AN INTRODUCTION TO

TREND SURFACE ANALYSIS

D.Unwin

CAT MOG

5

CATMOG

(Concepts and Techniques in Modern Geography)

CATMOG has been created to fill a teaching need in the field of quantitative methods in undergraduate geography courses. These texts are admirable guides for the teachers, yet cheap enough for student purchase as the basis of class-work. Each book is written by an author currently working with the technique or concept he describes.

1. An introduction to Markov chain analysis - L. Collins
2. Distance decay in spatial interactions - P.J. Taylor
3. Understanding canonical correlation analysis - D. Clark
4. Some theoretical and applied aspects of spatial interaction shopping models - S. Openshaw
5. An introduction to trend surface analysis - D. Unwin
6. Classification in geography - R.J. Johnston
7. An introduction to factor analytical techniques - J.B. Goddard & A. Kirby
8. Principal components analysis - S. Daultrey
9. Causal inferences from dichotomous variables - N. Davidson
10. Introduction to the use of logit models in geography - N. Wrigley
11. Linear programming: elementary geographical applications of the transportation problem - A. Hay
12. An introduction to quadrat analysis - R.W. Thomas
13. An introduction to time-geography - N.J. Thrift
14. An introduction to graph theoretical methods in geography - K.J. Tinkler
15. Linear regression in geography - R. Ferguson
16. Probability surface mapping. An introduction with examples and Fortran programs - N. Wrigley
17. Sampling methods for geographical research - C. Dixon & B. Leach
18. Questionnaires and interviews in geographical research - C. Dixon & B. Leach

Other titles in preparation

AN INTRODUCTION TO TREND SURFACE ANALYSIS

by

David J. Unwin
(University of Leicester)

CONTENTS

I INTRODUCTION

(i) Pre-requisites and Purpose

    This monograph is concerned with one of a number of methods that the geo-
grapher can use in the analysis of change over space. The method is called
trend surface analysis because the basic model used attempts to decompose each
observation on a spatially distributed variable into a component associated
with any regional trends present in the data and a component associated with
purely local effects. This separation into two components is accomplished by
fitting a best-fit surface of a previously specified type using standard re-
gression techniques. The values predicted by this *trend-surface* are assigned
to the regional effects wherea- the local departures of the observed data from
it, or *residuals,* are assigned to the local effects. It is assumed that the
reader will have some familiarity with normal regression analysis, particul-
arly the idea of fitting a mathematical function to data. Because trend sur-
face analysis highlights numerous problems in regression analysis associated
with the suitability of data, the validity of the assumptions, and of statis-
tical significance, in Section II we shall go a little further in regression
theory than is usual in geographic texts. For a full understanding of this
section some knowledge of partial differential equations and matrix algebra
will prove useful but is by no means essential.

(ii) Spatial Series

    In order to plot values on a map the geographer needs three pieces of in-
formation, the x,y spatial co-ordinates of each point together with the heights
above some datum, the z co-ordinate.   The z values might relate to any var-
iable that has a spatial distribution - in physical geography, temperature,
rainfall, soil grain size or pH, in human geography, population potential,
land rent and so on - but the whole operation defines a *spatial series* in which
the z observations are ordered with respect to the two spatial co-ordinates,
x and y. The map would be completed by drawing lines of equal z value (con-
tours or isolines) through the points. The resulting contour-type map defines
a complex surface which in most cases will reveal a *spatial structure* or form.
On the map, spatial form can be very simple and regular, as in the example
of land values declining away from a city centre, or very complex, as in the
case of rainfall at the warm front of a depression, but if it did not exist
and the z values were spatially random it would be difficult to speak of any
'geography' of that variable. The detection and analysis of spatial form is
clearly at the very heart of geography as a science (King, 1969) but spatial
series are also studied by geologists, ecologists, and econometricians.

The geographer might have three motives in mapping and analysing a spatial
series. At the simplest level he might be interested in the economical *des-
cription* of the series, perhaps in order to predict the z value at some un-
known location. Secondly, he might wish to *remove the spatial form* as a pre-
liminary to further aspatial analysis. Thirdly, and most sophisticated of all,
he might wish to use the observed spatial form as a 'model' constructed in
order to *test some hypothesis* about spatial variation.

    Remembering that a 'model' is usually defined as a 'simplification of
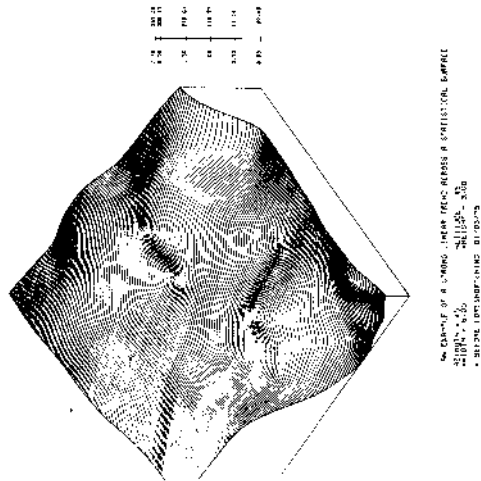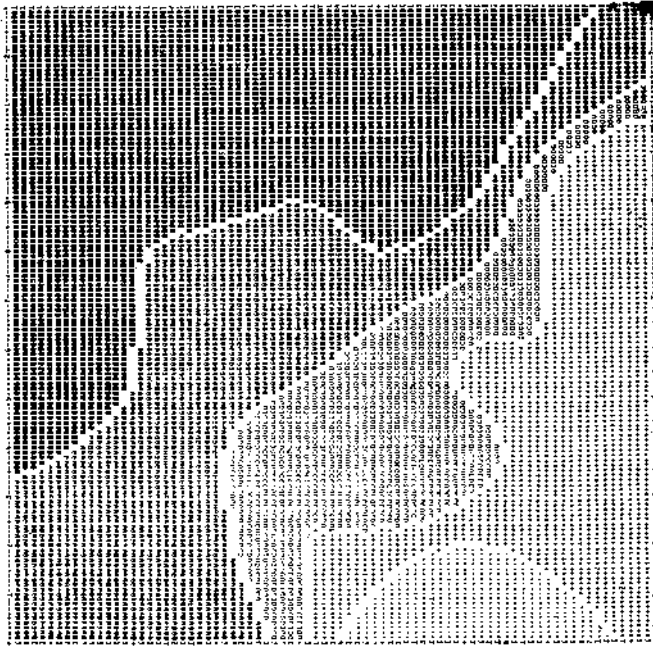
reality and that most contour-type maps show very complex patterns, how can we produce models of spatial form? Perhaps the simplest approach would be to assume that the series exhibits a regional trend, so that each observed value can be regarded as the sum of two components, the systematic trend and the departures from this. Figure 1 shows an example of a spatial series as a contour-type map and as a block diagram. There is a very strong trend in surface height which rises systematically from southwest to northeast.

Trend surface analysis makes this separation into a regional and a local component. Its use appears to have been first suggested by the statistician Student, but the first applications were in geology (Krumbein, 1959; Grant, 1961). In geography its use dates from the mid 1960's (Tobler, 1964; Chorley and Haggett, 1965) and it has been used to examine spatial trends in phenomena as diverse as erosion surfaces, glacial cirques, urban service functions, agriculture, soil pH, mental maps, rainfall totals and atmospheric pollution. Although this monograph is concerned with geographic trends, similar techniques of response-surface analysis have been used for many years in industry (Davies, 1954; Hill and Hunter, 1968).

II MATHEMATICAL DEVELOPMENT

(i) The method of least squares

In this section we shall show how we can use a knowledge of elementary mathematics to derive a simple trend surface model, illustrating this with a very simple numerical example before going on to demonstrate the properties of this model and possible extensions to it.

A trend surface analysis assumes that each mapped value can be decomposed into two components that arise from two scales of process:
A) A large scale process that generates a *trend component* operating over a large-6"n. According to Krumbein and Graybill (1965) this trend is 'associated with 'large scale' systematic changes that extend from one map edge to the other'. Similarly, Grant (1961) defines trend as '. . . that part of the data that varies smoothly. In other words, it is the function that behaves predictably'.

B) The combined result of two processes that operate over an area substantially smaller than the study area, random fluctuations and errors of measurement. This forms an assumed error, *local component,* or *residual* defined by Krumbein and Graybill as '. . . apparently non-systematic fluctuations that are super-imposed on the large scale patterns'. It is important to notice that at the scale of observation, these residuals appear to be spatially random; they may prove to be systematically related to an aspatial process but at this scale they do not vary systematically over the mapped area. Mathematically,

observed value of   trend component + residual at
    surface   at that point   that point

If component (A), the trend, varies smoothly over space its value (height) at any particular point can be expressed in terms of the spatial co-ordinates of that point, so that the basic equation of any trend analysis becomes

Fig. 1   An example of a spatial series showing a systematic trend.

$$Z_{obs_i} = f(x_i, y_i) + u_i \qquad \dots \ (1)$$

Z_{obs_i} = the observed value of the surface at the ith point.

x_i = the co-ordinate on the x-axis (northings) of the ith data point.

y_i = the co-ordinate on the y-axis (eastings) of the ith data point.

u_i = the residual at the ith data point.

f denotes 'some function', and thus the term $f(x_i, y_i)$ indicates the trend component. By *function* we simply mean that if we know the location of any point, i, as a pair of spatial co-ordinates, x_i and y_i, then the height of the trend component at this point can be found by simply plugging these x_i and y_i into a known equation (or function). It follows that we can calculate a trend component for any combination of x and y and that $f(x_i, y_i)$ denotes a complete *surface* of trend components called a *trend surface* .

There are a great many ways of representing this function - indeed almost any equation linking z to x and y would do - but to date most work has involved the use of very simple functions. We shall illustrate the development of a trend surface using the simplest possible of all functions to represent the trend component, leaving more complex representations until the reader has a firm grasp of essentials.

The simplest possible shape that we can imagine is an inclined plane as for example that given by a tilted, but unfolded geological stratum, illustrated in Figure 2. A moment's thought will indicate that in order to specify any such plane surface we need to know only three values that are constant over the entire surface area. These are (see Figure 2):-

A) ao, representing the height of the surface at the map origin where x=y=0
B) al representing the rate of change of surface height along the direction of the x axis. On Figure 2, al has a positive value, indicating that the surface rises in this direction.
C) a2 representing the rate of change of surface height along the direction of the y-axis. On Figure 2, a2 has a negative value, indicating that the surface falls in this direction.

To calculate the height of the trend surface at any point on it we simply need to evaluate

```
trend           value at         (rate of change      distance along)
component  =    map origin   +   (in x-direction   x  x-axis        )

                             +   (rate of change      distance along)
                                 (in y-direction   ι  y-axis        )
```

More succinctly this can be expressed:

$$z_{t_i} = a_0 + a_1 x_i + a_2 y_i$$

z_{t_i} = height of trend component at i^{th} point.
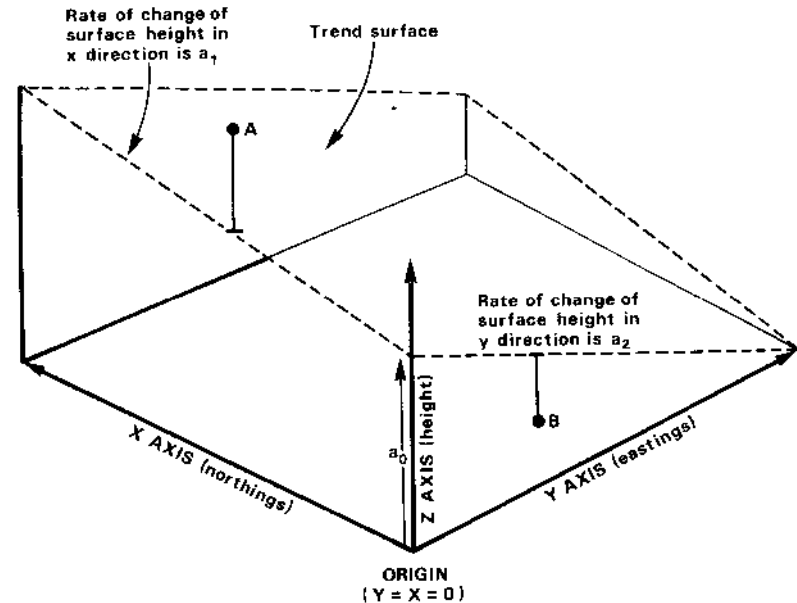
Fig. 2   A simple linear trend surface.

Hence, if we have specific numerical values for the constants a_0, a_1 and a_2 and know any location by its x_i, y_i co-ordinates, we can immediately calculate z_{t_i} at that point.

The equation defines a so-called first order, or *linear* trend surface, but in the real world it is very unlikely that any observed surface would exactly follow such an idealised trend. In practice, our observed data will lie above or below the trend, having a resid*ual* at each point. In Figure 2, the point A lying on the western map edge has an observed value higher than that predicted by the trend surface. The residual at this point is the difference between the observed surface height at this point (z_{obs_i}) and the predicted trend component (z_{t_i}):

residual at point	=	observed surface height at point	−	trend component at point
or	u_i	=	z_{obs_i}	− z_{t_i}
or	u_i	=	z_{obs_i}	− $f(x_i, y_i)$

In our example the residual is positive (above zero) indicating that the trend surface lies below the observed surface at this point, whereas on the southern edge of the map area, lies below the predicted trend surface and therefore has a negative residual.

From the above, it can be seen that for our simple inclined plane, equation (1), can now be restated more fully as follows:

$$z_{obs_i} = f(x_i, y_i) + u_i$$

i.e. $z_{obs_i} = (a_0 + a_1 x_i + a_2 y_i) + u_i$ .... (2)

Each combination of a_0, a_1 and a_2 would generate a different inclined plane. Some of these surfaces would be "good in that the observed points would lie close to them to give low residual values, whereas other surfaces would be 'poor' and have the observed values distant from them. It would be useful if we were able to find a method of determining the *very best possible combination* of a_0, a_1 and a_2 and it is usual (but at a higher level of sophistication than we will enter into here by no means essential) to chose those constants according to the *least squares criterion of goodness of fit*. This States that they are to be determined so that in combination they define a surface from which the sum of all the residual values squared is as low as it possibly can be for that surface shape. It is probable that the reader will have already met this minimum sum of residuals squared when fitting a *line* to a plot of data in two dimensions; all that we are doing is to extend it to the fitting of *surfaces* in *three*. It will be shown later that these *least-squares* values for the surface constants have, subject to certain assumptions, properties that make them intuitively desirable.

The least squares criterion can now be stated mathematically as 'find the combination of a_0, a_1 and a_2 that in relationship to the observed values minimises S, the sum of squares of the residuals'.

Defining S in symbols we have

$$S = \sum_{i=1}^{N} u_i^2$$ .... (3)

(the notation $\sum_{i=1}^{N}$ simply means 'sum the N values')

Remembering equation (2), each residual u_i becomes

$$u_i = z_{obs_i} - (a_0 + a_1 x_i + a_2 y_i)$$

and we can re-write equation (3) more fully as

$$S = \sum_{i=1}^{N} u_i^2 = \sum_{i=1}^{N} \{z_{obs_i} - (a_0 + a_1 x_i + a_2 y_i)\}^2$$ .... (4)

To understand how we can solve this equation requires some knowledge of equations that in calculus are called partial differential equations. The reader without such knowledge will have to take on trust the result that S is minimised only when

$$\frac{\delta S}{\delta a_0} = \frac{\delta S}{\delta a_1} = \frac{\delta S}{\delta a_2} = 0$$

8

Each term in this represents a rate of change of S measured relative to each of the required constants and S will be minimised if, and only if, all are zero simultaneously. Even if you do not have the necessary mathematical background, it should be possible to see that there are three rates of change of S for each of the three constants that need to be considered simultaneously. The required rates of change are found by *differentiating* equation (4) with respect to each coefficient in turn. Setting each resultant expression equal to zero we obtain the three equations:

$$\delta S/\delta a_0 = \sum 2 (z_{obs_i} - (a_0 + a_1 x_i + a_2 y_i)) \cdot (-1) = 0$$

$$\delta S/\delta a_1 = \sum 2 (z_{obs_i} - (a_0 + a_1 x_i + a_2 y_i)) \cdot (-x_i) = 0$$

$$\delta S/\delta a_2 = \sum 2 (z_{obs_i} - (a_0 + a_1 x_i + a_2 y_i)) \cdot (-y_i) = 0$$

The $\sum$ is understood to be for all i up to N, the number of points as before. As an exercise. the reader should divide each equation by 2, multiply out the (-1), $(-x_i)$ and $(-y_i)$, place before each term that has a subscript i its own sum sign and then gather all the z's on the right hand side. Hopefully, this will show that these three equations simplify to a very simple set of three simultaneous equations called the *normal equations*:-

$$a_0 N + a_1 \sum x_i + a_2 \sum y_i = \sum z_{obs_i}$$

$$a_0 \sum x_i + a_1 \sum x_i^2 + a_2 \sum x_i \cdot y_i = \sum z_{obs_i} \cdot x_i$$

$$a_0 \sum y_i + a_1 \sum x_i \cdot y_i + a_2 \sum y_i^2 = \sum z_{obs_i} \cdot y_i$$ .... (5)

N = the total number of points for which a z_{obs_i} exists.

There are now three equations, and since we can find all the terms in x, y, x.y, z_{obs}.x and z_{obs}.y from our observed data locations and heights there are only three unknowns remaining, a_0 a_1 and a_2. To determine the best-fit values for a_0, a_1 and a_2 all that need be done is to solve these three normal equations.

As the next section shows, this solution can be obtained using pencil and paper or a desk calculator, but there is a lot of arithmetic to be gone throught even for a simple first order surface with only a few data points. To save effort and to ensure that the obtained solutions are reasonably accurate, it is both normal and sensible to use a digital computer.

Suitable computer programs have been described and published by a number of authors and a selection are listed in the bibliography.

(ii) A simple example

Consider the data distribution shown in Figure 3; there are 10 data points each with observed values of z as indicated and we intend to fit a simple first order surface using the least squares method outlined in (i). To solve the normal equations (5), we need to use these data to find the $N, \sum x_i, \sum y_i, \sum x_i \cdot y_i, \sum y_i^2, \sum x_i^2, \sum z_{obs_i}, \sum z_{obs_i} \cdot x_i, \sum z_{obs_i} \cdot y_i$.

9
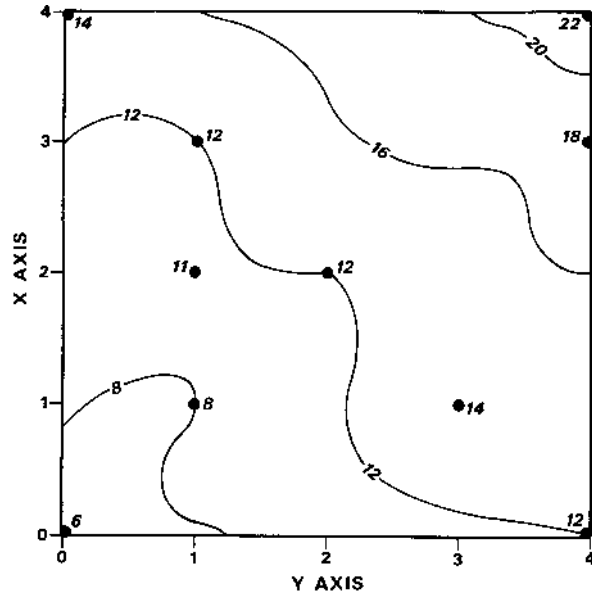
Fig. 3   Data for a simple trend surface example

It is best to proceed to find these quantities systematically, as is shown in Figure 4. Substituting the resulting sums into the normal equations we get

$$a_0.10 + a_1.20 + a_2.20 = 129 \qquad \dots (6a)$$
$$a_0.20 + a_1.60 + a_2.41 = 302 \qquad (6b)$$
$$a_0.20 + a_1.41 + a_2.64 = 305 \qquad \dots (6c)$$

Almost everyone will recognise that the combination of a_0, a_1 and a_2 which satisfies these three simultaneous equations can be found by systematically eliminating each constant.

In (6a) rearrange to give

$$a_0 = \frac{129 - 20a_1 - 20a_2}{10} = (12.9 - 2a_1 - 2a_2)$$

Now substitute this result for the a_0's in both (6b) and (6c):-

$$20 (12.9 - 2a_1 - 2a_2) + 60a_1 + 41a_2 = 302$$
$$20 (12.9 - 2a_1 - 2a_2) + 41a_1 + 64a_2 = 305$$

10

Fig. 4 Calculation of the quantities needed to solve the normal equations.

Point No.	Co-ordinates							
	x	y	z	x^2	y^2	xy	zx	zy
1	0	0	6	0	0	0	0	0
2	1	1	8	1	1	1	8	8
3	2	1	11	4	1	2	22	11
4	3	1	12	9	1	3	36	12
5	4	0	14	16	0	0	56	0
6	2	2	12	4	4	4	24	24
7	1	3	14	1	9	3	14	42
8	0	4	12	0	16	0	0	48
9	3	4	18	9	16	12	54	72
10	4	4	22	16	16	16	88	88
SUMS	20	20	129	60	64	41	302	305

$$\sum Z^2 = 1853$$

There are 10 points: N =10

These simplify to

$$20a_1 + 1a_2 = 44 \qquad \dots (7a)$$
$$a_1 + 24a_2 = 47 \qquad (7b)$$

Continue by finding a1 in (7a) as:

$$a_1 = \frac{44 - a_2}{20} = (2.2 - 0.05a_2)$$

and substitute this into 7(b) to find a2 as:

$$24a_2 = 47 - (2.2 - 0.05a_2)$$
$$a_2 = 44.8/23.95$$
$$a_2 = 1.8705637$$

Just as this value was found by forward (i.e. down from equation to equation) elimination of unknowns, so we can now back substitute to find a1 and ao.

In 7(a)

$$20a_1 + 1.8705637 = 44$$
$$a_1 = 42.12943/20$$
$$a_1 = 2.1064718$$

11

Finally, both results can be substituted back into equation (6a) to find a_0

$$a_0 = 4.9459289$$

We conclude that our first-order, or linear, trend surface fitted to these data by the least squares method has the equation

$$z_{t_i} = 4.9459289 + 2.1064718x_i + 1.8705637y_i$$

(if doing these calculations by hand it is a good idea to stop at this point and check that the values obtained really do satisfy each of the normal equations).

It is now possible to produce an isoline map of the trend surface that this equation represents by substituting suitable x and y values into it. The eeSulting trend surface map is shown as Figure 5 on which isolines have been drawn for constant z_t values of 8, 12, 16 and 20 units. The best fit trend surface rises steeply from the southwest of the mapped area towards the northeast, mirroring and simplifying the trend evident in the original data.

Residuals above and below this surface can be calculated at each data point by substituting in turn the x, y co-ordinates of these points into the trend equation and subtracting the resulting z_t from the observed surface value z_{obs}. For example, data point 3 is located at x = 2, y = 1 and has a z_{obs} of 11 units. The residual is therefore:

Residual at
point No. 3 $= 11 - (4.9459289 + (2 \times 2.1064710) + (1 \times 1.8705637))$

$= 11 - 11.02944$

$- 0.02944$
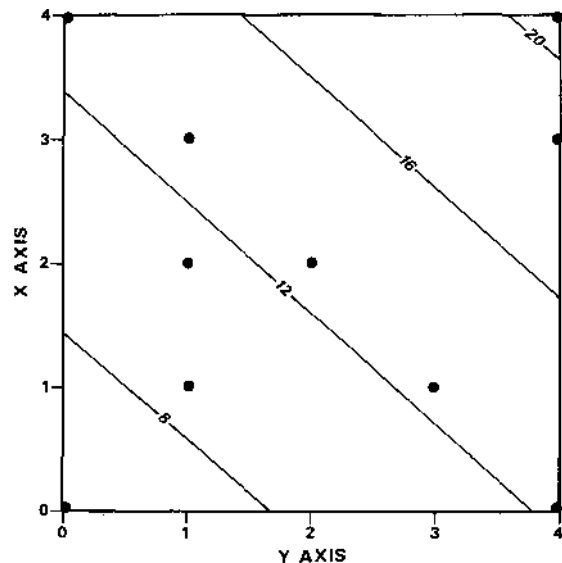
First order (linear) trend surface defined by the equation
$$Z = 4.95 + 2.11X + 1.87Y$$

Fig. 5 Trend surface map for simple example.

12

Fig. 6 Calculation of predicted trend values (z_t) and residuals (u_i) for each data point.

Point No.	x	y	z	z_t (computed)	u_i (residual)
1	0	0	6	4.94593	+ 1.05407
2	1	1	8	8.92296	- 0.92296
3	2	1	11	11.02943	- 0.02944
4	3	1	12	13.13591	- 1.13591
5	4	0	14	13.37182	+ 0.62818
6	2	2	12	12.90000	- 0.90000
7	1	3	14	12.66409	+ 1.33591
8	0	4	12	12.42818	- 0.42818
9	3	4	18	18.74760	- 0.74760
10	4	4	22	20.85407	+ 1.14593

Sums $\sum z_t = 128.99999$ $\sum u_i = 0.00000$
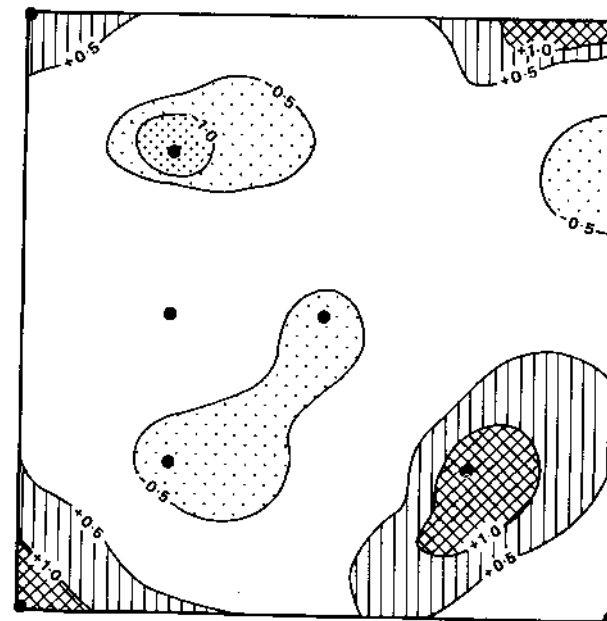
$\sum z_t^2 = 1844.6983$

Fig. 7   Residuals over the linear trend surface example.

13

The complete calculations for all 10 points are shown in Figure 6 and in map form as Figure 7. Notice that they show no obvious systematic variation over the map area, indicating that this simple linear surface is probably an adequate 'model' for these data, but how closely does it fit the original observations?

A *goodness of fit* statistic suitable for use in trend surface analysis is the *percentage reduction in sum of squares achieved,* or *%RSS.* This is simply the ratio, expressed as a percentage, of the corrected sum of squares of the computed trend calues (z_t) to the corrected sum of squares of the observations (z_{obs}), or

$$\%RSS = \frac{\text{corrected sum of squares of computed trend}}{\text{corrected sum of squares of observed values}} \times 100 \; (\%)$$

The corrected sum of squares of the computed values is

$$= \sum_{i=1}^{N} z_{t_i}^2 - \left(\sum_{i=1}^{N} z_{t_i}\right)^2 / N$$

and that for the observed values is

$$= \sum_{i=1}^{N} z_{obs_i}^2 - \left(\sum_{i=1}^{N} z_{obs_i}\right)^2 / N$$

The reader who is familiar with the 'short cut' method of finding a standard deviation will immediately recognise these formulae. If the quantities are equal, indicating a perfectly fitting trend surface that passes through every data point without leaving any residual, then the %RSS will be 100 whereas Tesser fits will give values between 0 and this maximum. The %RSS statistic used in trend surface analysis is analogous to the coefficient of multiple correlation. This suggests that we can transform %RSS values into equivalent correlation coefficients and that the relative strengths of trends might be described using the adjectives listed in Figure 8.

Fig. 8   Suggested terms for describing the strength of a trend based upon its %RSS.

%RSS	Equivalent coefficient of correlation ($\sqrt{(\%RSS/100)}$)	Adjectival description
< 4%	< 0.2	slight; almost negligible trend
4 - 16	0.2 - 0.4	low; definite but small trend
16 - 49	0.4 - 0.7	moderate- substantial trend
50 - 80	0.7 - 0.9	high; marked trend
81- 100	0.9 - 1.0	very marked trend

N.B. these terms should always be qualified by the type of trend fitted: 'almost negligible linear trend' etc.

For the data in our example, figures 4 and 6 contain all the information needed to calculate a %RSS as follows:-

Corrected sum of squares of computed trend
$$1844.6983 - \frac{129^2}{10} = 180.5983$$

Corrected sum of squares of observed values
$$1853 - \frac{129^2}{10} = 188.9$$

$$\%RSS = \frac{180.5983}{188.9} \times 100 = 95.60524 \; (\%)$$

The surface fitted would thus be described as having a very marked linear trend. Although this example has been worked through by hand, the data were deliberately chosen for their simplicity so that the reader might gain some insight into the mechanics of the method; in any practical application using real data a computer would invariably be used to perform the arithmetic and perhaps even to draw the maps.

(iii) A matrix formulation.

It is cumbersome to have to write out the normal equations (5,6) in full each time we wish to refer to them; fortunately we can compress them into a single equation using matrix notation. The reader who is unfamiliar with this can gain an adequate background from basic texts such as the summaries presented in Krumbein and Graybill (1965), King (1969) and Davis (1973).

If we look back to the full expression for the normal equations (5), we can define the following matrix and vectors:

1) $\underline{X} = \begin{bmatrix} N & \sum x_i & \sum y_i \\ \sum x_i & \sum x_i^2 & \sum x_i \cdot y_i \\ \sum y_i & \sum x_i \cdot y_i & \sum y_i^2 \end{bmatrix}$

This matrix is symmetric about its principal diagonal and for reasons that can readily be appreciated by examining its elements it is often referred to as the *matrix of sums of squares and cross-products.*

2) $\underline{b} = \begin{bmatrix} \sum z_{obs_i} \\ \sum z_{obs_i} \cdot x_i \\ \sum z_{obs_i} \cdot y_i \end{bmatrix}$

This column vector, b, is formed from the summed products of the observed data z-values and the locitional co-ordinates, (x,y).

3)
$$\underline{a} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$$

The column vector a contains the unknown coefficients of the trend equation.

Remembering the 'row times column rule of matrix multiplication, equation (5) can be written

$$\begin{bmatrix} N & \sum x_i & \sum y_i \\ \sum x_i & \sum x_i^2 & \sum x_i y_i \\ \sum y_i & \sum x_i y_i & \sum y_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum z_{obs_i} \\ \sum z_{obs_i} \cdot x_i \\ \sum z_{obs_i} \cdot y_i \end{bmatrix}$$

i.e. $\underline{X} \quad \underline{a} = \underline{b}$ .... (8)

The solution of this 'system' of equations is a well known mathematical problem, and the formulation in this way suggests a method by which many computer trend-surface programmes go about finding the required solution. As the data are read in the matrix X and the vector b are formed. Next, equation (8) is solved for the unknown vector a. A number of possible solution methods exist. Early computer programmes, su6F as that by Whitten (1963), often found the inverse of $\underline{X}$ denoted $\underline{X}^{-1}$ and premultiplied both sides of (8) by this:-

$$\underline{X}^{-1}.\underline{X}.a = \underline{X}^{-1}.\underline{b}$$

i.e. $\underline{I} . \underline{a} = \underline{X}^{-1}.\underline{b}$

$\underline{I}$ denotes the identity matrix $= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Various computer algorithms are available to find inverse matrices and usually this solution method will give reasonably accurate results but it should be noted that in general this is a notoriously difficult problem in automatic computation (Unwin, 1975). Most modern computer programs prefer to use the so-called 'direct' method outlined in section (ii) that does not involve calculation of the inverse matrix.

(iv) Extension to other surface shapes.

An inclined plane of the type we have fitted has the merit of simplicity, but when the original series of values is such that its variability is not adequately accounted for by the linear surface, or when geographic theory leads the investigator to expect a different surface'shape, then other, more complex, surfaces may be fitted. Exactly the same principles and techniques apply, but the calculations rapidly become lengthy and difficult so that a computer is

almost invariably used.

Suppose, for example, that we wish to describe a dome or a trough-like trend running over the study area. Such a sa--rface shape can be specified by *a quadratic polynomial,* with the trend equation:

$$z_{t_i} = b_0 + b_1 x_i + b_2 y_i + b_3 x_i^2 + b_4 x_i y_i + b_5 y_i^2 \qquad \text{.... (9)}$$

There are now 6 constants to be evaluated, b_0 - b_5, and in matrix form the normal equations are:

$$\begin{bmatrix} N & \sum x_i & \sum y_i & \sum x_i^2 & \sum x_i y_i & \sum y_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i y_i & \sum x_i^3 & \sum x_i^2 y_i & \sum x_i y_i^2 \\ \sum y_i & \sum x_i y_i & \sum y_i^2 & \sum x_i^2 y_i & \sum x_i y_i^2 & \sum y_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^2 y_i & \sum x_i^4 & \sum x_i^3 y_i & \sum x_i^2 y_i^2 \\ \sum x_i y_i & \sum x_i^2 y_i & \sum x_i y_i^2 & \sum x_i^3 y_i & \sum (x_i y_i)^2 & \sum x_i y_i^3 \\ \sum y_i^2 & \sum x_i y_i^2 & \sum y_i^3 & \sum x_i^2 y_i^2 & \sum x_i y_i^3 & \sum y_i^4 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix} = \begin{bmatrix} \sum z_{obs_i} \\ \sum z_{obs_i} \cdot x_i \\ \sum z_{obs_i} \cdot y_i \\ \sum z_{obs_i} \cdot x_i^2 \\ \sum z_{obs_i} \cdot x_i y_i \\ \sum z_{obs_i} \cdot y_i^2 \end{bmatrix}$$

Notice that the first three rows and columns are similar to those for the linear surface but that the associated constants b_0, b_1 and b_2 will not be the same. The addition of the extra 'quadratic' terms, x_i^2, $x_i y_i$ and y_i^2 greatly increasesthe number of summations that we need to form trom the original data and vastly complicates the solving of the normal equations but when values for b_0 to b_5 have been found they can be substituted in the trend equation (9) to calculate preaictea surrace neignts, resiauals ana tne k F(3. These added terms give the surface more flexibility, so that it is inevitable that a quadratic surface fitted to some data will always give a better fit, as evidenced by a % RSS nearer to 100, than a linear surface fitted to the same data.

A further extension, now involving 10 constants to be found, is the *cubic polynomial* whae—equation is

$$z_{t_i} = c_0 + c_1 x_i + c_2 y_i + c_3 x_i^2 + c_4 x_i y_i + c_5 y_i^2 + c_6 x_i^3$$
$$+ c_7 x_i^2 y_i + c_8 x_i y_i^2 + c_9 y_i^3 \qquad \text{....(10)}$$

It is possible to go on adding further terms to the trend equation to produce quartic, quintic and so on surfaces. In the literature surfaces up to the octic have been described (Bassett and Chorley, 1971), but there are two difficulties associated with their practical use. Initially, it is difficult to imagine any *a priori* geographical theory that might predict surfaces of order higher than the quadratic. Alternatively, and having fitted such surfaces, it is equally difficult to come up with any *a posteriori* theory to account for them so that they tend to be used solely as descriptive devices. Secondly, there are a great many terms in the trend equations and hence even more summations to be formed when assembling the X matrix of sums of squares

and cross-products so that such surfaces tend to be extremely demanding in their data requirements. Unless there exist a very large number of very evenly spaced control points and unless very great care is taken in the computer programming their fitting should not be attempted.

Power series polynomials of the type we have outlined are best suited to modelling smoothly varying surfaces without frequent regular changes in direction and slope; where it is suspected that the spatially-distributed variable behaves in an oscillatory manner, as for example the height of a tightly fol-ded geological stratum, a two-dimensional *Fourier* model may be used Following similar notation to that used for polynomial models, this can be written:

$$z_{t_i} = b_{00} + \sum_{i=1}^{m} \sum_{j=1}^{n} f(b_{ij} P_i Q_i) \qquad \ldots(11)$$

The function 'f' is now expanded as

$$f(b_{ij} P_i Q_i) = cc_{ij} \cos P_i \cos Q_i + cs_{ij} \cos P_i \sin Q_i$$
$$+ sc_{ij} \sin P_i \cos Q_i + ss_{ij} \sin P_i \sin Q_i \quad \ldots(12)$$

The locational co-ordinates in this expression are P_i and Q_i, but in this case they are expressed as fractions of two fundamental wavelengths, m in the direction of P and n in the direction of Q_i according to the equations

$$P_i = 2\pi x_i / m \quad , \quad Q_i = 2\pi y_i / n \qquad \ldots(13)$$

If the data are not spaced on a regular grid, these m and n values must be specified before the various constants b_{00}, cc_{ij}, cs_{ij}, sc_{ij} and ss_{jj} are estimated using the least squares approach. These equations may look forbidding but the ideas they express are not, and in principle they are identical to our earlier polynomials. The (P_i, Q_i) are still measures of spatial location and the constants of the polynomial have been replaced by constants which instead of referring to terms in a power series now refer to terms in the geometric sines and cosines of these location measures. A computer program has been presented by James (1966).

A comparison of the Fourier and polynomial models when applied to the same data has been published by Krumbein (1966). From a computational viewpoint the Fourier model can be easily applied to data collected over a regular grid, and even if the data are irregularly spaced the coefficients are generally easy to calculate. If the spatial series being examined has complex trends, Krumbein concludes that Fourier models produce surfaces that appear to be more in accord with the real world than do polynomials. There is, however, no low order Fourier equivalent to the linear and quadratic polynomials which interposes severe limitations on the analysis of simple trends.

This model has seldom been applied to socio-economic geographic data, finding most applications in geophysics. In physical geography it has been used to analyse topographic surfaces suspected of showing significant periodictties (Harbaugh and Preston, 1968). Elsewhere in geography its use seems to have been neglected in favour of other, more powerful modelling techniques (Unwin and Hepple, 1975, p.221).

Finally, other possible trend surface models based upon other series have been suggested by Miesch and Connor (1968) but have yet to find practical application.

18

The choice of surface type to be fitted is in many ways the *Achilles heel* of trend surface analysis and we will return to it later in this monograph.

(v) Properties of least-squares estimates and assumptions made.

We have not so far considered one obvious difficulty in our analysis - why did we chose our best fit values for the constants in the trend equations according to the minimum sum of squares of residuals criterion? Surely, other criteria, such as the minimum sum of the absolute values or the least mean square error could be considered just as good? The answer to these questions lies in the desirable properties that least squares solutions possess provided that a limited number of assumptions can be justified. A very readable and straightforward proof of these properties, together with an introduction to more recent work in the field is given in Koerts and Abrahamse (1971).

In our example, but not necessarily in all geographic applications (Unwin 1973), the ten data points at which z_{obs} were measured are obviously a sample from all possible points at which the original surface could have been measured. This latter forms the *population* for our analysis, and, as is so often the case, we are actually using our sample of ten to make inferences about trends in this total population. It follows that the numerical values for the constants (a0, 41 and a2 in the example) are actually *estimates* of the 'true but -unknown population values. It is not feasible, after all, to sample the original spatial series at every single location. These true population values are called *population parameters*. Manifestly, other samples drawn from tlie-seme population might give slightly different numbers for the estimates of the parameters so that we *must retain a clear distinction between a parameter and an estimate of it.*

Notationally, an estimated value is often signified by writing a small 'hat' above it, as for example $\hat{a}_0$, $\hat{a}_1$ and $\hat{a}_2$.

Had we repeatedly sampled the original spatial series at points other than our original ten and performed a trend surface analysis, then each exercise would have yielded a different trio of $\hat{a}_0$, $\hat{a}_1$ and $\hat{a}_2$, so that it is possible to think of a *distribution* of estimates of a parameter, as in Figure 9. In this a_x represents the true parameter value whilst the curves A, B and C are each for a number of estimates of that parameter, $\hat{a}_x$.

If we treat the problem of selecting the 'best' estimator as a problem in gambling in which a hypothetical gambler is required to make guesses at some true population value then a good gambler would require his guesses to be as right as they possible could be. One way of doing this would be to ensure that in the long run his estimates (guesses) averaged out to the true parameter value. In statistics, this 'long run' is called the *expectation,* written 'E' so that he would specify that

$$E(\hat{a}_x) = a_x$$

i.e. expectation of estimates = true parameter value.

If this condition holds, the estimates are said to be *unbiased*. In Figure 9, the curves A and C are for unbiased sets of estimates whereas curve B is for estimates that are biased downwards. Looking at the difference between curves A and C, the estimates summarised by 'A' are clearly a better 'bet' than those summarised by C because they have much lower variability around the true value.
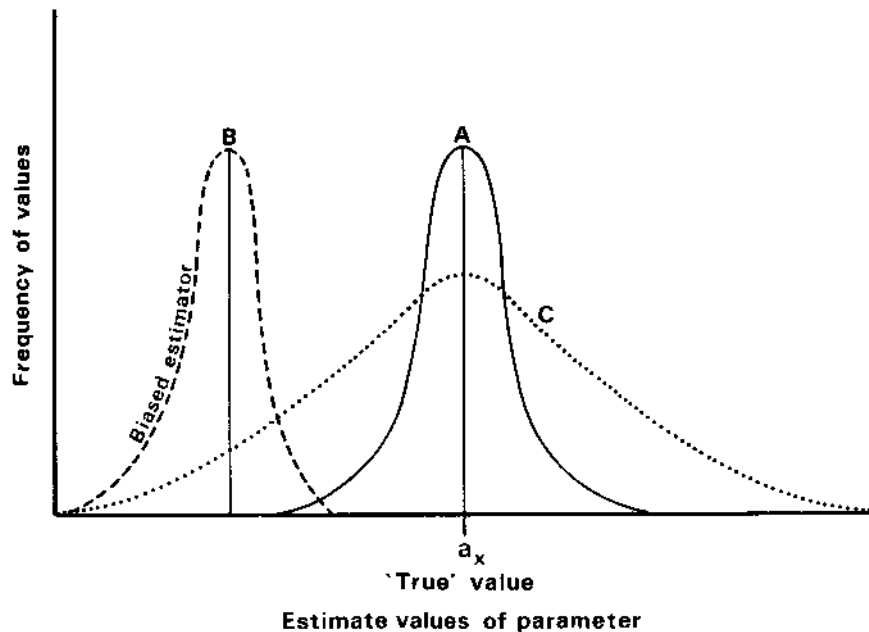
19

Fig 9   Distributions of estimates of a parameter a_x.

Most desirable of all would be a curve of estimates that had the minimum poss-
ible spread around a_x; this is called the property of *minimum variance. We
chose least squares estimates because it can be shown that these are the best
(i.e. have minimum variance) linear unbiased estimates (B.L.U.E.) of the pop-
ulation parameters, if and only if, certain assumptions hold true.*

   As Poole and O'Farrell (1971) point out, and contrary to what is often
held in the geographical literature, these assumptions are made about the
residuals and not directly about the original data. They can be stated as
follows:
A) The residuals have an expected mean value of zero.
B) They are uncorrelated with a constant variance σ^2. The assumption of
   constant variance is that of *homoscedasticity*. It implies, for example,
   that the variation of the original spatial series about the fitted trend
   surface is roughly the same no matter what part of the map we examine.
   The assumption of uncorrelated residuals is also a very strong one, es-
   pecially in our spatial series analysis where it is likely that the res-
   iduals will show *spatial autocorrelation* (Cliff and Ord, 1973) evidenced
   by the grouping together of similar residual values above or below the
   trend. Recently, tests for spatial autocorrelation in least squares res-
   iduals have been developed and could be used to test whether or not this
   assumption is valid.
C) The matrix X consists of non-stochastic elements and has a rank less than
   the number of data points. By 'non-stochastic elements' we mean that
   there is no error in the various summations that make up the X matrix

which is equivalent to specifying that the locations (x_i, y_i) of the data
points can be measured with only negligible measurement error. By 'rank
less than the number of data points' we mean that there must be at least
as many observations (N) as there are terms in the trend surface equation
(i.e. rows or columns - rank - in the matrix). Thus, to fit a linear
surface, we must have at least 3 and preferably many more observations.
If these conditions are not met, it is impossible to solve (5) and so es-
timate the trend equation constants. The requirement of negligible meas   -
urement error in (x_i, y_i) is worthy of some comment, because the author
can think of situations in which the researcher might be tempted to use
a trend surface analysis when the data points cannot be exactly located,
as for example values determined from satellite photography, or at sea,
or even on objects that themselves move! Any such observations would in-
evitable involve a stochastic (random) element in them. Assumptions (1)
to (3) above are sufficient to ensure that any least squares estimates
will have the 'best-linear unbiased' property, but if we intend in add-
ition to make probability statements about the underlying 'population'
it is necessary to make one further assumption:
D) The total population of residuals, or *disturbances* have a normal distri-
   bution. This assumption of normality is vital to any statistical testing
   in trend surface analysis, but clearly it need not always be completely
   justified. It can be defended on both theoretical and practical grounds.
   Theoretically, if the disturbances are interpreted as the effect of a
   large number of independent factors, each with a minor influence, then
   by the central limit theorem a normal distribution seems to be the most
   likely result. Practically, mathematical statisticians know a great deal
   about normal distributions and in practice if we are not prepared to make
   this assumption then our analysis cannot proceed very far. Certainly,
   none of the significance tests to be discussed in the following section
   could be applied were this assumption not to be made. The topic of least
   squares estimation when (A) to (D) above cannot be justified is a complex
   one, but a number of methods are now available and are detailed in advanced
   texts.

(vi) <u>Significance testing in trend surface analysis.</u>

   If the assumptions outlined in (v) seem justified, then it is possible to
use the estimates of the various population parameters obtained to make infer-
ences about that population, but there exists a considerable debate and dispute
as to the role and relevance of such inferential work in trend analysis.

   In our example (iv) we used sample data to estimate the population para-
meters a_0, a_1 and a_2 but in addition, the % RSS obtained for this sample is
also an estimate of a 'true' population value and in many applications the re-
searcher will want to know whether this obtained fit is simply a chance phen-
omenon or a significant result indicating a real spatial trend in the under-
lying population. Irrespective of how close to 100 the % RSS is, we still need
to know if this fit is significant. In a trend surface analysis, we can ask
at least four questions regarding the statistical significance of a fitted
surface:
A) <u>Is the total trend surface giving a fit that is significantly different</u>
   <u>from zero?</u>
   In effect, this gives a test of the hypothesis that all the population
   parameters are zero and is tested by separating out the two sources of
   variation - trend and residuals - into which the spatial series has been
   decomposed to form a variance ratio, F, according to the relation:

$$F = \frac{\% \text{ RSS} / df_1}{(100 - \% \text{ RSS})/df_2} \qquad \qquad \dots \text{(14)}$$

df_1 = degrees of freedom associated with the surface, equal to the number of constants in the trend equation *less one* for the base term (a_0).

df_2 = degrees of freedom associated with the residuals. This is given by the total degrees of freedom in the data, less df_1, those associated with the trend. As usual, the total degrees of freedom is N (sample size) less 1, so that $df_2 = N - 1 - df_1$.

It will be remembered that our fitted linear surface (II(ii)) had three constant terms, a % RSS of 95.60524, and was based upon a sample of 10 points. Hence

$$df_1 = 3 - 1 = 2$$

$$df_2 = 10 - 1 - 2 = 7$$

$$F = \frac{95.60524/2}{(100-95.60524)/7} = \frac{47.80262}{0.62782} = \underline{76.1406}$$

The tabulated critical value of F with 2 and 7 degrees of freedom at the 99.9% confidence level is 21.69 (Lindley and Miller, 1962) a value that is greatly exceeded by the observed F and we would therefore infer that the trend is a statistically significant one.

B) **Does a trend surface of order n + 1 give a significant** *improvement* **over one of order n?**

This second type of significance test is more complex to understand but it is one which is necessitated by the common practice, particularly in published computer programs, of fitting not one, but a series of trend surfaces of successively higher order (linear, quadratic, cubic and so on) to the same data points and then selecting from these the surface which seems to give the 'best' result. In such an analysis there is obviously little point in fitting, say, a quadratic (order, n=2) when the extra three 'quadratic' components added to the trend equation do not significantly improve on the fit obtained for the linear surface (order, n=1). Similarly, if the quadratic does give a significantly better fit, then does the addition of further cubic terms improve the fit still further? The analysis is similar to that presented under (A) except that we deal with the extra *increment* in % RSS given by the *quadratic components* rather than the % RSS given by the whole surface:

$$F = \frac{\text{extra \% RSS given by surface of order n+1/}df_3 \text{ over one of order n}}{(100 - \text{total \% RSS accounted for by surface})/df_4} \dots \text{(15)}$$

df_3 = degrees of freedom associated with the *added* components, 3 for a quadratic over a linear, 4 for a cubic over a quadratic and so on.

df_4 = degrees of freedom associated with the residuals calculated as before as total degrees of freedom less the number of constants in the trend equation.

(Notice that if we test a linear surface against no surface at all, $df_3 = 2$, the extra % RSS = the actual % RSS obtained, so that equation (14) is seen to be merely a special case of (15). In our simple example

(II(ii)) we could have gone on to fit a quadratic surface given by the trend equation:

$$z_{t_i} = 5.825 + 0.115x_i + 2.500y_i + \underbrace{0.457x_i{}^2 + 0.094x_iy_i - 0.199y_i{}^2}_{\text{quadratic components}}$$

This surface improves the linear % RSS by 1.6178% to 97.2230%.

Fig. 10 <u>Complete Analysis of variance for the linear and quadratic surfaces in the simple example.</u>

Source of variation	Degrees freedom	% RSS	Mean square
Total, 10 data points	9		
Due to linear surface with 3 constants	2	95.60524	47.8026
			76.140
Due to residuals over linear surface	7	4.39476	0.6278
Due to added quadratic components	3	1.6178	0.5393
			0.777
Due to residuals over quadratic surface	4	2.7770	0.6943

A full analysis of variance is presented as Figure 10. The first part of the table, for the linear surface, we have seen before, but the F ratio <u>associated with</u> the extra quadratic components is only 0.777 which is not significant at the 95% confidence level. We therefore infer that the quadratic surface doe-s not add significantly to the **fit** already obtained for the linear and there is no point in considering it further. It is interesting and salutory to n-ote that had we incorrectly tested not the quadratic increment but the entire surface against an alternative of no fit at all as in method (A) above than the quadratic fit is seen to be statistically significant. The apparent conflict between the two forms of analysis of variance is discussed in papers by Chayes and Suzuki (1963), Whitten (1963b) and Chayes (1970).

Because they are easy to calculate (A) and (B) are by far the most frequently applied significance tests in trend surface analysis, but there are a series of possible alternative tests.

C) <u>Is a single specified trend parameter significantly different from zero?</u>

If a term in the polynomial that makes up a trend equation has associated with it a parameter (a_0, a_1 etc) that does not differ significantly from zero then that term does nothing to improve the surface fit and might as well be omitted from the equation. The hypothesis that a single parameter is zero can be tested using an analysis of variance lay-out:

Source of Variation	degrees of freedom	% RSS (fit)
Due to all p terms in trend equation	p − 1	% RSS of entire surface ([1])
Due to equation less the specified parameter	p − 2	% RSS obtained if fit recalculated with parameter set at zero (2)
Extra due to single parameter	1	(3) = (1) − (2)
,Residual	N − 1 − p	(4) = 100 − (1)

The F ratio is now

$$F = \frac{\text{mean square of extra due to single parameter}}{\text{mean square of residuals}}$$

This analysis of variance is equivalent to comparing the estimate of the parameter with its standard error by means of a t-test (Davies, 1954) and is used in the published computer programs that do a so-called 'step-wise' trend analysis (Miesch and Connor, 1968). Instead of adding all the terms of a polynomial trend equation at once into the analysis, the computer program successively adds individual terms of increasing complexity. At each addition a complete trend analysis is undertaken and an F-ratio calculated for the added term. Should this prove statistically significant the term is retained but if it is not then the term is dropped from the trend equation. At the conclusion of such an analysis the trend equation contains only those terms that add significantly to the fit and can thus be thought of as the 'best' trend for those data

A recently reported extension of this one parameter test is to ask whether or not the parameters of any two trend surfaces (of the same order and calculated for the same number of data points) are equal (Earickson, Jones and Murton, 1972). Effectively this tests whether or not any two surfaces differ significantly in their shape and is thus a useful test when looking at trend maps derived from different spatially distributed variables.

D) where, in the mapped area is the fitted trend surface most reliable as an estimate of the population trend and where should it be treated with caution?
This final problem in .assessing the significance of a fitted surface can be answered by calculating confidence surfaces at a specified confidence level above and below the trend in a manner analogous to the confidence limits usually reported in ordinary two variable regression. Full details of the theory and calculation of these surfaces are to be found in Krumbein's expository paper (Krumbein, 1963) and an easy-to-follow worked example in Krumbein and Graybill (1965).

An example of a confidence surface applied to a practical problem is shown in Figure 11. Figure 11 (a) shows the best-fit linear trend surface calculated for the lip-altitudes of 84 cirques in northern Snowdonia, Wales (Unwin, 1973). Figure 11 (b) shows the 95% half confidence inter-
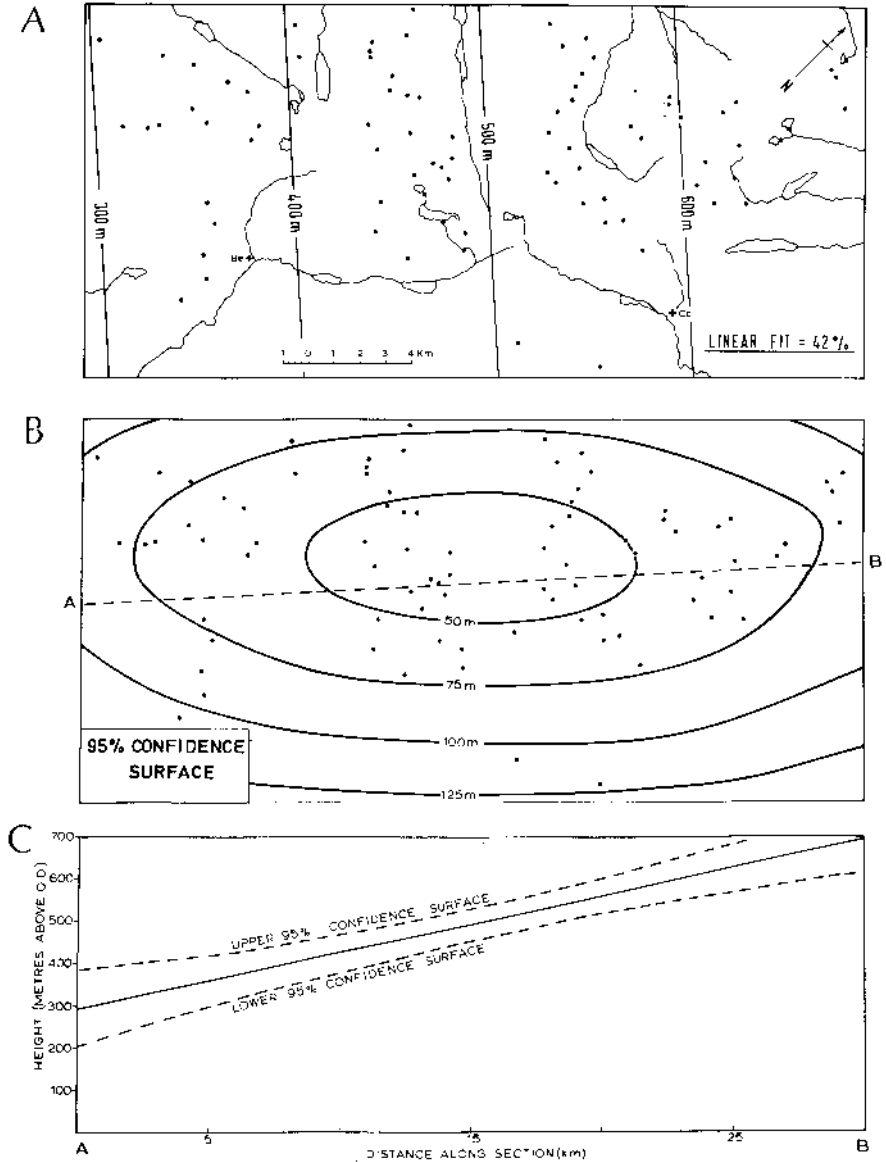
Fig. 11   Confidence surfaces for the Snowdonian cirques linear surface.

val width for this surface and is a good index of the degree of trust we can place in the fitted linear surface. It is clear that the minimum half confidence width occurs in the centre of the study area at the location $(\bar{x}_i, \bar{y}_i)$ and that as we move away from this area of maximum reliability se the half confidence interval width increases indicating a decrease in surface reliability. Figure 11 (c) shows a section along the line of maximum dip of the fitted surface on which the trend and both upper and lower confidence surfaces have been plotted. Again, it is clear that the fitted trend is a real effect.

III AN EXAMPLE OF A PRACTICAL APPLICATION: DAVIES AND GAMM (1969).

(i) Introduction.

Given the fundamental importance of spatial change to geography, it is hardly surprising that there have been a great many applications of trend surface analysis since the method was introduced in the mid 1960's. An interesting and relatively straightforward application that illustrates the method very clearly is that presented by Davies and Gamm (1969) in a paper Trend surface analysis applied to soil reaction values from Kent, England (Geoderma, 3, 223-31). They report the application of the method to soil reaction (pH) data for a small (2 x 1km.) area crossing the cuesta of the North Downs 1 km. north of Westerham in Kent, England. Figure 12 shows the location of this study area. The soils are mostly brown earths on the plateau to the north, rendzinas on the exposed chalk with gleyed brown earths on the Gault Clay in the south, and were sampled at 213 locations. Following appropriate treatment, surface soil pH was determined in the laboratory by electrometric methods.

pH values reflect the base status of the soil, and in England this often depends upon the soil parent material. Alkali values (pH>7) occur when calcium carbonate occurs free in the soil, acid values (pH<7) when free hydrogen ions are present.

(ii) Hypothesis under test.

In southern England there is a repetition of calcareous and non-calcareous soil parent materials which leads to a repetition of differing soil reaction values. As Figure 12 shows this is true of the study area. In the south, soils are developed upon Gault Clay and Upper Greensand. Crossing the centre of the area is a band of chalk exposed on the cuesta of the North Downs which in the extreme north of the area is masked by clay-with-flints ona plateau-like surface. Given this distribution of soil parent materials, Davies and Gamin hypothesised that 'pH values would reach an alkali maximum over the chalk outcrop whilst to the north and south, over non-calcareous deposits, they would change to moderately acid values'. This hypothesis would lead one to expect a quadratic or cubic polynomial as the most appropriate model, leaving residuals that might be expected to pick out locally important environmental influences on soil pH. The collected data were in sufficient number and measured with sufficient accuracy to make a surface analysis possible.

(iii)Results

Best fit, least-squares linear, quadratic, and cubic polynomial trend
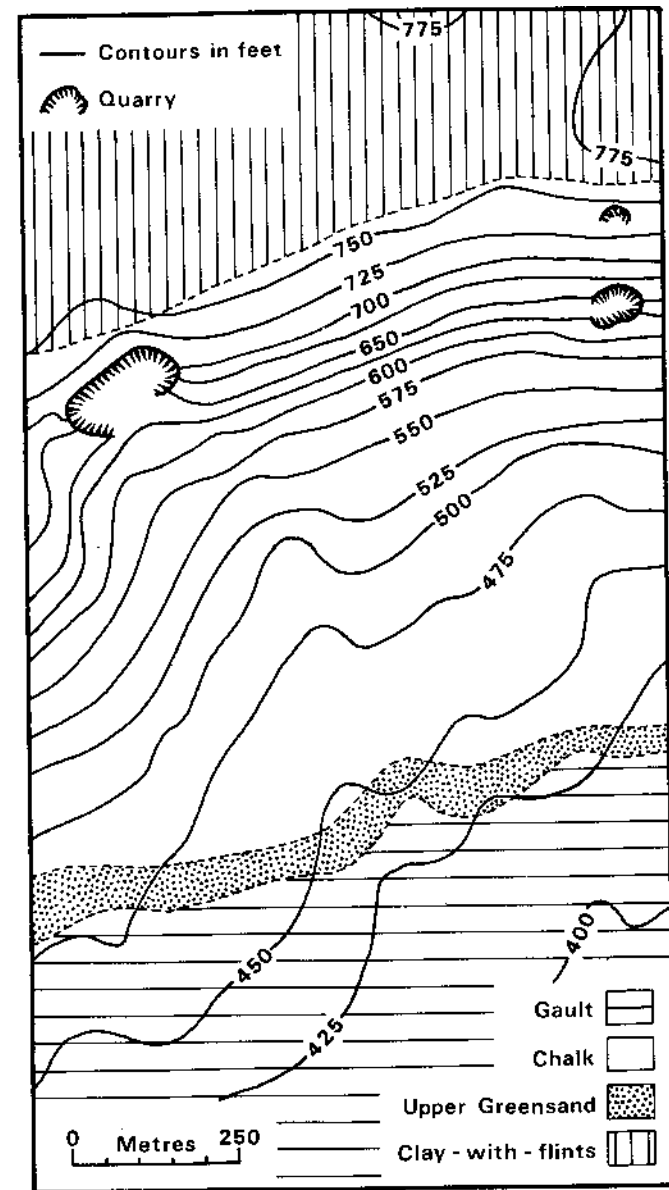
Fig. 12 Location and geology of study area (redrawn after Davies and Gamm, 1969, with permission).

surfaces were fitted to the 213 pH values. The results are summarised in Figure 13 which gives the % RSS values and a complete analysis of variance for all three surfaces. Examination of the residuals suggests that none of the assumptions listed in II(v) have been seriously violated. The linear surface had a very low fit, indicative of (Figure 8) a slight, almost- negligible linear trend. The quadratic surface showed a marked improvement in fit giving a value of 40.97% indicative of a substantial quadratic trend in the data whilst the cubic, with a % RSS of 48.05 could aTmost be described as having a 'marked' trend and accounted for almost half the observed variation in soil pH.

Fig. 13   % RSS and analysis of variance for the Davies and Gamm example.

Surface	% RSS obtained
Linear	3.27
Quadratic	40.97
Cubic	48.05

Source of Variation	Degrees Freedom	% RSS	Mean ire	F
Total, 213 data points	212			
Due to linear surface with 3 constants	2	3.27	1.635	} 3.549
Due to residuals over linear surface	210	96.73	0.461	}
Due to added quadratic components	3	37.7	12.566	} 44.067
Due to residuals over quadratic surface	207	59.03	0.285	}
Due to added cubic component	4	7.08	1.770	} 6.916
Due to residuals over cubic surface	203	51.95	0.256	}

   The analysis of variance table has been reworked from the original Davies and Gamm table using % RSS values instead of the actual sums of squares they used. Because 'a succession of surfaces was fitted the analysis proceeds according to method B in II(vi) and tests the significance of the increment obtained at each step. For the linear surface the F-ratio is 3.549 with df = 2, 210 which is just significant at the 95% level but is not significant at 97.5%. The addition of three quadratic components allowed the trend surface to assume a dome-like shape centred on the chalk outcrop and led to a very highly significant F value of 44.067 which with df   3 and 207 is significant at 99.9% . The added flexibility of the cubic improved the fit still further giving an F value of 6.916 which is also significant at 99.9%.

   Accordingly, Davies and Gamm selected the cubic surface for further examination. Its full equation (cf equation 10) was
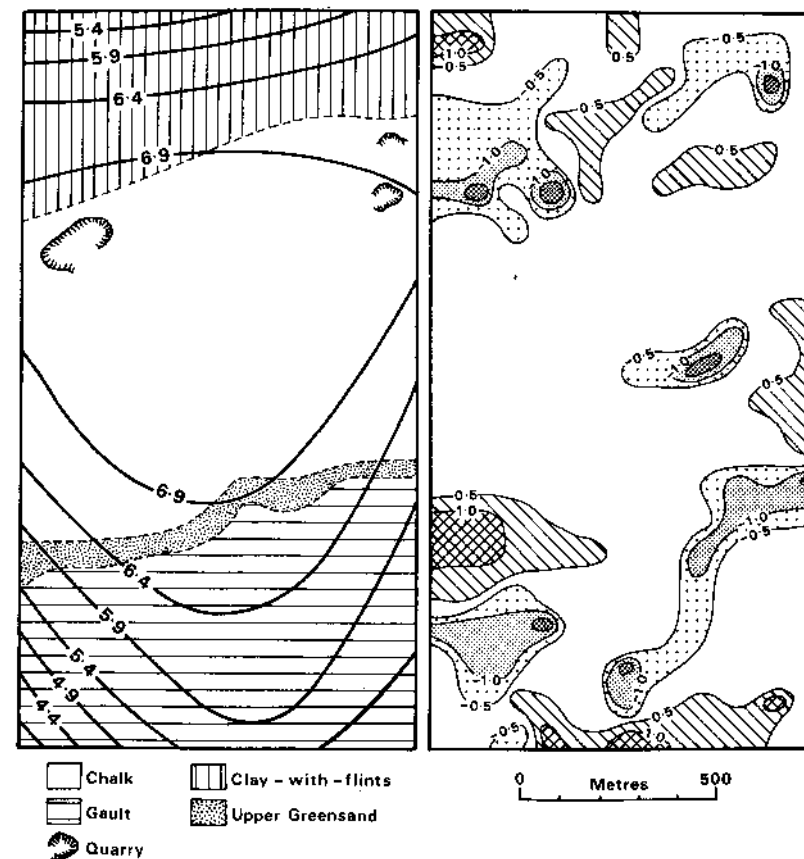
Fig. 14   Cubic trend surface and residuals for Davies and Gamm example.

$$pH_{obs_i} = 0.507 + 0568x_i + 0.484y_i - 0.453x_i^2 + 0.531x_iy_i$$

$$+ 0.490y_i^2 + 0703x_i^3 + 0141x_i^2y_i - 0.322x_iy_i^2 + 0.506y_i^3$$

A contour map of this surface together with its residuals drawn to the same scale is shown as Figure 14. The surface form verified the original idea that there should be an alkali maximum over the chalk cuesta with a trend to acid values away to the north and south. In the north, the isoline for pH = 6.4 corresponds with the top of the cuesta dip-slope and the isolines for pH less than this are sub-parallel trending WSW to ENE in sympathy with the general strike of the chalk outcrop. Davies and Gamm point out that in the field this area of increasingly acid soils corresponds to a non-calcareous superficial_ deposit which thickens away from the scarp edge. The zone of alkali maximum, pH greater that 6.9 is that of the rendzina soils on

the cuesta itself whilst the lobe-like feature of alkalinity which runs south-wards from this corresponds with the axis of a dry valley whose floor is mantled with chalky 'head' deposits. In the southwest values fall rapidly to mildly acid soils developed on the Gault Clay. This pattern of regional variability is seen to be closely in accord with the predicted pattern and illustrates the importance of soil parent material in determining base status in this area.

Further site-by-site analysis of the residuals also shows that they are successfulin isolating local pedogenetic variations due to changes in land us─. Figure 14(b) shows the distribution of residuals over the cubic surface with isolines drawn at +1.0, +0.5, -0.5 and -1.0pH units. Considering residuals in excess of ±1.0 unit as being truly anomolous, Davies and Gamm were able to identify five areas of soils more acidthan the regional trend predicts. These all have negative residuals and are identified by stipple on Figure 14. In 4 of these areas the soils were developed upon uncultivated woodland or old per-manent pasture whilst the fifth area coincided with an isolated outcrop of non-calcareous silty-clay head deposits. Similarly, areas more alkali than the trend predicts give positive residuals and are identified by a cross-hatching on Figure 14(b).With few exceptions, these areas coincided with the cultivation of barley or wheat whilst similar soils conforming with the reg-ional trend were all under pasture.

In conclusion, Davies and Gamm suggest that their analysis was successful since "statistically significant surfaces were generated which accorded with predictions based on the known properties of soils formed on parent materials similar to those of the study area. Furthermore, the residual values were generally explicable by commonly accepted influences on soil reaction. . .". They go on to suggest a more widespread use of the technique to help improve the reliability of soil survey and the objectivity of regional description.

IV PROBLEMS AND PITFALLS IN TREND SURFACE ANALYSIS

(i)   Introduction.

There are a number of practical and theoretical problems confronting the would-be trend analyst. Unwin and Hepple (1975) consider these problems in summary, and what follows is largely an elaboration on that work. In under-taking a trend surface analysis the researcher has to collect suitable data, specify and fit the correct trend model and make statistical inferences about the underlying population. At each step there are pitfalls that can seriously upset the analysis, ranging from a minor distortion in the fitted surface to a total invalidation of the results. It is not our purpose to argue that all trend analyses must conform to a series of rigid rules before they are accepted as useful substantive tools in geographic research; rather we argue that if an investigator choses to 'bend' the rules a little then this bending should be explicitly recognised and justified rather than being hidden beneath a mass of numerical results. For convenience, the pitfalls will be dealt with in the order they are likely to be met with in practice but it should be emphasised that they are not independent of each other. Inadequate data make trend specification, fitting and testing almost impossible; similarly, a badly for-mulated model fitted to good data makes inferential work difficult and an in-correctly applied significance test can ruin even a good model fitted to good data:

(ii) Data suitability.

At first sight any spatial series might be thought suitable for trend analysis, but if we go back to equation (1) it is apparent that this assumes spatial *continuity*. At every possible location there is exists a value for the elevation of the series z_{obs}, yet geographic data on phenomena such as settlement populations, cirque lip altitudes and factory production are in-trinsically *point-valued* and in no sense are drawn from a continuous surface such as those given by atmospheric pressure, annual rainfall and temperatures which in principle might be measured anywhere. Alternatively, other common spatial series such as crop acreages or population densities are *area-valued*, relating to a regular or irregular lattice of grid squares, counties, states or other administrative units.

The first type of discontinuity, that of point-valued data, has received most attention in the literature because it is difficult to justify using con-tinuous functions to describe inherently discontinuous data (Robinson, 1970). If we fit a trend surface to city sizes in England is it possible to use the results to argue that the population of Newport Pagnell should be 4.5m simply on the basis of its lying midway between Birmingham (1m) and London (8m)? Obviously the results are nonsense unless the investigator is able to provide a theoretical link between the discrete points and the trend surface threaded through them.

The second type of discontinuity, that related to area-valued data is extremely difficult to justify in a trend surface analysis and should be avoi-ded. In order to fit surfaces some reference point within each area must be chosen and the data for that area treated as if it had originated at that point. Nordbeck (1962) provides a discussion of how to locate areal data for computer processing and Tarrant (1969) a practical example, but the results obtained must depend to a greater or lesser extent on the sizes and shapes of the areal units and the choice of reference point. The use of similarly shaped areas of equal size that is small relative to the total mapped area will do much to minimise these effects but only seldom are such data sets available. In physical geography, problems of data discontinuity do not occur often yet in human geography, as for example when census data are analysed, they are likely to be the norm rather than the exception and in its unmodified form trend surface analysis should not be used (see Cerny (1973) for a discussion of this point).

There must also be a sufficient number of data points available for anal-ysis. At an absolute minimum (II(v), assumption C) there must be at least as many data points as there are constants to be estimated, but in practice many more are required to ensure that meaningful inferential tests can be carried out.

Assuming that the data are basically suitable for trend analysis, there remains the problem of their distribution in space and its effects upon the surfaces (Miesch and Connor, 1968; Unwin, 1970; Doveton and Parsley, 1970). Usually, the trend analyst does not have control over the sample locations and his samples are therefore *given* or *unplanned* (Box, 1966). In the soil reaction study, Davies and Gamm were able to sample the spatial series according to any suitable design and used a grid sample, but in many analyses no such choice is available. If point-valued data are analysed the 'sample' is obviously fixed by the location of the objects under investigation and for continuous series often the only available data are located by the fixed positions of monitoring

stations. Technically, the trend surface model outlined in II(i) is of the *regression* type that treats the data locations as fixed variables with no requirement that their spatial distribution be of a particular type but Doveton and Parsley's (1970) experiments show that the arrangement of points can greatly distort the results.

One noticeable effect occurs at the edges of the mapped area where high-order surfaces (quadratic and above) lack data point control and can give pronounced 'edge effects' as whatever slope exists in the controlled area is extrapolated outside the control at the map edges. To guard against these effects it is sensible to form a 'buffer region' (Davis, 1973) of control points outside the actual area of interest. If the map area contains many points then this can be quite narrow, but if the point-density is low a much wider zone becomes necessary.

A second effect is that of the shape of the mapped area. If the area under consideration is markedly rectangular rather than square there is a pronounced tendency for contours on higher order surfaces to become elongate parallel to the longer side.

Finally, great care should be exercised in dealing with the data point distributions that are in any way peculiar, such as points distributed along lines or grouped in pronounced clusters. At least two studies (Miesch and Connor, 1968; Unwin, 1975) show that certain types of distribution can produce very badly conditioned matrices of sums or squares and cross-products making the least squares solution very difficult to find and which, even if found, may be so unstable to small changes in the initial data as to be worth-less. Although the fitted surfaces may not be too severely distorted by the presence of clustered data the estimate of the true % RSS obtained can be extremely unreliable and the 'effective' degrees of freedom almost impossible to determine so that conventional significance tests become invalid.

The seriousness of these effects depends strongly upon the degree of inadequacy of the data cover, the type of surface to be fitted and the true population fit, being least for the low order surfaces with a good fit and highest for poorly-fitting high order surfaces, but studies have shown that in practical applications provided all parts of the mapped area have some control they are unlikely to be as severe as was once feared. (Doveton and Parsley, 1970; Robinson, 1970, 1972).

(iii) Trend specification.

In section II(iv) a number of possible trend surface equations of increasing complexity were outlined and the selection from these of the most appropriate model is without doubt the most difficult problem in surface analysis. It is compounded by a lack of clarity in the *aims* of the analysis (see section I(ii)). Some authors, notably Norcliffe (1969), regard it as a *formal* hypothesis-testing procedure in which a theory is propounded and a specific trend surface model for that theory chosen *a priori*. Data are collected and the model fitted. All that remains are significance tests to establish whether or not the observed fit is significant. If it is the theory can be accepted, but if not then it must be revised and a new model formulated. In an ideal world in which spatial theory was abundant such an approach would be common, but in practice the relative absence of true spatial theory has meant that very few studies explicitly use trend functions derived from theoretical considerations. A second, philosophic problem arises from the acknowledged difficulty in using spatial form as a test of any spatial theory (King,

32

1969). In particular, the observation that different spatial processes can produce similar spatial structures means that the formal approach is always dealing with trend surface models that are over-identified. An alternative, *informal* approach to trend specification is suggested by Robinson (1970). In this, trend surfaces are used as simple descriptive devices for spatial generalisations that might suggest *a posteriori* theory. The approach is therefore concerned to find a trend-equation that gives a reasonable fit to the observed data without worrying overmuch about its theoretical implications but, as Unwin and Hepple (1975) point out, this is likely to degenerate into an unsatisfactory hit or miss procedure. There is a variety of possible trend surface equations and there might well be several possible surfaces that give equally good fits to the observations. In the absence of any theoretical insight it is not immediately clear how the 'best' of these surfaces can be recognised. The 'stepwise' regression procedure adopted by Miersch and Connor (1968) is a possible but scientifically unsatisfactory solution to the problem.

(iv) Significance testing.

Although a number of significance tests may be applied to the results of a surface analysis (Davies, 1954; Mandelbaum, 1963; Agterberg, 1964; Tinkler, 1969) it is only rarely that their assumptions will be justified. It must be assumed that the residuals have zero expectation, are uncorrelated and have a constant variance that is independent of the observed values. The problems of meeting these assumptions in practical applications of conventional regression are well known, but in surface analysis there are a number of special problems that have received attention.

The first concerns the relevance of statistical tests in applications that use intrinsically point-valued data. These are not easily thought of as a sample of anything; rather they form the total identified and identifiable population so that the fitted surfaces must be 'significant'. At first sight there is little point in appeals to statistical theory unless the data are regarded as a sample of some underlying potential or unless inferences are made about the extent to which the results can be regarded as *unusual* (Cliff, 1973).

A second problem concerns the use of data whose distribution is more clustered than random. Surfaces fitted to such data tend to give undue influence to the sparsely distributed points outside the clusters and so become distorted. In such circumstances it is virtually impossible to determine the effective degrees of freedom involved necessary to the calculation of the F - ratios.

A third problem is that of spatial autocorrelation in the residuals which is almost always present to a greater or lesser extent. Positive autocorrelation, as evidenced by the grouping together of similar residual values results in the estimates of the 'unexplained' variance being smaller than they should be which in turn alters the F-distribution from its tabulated form. If this bias is serious and the investigation takes the published F tabulations at their face value, there is an increased probability of wrongly rejecting the null hypothesis of no trend and so inferring a real effect where none exists. In many practical applications of statistical methods to geographical problems the methods are sufficiently robust to allow violations of their assumptions but despite some preliminary studies using randomisation techniques (Haworth, 1967; Unwin, 1970) at the time of writing we

33

simply do not know how robust trend surface analysis is. What is clear
however, is that *surfaces that give poor,"just significant': fits, and the re-
siduals associated with them, should be* treated *with extreme caution.*

V DEVELOPMENTS IN TREND SURFACE ANALYSIS

(i) Introduction.

The basic trend surface model deals with data that have a primary char-
acteristic of spatial distribution, but it may well be that these spatial data
have other characteristics that make a conventional analysis difficult or im-
possible to apply or that necessitate alternative surface fitting techniques.
The basic model is very flexible and is capable of a great many variations
and extensions. In this final section we shall examine some of the variations
of the polynomial and Fourier models to enable them to incorporate non-stan-
dard data before going on to briefly examine some alternative surface-fitting
techniques.

(ii) Data with special attributes.

It may sometimes be necessary to locate each data sampling point by three
rather than two spatial co-ordinates in which the extra co-ordinate is some
other locational measure such as height above sea level and depth below sur-
face. Mathematically, this creates only a slight added complication to the
basic model (Equation 1) which is now modified to define a <u>four-variable trend
surface:</u>

$$z_{obs_i} = f(x_i, y_i, w_i) + u_i$$

As before, the function f can be expanded in any suitable way but to-date all
work has used polynomial expansions of which the first is:

$$z_{obs_i} = a_0 + a_1 x_i + a_2 y_i + a_3 w_i + u_i$$

The mathematical development of the least squares solution proceeds in exactly
the same way as before except that there are now four constants to be estim-
ated (a_0 to a_3) and hence four normal equations to be solved simultaneously.
Extension to higher orders is performed by expanding the polynomials as before
but the number of constants to be estimated, and hence the data requirement,
rises very rapidly indeed. In all cases the effect is to define *'hypersurfaces*
of best-fit in which the calculated trend in z is 'explained' by variation in
its three-dimensional location. This model was originally developed in get-
rology by workers such as Peikert (1962) for studies in which the z_{obs} are
some measured rock property (specific gravity, mineral composition etc) and
the (x_i, y_i, w_i) represent its location within a three-dimensional rock body
such as an igneous intrusion but it may have considerable application in geo-
graphy as a means of modelling spatial change using wi as a measure of time.
Computer programs for'hypersurface fitting using polynomials have been pub-
lished by Peikert (1963) and by Esler, Smith and Davis (1968) whilst a com-
prehensive review of geological work is to be found in Davis (1973).

Another difficulty often faced in practical applications is that the data
are best represented as located *vector* quantities that have both a magnitude

34

and a direction (usually measured in ^oE of N). The dip (magnitude) and orien-
tation (direction) of cobble long axes in sedimentary deposits is a familiar
example of vector data, but others might include wind speed and direction or
drumlin size and orientation. Such data are periodic in that every 360° there
occurs a repetition of the sequence and do not obey the normal rules of arith-
metic. The trend analysis of such data has been developed by Fox (1967) and
is called *vector trend analysis.*

Alternatively, the spatially distributed data might be multivariate so
that we are interested in spatial trends in several variables treated simul-
taneously. One possible solution is to trend analyse component or factor
scores derived from a previous study, but a more elegant technique makes use
of *canonical trend surfaces* (Lee, 1969). The theory and application of can-
onical methods are dealt with in a previous CATMOG (Clarke, 1975); all that
the trend surface variant does is to replace one side of the canonical equation
by the usual polynomial trend surface functions.

Whilst the above four variants of surface analysis are those most likely to
be encountered, mention ought to be made of the possibility of using *logit
models* for the trend surface analysis of categorised data such as those ob-
tained from social survey (Wrigley, 1973, 1975), and of *orthogonal polynomial*
models for irregularly-spaced data (Whitten, 1970).

(iii) Alternative techniques.

We have now progressed some way from the simple trend surface model in-
troduced at the start of this monograph. Most of the study has been an expo-
sitory account of one simple class of model for describing and testing hypoth-
eses about spatial variation which involves the decomposition of a spatial
series into just two components associated with a trend of assumed form and
residuals. At this point it is necessary to conclude by drawing the readers
attention to a selection of alternative techniques that might be used and so
place the method into a broader framework of spatial analysis.

If the analyst's aim is to produce an *objectively-defined contour map* from
his sample data and if these Bata cannot be easily represented using an ordinary
polynomial, the method known as *Kriging* is usually superior to trend analysis.
It uses a knowledge of the spatial autocorrelation structure revealed in the
data to estimate the values of a spatially distributed variable over a contin-
uous surface but in addition assesses the probable error of these estimates.
The results can be used to contour the data and to make statements about the
reliability of the contours and have been much used in ore evaluation which puts
a premium on having such precise estimates. Krige estimates are quite complex
to calculate and the reader is referred to summary reference by Matheron (1963)
for further details.

If the objective is simply to *generalise* a spatial series then matrix
smoothing and filtering methods of the type discussed by Tobler (1966; see
Basset and Chorley, 1971) ought to be considered as a possible alternative.

Finally, the *decomposition* of a spatial series into components associated
with differing scales of process need not simply use a 'trend plus residuals'
assumption. Instead, the methods of spectral analysis can be used to detect
periodicities at a number of scales pf variation (Rayner, 1971).

35

Trend surface models of the type outlined here are the oldest and in many ways the simplest of available techniques for handling spatial series, and if used correctly they remain one of the most powerful. In consequence they are the most frequently used model in spatial analysis, but there are many situations in which they are not appropriate and for which other methods are more suitable. Their study provides a useful starting point in the development of an understanding of much of the most recent work in spatial statistics.

VI BIBLIOGRAPHY

A. Theory.

Agterberg, F.P. (1964), Methods of trend surface analysis, *Colorado School of Mines Quarterly,* 59, 111-30.

Bassett, K. (1972), Numerical methods for map analysis. In *Progress in Geography, Volume 4,* (Board, Chorley, Haggett and Stoddart (Eds)), London: Arnold, 1972, 219-254.

Bassett, K.A. and Chorley, R.J. (1971), An experiment in terrain filtering. *Area,* 3(2), 78-91.

Box, G.E.P. (1966), Use and abuse of regression. *Technometrics,8,625-9.*

Cerny, J.W. (1973), Social data and trend surfaces: a comment. *Geographica Analysis,* 5, 156-159.

Chayes, F. (1970), On deciding whether trend surfaces of progressively higher order are meaningful. *Geological Society of America, Bulletin,* 81, 1273-78.

Chayes, F. and Suzuki, Y. (1973), Geological contours and trend surfaces; a discussion. *Journal Petrology,* 4, 307-312.

Chorley, R.J. and Haggett, P. (1965), Trend-surface mapping in geographical research. *Transactions, Institute British Geographers,* 37, 47-67.

Clarke, D. (1975), An introduction to canonical correlation. *CATMOG,* 3,

Cliff, A.D. (1973), A note on statistical hypothesis testing. *Area,* 5(3), 240.

Cliff, A.D. and Ord, J.K. (1973), *Spatial Autocorrelation,* London: Pion.

Davies, O.L. (1954), *The design and Analysis of Industrial Experiments,* New York: Hafner.

Doveton, J.H. and Parsley, A.J. (1970), Experimental evaluation of trend surface distortions induced by inadequate data point distributions. *Transactions, Institute Mining and Metallorgy,* Section B, B197-B208.

Earickson, R., Jones, R.H., and Murton, B.J. (1972), An extension of spatial analysis to social behaviour. *Geographical Analysis, 4,* 65-80.

Grant, F. (1961), A problem in the analysis of geophysical data. *Geophysics,* 22, 309-44.

Harbaugh, J.W. and Merriam, D.F. (1968), *Computer Applications in Stratigraphic Analysis.* New York: Wiley.

Harbaugh, J.W. and Preston, F.W. (1968), Fourier analysis in Geology. In Berry B.L.J. and Marble, D.F. (Eds.), *Spatial Analysis: a reader in Stetistical Geography,* Englewood Cliffs, N.J.: Prentice Hall.

Hill, W.J. and Hunter, W.G. (1968), A review of response surface methodology: a literature survey. *Technometrics,* 8, 571-90.

Howarth, R.J. (1967), Trend surface fitting to random data - an experimental test. *American Journal Science,* 265, 619-25.

King, L.J. (1969), The analysis of spatial form and its relation to geographic theory. *Annals Association American Geographers,* 59, 573-95.

Koerts, J. and Abrahamse, A.D.J. (1971), *On the Theory and Application of the General Linear Model,* Rotterdam.

Krumbein, W.C. (1959), Trend-surface analysis of contour-type maps with irregular control point spacing. *Journal Geophysical Research,* 64, 823-34.

Krumbein, W.C.(1963), Confidence intervals on low-order polynomial trend surfaces. *Journal Geophysical Research,* 68, 5869-5878.

Krumbein W.C. (1966), A comparison of polynomial and Fourier models in map analysis. *Northwestern University, Department of Geography, Technical Report,* 2 of *O.N.R. Task,* 388-078.

Krumbein, W.C. and Graybill, F.A. (1965), *An introduction to Statistical Models in Geology,* New York: McGraw Hill.

Lee, P.J. (1969), The theory and application of canonical trend surfaces. *Journal of Geology,* 77, 303-18.

Lindley, D. V. and Miller, J.C.P. (1962), *Cambridge Elementary Statistical Tables* Cambridge C.U.P.,Tables 7(a) - 7(c).

Mandelbaum, H. (1963), Statistical and geological implications of trend mapping with non-orthogonal polynomials. *Journal Geophysical Research,* 68, 505-19.

Matheron, G. (1963), Principles of geostatistics. *Economic Geology,* 58, 1246-1266.

Matheron, G. (1967), Kriging or polynomial interpolation procedures. *Canadian Institute of Mining Bulletin,60,* 1041-45.

Norcliffe, G.B. (1969), On the uses and limitations of trend-surface models. *Canadian Geographer,* 13, 338-48.

Nordbeck, S. (1962), The location of areal data for computer processing. *Lund Studies in Geography.(C),* 2. 41 pp.

Peikert, E.W. (1962), Three dimensional specific gravity variations in the Glen Alpine Stock, Sierra Nevada, California. Geological *Society of America Bulletin,* 73, 1437-42.

Poole, M.A. and O'Farrell, P.N. (1971), The assumptions of the linear regression model. *Transactions, Institute British Geographers,* 52, 145-158.

Rayner, J.N. (1971), *An introduction to Spectral Analysis.* London: Pion.

Robinson, G. (1970), Some comments on trend surface analysis, *Area,* 1970, 3, 31-6.

Robinson, G. (1972), Trials on trends through clusters of cirques,Area, 4, 104-13.

Tinkler, K.J. (1969), Trend surfaces with low "explanations"; the assessment of their significance. *American Journal of Science,* 267, 114-23.

Tobler, W.R. (1966), Of maps and matrices. *Journal of Regional Science,* 7, 234-52.

Unwin, D.J. (1970), Percentage RSS in trend surface analysis, *Area,* 1, 25-8.

Unwin, D.J. (1973) , Trials on trends.Area, 5(1), 31-33.

Unwin, D.J. (1975), Numerical error in a familiar technique: a case study of polynomial trend surface analysis. *Geographical Analysis,* 7(2), 197-203.

Unwin, D.J. and Hepple, L.W. (1975), The statistical analysis of spatial series. *The Statistician,* 23, 211-227.

Whitten, E.H.T. (1963) (b), A reply to Chayes and Suzuki, *Journal of Petrology,* 4, 313-16.

Whitten, E.H.T. (1970), Orthogonal polynomial trend surfaces for irregularly spaced data. *Mathematical Geology,* 62, 141-52.

Wrigley, N. (1973), The use of percentages in geographical analysis. *Area,* 5, 183-6.

Wrigley, N. (1975), Analysing multiple alternative dependent variables. *Geographical Analysis ,* 7(2), 187-196.

B. <u>Computer programs.</u>

Davis, J.C. (1973), *Statistics and Data Analysis in Geology,* New York: Wiley.

Esler, J.E., Smith, P.F., and Davis, J.C. (1968), KW1KR8, a FORTRAN IV program for multiple regression and geologic trend analysis. *computer Contribution,* (State Geol. Survey, Univ. of Kansas). 28, 31 pp.

Fox, W.T. (1967), FORTRAN IV program for vector trend analysis of directional data. *Computer Contribution,* (State Geol. Survey, Univ. of Kansas.) 11, 36 pp.

James, W.R. (1966), FORTRAN IV program for using Double Fourier series for surface fitting or irregularly spaced data. *Computer Contribution,* (State Geol. Survey, Univ. of Kansas) 5, 19 pp.

McCullagh, M.J. (1973), Trend surface analysis. computer *Applications,* (University of Nottingham), 15, 1-47.

Miesch, A.T. and Connor J.J. (1968), Stepwise regression and non-polynomial models in trend analysis. *Computer Contribution,* (State Geol. Survey, Univ. of Kansas), 27, 40 pp.

O'Leary, M., Lippert R.L. and O.J. (1966), FORTRAN IV and map program for computation and plotting of trend surfaces for degrees 1 through 6. Computer *Contribution,* (State Geol. Survey, Univ. of Kansas), 3, 48 pp.

Peikert, E.W. (1963), IBM 7090 program for least-squares analysis of three-dimensional geological and geophysical observations. *United States Office Naval Research Geographical Branch, Technical Report,* 4, Task, 389-135, 72 pp.

Whitten, E.H.T. (1963) (a), A surface fitting program suitable for testing geological models which involve areally distributed data. *United States Office of Naval Research, Geographical Branch, Technical Report,* 2, of Task, 389-135.

C. <u>Some applications.</u>

Anderson, P. (1970). The uses and limitations of trend surface analysis in studies of urban air pollution. *Atmospheric Environment,* 4, 129-47.

Davies, B.E. and Gamm, S.A. (1969). Trend Surface analysis applied to soil reaction values from Kent, England. *Geoderma,* 3, 223-31.

Gould, P.R. (1966). On mental maps. *Discussion Paper, Michigan Inter-University Comm. of Mathematical Geography,* 9, 54 pp.

Harbaugh, J.W. (1964). A computer method for four-variable trend analysis illustrated by a study of oil-gravity variations in south-eastern Kansas. *Kansas Geological Survey, Bulletin,* 171, 58 pp.

King, C.A.M. (1969). Trend-surface analysis of Central Pennine erosion surfaces. *Transactions, Institute British Geographers,* 47, 47-59.

Mandeville, A.N. and Rodda, J.C. (1970). A contribution to the objective assessment of areal rainfall amounts. *Journal of Hydrology, (New Zealand);* 9, 281-91.

Petersen, J.A. and Robinson, G.F. (1969). Trend surface mapping of cirque floor levels. *Nature,* 222, 75-6.

Robinson, G. and Fairbairn, K.F. (1969). An application of trend surface mapping to the 'distribution of residuals from a regression. Annals *Association American Geographers,* 59, 158-70.

Robinson, G. and Salih, K.B. (1974). The spread of development around Kuala Lumpur: a methodology for an exploratory test of some assumptions of the growth-pole model. *Regional Studies,* 5, 303-14.

Rodda, J.C. (1970). A trend-surface analysis trial for the planation surfaces of north Cardingshire. *Transactions, Institute British Geographers,* 50, 107-14.

Tarrant, J.R. (1969). Some spatial variations in Irish agriculture. *Tijdschrift voor Economische en Sociale Geographie,* 60, 228-37.

Tobler, W.R. (1964). A polynomial representation of the Michigan population. *Proceedings Michigan Academy Science, Arts and Letters,49,* 445-52.

Unwin, D.J. (1969). The areal extension of rainfall records: an alternative model. *Journal of Hydrology.* 7, 404-14.

Unwin, D.J. (1973). The distribution and orientation of cirques in Northern Snowdonia, Wales. *Transactions Institute of British Geographers,* 58, 85-97.

1