 points in particular are indicative of weaker power. Increasing the noise level results in a falling off of the overall power i.e. a lateral shift in the curves has resulted, hence greater areal changes are required for differences, between contour areas, to be defined as statistically significant and not simply a result of the surface fitting procedure.

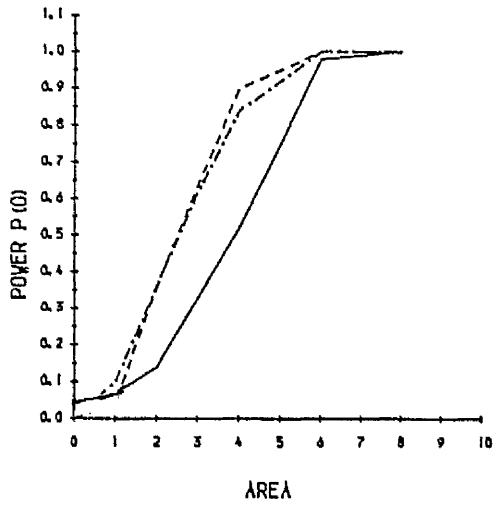


Figure 6.20 (a) :- 5% random noise

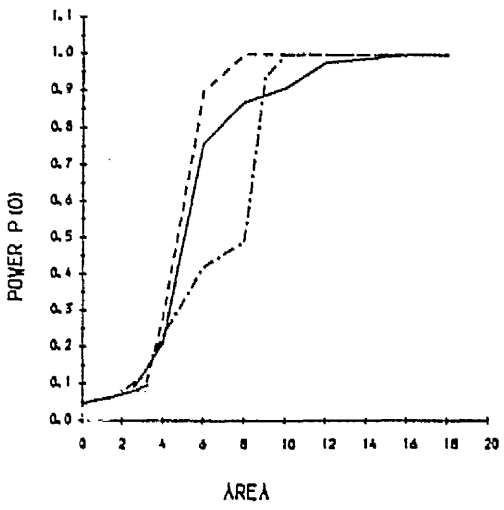
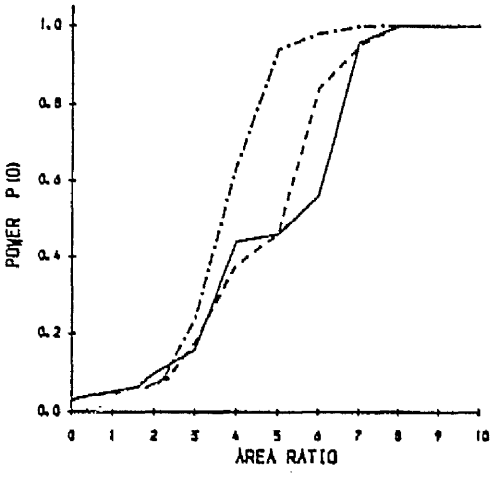


Figure 6.20 (c) :- 25% random noise



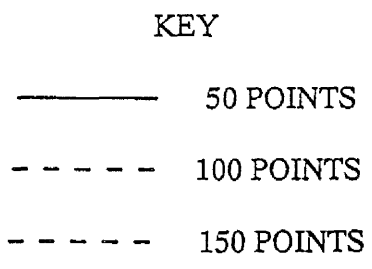


Figure 6.20 (b) :- 15% random noise

Figure 6.20 :- Power curves for area in presence of random noise.

§6.4.2.2 Perimeter

A strong pattern emerges in the power curves for perimeter, figure 6.21. The curves for 50 points consistently display a weaker level of power than either 100 or 150 points, whilst only marginal differences emerge between the results for 100 and 150 points.

Increasing the noise level, results in a sharp decrease in the power. For noise levels of 25% compared to those of 15% an approximate doubling of the perimeter ratio is recorded for a power of 95%.

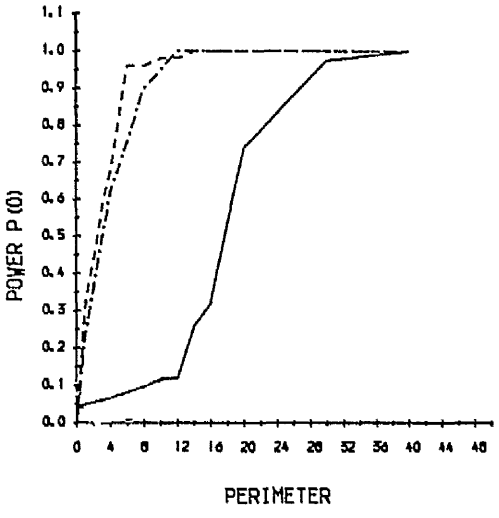


Figure 6.21 (a) :- 5% random noise

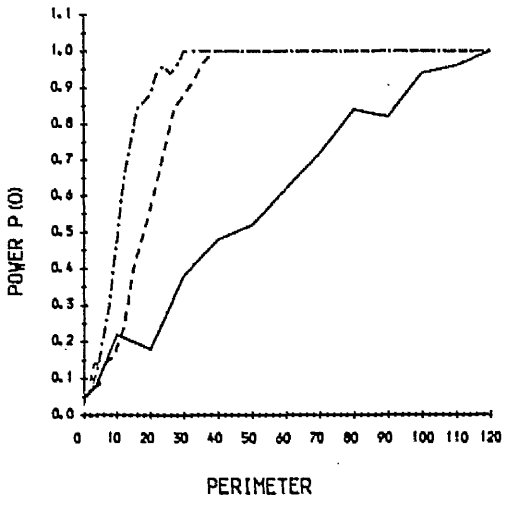
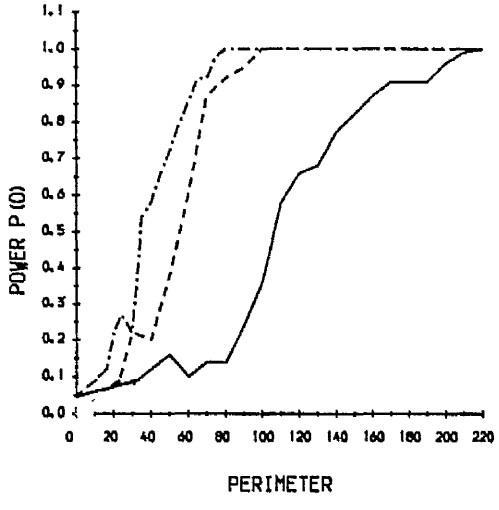


Figure 6.21 (b) :- 15% random noise



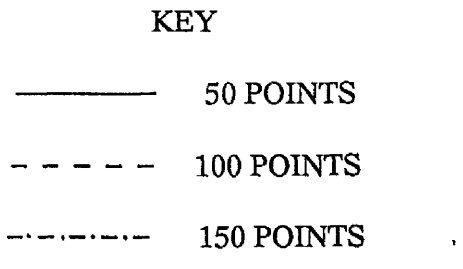


Figure 6.21 (c) :- 25% random noise

Figure 6.21 :- Power curves for perimeter in presence of random noise.

§6.4.2.3 Orientation

The power curves for orientation change are given in figure 6.22 for all three levels of noise. The behaviour for 5% differs to that for both 15% and 25%. A fairly strong accord exists between the results for 0% and 5% random noise. Generally, a slightly greater shift is required for change to be recognised as the result of external influences for 5%. The trends across the data sets are similar.

For 15% and 25% noise a different pattern emerges, the behaviour of the results for 100 and 150 points being similar, whilst for 50 data points, a reduction in the power occurs. Rotations of the order of 90^0 and 45^0, for 50 and 100 and 150 points respectively, are required, for a power of 0.05, before it is safe to say that change is not solely a facett of the surface fitting procedure.

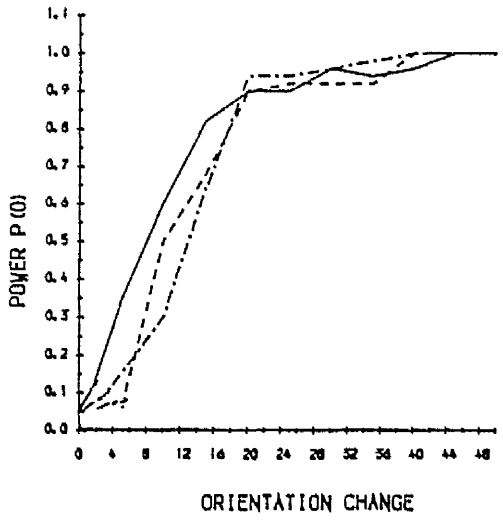


Figure 6.22 (a) :- 5% random noise

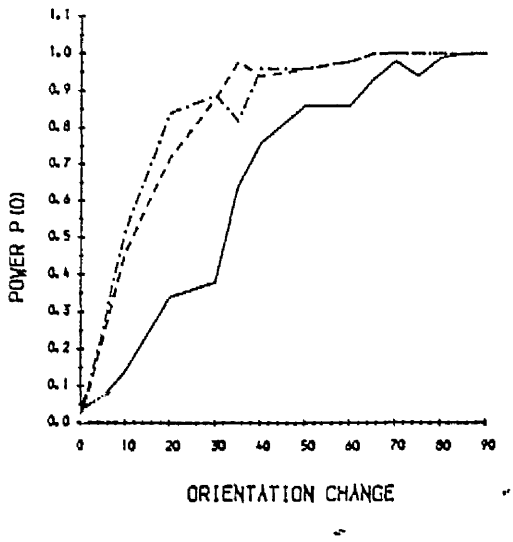


Figure 6.22 (b) :- 15% random noise

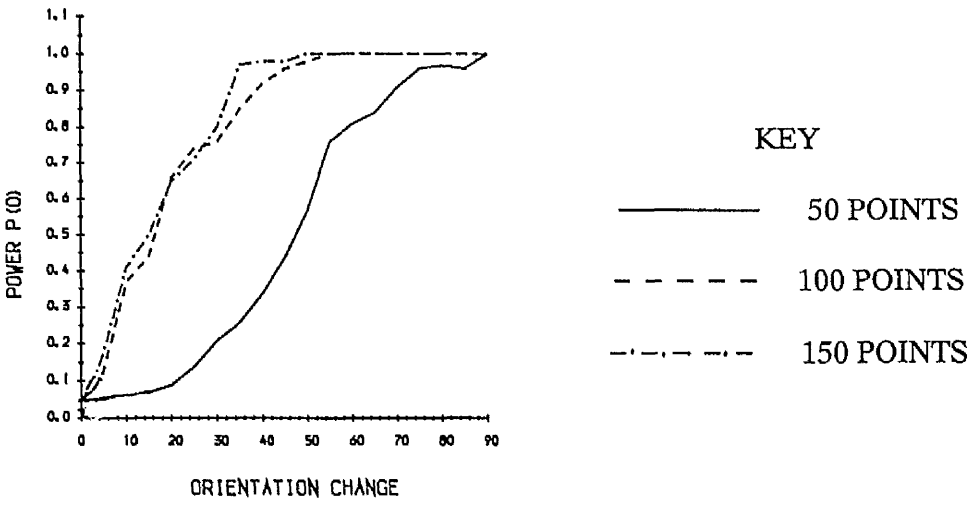


Figure 6.22 (c) :- 25% random noise

Figure 6.22:- Power curves for orientation change in presence of random noise.

§6.4.2.4 Centroid displacement

The ability of the test statistic to reject the null hypothesis was examined for various alternative hypotheses, the resultant power curves are given in figure 6.23.

Noise levels of 5% resulted in power curves which display little difference to those for the situation of no noise; the results for 100 and 150 points display strong accord whilst for 50 points, a much greater shift is required before the null hypothesis is rejected.

Increasing the noise level to 15% and 25% confirms the earlier suspicions that complications may arise in separating change attributable to the surface fitting procedure from that due to 'real' change, especially for the latter case where displacement of the order of two standard deviations is required before we can categorically state that an external process has been acting in effecting a change.

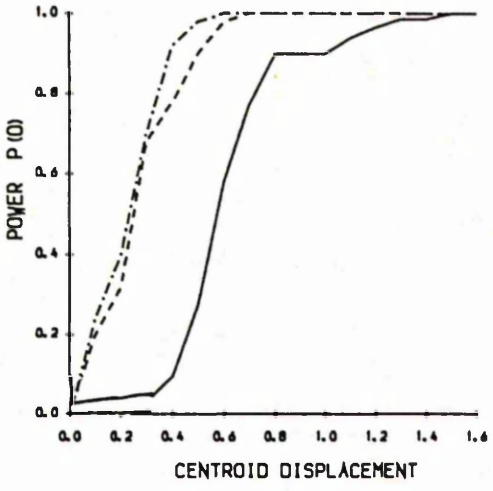


Figure 6.23 (a) :- 5% random noise

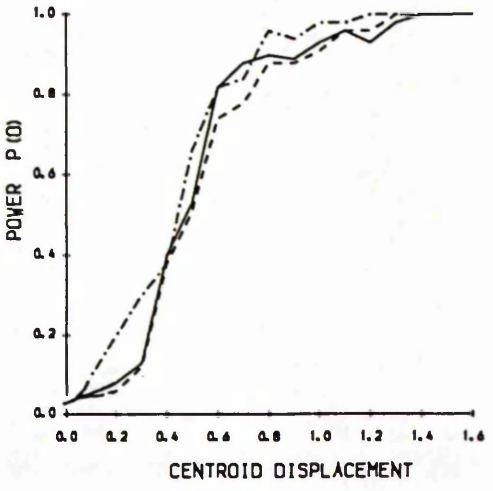
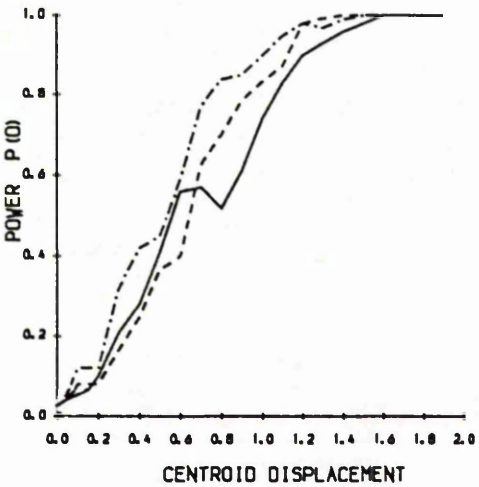


Figure 6.23 (b) :- 15% random noise



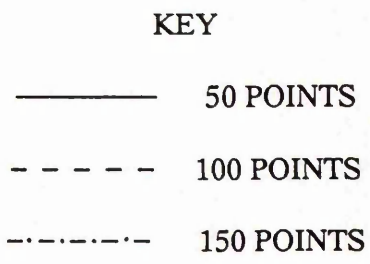


Figure 6.23 (c) :- 25% random noise

Figure 6.23 :- Power curves for centroid displacement in presence of random noise.

§6.4.2.5 Overlap

The behaviour of the power curves given in figure 6.24 is much less well defined than for any of the previous examples. For 5% noise, the results for 100 and 150 points are well defined, with an overlap percentage of the order of 60%, for a confidence level of 95%, required to ensure H_0 is rejected, whilst for 50 points, a value of the test statistic in the region of 40% would result in the rejection of the null hypothesis.

Noise levels of 15% and 25%, give much less smooth pictures, the results for 100 and 150 points still respond in a similar manner, but the power falls away considerably. For 50 points, the two contours have to be virtually disjoint before we can be certain that change is not solely indicative of variability attributable to the surface fitting procedure.

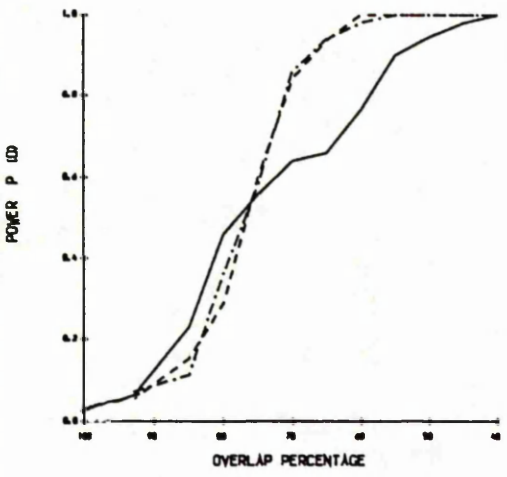


Figure 6.24 (a) :- 5% random noise

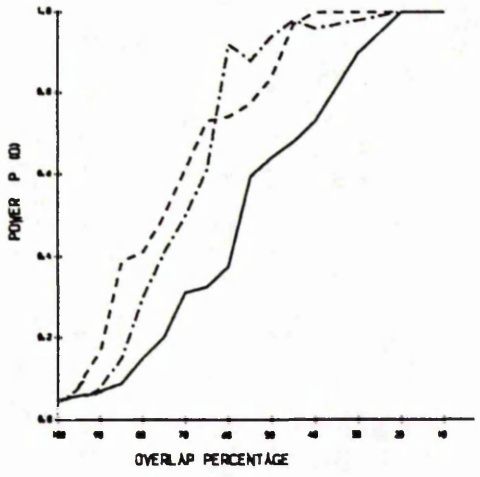
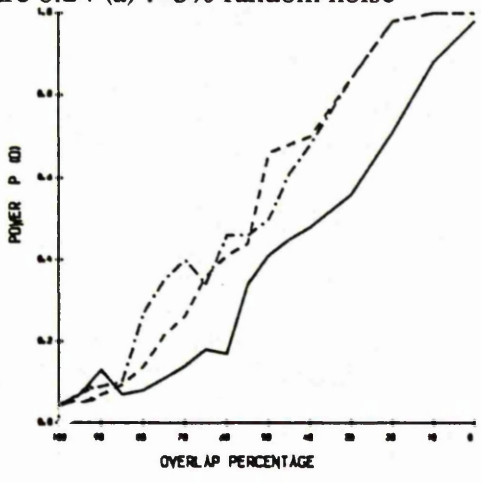


Figure 6.24 (b) :- 15% random noise



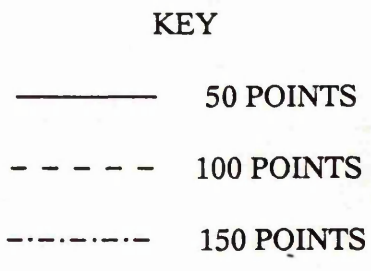


Figure 6.24 (c) :- 25% random noise

Figure 6.24 :- Power curves for overlap in the presence of random noise.

§6.4.2.6 Summary

Overall the introduction of random noise resulted in the underlying probability distribution function of the test statistic behaving as anticipated, increasing the level of random noise resulted in longer-tailed distributions and hence greater changes were required before change due to the external process could be separated from those due to noise in the variable of interest.

In terms of the power curves, the behaviour of the data sets for 5% random noise, responded in a similar fashion to those of no random noise i.e. where the data set size was influential in controlling the probability distribution function for no noise, a similar pattern was reported in the presence of low levels of noise. Also the percentage points recorded were similar in value.

This generalisation does not carry through for 15% and 25% noise. With increasing noise, the power curves for 50 points generally divest themselves from those for 100 and 150 points. The main exception to this was for the test statistic describing centroid displacement where a collation of the curves appeared to result.

One observation to emerge from the preceding analysis was the question of the validity of defining a translation statistic, particularily for 50 points where the level of noise was of the order of 25%. The variability due to the level of measurement noise was liable to dominate any measure of change recorded.

§6.5 Multivariate Hypothesis Testing Procedure

Univariate techniques may be utilised to analyse changes between single contours but we may wish to consider several test statistics simultaneously or to evaluate the same test statistic on a number of different contours. A univariate analysis effectively assumes the parameters are mutually independent between contours or that inter-relationships are unimportant. In the case of overlap and areal ratio, there are situations where the two are identical, hence correlation exists between the measures, unless the analyst is aware of the situation at the subjective stage and substitutes the statistic based on the centre of gravity.

Generally the measures described will display only weak correlation because of the complex nature of the contours. Table 6.14 describes the correlations between the various statistics, for the three levels of noise. On this premise, the assumption of independence is not unrealistic except possibly between the scalar measures area and perimeter and location shift and orientation..

	Area	Perimeter	Orientation	Location shift
Perimeter	0.949			
Orientation	-0.027	-0.049		
Location shift	-0.066	-0.077	0.585	
Overlap	-0.010	-0.014	-0.114	-0.061

Table 6.14 (a) :- Correlations for the various test statistics for 5% noise.

	Area	Perimeter	Orientation	Location shift
Perimeter	0.960			
Orientation	-0.239	-0.224		
Location shift	-0.078	-0.041	0.500	
Overlap	0.346	0.349	-0.274	-0.121

Table 6.14 (b) :- Correlations for the various test statistics for 15% noise.

	Area	Perimeter	Orientation	Location shift
Perimeter	0.951			
Orientation	-0.060	-0.025		
Location shift	-0.144	-0.174	0.078	
Overlap	0.161	0.210	0.074	0.033

Table 6.14 (c) :- Correlations for the various test statistics for 25% noise.

Table 6.14 :- Correlation between the various test statistics for all levels of noise.

The strength of the relationship between location shift and orientation may be attributable in part to the evaluation of the centre of gravity for each of the measures, hence a mathematical relationship exists between the two. Area and perimeter are known to be strongly related for simple structures and in the simulation study, the contours are always defined to be pseudo- elliptical or circular hence the resultant, possibly misleading, strength of the relationship.

A multivariate test may effectively be performed by evaluating for each level of interest, the appropriate measure of overlap, centroid displacement, scalar change and angular change for each set of comparable contours. Each individual contour in surface A being assessed against the 'corresponding' contour from surface B. An order of subjectivity is introduced at this stage, interest only being in those contours which are recognisably comparable in some form size, shape or position.

For a global analysis, the parameter values are not necessarily assumed to be independent. An analysis of this form will be based on a multivariate technique which explicitly models the covariance structure of the data. Theoretically the results are more powerful, but a number of restrictions limit the potential power of the test, section 6.6. A parametric form of analysis for a global test of change is Hotelling's one sample T^2-test, where we are testing :-

$$H0 \; : \; \underline{\mu} = \underline{m}$$

$$H1 \; : \; \underline{\mu}, \Sigma \text{ unconstrained}$$

The vector of fixed constants will depend on the form of investigation. Generally we will be concerned with comparing all aspects of the structure of the two surfaces, hence:-

$$\underline{m} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{i.e.} \quad \begin{pmatrix} \text{no change in area} \\ \text{no change in perimeter} \\ \text{no change in orientation} \\ \text{100\%overlap} \\ \text{no change in centroid displacement} \end{pmatrix} \text{,say.}$$

Where interest is only in specific, pre-defined aspects of a surface, $\underline{m}$ can be accordingly modified. An illustration of the implementation of a global test is given in chapter 7.

§6.6 The Technique

1. Interpolate grid points on a regular grid comprising 4000 nodes.
2. Select number and bounds of class intervals using empirical rules described earlier.
3. Construct contour plot.
4. Calculate the basic geometric quantities area, perimeter, centroid ordinates and orientation.
5. Evaluate the required test statistics.
6. (a) For a local analysis :- carry out the corresponding hypothesis test.
   (b) For a global approach :- collate the required information and perform a Hotelling's one - sample T^2.
7. Interpret the results.

§6.6 Limitations of the Technique

A number of problems are currently associated with both the univariate and multivariate forms of the analysis described in this chapter,which may bias the results.

Restriction 1 :- How should the test statistics be defined for a contour which was whole for surface 1 and becomes disjoint for surface 2 and vice-versa.

Solution 1 :- For area and overlap the total area for each level of interest may be summed for the subsequent analysis, for both the univariate and multivariate approaches.

One plausible method for dealing with the statistic for centroid displacement is to evaluate the statistic for each separate contour and then take a global average. This is particularily suited to the situation given in figure 6.25 (a).

In terms of orientation, the angular change may be equated for each of the disjoint sets of contours and hence compared. Realistically, for perimeter, there is no valid alternative which enables a valid perimeter ratio to be evaluated; combining the

perimeters will result in an over-estimation of the ratio whilst evaluating the ratio separately for each of the areas, tends to cause an under-estimation in both cases. The complexity of the shapes of the contours hinders any proportional combination being undertaken. However as described in the introduction to the section, area and perimeter may be strongly related, hence the inclusion of perimeter as a measure may not be essential.

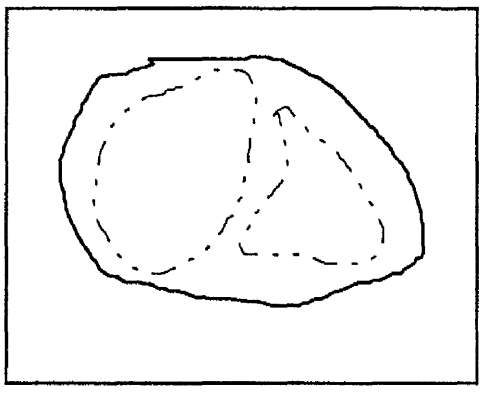


Figure 6.25 (a) Disjoint contour

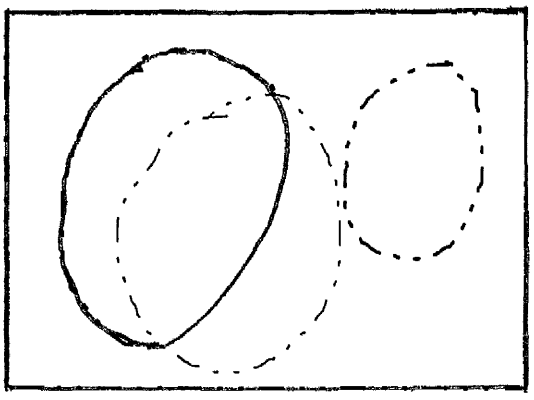


Figure 6.25 (b) Split contours

Figure 6.25 :- Problematical contour shapes.

Restriction 2 :- The second restriction arises where an additional disjoint contour appears in one of the surfaces, figure 6.25 (b).

Solution 2 :- For the univariate approach it may be included in the total area whilst for the multivariate case, unless interest is solely in the entire contour level and not individually comparable contours, it should be omitted from the analysis, hence potentially biasing the result.

Restriction 3 :- A further limitation with the methodology is linked to the general criticism of hypothesis testing, it only answers the question, 'Has there been a change?' or 'Is there an association between variables A and B?'. No account is taken of the size of the difference or asssociation.

Solution 3 :- For practical problems, alternative approaches may enable these questions to be answered, once we have assessed whether an association/change either exists or

has occurred. Secondly, having established a hypothesis testing procedure, the close links between it and interval estimation will enable a confidence interval to be placed on the level of change.

Restriction 4 :- A further limitation of the technique is in terms of the size of the data set and the level of random noise present within the system, these factors apparently control how small a change may be detected over and above that of 'natural' variability.

Solution 4 :- The only viable solution is for the practitioner to collect more data, or alternatively ensure the data collected is of a high quality and errors are minimised.

Restriction 5 :- Related to the previous limitation of noise level is that of the incorrect assessment of the level of noise within a series of measurement. Over or under-estimation of the noise factor may inadvertantly alter the result.

Solution 5 :- One possible method to eliminate this potential error source is instead of defining a p-value or other appropriate value, the maximum level of noise at which the statistic is significant may be quoted, for a particular significance level. Although unorthodox, it eliminates the potential for incorrectly defining the noise level.

§6.7 Advantages of the Technique

The main advantages of the methodology are :-

1. Its versatility and intuitiveness. The technique may be applied to a whole wealth of situations ranging from detecting changes in pollen levels, to assessing the relationship between the geology of an area and its background radiation levels.
2. The non-reliance on dimensionality ensures that two or more non-similar measures may be examined for changes.
3. It is simple to implement and the test statistics are easily calculable, once the surfaces have been constructed.
4. The lack of restriction placed on the methodology required to develop the surfaces, bar that of consistency, widens its appeal.
5. In terms of a local approach it enables the internal spatial complexities to be investigated more specifically.

CHAPTER 7

THE APPLICATION OF THE METHODOLOGY TO TWO ENVIRONMENTAL CASE STUDIES

§7.1 Why The Studies Were Selected

Within this chapter, two examples have been selected to illustrate the potential diversity and complexity of the applications to which the methodology developed in chapters five and six may be applied.

The first example deals with one of the simplest situations; the investigation of climatic change given the results from a monitoring network. The observations were recorded at fixed localities and different time points, t. The example addresses the question of whether a seasonal temperature change has resulted within the contiguous United States of America during the fifty year span, 1930 to 1980. Interest has been expressed in such a question due to the high public profile of global warming and its possible consequences. Attention will focus on the implementation of the hypothesis method, on both a local and global scale, with the results of the global test being compared to those of other global techniques, described in sections 4.3.

The second example was primarily selected to illustrate that the methodology was not only applicable to simple problems, but also to others where the level of spatial variability differed between variables. The motivating question was whether background radiation is a causative factor in the induction of leukaemia. A pilot study, commissioned by the Leukaemia Research Fund, examined this question for a region of south-west England. In addressing this problem, a secondary issue was examined; 'Was the underlying population the sole factor in controlling the distribution of cases?' The data comprised leukaemia cases, population figures and radiation values which all differed in spatial resolution and quantity. The sparsity of the leukaemia data served to curtail the implementation of the methodology directly but it provided the motivation for the implementation of alternative forms of analysis.

§7.2 Case Study 1 :- The Investigation of Climatic Change

In recent years there have been numerous press releases and technical articles concerned with the ideas of climatic change and potential global warming, including Woodward and Gray (1992), Karl et al. (1991), Tsonis and Elsner (1989). Of central focus has been the phenomenon called the 'greenhouse effect'.

The planet is made habitable by the presence of certain gases. These gases trap long-wave radiation emitted from the Earth's surface and result in a global mean temperature of 15^0C, as opposed to -18^0C, in the absence of an atmosphere. By far the most important greenhouse gas is water vapour. However there is a substantial contribution from carbon dioxide and smaller contributions from ozone, methane and nitrous oxide.

The concentrations of carbon dioxide, methane and nitrous oxide are believed to be increasing and in recent years other greenhouse gases, principally chlorofluorocarbons (CFC's), have been added in significant quantities to the atmosphere. There are many uncertainties in deducing the consequential climatic effects. Typically it is estimated that increased concentrations of these gases since 1860, may have raised global mean surface temperatures by 0.5^0C or so, and the projected concentrations could produce a warming of about 1.5^0C over the next 40 years.

Numerical climate models indicate that other changes in climate would accompany the increase in globally averaged temperatures with potentially serious effects on many social and economic activities.

Much of the evidence for a global warming effect has been based on large-scale Global Circulation Models (GCM's). Table 7.1 summarises the results from five of the most commonly cited methods.

Within these studies the global and annual average warming varies from 2.8K to 5.2K. The warming is accompanied by an increase in evaporation and precipitation. The enhanced radiative heating of the surface due to increases in carbon dioxide and water vapour is balanced by increased cooling due to evaporation, producing a more intensive globally averaged hydrological cycle. The change in temperature is not uniformly distributed in time or space. A concise review of how these changes may be effected is given by Mitchell (1989).

Study	Source	Surface Temperature Change (K)	Precipitation Change (%)
GISS	Hansen et al. (1984)	4.2	11.0
NCAR	Washington and Meehl (1984)	4.0	7.1
GFDL	Wetherald and Manabe (1986)	4.0	8.7
MO	Wilson and Mitchel (1987)	5.2	15.0
OSU	Schlesinger and Zhao (1987)	2.8	7.8

Table 7.1 :- Global mean changes in five carbon dioxide doubling studies.

The GCM's are all based on multi-level mathematical representations of the atmosphere. Given the complexity of the environment and the relative simplicity of the models there is much controvorsy concerning their validity.

Other numerical studies have focused on the idea of climate change and long-term patterns in temperature and precipitation, (Karl, Heim and Quayle (1991), Jones et al. (1986), Diaz and Quayle (1980)).The critical question of climate variability has also been addressed, (Karl et al. (1984), Agee (1982)).

A more recent study by Woodward and Gray (1992) raises the question of whether trend based analysis is valid for temperature data since for data of this form, it is common to observe trends that increase over one time span and then decrease over the next, more than likely because of the correlation imposed on the data by the physical phenomena that drive them. Although all the preceeding analysis invoking the use of a trend based procedure produce statistically significant results, the authors ask, 'If conditions remain the same, should we predict the temperature to increase in the future for an extended period of time?'

None of these studies have incorporated the spatial aspects of the temperature or precipitation field. Early papers which focused on the spatial-temporal change of

temperature and precipitation over the U.S.A. were either regional in nature, Sellers (1968), or restricted to relatively short time periods of the order of a decade or two, Skaggs (1975).

More recently Handcock and Wallis (1990) developed a comprehensive model for the spatial dimension in conjunction with the temporal component. Since the model was for the meteorological field as a whole, this facilitated its direct comparison with GCM's and allowed prediction of derived quantities throughout the region and over time.

For a gradual increase of 5^0C over 50 years, it has been suggested that it will take between 20 and 30 years before the change is discernable above the natural variation in temperature.

The quality of all these studies depends on the climatic records being both reliable and accurate. Many of the data bases used in the past were not believed to be either free of bias and error or suitable for long term climate studies.

Some of the more recent studies have made use of what is believed to be an accurate, unbiased, modern historical climate record set up by the Carbon Dioxide Research Program of the United States Department of Energy and the National Climatic Data Centre (NCDC) of the National Oceanic and Atmospheric Administration (NCAA), Quinlan et al. (1987).

Utilising the above data set, investigation of a very simple question was undertaken to illustrate the applicability of the methodology developed. The question of interest examined the seasonal changes in temperature during the 50 year span 1930 to 1980 and was 'Has a change occurred in temperatures between 1930 and 1980?'. The results obtained were compared to those evaluated on the basis of a number of other existing techniques. The seasons were defined as spring:- March, April and May; summer:- June, July and August; autumn:- September and November; and finally winter:- December and January and February of the next/ upcoming year.

§7.3 United States Historical Climatology Network

A network of 1219 stations (HCN network) within the contiguous United States of America was set up for the specific purpose of compiling an accurate, serially complete,

modern historical climate data set suitable for detecting and monitoring climate change over the past two centuries. The data base comprises station histories, monthly temperature (maximum, minimum and mean) data and total monthly precipitation. Potentially it is the most reliable and continuous sequence to have been collected in recent times, a whole wealth of sources gave rise to the final data set including climatological publications, universities, federal agencies, individuals and data archives.

All stations were quality controlled by NCDC with the use of outlier and areal edits, each station being corrected for time of observation differences, instrument changes and moves, station relocation and urbanisation effects, Karl et al (1986), Karl and Williams (1987), Karl et al (1988). A number of features associated with the data are :-

1. Confidence factors for each adjusted estimate.

2. Only a small portion of the data is missing with some missing values being estimated from neighbouring stations.

3. The data is constantly being updated and enhanced and hence provides a unique source for the evaluation of greenhouse effects.

§7.4 Subjective Impression

Before proceeding to the subjective analysis of comparison, a few basic summary statistics are cited in table 7.2 for the four seasons and the two years in question.

The first tool for assessing subjectively the question of change was based on an univariate form of analysis, the box-and-whisker plot. Specifically the plot examines the differences between the 1980 and 1930 temperatures, figure 7.1. Spring and winter display a similar trend across the results, the mean temperature generally being cooler in 1980 than 1930 whilst the reverse is true in summer and autumn. In terms of range greatest variability amongst stations is demonstrated in the winter months where considerable differences are displayed for the two years in question.

Year	Season	Number	Number (missing)	Mean	Stdev.	Min.	Max.
1930	Spring	1174	25	51.400	8.380	30.29	75.74
1980	Spring	1184	15	50.862	7.956	21.24	77.81
1930	Summer	1170	29	71.994	6.612	50.05	97.72
1980	Summer	1172	27	72.563	7.830	48.81	96.87
1930	Autumn	1169	30	53.294	8.187	33.76	78.78
1980	Autumn	1161	38	53.895	7.970	35.83	80.22
1930	Winter	1170	29	34.141	9.959	10.74	67.30
1980	Winter	1191	8	34.028	10.572	5.790	65.98

Table 7.2 :- Summary statistics for the four seasons.

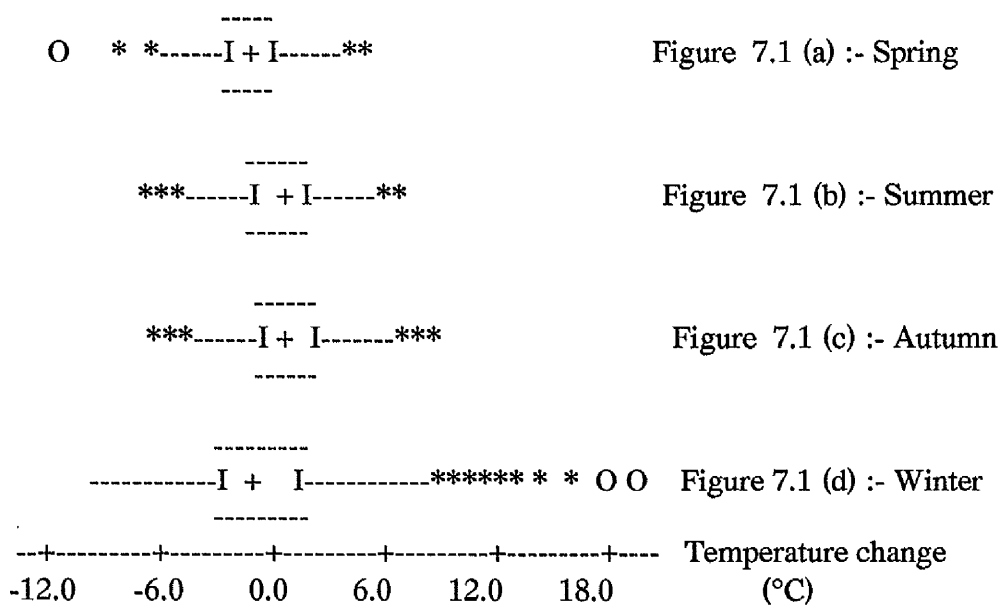


Figure 7.1 (a) :- Spring

Figure 7.1 (b) :- Summer

Figure 7.1 (c) :- Autumn

Figure 7.1 (d) :- Winter

Figure 7.1 :- Box-and whisker difference plots (1980-1930).

The major problem is as for all univariate approaches to spatial statistics, the spatial context of the data is ignored and no feel is attained for how these differences are distributed over the United States of America.

Bivariate plots are an alternative means for describing the data. These enable a better assessment of the relationship between the two data sets to be made. For spring and summer a fairly narrow ellipse encloses all the results, whilst for autumn the points are less well confined especially at the lower range of temperatures. Finally for winter, a separation of the enclosing ellipse occurs at the cooler end of the scale, indicating a greater diversity of results between the two years in question, this confirms the maximal change of 18^0C reported in the box-and-whisker plot for the differences.

However, although the bivariate plot is slightly more satisfactory than the univariate approach, the spatial dimension is still ignored.

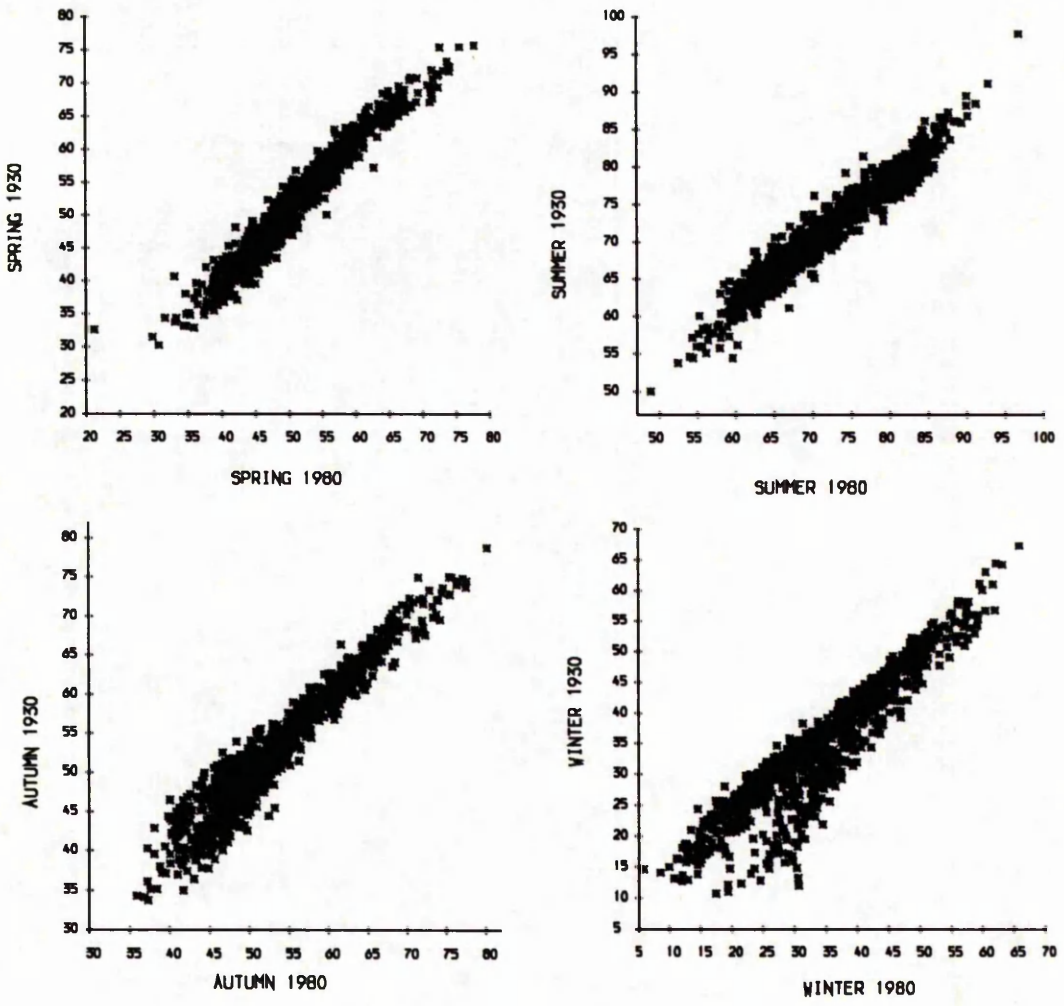


Figure 7.2 :- Bivariate plots of temperature for the contiguous U.S.A. for 1980 and 1930.

We will now consider two of the spatial techniques which formed the cornerstone of the methodology developed in the preceding chapters, to test whether change has resulted between the two variables of interest. Figures 7.3 to 7.6 illustrate the superposition of the two surfaces of interest and the differences between the two data sets i.e. isopach maps, for a selected set of contours for each of the four seasons. The selection of the contours was based on a set of pre-defined temperature levels since interest was in whether change, in the form of an increase or decrease, had resulted during the two years in question, not whether changes had occurred in the spatial distribution of the two surfaces, i.e. we are comparing the same temperature contours and assessing by eye any apparent differences over time. For clarity only five contours have been depicted for each surface.

Based on a recommended grid of 4000 points, the surfaces were plotted using kernel density estimation, with selection of the smoothing parameter undertaken using least-squares cross validation.

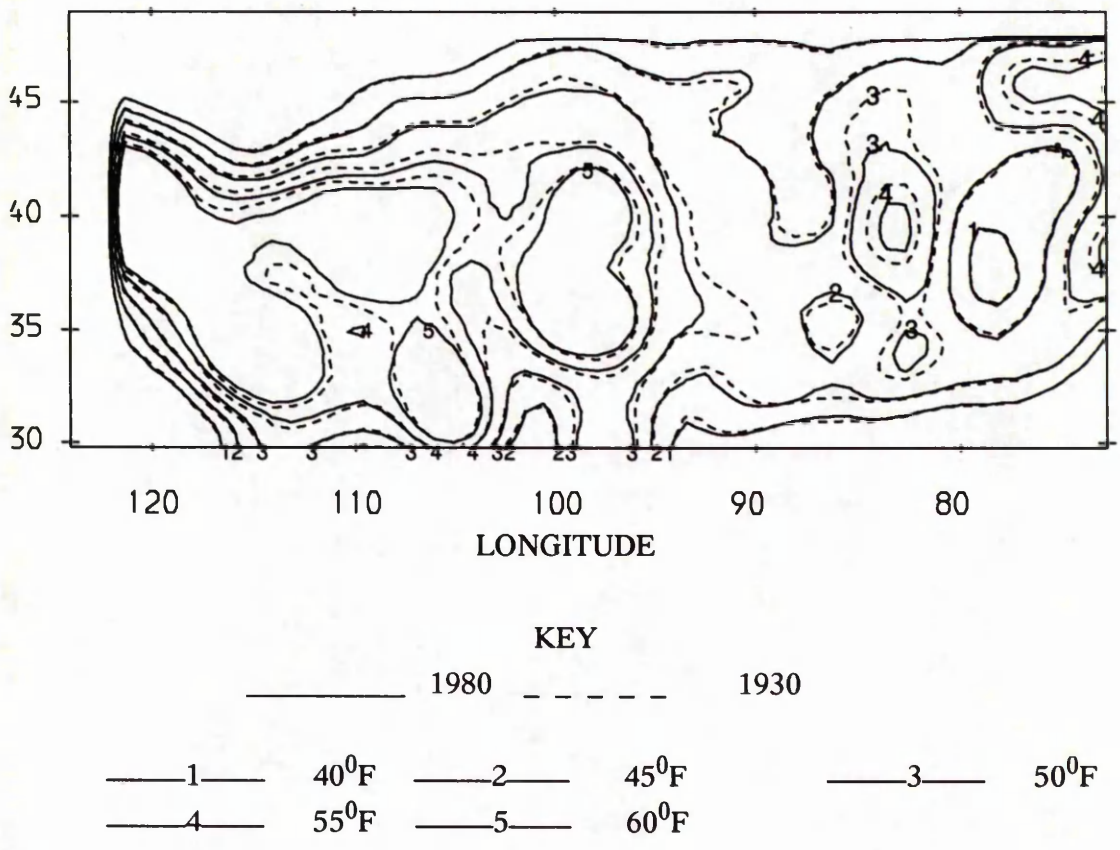


Figure 7.3(a) :- Superposition of temperature surfaces for spring.

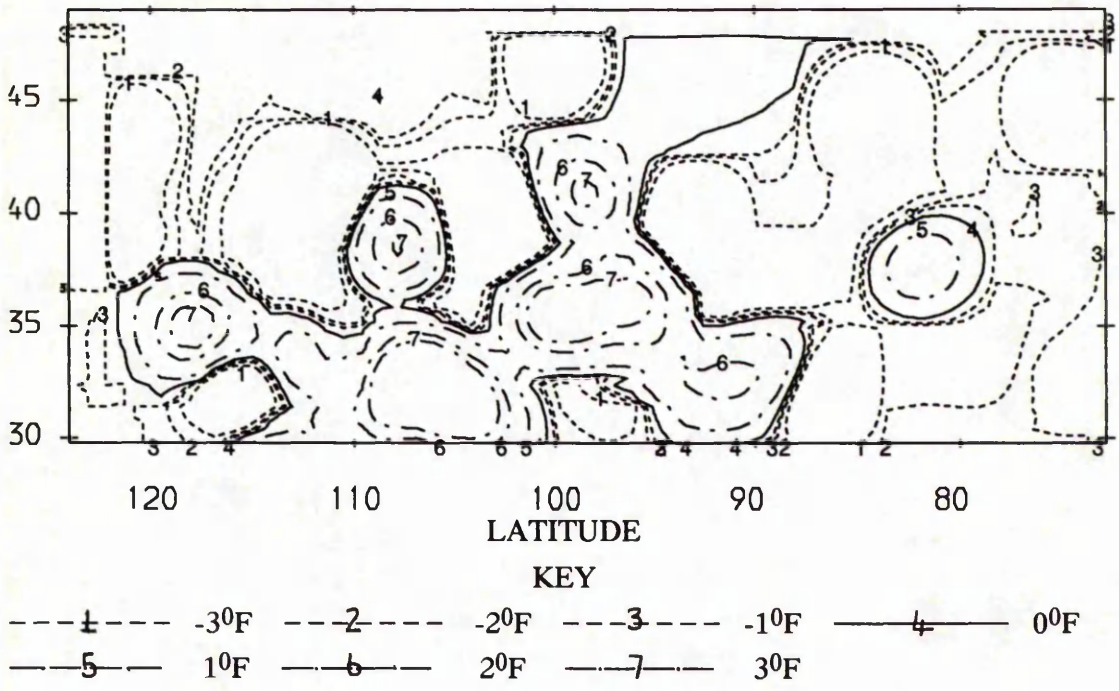


Figure 7.3(b) :- Isopach map for Spring (1980 -1930)

Figure 7.3 :- Spatial subjective analysis for spring.

For temperatures up to 50^0F almost total accord is expressed between the isotherms, above this level indications are that 1980 tended to be slightly cooler, this is indicated by the contours depicting higher levels of temperature generally being smaller in areal dimension for 1980 than 1930.

Moving to the residual map, a difference in the geographical dispersion of temperature change is suggested. A region in the south-west of the United States appeared to experience an overall increase in temperature levels during the 50 year span whilst the rest of the continent was undergoing a reversal of this apparent trend. The situation is relatively more complex to analyse than suggested from the superposition of the two surfaces.

The two techniques describe change in a different way, the first specifically examines a temperature level x^0F, for the whole of an area, whilst the isopach map ignores the underlying level of temperature and considers differences at each specific station.

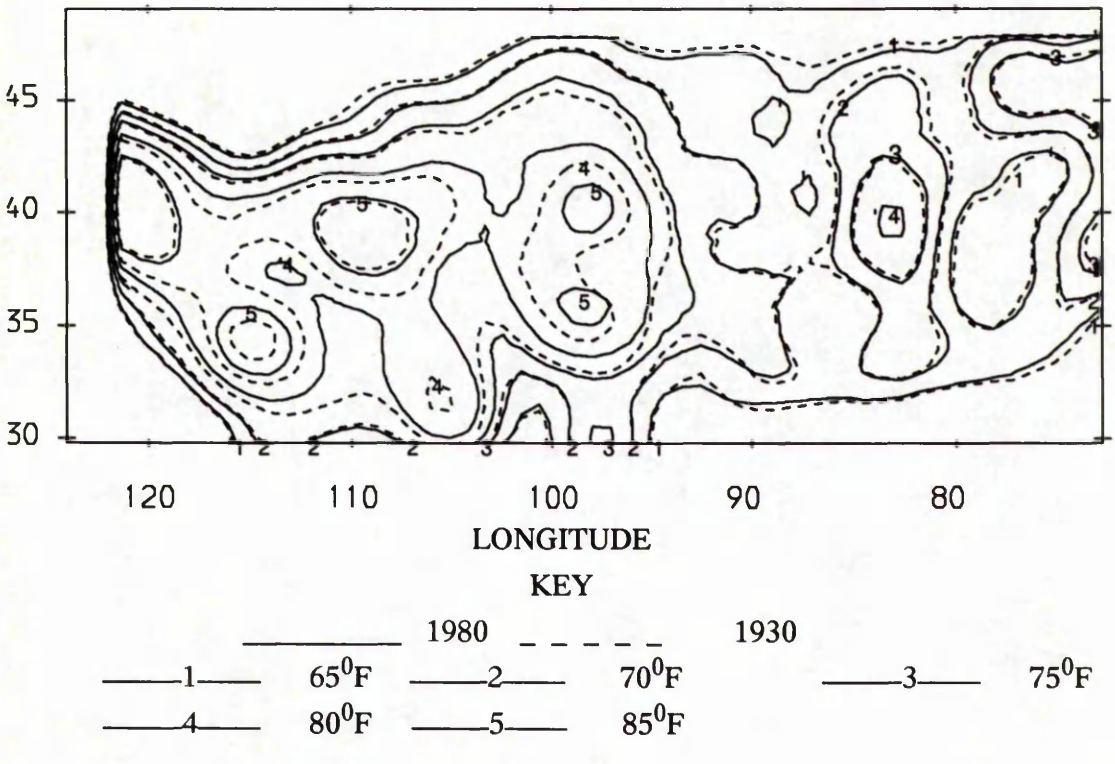


Figure 7.4 (a) :- Superposition of temperature surfaces for summer.

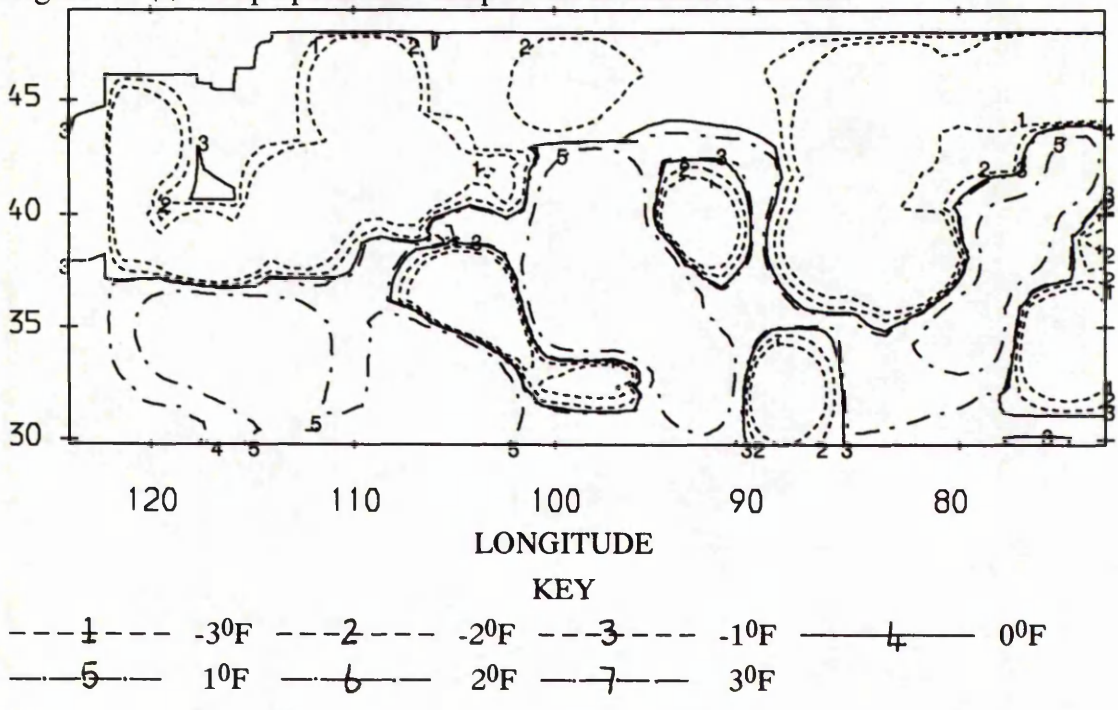


Figure 7.4 (b) :- Isopach map for summer (1980 - 1930)

Figure 7.4 :- Spatial subjective analysis for summer.

Once again greater concordance is displayed between the temperature contours at the cooler end of the range whilst at the upper end the area covered by the contours is correspondingly greater for 1980 than 1930, indicative of the climate becoming warmer over time.

Moving to the isopach map for summer, an approximate north/south divide is indicated. Generally for the more southerly latitudes the temperatures recorded are higher in 1980, whilst the reverse is true for the northerly latitudes. Anomalies do appear to dispel the overall validity of such generalisations, however, this general trend is in accord with other papers in this field.

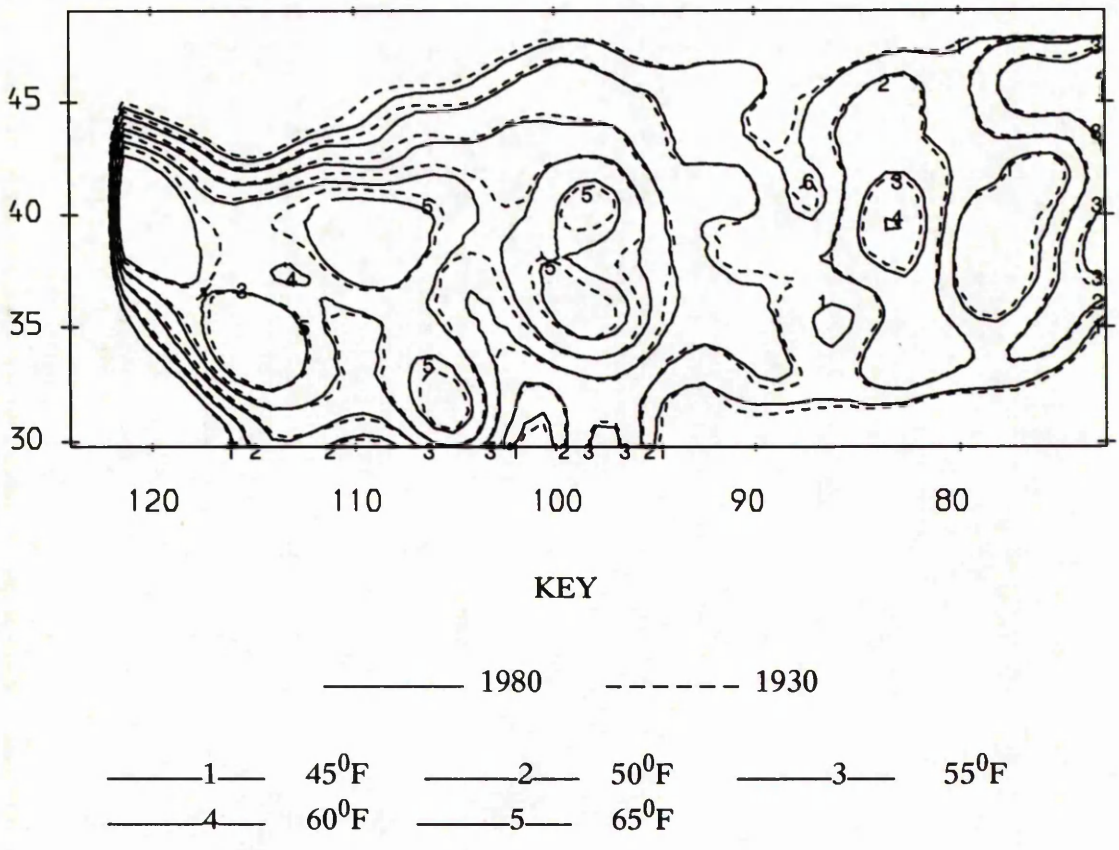
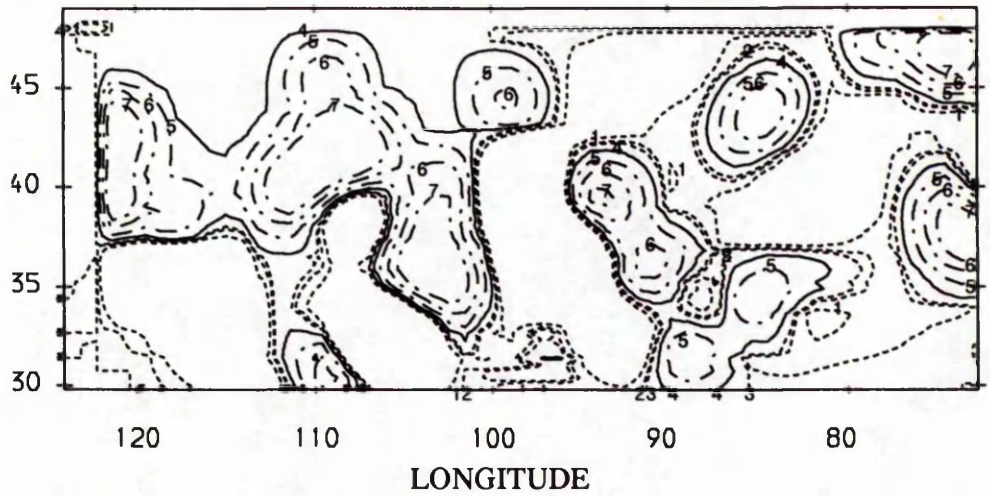


Figure 7.5 (a) :-Superposition of temperature surfaces for autumn.



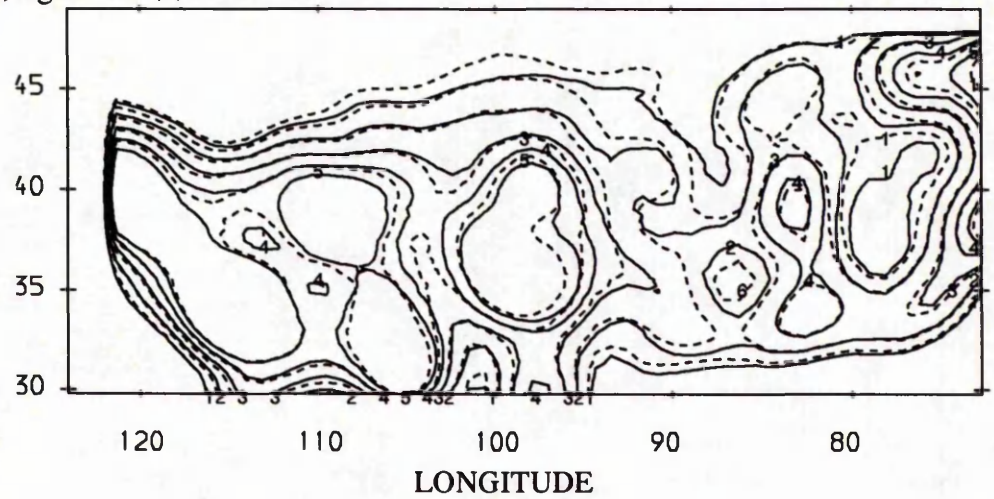
Key :- ---1--- -3^0F ---2--- -2^0F ---3--- -1^0F

—4— 0^0F —5— 1^0F —6— 2^0F

Figure 7.5 (b) :- Isopach map for autumn (1980-1930)

Figure 7.5 :- Spatial subjective analysis for autumn

Examining figure 7.5 (a) closely, it can be seen that no clear pattern emerges, with trends being less apparent than for either spring or summer. For some contours with temperature level x^0C, the areal range is greater for 1980 than 1930, whilst for others, the reverse is true. This extremely complex set of changes is borne out by the isopach map, figure 7.5 (b).



KEY :- ——— 1980 - - - - - - - 1930

———1——— 25^0F ———2——— 30^0F ———3——— 35^0F

———4——— 45^0F ———5——— 45^0F

Figure 7.6 (a) :- Superposition of temperature surfaces for winter

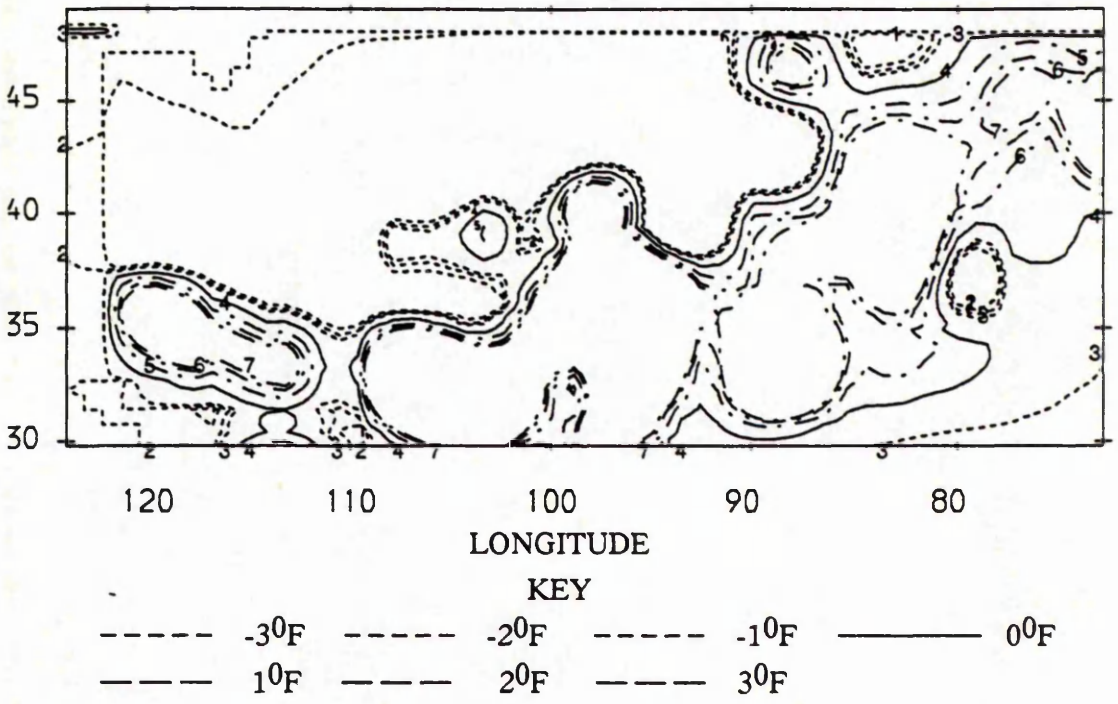


Figure 7.6 (b) :- Isopach map for winter.

Figure 7.6 :- Spatial subjective analysis for winter

Of all the seasons the greatest temperature changes are apparent during winter for the two years in question. There appears to be a strong indication that temperatures have become more extreme during this time. This is particularily apparent from the isopach map. The north-west appearing cooler during 1980, whilst the south-east has witnessed a slight increase in the temperatures during the 50 year span.

§7.3.1 Summary of subjective analysis

From the various forms of subjective analysis, temperature change may be summarised as follows:-

1. Temperatures appear to have risen during the 50 year span for summer and autumn but the cause of the change is unknown i.e. whether it is due to climatic changes or variability in the data.
2. For spring and winter, in general, a drop in the overall temperature has been recorded with the latter season recording the greatest fall of all the seasons.
3. For autumn, an extremely complex set of changes have occurred over time.

In terms of a subjective analysis, a contoured map illustrates the underlying reality of the complexities of change whilst the univariate or bivariate approach mask many of the interesting features and fails to provide the analyst with a feel for possible causative factors.

§7.5 Local Hypothesis Testing Procedure

Many of the techniques described in chapter 4.4 for performing a local analysis are highly convoluted and unsuitable for this type of analysis, the main thrust of this example is to illustrate the apparent simplicity of the hypothesis technique and how it circumvents the problem of pre-defining the inherent level of random noise within the data.

The basis of the local hypothesis testing procedure are the diagrams illustrating the superposition of the temperature surfaces for the two years in question, figures 7.4(a), 7.4(b), 7.4(c) and 7.4(d), for spring, summer, autumn and winter, respectively. For each of the diagrams, the various geometric properties for quantifying the test statistics were evaluated i.e. area, perimeter, orientation and centroid displacement. The last of which was broken down into latitude and longitude displacement. Appendix 1 details the individual results for each contour. The test statistics evaluated using these values are given in table 7.3.

The disparity in the number of entries for each of the temperature levels is solely a feature of the overall contour structure which for some levels is more disjoint than for others. In terms of the test statistics, only those contours within each level which were recognisably comparable were included. For this particular problem, only a couple of small contours were omitted from the analysis.

Season/ temp.	Area	Perimeter	Orientation	Standard deviation displacement	
				x	y
Spring 40°F	1.0019	1.00218	-5.8×10^{-4}	-3.635×10^{-3}	0.04086
	0.93136	0.97296	0.0355	0.13955	-5.2071
45°C	0.9780	1.11360	-1.8×10^{-3}	0.30704	0.25152
	1.0284	1.02239	-0.0621	-0.70175	0.28630

50°F	0.94278	0.96553	0.3292	0.06645	0.26064
	0.81413	0.96790	8.33×10^{-3}	0.37520	0.416832
55°F	1.0757	0.91162	-0.0405	-0.64094	-0.63024
	0.82558	0.58925	-0.0261	0.51950	3.0044
	1.93404	1.2025	1.30525	-0.1060	-0.22030
60°F	0.9524	0.95500	0.14496	0.21021	7.6538
	0.6826	0.81996	-0.28203	2.5099	2.0117

Summer 65°F	0.6316	0.79948	0.02231	4.88674	27.3777
	0.9710	0.99610	-0.626×10^{-3}	0.01747	0.29005
70°F	0.9054	0.77191	0.41104	-2.97786	1.34336
75°F	1.1460	1.04100	-5.70×10^{-3}	-0.28812	-0.43094
	0.4004	0.96455	-0.0205	0.25537	0.99326
	1.0830	0.61902	-0.3170	3.92562	-4.40771
80°F	1.9531	1.54463	-0.0670	-3.9866	0.393133
85°F	2.205	1.49793	-0.0645	-13.213	-9.72238

Autumn 45°F	2.0384	1.4272	2.1746	-19.988	-30.0176
	1.10813	1.0375	-0.03423	-0.7295	-2.95231
	0.97157	1.00264	-6.82×10^{-3}	-2.20×10^{-4}	0.04103
50°F	1.01993	1.00522	0.34103	-0.08018	-0.07099
	1.1015	1.05752	0.07033	-0.35134	-1.6495
	1.2139	1.07564	0.02168	-0.36916	-8.345t
55°F	0.9679	1.07260	-0.29228	-6.02269	-20.1170
	0.96135	0.89430	-9.28×10^{-3}	0.51844	0.33500
	1.53413	1.2233	0.02482	1.17628	-3.86028
60°F	1.02073	1.0220	-0.05399	-0.94467	-1.78106
	0.71339	0.79512	0.39070	-0.13559	3.3182
65°F	0.97220	0.98311	-0.02078	-11.4984	-36.5372
	1.08070	1.04255	0.02806	-1.13102	-1.47291
	1.2358	1.04862	-0.07368	-7.8949	-1.81121
	1.9770	1.53175	0.09076	-27.65119	-39.220

	0.74826	0.826653	0.06799	2.6625	7.2238
	0.79049	0.89743	-3.257×10^{-3}	2.73579	3.09780

Winter 25°F	0.97678	0.93888	-3.02×10^{-3}	-0.04095	-0.20802
	0.66336	0.74910	-2.32220	3.36146	10.6727
	0.03390	0.19516	-0.22166	137.29	471.32
30°F	1.25450	1.8898	-0.02368	0.50791	-1.176
35°F	0.93717	1.00449	0.0486	0.36145	-2.29×10^{-3}
	1.21001	1.02836	-0.02758	-0.65218	-7.0473
	57.7506	6.69628	-0.02585	-86.3656	-15.8296
40°F	0.92218	0.9789	-0.01228	0.21380	-0.03105
	3.0485	1.85556	-0.95204	-8.52874	-137.92
	1.1272	1.2917	0.05763	-2.51320	-17.1397
	1.00273	0.97947	0.11944	0.54803	-1.58112
45°F	0.34335	0.44868	-0.51460	15.8644	19.2410
	1.10012	0.94118	0.30433	10.93002	-10.9958
	1.1510	0.9855	-0.0714	-0.4002	-5.2148
	1.6535	1.51163	0.2633	-20.1513	-7.8087
	11.000	2.6401	-0.5273	-55.131	-956.13

Table 7.3 :- Observed test statistics describing the various forms of change for the contours of interest.

For the local analysis, the null hypothesis for each test statistic describing the situation of no change in the geometric quantity, whilst the alternative was that change of some form had occurred.

The results of these hypotheses tests, table 7.4. were reported in terms of the upper bound of the random noise at which the observed value of the test statistic was significant for a 5% significance test i.e. N.S. (non-significant), 0%, 5%,15% and 25% i.e. for a reported value of x%, say, the null hypothesis may be rejected in favour of the alternative if the upper bound of noise in the original data is less than or equal to x%. Alternatively if the data has an inherent noise level greater than x%, then change maybe due to random noise and not some underlying physical process.

Season	Temperature	Area	Perimeter	Orientation	Centroid Displacement	
					x	y
Spring	40°F	NS	NS	NS	NS	NS
		15%	15%	NS	NS	25%
	45°F	15%	5%	NS	NS	NS
		15%	5%	NS	15%	NS
	50°F	15%	5%	25%	NS	NS
		25%	5%	NS	5%	15%
	55°F	25%	15%	NS	15%	15%
		25%	25%	NS	15%	25%
		25%	25%	25%	NS	NS
	60°F	15%	15%	15%	NS	25%
		25%	25%	25%	25%	25%
Summer	65°F	25%	25%	NS	25%	25%
		15%	5%	NS	NS	NS
	70°F	25%	25%	25%	25%	25%
	75°F	25%	5%	NS	NS	15%
		25%	15%	NS	NS	25%
		25%	25%	25%	25%	25%
	80°F	25%	25%	NS	25%	15%
	85°F	25%	25%	NS	25%	15%
		25%	25%	NS	25%	25%
Autumn	45°F	25%	25%	25%	25%	25%
		25%	5%	NS	15%	25%
		15%	NS	NS	NS	NS
	50°F	15%	NS	25%	NS	NS
		25%	15%	NS	NS	25%
		25%	15%	NS	25%	25%
	55°F	15%	15%	25%	15%	NS
		15%	25%	NS	25%	25%

		25%	25%	NS	25%	25%
	60°F	15%	5%	NS	NS	25%
		25%	25%	25%	25%	25%
	65°F	15%	5%	NS	25%	25%
		25%	5%	NS	25%	25%
		25%	5%	NS	25%	25%
		25%	25%	NS	25%	25%
		25%	25%	NS	25%	25%
		25%	15%	NS	25%	25%

Winter	25°F	15%	15%	NS	25%	25%
		25%	25%	25%	25%	25%
		25%	25%	25%	25%	25%
	30°F	15%	25%	NS	25%	25%
	35°F	15%	NS	NS	5%	NS
		25%	5%	NS	15%	25%
		25%	25%	NS	25%	25%
	40°F	25%	5%	NS	NS	NS
		25%	25%	25%	25%	25%
		25%	25%	NS	25%	25%
		NS	5%	15%	15%	25%
	45°F	25%	25%	25%	25%	25%
		25%	15%	25	15%	25%
		25%	5%	NS	15%	25%
		25%	25%	25%	25%	25%
		25%	25%	25%	25%	25%

Table 7.4 :- Results for local hypothesis testing procedure.

§7.5.1 Summary of local results

The following general comments may be drawn from the results of the local analysis. For the separate test statistics, greatest change is witnessed in those statistics expressing scalar change i.e. area and perimeter.

The orientation of the contours appear fairly static for all seasons and temperature levels, although a number of anomalous results are recorded. Referring to table 7.3, these changes appear to be linked to those situations where either, the areal dimension of the contour is small, the contour is near cyclic in shape or finally, more seriously, where the contour describing a specific temperature level has become disjoint over time between 1980 and 1930 or vice-versa. This observation is potentially worrisome and an alternative method for collating contours may be required.

Finally in terms of centroid displacement, although not geographically universal, the results indicate a general cooling which is most pronounced in the east and a general warming in a southerly direction, this bears out the findings of Diaz and Quayle (1980).

More specifically, examining each of the seasons in turn :-

Spring :- The changes in the scalar statistic may be succinctly summarised by noting the more disjoint appearance of the contour structure at higher temperature levels for 1980. In accordance with the previous findings this suggests that a drop in temperature has resulted. In general a higher level of agreement is observed at cooler temperatures where the areal dimension of the contours is greater than for physically smaller contours at the same level.

Overall the changes reported in the other statistics are not of great interest with no trends evident. The changes in spring are not very strong and this confirms the subjective impressions.

Summer :- The scalar trends for summer are similar to those for spring with warmer temperatures once again being the source for greatest change. For both area and perimeter the changes are significant for noise levels of 25% and possibly greater.

The centroid displacement results appear to suggest the possibility of a north/south shift in temperature level. Referring specifically to the test statistics, the indications are that an increase in temperature has occurred for the north between 1930 and 1980. In a similar vein, an increase in temperature is more liable to have been reported in the west during this time span, this is less apparent than for the latitude results.

Autumn :- Greater changes are evident during the autumnal period than for the previous seasons. In terms of scalar change no pattern emerges, an areal increase is indicated for some contours during the 50 year span whilst, for the same temperature level, the reverse is true.

A semblance of a trend in terms of directional change is indicated. Once again the south and west both appear to have witnessed a warming of the temperatures during the fifty years of the study.

Winter :- Overall the changes are the most diverse for all of the seasons. In terms of the scalar changes a very complex structure emerges and generalisations are not plausible. The north/south, east/west divide once again appears to materialise as described for the preceeding seasons.

§7.6 Global Analysis

§7.6.1 Existing Techniques

Some of the more common methods cited in the literature for comparing two sets of spatial data on a global basis were used to examine the question of change between the two data sets :-

1. Correlation analysis
2. Paired t-interval
3. Trend-surface analysis
4. Regression analysis

Before proceeding to execute any of the above techniques, the assumption of normality was verified for all permutations of season and year. The second assumption of equality

of variances across temperatures for each season was checked using the standard F-test, the results indicated pooling of variances was valid for the data.

Each of the above methods may be implemented using one of the standard statistical packages e.g. Minitab, hence their wide appeal to users in other areas.

§7.6.1.1 Correlation analysis

The first method used the ideas of correlation to examine the strength of the relationship between the variables. The question of independence between the two data sets was examined by the standard hypothesis procedure.

$$H0: \rho = 0$$
$$H1: \rho \neq 0$$

and secondly, a confidence interval was derived for the correlation coefficient using the following result :-

$$\frac{z(r)-z(\rho)}{\sqrt{1/n-3}} \cdot N(0,1)$$

The implementation of the above two procedures enables the analyst to formulate an opinion as to the strength of the relationship. Table 7.5 collates the results :-

Season	Correlation (ρ)	Hypothesis test (p-value)	Confidence Interval for (ρ)	
Spring	0.975	0.00	0.9721	0.9776
Summer	0.969	0.00	0.9654	0.9721
Autumn	0.963	0.00	0.9587	0.9667
Winter	0.926	0.00	0.9170	0.9381

Table 7.5 Results for correlation based analysis.

The results display a strong association between the two variables. The presence of a strong physical connection reduces the potential for a weak association between the two variables.

§7.6.1.2 Paired t-interval

With the underlying assumptions of normality and the pooling of the variances having been checked, paired t-intervals were calculated for the four seasons. The hypothesis of interest being :-

$$H0 : \mu_d = 0$$
$$H1 : \mu_d \neq 0$$

i.e. under the null hypothesis, no change occurs in the mean value of the differences for the two data sets whilst under the alternative, some form of change has taken place over time. Table 7.6 summarises the results.

Season	Paired t- interval
Spring	(-0.792, -0.574)
Summer	(0.264, 0.512)
Autumn	(0.310, 0.571)
Winter	(-0.575, -0.115)

Table 7.6 :- Results for the paired t-interval.

From the results it is apparent that overall the temperatures in 1980 are cooler than for 1930 for spring and winter, the greater changes being in the mean value for spring. A reversal of the pattern is seen for the other two seasons, with little disparity between the mean changes.

§7.6.1.3 Trend surface analysis

The penultimate technique compared the trend surface for the two sets of data using the correlation coefficient basis for comparison, section 4.3.4. The most appropriate form of trend surface was that of a quadratic i.e. of order two :-

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 y_i + \beta_3 x_i y_i + \beta_4 x_i^2 + \beta_5 y_i^2$$

i.e. temperature is equal to the sum of a constant term related to the means of the geographic co-ordinates, plus a polynomial expansion of degree two of the geographic co-ordinates, plus a randomly distributed measurement error.

The same criteria as described in the section on correlation analysis allows the analyst to judge the strength of the relationship between the two surfaces. A large value of ρ is indicative of a low level of change and vice-versa, the results are given in table 7.7.

Season	Correlation value
Spring	0.977
Summer	0.891
Autumn	0.999
Winter	0.981

Table 7.7 :-Results derived for the correlation between the two trend surfaces.

A higher level of accord is suggested by the results of the trend surface analysis than for the standard correlation analysis. This is a manifestation of the type of analysis, the two surfaces will inevitably have similar trends due to the underlying physical phenomena which intrinsically control the temperature levels e.g. altitude, latitude, longitude etc. and these are constant for the two data sets.

§7.6.1.4 Regression analysis

In terms of a regression, two approaches are possible. Both methods are based on the model for simple linear regression :-

$$temp_{1980} = \alpha + \beta temp_{1930} + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2) \qquad (1)$$

Method 1 :- On the assumption that no temperature change has resulted if the slope β, of the fitted regression line equates to one, when α, the intercept, is forced to equal zero, a standard interval for β was constructed, table 7.8.

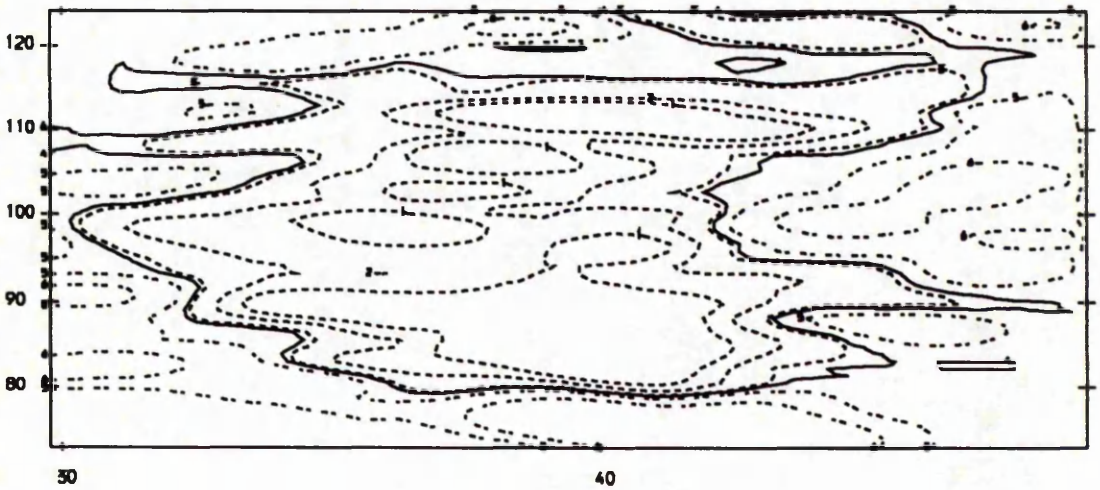
$$\text{i.e. } temp_{1980} = \beta temp_{1930} + \varepsilon_i \qquad \varepsilon_i \sim N(0,\sigma^2)$$

The point estimates for the slope confirm the similarity of the trends for spring and winter, and summer and autumn, i.e. for the former, the temperatures have fallen during the fifty year span, the reverse is true for the latter. However in terms of the confidence interval for the slope, the results for both summer and winter are indicative of no change over the time period of 50 years.

Season	Slope (β)	Interval estimate for β	
Spring	0.9850	0.9829	0.9871
Summer	1.0065	0.9997	1.0133
Autumn	1.0064	1.0023	1.0106
Winter	0.9890	0.9712	1.0067

Figure 7.8 : Results for regression slope analysis.

Method 2 :- The second technique fits regression equation (1) and from the fitted values, $\hat{y}$, a surface is produced which superimposes these values against the actual values of y. Alternatively, a residual map may be used to describe the results, figure 7.7 describes the surface for spring.



KEY

----1----	-6	----2----	-3.0	——3——	0.0
----4----	1.5	----5----	6.0		

Figure 7.7 :- Residual map for spring

This approach combines a theoretical and subjective based analysis for assessing change. The advantage of this technique is that the residual surface attempts to incorporate the spatial dimension which is ignored in the preceding methods. The plot underlines the simplifications imposed by utilising such techniques and that change cannot be truly expressed in terms of one parameter.

§7.6.2 Summary of the existing techniques

Correlation based methods provide a measure of the association/change between data sets but are generally uninformative. For data of this form a strong natural association exists due to the physical conditions which intrinsically control temperature. In terms of the trend surface analysis, the resultant measure suffers from the same drawbacks as for a standard correlation analysis.

The paired t-interval provides greater insight into the mean change particularly in terms of the direction of change. A potential problem related to ignoring the spatial dimension is that of auto-correlation and the possibility that the results are more extreme than in reality. The overall findings confirm the observations postulated in the subjective analysis: 1980 temperatures have fallen in spring and winter whilst the reverse has occurred during the other two seasons.

The final method based on the notion of regression saw two possible modes of analysis; the first gauged change as a function of the slope of the regression line. A slope of one, indicative of no change when the intercept was forced to equal zero. Once again the same reservations expressed for a correlation approach hold here. The second of the regression methods, although not providing a finite solution for change does consider the spatial dimension, although only subjectively.

Attempting to produce a 'true' global value for change is liable to be difficult using the standard statistical methods. Utilising the methodology developed in chapters four and five, to produce a global test goes part of the way to accounting for the spatial complexities of change, illustrated so clearly in figure 7.7.

§7.6.3 Global Hypothesis Testing Approach

The global hypothesis testing procedure used the test statistics describing angular, translational and scalar change namely angular change, latitude and longitude displacement and areal change respectively. The minimal number of contours of interest was five because of the condition imposed when implementing a Hotelling's One-Sample T^2 test, $n > p$, where n is the number of contour levels and p, the number of variables of interest. The number of temperature levels examined for each season was five. Within each temperature level, the surface was represented by more than a single contour.

In terms of the structure of the hypothesis testing procedure, the null hypothesis related to the situation of no change in all the measures described. The alternative was expressed in general terms of change.

As mentioned in the previous chapter (§6.7), a number of complications arise when implementing the multivariate testing procedure. These relate mainly to the definition of the various descriptors, where contour A has become disjoint over time or alternatively where two contours have amalgamated. A consistent approach to this problem was adopted:-

(a) For area, for the specific contour level of interest, the sum was taken.

(b) A global mean was taken for the centroid displacement.

(c) In terms of orientation, a mean difference was effectively used. The whole contour being split into its two component parts and the orientation for each segment evaluated, the differences with the original disjoint contour were then taken and finally, the mean calculated.

A collation of the necessary mathematical quantities are given in appendix 1, for all the contours depicting the various temperature levels of interest for each season with table 7.9 describing the results of the global analysis.

Season	Hotelling's One Sample T^2 Test Statistic	P-value
Spring	0.93476	0.465
Summer	0.31140	0.820
Autumn	1.04009	0.400
Winter	1.16014	0.360

Table 7.9 :- Results of global hypothesis testing procedure.

For all seasons, the null hypothesis cannot be rejected in favour of the alternative. Differences in the results therefore appear to be solely attributable to natural variability. This statement should be tempered due to a number of drawbacks of the methodology cited i.e. only those contours which are visually comparable are incorporated and secondly, restraints are imposed by the selection of the contour level. Selecting contours in the zone of greatest change i.e. the upper range for each of the seasons may cause a reversal of the results. Selection based on a pre-defined criteria eliminates this methodological drawback and area of potential bias.

§7.7 Conclusions

The local approach provides a more detailed insight into the changes which have resulted over time than the global analysis. However in general it may be stated that some form of change has occurred for all four seasons. This change is most apparent for winter and then on a sliding scale, autumn, summer and finally spring. This generalisation glosses over many of the intricacies of change and how regional differences are evident, the question of micro-climate possibly influencing the results.

The changes are primarily of a scalar nature and to a lesser extent, translational. In terms of orientation, it may be reported that almost no change has occurred.

This example has illustrated the simplicity of implementing both approaches and how the results generally conform to what has been postulated in the subjective impression. The main point of interest is in terms of the centroid displacement, where some of the test statistics are highly inflated. This primarily occurs when dealing with areally small

contours where the standard deviation is less than one, as a result, the statistic describing centroid displacement is considerably inflated. The resultant conclusions are unaffected since if the two standard deviations are comparable then this problem is eliminated in the ensuing test statistic, whilst if the two contours differ considerably, the results confirm the anticipated behaviour.

The drawback as indicated earlier is that we have no estimate of the size of the difference. This is a problem with all hypothesis based tests. Overall on a local scale, change of some form is indicated, even if the level of noise, inherent to the data set is fairly large. On this premise, it is essential that further investigations should be carried out to see whether this change may be linked to a physical process. Although Handcock and Wallis (1989) concluded 'It will be 20 to 30 years before change is discernible from the natural variability in temperatures', the use of methodology of this form shows that change can potentially be detected before this duration in a good quality data set.

The other factor to emerge from this example relates to the ability of the global test to take account of the spatial dimension, which to date the majority of techniques have ignored. Figure 7.7 illustrates the consequences of ignoring this factor. By adopting a consistent approach to the problematical contours, the global analysis should be a good indicator of change.

§7.8 Case Study 2 :- The Investigation Of A Possible Link Between Leukaemia And The Underlying Radiation Fields

For some years there has been considerable public concern about radiation in general and man-made radioactivity in particular. Public awareness of the impact of radionuclides in the environment has been heightened even more by the Chernobyl reactor accident.

Radiation of natural origin is widespread in the environment. The earth itself is radioactive and naturally occurring radionuclides are present in the air we breathe, in the food we eat and in our own bodies. Everyone is exposed to natural radiation and for most people it is the highest contributor to total dose. Table 7.10 provides an estimate of the breakdown of the total radiation dose received by the people of Thurso provided by the NRPB, Dionian (1986).

Source	Percentage of Total Dose
Natural Radiation	79.0
Fallout	12.0
Medical	7.5
Dounreay discharges	1.2
Sellafield discharges	0.3

Table 7.10 :- Summary of breakdown of total radiation dose received by the people of Thurso, Dionian (1986).

Man-made radionuclides have been distributed throughout the world as a result of nuclear weapons testing in the atmosphere. These radionuclides are inhaled, deposited on the ground giving rise to external exposure, and they can also be transferred through food-chains to our diet. Even though the period of intensive weapon testing occurred more than twenty years ago, residual activity from these tests and from occasional more recent explosions, still give rise to some small exposure of the population.

Radioactive materials are discharged from nuclear installations, some industrial premises and from medical and research institutes. Accidental releases of radionuclides

may also occur and, as Chernobyl has shown, a severe accident at a nuclear power station can lead to widespread contamination of the environment.

Radionuclides are subject to all the physical, chemical and biological processes of environmental transfer. No matter how complex the pathway by which the activity may reach man, the actual routes of human exposure are limited to: external irradiation; inhalation of airborne material; ingestion of activity in food or water. Measurements are related to these mechanisms of exposure. Thus, in the environment we measure external dose rates and activity concentrations in air, food and water.

In this section we are specifically interested in whether background radiation is a causative factor in the induction of various forms of lukaemia. To date, the spatial analysis of disease patterns has been used by investigators as one tool with which to address problems of disease causation. Four main approaches to this type of analysis have been taken :-

1. Ecological analysis
2. Mapping and estimation patterns of disease
3. Clustering
4. Regression/correlation analysis

An ecological study is one in which the unit of analysis is a group of individuals, often defined geographically, and their relationship between the incidence of a disease in spatial units and other covariates is examined. Making aetiological inferences about individuals from data on groups is potentially hazardous. However where data on individuals is unavailable it is useful to express group relationships between areas. The role of ecological studies in epidemiological research is discussed by Morgenstern (1982).

The second group, estimating and mapping disease rates, have generally been concerned with the production of rates relating to the incidence of cancer with the end-product being a cancer atlas. The rates are usually computed using mortality or morbidity data as the numerator, with census data providing the denominator. Walter and Birnie (1991) have reviewed the techniques of analysis, presentation and interpretation of atlases from various countries.

A more limited approach to the problem of spatial analysis of disease patterns is that of disease clustering. Hills and Alexander (1989) examine some of the approaches taken and the associated problems. Generally testing for clustering is aimed at tackling two issues. First, is there a tendency for clustering to occur and, if so, where? Second, do clusters occur in specific areas e.g. near suspected environmental hazards.

Finally the use of correlation and regression have featured strongly in testing whether natural radiation, radon in particular, is a causitive factor in the induction of particular forms of leukaemia. This has generated two schools of thought, firstly, those who advocate that ecologically low levels of ionising radiation is harmful to human beings. Henshaw et al (1990) suggested that in the United Kingdom, 6-12% of myeloid leukaemias may be attributed to radon. In Cornwall where radon levels are higher, this increases to 23-43%. This view supported the findings of Kneale and Stewart (1987) who demonstrated the existence of a correlation between childhood cancer and background indoor gamma radiation. The second school advocate the hormesis effect, i.e. low levels of radiation are beneficial. This has been substantiated by both animal and biological experimentation and to a lesser extent in a number of studies on humans in Japan, Ujeno (1983), the U.S.A., Hickey et al (1983) and more recently in a larger study in India, Nambi and Soman (1987,1990).

§7.9 Description Of The Problem And How It Differs To the Previous Example

The effects of pollution, industrialisation, social deprivation and so forth are increasingly perceived as the causes of geographical anomalies in health. Within this section we examine the effect of background radiation on man i.e. 'Is there a link between the disease pattern of certain forms of leukaemia and the underlying radiation fields?'

In 1989 the Leukaemia Research Fund commissioned a pilot study to investigate the feasibility of aerial radiometric survey for generating information on background radiation levels. The information garnered was to be used in juxtaposition with the epidemiological data to examine the pre-stated hypothesis. This problem differs from the previous climatic illustration in the following ways:-

1. It seeks to establish whether an association exists between two sets of differing spatial variables, cases and population and secondly, radiation levels and cases.

2. The spatial variables of interest differ in spatial resolution. This point is expanded upon in section 7.10.

Furthermore it differs in form to the more theoretical types of problem which underwrote the methodological development i.e. the comparison of two sets of punctual data points. This apart, the analysis addresses two separate issues, firstly the relationship of the case locations to the underlying population and secondly, the relationship between the case locations and the radiation fields. Before proceeding to the analysis stage a general overview of the pilot study is given with a short description of the three data sources :-

1. epidemiological data
2. population data
3. radiation data.

§7.10 Pilot study

The pilot study commissioned by the Leukaemia Research Fund in July 1989 required an aerial survey to be flown over three disjoint regions in South-West England covering approximately 2250km^2 in total. In September of that same year the survey was undertaken. The location of the three grids are shown in figure 7.8.

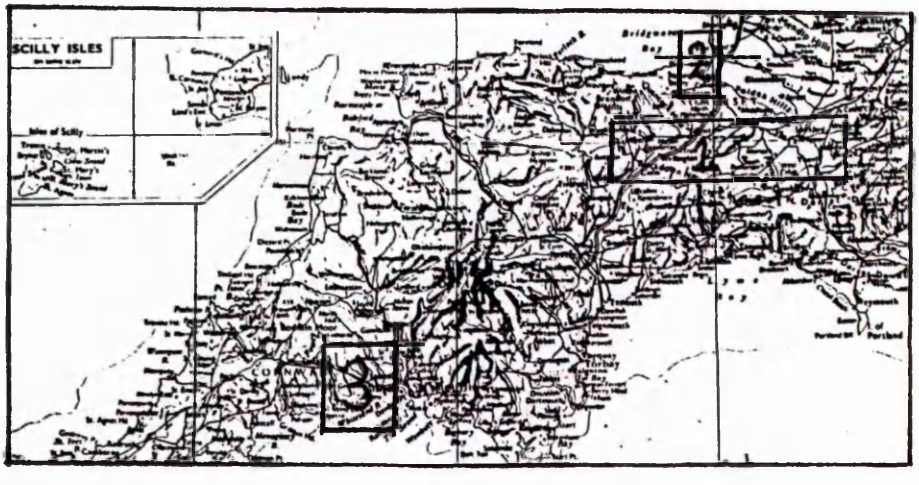


Figure 7.8 :- Location of the three grids.

Grid one encompasses approximately 900km^2. The main population centres being Yeovil, Wellington, Illminster and the southern edge of Taunton. The remainder of the area consists of small hamlets with the Black Down Hills forming the main geographical feature.

The second grid lies to the north of grid one and is approximately 20km×15km. Part of the north section lies in Bridgewater Bay. The area is predominately rural with two main towns, Bridgewater and Burnham-on-Sea. Crossing this region is part of the Sedgemoor Drain system.

Finally grid three which is located to the south-west of grids one and two and west of Plymouth has Liskeard, Saltash, Torpoint and Launceston as its principal towns. The southern fringes of Bodmin Moor encroach into this grid as does the estuary of the River Tamar.

§7.10.1 Radiation Data

The first primary source of information was that of the radiation data. Aerial radiation survey methods are based on the ability of gamma radiation to propagate up to a few hundred metres in air from the originating source of radioactivity. It is possible to monitor the flux of gamma radiation above ground or sea level using highly sensitive gamma ray spectrometry equipment mounted in aircraft flown close to the ground. Environmental radioactivity measurements using portable field based spectrometers may take from 15 minutes to 30 minutes per sampling site, or environmental soil cores extracted from single sites may take several days each to analyse. High volume aerial survey equipment can make sensitive readings every few seconds while moving on preset paths above the land surface.

The survey was conducted with 1km line spacing and 500m resolution along each flight line using an aerospatiale squirrel helicopter flown at 120km/hr. Raising a detector above ground opens the detection geometry so that the area on the ground being sampled increases very rapidly. Typical areas of investigation are such as to give 90% of the detected signal from a circle diameter 4-5 times the height above the ground. Thus each observation at a survey height of 100m is averaged over a circle of diameter of 500m. Details of how the analysis of the data set proceeds to produce values for the

various radiation fields is given in Sanderson et al (1990). A summary of the set of steps undertaken are as follows:-

1. Generation of summary files which involves collation of individual readings and their corresponding positions and altitudes along each flight line. Checks are incorporated at this stage for any anomalies.

2. Detector background rates are subtracted from the readings.

3. The counts are stripped i.e. the spectral interferences between adjacent channels is removed.

4. Altitude corrections are made to the stripped count rates.

5. The stripped counts are converted to calibrated data - so that for each location, equivalent uranium, eU, and thorium, eTh, and potassium, ^{40}K, in kBq/kg, and alpha, beta and gamma dose rates in mGy/a are calculated - using linear equations derived by regression analysis of ground level concentrations against aerial observations.

§7.10.2 Epidemiological Data

The epidemiological data provided by the Leukaemia Research Fund, (L.R.F.) comprised a set of all the recorded incidences of leukaemia during the five year period 1984 to 1988 for the three grids surveyed. The case locations were identified by the Ordinance Survey co-ordinates of the postal code area in which the person resided at the time of diagnosis. Additional information available on these people included age, diagnostic code and date of diagnosis.

The diagnostic codes of which there are ninety, serve as identifiers for the type of leukaemia contracted. Early in 1990, a Leukaemia and Lymphoma Atlas was launched by the L.R.F. for selected regions of England and Wales. The Atlas was based on ten medically defined categories of leukaemia of which the diagnostic codes formed the basis. These ten groups can be split into two broader categories, all lymphoproliferative diseases and all myeloproliferative disorders. The first group, all lymphoproliferative diseases contains six of the ten categories whilst the remaining four fall into the category of all myeloproliferative diseases, table 7.11.

All Lymphoproliferative Diseases	All Myeloproliferative Diseases
Acute lymphoblastic leukaemia	Acute myeloid leukaemia
Chronic lymphocytic leukaemia	Chronic myeloid leukaemia
Hodgkin's disease	Myeloid dysplasia
Low-grade non-Hodgkin's lymphoma	Other myeloproliferative disorders
High-grade non-Hodgkin's lymphoma	
Multiple Myeloma	

Table 7.11 L.R.F.'s diagnostic breakdown of the various forms of leukaemia.

A total of 377 cases were reported for the three grids. The breakdown of the cases for the ten diagnostic groupings is given in table 7.12. Neither cases of multiple myeloma nor myeloid dysplasia were extracted by the L.R.F. from their original records for analysis.

Diagnostic Group	Grid1	Grid 2	Grid 3
All lymphoproliferative diseases	155	33	74
Acute lymphoblastic leukaemia	3	3	5
Chronic lymphocytic leukaemia	53	15	22
Hodgkin's disease	22	5	13
Low-grade non-Hodgkin's lymphoma	46	5	19
High-grade non-Hodgkin's lymphoma	29	5	10
Multiple Myeloma	0	0	0
Acute myeloid leukaemia	31	10	19
Chronic myeloid leukaemia	8	2	5
Myeloid dysplasia	0	0	0
Other myeloproliferative disorders	18	6	13
All myeloproliferative disorders	57	18	37

Table 7.12 :- Breakdown of leukaemia cases into diagnostic groups.

By means of a chi-squared test of homogeneity there was no apparent significant difference between the grids as to the frequency of occurrence for each disease type.

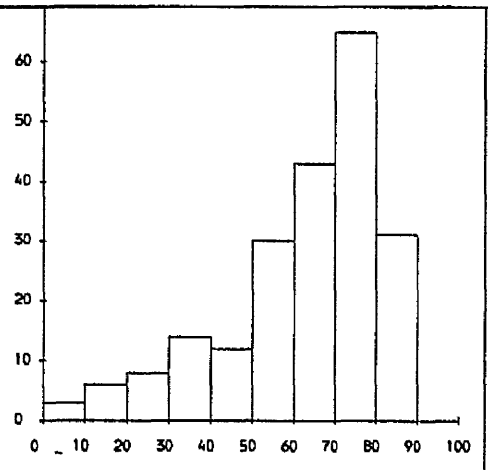


Figure 7.9(a):- Age distribution for grid 1

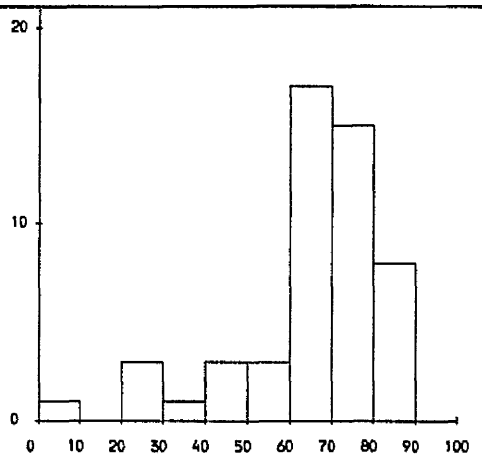


Figure 7.9(b):- Age distribution for grid 2

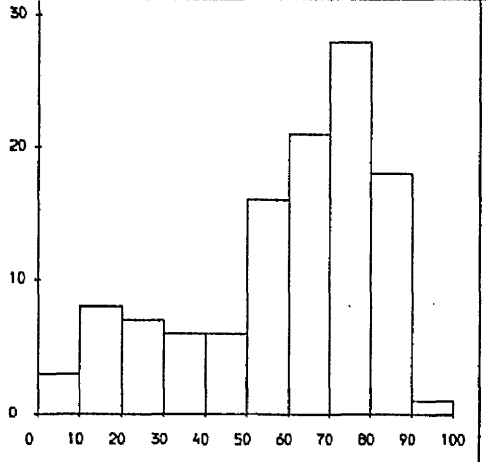


Figure 7.9(c):- Age distribution for grid 3

Figure 7.9 :- Age distribution for the three grids.

For the three grids, the distribution of age was similar with over 70% of the cases aged fifty plus. In all three instances the distribution was strongly skewed, figure 7.9. This is a facet of leukaemia since in general the risk for the majority of leukaemia types increases with age and for some strains, only those persons in their late middle and old age are afflicted e.g. chronic lymphoblastic leukaemia.

On the basis of previous experience within the L.R.F. particular interest was centred on three of these disease groupings namely,

1. all lymphoproliferative diseases
2. all myeloproliferative disorders
3. acute myeloid leukaemia.

The main reason for opting for these three categories was firstly that, high dose studies have tended to identify myeloid leukaemia as one consequence of exposure. Other reasons for the separation into lymphoproliferative and myeloproliferative groups is because, aeteologically, they are easily separable. The extraction of acute myeloid leukaemia is because it is a well recognised group which is highly malignant. Further breakdown was avoided since some of the categories were already very small.

For the latter two disease categories, a set of controls were supplied by the Leukaemia Research Fund. The all lymphoproliferative disease category was not supplied with a distinct control set but it was suggested, (F. Alexander, per comms.), the all myeloproliferative and acute myeloid leukaemia controls should be combined and serve as the controls. For categories two and three, controls were produced using a matching ratio of 3:1 for each grid and a multinomial allocation was used to assign the appropriate number to the available post-codes with probabilities proportional to :-

$$\sum_i v(i)n(i)$$

where i - age stratum 0-14, 15-64, 65-79
v(i) - overall L.R.F. age-specific incidence
n(i) - estimated stratum population for the post codes

§7.10.3 Population Data

Although the occurrence of the leukaemia cases of interest spanned the five years 1984-1989, the most readily available population data was from the 1981 small area statistics (SAS).

A wide range of different areas are available for describing the population base of interest including enumeration districts, electoral wards, local government districts, counties and civil parishes. Optimally the unit selected should be small enough to reveal patterns of interest and yet large enough to present reasonably stable data. Table 7.13 illustrates the difference in size between three areal divisions :-

	Districts	Wards	Enumeration Districts
Cornwall	7	130	1109
Devon	10	256	2334
Somerset	5	153	1046

Table 7.13 :- Comparison of the sizes of population areas for three regions.

Civil parishes as defined at the local government reorganisation in 1974 are either single enumeration districts or the amalgamation of two or more. An enumeration district, the smallest areal descriptor covers approximately 150 households. Finally a ward describes a local government district electoral ward as existed at census day (5^{th} April 1981).

The nature of the ensuing analysis prompted the utilisation of both civil parishes and enumeration district data. The enumeration district data gave access to the centroids of each district, hence it was more suitable for the generation of a population surface. In rural regions, enumeration districts are much larger than in an urban environment, however this apart, they enabled a better defined surface to be produced due to the larger number of districts within the area of interest as opposed to either wards or civil parishes.

§7.10.4 Data Summary

The preceeding three sections have described various aspects of the data. The most apparent feature being the difference in the spatial resolution for each of the data sets and the contrast in the sizes of the data sets, table 7.14.

Data Source	Spatial Resolution	Data Quantity		
		Grid 1	Grid 2	Grid 3
Radiation data	1km × 1km	3240	1116	2867
Leukaemia data	point location	212	51	101
Population data	civil parishes	60	32	47

Table 7.14 :- Summary of the spatial resolution and quantity of the three data sources.

The epidemiological data is extremely sparse by nature and concentrated in urban areas, depicting it as a continuous surface is liable to introduce large amounts of inherent error. The use of enumeration districts rather than the more sparse civil parish data enables a plausible surface to be constructed. Finally the radiation surface is generated using transect data and is near continuous in description.

These differences cause problems in the ensuing analysis since the methodology developed earlier requires the data to be not too sparse and preferably depicted as a series of point values rather than areal regions as in the case of the population data. However certain aspects of the methodology developed earlier provide possible directions of investigation.

§7.11 Analysis

§7.11.1 Introduction

As mentioned previously two basic comparisons were of interest, firstly the relationship of the case locations and the overall population and secondly, the relationship between the radiation fields and case locations.

The format of the data as summarised in the preceeding section required that in addition to spatial techniques, i.e. the mapping of case locations on the radiation fields and the population surface to assess their comparability, other approaches were required, namely,

1. Histograms of the number of cases and controls in the different radiation field levels.

2. Kolmogorov-Smirnov tests to compare the cumulative distribution functions for both cases and controls.

3. Description of various summary statistics for the different radiation variables for both cases and controls.

4. Basic statistical analysis of 'risk' against radiation level.

§7.11.2 Analysis of case locations and population data

Before proceeding to examine the main hypothesis of interest, 'Is there an association between the spatial dispersion of background radiation fields and the location of leukaemia cases?', it was essential to examine whether any anomalies existed between the location of leukaemia cases and the underlying population for the three regions of interest. The interpretation of areas of 'high' radiation levels where a large number of cases are located depends on the underlying population density of the area. The following graphical procedures examined this preliminary question:-

1. Lorenz curves.
2. A map of the population density with cases superimposed.
3. A weighted difference histogram.

Graphical results are presented for grid one for each of the three leukaemia groups. For the remaining two grids the results will be briefly summarised. Full details of the pilot study and the statistical analysis are given in Sanderson et al. (1992).

The first technique was that of the Lorenz curve, figure 7.10, described in section 4.3.1. There appears to be some disparity between the case population and the underlying general population. This increases from disease categories one to three. However caution should be expressed in the interpretation of these results. The major difficulty when implementing this and subsequent techniques stem from working with a grid whose bounds are clearly defined by straight lines. How should areal regions which are only partially contained within the grid be dealt with? Three plausible approaches are:-

1. Only include those parishes which are wholly contained within the grid.

2. Estimate the proportion of the population which falls within the grid.

3. Incorporate all civil parishes but display caution in the interpretation of the results.

The plots have all been presented with all the parishes contained within the bounds, either partially or entirely.

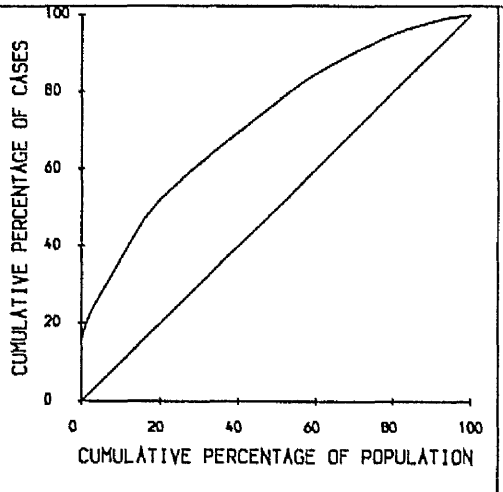


Figure 7.10 (a) :- All lymphoproliferative cases.

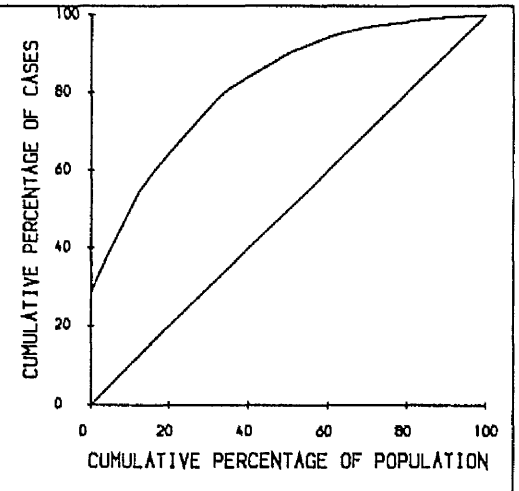


Figure 7.10 (b) :- All myeloproliferative cases.

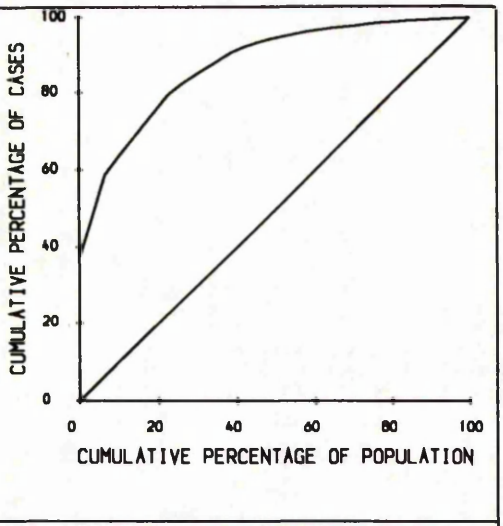


Figure 7.10 (c) :- Acute myeloid leukaemia cases.

Figure 7.10 :- Lorenz curves for the three diagnostic categories of interest for grid 1.

The three measures cited in chapter 4.3.1 for quantifying the apparent level of dissimilarity within a Lorenz curve are summarised in table 7.15. As anticipated from the preceeding discussion, the levels of dissimilarity are fairly high with acute myeloid leukaemia displaying the weakest association between cases and population.

Dissimilarity measure.	All lymphoproliferative diseases.	All myeloproliferative diseases	Acute myeloid leukaemia
D_m	32.2668	46.2941	57.0049
D_l	32.2671	46.2942	57.0051
D_f	44.3800	61.6900	73.8960

Table 7.15 :- Results for various dissimilarity indices for Lorenz curves.

Figure 7.11 illustrates a histogram based approach to describe the potential association/differences between the relative frequency of cases and population for the civil parishes contained within the region. A second histogram explicitly depicts the differences in the two frequencies which are shown superimposed in the first diagram. The problem of incorporating parishes only partially enclosed in terms of the number of cases still materialises, but unlike for Lorenz curves, it is possible to identify those parishes where the greatest disparity occurs and check their location. Any anomalies are

easily identifiable using this mode of presentation. A similar set of conclusions are reached as for Lorenz curves.

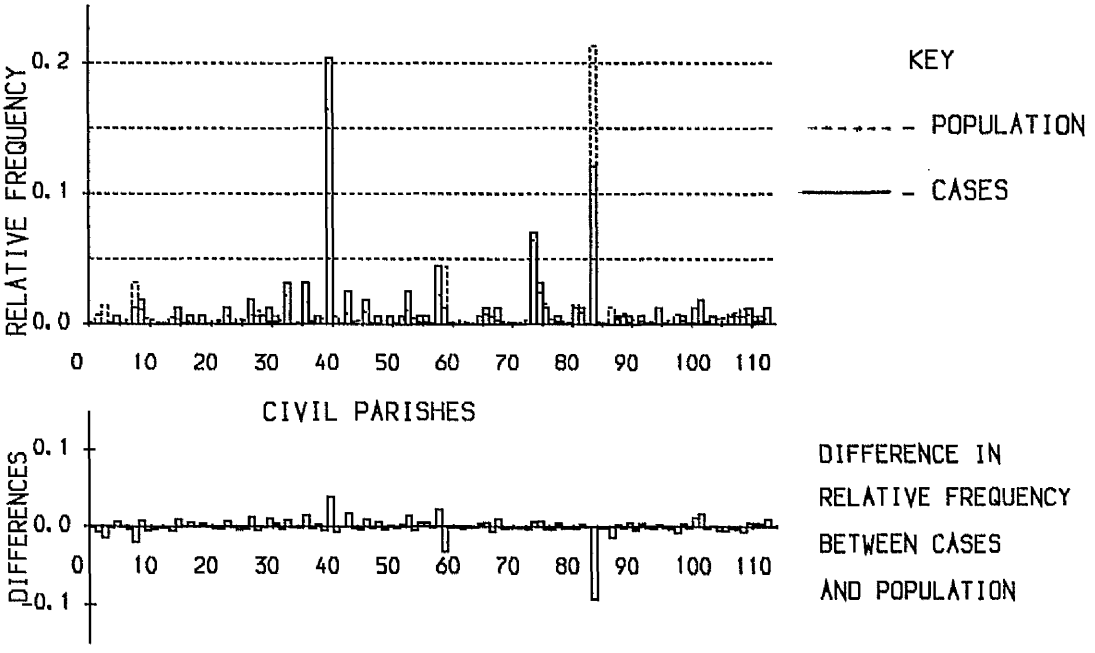


Figure 7.11(a) :- Comparison of civil parish population and number of all lymphoproliferative cases.

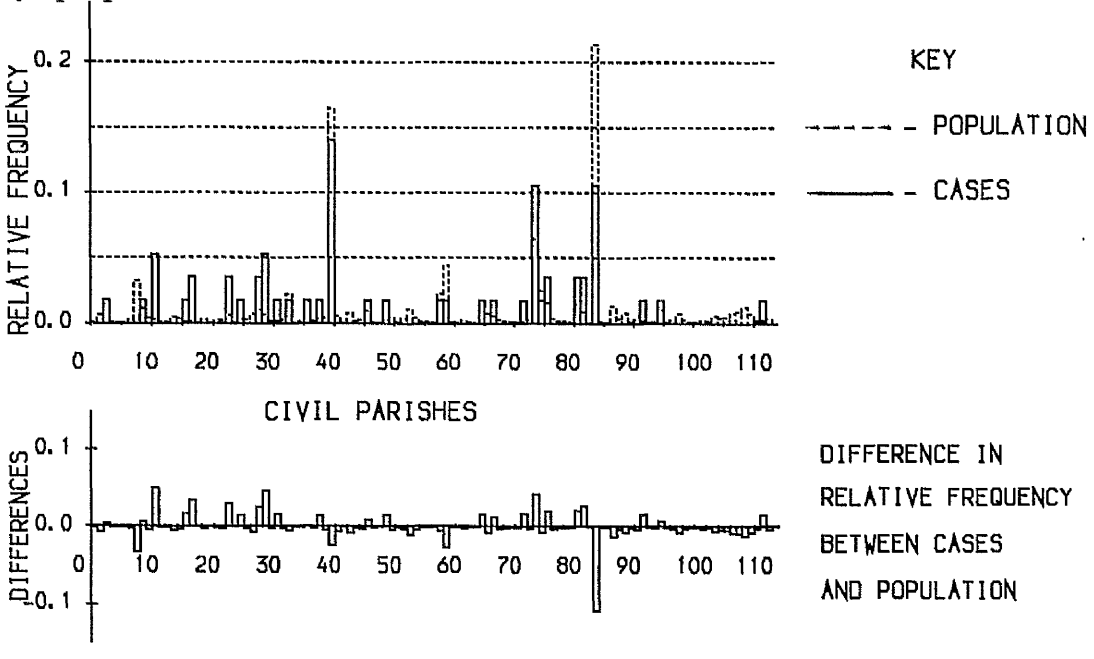


Figure 7.11 (b) :- Comparison of civil parish population and number of all myeloproliferative cases.

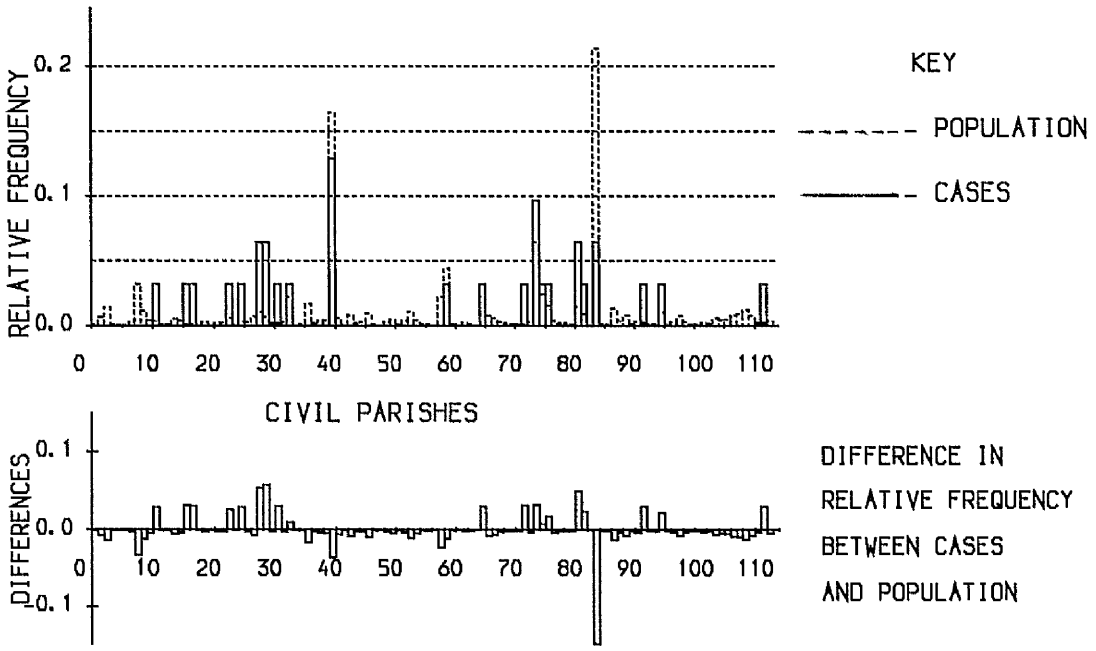


Figure 7.11 (c) :- Comparison of civil parish population and number of acute myeloid leukaemia cases.

Figure 7.11 :- Comparison of civil parish population and number of cases for various disease types.

The preceeding two approaches were applicable due to the choropleth style of the population data. The availability of centroid data for enumeration districts enabled a spatial based approach to the problem to be adopted. The population surface was constructed using the a-priori criteria of section 3.9. The sparsity and local concentration of cases in a small number of regions restricts the feasibility of fitting a surface through these points. Representing the cases as points conveys a good impression of the spatial dispersion of the cases in relation to the underlying population, figure 7.12, but curtails the implementation of the local hypothesis testing procedure.

Once again there do not appear to be any anomalous results, the majority of the cases being located in towns, suburbs or for the single cases in some of the smaller rural villages which typify this area.

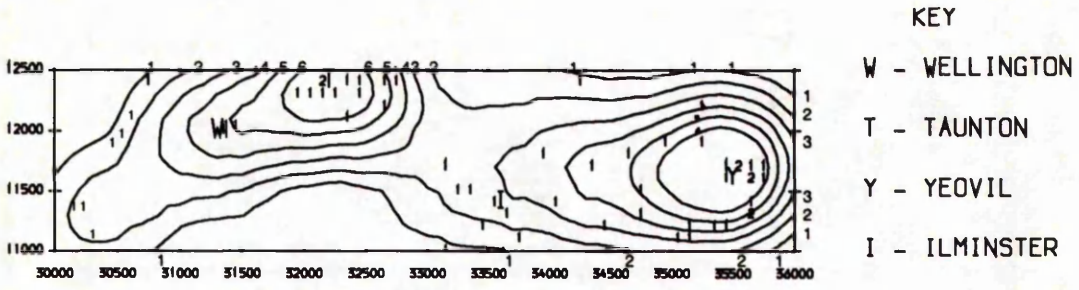


Figure 7.12 (a) :- Location of all lymphoproliferative leukaemia cases and population distribution

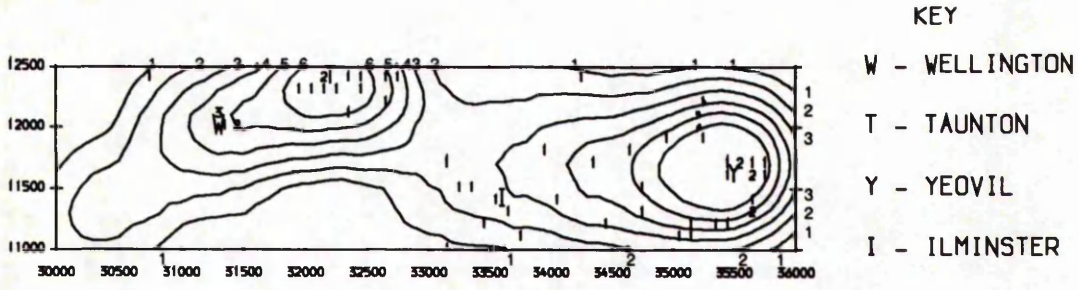


Figure 7.12 (b) :- Location of all myeloproliferative leukaemia cases and population surface

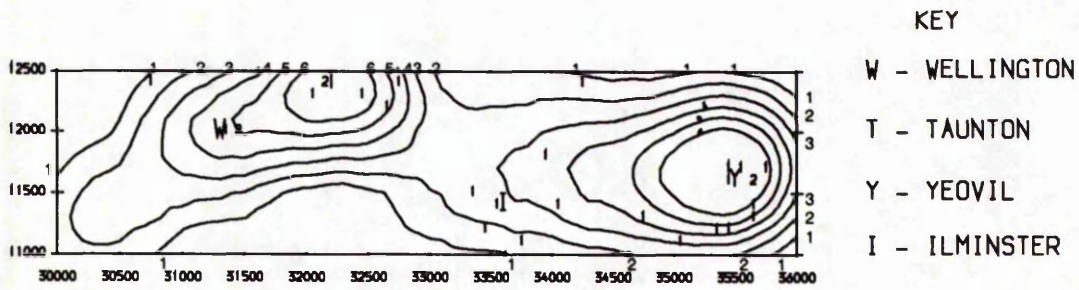


Figure 7.12 (c) :- Location of acute myeloid leukaemia cases and population surface

	Contour Level					
	1	2	3	4	5	6
Population level	1200	1700	2100	2600	3100	3600

Figure 7.12 :- Location of leukaemia cases in relation to population surface.

Unsurprisingly for all three methods, the preliminary analysis of case and population densities in the main is controlled by the population density. A similar set of general trends are apparent for grids two and three hence it may be concluded, that the behaviour for all the grids is similar in terms of case locations and population with no obvious differences apparent.

§7.11.3 Analysis of radiation fields and case locations

The previous section highlighted the difficulty of separating the case and population distributions. The next step was to examine whether background radiation is instrumental in determining the location of leukaemia cases i.e. are the cases of leukaemia primarily located in areas corresponding to high/low radiation levels.

One of the major drawbacks of an analysis of this type is the paucity of the case data set and secondly the matching of the cases and controls to a radiation grid cell. Due to the shortcomings of the epidemiological data, the grid cell is the one in which the case was located at the time of diagnosis, it may not represent the value of the radiation field at the time of conception when the cells are most sensitive nor will it describe the total radiation life history of the case.

In terms of the restricted size of the data set and the breakdown into three separate grids, major differences would be required between the case and control radiation levels for any change to be detected.

^{40}K			Gamma		
Difference ($kBqm^{-2}$)	cases (n_1)	controls (n_2)	Difference (mGy/a)	cases (n_1)	controls (n_2)
5	12383	37149	0.01	530	1590
10	3096	9287	0.015	236	708
15	1376	4128	0.02	133	399
20	744	2322	0.05	22	66
25	496	1488	0.07	11	33
30	344	1032	0.10	5	15

Table 7.16 :- Sample sizes to detect change of specific size.

Table 7.16 summarises the approximate sample sizes which would be required to detect a change in the radiation level between a group of cases and controls for a 5% significance level, assuming both normality and a value for the standard deviation.

In terms of typical values for natural variability over Britain, a change of 0.01mGy/a in the background level of gamma radiation corresponds to an increase/decrease of 3% in the natural variability, whilst the detection of a change of 0.1mGy/a equates to a 30% change in the natural levels over Britain.

Within this section the underlying population is represented by a set of control locations selected as described in section 7.10.1. Each case and control location, identified by its ordinance survey co-ordinate, was matched using a nearest neighbour technique to the appropriate radiation field.

Figure 7.13 displays the raw histograms of the radiation features, for the nuclide ^{40}K, for the three disease categories for grid one. The bins were selected in accordance with a scheme implemented by Sanderson et al (1990). For the case and, to a lesser extent, the control distribution, two features were apparent, bimodality and skewness.

For the remaining five radiation fields similar trends were reported. Also for grids two and three the range of radiation values was comparable between the cases and controls. Bimodality is weaker in grid two than for either of the other two grids, however the lack of symmetry property is present in both grids.

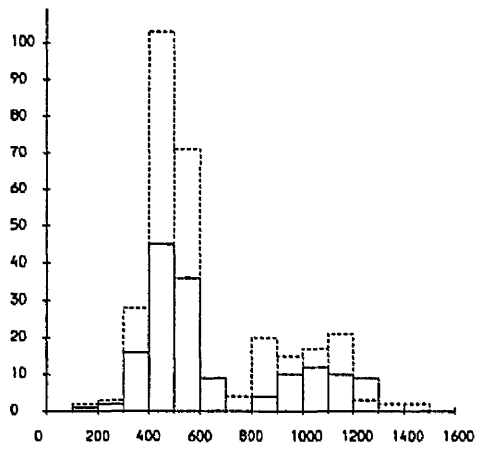


Figure 7.13 (a) :- All lymphoproliferative diseases.

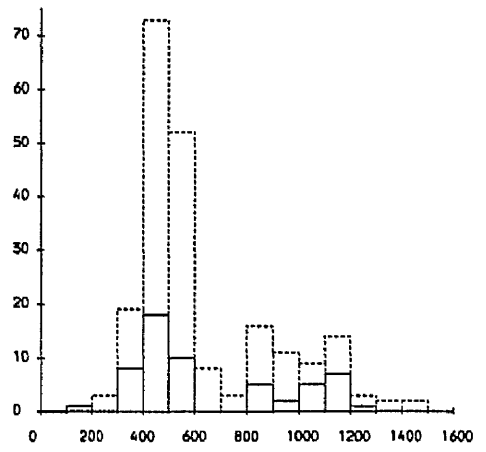


Figure 7.13 (b) :- All myeloproliferative diseases.

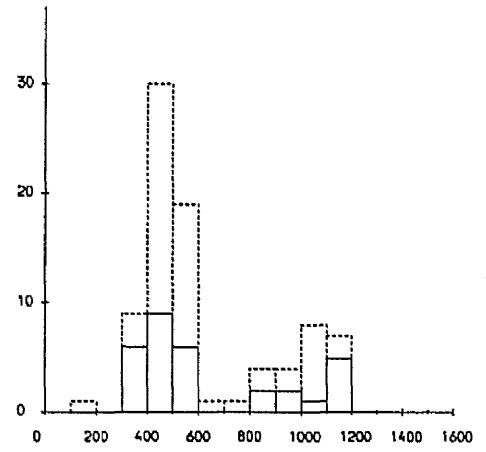


Figure 7.13 (c) :- Acute myeloid leukaemia.

KEY

Cases

Controls

Figure 7.13 :- Histogram of cases and controls for grid one for the three diagnostic categories of interest.

For both sets of information there is evidence of bimodality and lack-of-symmetry, hence performing the standard parametric statistical tests for comparing univariate data

sets would be inappropriate, due to the violation of the basic assumptions. On this premise, table 7.17 collates a set of summary statistics for grid one for the three conditions of interest and the six radiation fields.

All lymphoproliferative diseases

	N	Summary statistic	40_K	eTh	eU	Alpha	Beta	Gamma
Cases	154	median	529.1	39.98	49.78	18.225	2.260	0.400
		minimum	174.6	18.88	25.64	8.96	1.10	0.200
		maximum	1243.3	60.70	68.5	24.7	4.30	0.660
Controls	300	median	514.1	38.87	49.46	17.99	2.230	0.395
		minimum	195.0	18.44	23.11	8.74	0/930	0.190
		maximum	1422.2	66.61	72.23	25.04	4.68	0.670

All myeloproliferative diseases

	N	Summary statistic	40_K	eTh	eU	Alpha	Beta	Gamma
Cases	57	median	517.1	39.17	48.5	17.60	2.20	0.400
		minimum	184.2	23.74	28.61	12.03	1.31	0.270
		maximum	1223.4	52.47	62.47	22.68	4.21	0.640
Controls	216	median	516.2	38.81	49.13	17.86	2.23	0.395
		minimum	195.0	19.63	23.11	8.92	0.95	0.200
		maximun	1422.4	66.61	72.23	24.70	4.68	0.670

Acute Myeloid Leukaemia

	N	Summary statistic	40_K	eTh	eU	Alpha	Beta	Gamma
Cases	31	median	517.1	38.7	48.0	17.60	2.160	0.380
		minimum	331.0	23.74	35.7	12.03	1.470	0.290
		maximum	1190.3	52.28	61.3	21.96	4.210	0.640
Controls		median	503.6	39.06	50.66	18.42	2.225	0.395
		minimum	198.6	18.44	25.3	8.74	0.930	0.190
		maximum	1192.0	56.62	69.81	25.04	4.230	0.640

Table 7.17 :- Summary statistics for grid one.

For both all lymphoproliferative diseases and acute myeloid leukaemia the median value for ^{40}K is noticeably lower for the controls. For the other radiation fields, differences are less noticeable although, in general, for all lymphoproliferative diseases, the median values for the controls are lower. One potentially interesting feature arises for acute myeloid leukaemia where, with the exception of ^{40}K, the cases all have a lower radiation value than the controls.

These results indicate that the collation of the data for all three grids would possibly mask some interesting features, however as emphasised earlier to detect significant differences, the changes are required to be fairly large.

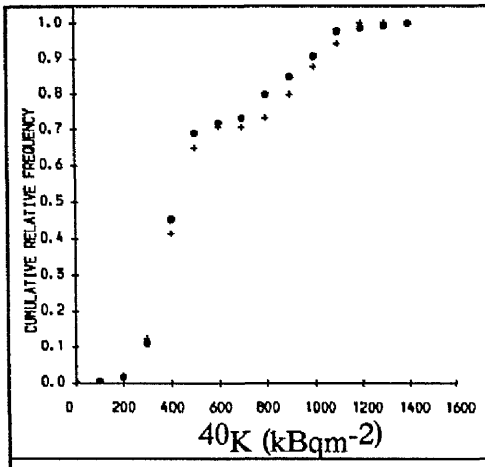


Figure 7.14 (a) :- All lymphoproliferative diseases.

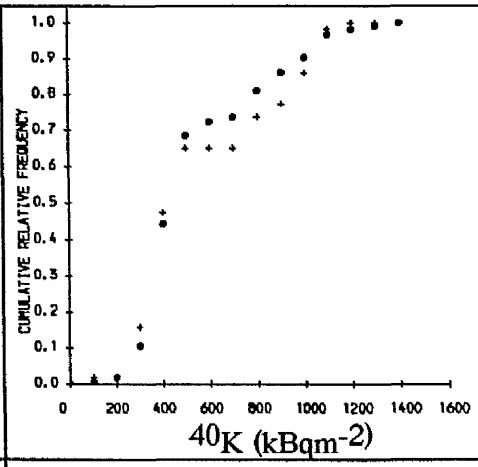


Figure 7.14 (b) :- All myeloproliferative diseases.

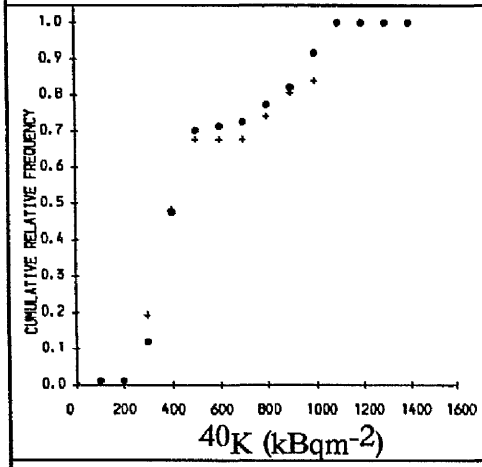


Figure 7.14(c) :- Acute myeloid leukaemia.

KEY

Cases

Controls

Figure 7.14 :- Cumulative distribution function for the three disease categories.

The third technique implemented was to compare the case and control figures using the Kolmogorov-Smirnov test, the cumulative distribution functions are given in figure 7.14. For all three grids and disease characteristics none of the results are significant, table 7.18.

Once again the bimodality feature described earlier is apparent from the diagrams.

Disease Group	Value of Test Statistic	Critical Value for a 5% significance level
All lymphoproliferative diseases	0.60526	1.36
All myeloproliferative diseases	0.67155	1.36
Acute myeloid leukaemia	0.47585	1.36

Table 7.18 :- Results of Kolmogorov-Smirnov test.

The preceeding techniques in this section have all ignored the spatial context of the radiation data and as seen from the previous example on climatic change, this may result in some of the underlying complexities being smoothed over. Within the remainder of the section, a spatial based approach will be adopted to assess the cause of the bimodality.

Initially, the generation of a radiation surface was hindered due to a problem in the selection of a suitable smoothing parameter, h, since minimisation of the score function was not possible. The source of the problem was traced to the near repeatability of points within the data set. Silverman (1978) stated that real data are nearly always rounded or discretised to a greater or lesser degree. For a discretised data set $x_1, x_2, \ldots, x_n$, let m be the number of pairs $i < j$ for which $x_i = x_j$. If a data set of size n is discretized to a grid of k points, no matter how the data points fall, it can be shown by Jensen's inequality, Feller (1966);-

$$\frac{m}{n} \geq \frac{1}{2}\frac{n}{k} - 1 \qquad (1)$$

If the data are all nonuniform, m will generally be much larger than the minimum value given by (1). If $\frac{m}{n}$ is larger than some theoretical value β, depending on the kernel k, then the least squares cross validation score function tends to minus infinity as $h \rightarrow 0$ and hence least squares cross validation as it stands will choose the degenerate value h=0 for the window width.

Not only is it dangerous to use least squares cross validation on discretised or repeated data but Silverman (1978) emphasises that the behaviour of the score function for small h is highly sensitive to a very fine small scale effect in the data.

One approach to avoid this problem is to perturb the data points by a small amount. The data points were gradually perturbed by an increasing amount until a realistic estimate of the smoothing parameter was achieved. This amount is individual to each data set. In practice the aim is to perturb the data set sufficiently that a value of h is selected which is non-infinite. For the case of the radiation surface, it was decided to ensure that by perturbing the data, the point did not encroach into a neighbouring grid cell, i.e. the limits of perturbation were 1km × 0.5km, the spatial resolution of the data.

In an attempt to locate the cause of the bimodality, the information garnered from the histograms and cumulative distribution functions was used. The case data was split into two groups at the point where the two distributions appear to split. For ^{40}K this was taken as 500kBqm^{-2}. Figure 7.15 (a) describes the radiation surface for grid one, with the location of the acute myeloid leukaemia cases superimposed, whose levels were less than 500kBqm^{-2}. Figure 7.15 (b) describes the scenario for cases with levels greater than 500kBqm^{-2}.

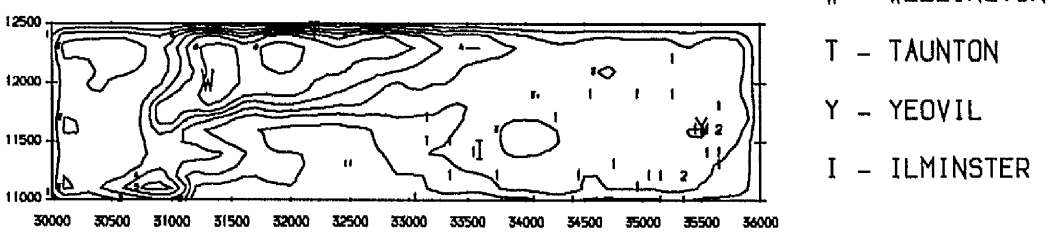
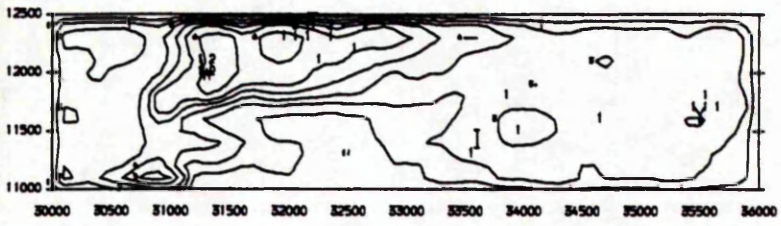


Figure 7.15 (a) :- ^{40}K radiation surface with cases whose levels < 500kBqm^{-2} superimposed.



KEY

W - WELLINGTON

T - TAUNTON

Y - YEOVIL

I - ILMINSTER

Figure 7.15(b) :- ^{40}K radiation surface with cases whose levels > 500kBqm^{-2} superimposed.

Figure 7.15 :- ^{40}K radiation surface with cases superimposed.

A fairly-conclusive explanation of the bimodality emerges. The first set of cases (^{40}K < 500kBqm^{-2}) were centred around Yeovil, where the radiation levels are generally lower than those in the Taunton region and the second group were focused in the Wellington region, where levels were higher. A number of anomalies are apparent but it appears the geology of the area is a major factor in causing the bimodality. Although not so pronounced the same pattern was repeated for the controls.

A similar diagnosis may be placed on the bimodality for grid three, but it is less striking. The cases with lower levels of ^{40}K occur in the Torpoint, Saltash area, whilst the cases with higher levels are scattered in the north of the region.

§7.11.5 Analysis Based On A Leukaemia Rate

Preliminary investigation of whether there is a link between the location of the cases and the radiation field has been graphical in nature to this point. For a more formal approach to the investigation of the hypothesis of interest, a leukaemia rate was defined:-

$$\text{Leukaemia rate} = \frac{\text{Number of cases in region x}}{\text{Population in region x}}$$

As mentioned previously, three sources of information are available:-

1. Case data for the three grids surveyed, identified by the ordnance survey co-ordinates of the postal district for that case.

2. Radiation data, in integrated form, typically corresponding to a spatial resolution of 500m ×1000m. Each radiation value was associated uniquely with the ordnance survey co-ordinates of the centroid of the grid cell.

3. Population data from the 1981 census, for each of the enumeration districts contained within the grids.

For this problem, the region of interest was sub-divided into grid cells of size 500m × 500m. For each cell a value was calculated for :-

1. number of cases
2. radiation level in region x
3. population in region x

The number of cases was easily derived for each cell by simply checking whether a point lay within the cells bounds. The radiation levels were also simple to evaluate since from the aerial survey values had been calculated for a resolution of 1km × 0.5km, by using a nearest neighbour technique, an appropriate value was calculated for the centroid of the cell.

For the population in region, x, a more complicated approach was required since an enumeration district spanned both complete and segments of grid cells. As an initial solution to the problem, a surface produced using kernel density estimation was fitted to the data and the population for a given small area was evaluated by numerically integrating under the estimated surface.

For the analysis based on a leukaemia rate, the radiation data was categorised into eight groups defined a-priori. For these groups, the population and case figures for all grid cells of a particular radiation level were summed. The leukaemia rate was then estimated. The resultant data was of of the form, (x_i,t_i): where x_i is the leukaemia rate, t_i the radiation value (mid-value) in category i. For some of the analysis, it was necessary to collapse the eight cells down to four due to the sparsity of the leukaemia

data, table 7.19 describes the results for grid one for ^{40}K, for all lymphoproliferative diseases, (ALL).

Radiation value ($kBqm^{-2}$)	Population total	No. of cases (ALL)	Rate (x_i)	Combined rate
100	13.77	1	72.60×10^{-3}	
300	3518	18	5.117×10^{-3}	5.380×10^{-3}
500	18813	84	4.465×10^{-3}	4.465×10^{-3}
700	995	5	5.025×10^{-3}	5.025×10^{-3}
900	5041	16	3.174×10^{-3}	3.174×10^{-3}
1100	4187	24	5.732×10^{-3}	5.732×10^{-3}
1300	712.7	3	4.209×10^{-3}	3.854×10^{-3}
1500	65.8	0	0	

Table 7.19 :- Example of leukaemia rates for grid one.

Performing any statistical test on a maximum of eight data points does not provide conclusive evidence of an association, but it may indicate that an analysis of this type is plausible for data sets of this form, or alternatively, there may be preliminary indications of an association, which may advocate the implementation of a larger study.

A regression analysis leukaemia rate against radiation level displayed a general lack of consistency between both grids and leukaemia types. For some situations a quadratic fit was the best, whereas a linear trend sufficed for others. The major problem lies in attempting to summarise a complex and variable situation in a simplistic manner. The error rate on the variables are known to be large especially for the case data, where error rates follow a Poisson distribution i.e. $\sqrt{n}$. The use of Poisson regression may be more applicable in this instance and it has been used in the study of cancer incidence, Gail (1978).

The second analysis of rate performed was an analysis of variance. For each of the six different radiation variables (eU, eTh, K^{40}, α, β and γ), the factors examined were grid, leukaemia type and level of radiation value. The response variable was leukaemia rate.

Model :- $$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} \qquad (1)$$

$$\sum_i \alpha_i = 0 \quad \sum_j \beta_j = 0 \quad \sum_k \gamma_k = 0 \quad \sum_i (\alpha\beta)_{ij} = 0 \ \forall i \quad \sum_j (\alpha\beta)_{ij} = 0 \ \forall j$$

where X_{ijk} = leukaemia rate for grid i, leukaemia type j, and radiation level k.

By adopting model (1), we are attempting to assess which factors are important in influencing the rate. Table 7.20 reports the results.

Radiation Field	Leukaemia type	Grid	Radiation Level	Leuk. type/grid
K^{40}	Significant	Non-Significant	Significant	Non-Significant
eU	Significant	Significant	Significant	Non-Significant
eTh	Significant	Non-Significant	Non-Significant	Non-Significant
α	Significant	Significant	Non-Significant	Significant
β	Significant	Significant	Non-Significant	Significant
γ	Significant	Significant	Significant	Significant

Table 7.20 :- Results of performing an analysis of variance on grid (1,2,3), radiation level (1-8) and leukaemia type (1,2,3,4).

From table 7.20, differences exist between the leukaemia types, but also between the grids for some of the radiation fields. The first of these results is unsurprising since it is well known that the rates for the three types of leukaemia differ, table 7.21, but the second result suggests that the rates differ between grids, this may indicate contrasting environmental reasons. The radiation variable was found to be significant for total gamma, ^{40}K and eU the last two being of likely geological origin, this again indicates differences between the grids. The confounding nature of the analysis clearly makes it

difficult to disentangle the individual effects, and the apparent differences amongst grids argue against increasing the sensitivity of the analysis by pooling the data from the three grids.

Disease Type	Relative risk for Cornwall
All lymphoproliferative diseases	102.5 - 110.0
All myeloproliferative diseases	97.5 - 102.5
Acute myeloid leukaemia	90.0 - 97.5

Table 7.21 :- Differences in rates between the three leukaemia types for Cornwall.

§7.12 Summary

The study and its analysis has demonstrated the potential for linking epidemiological data and radiation fields. The interpretation of the results is not however clear-cut, the three geographical regions demonstrate interesting differences, with grid two showing anomalous features, not least an indication that cases appear to be associated with lower radiation levels than controls.

Difficulties encountered include the paucity of the leukaemia data which serves as a restriction to the implementation of the hypothesis methodology and other forms of statistical analysis. The usage of three grids as oppoosed to one, dimished the potential power of the study, hence only tentative conclusions could be reached. The spatial framework of the analysis which supported many of the ideas in the preceeding sections would be plausible for diseases which are more common. By utilising a spatial analysis, a clearer picture of what is going on is obtainable. Addressing the problem, by ignoring the spatial dimension leaves many questions unanswered, as demonstrated by the location of the apparent source of bimodality using a spatial approach.

Appropriate developments for further work would be to :-

1. select two regions with very different radiation features

and/or

2. increase the number of cases.

§7.13 General Conclusions

The two case studies have demonstrated the pros-and-cons of the methodology developed. The climatic example has shown the simplicity of implementing the hypothesis transformation technique at both a local and global level. It also emphasises the need for a spatial framework on which to base an analysis. Many of the analyses carried out ignored the spatial dimension, illustrated a strong relationship between the two sets of data, but on closer examination, it was only to be anticipated from the nature of the data. The local hypothesis testing procedure enabled a much more detailed resumé of the changes to be portrayed with some interesting features concerning differences between regions being highlighted.

Although for the second example no formal analysis was undertaken to establish whether a link exists between radiation levels and case locations in the south-west of England, this was primarily a result of the design of the pilot study, however various graphical and semi-quantitative approaches were undertaken. Firstly the analysis of three disjoint regions resulted in the already limited number of cases being reduced to sizes less than desirable for any formal statistical analysis. The selection of one complete grid would have increased the power of the study and the problem of bimodality would possibly not have manifested itself. This apart, the methodology of hypothesis testing prompted the form of the subjective spatial analysis undertaken.

In general, from the examples examined, it appears that the local hypothesis testing procedure is a step in the right direction when comparing two sets of data which may or may not be compatible in terms of spatial resolution. It addresses the complexity of the data introduced by the spatial dimension and may also give the analyst a pointer as to the cause of the change, if they are unaware of the form of the change i.e. scalar, rotational or translational. A number of problems, especially in terms of disjoint contours, still exist with the method but hopefully these can be resolved .

REFERENCES

Abramson, I.S. (1982). On bandwidth variation in kernel estimates - a square root law. *Ann. Statist,* **10**, 1217-1223.

Agee, E.M. (1982). A diagnosis of twentieth century temperatures at West Lafayette, Indianna. *Climate Changes* **4**, 399-418.

Agterberg, F.P. (1964). Methods of trend surface analysis. *Quarterly Colorado of Mines,* **59**, 111-130.

Anderson, P. (1970). The uses and limitations of trend surface analysis in studies of urban air pollution. *Atmospheric Environment,* **4**, 129-147.

Armstrong, M. (1984). Problems with universal kriging. *Mathematical Geology,* **16(1)**, 101-108.

Armstrong, R.W. (1969). Standardised class intervals and rate computation in statistical maps of mortality. *Annals Association American Geographer,* **59**, 382-390.

Baddeley, A.J. (1987). A class of image metrics. *Proc. ANZ AAS Congress,* Townsville, Queensland, Australia.

Bengtsson, B.E. and Nordbeck, S. (1964) . Construction of isarithms and isarithmic maps by contour. *Report BIT4, University of Lund, Lund,* 87-105

Berezin, I.S. and Zhidkov, I.P. (1965). *Computing Methods,* **Vol. 1**, Chapter 2. Addison-Wesley, Reading, Mass.

Besag, J. (1986). On the statistical analysis of dirty pictures (with discussion). *Journal Royal Statistical Society B,* **48**, 259-302.

Besag, J. and Kempton, R.A. (1986). Statistical analysis of field experiments using neighbouring plots. *Biometrics,* **5**, 351-360.

Besel, P.J. and Jain, R.C. (1988). Segmentation through variable order surface fitting. *IEEE transactions on pattern analysis and Machine Intelligence,* **10**, 167-192.

Birkhoff, G. and Mansfield, L. (1974). Compatible triangular finite elements. *Journal Mathematical Analysis and Applications,* **47(3)**, 531-553.

Bithell, J.F. (1990). An application of density estimation to geographical epidemiology. *Statistics in Medicine,* **10**,691-701.

Blair, D.J. and Bliss, T.H. (1967). The measurement of shape in geography. *University of Nottingham, Department of Geography, Bulletin of Quantitative Data for Geographers,* **11**.

Blum, H. (1973). Biological shape and visual science. *Journal Theoretical Biology,* **38**, 205-287.

Blumenstock, D.I. (1953). The reliability factor in the drawing of isarithms. *Annals of the Association of American Geographers,* 290-304.

Bookstein, F. L. (1978). The measurement of biological shape and shape changes. In: Levin S. (ed) *Lecture notes in biomathematics 24.* Springer-Verlog New York.

Bookstein, F. L. (1984a). A statistical method for biological shape comparisons. *Journal of Theoretical Biology,* **107**, 475-520.

Bookstein, F. L. (1984b). Tensor biometrics for changes in cranial shape. *Annals of Human Biology,* **11**, 413-437.

Bookstein, F.L. (1986). Size and shape spaces for landmark data in two-dimensions. *Statistical Science,* **1**, 181-242.

Boots, B.N. and Lamoureaux (Jr.), M.S. (1972). Working notes and bibliography on the study of shape in human geography and planning. *Council of Planning Librarians,* Exchange Bibliography, **346**, 1-22.

Bowman, A.W. (1984). An alternative method of cross-validation for the smoothing of density estimators. *Biometrics,* **71**, 353-360.

Boyce, R. R. and Clark, W.A.V. (1964). The concept of shape in geography. *Geographical Review,* **54**, 561-572.

Bradley, R. (1970). The excavation of a beaker settlement at Belle Tout, East Sussex. *Proceedings of the Prehistoric Society,* **36**, 312-379.

Breiman, L., Meisel, W. and Purcell, E. (1977). Variable kernel estimates of multivariate densities. *Technometrics,* **19(2)**, 135-144.

Brooks, C.E.P. and Carruthers, N. (1953). *Handbook of Statistical Methods in Meteorology,* Her Majesty's Stationary Office, London.

Cacoullos, T. (1964). *Technical Report No. 40.* Dept of Statistics, University of Minnesota.

Cacoullos, T. (1966). Estimation of a multivariate density. *Annals Inst.Statist. Math.,* **18**, 179-189.

Clarke, I. (1979). *Practical Geostatistics*. Applied Science Publishers, London.

Cleek, R.K. (1979). Cancers and the environment: The effect of scale. *Society Science Medicine,* **13D**, 241-247.

Cliff, A.D. (1970). Computing the spatial correspondence between geographical patterns. *Transactions, Inst. British Geographer,* **50**, 143-154.

Cliff, A.D., Haggett, P., Ord, J.K., Bassett, K. and Davies, R.B. (1975). *Elements of Spatial Structure: A Quantitative Approach.* Cambridge University Press, London.

Cliff, A.D. and Ord, J.K. (1981). Spatial processes: Models and applications. London: Pion.

Cole, J.O. (1964). Study of major and minor civil divisions in political geography. *Paper presented at 20th International Geographical Congress, Sheffield, England.*

Coombes, A., Linney, A.D., Richards, R. and Moss, J. (1991). A method for the analysis of the 3-D shape of the face and changes in the shape brought about by facial surgery. *Biostereometrics Technology and Applications,* ed. Herron R., Procc. SPIE 1380, 180-189

Coons, R.L., Woolard, G.P. and Hersberg, G. (1967). Structural significance and analysis of mid-continent gravity high. *Bulletin of the American Association Of Petroleum Geologists,* **51**, 2381-2399.

Cottafava,G. and le Moli, G. (1969). Automatic Contour Maps. *CACM, 12,* 386-391.

Court, A. (1970). Map comparisons. *Economic Geography,* **46**, 435-438.

Crane, C.M. (1972). Contour plotting for functions specified at nodal points of an irregular mesh based on an arbitrary two parameter co-ordinate system. *The Computer Journal,* **15**, 382-384.

Dacey, M.F. (1964). Measures of contiguity for 2-color maps. *Technical Report,* Northweston University.

Davies, J. (1973). *Statistics and Data Analysis in Geology, (2nd edition).* John Wiley and Sons Inc.

Dayhoff, M.C. (1963). A contour map program for x-ray crystallography. *CACM,* **6**, 620-622.

Diaz, H.F. and Quayle, R.G. (1980). The climate of the United States since 1895: Spatial and temporal changes. *Monthly Weather Review,* **108**, 249-266.

de Boor, C., (1962). Bicubic spline interpolation. *Journal of Mathematics and Physics,* **41**, 212-218.

Dionian, J., Muirhead, C.R., Wan, S.L. and Wrixon, A.D. (1986). The risks of leukaemia and other cancers in Thurso from radiation exposure. *Report NRPB-R196.London,Her Mayesty's Stationery Office.*

Dobson, M.W. (1973). Choropleth maps without class intervals? A comment. *Geog. Anal.,* **5**, 358-360.

Efron, B. (1979). Bootstrap Methods: Another look at the jacknife. *The Annals of Statistics* **7**, 1-26.

Erlich, R., Baxter Pharr, R., Healy-Williams, H. (1983). Comments on the validity of Fourier descriptors in systematics: *A reply to Bookstein et al.* Syst. Zool., **32(2)**. 202.

Evans, I. (1976). The selection of class intervals. *Transactions, Institute of British Geographers,* **2(2)**, 98-124.

Evans, I.S., Catterall, J.W. and Rhind, D.W. (1975). Specific transformations are necessary. *Census Res.* Unit Working, Paper **4**, University of Durham.

Faraway, J.J. and Jhun, M. (1990). Bootstrap choice of bandwidths for density estimation. *Journal of the American Statistical Association,* **85(412)**, 1119-1122.

Feller, W. (1966). An introduction to probability theory and its applications. *Volume 11, New York, Wiley.*

Fisher, R.A. (1935). *The design of experiments.* Oliver and Boyd, Edinburgh.

Fix, E. and Hodges, J.L. (1951). Discriminatory analysis, non-parametric estimation: consistency properties. **Report No. 4**, Project no. 21-49-004, USAF School of Aviation Medicine, Randolph Field, Texas.

Folland, C.K., Karl, T.R. and Vinnikov K.Y. (1990). Observed climatic variations and change. In *Climate change: the IPCC Scientific Assessment*, Houghton, J.T., Jenkins, G.J. and Ephraums, J.J. (eds). Cambridge University Press.

Fryer, M.J. (1977). A review of some non-parametric methods of density estimation. *Journal Inst. Maths Applications,* **201**, 335-354.

Fukunaga, K. (1972). *Introduction to statistical pattern recognition.* New York, Academic Press.

Gail, M. (1978). The analysis of heterogeneity for indirect standardised mortality ratios. *J.R.Statistical Society A,* **141**, 224-234.

Geary R.C. (1930). The frequency distribution of the quotient of two normal variances. *J.R. Statistical Society,* **93**, 442.

Gesler, W.M., Todd, C., Evans, C., Casella, G. Pittam, J. and Andrews I.I. (1980). Spatial variations in morbidity and their relationship with community characteristics in Central Harlem District. *Society Science Medicine,* **14D**, 387-396.

Gibbs, J.P. (1961). A method for comparing the spatial shapes of urban units. In Gibbs J.P., editor, *Urban Research Methods.* Princeton, New Jersey. Van Nostrand Co., Inc, 99-106.

Gini, G. (1913-1914). Sulla misura dell concentraziono e della variabilitia dei caraterri. *Atti del Reale Institute Veneto di Scienze, Lettre ed Arti,* **53**, 2.

Gold, C.M., Charters, T.D. and Ramsden, J. (1977). Automated contour mapping using triangular element data structures and an interpolant over each irregular triangular domain. *Computer Graphics,* **11(2)**, 170-175.

Goodman, L.A. and Kruskal, W.H. (1954) Measures of association for cross-classification Part 1. *Journal American Statistical Association,* **49**, 732-764.

Gotshtasby, A., Stockman, G.C. and Page, C.V. (1986). A region based approach to digital image registration with sub-pixel accuracy. *IEEE Transactions on Geoscience and Remote Sensing.* **Vol GE 24(3)** 390-399.

Grant, F.A. (1957). A problem in the analysis of geophysical data. *Geophysics,* **22,** 309-344

Gray, J.M., (1972). Trend surface analysis: trends through clusters. *Area,* **4,** 102-103.

Haining, R. (1987). Trend surface models with regional and local scales of variation with an application to aerial survey data. *Technometrics,* **29(4),** 461-469.

Hall, P., Diciccio, T.J. and Romano, J.P. (1989). On smoothing and the bootstrap. *The Annals of Statistics,* **17(2),** 692-704.

Halley, E. (1686). A historical account of the trade winds and monsoons, observable in the seas between and near the tropics: with an attempt to assign the physical cause of said winds. *Philosophical Transactions,* 153-158.

Hamming, R.W. (1983). *Digital Filters.* Englewood Cliffs N.J. Prentice Hall (Signal processing Series) 2nd edition.

Handcock, M.S. and Wallis, J.R. (1990). An approach to statistical spatial-temporal modeling of meteorological fields. *Copenhagen General Assembly of the European Geophysical Society.*

Hansen, J., Lacis, A., Rind, D., Russell, G., Stone, P., Fung, I., Ruedy, R. and Lerner, J. (1984). Climate sensitivity analysis of feedback mechanisms in climate processes and climate sensitivity. *Geophysics, Monogr., Ser.,* **Vol29** ed. Hansen J. E. and Takahashi T. 130-163, AGU, Washington D.C.

Heap, B.R. and Pink, M.G. (1969). Three contouring algorithms. *DNAM Report 81,* National Physical Laboratory, Teddington.

Henshaw, D.L., Eatough, J.P. and Richardson, R.B. (1990). Radon as a causative factor in induction of myeloid leukaemia and other cancers. *The Lancet.* **335,** 1008-1012.

Hickey, R.J., Bowers, E.J., Spence, D.E., Zemel, B.S., Clelland, A.B. and Clelland, A.B. (1983). Radiation hormesis, public health and public policy: a commentary. *Health Physics.* **44,** 207-219.

Hills, M. and Alexander, F. (1989). Statistical methods used in assessing the risk of disease near a source of possible environmental pollution: a review. *J.R. Statistical Society A,* **152**, 353-363.

Horton, R.E. (1932). Drainage basin characteristics. *Transactions of the American Geophysical Union,* **13**, 350-361.

Hsu, M.L. and Robinson, A.H. (1970). The fidelity of isopleth maps, an experimental study. *University of Minnesota Press,* Minneapolis.

Hugg, L. (1979). A map comparison of work disability and poverty status in the United States. *Soc. Sci. Med* **13D**, 237-240.

Jenks, G.F. and Caspall, F.C. (1971). Error on choroplethic maps: definition, measurement, reduction. *Annals Association American Geographer,* **61**, 217-244.

Jenks, G.F. and Coulson, M.R.C. (1963). Class intervals for statistical maps. *International Yearbook of Cartography,* **3**, 119-134.

Jones, P.D., Raper, S. S. B., Bradley, R.S., Diaz, H. F., Kelly, P.M. and Wigley, T.M. (1986). Northern hemisphere surface air temperature variations 1851-1984. *Journal of Climate and Applied Meteorology* **25**, 161-179.

Journal, A.G. (1969). Rapport d'etude sur l'estimation d'une variable régionalisée. Internal Report No. N-156 CGMM.

Journal, A.G. and Huijbregts, G.J. (1978). *Mining Geostatistics.* Academic Press, London.

Karl, T.R., Diaz, H.F. and Kuklan, G. (1988). Urbanisation: Its detection and effect in the United States climate record. *Journal of Climate,* **1**, 1099-1123.

Karl, T.R. Heim, R.R. and Quayle, R.G. (1991). The greenhouse effect in Central North America: If not now, when? *Science,* **251**, 1058-1061.

Karl, T.R. Livezey, R.E. and Epstein, E.S. (1984). Recent unusual mean winter temperatures across the contiguous United States. *Bull. Amer. Meteor. Soc.,* **65**, 1302-1309.

Karl,T.R. and Williams (Jr.) C.N. (1987). An approach to adjusting climatological time series for discontinuous homogeneities. *Journal of Climate and Applied Meteorology,* **26**, 1744-1763.

Karl, T.R., Williams (Jr.), C.N., Young P.J. and Wendland W.M. (1986). A model to estimate the time of observation bias associated with monthly mean maximum, minimum and mean temperature for the United States. *Journal of Climate and Applied Meteorology,* **25**, 145-160.

Kendall, M.G. and Stuart, A. (1977). *The Advanced Theory of Statistics, Vol 1, Distribution Theory, (4th edition),* C. Griffin and Co., Ltd., London.

Kneale, G.W. and Stewart, A.M. (1987). Childhood cancers in the United Kingdom and their relation to background radiation. In Jones R.R., Southwood R. (eds.). *Radiation and Health: The Biological Effect of Low Level Exposure to Ionising Radiation,* Chichester: John Wiley, 203-220.

Krige, D.G. (1978). Lognormal de Wijsian Geostatistics For Ore Evaluation. *South African Inst. Min. Metall. Monograph Series: Geostatistics* **Vol. 1.**

Krumbein, W.C. (1956). Regional and local components in facies maps. *Bulletin American Association Petroleum Geologists,* **40**, 2162-2194.

Krumbein, W.C. (1959). Trend surface analysis of contour-type maps with irregular control point spacing. *Journal of Geophysical Research,* **64**, 823-834.

Leukaemia and Lymphoma. An area atlas of distribution within areas of England and Wales 1984-1988. (1990). Compiled by the Leukaemia Research Fund Centre for Clinical Epidemiology at the University of Leeds.

Link, R.F. and Koch, (Jr.)G.S. (1975). Some consequences of applying lognormal theory to pseudo-lognormal distributions. *Mathematical Geology,* **17(2)**, 117-128.

Loftsgaarden and Quesenbury, C.P. (1965). A non-parametric estimate of a multivariate probability density function. *Annals Math. Statist.*, 1049-1051.

Lorenz, M.C. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association,* **9**, new series, 209-219.

McCullagh, M.J. (1983). Transformation of contour strings to a regular grid based digital elevation model. *Euro-Carto,* 18pp.

McCullagh, M.J. and Ross, C.G. (1980). Delaunay triangulation of a random data set for isarithmic mapping. *Cartographic Journal,* **17(2)**, 93-99.

McGlashan, N.D. (1972). Geographical evidence on medical hypothesis. *Medical Geography: Techniques and Field Studies.* Methuen, London.

McLain, D.H. (1974). Drawing contours from arbitrary data points. *Computer Journal,* **17**, 318-324.

Mackay, J.R. (1951). Some problems and techniques in isopleth mappping. *Economic Geography,* **21**, 1-9.

Mackay, J.R. (1953). The alternative choice in isopleth interpolation. *Professional Geographer,* **5**, 2-4.

Maniya, G.M. (1961). *Soobshch.* Akad. Nauk Gruzin. SSR **27**, 385-390.

Matern, B. (1960). *Spatial Variation* Medd. Statens. Skogsforskningsinstitut.

Merriam, D.F. and Sneath, P.H.A. (1966). Quantitative Comparison of Contour Maps. *Journal of Geophysical Research,* **7(4)**, 1105-1115.

Miesch, A.T. and Conner, J.J. (1967). Stepwise regression and non-polynomial model in trend analysis. University of Kansas, State Geological Survey, *Computer Contribution,* **No.27**.

Miller, R.L. (1956). Trend surfaces : Their application to analysis and description of environment of sedimentation. *Journal of Geology,* **64**, 425-446.

Miller, V.C. (1953). A quantitative geomorphic study of drainage basin characteristics in the Clinch Mountain Area, Virginia and Tennessee. *Technical Report No. £., Department of Geology, Columbia University, New York.*

Mills, F.C. (1955). *Statistical Methods (3rd edition).* Holt Rinehart and Winston, New York.

Minnick, R.F. (1964). A method for the measurement of areal correspondance. *Papers of The Michigan Academy of Science, Arts and Letters,* **Vol XL1X**, 333-342.

Mirchink, M. F. and Bukhartsev, V. P. (1959). The possibility of statistical studies of structural relations. *Doklady of the Academy of Sciences of U.S.S.R.,* **126(5)**, 495-497.

Mitchell, J.F.B. (1989). The 'greenhouse' effect and climate changes. *Reviews of Geophysics,* **27**, 115-139.

Monmonier, M.S. (1973). Eigenvalues and principal components. A method for detecting natural breaks for choroplethic maps. *American Congress Survs. Mapp. Proc.* Fall Convention, 252-264.

Morgenstern, H. (1982). Uses of ecologic analysis in epidemiologic research. *American Journal Public Health,* **72**, 1336-1344.

Morrison, J.L. (1971). Method-produced error in isarithmic mapping. *American Congress on Surveying and Mapping,* Technical Monograph No. CA-5.

Mosteller, F. and Wallace, D.L. (1963). Inference in an authorship problem. *Journal American Statistical Association,* **58**, 275-309.

Nambi, K.S.V. and Soman, S.D. (1990). Further observations on environmental radiation and cancer in India. *Health Physics,* **59(3)**, 339-344.

Richardson, L.F. (1961). The problem of contiguity. An Appendix to '*Statistics of Deadly Quarrels*', Wright, Q. and Lenau, C.C. (editors).

Ripley, B.D. (1981). *Spatial Statistics.* J. Wiley and Sons, U.S.A.

Robinson, A.H. (1961). The cartographic representation of the statistical surface. *International Yearbook of Cartography,* 1, 53-63.

Robinson, G. (1972). Trials on trends through clusters of cirques. *Area,* **4**, 102-103.

Robinson, A.H. and Bryson, R.A. (1959). A method for describing quantitatively the correspondance of geographical distributions. *Annals Association American Geographer,* **47**, 379-391.

Robinson, A.H. and Sale, R.D. (1969). *Elements of cartography (3rd edition).* John Wiley and Sons, New York.

Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function. *Annals Mathematics Statistics,* **27**, 832-837.

Rothwell, M.A. (1971). A computer program for the construction of pole figures. *Journal Applied Cryst.,* **4**, 494.

Rudemo, M. (1982). Empirical Choice of Histograms and kernal density estimators. *Scand. Journal Statist.,* **9**, 65-78.

Sanderson, D.C.W., Allyson, J.D., Martin, E., Tyler, A.N., and Scott, E.M. (1990). An airborne gamma ray survey of three Ayrshire Districts. *Scottish Universities Research and Reactor Centre, East Kilbride.*

Sanderson, D.C.W., Martin, E., Scott, E.M., Baxter, M.S. and NiRiain, C. (1992). The use of radiometrics for epidemiological studies of leukaemia. A preliminary investigation in S.W. England. *Scottish Universities Research and Reactor Centre, East Kilbride.*

Schlesinger, M.E. and Zhao, Z. (1987). Seasonal climate changes induced by doubled carbon dioxide as simulated by the OSU atmospheric OCMI mixed layer model. *Report 70,* 73pp. Oregon State University, Climateic Institute, Corvallis.

Schmid, C.F. and MacCannell, E.H. (1955). Basic problems, techniques and theory of isopleth mapping. *Journal American Statistical Association,* **50**, 220-239.

Schultz, G.M. (1961). An experiment in selecting value scales for statistical distribution maps. *Surveying and Mapping,* **21**,224-230.

Schuster, E.F. and Gregory, C.G. (1981). On the nonconsistency of maximum likelihood non-parametric estimators. *Computer Science and Statistics: Proceedings of The 13th Symposium On the Interface.* New York: Springer-Verlag, 295-298.

Scott, D.W. (1985a). Frequency Polygons: Theory and Application. *Journal American Statistical Association,* **80(390)**, 348-354.

Scott, D.W. (1985b). Average shifted histograms: Effective non-parametric density estimators in several dimensions. *Annals of Statistics,* **13(3)**, 1024-1040.

Scott, D.W. and Factor, L.E. (1981). Monte Carlo study of three data-based non-parametric density estimators. *Journal American Statistical Association,* **76**, 9-15.

Scott, D.W. and Thompson, J.R. (1983). Probability density estimation in higher dimensions. In Gentle J.E. (ed.), *Computer Science and Statistics: Proceedings of the Fifteenth Symposium and The Interface,* Amsterdam: North Holland, 173-179.

Scriptor, M.W. (1970). Nested-means map classes for statistical maps. *Annals Association American Geographer,* **60**, 385-393.

Seber, G.A.F. (1977). *Linear Regression Analysis.* John Wiley and Sons.

Sellers, W.D. (1968). Climatology of monthly precipitation patterns in the Western United States, 1931-1966. *Monthly Weather Review,* **96**, 585-595.

Shephard, D.S. (1968a). A two-dimensional interpolation function for computer mapping of irregularily spaced data. *Harvard Theoretical Geography,* **Paper No. 15,** Laboratory for Computer Graphics, Harvard University Graduate School of Design, Cambridge, Mass.

Shephard, D.S. (1968b.). A two-dimensional interpolation function for irregularily spaced data (revised). *Proceedings of the 23rd National Conference of the Association for Computing Machinery,* Brandon/Systems Press, Inc., Princeton, N.J.

Siegel, S. (1956). *Non-Parametric Statistics For The Behavioral Sciences.* McGraw - Hill, New York.

Siegel, A.F. and Benson, R.H. (1982). A robust comparison of biological shapes. *Biometrics,* **38**, 341-350.

Silverman, B.W. (1978). Choosing the widow width when estimating a density. *Biometrika,* **65**, 1-11.

Silverman, B.W. (1986). *Density estimation for statistics and data analysis.* Chapman and Hall.

Skaggs, R.H. (1975). Drought in the United States 1931-1940. *Annals Association American Geographer,* 65, 391-402.

Sneath, P.H.A. (1967). Trend surface analysis of transformation grids. *Journal Zoology.,* London 151, 65-122.

Steffensen, J.F. (1927). *Interpolation.* Williams and Wilkins.

Stevens, S.S. (1946). On the theory of scales of measurement. *Science,* **103**, 677-680.

Stockman, G., Kapstein, S. and Bennett, S. (1982). Matching images to models for registration and object detection via clustering. *IEEE Trans Pattern Anal. Machine Intelligence.* **Vol PAMI-4 no.3**, 229-241.

Stoddart, D.R. (1965). The shape of atolls. *Marine Geology,* **3**, 369-383.

Stone, C.J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *Ann. Statist.,* **12**, 1285-1297.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions (with discussion). *J.R. Statistical Society B,* **36**, 111 - 147.

Student, I. (1907). On the error of counting with a haemacytometer. *Biometrika,* **5**, 351-360.

Switzer, P. (1973). Applications of random process models to the description of spatial distributions of qualitative geologic variables. *24th International Geological Congress, Montreal. Irving, E. (ed.).* 1-11.

Switzer, P. (1975). Estimation of the accuracy of qualitative maps. *Display and Analysis of Spatial Data,* J.C. Davies and M.J. McCullagh (eds.), Wiley, New York, 1-13.

Switzer, P., Mohn, C.M., and Heitman, R.E. (1964). Statistical analysis of ocean terrain and contour plotting. *Procedures, Project Trident Report No. 1440464,* Arthur D. Little, Inc., Cambridge, Mass.

Thompson W. D'Arcy (1917, 1961). *On growth and form.* Bonner J.T. (ed) Cambridge University Press, Cambridge.

Tipper, J.C. (1979). Surface Modelling Techniques. *Kansas Geological Survey Series on Spatial Analysis,* **No. 4.**

Tobler W.R. and Lau, J. (1970). Isopleth mapping using histoplanes. *Geographical Analysis,* **10**, 273-271.

Tsonis, A.A. and Elsner, J.B. (1989). Testing the global warming hypothesis. *Geophysical Research Letters.* **16**, 795 - 797.

Tukey, P.A. and Tukey, J.W. (1981). Geographical display of data sets in three or more dimensions. In Barnett, V. (ed.) *InterpretingMultivariate Data.* Chichester: Wiley, 189-275.

Ujeno, Y. (1983). Relation between cancer incidence or mortality and external natural background radiation in Japan. In *Biological Effects of Low Level Radiation,* Vienna, I.A.E.A. 253-262.

Unwin, D.J. (1981). *Introductory Spatial Statistics,* Methuen and Co., Ltd. London.

Walter, S.D. and Birnie, S.E. (1991). Mapping mortality and morbidity patterns: an international comparison. *Int. J. Epidemiology.*

Warntz, W. (1959). Toward a geography of price. Philadelphia.

Washington, W. M. and Meehl, G.A. (1984). A seasonal cycle experiment on the climate sensitivity due to a doubling of carbon dioxide with an atmospheric general circulation model coupled to a simple mixed layer ocean model. *J. Geophys. Res.,* **89**, 9475-9503.

Watson, G. (1971). Trend Surface Analysis. *Journal Mathematical Geology.* **3(3)**, 215-226.

Weatherald, R.T. and Manabe, S. (1986). An investigation of cloud cover change in response to thermal forcing. *Climatic Change,* 8, 5-24.

Webb 111, T., Bartlein, P.J. and Kutzbach, J.E. (1987). Climatic change in eastern North America during the past 18000 years: Comparisons of pollen data with model results. In Ruddiman W.F. and Wright H.E. Jr. (Eds.) *North America and adjacent oceans during the last deglaciation: Boulder, Colorado, Geological Society of America, The Geology of North America,* V K-3.

Werrity, A. (1969). On the form of drainage basins. *Papers in Geography, The Pennsylvania State University,* Pennsylvania, Papers in Geography.

White, R.R. (1972). Probability maps of leukaemia mortalities in England and Wales. In *Medical Geography : Techniques and Field Studies.* McGlashen, N.D. (ed.), Methuen, London, 173-185.

Whitten, E.H.T. (1957). Composite trends in granite: Modal variation and ghost stratigraphy in part of the Donegal Granite, Eire. *Journal of Geophysical Research,* **64**, 835-848.

Whitten, E.H.T. (1970). Orthogonal polynomial trends for irregularily spaced data. *Journal of Mathematical Geology,* **2**, 141-152.

Wilson, C.A. and Mitchell, J.F.B. (1987). A doubled carbon dioxide climate sensitivity experiment with a GCM including a simple ocean. *J. Geophy. Res,* **92**, 13315-13343.

Woodward, W.A. and Gray, H.L. (1992). 'Global Warming' and the problem of testing for trend in time series data. *Department of Statistical Science, Southern Methodist University,* 19 pages.

Young, I.T., Walker, J.E. and Bowie, J.E. (1974). An analysing technique for biological shape. *Medinfo. 74.* M.T.T 843.

APPENDIX

Year	Season	Temp. level	Area	Perime -ter	Orient- ation	Centroid displacement	
						x	y
1930	Spring	40^0C (a)	728.5	127.2	1.827	97.95 (232.0)	39.07 (51.39)
		(b)	19.23	19.55	0.525	115.9 (2.607)	115.9 (4.753)
		45^0C (a)	627.1	142.9	1.828	97.40 (184.4)	38.73 (37.89)
		(b)	38.83	24.50	0.6215	116.1 (5.260)	39.33 (7.663)
		(c)	8.604	10.77	6.229	107.9 (1.787)	35.59 (1.196)
		50^0C (a)	337.7	89.08	1.669	88.24 (95.42)	37.83 (29.16)
		(b)	28.59	33.83	2.867	118.5 (3.894)	44.37 (13.95)
		55^0C (a)	231.0	88.63	1.757	86.68 (64.71)	37.56 (16.38)
		(b)	8.062	11.84	1.040	89.38 (1.384)	49.36 (0.068)
		(c)	4.915	11.53	1.009	118.8 (2.045)	46.33 (0.432)
		60^0C (a)	30.87	24.16	3.065	96.22 (5.223)	38.11 (6.820)
		(b)	135.5	72.73	2.093	82.79 (13.91)	37.74 (30.98)

1980	Spring	40^0C (a)	728.5	127.2	1.826	97.94 (229.1)	39.06 (51.39)
		(b)	17.91	19.02	0.561	115.9 (2.527)	39.07 (4.666)
		45^0C (a)	613.3	159.1	1.829	97.13 (171.2)	38.80 (34.57)
		(b)	39.93	25.05	0.559	116.1 (5.203)	39.34 (8.487)
		50^0C (a)	318.4	86.00	1.998	87.87 (89.40)	37.75 (28.50)
		(b)	23.27	32.74	2.875	118.6 (3.437)	44.63 (13.26)
		(c)	24.29	12.985	2.985	110.8 (2.302)	40.72 (7.074)
		55^0C (a)	248.5	80.80	1.716	86.98 (73.51)	37.74 (19.08)
		(b)	6.656	6.979	1.014	89.35 (1.280)	49.36 (0.068)
		(c)	9.507	13.87	2.314	118.6 (2.851)	46.23 (1.172)
		(d)	4.847	8.224	2.792	110.9 (0.810)	40.00 (0.940)
		60^0C	29.41	23.07	3.206	96.33 (3.779)	37.98 (6.584)
		(b)	92.50	59.20	1.812	81.18 (23.11)	38.37 (23.11)
		(c)	14.48	14.26	2.796	89.60 (2.410)	32.56 (2.530)

1930	Summer	65^0C (a)	805.0	128.4	1.8388	98.38 (280.1)	39.40 (55.40)
		(b)	9.583	11.84	0.069	116.1 (1.489)	38.10 (1.730)
		70^0C (a)	576.5	189.4	1.483	87.03 (216.8)	38.61 (15.05)
		75^0C (a)	339.5	86.40	1.654	88.39 (94.01)	38.41 (29.60)
		(b)	17.28	30.53	0.2782	119.4 (4.040)	45.59 (12.87)
		(c)	6.156	16.25	0.144	111.5 (2.276)	40.32 (2.276)
		80^0C (a)	96.00	62.56	1.793	86.52 (63.98)	37.26 (7.490)
		(b)	13.43	23.58	2.685	96.64 (2.811)	38.96 (5.841)
		(c)	1.531	3.368	0.001	90.26 (0.314)	31.94 (0.298)
		(d)	0.2812	4.163	2.356	82.63 (0.013)	37.75 (0.023)
		85^0C (a)	6.844	9.748	2.469	75.62 (1.959)	40.89 (1.029)
		(b)	1.727	4.866	1.213	97.41 (0.435)	41.00 (0.192)
1980	Summer	65^0C (a)	782.0	128.3	1.833	98.33 (254.1)	39.38 (54.98)
		(b)	9.583	11.84	0.069	116.1 (1.489)	38.10 (1.730)

		70^0C	522.0	146.2	1.895	93.21 (163.2)	38.65 (30.71)
		(b)	28.37	41.28	0.042	105.3 (0.613)	38.97 (0.523)
		(c)	0.646	0.537	0.075	106.4 (0.078)	38.23 (0.078)

		75^0C (a)	339.5	86.40	1.654	88.39 (94.01)	38.41 (29.60)
		(b)	19.81	31.65	0.284	119.4 (3.910)	45.05 (12.07)
		(c)	6.156	16.25	0.1444	111.5 (2.276)	40.32 (3.861)
		(d)	2.969	5.240	0.001	97.91 (0.590)	30.64 (0.511)

		80^0C (a)	216.7	96.63	1.725	87.05 (60.31)	38.26 (15.80)
		(b)	1.793	5.079	6.242	111.6 (0.356)	40.14 (0.371)
		(c)	0.2320	0.273	0.012	93.16 (0.238)	40.10 (0.264)
		(d)	0.2812	4.164	2.356	82.63 (0.012)	37.75 (0.123)

		85^0C (a)	15.09	14.60	2.404	75.67 (2.918)	40.90 (2.280)
		(b)	3.875	11.78	0.001	86.76 (2.731)	39.55 (1.064)
		(c)	2.125	9.064	0.003	97.21 (0.049)	36.15 (0.536)
		(d)	0.2186	4.853	1.198	97.35 (0.312)	40.84 (0.296)
		(e)	1.250	3.014	0.0012	81.47 (2.731)	34.52 (1.075)

1930	Autumn	45^0C (a)	678.1	141.7	4.975	97.58 (192.3)	38.86 (41.19)
		(b)	30.21	21.90	0.5008	116.0 (4.134)	39.50 (6.215)
		(c)	3.133	6.522	0.9559	107.9 (0.7244)	35.39 (0.4139)
		50^0C (a)	387.0	100.2	1.633	89.10 (112.9)	38.03 (31.23)
		(b)	37.87	39.59	3.277	118.5 (3.091)	43.21 (21.93)
		(c)	46.36	34.52	3.070	110.9 (3.142)	39.49 (17.54)
		55^0C (a)	281.7	79.32	5.140	87.27 (79.77)	37.91 (23.22)
		(b)	29.03	12.99	2.914	118.9 (2.655)	44.82 (13.37)
		(c)	5.654	8.921	2.873	110.9 (0.9479)	39.96 (1.095)
		60^0C (a)	195.5	93.95	1.732	86.10 (55.25)	37.97 (14.25)
		(b)	45.32	24.98	3.105	95.99 (5.989)	37.91 (6.831)
		65^0C (a)	96.82	9.718	2.513	89.71 (0.580)	31.14 (0.536)
		(b)	14.39	13.83	2.174	34.60 (1.830)	34.60 (1.830)
		(c)	7.249	10.93	2.310	95.99 (1.976)	37.27 (1.063)
		(d)	3.312	6.075	0.813	97.23 (0.6398)	40.57 (0.328)

		(e)	23.76	18.97	1.778	86.37 (6.723)	39.46 (2.280)
		(f)	28.38	20.71	5.443	76.02 (4.985)	40.44 (4.111)
1980	Autumn	45^0C (a)	658.9	142.1	4.969	97.57 (190.0)	38.82 (41.09)
		(b)	33.48	22.64	0.4665	116.1 (4.605)	39.42 (6.844)
		(c)	6.386	9.308	6.273	108.1 (1.246)	35.28 (1.015)
		50^0C (a)	394.7	100.7	5.114	89.32 (115.4)	38.01 (31.96)
		(b)	41.70	41.88	3.347	118.4 (3.241)	42.91 (23.36)
		(c)	56.28	37.13	6.232	111.0 (4.196)	39.54 (19.04)
		55^0C (a)	272.6	85.08	4.848	87.26 (74.57)	37.79 (20.30)
		(b)	8.659	10.91	2.898	110.8 (1.399)	39.94 (1.542)
		(c)	27.91	11.62	2.905	118.9 (2.956)	45.14 (11.29)
		60^0C (a)	139.5	74.70	5.262	82.90 (30.95)	37.54 (14.32)
		(b)	46.26	25.53	3.051	96.47 (6.639)	38.30 (7.975)
		65^0C (a)	6.634	9.555	2.492	89.82 (1.222)	32.16 (1.073)
		(b)	15.55	14.42	2.203	81.23 (3.371)	34.62 (2.006)

		(c)	8.956	11.46	2.236	96.58 (2.111)	36.31 (1.771)
		(d)	6.548	9.306	0.9040	96.72 (1.379)	40.39(0. (0.760)
		(e)	17.77	15.68	1.846	86.02 (4.566)	39.24 (1.857)
		(f)	22.43	18.58	5.440	75.88 (4.174)	40.41 (3.317)
1930	Winter	40^0C (a)	658.4	134.5	4.958	98.13 (213.8)	38.25 (37.96)
		(b)	37.16	27.15	5.968	115.5 (4.866)	40.45 (8.513)
		(c)	5.616	8.710	0.2913	107.7 (1.153)	35.73 (0.870)
		45^0C (a)	379.3	96.55	1.666	89.19 (118.0)	37.61 (28.56)
		(b)	86.02	69.51	3.690	115.5 (13.70)	39.33 (21.51)
		(c)	5.216	8.764	0.7221	110.4 (1.215)	45.38 (0.8052)
		50^0C (a)	283.7	84.33	4.883	87.52 (87.12)	37.29 (23.37)
		(b)	20.43	32.62	3.156	119.1 (2.292)	45.57 (15.18)
		(c)	0.1778	1.819	6.107	111.5 (0.022)	34.07 (0.912)
		55^0C (a)	160.0	57.80	5.225	83.48 (41.96)	36.88 (17.84)
		(b)	3.156	8.067	3.478	120.0 (0.3220)	46.27 (1.040)

		(c)	7.222	2.259	3.335	120.2 (0.1364)	38.63 (1.035)
		(d)	50.02	26.58	6.236	96.24 (7.335)	37.93 (7.335)

		60^0C (a)	54.20	35.24	5.383	79.01 (13.16)	37.13 (11.30)
		(b)	44.91	35.70	2.626	87.63 (6.893)	36.70 (12.66)
		(c)	27.96	23.17	3.286	96.14 (3.642)	37.60 (5.783)
		(d)	3.156	3.038	2.975	120.5 (0.004)	38.70 (0.2783)
		(e)	0.4062	2.584	3.359	120.5 (0.0096)	46.38 (0.1920)

1980		40^0C (a)	643.1	134.5	4.958	97.79 (199.8)	38.54 (38.02)
		(b)	24.65	20.34	0.5066	115.9 (3.383)	39.55 (5.273)
		(c)	0.1904	1.700	0.070	108.0 (0.036)	35.68 (0.041)

		45^0C (a)	475.9	182.5	1.643	95.51 (184.4)	37.74 (25.01)

		50^0C (a)	265.9	84.7	4.932	87.68 (87.49)	36.83 (20.80)
		(b)	24.72	33.55	6.268	118.9 (2.755)	43.82 (15.74)
		(c)	10.26	12.18	2.939	110.9 (1.410)	39.54 (2.235)

		55^0C (a)	147.6	56.58	5.213	83.53 (42.22)	36.45 (16.59)

		(b)	9.623	14.97	2.526	118.9 (2.611)	46.15 (1.569)
		(c)	8.141	2.919	3.393	120.1 (0.1518)	38.64 (1.879)
		(d)	50.16	26.03	3.214	96.61 (8.097)	37.48 (7.802)

		60^0C (a)	18.61	15.81	1.728	86.68 (4.465)	38.95 (2.095)
		(b)	49.40	33.60	2.930	79.29 (12.55)	36.68 (10.64)
		(c)	32.19	22.84	3.215	96.58 (4.573)	37.18 (5.916)
		(d)	5.218	4.593	3.242	120.4 (0.031)	38.69 (0.529)
		(e)	4.469	6.824	2.832	120.1 (0.1937)	46.19 (0.8290)
		(f)	20.38	16.79	5.988	89.52 (3.464)	32.64 (3.317)