Trend-Surface Mapping in Geographical Research

R. J. Chorley, M.A. and P. Haggett, M.A.

(University Lecturers in Geography, University of Cambridge)

MS. received 6 March 1965

Although geographers are traditionally concerned with the description, analysis and explanation of areal distributions of phenomena, much of the most vigorous development of new techniques in this field has come from outside geography. The harnessing of the vast potential of computer systems to mapping problems has been pressed forward both in subjects like meteorology where mechanical graph-plotters, line printers, and cathode-ray tube displays are being used to map directly the output of digital computers (Sawyer, 1960; Wippermann, 1959), and in botany where field information on the occurrence of vascular plants are being rapidly processed and printed out (for example, the 1,623 maps of the Atlas of the British Flora (Perring and Walters, 1962)). The most striking adaptations of computer technology to traditional mapping problems are being employed in the earth sciences, however: for example, in geophysical prospecting, structural mapping and sedimentary petrology (Krumbein, 1958). Here the most fundamental developments in mapping techniques lie less in the automation of traditional techniques (a field reviewed for geographers by Tobler (1959)) but in the evolution of new ways of extracting more information from map data. This paper reviews developments in one of the most promising of these techniques, trend-surface mapping, and attempts to assess its significance for, and applications in, the wider field of geographical research.

The Nature of Trend-Surface Models

Considerable insight into mapping techniques has come from the mathematical theory of information, originally developed in the late 1940s by American workers such as C. E. Shannon in relation to communications engineering. In terms of information theory we may regard the map as a 'communications channel' into which bits of information are fed (for example, height values for individual control-points). The information is then coded (for example, by contours) and transmitted in the form of an isarithmic map. Because geographical problems, like those of geology, are characterized by areal sampling restrictions, by a multiplicity of variables, and by the interaction and simultaneous variation of most of the variables (Whitten, 1964), we cannot be certain how much of the information transmitted by a map may be regarded as a ‘signal’ and how much as random variations or ‘noise’.

Areal data always present such ambiguities (for example, those of definition (Chayes and Suzuki, 1963)) and the most obvious manner in which to treat them is to attempt to disentangle the smooth, broader regional patterns of variation from the non-systematic, local and chance variations (Krumbein, 1956); and then to ascribe mechanisms or causes to the different components. Thus the regional effect is commonly viewed as a smooth, regular distribution of effects, termed a trend surface or trend component (Whitten, 1959), which are too deep, too broad, and too great in ‘relief’ to admit of the purely local explanation which is reserved for the residuals or deviations (Nettleton, 1954).

Trend surfaces can thus be considered as response surfaces (Box, 1954; Krumbein, 1956)

47
from which aspects of origin, dynamics or process can be inferred, wherein variations in form may be thought of as responses to corresponding areal variations in the strength and balance of the controlling factors. As will be shown later, on this level trend-surface mapping represents an attempt to build up some generalized picture of areal variability in order to test some process-response model (Whitten, 1963A and 1964), in which an attempt is made to explain distributions in terms of sets of process factors.

Although the discussion throughout this paper is confined to isarithmic maps describing continuous surfaces, this does not mean that trend-surface models are restricted to such conventional isarithmic surfaces as terrain elevation or isobaric pressure. Bunge (1962, p. 46) has suggested that only ‘pure hideboundness’ keeps cartography from applying isarithmic techniques to a wide range of geographical phenomena. Population, like light, may be profitably regarded either as a series of discontinuous quanta or as a continuum. The choice is largely a matter of scale, convention and convenience; and over thirty years ago Jones (1930) urged the more general use of ratio-type maps in studies of land use and agricultural geography. Isarithmic maps represent, in cartographic terms, the highest level of the four-stage measurement sequence—nominal, ordinal, interval and ratio—and the great flexibility and utility of the ratio scale for both mapping and statistical analysis are strong arguments for converting the largest possible amount of appropriate geographical data into this form.

Construction of Trend Surfaces

The information from which trend surfaces may be derived is contained within the original input (that is, control-point data) and such surfaces are ultimately controlled as to their accuracy

Figure 1—Analysis of forest distribution in a 10,000 square kilometre section of the Tagus-Sado basin, central Portugal through (A) contour interpolation, (B) grid generalization, and (C) linear regression. The cross-sections are drawn orthogonal to the dominant regional trend in the area.
FIGURE 2—Schematic table of alternative methods of trend-surface analysis. Details of papers included in this table are given in the list of references at the end of the paper.
by the quality of this original data (Whitten, 1962A). However, trend-surface mapping differs from conventional contour-mapping (Robinson, 1960, pp. 178–94) in the use it makes of the data. Figure 1 shows a series of maps derived from the same control-point data from central Portugal (Haggett, 1965, Chapter 9) with a representative cross-section of control-point values. In conventional isarithmic mapping (Fig. 1A) each control point contributes information only for the area immediately adjacent to it, in the sense that contour lines are drawn between a control-point and its immediate neighbours, but it does not affect values outside the polygon formed by joining the control points which immediately surround it. In the two types of trend-surface mapping illustrated, this restriction is not maintained. In Figure 1B a method of ‘moving means’ has been adopted by which a series of control points is averaged to give a series of new values. Depending on the number of points averaged (that is, analogous to the grid size used), each control point is allowed to contribute information over a wider area. The effect of this, as shown in the two accompanying maps, is (1) to damp down local irregularities to give a clearer picture of the regional trend and (2) to allow the separation of local ‘residuals’. Figure 1C shows the logical development of this information-generalization sequence in which, by using best-fit regression analysis, each control-point contributes information to the whole of the area being mapped, individual local values being sacrificed to the overall trend.

Inspection of the three graphs shows that while, at different levels of sophistication, contour interpolation (Fig. 1A) and regression analysis (Fig. 1C) are relatively objective methods, the derivation of the intermediate type of surface leans heavily on the scale of grid selected. The construction of trend-surface maps from areal data may thus be broadly divided into those which are primarily selective and those which are more, but not exclusively objective in character. This division also broadly separates, as Figure 2 shows, (1) the graphical and the grid methods from (2) the computation methods.

(a) Selective Methods

Nettleton (1954) has discussed the ambiguity of areal data, when one is to identify regional trends, by pointing to the common lack of clear criteria for separating the whole field into its supposed component parts. Added to this any graphical method depends to some extent on subjective decisions on the part of the operator (Skeels, 1947). The first group of the graphical selective methods falls conveniently into the ‘eyeball’ method of drawing smoothed contours (for example, those on the Chiltern Pliocene bench drawn by Wooldridge and Linton (1955, p. 110)) and the various profile methods. The simplest of these is the parallel profile method, in which equally-spaced parallel profiles are constructed, smoothed out by eye, and used as the basis for the production of a three-dimensional map (Nettleton, 1954; Krumbein, 1956). A grid of intersecting profiles, adjusted at their intersections, can also be used to obtain a set of internally-consistent values on a rectangular grid, the vertices of which form the control points for a contoured map. Krumbein (1956) has also described a matrix method for the construction of orthogonal profiles.

The second group of selective methods, the grid/filter type, are largely restricted to centre-point and ring constructions. In the simplest form of these a circle of unit radius is centred on each grid point and the values obtained by interpolation on the circumference are averaged to give the central grid-point value (Griffin, 1949; Krumbein, 1956). More complex multiple-ring methods involving first- and second-derivative maps have also been devised (Elkins, 1951; Henderson and Zietz, 1949; Nettleton, 1954; Peters, 1949; Rosenbach, 1953) and widely applied to geophysical prospecting.
Haggett (1961) has used a variant of the ring method, the *hierarchical quadrat* method, in which a series of square cells (quadrats) of increasing area were drawn around each control point. Figure 3 shows the values obtained for a survey of forest distribution in central Portugal in which 960 control points were established at the vertices of a rectangular grid system with four sizes of quadrat (2500 square kilometres, 625 square kilometres, 156 square kilometres, and 39 square kilometres). Comparison of the four maps indicates the way in which the regional trend is closely related to the size of the grid, the '2500 filter' allowing only broad regional swells to be transmitted, whereas the '39 filter' allows considerably greater detail.

![Figure 3](https://example.com/figure3.png)

**Figure 3**—Analysis of forest distribution in a 15,000 square kilometre section of the Tagus-Sado basin, central Portugal through a hierarchy of quadrat sizes: (A) 2500 square kilometres, (B) 625 square kilometres, (C) 156 square kilometres, and (D) 39 square kilometres. Contours are given at intervals of ten angular units and areas above the mean are shaded (from Haggett, 1961).

*(b) Objective Methods*

The so-called objective methods are also anomalous in a number of respects. Obviously the selection of the number and spacing of the control-point array vitally influences the form of the resulting map, and Merriam and Harbaugh (1964) have demonstrated the effects of areal extent and control-point density for five nested areas in Kansas, varying from the whole state to a locality of about 13 miles square. Even allowing for a high density and even scatter of control points, however, there are two other problems involving selective anomalies—grid spacing, and the mathematical families of surfaces employed.

Where the areal array is in rectangular grid form, Nettleton (1954) has shown that many of the results hinge on their spacing, for the grid acts like an electric filter which will pass components of certain frequencies while excluding others. Indeed, Holloway (1958) views all smoothing attempts as forms of filtering, and points out that on hemispherical meteorological maps a grid-point spacing of 500 miles completely attenuates features on a scale of about 1500 miles, whereas 4000-mile features are retained at about 75 per cent of their original amplitude.
The question of the choice of the mathematical form of fitted surface will be returned to in the subsequent discussion of regression methods.

Included within the more objective mathematical methods of trend-surface fitting, both the use of matrices of expected values (Krumbein, 1956) and the fitting of sine curves (Swartz, 1954; Wiener, 1948) were initially useful. In the expected-values method the mean values of the rows and columns of rectangular arrays of control points are used to compute the expected value of a point in terms of the characteristics of both the row and column in which it lies. These computed values give the regional trend surface, which may be compared with the original values at each intersection to give maps of local residuals. Although the use of sine curves has been largely confined to time series where a linear sequence of values may be 'decomposed' into constituent parts (for example, into the diurnal curves of meteorological observations or the business cycles of econometric observations), it can be extended to areal distributions. Using Fourier series, the isarithmic map may similarly be broken down into a series of waves of varying harmonic characteristics. These methods, however, have been found to be too cumbersome and restrictive and are being replaced.

Of all the methods described, that of fitted regressions is by far the most flexible and promising from the geographical viewpoint. In the same manner as two-dimensional regressions are fitted to graph data, so it is possible to calculate best-fit three-dimensional regressions related to areally-distributed data. Similarly the degree of 'fit', or 'explanation', can be expressed in conventional terms—the percentage reduction in the total sum of squares (Krumbein, 1959) (Fig. 4). The choice of the family of mathematical surfaces to be so fitted rests upon the versatile orthogonal polynomials, which have the special advantage of adaptability to electronic computing. Their value in this respect was first demonstrated some ten years ago for regularly-spaced control points (Simpson, 1954; Oldham and Sutherland, 1955; Brown, 1955) and a

![Fourier Series Diagram](image-url)
definitive paper on their application to trend-surface analysis was later prepared by Grant (1957). *Abbreviated regression* methods employing polynomials were used by Krumbein (1936) and by Miller (1956), the latter fitting first- and second-order polynomials to areal data by a graphic method of successive approximations, as earlier described by Ezekiel (1930).

Although Grant (1957) was mainly concerned with rectangular grid data, he indicated how irregularly-spaced data could be utilized, and more refined methods of constructing trend surfaces from such information have been more recently described (Krumbein, 1959; Whitten, 1962a and 1963a; Harbaugh, 1963). The simplest regression methods involve the fitting of *linear surfaces* to areal data, as has been done by Robinson and Bryson (1957) to rural population and rainfall in Nebraska, by Lippitt (1959) to sedimentary parameters in New York and Southern Ontario, and by Haggett (1964) to forest distribution in south-eastern Brazil. This method of simplified areal description has been carried a stage further into the analytical realm by Robinson and Caroe (In Press) who attempted to test correlations or coincidences between what were considered to be *superimposed surfaces*, and by Haggett (1965, Chapter 9) who mapped successive residuals from linear surfaces to obtain *sequential surfaces* each representing the areal activity of progressively less regionally-significant controls. However, such work brings us into the field of the exploitation of trend surfaces which is the subject of the last part of this paper.

Obviously the descriptive restrictions imposed by having only linear surfaces to fit to complex distributions has, despite their relatively simple computation, required that more complex trend surfaces be fitted involving the use of high-speed electric computers (Krumbein, 1962a and 1962b; Foragton, 1963). For this purpose progressive expansions of linear, quadratic and cubic polynomial functions have been used, of the general form:

\[ Z = a + bU + cV + dU^2 + eUV + fV^2 + gU^3 + hU^2V + iUV^2 + jV^3 \]

where \( Z \) is the areally-distributed variable, and \( U \) and \( V \) are the locational rectangular coordinates. The general forms of these surfaces are indicated by Figure 4, although considerable variation is possible within this versatile family (for example, Fig. 15) and, indeed, attempts are being made to introduce higher degree functions into geological work (Baird, et al., 1963; McIntyre, 1963; Conner and Miesch, 1964). Broadly, it is possible to separate those computer programmes which give a *control-point output* only, from those which also supply some form of *printed contour map*, although mathematically there is no difference between them. In the former group is the FORTRAN programme prepared for the IBM 7090 (Whitten, 1963a; see also Peikert, 1963) which (1) computes linear plus quadratic plus cubic polynomial surfaces by least squares, (2) gives the reduction in sums of squares achieved by each surface, and (3) produces an output matrix of up to 35 x 35 point values for each surface, together with the residuals. In the latter group are the FORTRAN II programme for the IBM 1620 (Good, 1964) and the more elaborate BAGOL programme for the IBM 7090 (Harbaugh, 1963). This latter produces linear plus quadratic plus cubic polynomial surfaces both as equations and contour maps, residuals for each surface, and certain properties of the trend and residual surfaces (for example, an error measure, reductions in the sums of squares, the volume beneath the surface, spatially weighted averages, and the arithmetic mean). The computations for a programme involving 200 data points (excluding the printing of the map) takes some 48 seconds at a computer cost of $3, and is approximately equivalent to about 100 hours of work by a skilled desk calculator operator. An example of a printed map, which commonly measures up to 60 square inches, is given in Figure 5.
Information Output from Trend-Surface Analysis

Figure 6 exemplifies a more formal analysis of the information output from the trend-surface programme which, for both regional surfaces and residuals falls broadly into descriptive and interpretative information.

Descriptive information relating to regional trend surfaces includes details of their morphometry and accuracy in relation to the original data. For linear surfaces, morphometry includes the equation giving the dip and strike of the plane and, for higher order surfaces, the equation and volume beneath, together with some spatially-averaged values. Examples of the value of generalizations obtained by the use of regional trend-surface analysis in terms of the composition of granite massifs have been given by Whitten (1961A, 1961B and 1962B). In addition to the morphometry, however, it is necessary to relate these best-fit trend surfaces to the original data from which they were derived. Besides the reduction in the total sum of squares effected by the surface, some indication of confidence limits (Krumbein, 1963) is required. It is necessary to inquire whether the fitted surface is a valid expression of large-scale variations in the mapped variable, or whether it may have arisen purely by chance from the
sample of map points used, and to what extent the residuals may 'contain locally important signals'. As Krumbein (1963) has stressed, mappable data contain variations at several geographic levels. Just as tests of significance of trend are applied to two-dimensional best-fit regressions, so Krumbein (1963) has described a three-dimensional test of significance of trend for fitted surfaces, giving levels of confidence that the improvement of fit offered by the regional surface is due to a real trend and not to chance alone (Fig. 7). Even where the fit of the surface (that is, the reduction in the sum of squares effected by it) is poor, the form of the bounding confidence surfaces may demonstrate that the trend of the surface is 'real'—in other words 'having some flexibility in its angle and direction of dip but without sufficient freedom so that its direction of inclination could be completely reversed' (Krumbein, 1963, p. 5874).

The interpretation of regional trend surfaces resolves itself into three main considerations—the major trend, discrete trends and interlocking trends. Little more need be said here regarding the obvious advantages of identifying the sort of major regional trends referred to above, but the other two types of interpretation also have considerable geographical potential. Figure 8 shows the general isopleths of the colour index for the Lacorne granite massif in Quebec, to which single trend surfaces of low explanation have been fitted (Dawson and Whitten, 1962; Whitten, 1963A). When the region is divided into two areas, however, two sets of concentric surfaces of much higher explanation appear, supporting the idea of discrete multiple granitic intrusion. It is readily apparent that such a test involves the same reasoning as the covariance analysis commonly applied to two-dimensional regressions. Allen and Krumbein (1962) have also used trend-surface analysis in an attempt to disentangle interlocking trends of facies components in the Top Ashdown Pebble Bed of the Weald (Fig. 9). A surface fitted to the garnet

---

**Figure 7**—Trend-surface analysis and confidence intervals applied to the zircon size index (in microns) of the Top Ashdown pebble bed, south-eastern England (from Allen and Krumbein (1962) and Krumbein (1963)). (A) Sampling points and generalized isopleths (10 micron intervals). The cross indicates the areal centre of gravity of the sampling points. (B) Best-fit linear surface fitted to the data, effecting a reduction in the sum of squares of 27.1 per cent. (C) Deviations (in microns) from the best-fit linear surface, with positive deviations stippled. (D) The upper 95 per cent confidence surface. (E) The lower 95 per cent confidence surface. (F) Cross-sections of the linear surface (X comp.) and of the upper (hicon) and lower (locon) 95 per cent confidence surfaces along the line AB (Fig. 7B). The form of the bounding confidence limits indicates that, although the linear surface has considerable flexibility, it is very improbable that the indicated trend could be completely reversed.
FIGURE 8—Trend-surface analysis applied to the composition of the Lacorne granitic massif, Quebec (from Whitten, 1963A). 

(A) Isopleths applied to colour index data. (B) Best-fit linear surface fitted to the data in (A). (C) Best-fit quadratic surface fitted to the data in (A). (D) Best-fit cubic surface fitted to the data in (A). (E) Two areally-separated best-fit cubic surfaces fitted to the data in (A). These give marked improvements in the sums of squares so effected (SS), and suggest two separate circular intrusions. (F) Two similarly areally-separated best-fit cubic surfaces fitted to quartz percentages of the Lacorne granitic massif, supporting the multiple-intrusion hypothesis suggested by figure 8E.

content of the bed (sum of squares reduction 21.8 per cent) suggests a deltaic source in the northeast, with the residuals appearing as proto-channels in the delta front, whereas other regional components indicate a second interlocking and contemporaneous deltaic source in the northwest.

It is however, to the interpretation of residuals that most effort has been directed. Simpson (1954) discussed the general construction of residual maps, and Swartz (1954) considered some of their geometrical properties. Thomas (1960) has examined the geographical implications of...
Figure 9—Cubic trend surface fitted to garnet percentages sampled in the area of south-eastern England shown in Figure 7 (reduction in sum of squares 21.8 per cent), together with deviations from it. The former suggests a garnet-rich north-eastern source and the latter proto-channels in the delta front. Analysis of the garnet component enables the separation of a medium-grained, arenaceous, garnet-rich north-eastern source from a north-western source of chert pebbles with silt and clay (from ALLEN and KRUMBEIN, 1962).

Figure 10—The geological structure of Kansas illustrated by the trend-surface analysis of the top of the Arbuckle Group (from MERRIAM and HARBAUGH, 1964). (A) Isopleths drawn on the top of the Arbuckle Group. (B) Quadratic trend surface fitted to the elevations of the top of the Arbuckle Group. (C) Deviations from the trend surface shown in (B). (D) The major structural features of Kansas deriving both from (B) and (C).
maps of residuals from conventional regression analysis, but their value in outlining oil structures has been the primary stimulus for their development. Merriam and Harbaugh (1964) and Merriam (1964) have, for example, depicted the more local structures in Kansas by means of residuals from the Arbuckle Group regional trend (Fig. 10) and have shown how the Lost Springs oilfield is associated with a gentle local flattening or structural terrace on a westward regional dip (Fig. 11). Another productive field wherein interpretation has been assisted by means of the identification of residuals is in the character of granitic intrusives; and Whitten (1959, 1960, 1961A, 1963A, and 1963B) has shown how deviations in the composition of the Thorr granite of Donegal appear as a 'palimpsestic ghost stratigraphy' probably related to Precambrian meta-sedimentary rocks preserved throughout this granite in their 'pregranite positions'. This gives some evidence regarding the manner of formation of the mass (Fig. 12). Chorley (1964)
Figure 12—Trend-surface analysis and deviations applied to the microcline/plagioclase ratios of the ‘older granite’ of the Thorr district, Co. Donegal (Whitten, 1961 A). (A) Best-fit quadratic trend surface. (B) Positive deviations (0.0–0.2 isopleths; 0.2–0.3 shaded; greater than 0.3 black) from the quadratic trend surface, suggesting a ‘ghost stratigraphy’ of metasedimentary relics.

Figure 13—Trend-surface analysis and deviations applied to median grain sizes (mm) of the surface soils of the Lower Greensand outcrop between Ely (E) and Leighton Buzzard (LB) east-central England (from Chorley, 1964). (A) Best-fit cubic trend surface (SS 38.03 per cent). (B) Negative deviations from the cubic trend surface (stippled). Dots indicate soil sampling points. Solid dots show where transported flints were encountered, suggesting that glacial drift admixtures might be generally associated with the localities where abnormally fine-grained contaminations occur, so confusing the regional soil grain-size pattern which is broadly related to the bedrock facies of the Lower Greensand.

has also used this residual technique to distinguish possible glacial admixtures from the in situ Lower Greensand soil facies of east-central England (Fig. 13); while Haggett (1964) has attempted an explanation of the distribution of residuals of forest distribution in south-eastern Brazil. It is thus readily apparent that both the geometry and the areal associations of residuals often give valuable clues as to their controlling factors and mechanisms.
Haggett (1965) has suggested that there may be hierarchies both of regional components and of different orders of residuals present in any single pattern of areal distribution. However, there always remains, even after the most complete 'areal filtering' a randomly-distributed, unexplained component commonly termed 'noise'. Noise itself emanates from a number of different sources, such as errors of identification and measurement, but in most geographical distributions there is a real random element the existence of which must be recognized. What might be thought of as noise on one level, however, may be subsequently identified as partly-explicable local components on another level. The high proportion of unexplained residuals (78.49 per cent) from the third-degree regional surface fitted to the distribution of Breckland surface-sand size facies (Fig. 14) contains an explainable local variation, due perhaps to periglacial sorting processes. (Chorley, Stoddart, Haggett and Slaymaker; in preparation).

![Figure 14. Trend-surface analysis and deviations applied to median grain sizes (mm) of the surface sands of the Breckland, eastern England (Chorley, Stoddart, Haggett and Slaymaker; in preparation). (A) Rectangular grid (2 km. intervals) of the 149 sampling points. (B) Best-fit cubic trend surface. (C) Deviations from the cubic trend surface (positive deviations stippled), indicating a high local 'noise level'.](image)

**Further Exploitation of Trend-Surface Analysis**

Although there have been, as yet, comparatively few applications of trend-surface analysis to conventional geographical problems, the foregoing discussion suggests a number of possible utilizations. These may be summarized under four headings: (i) improved isarithmic mapping; (ii) simplified description and comparison of complex geographical patterns; (iii) comparative areal analysis and (iv) construction of process-response models.

(i) The speed and objectivity of contour-type mapping by computers has already led to its widespread use in meteorology and geophysics, and its adaptation to geography would seem to be only a matter of time. Like the cartographer, the computer carries out simple linear interpolation between control points, and is subject, in certain cases, to the same indeterminacy to which Mackay (1953) has drawn attention. This problem of indeterminacy may be overcome by re-arranging the pattern of control points into a triangular lattice which has the advantage over square systems in that '... linear interpolation is unambiguous, in extending isarithms, since only one single plane can go through three points in space' (Nordbeck, 1964,
Nordbeck has derived a computer programme (NORKI) which transforms the square network into a triangular one. In the NORKI programme, however, there is some loss of detail in that it takes four control points arranged in a square to give one control point in the triangular system. Ojakangas and Basham (1964) have overcome this problem with their INGRID programme written in FORTRAN for an IBM 7090 computer. In this programme the input consists of a series of control points (up to a maximum of 4000) given in terms of their $UV$ co-ordinates (for example, National Grid reference numbers) and $Z$ values (for example, elevation). Since the control-points may be irregularly distributed over the area, the first stage of the INGRID programme is to interpolate values for the grid intersections of any specified grid-size to give a square network of control points. This stage is followed by contouring in which a quadratic surface of the form:

$$Z = a_0 + a_1U + a_2V + a_3UV$$

is fitted to the four corners of each grid square: the $Z$ value of each unit cell within the square is then computed from this surface and printed out as a contour map on an IBM 1403 printer. By using the quadratic surface, Ojakangas and Basham have outflanked Mackay's indeterminacy problem and ensured consistency of contours over the whole map area. The programme was tested on geophysical data over the west Los Angeles basin, California with 200 original control-points, and compared with the hand-contoured U.S. Geological Survey maps of the same area. The number of contours in the computer maps was limited by the number of symbols available (thirty-eight) and by the size of the print-out unit; so that some contour lines were lost in high-gradient areas where contours were closer together than

---

**Figure 15**—Four basic patterns generated by quadratic equations: (A) maximum, (B) stationary ridge, (C) rising ridge, and (D) minimax (from Davies, 1956).
the printed characters allowed. Moreover, all lines were plotted as continuous contours while the hand-contoured map showed breaks, for example, along the Palos Verdes fault-line. Against this, the speed of the system (the whole computation from the input of original data to the final maps took 255 seconds), the relatively low cost ($25.00), and the greater accuracy suggest that isarithmic mapping with contour interpolation by quadratic trend surfaces might be of considerable use in the rapid first-stage investigation of trends in an area.

As in geophysics, so in geography there are many fields of investigation in which the distribution of control points is 'spotty' in character, and the individual Z values may be of very variable accuracy (Blumenstock, 1953). For example, in historical geography Henderson (1952) has shown how the 1801 crop returns are abundant in some areas of England and scarce in others. In such instances, there is much to be said for fitting regression surfaces to the whole distribution. By this device the errors of individual control-points are distributed over the whole map rather than violently distorting a small area, and regional trends of a given level of significance may be culled from the data.

(ii) Progressive reduction of complex isarithmic maps to a series of simpler components is being often carried out at present to assist both teaching and research. Thus we may talk of 'ridges' of high population density in the north-eastern United States or of 'gradients' in agricultural intensity in the Middle West. Trend surfaces lend themselves readily to the relatively simple description of regional characteristics. Thus, using linear surfaces, Haggett (1964) was able to describe the distribution of forest in south-eastern Brazil in terms of a simple 'shed-roof' gradient dipping inland at right-angles to the coastline. Such a description might be readily compared in terms of dip and strike with environmental gradients in the same areas, or with forest gradients in other areas. Using scarcely more complex quadratic surfaces, Davies (1956, p. 523) showed the four main types of system that might be generated by such second-degree equations, viz. a maximum, stationary ridge, a rising ridge and a minimax (Fig. 15). Since each of these surfaces may be inverted (the maximum to a minimum, the ridge to a trough, and so on) there are eight basic forms which may be identified and compared. Even though any single quadratic trend map may show only part of one of the basic eight forms, some useful yardstick for describing geographical patterns is clearly provided.

Again trend surfaces may be used in research to test prevailing descriptions of physiographic and other forms. Thus Svenson (1956) suggested that the application of trend-surface techniques to the mapping and separation of possible erosion surfaces should be further explored. In view of the wide range of alternative forms which have been recognized by different workers in areas like Wales, it would be profitable to subject such disputed areas to successively more complex trend-surface analyses.

Description of distributions leads inevitably to a recognition of the anomalous problem areas which depart from the overall trend. While in the case of first-order analysis we might expect such anomalies to be widespread, local pockets which persist after higher-order analysis may suggest the location of more significant anomalies. Thus the trend surface may act, by its filtering effect, so as to identify or 'sieve out' areas for special study.

(iii) Trend-surface analysis has already been used as an adjunct to forms of areal multivariate analysis (Krumbein and Imbrie, 1963), in attempts to apportion out 'explanation for residuals' (Thomas, 1960; Robinson, Lindberg and Brinkman, 1961; Haggett, 1964) and to explain the regional variability of one areal feature by means of corresponding areal variations in assumed controlling factors. All these methods, however, simply use the trend surfaces to generate data which are treated by the conventional multivariate techniques; and there is
need for the development of direct methods by which trend surfaces themselves can be correlated and the features of one surface explained directly in terms of others. First steps in this direction have been provided by Robinson and Caroe (In Press) who developed a correlation technique between linear trend surfaces based upon a comparison of vectors, and by Miller
(1964) who has used certain properties of the best-fit polynomial surfaces in the comparison of contour maps.

(iv) Finally, it is possible to view the formalized description afforded by trend-surface analysis not merely as an aid to the comparison of different areal patterns, but also to adapt the response-surface concept in the building of process-response models. These represent actual or conceptual frameworks with reference to which information is related, or 'structured', as an aid to generalization in an attempt to explain areal distribution in terms of sets of process factors, and perhaps as a basis for prediction (Krumbein and Sloss, 1963; Whitten, 1964). These models attempt to embrace the broad outlines of the interrelationships of areal-process factors (Krumbein, 1964), and have been constructed for the analysis of beach phenomena (Miller and Ziegler, 1958; Krumbein, 1964), stratigraphic units (Krumbein, 1962A; Allen and Krumbein, 1962 Sloss, 1962), granite plutons (Dawson and Whitten, 1962; Whitten, 1963A), and the earth's mantle (Ringwood, 1962).

Figure 16 shows two such process-response models evolved on the basis of trend-surface analysis. The first, developed by Allen and Krumbein (1962), shows the general sedimentary processes which may explain the stratigraphic sequence revealed by trend-surface analyses of the Top Ashdown Pebble Bed, south-eastern England. The second, developed by Haggett, shows the general settlement processes of Portuguese colonization in south-eastern Brazil as a function of a simple linear surface dipping inland and of a quadratic 'minimax' form of surface. In this quadratic case there appear to be two main source areas (the Tiête basin around São Paulo and the Baixada Fluminense around Rio de Janeiro) with negative areas between and to the north. Although the two models are elementary, they suggest possible diffusion processes (respectively of sediments and population) that can be further tested by conventional statistical analysis.

Conclusion

Striking developments in trend-surface mapping have been made in disciplines like meteorology and geophysics, where the demand for generalization, interpolation and prediction have provided major financial incentives for research and development, and where the establishment of quantification and rapid computational aids have provided a receptive environment for experimentation. This paper has reviewed some of the more successful experiments that have been made in trend-surface mapping which, with their concern for regional trends and local anomalies, promise to throw new light on long-standing geographical problems both of teaching and research. Both very simple graphical or grid techniques and also very complex computer programmes are now available and there seems no reason, other than convention and lethargy, why they should not be very widely adapted for use in all branches of geography, both physical and human, in the immediate future.

REFERENCES


Davies, O. L. (Ed.) (1956) *The Design and Analysis of Industrial Experiments*.


Haggett, P. (1965) *Locational Analysis in Human Geography*.


This content downloaded from 35.8.15.185 on Sun, 6 Oct 2013 12:22:41 PM
All use subject to JSTOR Terms and Conditions
TREND–SURFACE MAPPING


ROSEN BACH, O. (1953) 'A contribution to the computation of the “second derivative” from gravity data', Geophys., 18, 894–912.

SAYYER, J. S. (1960) 'Graphical output from computers and the production of numerically forecast or analysed synoptic charts', met. Mag., 89, 187–90.


