TREND SURFACE ANALYSIS — A REVIEW

A. EASTON WREN*

"... while computers are not necessarily labour-saving devices in geology, they can be regarded as "intelligence amplifiers" which can have great potential in extending an individual's thought processes — Harbaugh and Merriam, 1968".

INTRODUCTION

Earth scientists have always been interested in the description and analysis of surfaces, particularly those which serve to describe mapped data. The methods used in this context fall into two categories, written descriptions (using terms such as dip, strike, plunge, isoclinal folding, attitude of axial plane, foliation etc.) and mathematical approximations. To overcome the inadequacies of verbal descriptions, earth scientists have turned to the more consistent and objective techniques of numerical approximation. The transition to the mathematical approach was essentially synchronous with the advent of fairly sophisticated computer programs which made practical the high-speed computation of large matrices and the generation of complex functions.

It is the purpose of this paper to describe, in some detail, the technique of trend surface analysis, its advantages and limitations, together with a reference to more recent developments using Fourier analysis.

TREND SURFACE ANALYSIS

Since 1955, a vast literature on the application of mathematical surfaces to the interpretation of data in the geological sciences has appeared. Their main use has been to simplify and analyze data so that "trends" which were not particularly obvious in the data have been revealed, and to employ such trends in correlation and interpolation. Trend analysis deals with the recognition, isolation and measurement of trends that can be represented by lines, surfaces or "hypersurfaces". It should be emphasized however, that trend analysis as practiced by geologists constitutes only a segment of the larger statistical field of regression analysis. In this context, many types of functions other than integer power series have been used in regression and could probably be used with advantage in geology. Trend analysis is now commonly referred to as a mathematical method of analyzing data in map form. The technique is designed to separate an observed contour map surface into two components: the REGIONAL component expresses the large scale effects or trend influencing the entire map area; the RESIDUAL component is the difference between the observed and regional values and expresses local effects or anomalies which influence only parts of the map area. It must be stressed that the terms REGIONAL and RESIDUAL are arbitrary and are used only for convenience of interpretation. The technique of "surface-fitting" is often confused with trend analysis and it may be appropriate at this point to emphasize the differences between them. Trend surface analysis is an analytical technique designed to separate observed data into regional and residual components. The regional component is obtained by fitting a low order surface to the data using techniques such as least squares. Hence, one of the basic assumptions in trend analysis is that the regional can be reasonably approximated by a smooth con-

*Amoco Canada Petroleum Company Ltd., Calgary, Alberta.

tinuous surface, a surface which can be expressed by a simple mathematical function called a polynomial. However, trend surface analysis cannot decide which low order surface constitutes the trend. This decision must be made by the interpreter.

Grant (1957, p. 310) has defined trend as being the polynomial of "best fit" to the data and has stated that the success of the method depends on the validity of the assumption that any trend in the data can be adequately described by a low order polynomial. He also stated that with most geological data this assumption should rarely fail.

The Use of Polynomials in Trend Analysis

We can use 2-variable polynomial trend analysis (or curve-fitting) where a straight line or curve is fitted to data points, adhering to least squares criteria. This approach may be applied in the objective description of trends in vertical variations in sedimentary successions, well logs or profiles such as elevation or potential field values. Trend analysis involving 3-variable polynomials is used in the fitting of surfaces to mapped data, the type of data being of no consequence. 4-variable polynomial analysis (or space-fitting) involving "hypersurfaces" can be used in the description of a parameter in 3-D space; e.g. porosity variations in a rectangular block of rock. Here we could regard porosity as the dependent variable and the spatial coordinates of a point within the block as the independent variables.

Reference has already been made to the mathematical functions known as polynomials and indications have been given of their usefulness in the field of trend analysis. Grant has stated that ". . . polynomials with their great flexibility and computational advantages are among the best of mathematical tools for analyzing geological data." One of their major advantages is the ease with which they can be adapted to the computer as an iterative process.

A polynomial of order n in the variable x is a linear combination of the n + 1 terms, 1, x, x^2, x^3, . . . x^{n-1}, x^n. If the coefficients are denoted by a_0 a_1, a_2, . . . a_{n-1}, a_n, and the polynominal by f(x) then

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \dots a_{n-1}x^{n-1} + a_nx^n.$$

The order of a polynomial expression is the highest degree of the terms making up the expression. The degree of a term is the sum of the exponents of its variables. For example, the function

$$y = 3x^3 - 2x^2z + 4xz - 7$$

is of order 3 and the degree of its terms, taken in order, are 3, 3, 2 and zero.

A convenient method of illustrating the basic concept of fitting polynomials is given below. Here the discussion is restricted to curve-fitting, i.e. the 2-dimensional case in the x-y plane, and in general terms, y = f(x).

1(a) is the case of a straight line graph, where $y = ax + b$. This is the linear case and the order of the function is one.

1(b) is a curve of the form $y = ax^2 + bx + c$. This is the quadratic case and the order of the function is 2.

1(c) is a curve of the form $y = ax^3 + bx^2 + cx + d$. This is the cubic case and the order of the function is 3.

The standard and well-known straight-line relationship, $y = ax + b$ may also be written as $f(x) = ax + b$ since y is a function dependent on the value of x. The equation may therefore be considered as a profile in the x-y plane . If the concept is now extended by one dimension into three-dimensional space, then the z-axis is introduced to represent the value of the dependent variable at any point defined in the (x, y) plane. There are now 3 planes, the x-y which is horizontal and the x-z and y-z planes which are both vertical but at right angles to each other. We can write the general equation of these planes as

$$Ax + By + Cz + D = 0.$$

We can now establish a plane which best represents some set of data having the value z as a function of x and y. The data may be described as f(x, y). Since the coefficients A, B, C, D may be of any algebraic sign and value, the equation of any plane may be written as

$f(x,y) = Ax + By + D$ (This is linear, i.e. of first order.)

This can be extended through the quadratic or second order case giving

$$f(x,y) = (A + Bx + Cy) + (Dx^2 + Ey^2 + Fxy)$$

and the general representation is

$$f(x,y) = a_0 + a_1x + a_2y + a_3x^2 + a_4y^2 + a_5xy + a_6x^3 + a_7x^2y, \ldots$$

The foregoing discusion has been limited to the two-dimensional representation of a surface where the three coordinate axes are used to depict the two independent variables x, y and the dependent variable z or f(x, y). It is possible to extend this notation into three-dimensional space where the three independent variables x, y, z represent position and the value of f(x, y, z) is to be determined. The general equation of such a notation is of the form

$$\begin{aligned} f(x,y,z) = {} & a_0 + a_1x + a_2y + a_3z + a_4x^2 + a_5xy + a_6xz + a_7y^2 + \\ & a_8yz + a_9z^2 + a_{10}x^3 + a_{11}x^2y + a_{12}x^2z + a_{13}xy^2 + a_{14}xz^2 + \\ & a_{15}xyz + a_{16}y^3 + a_{17}y^2 + a_{18}yz^2 + a_{19}z^3 + \ldots\ldots\ldots \end{aligned}$$

Least Squares Surface of Best Fit: Normal Equations

A brief review of the method of least squares is now appropriate. The method of least squares is a method of curve fitting suggested early in the 19th Century by the French mathematician, Adrien Legendre. If we have a point distribution on a graph that do not all fall on a straight line, we might say that a straight line provides a reasonably good fit, and the overall pattern sugggests a linear regression. If we decide that the straight line gives a good approximation, we are then faced with determining which linear equation provides the best possible fit; i.e. estimating the regression coefficients, a and b, in $y = a + bx$. We use the criterion of least squares for deciding the best fitting line. Computed regional surfaces satisfying the "least-squares" criteria are positioned in such a way that the sum of the squares of the distances of each control point from the surface is a minimum. Only one surface of each degree has this property; i.e. the optimum representation makes the function,

$$f(a_o, a_1, a_2, \ldots. a_n) = R_1^2 + R_2^2 + R_3^2 + \ldots. R_m^2 \qquad \ldots.1$$

a minimum

i.e. to make $\sum_{k=1}^{m} R_k^2$ a minimum

This happens when the first derivatives are zero, i.e.:

$$\frac{\partial f}{\partial a_i} = 2\sum_{k=1}^{m} R_k \cdot \frac{\partial R_k}{\partial a_i} = 0 \qquad \ldots.2$$

where $i = 0, 1, 2, \ldots. n$

$$\text{Now } \frac{\partial R_k}{\partial a_i} = X_k^i \qquad \ldots.3$$

Therefore equations (2) become:

$$\sum_{k=1}^{m} R_k X_k^i = 0 \qquad \ldots.4$$

where $i = 0, 1, 2, \ldots. n$

Substituting for R_k, the difference between observed y and the polynomial values of y,

$$R_k = \left[\sum_{i=0}^{n} a_i X_k^i\right] - y_k \qquad \ldots.5$$

where $k = 1, 2, 3, \ldots. m$

$$\text{Then } \sum_{k=1}^{m} X_k^i \left[\left(\sum_{i=0}^{n} a_i X_k^i\right) - y_k\right] = 0 \qquad \ldots.6$$

where $i = 0, 1, 2, \ldots. n$

Equations (6) are known as the normal equations. All the summations can be determined so that there is a system of $n + 1$ linear equations in the $n + 1$ unknowns a_i $(i = 0, 1, 2 \ldots n)$. Solution of the $n + 1$ linear equations yields the coefficients a_i so that the polynomial

$$y = a_0 + a_1x + a_2x^2 + \ldots a_nx^n$$

can be determined. The criterion of least squares demands that

$$\Sigma\left(X_{obs} - X_{trend}\right)^2 = \text{minimum}$$

Because we are dealing with a regression function in which we seek the best estimate of one variable with respect to another, either y with respect to x or x with respect to y, it is important to decide on the dependent variable.

Stastical Measures "Goodness of Fit"

The computation of the coefficients of a trend function is only part of trend analysis. In addition, it is necessary to compute meaures of the goodness of fit of the function to the data and to determine if the function components are statistically significant. One measure is the amount of deviation of a variable z from its mean value $(\bar{z})$, which is an index of the total variation within the entire data set. This is calculated by summing the squares of deviations from the mean. The simplest expression is

$$S = \Sigma\left(z_{obs} - \bar{z}_{obs}\right)^2$$

where S = total corrected sum of squares of deviations from mean and

z_{obs} = observed value of variable Z

$\bar{z}_{obs}$ = arithmetic mean of observed values of Z

or, more simply,

$$S = \Sigma z_{obs}^2 - \frac{(\Sigma z_{obs})^2}{n}$$

where n = number of data points

If we consider S as being comprised of 2 components, viz. S_t (that contributed by the trend) and S_d (that contributed by deviations from the trend), i.e. $S = S_t + S_d$ assuming that least squares criteria have been satisfied and that the coefficients of the trend functions are linear. The sum of squares contributed by the trend function S_t represents the squared difference between the predicted (or trend) values of z from the mean value of z,

$$S_t = \Sigma\left(z_{trend} - \bar{z}_{obs}\right)^2$$

The sum of squares due to deviations is a reflection of the failure of the trend values to coincide with observed values.

$$S_d = \Sigma\left(z_{obs} - z_{trend}\right)^2$$

In assessing the goodness of fit it is convenient to express that part of the total sums of squares accounted for by the trend function as a percentage. This permits ready comparison of trend functions fitted to different sets of data (Harbaugh and Merriam, 1968, p. 67). Another convenient measure is the percent of total sum of squares,

$$100\left[1 - \frac{S_d}{S}\right]$$

or equivalently,

$$100\left[1 - \frac{\Sigma(z_{obs} - z_{trend})^2}{\Sigma(z_{obs} - \bar{z}_{obs})^2}\right]$$

A value of 100% indicates a perfect fit of a function to the data, with zero deviations. This is uncommon, and a perfect fit will only be obtained if the number of terms in the fitted function equals the number of data points. There is little point in fitting surfaces of this magnitude to data sets. If the goodness of fit is low, say around 15% to 30%, then most of the variation in the data is not represented by the function, although the residuals may still be geologically significant.

Other statistical approaches may be applied as goodness of fit indicators; for example, analysis of variance and the setting up of confidence levels about fitted lines and surfaces. These are not described in this paper but are readily accessible in the vast library of publications on statistics.

Evaluation of Trend Analysis Results

The first step in trend analysis should be the setting up of a geological model and to state the objectives of the analysis. Failure to do this properly is the chief cause for much of the confusion, misconception and frequent abuse of trend analysis. The subsequent evaluation of trend analysis results is of equal significance. Any evaluation requires the acceptance of a particular model, or its rejection. Such decisions form the basis for all subsequent geological interpretation. For proper evaluation, the interpreter must understand the geological significance of the regional model used, of the residual anomalies and of the validity of the applied mathematical model. The regional and residual surfaces obtained must be regarded in the light of other types of information available.

Recent Developments in Trend Analysis

Over the last few years many well-conceived methods have been presented to determine the order of equation necessary to describe major elements of a surface. Despite impressive efforts in trend analysis there is growing uneasiness as to the appropriate interpretation of "trend", particularly when surfaces of high order are generated. For example, it is difficult to rationalize a 5th order residual with geololical reality even though it may be mathematically meaningful. The basic problem is distinguishing between "trend" and "random components". We can call these two elements "signal" and "noise". If one wishes to extract one from the other, one must know, a priori, the inherent nature of either the signal or noise. This a priori insight does not seem to be available for most geological surfaces (Esler and Preston, 1967, p. 1). Zurflueh (1967) has outlined some of the problems inherent in trend surface analysis techniques and has compared the use of polynomials with the application of Fourier analysis. Fourier series has recently come into use as an alternative model to the polynomial for trend surface analysis.

In geology, many processes are oscillatory in nature e.g. folded rocks, dendritic drainage patterns in badlands, varve laminae in lake sediments and variations in the earth's magnetic field. Harmonic analysis can be used to both analyze and synthesize such phenomena. In its objectives, harmonic trend analysis is very similar to polynomial trend analysis. One of its objectives is the representation of phenomena by periodic functions. A second objective is to separate periodic oscillations from random components; i.e. signal from noise. If the Fourier approach provides a good representation of the data then it is probable that the variation in the data is due to periodic causes. If the representation by Fourier series is not good, then other functions such as polynomials or exponentials may be more appropriate.

Conclusions

Subjectively constructed maps and subjective conclusions about mapped data are common in geology (Whitten, 1963). In many fields such subjective evaluations are no longer adequate, though objective evaluation is not always easy. Trend analysis is one approach to statistical geology that permits objective evaluation although it may be criticized for having a somewhat subjective basis. Also it should not be considered as a geological panacea and it is important to hold the limitations of this valid and potentially informative method of data summary in perspective.

Trend Surfaces are therefore simplified representations of "actual" surfaces; they have the advantage of being parametric, i.e. expressable as a mathematical function [f (x, y)] as discussed before. Such surfaces "fit" the observed data with varying degrees of accuracy and precision which can and should be assessed statistically. Since there exists a functionally determinable value of z for every x and y point on the map the calculated surface can serve as a device for interpolating between the data points but only with the greatest caution since the form of the surface is constrained by the algebraic nature of its characterizing function. Trends and gradients in noisy data may be highlighted by the use of lower order (simple) trend surfaces but this usage can create problems of its own. There is a certain paradox involved in fitting noisy data because if the data are noisy a simple polynomial surface will have to be a weak approximation of it, which automatically inflates the likelihod of highlighting meaningless, accidental trends. Also, the dependency of surface fit on density and location of data points has by knowledgeable authors been suggested. (E. C. Dahlberg, personal communication).

References

Esler, J. E. and Preston, F. W., 1967. Fortran IV Program for the GE625 to compute the power spectrum of geological surfaces. Computer Contrib. 16, State Geol. Surv. and Univ. of Kansas.

Grant, F. A., 1967. A problem in the analysis of geophysical data. Geophysics, 22, 309-344.

Harbaugh, J. W. and Merriam, D. F., 1968. Computer Applications in Stratigraphic Analysis. John Wylie, New York.

Whitten, E. H., 1963. A reply to Chayes and Suzuki. J. Petrology, 4, 313-316.

Zurflueh, E. G., 1967. Applications of two dimensional linear wavelength filtering. Geophysics, 32, 1015-1040.

Recommended Reading

Fox, W. T., 1967. Fortran IV Program for vector trend analysis of directional data. Computer Contrib. 11, State Geol. Surv. and Univ. of Kansas.

James, W. R., 1966. Fortran IV Program for double Fourier series for surface fitting of irregularly spaced data. Computer Contrib. 5, State Geol. Surv. and Univ. of Kansas.

Krumbein, W. C., 1956. Regional and local components in facies maps. Bull. Am. Assoc. Petrol. Geologists, 40, 2163-2194.

——, 1959. Trend surface analysis of contour-type maps with irregular control-point spacing. Jour. Geophys. Res., 64, 823-834.

Mandelbaum, H., 1963. Statistical and geological implications of trend mapping with non-orthogonal polynomials. Jour. Geophys. Res., 68, 505-519.

Miller, R. L., 1956. Trend surfaces: their application to analysis and description of environments of sedimentation. J. Geol., 64, 425-446.

——, 1963. Comparison · analysis of trend maps. From Computers in the Mineral Industries, Part 11. Stanford Univ. Pub. in Geol. sciences., 9, 669-684.

Oldham, C. H. G. and Sutherland, D. B., 1955. Orthogonal polynomials: their use in estimating the regional effect. Geophysics, 20, 295-306.