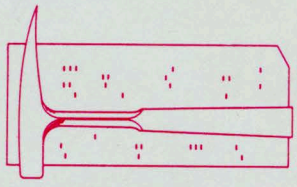
SYMPOSIUM ON COMPUTER APPLICATIONS IN PETROLEUM EXPLORATION

Edited By

DANIEL F. MERRIAM



COMPUTER CONTRIBUTION 40

State Geological Survey
The University of Kansas, Lawrence
1969

in cooperation with the
American Association of Petroleum Geologists
Tulsa, Oklahoma

Editor's Remarks

The papers included in this Symposium were presented at a session of the annual meeting of the American Association of Petroleum Geologists in Dallas, Texas, on 15 April 1969. The purpose of the Symposium was best stated in the preliminary announcement:

> A group of noted earth scientists will discuss aspects of computer applications useful in petroleum exploration and exploitation. All speakers are or have been closely associated with the petroleum industry; they will consider new concepts and approaches which may be used in the search for hydrocarbons. Several outstanding scientists will discuss with the speakers the future of computers in geology and petroleum geology in particular. The program is designed for maximum benefit to practicing petroleum geologists.

Computer applications in the earth sciences now play an important part in petroleum exploration and exploitation. This is true in both data processing and the utilization of problem-solving techniques. These papers extol some techniques which, when applied, may offer exceptional promise.

The authors represent different aspects of industry and academia, and represent a major company, geophysical contract company, consultant, University, Survey and an editor. Each has a different background and experience in computer applications. All have had considerable experience in the petroleum industry.

These new techniques deserve testing and it is hoped that they may prove to be of use and have economic viability.

---

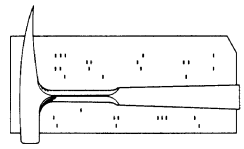
* Active Member, ° Associate Member, + Junior Member, American Association of Petroleum Geologists.

SYMPOSIUM ON COMPUTER APPLICATIONS IN

PETROLEUM EXPLORATION

Edited By

DANIEL F. MERRIAM



1969

CONTENTS

COMPUTER UTILIZATION BY GEOLOGISTS

by

Daniel F. Merriam[1/]

"Quantification and the advent of the computer open new vistas in a science traditionally qualitative."
W.C. Krumbein, 1962.

The application of computer-oriented techniques by earth scientists are numerous. These methods in problem solving and data processing have proved of value, especially in petroleum exploration and exploitation (Merriam, 1967a). Some methods, unknown a few years ago, now are used routinely. Many geologic applications have been in the areas of stratigraphy, sedimentalogy, and paleontology. The techniques have been oriented to statistics, trend analysis, correlation and classification, or simulation (Table 1).

Table 1. - Computer-oriented techniques presently utilized in earth sciences.

Statistics	Trend Analysis
Correlation and Classification	Simulation

Statistical techniques were used before advent of the computer and therefore were adapted easily when computers became available. Most statistical techniques were applied in sedimentology, and development was mainly in the 1950's. Trend analysis, correlation procedures, and classification techniques developed in the 1960's. These techniques were too involved or complicated for use prior to the computer age. Simulation in the earth sciences is developing now and undoubtedly will be utilized widely in 1970's.

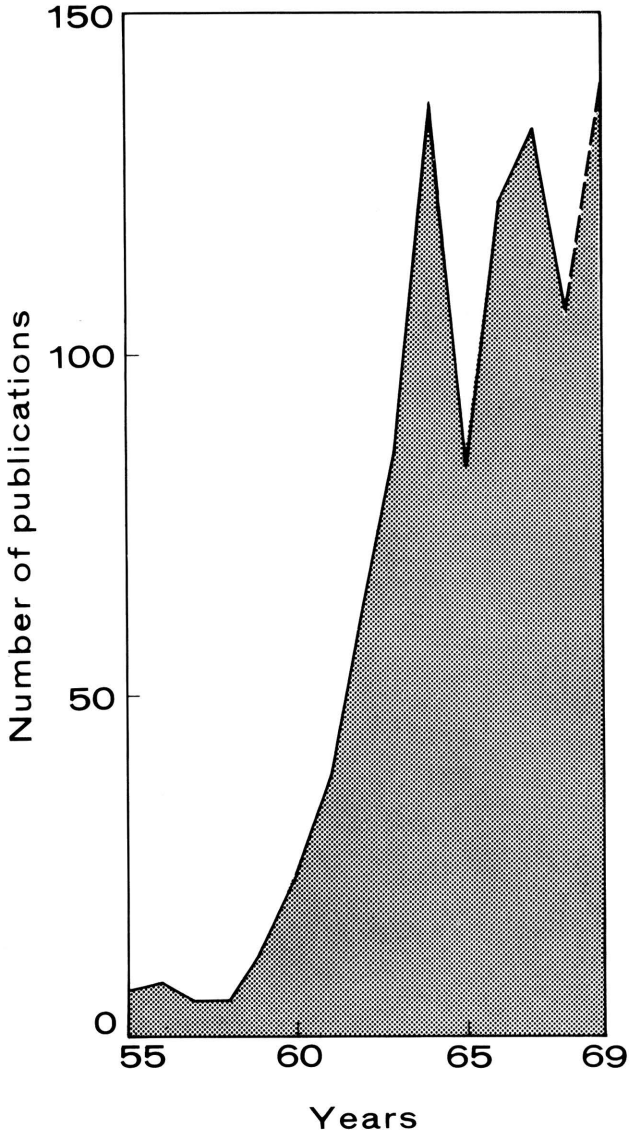


Figure 1. - Number of articles published dealing with computer applications in geology from 1955. Figure for 1969 estimated.

[1/] Kansas Geological Survey, The University of Kansas, Lawrence, Kansas

Other yet unknown techniques await discovery and adaptation by earth scientists.

The literature which deals with computer applications has blossomed in recent years (Fig. 1). It is difficult to keep up because the developments are rapid.

> "Communication in the earth sciences as in all other sciences, is threatened by the savage and accelerating increase in volume of printed matter...We must do something about it!" P.C. Sylvester-Bradley, 1969.

Witness for example the Computer Contribution series of the Kansas Geological Survey and the new Journal of the International Association for Mathematical Geology which are the direct result of demand for special publications in the "new" geology. In addition most established journals willingly accept papers on quantitative geology. KWIC (keyword-in-context) indices are helping alleviate the problem to some extent (Merriam, 1966).

To determine what computer equipment is available to earth scientists, a questionnaire was sent recently to everyone on the Kansas Geological Survey Computer Contribution distribution list. Results of these data are presented here.

Of those answering the questionnaire, 77 percent are concerned with research, and 34 percent with exploration (Table 2). About half of the respondents are associated with academic institutions, 15 percent with government agencies and 35 percent with industry. Of the 480 questionnaires returned, 91 percent have access to a computer and only 9 percent have no access (Table 3). Most workers with no computer facility were scientists working in foreign countries. Of those with access to the computer, 64 percent had plotting facilities and 15 percent had CRT terminals. An overwhelming majority, 81 percent, of the respondents use FORTRAN. Only 11 percent use ALGOL, and that is mainly in Europe (Merriam, 1967b). A small percentage are using PL/1. Other languages reported are machine-dependent or of a specialized nature.

Table 2. - Affiliation of those answering Kansas Geological Survey questionnaire.

Research	77%
Exploration	34%
Academic	50%
Government	15%
Industrial	35%

Table 3. - Accessibility of computer equipment to earth scientists and programming language used.

Access to Computer	91%
no access	9%
Plotter	64%
CRT	15%
FORTRAN	81%
ALGOL	11%
PL/1	5%

It is interesting to note that 52 percent of the respondents had access to IBM equipment, 13 to CDC equipment, 12 percent to English-made hardware, 5 percent to Sperry-Rand, 3 percent to GE, 2 percent to SDS, 1 percent to Burroughs, and others (mainly of Russian, Swedish, German and Japanese make), 12 percent (Table 4). In regard to IBM, this is a considerably less share of the market than normally given to them. For example, in another survey it was found that of 20 major oil companies, 90 percent use IBM equipment.

Geologists often ask how they can use the computer and where computer programs are available. Both applications and programs are published in the Computer Contribution series of the Kansas Geological Survey at The University of Kansas. The series is now co-sponsored by the American Association of Petroleum Geologists. Several books also are available (e.g. Griffiths, 1967; Harbaugh and Merriam, 1968; Krumbein and Graybill, 1965; Smith, 1966; Vistelius, 1967). Of the programs published by the Kansas Survey 39 percent of the respondents are using them; 61 percent are not (Table 5). Of those not using programs, 35 percent said they plan to use them soon; 65 percent said they did not plan to use them.

Table 4. - Type of computer facilities available to earth scientists.

Manufacturer	percent
IBM	52
CDC	13
ICT and English Electric	12
Sperry - Rand	5
GE	3
SDS	2
Burroughs	1
others	12
	100%

Table 5. - Respondents using (or not using) Kansas Geological Survey Computer Contribution programs.

Using KGS programs	39%
Not using	61%
plan to use soon	35%
do not plan to use	65%

Many others, however, said they formed a base for development of their own programs. On the whole, the respondents considered the programs documented adequately (76 percent).

The respondents were asked for an indication as to the areas they would like programs made available. This is interesting because simulation and modeling rated top priority with 68 percent (Table 6). Sixty-two percent indicated they would like additional plotting and graphic display programs; 62 percent are interested in correlation techniques; 61 percent wanted additional programs in multivariate statistics; and 57 percent want additional trend-surface programs.

Table 6. - Types of programs wanted by respondents.

Programs Wanted	
Simulation and modeling	68%
Plotting and graphic display	62%
Correlation Techniques	62%
Multivariate Statistics	61%
Trend — surface Analysis	57%

More than 50,000 computers are presently in use and organizations in the mineral sciences (mainly the petroleum industry) are major consumers of these services. Although geologists have been slow to utilize the computer, geophysicists, petroleum engineers and others have used it intensively for many years (Merriam, 1965). For the most part earth scientists have been using programs adapted from other disciplines, especially from biology, psychology, statistics, and engineering.

In 1964, Weber wrote:

> "It is quite clear from the nature of currently published research that the potential of the digital computer has not yet been realized, and that the enormous impact of this revolutionary device on the nature of geological investigations is still to come".

Five years later this observation is true and I expect it will be true five years from now. There are many exciting possibilities of this new and most interesting development of computer utilization in the earth sciences. I hope everyone will take the opportunity to learn of them.

REFERENCES

Griffiths, J.C., 1967, Scientific method in analysis of sediments: McGraw-Hill Book Co., New York, 508 p.

Harbaugh, J.W., and Merriam, D.F., 1968, Computer applications in stratigraphic analysis: John Wiley & Sons, New York, 282 p.

Krumbein, W.C., 1962, The computer in geology: Science, v. 136, no. 3522, p. 1087-1092.

Krumbein, W.C., and Graybill, F.A., 1965, An introduction to statistical models in geology: McGraw Hill Book Co., New York, 475 p.

Merriam, D.F., 1965, Geology and the computer: New Scientist, v. 26, no. 444, p. 513-516.

Merriam, D.F., 1966, Geologic use of the computer: Wyoming Geol. Assoc. 20th Ann. Conf., p. 109-112.

Merriam, D.F., 1967a, Computer aids exploration geologists: Oil and Gas Jour., v. 65, no. 4, p. 118-121.

Merriam, D.F., 1967b, Computer use in Europe: Geotimes, v. 12, no. 9, p. 14-16.

Smith, F.G., 1966, Geological data processing: Harper & Row, New York, 284 p.

Vistelius, A.B., 1967, Studies in mathematical geology: Consultants Bureau, New York, 294 p.

Weber, J.N., 1964, Impact of the digital computer on the geological sciences: Jour. Geol. Educ., v. 12, no. 2, p. 41-48.

MARKOV FORECASTING TECHNIQUES IN EXPLORATION

by

Graham Lea[1/]

ABSTRACT

When preceding events influence succeeding events a certain probability can be calculated for the process, which is said to possess the Markov property. An increasing number of geological processes have been described that demonstrate this property, and the behavior of exploration geologists is frequently no exception. Markov methods allow a reasonably limited number of exploration factors to be considered together on a probability basis. A particular advantage is that factors having different dimensions - such as barrels of oil, density of seismic coverage, or cost of drilling - can be evaluated together for forecasting purposes.

Small Markov studies can be undertaken without a computer, but for larger models it is both simple and desirable to use a computer. A forecasting model should include consideration of environmental conditions (the historical events leading to the play), alternative choices (i.e. the possible outcomes from which the optimum forecast may be derived), and weights attached to each factor.

Forecasts can be made for two general areas. Within a company, the exploration environment conducive to success is worthy of investigation as is the efficiency of the exploration process. From a competitive viewpoint, the behavior of other companies is of interest because advantage can be taken of any known Markov tendencies in their exploration policies. Geologists should use Markov methods to reduce the uncertainty of decision-making to finite probability. This will result in an increased success ratio.

[1/]Editor, GEOCOM Bulletin, London, England.

MULTIVARIATE FACIES MAPS

by

James M. Parks[1/]

INTRODUCTION

Facies maps are used by geologists to depict and to study the obvious and the subtle differences in a rock stratum over a regional map area. In the search for nonstructurally controlled or stratigraphic traps, the patterns of depositional environments and of diagenetic changes may provide clues to the location of potentially productive reservoir rocks, and may help eliminate from consideration areas more likely to be barren. The criteria for facies maps are usually not one simple characteristic, such as mean grain size, or degree of sorting, or percentage of a certain mineral, or the presence or absence of a particular fossil. No consistent combination of features are sufficient criteria for meaningful facies maps. Simple facies maps, that rely on sand-shale ratio or clastic-carbonate ratio or an arbitrarily subdivided three-component diagram, are seldom adequate: such maps use only a small part of the available data. If, in an attempt to use more of the data, several types of facies maps are constructed, the interpretation becomes difficult. A single facies map that makes use of all the available data should be more useful.

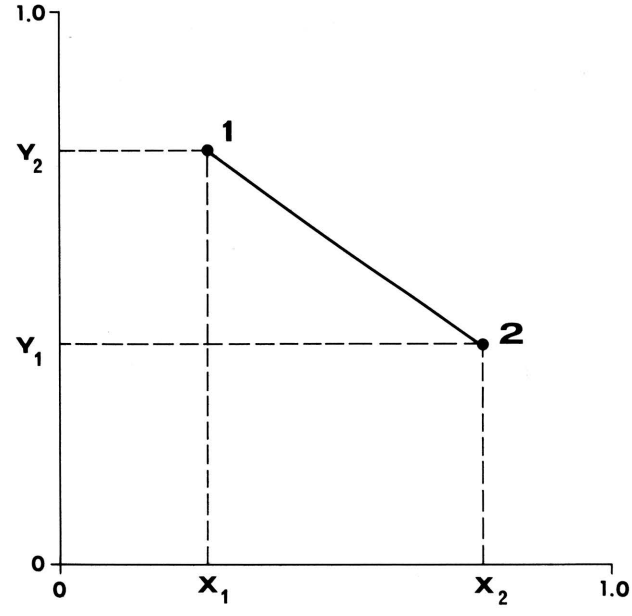
Several types of facies maps have been developed in recent years based on multivariate classification techniques, such as factor analysis (Imbrie and Purdy, 1962), principal components analysis (McCammon, 1966, 1968), and cluster analysis (Bonham-Carter, 1965; Parks, 1966). Cluster analysis starts with a comparison of each sample (or location) with all other samples in the study, using some measurement of similarity, such as the product-moment correlation coefficient, cosine theta, one of several coefficients of association, or a distance function.

One of the limitations with most coefficients is that each is useful on only one type of data. Some types of the coefficient of association are for coded present-or-absent data, and some will handle multiple-coded states. The product-moment correlation coefficient and cosine theta are designed specifically for continuously variable measurements. It is virtually impossible with most similarity measurements to mix data modes, and this can become a serious limitation in many situations where data exist in more than one mode. One of the goals I set in attempting to develop yet another multivariate facies map technique was to find a measure of similarity that could handle mixed-mode data.

THE SIMPLE DISTANCE FUNCTION

The simple distance function seems to be capable of handling mixed-mode data. As used in the technique described in this paper, raw data are first normalized, or transformed to range from 0.0 to 1.0. For two-state presence/absence data, presence is coded as 1.0 and absence as zero. Multicoded states or whole number integer counts in an ordered sequence (for instance, to code grain shape, one is for angular, two for subrounded, three for rounded, and four for well-rounded) are normalized to become in this example 0.0, 0.33, 0.67, and 1.0. Continuously variable measurements and percentages also are transformed to range from zero to one.

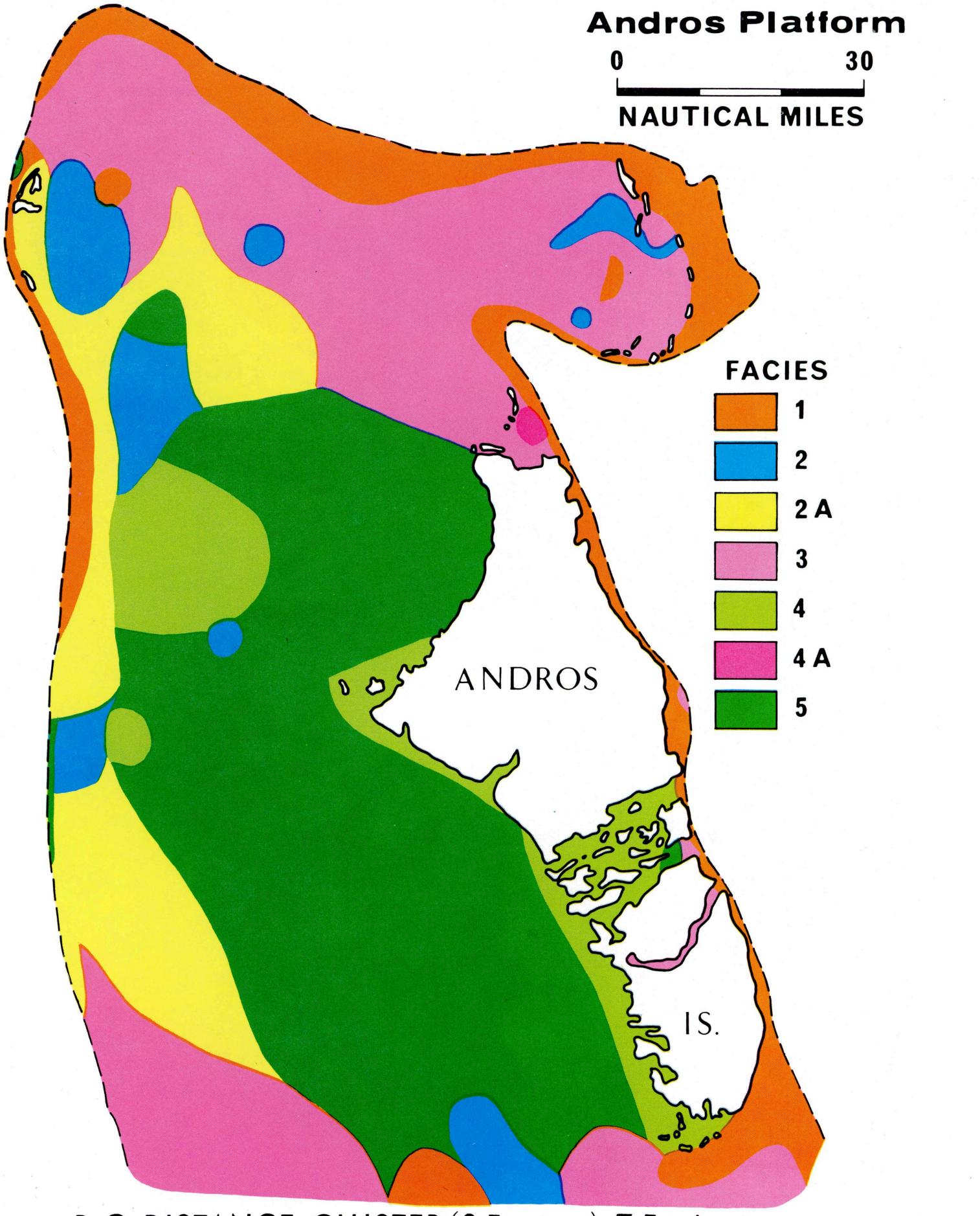
In the simple situation of two dimensions, two samples are plotted according to the values of two variables, X and Y, measured or counted on each (Fig. 1). The distance between these two points is,



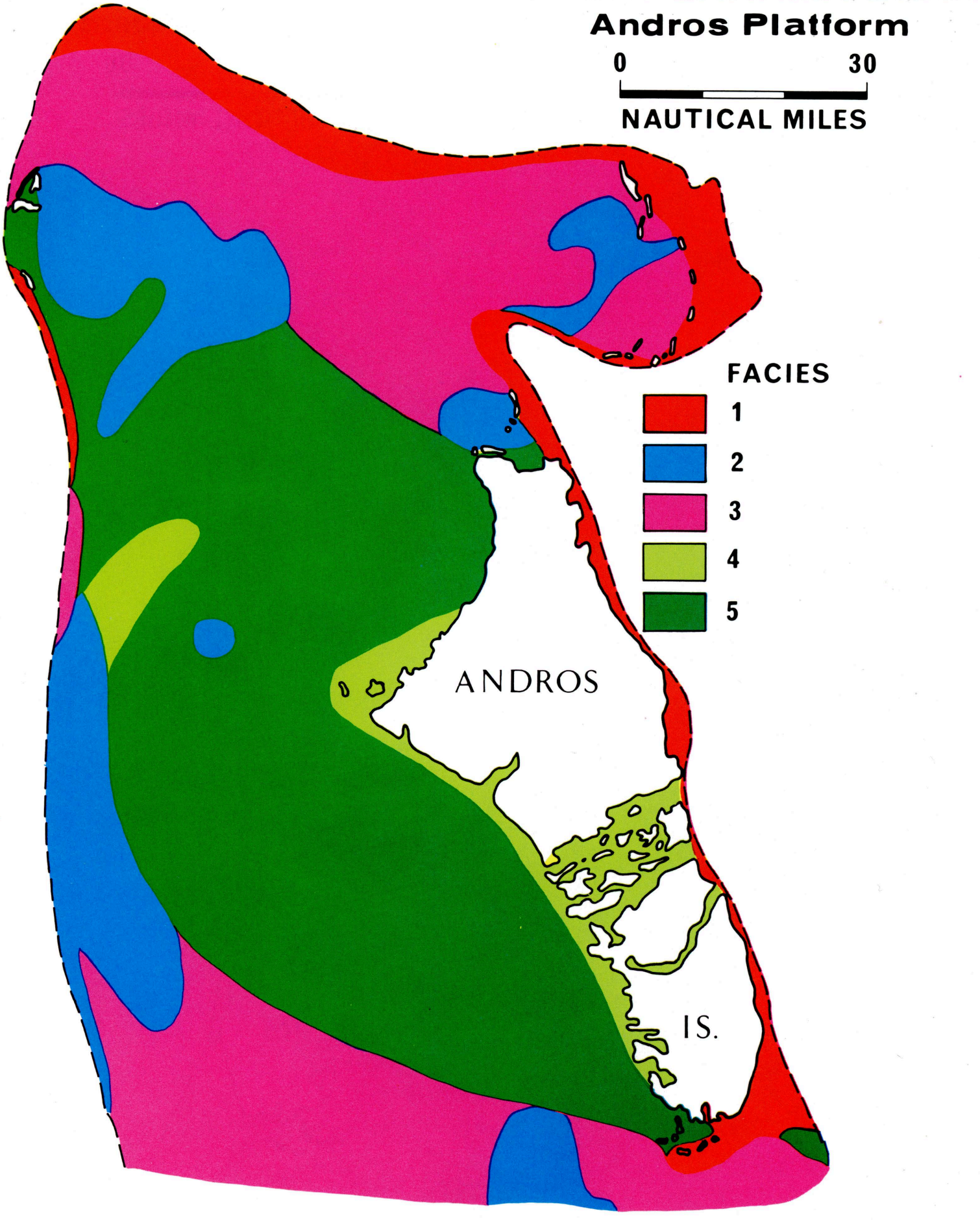
$$D_{1,2} = \sqrt{\frac{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}{M}}$$

Figure 1. - The simple distance function.

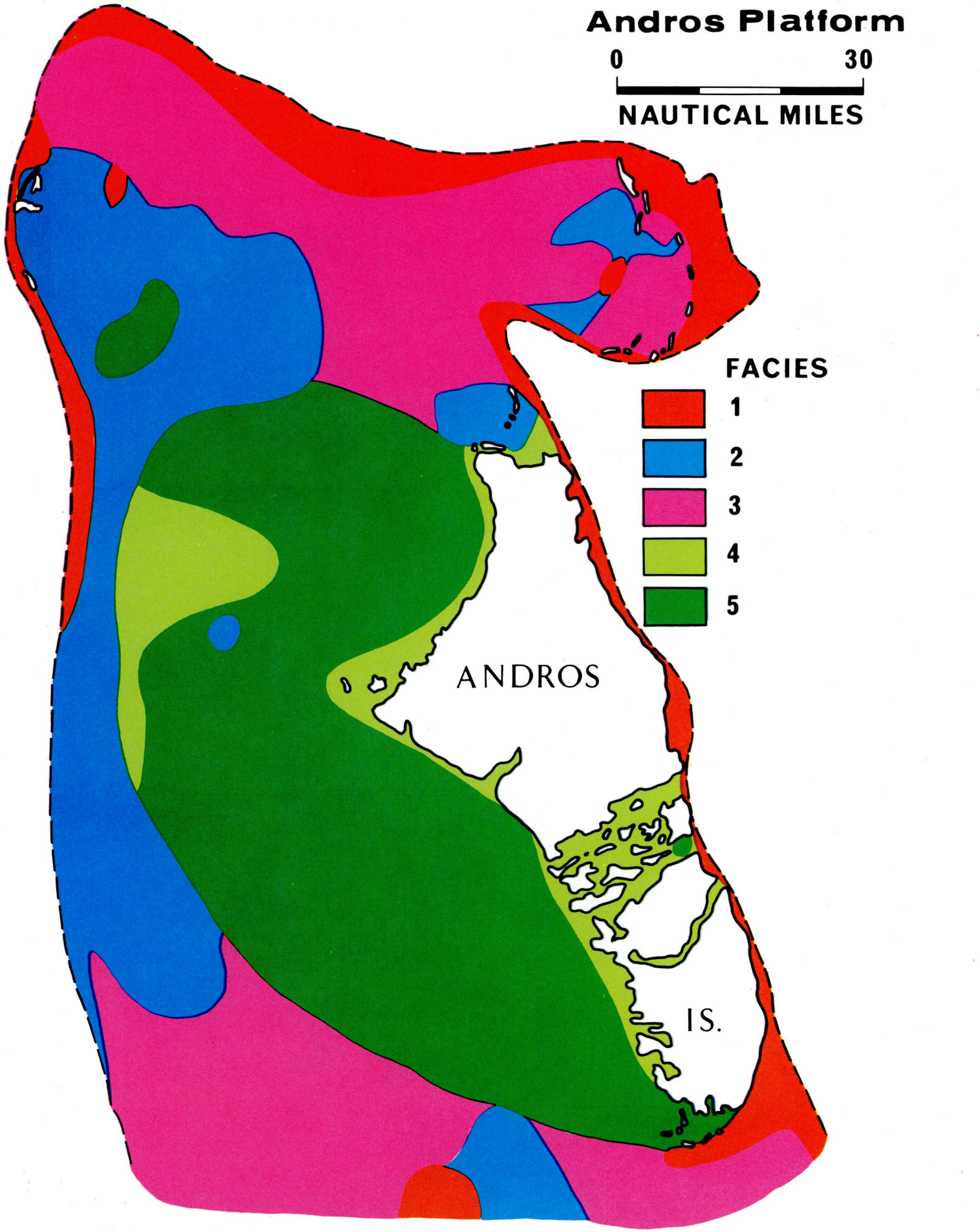
[1/] Center for Marine and Environmental Studies, Lehigh University, Bethlehem, Pennsylvania



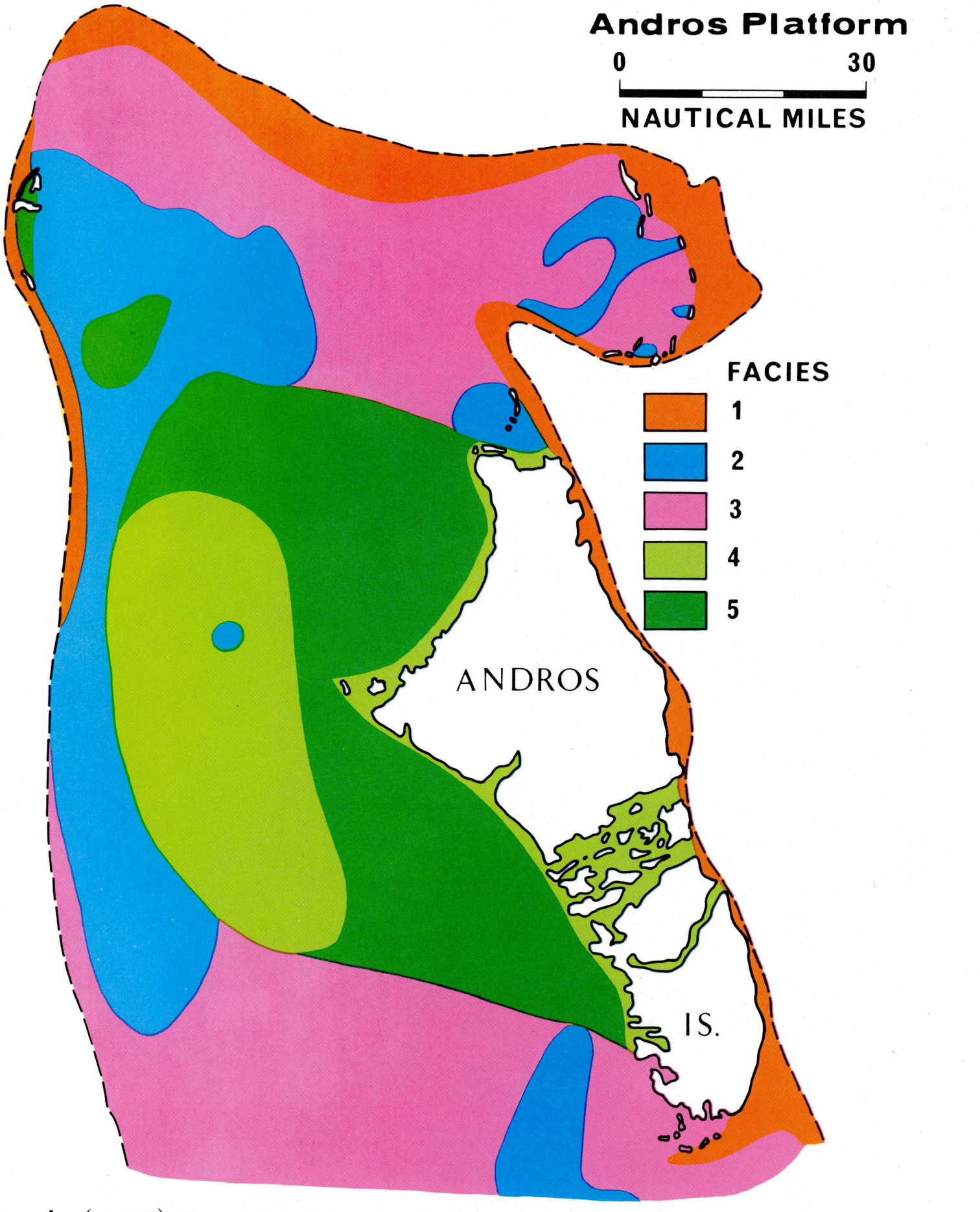
R-Q DISTANCE CLUSTER (3 Factors) 7 Facies



McCammon (1968) 5 MULTIPLE COMPONENTS



R-Q DISTANCE CLUSTER (3 Factors) 5 Facies



Purdy (1963) Q FACTOR ANALYSIS, 5 GROUPS

by simple geometry, the square root of the sum of the squared differences between the X and Y values of the two points; as, in a right triangle the square of the hypotenuse is equal to the sum of the squares of the two sides of the triangle. This is expanded easily to 3, 4, or more dimensions or variables. With normalized variables, the maximum distance between two points determined by two variables is equal to the square root of $(1.0)^2 + (1.0)^2$ or 1.414; for 3 dimensions the maximum distance is equal to the square root of three or 1.732, etc. To keep the distance values from being dependent on the number of variables used, the sum of the squared differences is divided by the number of variables (M) so that the maximum distance is always 1.0. Thus we have a mean squared difference, similar to the variance of the differences between the measurements on the two samples; and as the simple distance (D) is the square root of this mean squared difference, it is similar to the standard deviation of these differences.

This calculation of the simple distance function assumes that the input variables (or the axes from which they are measured) are uncorrelated, that is, orthogonal or at right angles to each other. However, most raw variables are correlated to differing degrees so that the coordinate axes would not be at right angles and the simple Euclidean distance formula would be inaccurate. To overcome this difficulty, the original variables can be transformed to orthogonal (uncorrelated) variables by an R-mode principal components analysis. This has the additional advantage of usually reducing the number of axes or variables needed.

CLUSTER ANALYSIS

A two-dimensional distribution of points such as shown in Figure 2A would be difficult to classify into its natural groups or clusters by most methods.

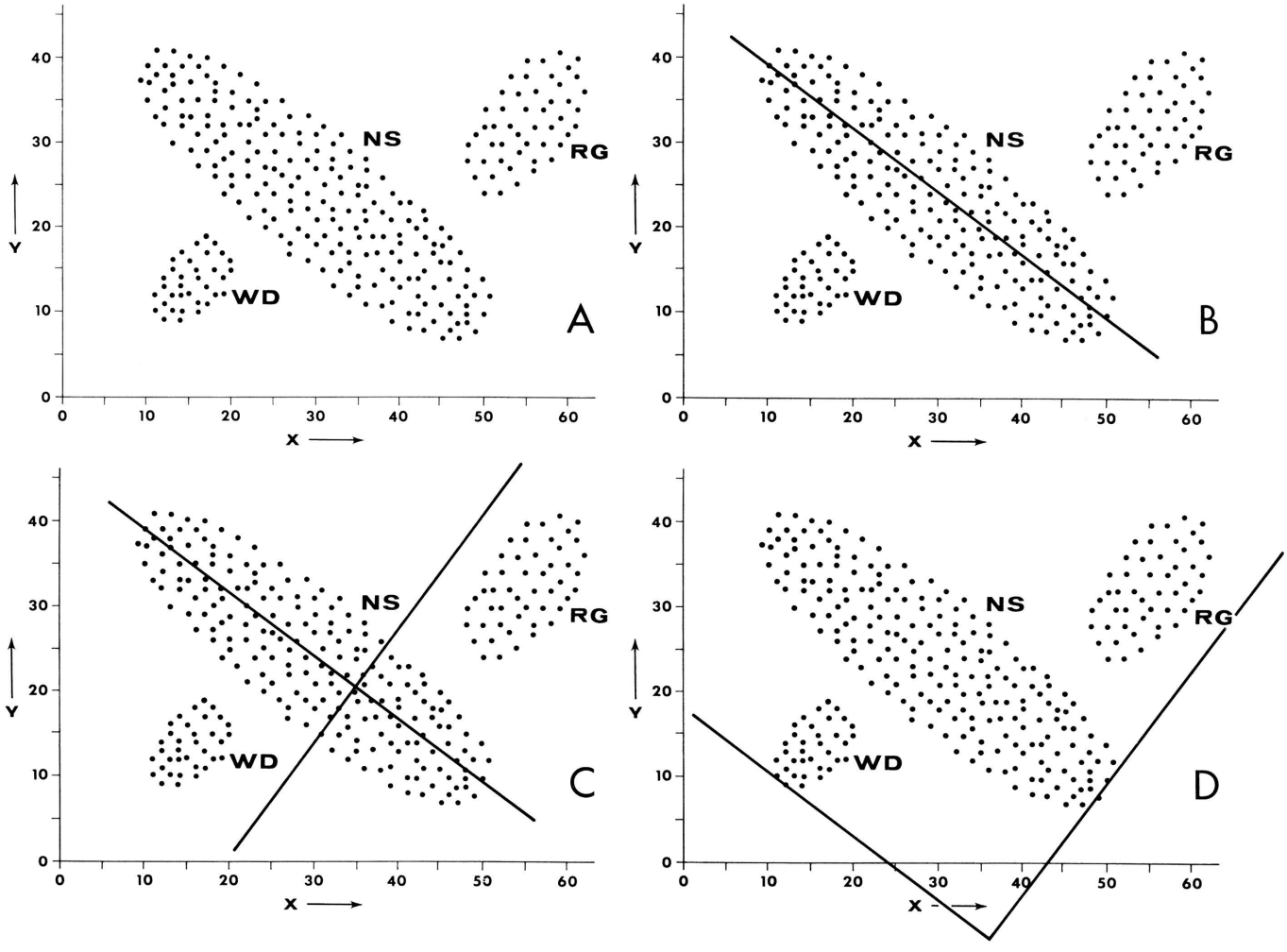


Figure 2. - Two-dimensional distribution of points (A); axis through data accounting for maximum variance (B); second axis which accounts for next largest amount of variance (C); translated axes (D).

In an oversimplified graphical fashion, the R-mode principal components analysis is somewhat equivalent to passing a new axis through the data in such a fashion as to account for the largest portion of the variance (Figure 2B), and then passing a second axis through the data at right angles to the first axis so as to account for the next largest amount of variance (Figure 2C), and so on. In this instance two axes are sufficient to account for all the variance. Now the positions of the points are measured from the new transformed axes, not from the original coordinates. If the original coordinates were not orthogonal, this transformation would produce orthogonal coordinates. It is not necessary to perform a rotation of these new axes to so-called simple structure, because the distances between points will remain the same no matter what the rotation. In practice, these new axes are translated laterally, by normalizing the new factor measurements, so that all measurements are positive and range from zero to one (Figure 2D).

Normally an R-mode principal components analysis uses a correlation coefficient matrix for the variable-by-variable comparison. A simple distance function matrix often gives more meaningful results on geological data. As the distances differ from zero (closest similarity) to 1.0 (greatest dissimilarity), it is only necessary to reverse this relationship by subtracting each value from 1.0 to achieve a matrix amenable to principal components analysis.

Finally, a Q-mode (sample-by-sample) cluster analysis is performed, using distance function computed from the factor measurements.

The cluster analysis computer program developed in this study has several unique features, one of which is a computer printed output of the cluster diagram or dendrogram. Figure 3 is a subset of 40

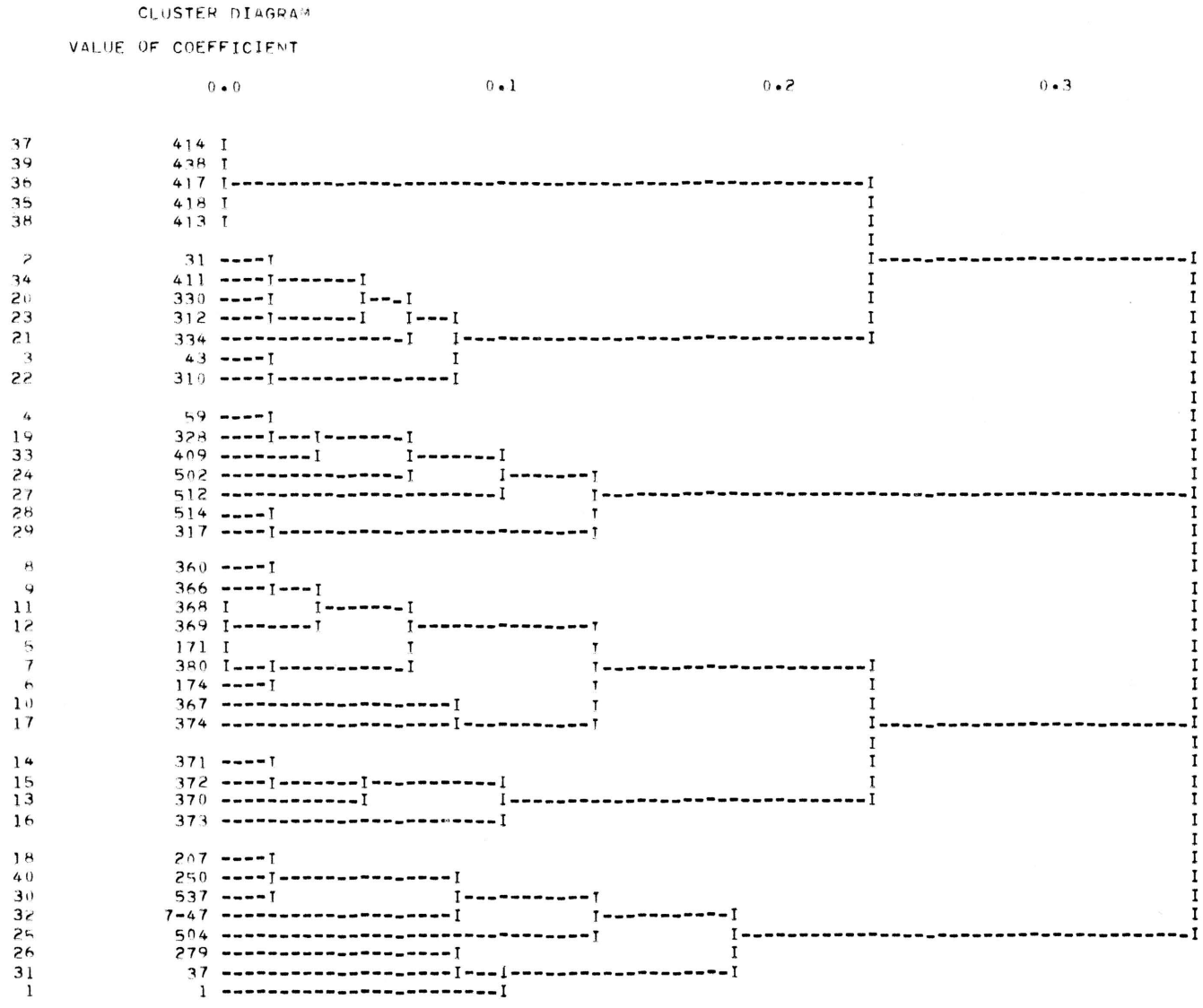


Figure 3. - Dendrogram output of cluster program.

samples from a larger study used later in this paper as an example. The sequential sample numbers are listed in the left column, and the next column contains the actual sample designators which may be alphabetic as well as numeric. The X-axis of the diagram is the value of the distance function and the scale is expanded to full page width for the largest distance encountered. From such a diagram, clusters or groups of similar samples can be selected at any desired level of similarity, and each cluster then represents a facies which can be plotted on a map as a number, a pattern or a color.

MULTIVARIATE FACIES TECHNIQUE

The ultimate test of any multivariate facies technique is that it can produce a meaningful map pattern from a set of data. This is difficult to demonstrate except by comparison with known examples. Several of the recently proposed multivariate facies methods have used the published data from the study of recent carbonate sediments by Purdy (1960, 1963a, 1963b) on the Bahama Platform around Andros Island. Purdy obtained data on 12 compositional variables from 216 locations and used the Q-mode factor analysis technique (Imbrie and Purdy, 1962) to develop a map pattern similar to Plate 1 with 5 facies.

McCammon (1968) pointed out that the Q-mode factor analysis method is restricted by the memory size of most present large computers to about 200 samples, whereas many real world problems involve as many as 1000 samples. McCammon's multiple component method, not limited in sample number, produced the similar pattern of Plate 2, but with a more restricted area of pellet mud (facies 4).

The R-Q distance cluster method used in this paper, using factor measurements derived from three principal components accounting for 83.4 percent of the variance, produced the slightly different map pattern of Plate 3, if five facies are used.

Table 1 shows a quantitative comparison of Purdy's map with the R-Q distance cluster map, according to a method described by McCammon (1968). Along the main diagonal are the number of samples assigned to the same facies by both methods. The off-diagonal numbers are the samples assigned to one facies by one method but to another facies by the other method. One hundred eighty-six out of 216 samples are in agreement.

Actually, by the R-Q distance cluster method, there seem to be 6 or 7 distinctly different facies, not just 4 or 5. In Figure 3, a subset from Purdy's data, 6 distinct clusters are shown. Whether or not the more complicated map pattern of Plate 4 has any real meaning might depend upon more knowledge of the area, more data, or further interpretation.

Another test used by McCammon (1968) to see for which classification method the samples are most tightly clustered for a given number of groups is the partition variance or total within-groups sums of squares. For N samples, characterized by P variables partitioned into R groups, the partition variance (TSS) is defined by the formula shown in Table 2, where $\bar{x}_{i.k}$ is the mean of the K^{th} variable of the i^{th} group

Table 1. - Quantitative comparison of Purdy's map with the R-Q distance cluster map, according to methods of McCammon.

M=186, N=216

R-Q DISTANCE CLUSTER

PURDY (1963) FACIES	1	2	3	4	5
1	23	1	2	3	2
2	3	67	1	0	0
3	2	1	61	0	0
4	0	0	0	14	13
5	0	2	0	0	21

(3 FACTORS, 83.4% VARIANCE)

Table 2. - Statistical data for classification method comparisons.

PARTITION VARIANCE

(TOTAL WITHIN GROUPS SUM OF SQUARES)

$$rSS = \sum_{i=1}^{r}\sum_{j=1}^{ni}\sum_{k=1}^{P}\left(x_{ijk}-\bar{x}_{i.k}\right)^2$$

	NO. OF GROUPS 4	5	6
R-Q Distance Cluster (3 Factors)	1231.5	1095.3	1021.8
Multiple Components (McCammon,68)	1245.4	1085.6	—
Q-Factor (Purdy,63)	1305.2	1206.3	—
Principal Components (McCammon,68)	1264.4	1168.6	—
Optimize Objective Function (Ward,63)	1356.3	1136.1	—

into which N_i samples have been classified. A lower value means a better grouping or classification. The R-Q distance cluster method and McCammon's multiple components method are similar for 4 and 5 facies, and both of them are better than Q-factor or principal components or the optimized objective function of Ward (1963). For 6 facies by the R-Q distance cluster method, the partition variance was even lower.

The R-Q distance cluster method, embodied in a FORTRAN IV program written for the CDC 6400 computer, can handle up to 200 variables on as many as 1000 samples. A complete computer run from raw data to dendrogram, for the 12 variable 216 sample Bahama carbonate problem using 3 principal components, lasted about 100 CDC 6400 central processor seconds.

An example of more practical interest to petroleum exploration geologists involves ancient rocks. Many types of lithologic and paleontologic data on geologic rock formations are available to geologists in lithologic sample logs. The American Stratigraphic Company, Denver, Colorado, has completed re-

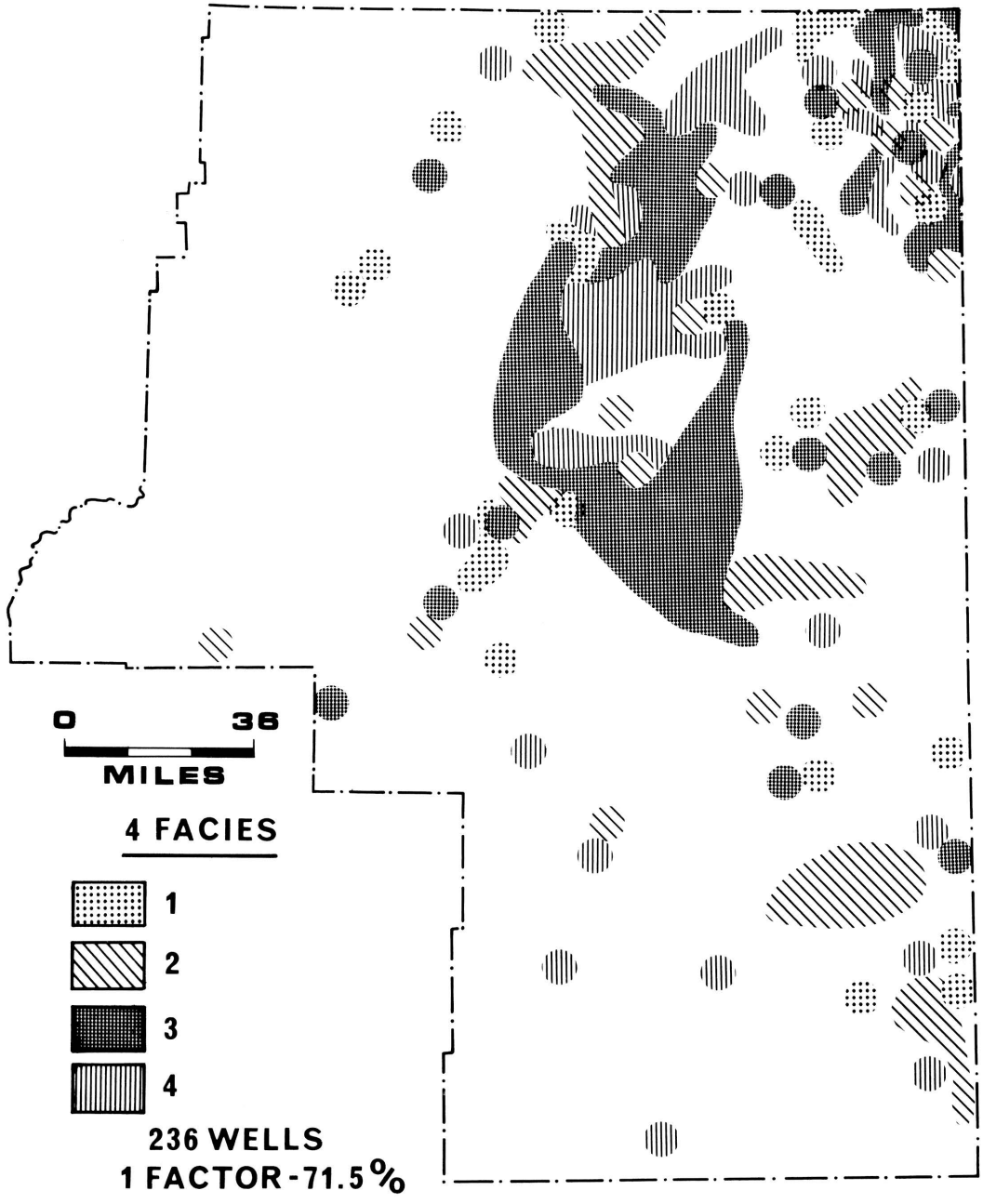


Figure 4. - Distribution of facies of Lower Kibbey Limestone (Mississippian), for AmStrat Area I in eastern Montana.

cently coding on magnetic tape for computer use their sample logs for Area I in eastern Montana. There is a wealth of data on these logs, so much data that an electronic digital computer is needed to make adequate use of the information. Approximately 30 rock types have been coded; as well as grain size on a scale from 1 through 10 plus additional carbonate, clay and coarse sizes; 4 multistate codes for grain shape; 3 for sorting; framework by percent; 48 fossil framework builders; 10 types of carbonate diagenesis; a 3 part color code; porosity percent and type; and presence/absence coding for minor accessories, fossils and minerals. According to the mixed-mode decoding system used in this study, there are 180 possible variables. Not all the possible variables are used in any one study.

For this study, AmStrat generously loaned me a magnetic tape of coded sample log data from Area I for the Lower Kibbey Limestone, a formation of little present oil potential, to use in testing multivariate facies techniques. Two hundred thirty-six wells had data on the Lower Kibbey Limestone. There are many possible ways of decoding and of using the AmStrat data, and only a few have been tested here. Using one approach, 4 facies were found, with a map distribution as shown in Figure 4. In areas of sparse data, no attempt was made to interpolate the facies, but in the northeastern portion of Area I, some definite patterns of distribution of the four facies are evident. No attempt has been made to characterize the different facies or to interpret the patterns. Much more needs to be done on data such as that in the coded AmStrat sample logs.

None of this work, involving so many variables and so many samples, could even be attempted without the aid of a modern electronic digital computer with a large memory core. The data files, the decoding of these files, the multivariate analysis, all are performed easily on a computer (Figure 5). The difficult task, that of interpretation of the resulting maps, is the task of the exploration geologist.

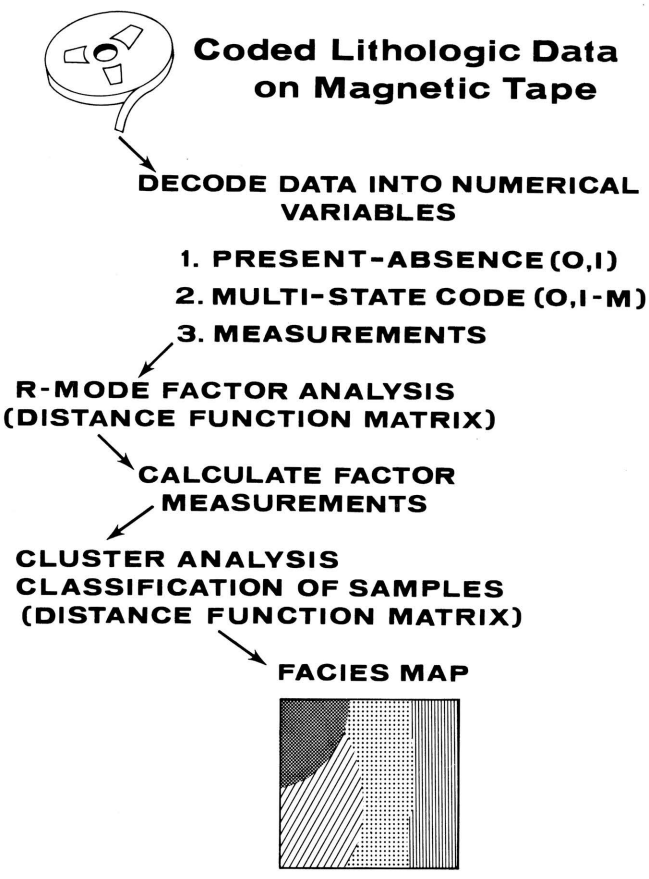


Figure 5. - Flow diagram of data processing using computer methods.

REFERENCES

Bonham-Carter, G.F., 1965, A numerical method of classification using qualitative and semi-quantitative data, as applied to the facies analysis of limestones: Canadian Petroleum Geology Bull., v. 13, no. 4, p. 482-502.

Imbrie, J., and Purdy, E.G., 1962, Classification of modern Bahamian carbonate sediments: Am. Assoc. Petroleum Geologists Mem. 1, p. 253-272.

McCammon, R.B., 1966, Principal component analysis and its application in large scale correlation studies: Jour. Geology, v. 74, no. 5, pt. 2, p. 721-733.

McCammon, R.B., 1968, Multiple component analysis and its application in classification of environments: Am. Assoc. Petroleum Geologists Bull., v. 52, no. 11, pt. 1, p. 2178-2196.

Parks, J.M., 1966, Cluster analysis applied to multivariate geologic data: Jour. Geology, v. 74, no. 5, pt. 2, p. 703-715.

Purdy, E.G., 1960, Recent calcium carbonate facies of the Great Bahama Bank: unpubl. doctoral dissertation, Columbia Univ., 174 p.

Purdy, E.G., 1963a, Recent calcium carbonate facies of the Great Bahama Bank. 1. Petrography and reaction groups: Jour. Geology, v. 71, no. 3, p. 334-355.

Purdy, E.G., 1963b, Recent calcium carbonate facies of the Great Bahama Bank. 2. Sedimentary facies: Jour. Geology, v. 71, no. 4, p. 472-497.

Ward, J.H., 1963, Hierarchical grouping to optimize an objective function: Am. Statistical Assoc. Jour., v. 58, p. 236-244.

SPATIAL FILTERING ILLUSTRATES RELATIONSHIP BETWEEN TECTONIC STRUCTURE AND OIL OCCURRENCE IN SOUTHERN AND CENTRAL ALBERTA

by

J.E. Robinson[1] and H.A.K. Charlesworth[2]

ABSTRACT

Structural contour maps commonly express a combination of features whose relief and spacing differ widely. Analysis of individual features or trends is most successful if the effects of conflicting features are suppressed. Spatial filtering permits the rigorous suppression of all defined features that fall outside a specified range of size or trend without distortion of the main attributes of the desired features. The process is a practical application of Fourier theory that allows any contourable feature to be expressed in terms of a limited range of sinusoidal surfaces of specific frequencies, phases, amplitudes and orientations. It requires the use of a digital computer and involves determining the wavelengths present, designing a spatial filter to remove the unwanted wavelengths and filtering the map.

The application of spatial filtering techniques to structural contour and topographic maps in central Alberta clearly indicates the presence of a series of features that are mathematically rigorous in their length, breadth and trend. These features form two mutually perpendicular structural trends with the most prominent aligned NE-SW. The structures extend throughout the sedimentary section and are interpreted as resulting from repeated adjustments of tabular blocks along Churchillian-like patterns in the Precambrian basement. The orthogonal NW-SE trend is subordinate generally to the main trend except in an area northeast of Edmonton where there is evidence of a pre-Cretaceous wrench fault. Both trends seem to have exerted an influence on the biological, depositional and erosional environments that have resulted in oil and gas accumulations.

INTRODUCTION

A stratigraphic surface usually presents a complex and interesting array of features to the geologist. Not only is it affected by depositional, compactional and possibly erosional irregularities but each tectonic movement may set up its own set of faults or folds. All such irregularities are structures and are combined to produce the present attributes of the observed surface. A structural contour map of a stratigraphic surface therefore contains the algebraic sum of all structures displayed by the surface that can be defined within the limits of the original sample spacing. The structures range from large-scale regional gradients to small undulations that are beyond the resolving power of the sample intervals and that must be considered as part of the error inherent in every map. Interpretation of individual structures or trends is most effective if the desired features can be isolated and examined free from distorting effects of both error and conflicting structures. The process that permits the mathematically rigorous extraction of a specific size range of structures from a heterogeneous background is called spatial filtering (Robinson, Charlesworth, and Kanasewich, 1968). It permits the breaking down of a conglomerate of structures into smaller parts so that each segment or event can be considered separately.

Spatial filtering is most feasible if carried out on a high-speed digital computer. The computer accepts a uniformly digitized geological map, then transforms it into a new map containing only those structures that fall within the size range that is to be examined in detail. If the spatial filter is designed correctly, the filtered map will contain only "real" structures with parameters such as length, width and trend that can be defined rigorously. Filtering permits the geological analysis to be separated into two parts, one mathematically rigorous and one that is based on geological experience and intuition.

SPATIAL FILTERING PROCEDURES

The basic principles leading to spatial filtering were originated over 150 years ago when Fourier was able to show that any finite continous function of distance, such as a topographic cross section, can be described completely by the sum of a series of sinusoidal waves (Lee, 1960). For instance, a square wave (Fig. 1) can be approximated reasonably by the summing of two of its component sine waves. If more of the components are added the approximation is improved until finally the desired form is reproduced exactly. More complex sections may require a greater number of component wave forms but can likewise be constructed from simple sine and cosine waves.

[1] Union Oil Company of Canada Limited, Calgary, Alberta

[2] The University of Alberta, Edmonton, Alberta

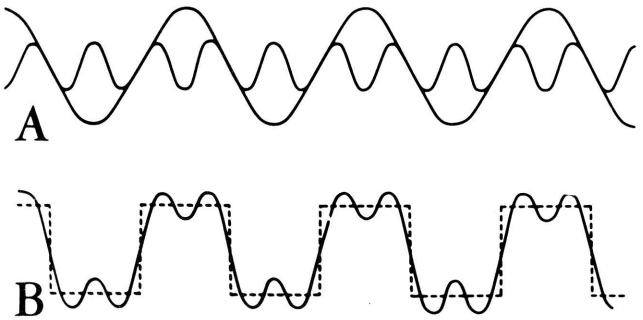


Figure 1. - Any function of distance can be described by sum of its sinusiodal components. Sum of two sine waves (A), approximates square wave (B).

Similarly any contourable map can be represented by the sum of its component sinusoidal surfaces (Fig. 2). These surfaces are similar to sheets of corrugated iron and are really two-dimensional sine

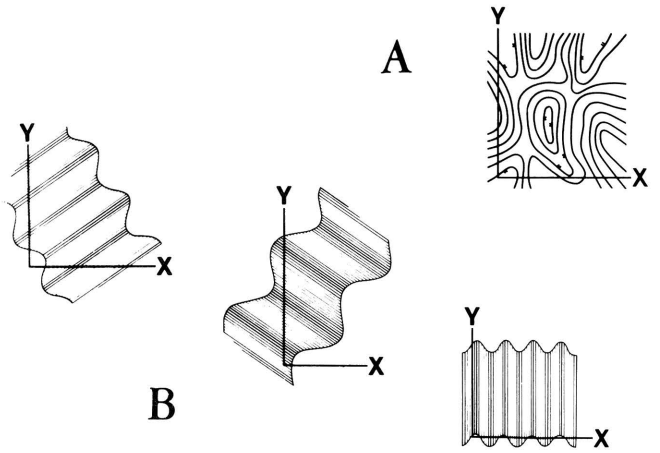


Figure 2. - Any contour map (A) can be described by sum of its two-dimensional sinusoidal components (B).

waves. The amplitude, direction, position and wavelength of these component spatial frequencies determine the size and location of all structures that make up the mapped surface (Bhattacharyya, 1965; Dean, 1958, Swartz, 1954). Individual structures are composed of a discrete range of spatial frequencies with the large regional features dominated by those with long wavelengths, and the small features represented by the short wavelengths. The spatial filtering process deletes unwanted features by rejecting their component spatial frequencies while permitting those of the desired structures to pass unchanged. Filters can be considered as strainers because the earliest filters were felt mats for removing inpurities from liquids. The process can be described mathematically, electronically or mechanically.

Tuning a radio to a specific program illustrates electronic filtering. The station selector of a radio is a variable filter that can be adjusted to delete all frequencies except those making up the desired program. A radio will not tune in a station that is off the air, and spatial filtering as applied to geological maps will not introduce any structures that are not in the original map. The spring suspension of a motor vehicle is a mechanical filter (Fig. 3). It allows the passengers to feel only a moderate up and down motion when passing over sharp bumps. However, because the main effect of the bump is not felt until after the vehicle has gone over it, the spring filter has the effect of apparently displacing the position of the bump. This is a fault that could not be tolerated in a geological map and therefore spatial filters are designed especially so that all structures are retained in their correct location.

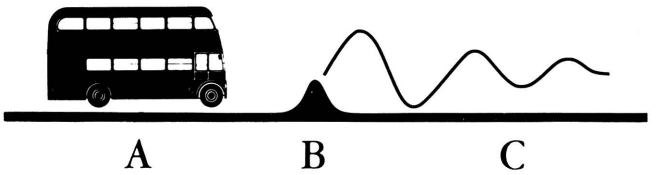


Figure 3. - Suspension of vehicle is mechanical filter (A). Bump is input (B) and motion of passengers is filter output (C).

Spatial filtering of geological maps requires the construction of the best filter for the desired feature and applying it mathematically to the original maps. The filter is simply the digitized sum of the individual sinusoidal components that compose the structures that are to be retained. The mathematical process involved in applying the filter to the original map is called convolution (Lee, 1960). For a topographic section (Fig. 4) it is performed by superposing the reversed filter on the section, multiplying adjacent map and filter values and adding the results. This sum is the filtered value for the center position of of the filter. After each multiplying and summing operation the filter is moved one digital position to

SECTION	S_1	S_2	S_3	S_4	S_5
FILTER	F_3	F_2	F_1		
	→	F_3	F_2	F_1	
		→	F_3	F_2	F_1
OUTPUT		O_1	O_2	O_3	

$$O_1 = S_1F_3 + S_2F_2 + S_3F_1$$

Figure 4. - Digital filtering involves multiplying corresponding filter and section values then summing results. After each operation filter is shifted one position to right.

the right and the process repeated until the entire section has been covered. Filtering of maps is similar except the data are in matrix rather than vector form (Holloway, 1958; Zurflueh, 1967).

The result of the spatial filtering process is a new map containing only the desired structures in a form in which their position, length, width and trend are mathematically rigorous. The absolute shape and the amplitude of the features are not rigorous and must be interpreted because the original geological information usually is insufficient for exact definition of all parameters.

APPLICATION OF SPATIAL FILTERING

Spatial filtering techniques were applied to structural contour and topographic maps covering a 100,000 square mile area of central and southern Alberta (Fig. 5). The structural maps were compiled from elevations determined from 7,500 randomly distributed exploratory wells. Horizons included top of the Devonian, pre-Cretaceous unconformity, base of the Fish Scale Sandstone and top of the Upper Cretaceous First White Speckled Shale. National topographic system, 1:250,000 sheets were used for the surface topography. All maps contain a variety of features with the structural maps dominated by the strong westward regional gradient and large structures such as the Sweet Grass Arch in the southern portion of the area.

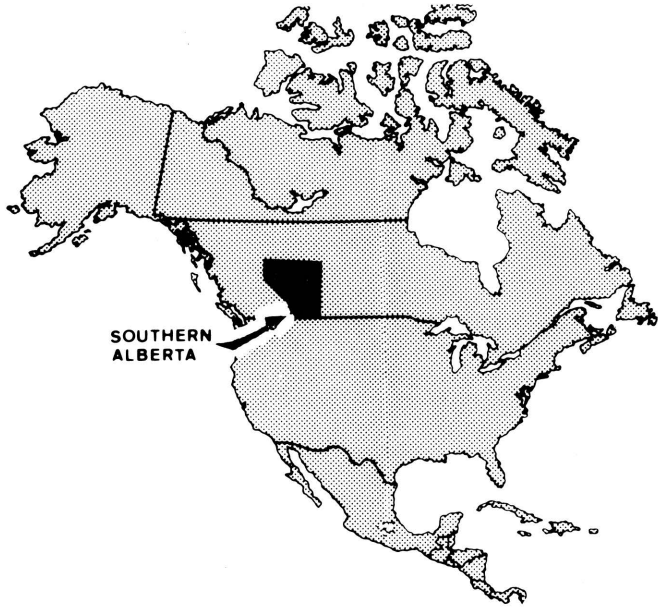


Figure 5. - Index map showing location of southern Alberta.

The unconformity (Fig. 6A) seemingly contains a number of smaller structures but the majority of these are so distorted that effective interpretation is almost impossible. Small features are more subdued in the overlying Base Fish Scales (Fig. 6B) and appear only as minor undulations in the regional contours.

A spatial filter was designed to retain all structures with a minimum dimension of between 10 and 40 miles and was convolved with the structural maps. This size range allowed the filter to pass a relatively wide variety of small and intermediate structures, while deleting both the large structures and the high frequency noise that is inconsistant with the original well spacing. The same filter also was applied to the topographic map. The new maps indicate clearly the presence of two mutually perpendicular structural trends with the most prominent aligned NE-SW.

STRUCTURAL INTERPRETATION

The structural patterns of the Precambrian basement under the interior plains of Alberta, similar to those in the exposed portions of the Canadian Shield (Wilson and Brisbin, 1962), are reflected in the anomalous magnetic features. These (Fig. 6C) show that the Precambrian of central Alberta has the same trends, generally NE-SW with occasional orthogonal interruptions, as those in the Churchill Province of the Shield. In the south, the features tend to show more preference for alignment along the buried Precambrian rift valley noted by Kanasewich (1968).

The filtered maps on the top Devonian and the unconformity (Fig. 6D) contain similar structures that have a marked resemblance both to each other and to those of the Precambrian basement. This resemblance is so strong that there can be little doubt that the features are related. Those on the unconformity have a relief of 200-400 feet and a width ranging from 15 to 20 miles. Lengths may be as much as 100 miles and trend predominantly NE-SW. They are intercepted occasionally by NW-SE trends. Because the structures on the depositional portion of the Devonian are similar but average only half the relief of those on the unconformity, it is interpreted that erosion has enhanced the relief, but retained the original structural pattern. This pattern could have originated only by compaction over existing structures or by tectonic movements. Because detailed seismic studies of good reflecting horizons in the Middle Devonian evaporite section have failed to indicate structures with the necessary amplitude and there is only a partial correlation with known compaction structures over upper Devonian reefs, it is concluded that the filtered structures are not compaction features. They are probably fault-bounded tabular structures that resulted from post-Mississippian-pre-Cretaceous tectonic movements. There is additional evidence of a tectonic origin in the north-central part of the map where there is a prominant northwesterly trending linear feature. The orthogonal NE-SW trending structures have the same relative relief on either side of this feature but are misaligned so that highs are opposite lows. Although the sense of displacement cannot be determined absolutely, the misalignment suggests the presence of a northwesterly trending

wrench fault with a minimum displacement of 10 miles. However, the possibility of an offset in the Leduc-Morinville reef trend (Bassett and Stout, 1967) suggests the movement may have been left handed.

The filtered map on the Base Fish Scale (Fig. 7A) exhibits the same general pattern as they underlying unconformity, but the relief is considerably less and NW-SE trending features are confined to the

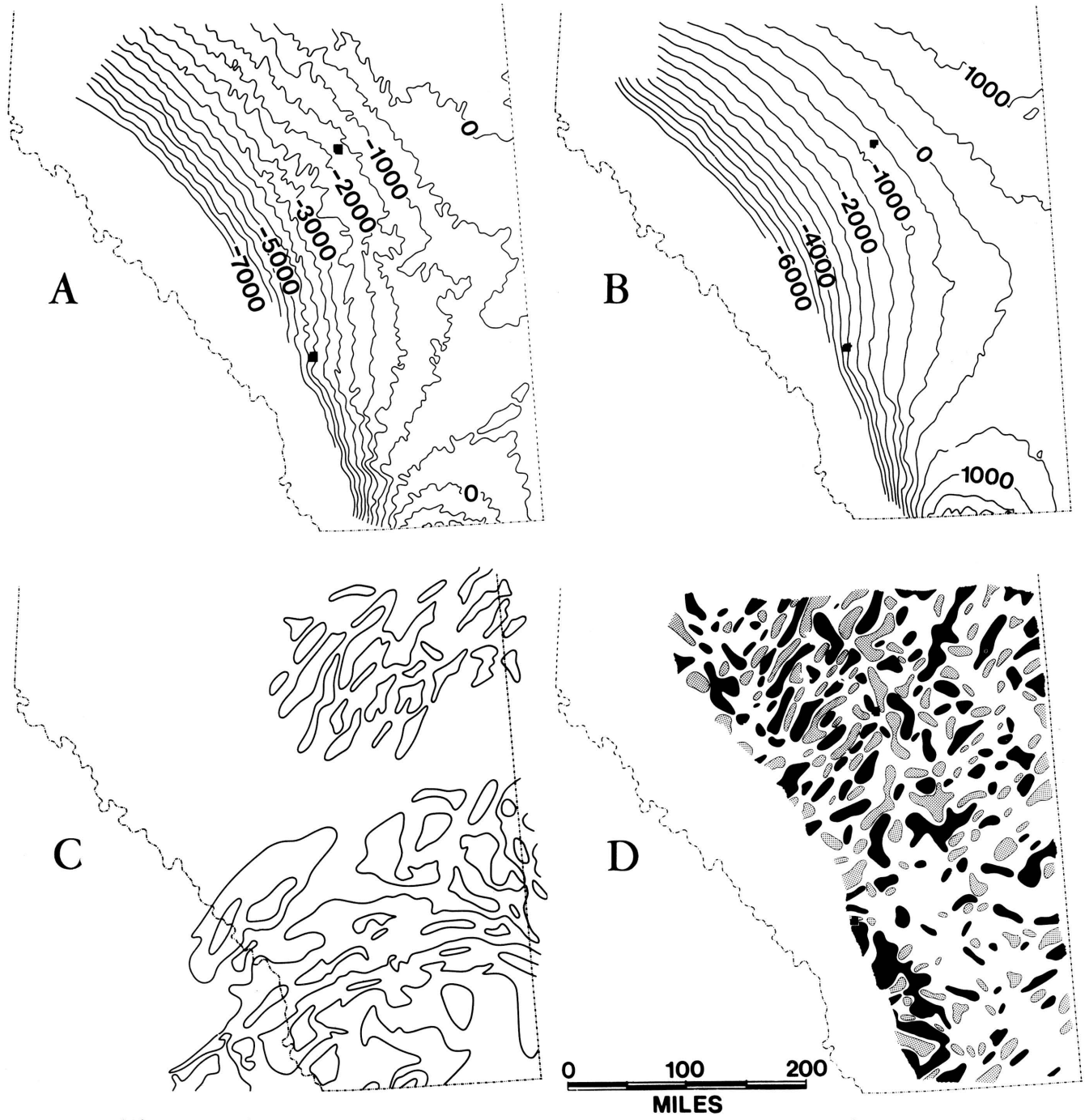


Figure 6. - (A) Structural contours on pre-Cretaceous unconformity. Contour interval - 500 feet; datum - sea level. (B) Structural contours on base of Fish Scale Sandstone. Contour interval - 500 feet; datum - sea level. (C) Basement magnetic anomaly trends (Agarwal, 1960; Kanasewich, 1968). (D) Filtered structural contours on pre-Cretaceous unconformity. Only features with minimum dimension between 10 and 40 miles with relief of greater than 100 feet are retained. Positive structures are shown in solid black.

westernmost part of the area. Because the relief on this map coincides in trend but not in position with that on the unconformity, seemingly Cretaceous features are possibly due not to compaction but to renewed tectonic activity. However, the movements are more modest and are confined generally to a singular

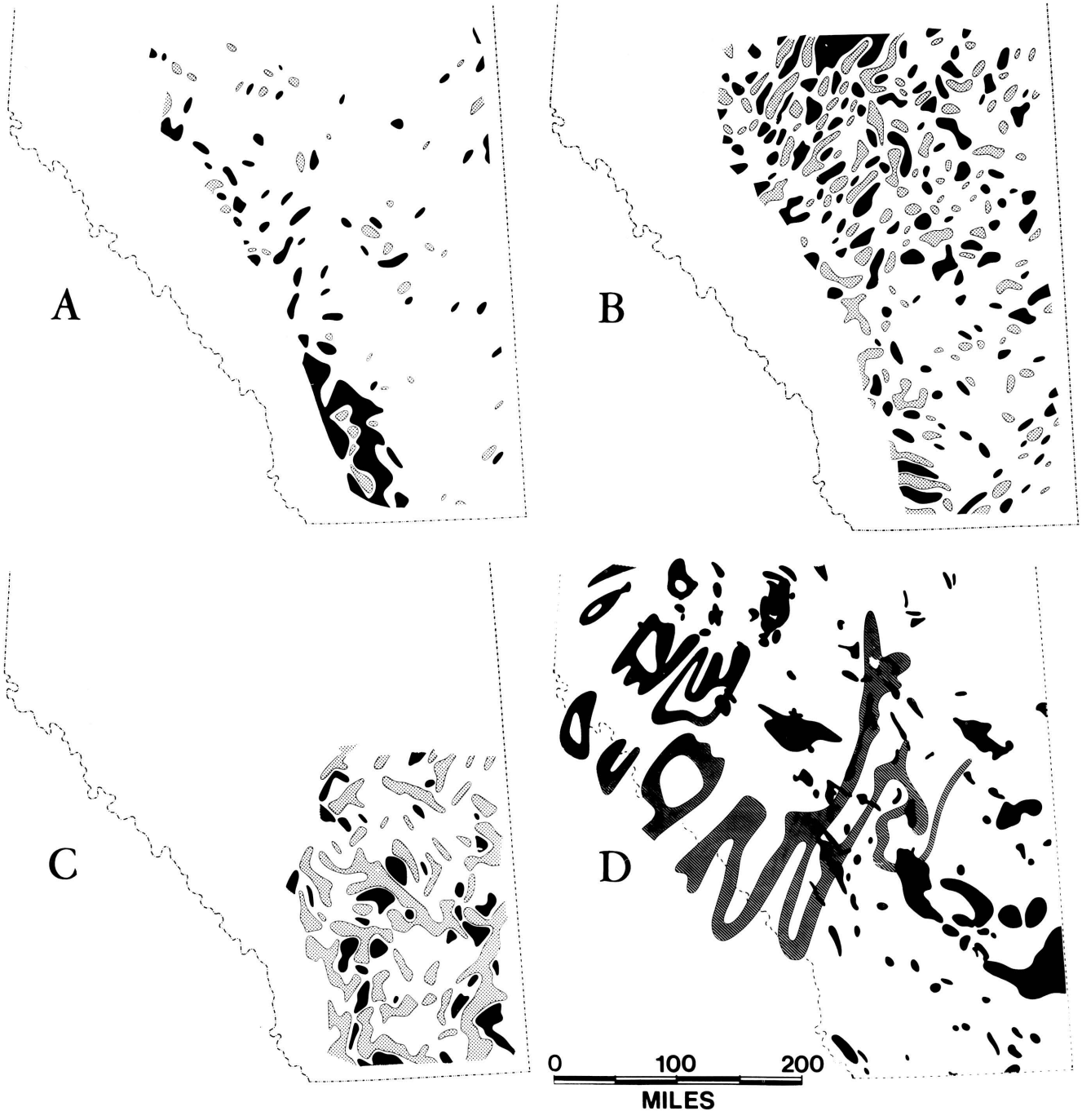


Figure 7. - (A) Filtered structural contours on Base Fish Scales. Only features with minimum dimension between 10 and 40 miles with relief greater than 100 feet are retained. Positive structures shown in black. (B) Apostreptic Base Fish Scales to pre-Cretaceous unconformity map constructed by subtracting one filtered map from another. Features shown are both positive (black) and negative with over 100 feet departure. (C) Filtered topographic contours. Only features with minimum dimension between 10 and 40 miles with relief of over 100 feet are retained. Positive structures shown in solid black. (D) Main gas and oil field (black) and major reef trends (shaded) (Bassett and Stout, 1967) in map area.

trend. Confirmation of this interpretation is suggested by the relief on the apostreptic map (Fig. 7B) which describes the difference between the two filtered maps. The word, apostreptic, is from a Greek adjective meaning divergence from the normal, and is used to describe what would be normally an isopach except there can be both positive and negative differences between two filtered surfaces. This map shows that in many instances the significant relief on the unconformity has been obliterated by Lower Cretaceous sedimentation.

The filtered map on the Upper Cretaceous First White Specks shows structures with the same pattern but they are not as abundant and have less relief. The filtered topographic map (Fig. 7C) has the same pattern but the relief is dominated by erosional rather than tectonic agencies. The maps show that intermittent tectonic activity occurred during the Cretaceous and possibly into recent time. The basement tectonic trends can be discerned in the surface topography and seem to have resulted from a series of adjustments along zones that have been present since Precambrian time. Because a consistent pattern of repeated tectonic movement has been impressed on the sedimentary section, it would seem likely that there should be some effect on contemporaneous depositional and ecological patterns.

The structures in the filtered maps are large if compared to the size of individual oil-producing reservoirs. Therefore the structural influence should be in alignment of prospective trends rather than in the position of individual fields. On this basis, the positions of known reefs and oil and gas fields (Fig. 7D) show a striking agreement with the tectonic trends. Some, such as the Leduc-Morinville reef trend, have an angular divergence and may originate from causes that are not apparent on the maps, but there is sufficient correlation to justify the hypothesis that tectonic movement has influenced strongly the environmental patterns leading to oil and gas accumulation. If new information from recent wells and seismic profiles are included, filtering techniques may be used, not only as an aid in the structural analysis of the western Canadian sedimentary basin, but also to localize specific oil and gas prospects.

ACKNOWLEDGMENTS

Financial assistance was provided by the National Research Council of Canada, The Geological Survey of Canada, Hudson's Bay Oil and Gas Co. Ltd., and The University of Alberta. Imperial Oil Ltd. supplied the original well information and Union Oil Company of Canada Ltd. drafted the figures. J. E. Hogg and J. W. Keith critically read the manuscript.

REFERENCES

Agarwal, R.A., 1960, Interpretation of aeromagnetic data in west central Saskatchewan and east central Alberta: Saskatchewan Dept. of Min. Res., Rept. no. 63, 17 p.

Bassett, H.G., and Stout, J.G., 1967, Devonian of western Canada, in International symposium on the Devonian System: Alberta Soc. Petroleum Geol., p. 717-752.

Bhattacharyya, B.K., 1965, Two-dimensional harmonic analysis as a tool for magnetic interpretation: Geophysics, v. 30, no. 5, p. 829-857.

Dean W.C., 1958, Frequency analysis for gravity and magnetic interpretation: Geophysics, v. 23, no. 1, p. 97-127.

Holloway, J.L., 1958, Smoothing and filtering of time series and space field: Advances in geophysics, v. 4: Academic Press Inc., New York, p. 351-389.

Kanasewich, E.R., 1968, Precambrian rift: Genesis of strata bound ore deposits: Science, v. 161, p. 1002-1005.

Lee, Y.W., 1960, Statistical theory of communication: John Wiley & Sons, New York, 509 p.

Robinson, J.E., Charlesworth, H.A.K., and Kanasewich, E.R., 1968, Spatial filtering of structural contour maps: 23rd Session Internat. Geol. Cong. (Prague), Sec. 13, p. 163-173.

Swartz, C.A., 1954, Some geometrical properties of residual maps: Geophysics, v. 19, no. 1, p. 46-70.

Wilson, H.D.B., and Brisbin, W.C., 1962, Tectonics of the Canadian Shield in northern Manitoba, in Tectonics of the Canadian Shield: Royal Soc. of Canada Spec. Publ. 4, Univ. of Toronto Press, p. 60-75.

Zurflueh, E.G., 1967, Application of two-dimensional linear wavelength filtering: Geophysics, v. 32, no. 6, p. 1015-1035.

AMBIGUITY FUNCTIONS AND THE CONCEPT OF GEOLOGICAL CORRELATION

by

Norman S. Neidell[1/]

ABSTRACT

A profitable approach to quantitative geology is identifying the methods currently employed. It is clear for instance that the intuitive notion of correlation is more general than the mathematical operation. The most obvious difference is the lack of provision for stretching or shrinking of scales during comparison.

Ambiguity functions, an elementary extension of cross-correlation, includes a scale variable. The name was taken from the parlance of radar engineers, and the technique was devised originally for measuring target velocities or detecting fast moving targets. The same principles offer an effective means for identifying thinning or thickening of stratigraphic intervals by their well-log signatures, for matching magnetic profiles and following trends or for estimating dispersion from seismic results.

Other applications will, of course, come to mind as the method is exposed and simple examples are studied. The principal accomplishment, however, is in bringing the mathematical model of the concept of correlation one step nearer to the definition implied by practice.

GEOLOGICAL CORRELATIONS AND CONSISTENCY OF HYPOTHESES

One of the most frequently occurring breaches in the logic of science is the use of hypotheses which lack consistency. No group of scientists should be singled out for committing these types of errors, however, they are made most frequently in the evolution of quantitative frameworks for qualitative or intuitive concepts.

The notion of correlation in geology, or in the earth sciences, has itself been a cause of many mistakes in this vein. Indeed, these mistakes are one of the prime motivations for the presentation of ambiguity functions and related quantities.

We can illustrate easily the widespread misuse of correlation in the earth sciences by drawing from a wide spectrum of examples - a few however, will suffice to make the point. In geology for instance, the existence of uniformly convergent or divergent bedding is accepted as commonly occurring in a number of environments. Hence the possibility of this condition must be kept in mind while making stratigraphic correlations from well-log data. The human eye will identify similar or anomalous signatures despite their scaling in amplitude or depth (sometimes time). But all of us probably are familiar with investigations in which well-log data were compared using ordinary cross-correlation methods. Because these methods can detect only shifts and have no provision for stretching or shrinking of the depth scale, most results have been unspectacular or inconclusive. A clear cut inconsistency in hypotheses exists in every such study.

Aeromagnetic or ship-towed magnetometer profiles are an abundant and currently popular type of data for analysis. Whereas earlier studies of such data were directed toward commercial exploration or else separation of core and crustal components, the newer studies given fascinating insights into the large-scale crustal tectonics. Such works were initiated with Vine and Matthews (1963) who suggested that the cyclic reversals of the earth's magnetic field were recorded permanently in the solidified crust which had spread from mid-ocean rifts and ridges. The hypothesis, as originally conceived, implied more or less uniform rates of spreading, and made ordinary mathematical cross-correlation an excellent tool for matching sequences of events on parallel lines which had become displaced owing to other processes. In fact, we have become used to viewing correlations of adjacent profiles as shown in Figure 1.

But the work of Vine and Matthews has been further developed and expanded by Cox (1969), Dickson, Pitman, and Heirtzler (1968), Godby, Hood, and Bower (1968), and Morgan (1968). Global tectonics involving seafloor spreading although simple, is no longer as simple as it was. Transform faults had to be introduced (Wilson, 1965) and differential rates of spreading, even from the same lineament have become accepted. Here again we see a basic inconsistency in the use of simple cross-correlation in comparing profiles. Clearly a scale stretch or shrink parameter is necessary to compensate for different spreading rates. We must further comment that such a tool ought to provide an excellent method for measuring the relative spreading rates.

Lastly, let us turn toward reflection seismology perhaps the most sophisticated element within the

[1/] Seismic Computing Corp., Houston, Texas

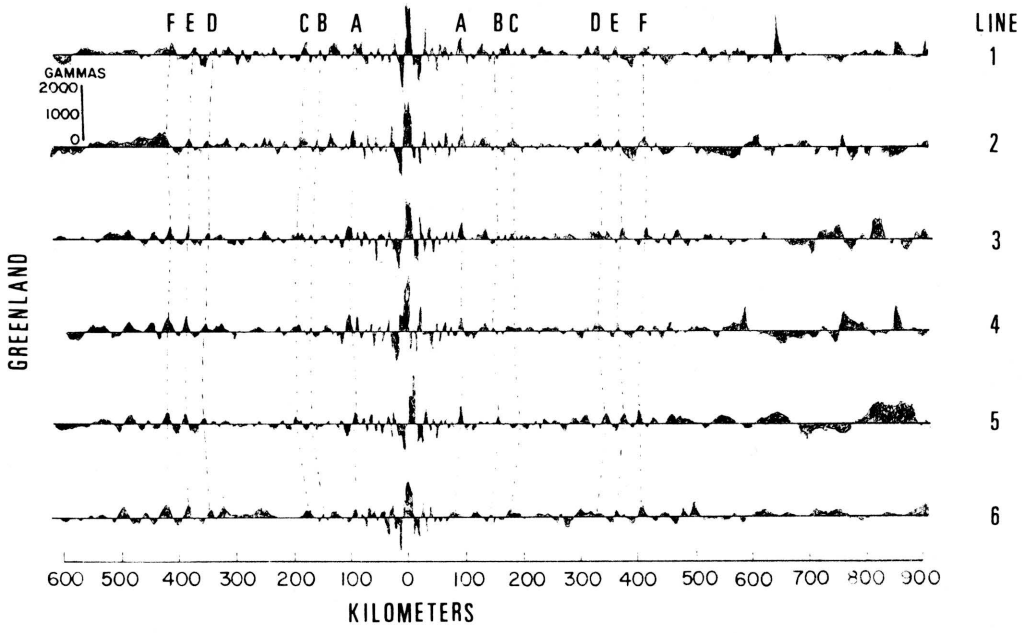


Figure 1. - Correlation of magnetic profiles (after Godby, Hood, and Bower, 1968).

entire sphere of the earth sciences. The Vibroseis[1/] seismic source is perhaps its most sophisticated tool - a digitally controlled source (Neidell, 1968). Conventionally, long frequency sweeps or chirps (7 seconds on the average) are put into the ground. A vital part of signal recovery in processing is a correlation of the returning signal with the outgoing or control signal. This usage is similar to radar signal processing. In that discipline it is recognized that moving targets stretch or shrink the return depending on the sense of target velocity hence degrading the correlations. For this reason ambiguity functions were introduced.

The earth is a stationary target it is true, but there are nevertheless mechanisms by which returning signals can be stretched or shrunk. For one thing, misalignment of the vibrators such that they fail to follow the control signal is a possibility. Dispersive mechanisms in the uncompacted near surface or weathered layers offer another source for stretching or shrinking signals. The degradation of correlation with stretching was first pointed out by Taner, Cook, and Koehler (1968) with reference to Vibroseis processing. Their experiment is shown in Figure 2, until then scale adjustments were not employed in Vibroseis processing, another inconsistency in hypothesis.

By this point most readers would agree that there is a genuine need for ambiguity functions, a cross-correlation analysis with a stretch allowance, in the earth sciences as well as in other fields. In the next sections we introduce and describe such an analysis and illustrate its use.

INTRODUCING AMBIGUITY AND SIMILARITY FUNCTIONS

Computing a cross-correlation analysis which incorporates a stretch/shrink parameter is straightforward. In fact, the calculation follows the simple description. Ambiguity functions have been used for some time in radar signal processing, however the pertinent literature is specialized and borders on the esoteric (for example, Gassner and Cooper, 1967). Ambiguity functions are introduced here and defined for what they are - cross-correlation analyses with the inclusion of a stretch/shrink factor. A similarity function also will be presented as being useful for eliminating an inconsistency in hypothesis which has been overlooked up to this point. The similarity function is an obvious extension of the ambiguity function which has not been employed before.

In discrete mathematical notation (which is the only kind applicable to a digital computer) the ambiguity function $A(\gamma,\alpha)$ is taken to be

$$A(\gamma,\alpha) = \frac{1}{N} \sum_{i=-\left(\frac{N-1}{2}\right)}^{\left(\frac{N-1}{2}\right)} U_{i+\alpha} V_i(\alpha) \qquad N \text{ odd.}$$

[1/] Registered trademark and service mark of Continental Oil Company.

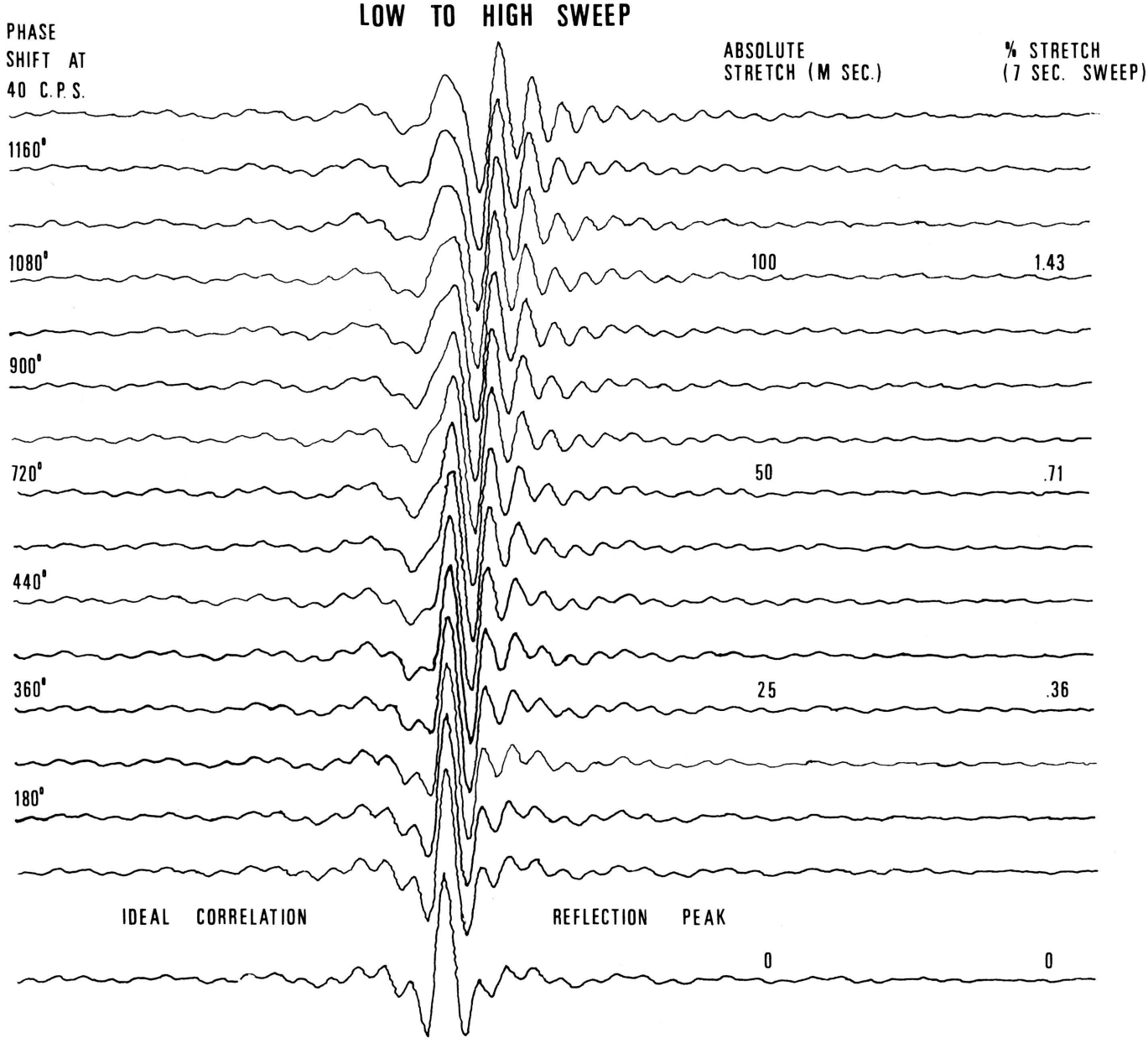


Figure 2. - Effects of stretch on Vibroseis correlations (after Taner, Cook, and Koehler, 1968).

The peculiar notation $V_{i\,(\alpha)}$ merely denotes that the series of elements V_i has been stretched by a factor α and resampled so that the indicated operations may be performed. We comment here that the mathematical expression glosses over many practical details and difficulties. These will be treated in the following section and in the computational Appendix.

Consistency of hypothesis has been stressed repeatedly, yet we have deliberately allowed an inconsistency to pass to this point. The purpose for such an intentional oversight is to illustrate the degree of subtlety which may be encountered. Let us now unmask a culprit.

In discussing well-log correlation, the situation of uniformly convergent or divergent dips was noted. But what becomes of a correlation technique in the face of discordant dips? After all, here is another circumstance which is present routinely in nature. Similarly, in talking about the new global tectonics and the associated significance of magnetic data, we neglected to mention the discontinuities in spreading rate which have been identified from epoch to epoch even from the same lineament. Once again the matter calls for treatment of discordant dips.

Figure 3 shows what discordant dips do to a

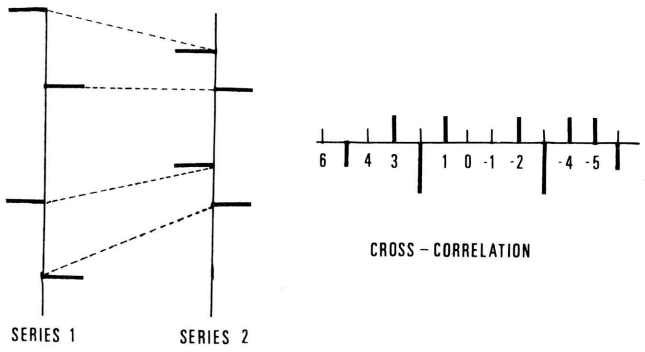


Figure 3. - Cross-correlation in presence of discordant dips.

correlation analysis - it becomes nonunique and meaningless. Inclusion of a scale stretch parameter does nothing to improve matters. One solution for this difficulty is to provide also a range parameter over which homogeneity of dips or trends is assured. Hence we define the similarity function to eliminate this inconsistency. Specifically, $S(\gamma, \alpha, K)$

$$S(\gamma, \alpha, K) = \frac{1}{K} \sum_{i=-\left(\frac{K-1}{2}\right)}^{\left(\frac{K-1}{2}\right)} U_{i+\alpha} \cdot V_i(\alpha) \qquad K \text{ odd.}$$

The meaning and usage of similarity functions follows directly from the ambiguity function. Although we have not employed similarity functions in an organized manner they do deserve at least a comment on their existance. We can now move on to the details of calculation.

CALCULATING AMBIGUITY AND SIMILARITY FUNCTIONS

At this stage in the adaptation of ambiguity and similarity functions to earth science problems there arise a number of details which require legitimately some discussion in depth before any computations are made. First among these is a practical means for accomplishing scale changes. In this context it will be necessary to talk about methods of sampling and interpolation. We will be alerted to the presence and nature of interpolation errors, sampling inadequacy and the aliasing phenomenon.

There are also secondary matters such as appropriate normalization (Purdy and Cooper, 1968) upon which comment should be deferred until our particular uses for this technology have matured, and taken us beyond results which merely affirm our intuition. A normalization question arises because unlike ordinary cross-correlation we are computing functions of several parameters. It is clear that the method of normalization must depend on the way in which the functions are employed. In order to make quantitative comparisons of neighboring ambiguity functions for instance, it is absolutely necessary to introduce some normalization. For our purposes however, where we seek only to identify maxima on individual functions, no scaling is needed.

We choose to calculate ambiguity and similarity functions using a stretch parameter only. By

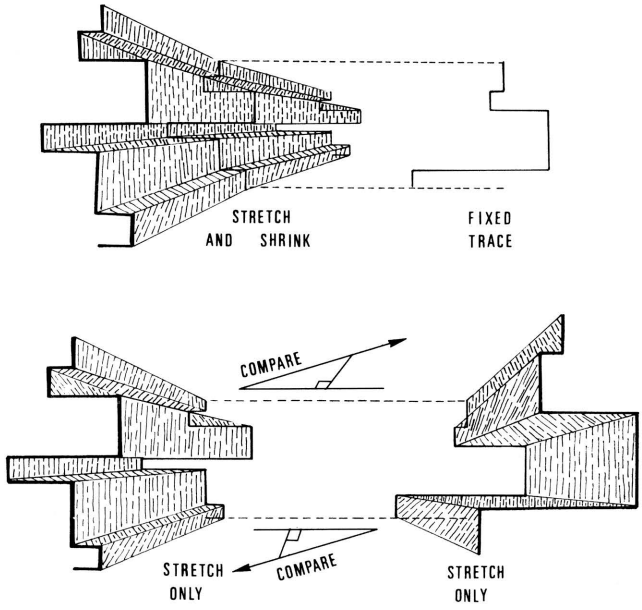


Figure 4. - Alternate methods for computing ambiguity functions.

stretching each series in turn holding the other fixed for comparison the net effect is identical to the stretching and shrinking of one of the series. Figure 4 illustrates this principle. This approach was taken in order to simplify conceptually the filtering operation which should follow the resampling of time-scale deformed data in order to avoid aliasing effects. This important point will be raised again and in more detail during the discussions on sampling and interpolation.

It would be unfair to burden earth scientists with a full blown discussion of sampling theory, particularly as a large number of the functions with which they routinely must deal can not be sampled properly. Nevertheless, the elements of sampling theory are valid and its principles useful. Hence, despite our best intentions we shall be forced to live with compromise. Philosophically we can rationalize the situation by noting that a bit of discomfort in the handling of sampled data is a healthy state of affairs.

Sampled data may be regarded as ideal continuous data of infinite numerical accuracy multiplied in a term-by-term sense with a sampling function.

The sampling function has a limited range, limited digital resolution - or quantization levels - and a finite interval between samples. In essence, sampled data are the ideal data, but viewed through a defective "window". Figure 5 shows some ideal data and a discrete sampling. Note how the rounding of numbers to the quantization levels (or equivalently the finite digital accuracy) has deformed the smooth character of the data. Explicit relations between the ideal data and a sampled form are codified by sampling theory (Lee, 1963).

According to sampling theory the "informa-

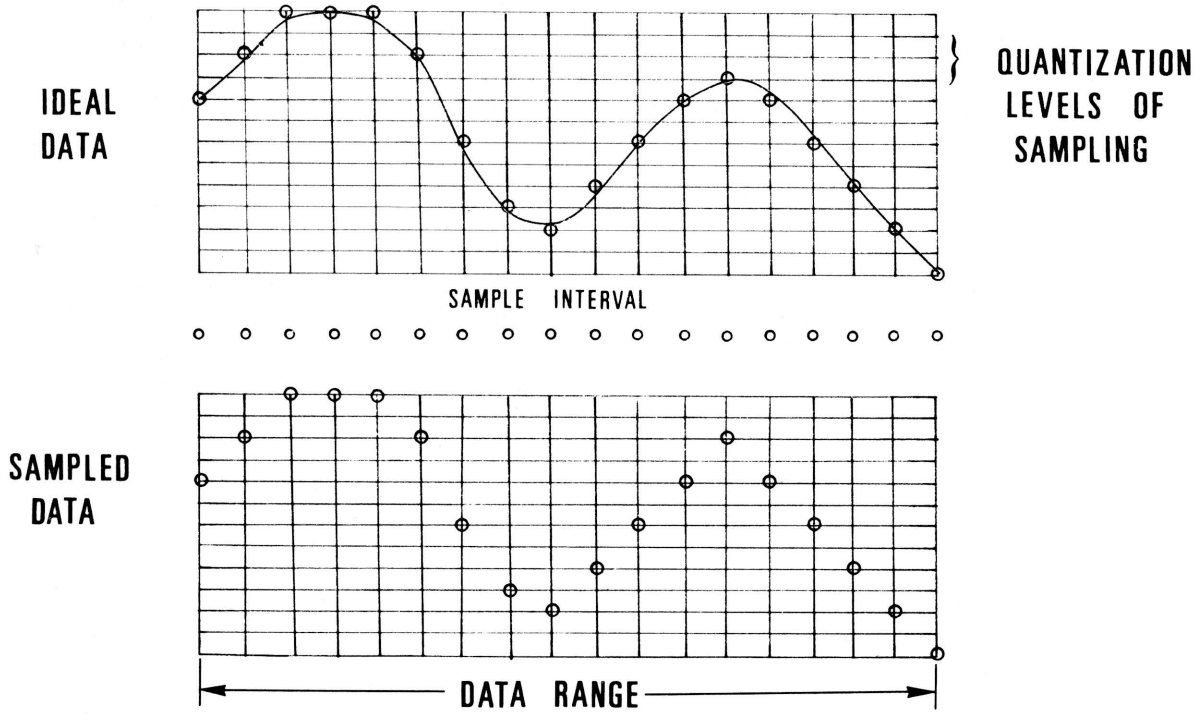


Figure 5. - Ideal and sampled data.

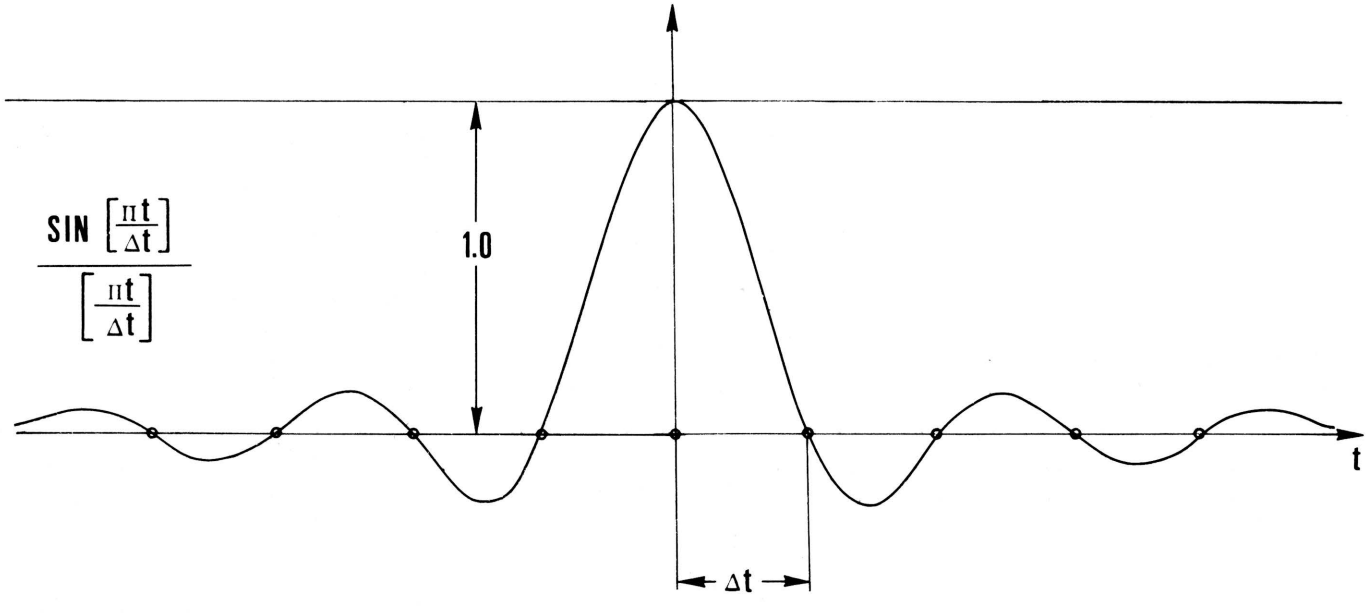


Figure 6. - The Wiener interpolator.

tion" of a signal can be measured and compared in the Fourier frequency domain, this domain is reached of course by the familiar Fourier Transform. If in sampling, none of the significant Fourier components of the ideal data have been deleted or altered, then the sampled data contain all information of the ideal data and the sampling has been accomplished properly. In order to preserve the information content of the sampled data interpolation must be performed using the optimal interpolation function

$$\frac{\sin\left[\frac{\pi t}{\Delta t}\right]}{\left[\frac{\pi t}{\Delta t}\right]}.$$

This function is shown in Figure 6, and Δt is the sampling interval. Note that this "Sine-X-over-X" function is a bonafide interpolation function and if centered on a sample gives it unit weight with zero weight assigned to all other samples.

We use optimal interpolation exclusively in all calculations. There are techniques by which simpler interpolation schemes may be used, but these are usually not worth the added mental effort. As a footnote we call attention to the fact that the optimal interpolator as specified handles only regularly sampled data. There does not exist now a unified and comprehensive body of theory for nonuniformly sampled data.

Figure 7 depicts the Fourier amplitude spectrum of some ideal data. Sampling theory tells us that the sample interval Δt limits the highest frequency in the data which can be properly represented. This is the Nyquist limit. In angular frequency units, the Nyquist frequency W_N is

$$W_N = \frac{\pi}{\Delta t}.$$

Figure 7 shows a zone in which any position would be an acceptable Nyquist limit.

Suppose for example, that the Nyquist frequency is at A as shown in the Figure 7. Clearly all of the frequency components of the ideal data with significant amplitudes will be represented properly by the sampling. Any frequency components beyond A are folded back into the range 0 - A accordian style; this is the high frequency aliasing phenomenon. In the situation under study no components of any consequence would be folded back, hence high frequency aliasing may be ruled out as a source of sampling errors.

If the data were stretched by a factor of 2 however, the Nyquist frequency would be cut in half owing to its inverse dependence on the sampling rate. $B = A/2$ (Fig. 7) is now the highest frequency being handled properly. Of course the entire frequency spectrum is compressed into the range from 0 to B. However, we must assure that significant components beyond B are not present. These can be introduced by the limits on numerical accuracy in the computations. High frequency aliasing, in this instance, poses a real problem.

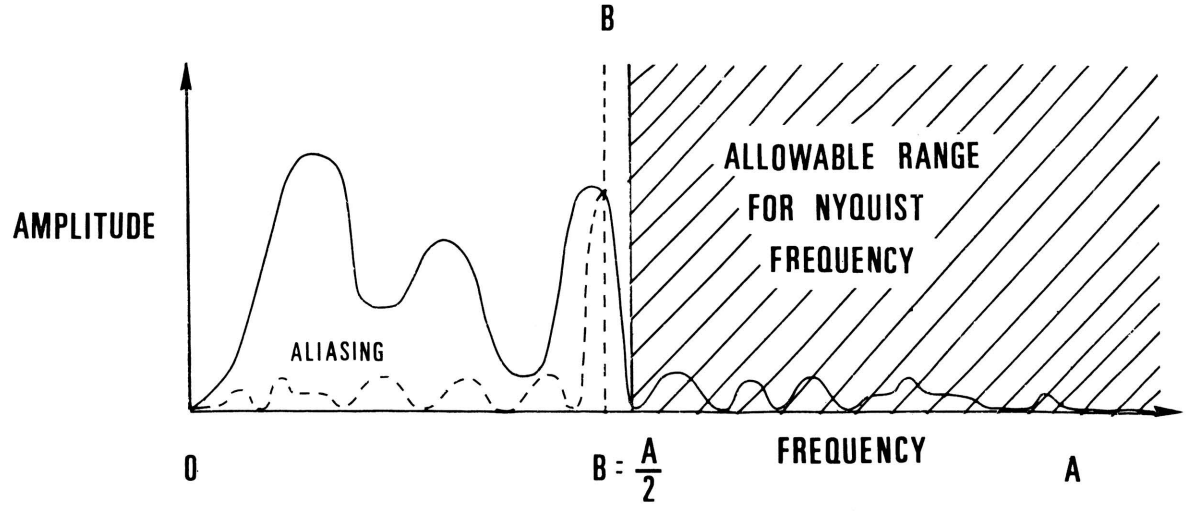


Figure 7. - Fourier transform view of sampling.

The "standard" cure for this type of aliasing is to remove those frequencies which the sampling cannot represent by filtering techniques (Mooney, 1968; Otnes, 1968; Wood, 1968). Hence our computing procedure as a matter of course applies a standard type zero-phase low pass seismic filter which is appropriate for the particular stretch factor. (A zero-phase filter does not change the time origin of

those frequency components passed.) As we indicated earlier, this precautionary digital operation motivated the method of calculation using stretch factors only. It is clear that the larger the stretch factor, the less frequency components and hence data are entering subsequent computations. As a practical matter stretching of time scales by more than a factor of 3 is not usually worthwhile owing to this loss of information.

Returning to Figure 7 we direct attention again to the circumstance when the Nyquist limit is at A. Because all the information content of the data is well below A, simple interpolation schemes similar to second difference methods or linear interpolation may be employed. For other series, however, the situation can be different and necessitates use of the Sine-X-over-X interpolator. Hence in our calculations it was employed routinely.

Lastly, we have been precise in discussing aliasing calling it always high frequency aliasing. As might well be anticipated there is also a low frequency aliasing phenomenon. In this situation the range of the sampling function does not allow adequate representation of low frequency content of the data. Figure 8 shows a typical mathematical form which gives rise to this difficulty - the step function or simple discontinuity. As is stands there is no way to sample properly such a function. Trend removal or high-pass filtering are used in these examples to improve matters. But recognize how common a function the simple discontinuity is in the earth sciences. In fact, the current interest in Markov methods (Lea, 1969, for instance) arises directly from the repeated occurrence of such functions.

In closing this section we should appreciate the origins of our discomforts in handling sampled data. Hence unlike engineering where a problem begins with the statement of an objective, earth science problems begin with the representation of the data themselves.

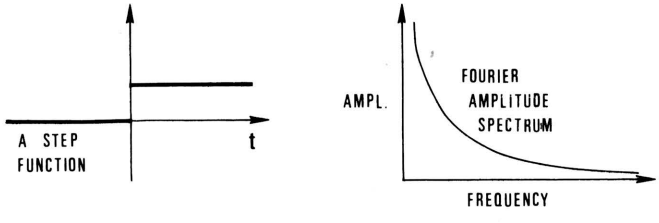


Figure 8. - A sampling problem.

SAMPLE CALCULATIONS

Use of ambiguity functions or equivalently cross-correlation analysis with a stretch parameter, is similar to other mathematical techniques and best appreciated after studying specimen calculations. In this section we shall apply the principles of ambiguity/similarity functions to aeromagnetic data and well-log surveys. No attempt will be made at a serious investigation of the particular data. Rather, emphasis will be placed on understanding the techniques employed, the computed results, and the nature of the questions to be answered.

Figures 9 and 10 each present a trio of data profiles, the magnetic data shown in Figure 9 are selected portions of Project Magnet lines which were flown over the South Atlantic mid-ocean ridge area. Digitized Libyan well velocity surveys are shown in Figure 10. Amplitude scaling in both diagrams is arbitrary as indicated.

The two types of data were selected because their difference in character illustrates the diversity of circumstances which may be encountered. For instance, the well-log data has a high content of high frequency information relative to its sampling rate - so high in fact that one may suspect that the original sampling is aliased. Aeromagnetic data on the other hand has a more regular character dominated by low frequency information - low enough to create problems of low frequency aliasing. Profiles A and B of the magnetic data were arranged such that there appears a small shift in their correlation. Note also the suggestion that the scale of B is stretched with respect to A, although this is difficult to measure owing to its variability. Both of these relationships will be examined further.

As a first step it is a good idea to acquire some experience in stretching data. Figure 11 shows the stretching of magnetic profile A by factors 1.25, 1.50, 1.75 and 2.00. Log profile B is stretched by these same factors in Figure 12. Some basic observations are in order.

Because the stretching is accomplished relative to the center of the data window it is evident that events migrate symmetrically away from this center point as stretch increases. This occurrence necessitates rethinking of what we mean by a shift or lag parameter in ordinary cross-correlation. Unless the data window includes only a single feature being stretched about an axis through its centroid, the lag or shift scale must be compensated by dividing by the stretch parameter. When interpreting ambiguity functions this principle must be kept in mind.

Figure 12 showing the log data also illustrates nicely the effects of frequency, content on appearance. Halving the data's frequency content by applying a stretch of two markedly changes the appearance of the data. Trends and low frequencies are affected by stretching in an interesting manner. A linear trend for instance will have its slope halved by a stretch of two. In effect, the basic low frequency component has its wavelength doubled, however, its amplitude is cut in half. Hence the net effect of stretching on the low end of the frequency spectrum is difficult to assess.

It is becoming clear that the calculation and interpretation of correlation analyses using stretch

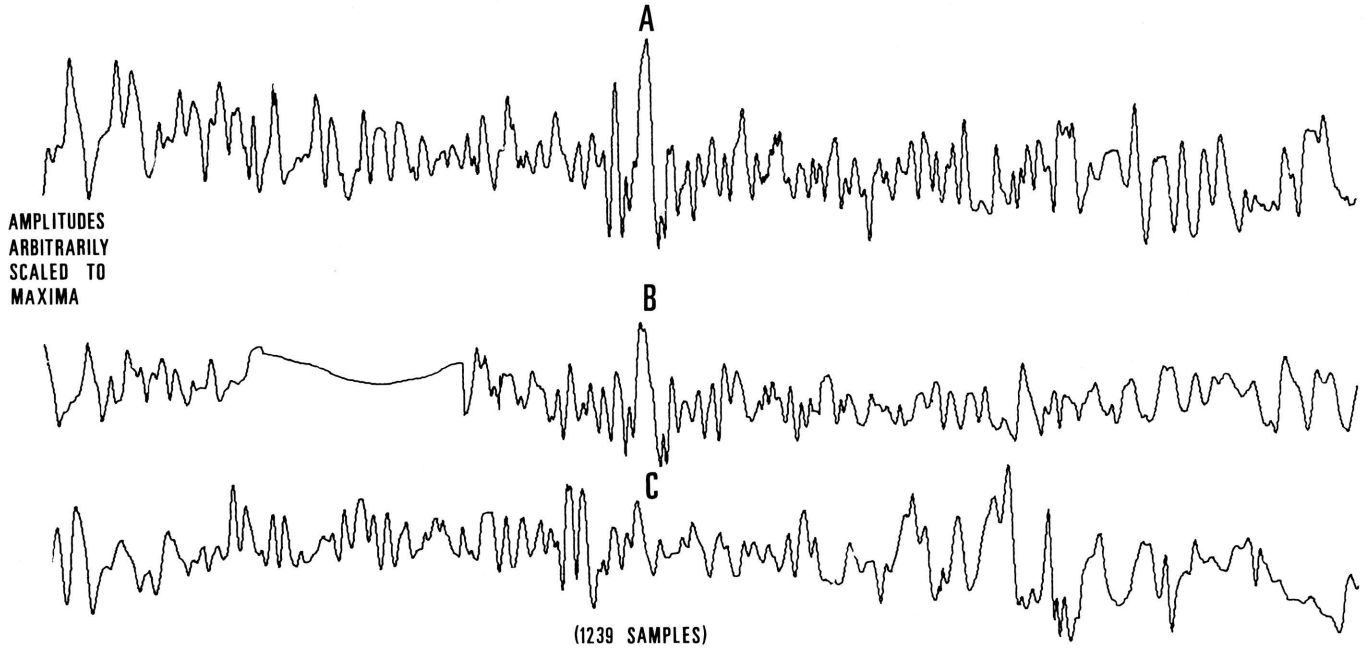


Figure 9. - Segments of Project Magnet profiles from South Atlantic Ridge area.

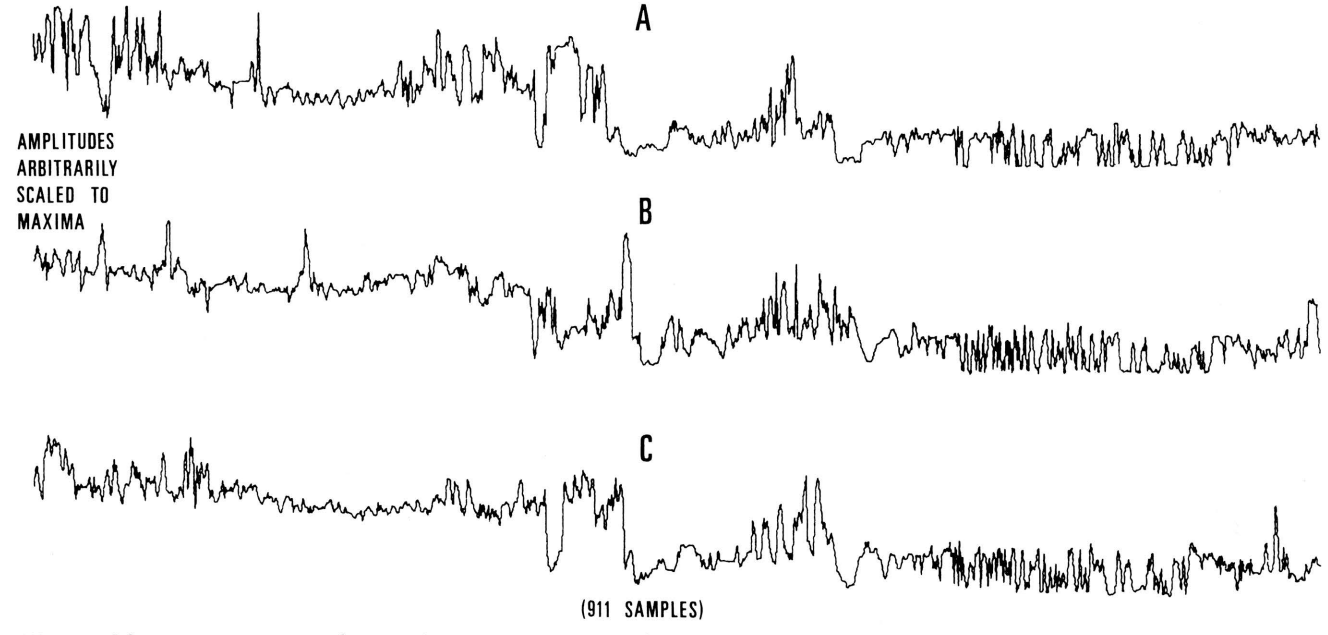


Figure 10. - Digitized velocity log data from neighboring wells in Libya.

parameters is itself a legitimate and fertile area for study. Observations up to this point indicate that shifting the data to eliminate a lagged relationship might well be in order before computing ambiguity or similarity functions for determining any stretching. Of course, this approach necessitates identifying the proper lag before stretching. This may or may not be possible depending on the individual problem and data.

Now let us look at a preliminary ambiguity calculation made between magnetic profiles A and B. The computed results are displayed in Figure 13. So much low frequency content was present in the data that means had to be removed from the cross-corre-

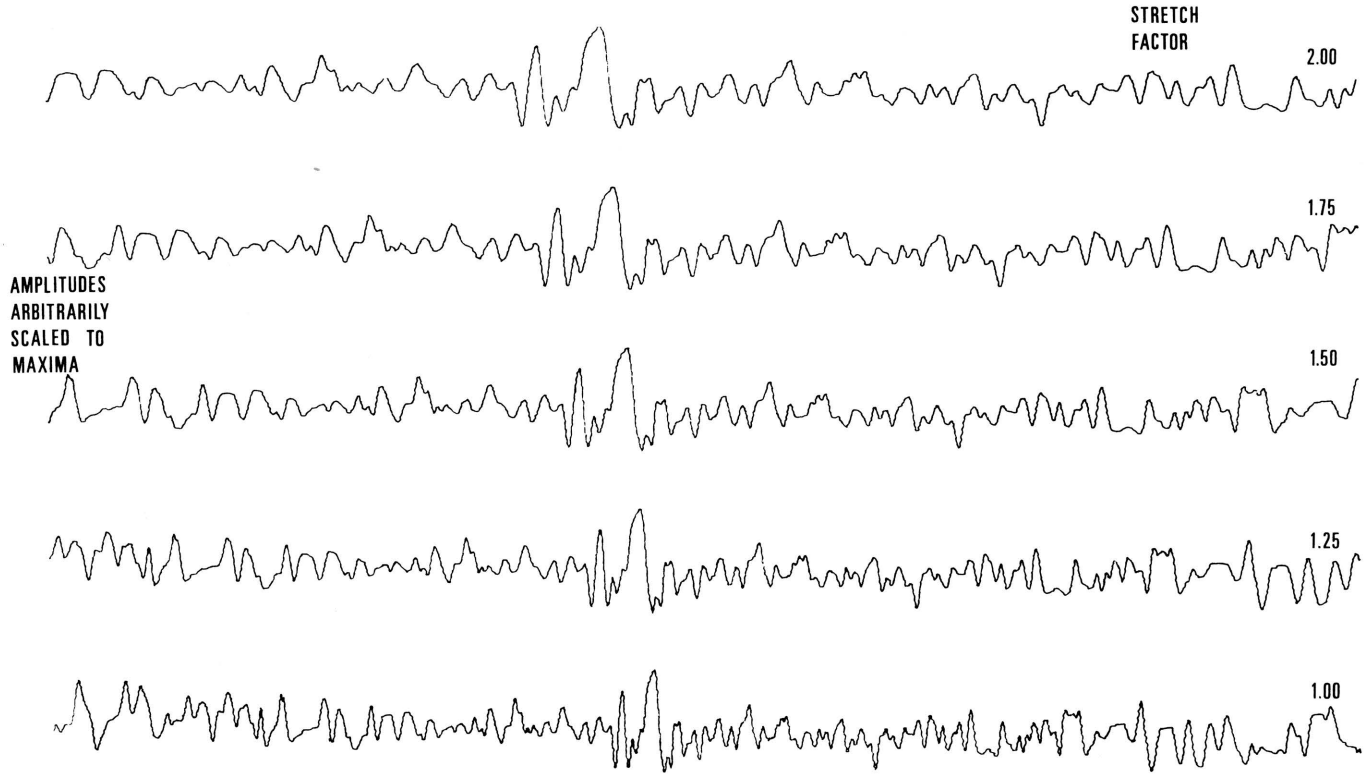


Figure 11. - Stretching magnetic profile A.

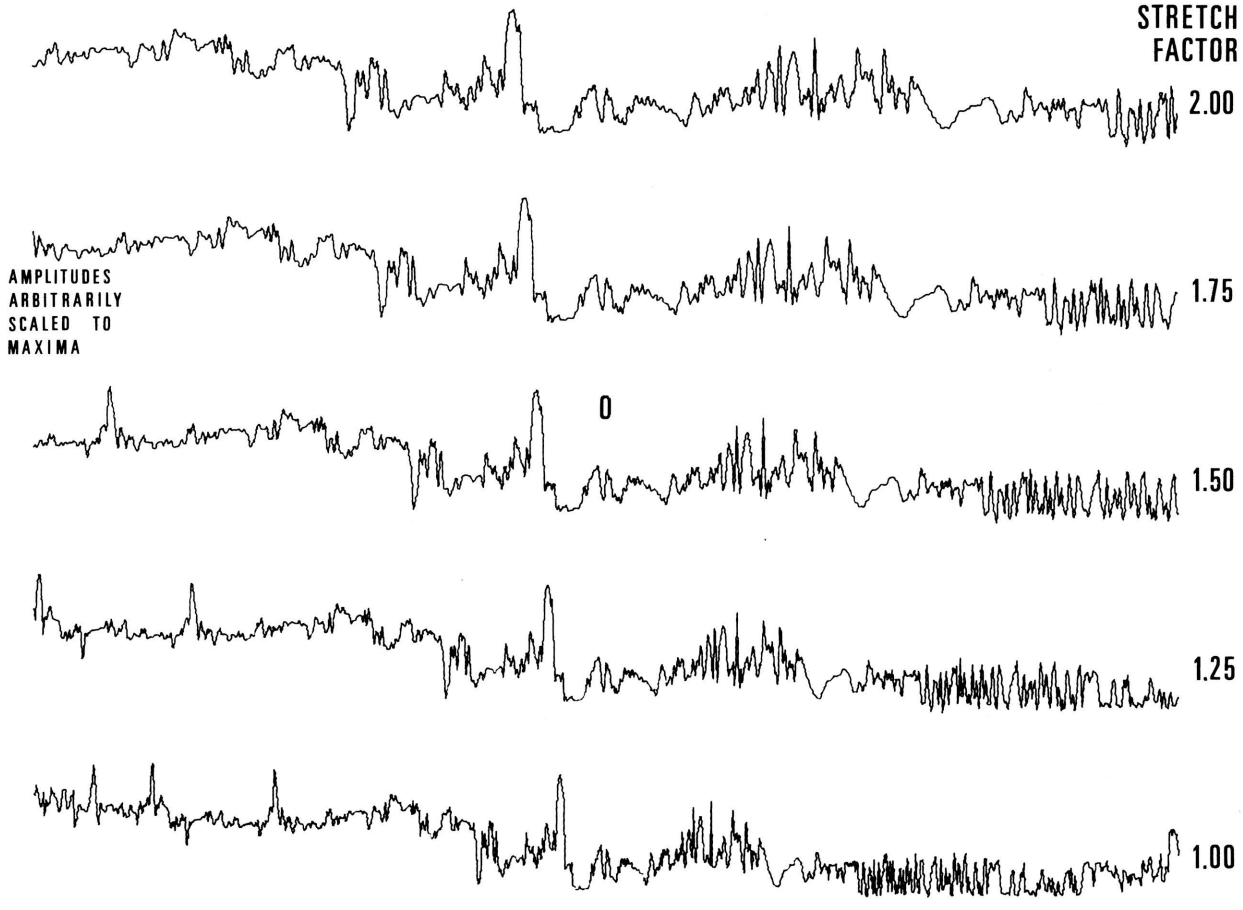


Figure 12. - Stretching magnetic profile B.

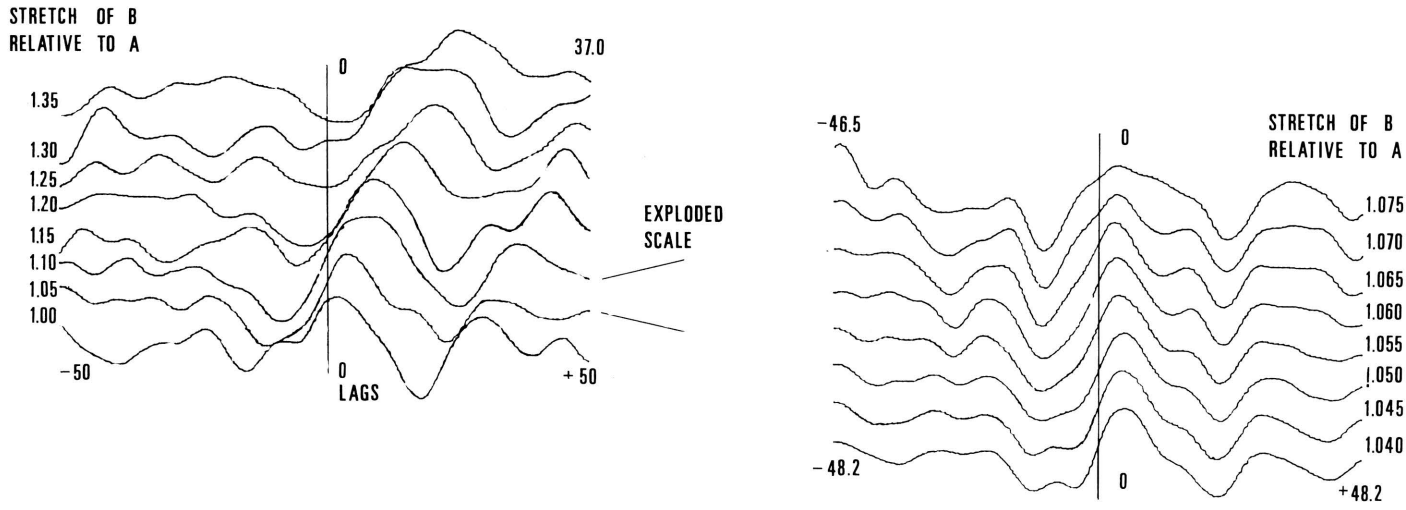


Figure 13. - Preliminary ambiguity study of magnetic profiles A and B.

lations before they could be conveniently displayed.

Two scales of stretching were used in the calculations in order to illustrate the sensitivity of the method. Profile B was stretched with respect to A in increments of 5 percent to 35 percent (a stretch factor of 1.350). Then B was stretched from 4 to 7.5 percent in increments of 0.5 percent. Our purpose was to note how the correlation character varied with the stretch. It would be unwise to draw general conclusions from the study of a single type of data, however some broad principles appear to be indicated.

Because the ambiguity functions were not normalized, recall there is some argument to be resolved concerning this matter - comparisons of peaks must be made within the correlation function corresponding to a single stretch and not among correlations appertaining to other stretches. Two suitable and simple criteria for rating correlation peaks would be to measure their width or sharpness of resolution and to note the percentage by which the maximum exceeds any secondary correlation.

If we apply these criteria to the study at hand, the coarse scale calculations suggest a 5 percent stretching of B relative to A. This figure is reinforced in a quantitative sense by examining the finer scale calculations. We have here direct evidence of a resolving power in the calculation of at least $\frac{1}{2}$ percent in stretching. It seems likely therefore that stretches of 0.1 percent ought to be measurable by this same technique if the variation of stretch in the data themselves is not too unreasonable.

Lastly, we mention the identification of the proper lag or data shift. Where it is possible to study the evolution of a correlation peak relating to a lag, its centroid at best correlation according to our two criteria (resolution and relative height) is probably the best lag estimate. The variability of shift and stretch no doubt interact, hence as we indicated earlier, shifting the data before computing ambiguity functions may have advantages in the parameter estimation. These considerations led the author to provide flexible subroutines rather than a rigid computational package. Displaying the results also deserves mention. There are situations where contouring the two-variable surface would be appropriate.

We have reached the stage where there is little more that can be said without referring directly to the features of the particular data in question. As this becomes the domain of specialists we shall close our discussions here. Many doors have been set ajar and the need for further work is clear. For example, the entire field of a cross-spectral analysis appropriate to this extended notion of correlation lies virtually untouched. Subsequent studies should be convenienced in some measure by the computer subroutines provided by this work.

SUMMARY

This paper has presented and discussed the principles and theory requisite for calculating ambiguity and similarity functions. Basically, these are nothing more than generalizations of the cross-correlation (or cross-covariance) function where a scale stretch (or shrink) parameter has been introduced. The similarity function employs a range limiting parameter to allow for inconsistent trends.

Use of these functions makes it possible to attach a quantitative meaning to the geological notion of correlation. The inconsistency between the mathematical - statistical definition of correlation and the more general intuitive meaning of the same word is eliminated.

Synthetic and real calculations were discussed to illustrate the method. Several of the key subroutines used in the computing are also listed and explained. It is hoped that more meaningful works will be stimulated by this exposition and clarification of ideas.

ACKNOWLEDGMENTS

I would like to thank Mr. M.T. Taner of Seiscom for moral support and valuable discussions about

the theory and computing. Mr. R.L. Clark also rendered valuable programming assistance. Dr. J. Mudie of the Scripps Institute of Oceanography provided helpful comments about the magnetic data. The Project Magnet data were contributed by Dr. J. Morgan of Princeton University whose discussions were appreciated.

REFERENCES

Cox, A., 1969, Geomagnetic reversals: Science, v. 163, p. 237-245.

Dickson, G.O., Pitman, W.C., III, and Heirtzler, J.R., 1968, Magnetic anomalies in the South Atlantic and ocean floor spreading: Jour. Geophysical Res., v. 73, no. 6, p. 2087-2100.

Gassner, R.L., and Cooper, G.R., 1967, Note on a generalized ambiguity function: I.E.E.E. Trans. Info. Th., v. IT-13, p. 126.

Godby, E.A., Hood, P.J., and Bower, M.E., 1968, Aeromagnetic profiles across the Reykjames Ridge southwest of Iceland: Jour. Geophysical Res., v. 73, no. 24, p. 7637-7649.

Lea, G., 1969, Markov forecasting techniques in exploration (abs.): Am. Assoc. Petroleum Geologists Bull., v. 53, no. 3, p. 728.

Lee, Y.W., 1968, Statistical theory of communication: John Wiley & Sons, New York, 509 p.

Mooney, H.M., 1968, Pole-and zero design of digital filters: Geophysics, v. 33, no. 2, p. 354-360.

Morgan, W.J., 1968, Rises, trenches, great faults and crustal blocks: Jour. Geophysical Res., v. 73, no. 6, p. 1959-1982.

Neidell, N.S., 1968, Data processing for controlled energy acoustic sources: UnderSea Tech., reprint R231.

Otnes, A.K., 1968, Recursive bandpass filter design: Proc. I.E.E.E., v. 56, no. 2, p. 207-208.

Purdy, R.J., and Cooper, G.R., 1968, A note on the volume of generalized ambiguity functions: I.E.E.E. Trans. Info. Th., v. IT-4, no. 1, p. 153-154.

Taner, M.T., Cook, E.E., and Koehler, F., 1968, On processing Vibroseis data: Soc. Expl. Geophysicists, preprint.

Vine, F.J., and Matthews, D.H., 1963, Magnetic anomalies over oceanic ridges: Nature, v. 199, p. 947.

Wilson, J.T., 1965, A new class of faults and their bearing on continental drift: Nature, v. 207, p. 343-344.

Wood, L.C., 1968, A review of digital pass filtering: Rev. Geophys., v. 6, no. 1, p. 73-97.

APPENDIX - THE "NITTY-GRITTY" OF COMPUTING

The computation of ambiguity functions (and similarity) functions can be accomplished by using the subroutines included. Here we present an overview of one way in which the subroutines may be pieced together.

a (constant) STRTCH	Set stretch parameter (greater than 1 please!) and decide which series is stretched relative to the other.
RSMPL	Stretch data and resample at former interval. This subroutine calls SINXOX - optimal interpolation subroutine.
FILTGEN	Generate antialiasing filters appropriate to value of STRTCH.
CORFLT	Filter data after stretching using filter created by FITGEN.
CORFLT	Correlate streiched-resampled output of RSMPL with undeformed series.

CALL CORFLT (A, NA, B, NB, C, NC)

NA, NB, NC must be odd

A Input Series	NA Terms
B Input Series or Filter	NB Terms.

If NA = NB Correlation is Performed

C Output Series	NC Terms (Lags go from $-(\frac{NC-1}{2})$ to $(\frac{NC+1}{2})$)

If NA $\neq$ NB Unnormalized Correlation (time reversed convolution) is Performed.

C Output Series	NA Terms.

CALL SINXOX (I, D, NT, AA, SUM)

NT must be even

I	Index of AA array just smaller than desired Interpolation position
D	Desired Interpolation position (Measured from the start of the series 1.0)
NT	Number of values of interpolator. An even number usually about 10% of length of AA
AA	Series for Interpolation
SUM	Interpolated Value.

Warning: In calling SINXOX it is assumed that AA is loaded on the front end and tail end with zeros. Note the proper calls to SINXOX made in RSMPL.

CALL RSMPL (A, NA, B, STRTCH)

NA must be odd

A	Input Series NA Terms
B	Output Series NA Terms
STRTCH	Stretch Parameter. Greater than 1.0 Please.

CALL FILTGEN (STRTCH, FF, NF)

NF must be odd

FF	Output Filter Series NF Terms
STRTCH	Stretch Parmameter.

```
    1:       SUBROUTINE CORFLT (A,NA,BB,NB,C,NC)
    2:       DIMENSION A(2000),BB(2000), C(2000),AA(2500)
    3:       COMMON    AA
    4:       IF (NA.EQ.NB)  LIM = NC; GO TO 600
    5:       LIM = NB
    6:   600 LL = LIM/2
    7:       DO 700 I = 1, NA+LIM-1
    8:       IF((I.LE.LL).OR.(I.GT.(NA+LL)))AA(I) = 0; GO TO 700
    9:       AA(L) = A(I-LL)
   10:   700 CONTINUE
   11:       LM = NA
   12:       IF(NA.EQ.NB) LM = NC
   13:       DO 900 I = 1,LM
   14:       SUM = 0.0
   15:       DO 800 J = 1,NB
   16:   800 SUM = SUM +BB(J)*AA(I+J-1)
   17:   900 C(I) = SUM
   18::      NC = LM
   19:       IF (NA.NE.NB ) GO TO 1200
   20:       DO 1000 I =1,LM/2 +1
   21:       U = NB - (LM/2 + 1) + I
   22:       C(I) = C(I)/U
   23:       IF ( I.EQ.LM/2 +1 ) GO TO 1200
   24: 1000 C(LM - I + 1 ) = C(LM -I + 1 )/U
   25: 1200 RETURN
   26:       END
***GLOBAL VARIABLES***
***LABELED COMMON BLOCKS***
***ENTRY POINTS***
00000 CORFLT
***EXTERNAL REFERENCES***
00001 9SETUPN          00205 9IDOSET
```

```
    1:       SUBROUTINE SINXOX (I,D,NT,AA,SUM)
    2:       DIMENSION AA(2500)
    3:       L = NT/2
    4:       SI =I
    5:       S = SI - D
    6:       IF (ABS(S) .LE.0.0001) SUM =AA(I);GO TO 250
    7:       T = 1.0+S
    8:       PI = 3.1415927
    9:       SUM = 0.0
   10:       DO 200 J = 1,L
   11:       SJ = J-1
   12:       TU =AA(I-J+1)*SINF(PI*(S-SJ))/(S-SJ)
   13:       TD = AA(I+J)*SINF(PI*(T+SJ))/(T+SJ)
   14:   200 SUM = SUM + (TU+TD)/PI
   15:   250 RETURN
   16:       END
***GLOBAL VARIABLES***
***LABELED COMMON BLOCKS***
***ENTRY POINTS***
00000 SINXOX
***EXTERNAL REFERENCES***
00001 9SETUPN          00057 9IDOSET         00127 SINF
```

```
    1:       SUBROUTINE RSMPL (A,NA,C ,STRTCH)
    2:       DIMENSION A(2000),AA(2500),                C(2000)
    3:       COMMON AA
    4:       IF ( ABSF(STRTCH).LE.1.0) STOP
    5:       NINT = 2*(NA/20)
    6:       MID = NA/2 + 1
    7:       NUM = MID-1
    8:       NIN = NINT/2
    9:       NR = MID + NIN
   10:       SR = NR
   11:       DO 500 I = 1,NA+NINT
   12:       IF((I.LE.NIN).OR.(I.GT.(NA+NIN)))AA(I) =0; GO TO 500
   13:       AA(I) = A(I-NIN)
   14:   500 CONTINUE
   15:       DO 700 I = 1,NUM
   16:       SU = I
   17:       DD = SU/STRTCH
   18:       ID = DD
   19:       DU = SR - DD
   20:       IDU = DU
   21:       CALL SINXOX (IDU,DU,NINT,AA,SUM)
   22:       C (MID-I) = SUM
   23:       IDD = NR + ID
   24:       DD = SR + DD
   25:       CALL SINXOX (IDD,DD,NINT,AA,SUM)
   26:   700 C (MID+I)= SUM
   27:       C (MID) =AA(MID+NIN)
   28:       RETURN
   29:       END
***GLOBAL VARIABLES***
***LABELED COMMON BLOCKS***
***ENTRY POINTS***
00000 RSMPL
***EXTERNAL REFERENCES***
00001 9SETUPN          00016 ABSF           00025 9STOP
00202 SINXOX
```

```
    1:       SUBROUTINE FILTGEN (STRTCH,FF,NF)
    2:       DIMENSION FF(250)
    3:       PI = 3.1415926 / STRTCH
    4:       NN = NF/2
    5:       DO 600 I = 1,NN
    6:       SU = I
    7:       FF(NN+1+I) = SINF(PI*SU)/(PI*SU     )
    8:   600 FF(NN+1-I) + FF(NN+1+I)
    9:       FF(NN+1) = 1.0
   10:       RETURN
   11:       END
***GLOBAL VARIABLES***
***LABELED COMMON BLOCKS***
***ENTRY POINTS***
00000 FILTGEN
***EXTERNAL REFERENCES***
00001 9SETUPN          00025 9IDOSET        00047 SINF
```

DISTRIBUTION OF HYDROCARBONS IN THREE DIMENSIONS

by

John C. Davis [1]

ABSTRACT

The areal distributions of hydrocarbons and other components of rock bodies have been mapped in an attempt to relate their distribution to economic concentrations of petroleum and other components. These components are dispersed three-dimensionally; their vertical distribution may be as important as (or more important than) their areal concentration. Distribution patterns in three dimensions are difficult to portray; this has handicapped efforts to relate the distribution patterns to economic drilling objectives.

Response surface analysis provides a rapid method of displaying three-dimensional relations within rock bodies. The variable of interest—for example, percent organic carbon—is regressed upon a linear combination of the three geographic axes. The resulting linear equation is a least-squares expression of the relation between the dependent variable and the spatial coordinates. In practive, a polynomial expansion of the linear equation commonly is used to provide a better representation of the data. Other linear equations may be more appropriate in specific cases; trigonometric functions, for example, may be introduced to simulate the effect of bedding.

Measurements of organic carbon from the Lower Cretaceous Mowry Shale in Wyoming provide an illustration of the effects that can be obtained using different linear models in the regression. Relations between the distribution patterns and location of Lower Cretaceous hydrocarbon fields in this region can be shown graphically on the models.

INTRODUCTION

There are two principle philosophies of petroleum exploration; one is the older more traditional approach in searching for petroleum traps. Historically, exploration geologists have sought anticlines, fault structures, or stratigraphic pinchouts in which hydrocarbons could accumulate. The opposing operational philosophy is searching for source beds or rock units which could be originators of petroleum. Although the trap-seeking approach has been used extensively and has proved to be of undeniable utility, examination of source areas has become more prominent in exploration in recent years. In older productive areas such as the United States, where the more obvious structural traps have been drilled and the cost of exploring is great, the source rock approach is becoming more prevalent. It is expensive to prospect in the United States and a driller has only slim chances for success; therefore it is imperative that he concentrate his resources in those areas which are most apt to be petroliferous. If traps are found only with the greatest difficulty, it is important that those traps which are found do indeed contain oil.

Unfortunately, the exploration for source rocks and the study of hydrocarbon migration routes is considerably less straightforward than the search for stratigraphic or structural traps. Our knowledge of the nature of source beds is comparatively meager; it can be summarized by the statement that they seem to be marine, fine grained, and rich in hydrocarbons.

A research project to identify actual or potential source beds might proceed in the following manner. The stratigraphic section of a known petroliferous province or basin would be examined, generally with most concentration placed on the more basinward facies. Fine-grained units would be sampled carefully, especially if they were black or showed other evidence of having unusual hydrocarbon content. Once organic-rich units were identified, a systematic sampling and geochemical analysis project would be undertaken. All this research effort requires tremendous amounts of money and it is important that the maximum amount of information be obtained from the analyses.

There are many problems associated with this type of study. For one, we are attempting to trace extremely subtle changes in composition through three-dimensional space; space which we can sample only incompletely. The patterns which we see emerging from the analyses are difficult to visualize as they are complex and contain error introduced by the inexactness of our sampling and chemical analyses. Finally, whatever patterns we perceive or impose upon the data must be made to correlate with our geologic knowledge of the nature of the unit, if the results are to have any meaning to the geologist who is conducting exploration.

TREND SURFACE ANALYSIS

Geologic trend-surface analysis is a computer-based technique which has been proven valu-

[1] Kansas Geological Survey, The University of Kansas, Lawrence, Kansas

able in exploration for structural traps (Merriam and Harbaugh, 1963, 1964). A simple extension of the procedure potentially is equally useful in problems associated with exploration for source rocks. The technique is called three-dimensional trend-surface analysis (Harbaugh and Merriam, 1968). Its application to petroleum source exploration will be illustrated using an example from the Lower Cretaceous of the Rocky Mountains. First however the technique of trend analysis should be explained for the benefit of those not familiar with it (see especially Draper and Smith, 1966).

It may help to understand the method if we reduce the number of variables which we are considering. Figure 1 shows one variable, X, represented along the vertical axis of the plot. Of all possible values within the range of the data, the mean, or average, shown by the heavy line and symbolized by $\overline{X}$, is the best estimate of X. It is the number about which there is the smallest variance. That is,

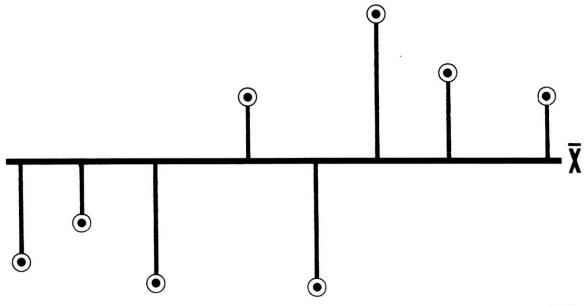


Figure 1. - Points distributed about the mean, $\overline{X}$.

the spread of data around the mean is smaller than the spread around any number.

We now introduce a second variable and to conform to statistical conventions we will refer to these as X and Y. We are concerned principally with the variation in Y and are interested in describing the behavior of Y as X changes. This type of problem is called a regression and can be defined simply as "finding the line which most efficiently describes the variation in Y for a given X". Shown in Figure 2 is the line of regression of Y on X and, similar to the mean,

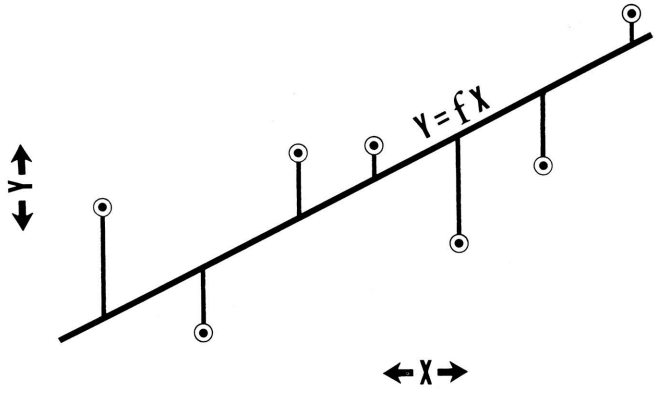


Figure 2. - Points distributed about regression line.

it possesses several useful characteristics. This line is calculated by least squares, which insures that variation about the line is the least possible. The regression line gives the most efficient estimates of Y for any given values of X.

If we now extend our consideration to three variables (Fig. 3), which we will call Y, X_1, and X_2, we have a problem which may be familiar to you as a trend analysis. Instead of computing the regression line of Y on X, we are now interested in a regression or trend surface of Y on X_1 and X_2. The extension from the line shown previously to this surface should be obvious. We now are computing the plane from which the squared deviations of Y will be a minimum. The computational procedures are a straight-forward extension of methods used to find the regression of Y on X. In geological parlance, this is known as a first-degree trend surface.

Until this point we have not discussed what the variables in the diagrams might be. In trend-surface analysis, X_1 and X_2 are the geographic coordinates of a sample point. If we are concerned with structure, the dependent variable Y is depth or elevation of some stratigraphic horizon. What we are in essence saying with trend-surface analysis is that the elevation of a horizon is in part a function

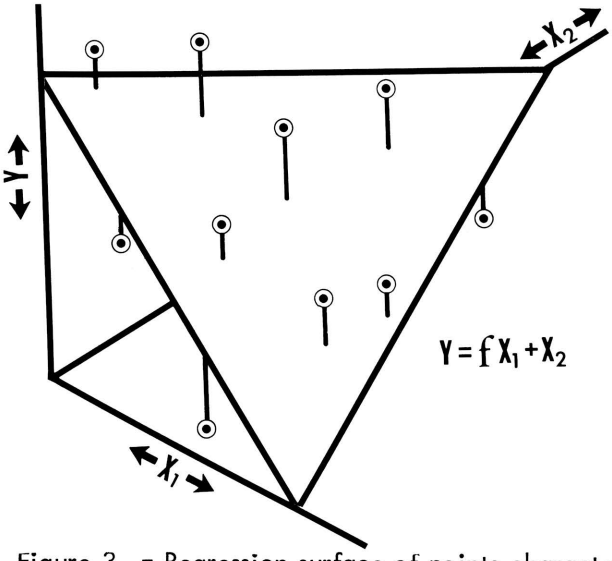


Figure 3. - Regression surface of points characterized by two independent variables.

of its geographic location. Whether this is true or not in some absolute sense is not important. What is important is that we can consider this to be the situation and we can experiment within this operational definition.

In Figure 4 we have carried the concept one step further. We have three independent axes, X_1, X_2 and X_3. For our purposes we may consider the three axes as geographic coordinates; X_2 is east-west, X_3 is north-south and X_1 is depth. We now have a fourth variable for the dependent part of our equation.

The dependent variable may be some element of composition at a particular sample point identified by its spatial coordinates (Peikert, 1965). The regression does not appear as a line or plane as it did in previous examples, but rather as what has been called a hypersurface or hyperenvelope (Harbaugh, 1964; Smith and Harbaugh, 1966).

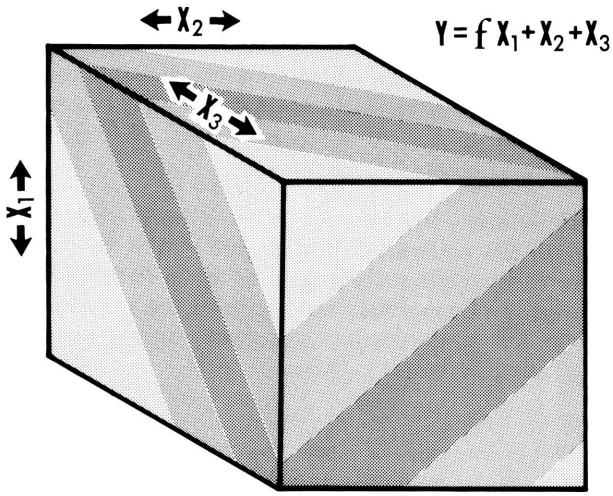


Figure 4. - Regression envelope around points having three independent variables.

Contour lines may have the form of sheets or sleeves which enclose volumes of equal composition. The volumes of equal composition are denoted by shading. Just as the area between two successive contour lines is occupied by points which have the same range in elevations, so the volume between two successive contour sheets in this diagram is occupied by points which have the same range in composition.

All illustrations have been of linear equations. That is, they graph as straight lines. We can introduce flexibility into trend surfaces by adding powers of the coordinates to the equations. This has the effect of creating curvature in the surface. Just as a simple regression can be considered as the best line that could be drawn through the data with a ruler, more complex equations produce the best lines that could be drawn with a French curve. In trend analysis, we are fitting the data with multidimensional French curves.

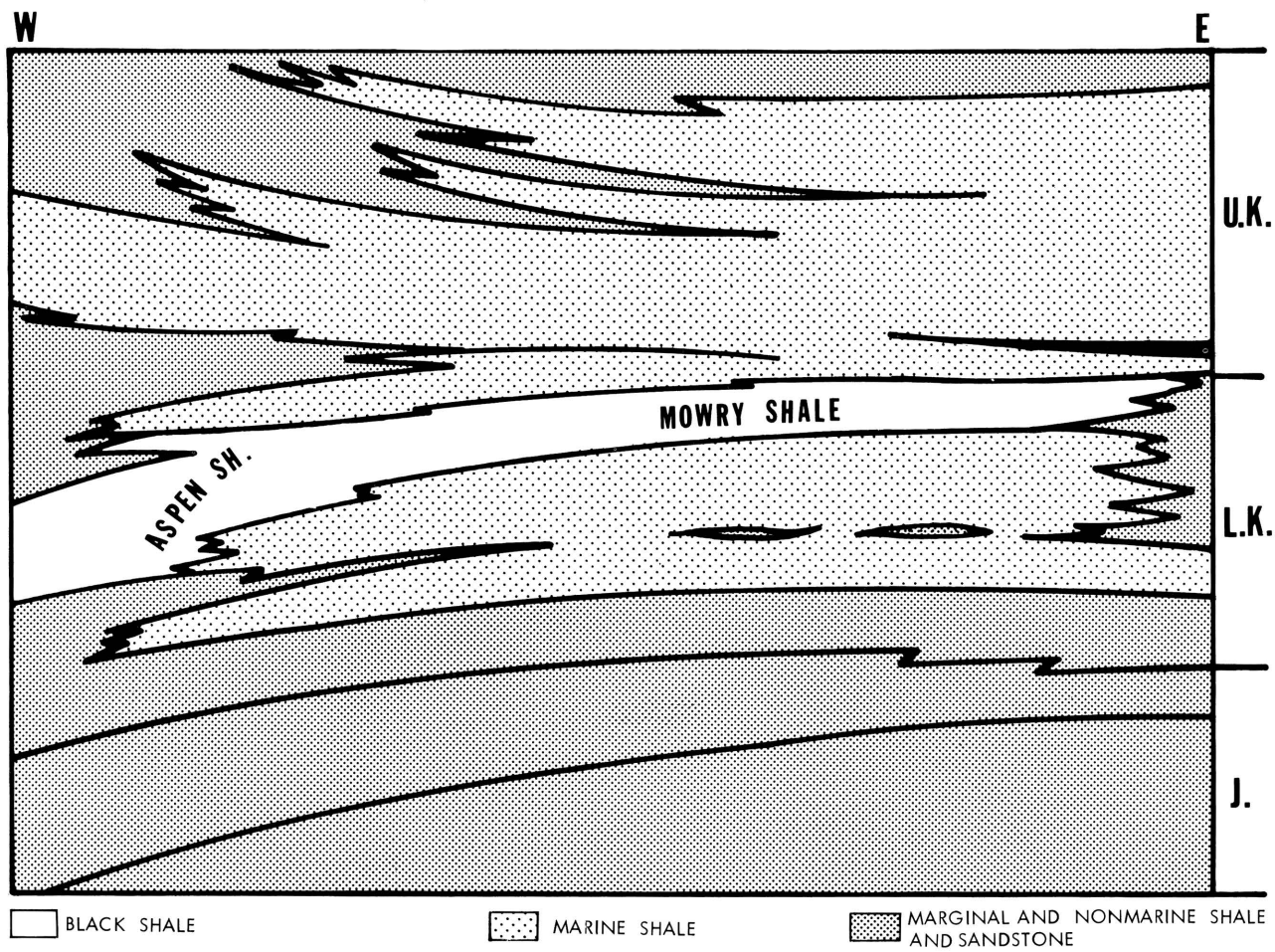


Figure 5. - Generalized cross section of Wyoming showing relation of Mowry Shale to other stratigraphic units.

GEOLOGIC EXAMPLE

I would like to turn now to the geologic example which will illustrate the application of this technique. Figure 5 is a highly diagrammatic cross section from west to east across central Wyoming (Burk, 1956; Mills, 1956; Haun and Barlow, 1962). The fine stipple pattern outlines the terrestrial or coarse clastic sediments, whereas the coarse stipple shows units which are marine and generally fine grained. The formation we are interested in is the Mowry Shale. The Mowry and its equivalent in western Wyoming, the Aspen Shale, is a black, dense siliceous shale (Davis, 1963). One of the most notable attributes of the unit is that it is extremely rich in organic carbon.

Gulf Research and Development Company undertook an extremely extensive source rock study of this unit (Shrayer and Zarrela, 1963, 1966). The objective of the study was to examine the relationship between organic carbon and hydrocarbons in the Mowry Shale and the location of Lower Cretaceous producing oil fields. It was postulated that petroleum found in traps in the Muddy Sandstone was drawn at least in part from the overlying Mowry. Gulf released 560 organic carbon analyses for use in this study; the samples had been collected through the Mowry Shale at 32 measured sections in central and western Wyoming. Figure 6 is adapted from one by Shrayer and Zarrela. Contour lines show the general regional variation in organic carbon content within the Mowry. Although the contours conform reasonably well with what we can see in the raw data, a presentation of this type cannot take into consideration vertical variation in organic content in the unit. The map was made by computing weighted averages of the analyses at each sample locality. Averages then were contoured. Unfortunately, the amount of information contained in the map is limited and it is difficult to combine with stratigraphic and geologic information about the unit.

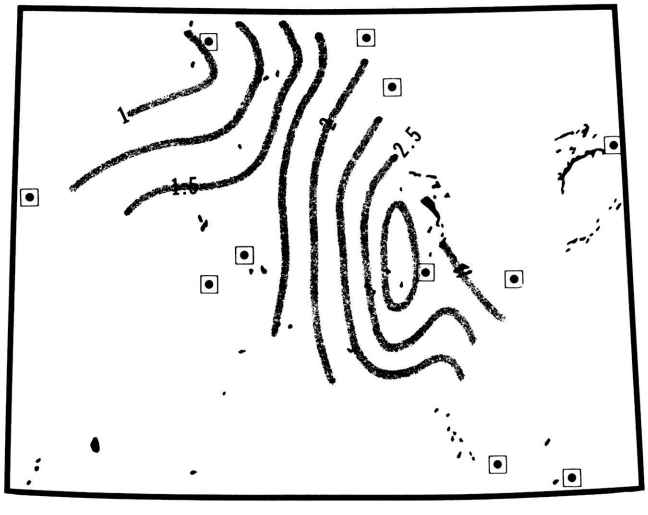


Figure 6. - Two-dimensional distribution of hydrocarbons in Mowry Shale in Wyoming (from Shrayer and Zarrela, 1963).

It was felt that for this investigation, three-dimensional trend-surface analysis might provide a more efficient way of showing the distribution of organic carbon in the Mowry Shale. By using a three-dimensional portrayal, the relationship between organic carbon and geologic parameters which operated within the unit could be seen more clearly.

Lower Cretaceous oil fields are shown in black in their relative geographic positions. The coincidence of fields in central Wyoming with the carbon high found by Gulf was a point of obvious interest. By looking at the distribution of organic material in three dimensions, it was hoped to gain some further insight into this similarity.

The first step was to transfer the carbon data to a coordinate system in which X_1 represents east-west, X_2 north-south, and X_3 the thickness or depth coordinate. The initial investigations used a trend-surface program called KWIKR8 published by the Kansas Geological Survey (Esler, Smith, and Davis, 1968). Some complicated analyses were performed using a program called UFF developed at the Geological Survey by Robert J. Sampson.

Figure 7 is a map of Wyoming, similar to the previous one except that it has been laid nearly flat in an oblique projection. This is the same projection used in the series of block diagrams (Pl. 1-4); because the presence of town names would obscure detail on the diagrams, the index map gives general location of features.

The linear three-dimensional trend surface on organic carbon in the Mowry Shale presents a simple pattern of values which increase upward and to the east. This relatively simple model accounts for only a small amount of the variation in the unit and was not considered adequate.

The second-degree surface was more efficient (Pl. 1). The green folded sheet represents the two-percent organic carbon surface. The blue sheet represents the one-percent surface. The volumes of shale to the right of the green sheet contain organic carbon contents estimated to be greater than two percent. Those between the green and blue sheets con-

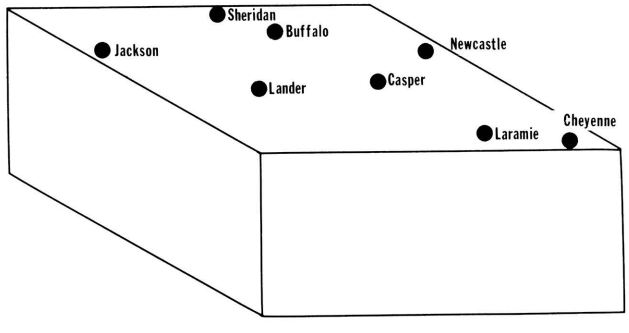


Figure 7. - Index map of Wyoming.

tain estimated contents between one and two percent, and those to the left of the blue sheet contain less than one-percent organic carbon. As you can see, the general form of the trend surface is that of a folded envelope containing higher carbon values which increase upward and eastward. It is significant also that low carbon values are indicated near the bottom of the shale unit.

One of the problems of using polynomials or power terms in trend analysis is that the surfaces cannot form a repetitive design. It is difficult therefore to imitate patterns such as those created in sediments by transgressing and regressing environments. Because variation in organic carbon content in the shale might be dependent upon features related to bedding, any variation parallel to bedding would be smoothed by a conventional trend surface. In an attempt to discern this type of variation, the power terms were combined with trigonometric terms, which are sensitive to wavelike oscillations (James, 1966). This allows for the possibility of cyclic repetitions in concentration through the unit.

A trend surface whose equation contains squared terms in the geographic coordinates and trigonometric terms in the X_3 or depth dimension is shown in Plate 2. As can be seen the trigonometric components have produced a crenulation or fold in the contour sheets, representing an intertonguing relationship in organic carbon content. Each of the two model equations are almost equally efficient in terms of the amount of variation in carbon content that they explain. The model containing trigonometric functions is somewhat more efficient in that fewer terms are necessary in the equation.

In this study two criteria were used for judging the adequacy of a trend surface. One of these was how well the equation accounted for variation in the dependent variable. That is, how much of the total variation in organic carbon could be attributed to the trend. This is a gross estimate of efficiency as it is based on total overall goodness-of-fit of the trend surface.

The second criterion for choosing between alternative equations was the amount the model reduced the volume of autocorrelated residuals. This can be explained with the help of a diagram which shows only two variables, X and Y (Fig. 8). Here the trend line accounts for most of the variation in Y. Data points are scattered uniformly on alternate sides of the regression. In two small areas, adjacent points lie on the same side of the line. One of these areas, indicated by the plus, contains observations greater than the trend. The other area is negative with points below the trend. The pair of adjacent points which deviate from the trend in the same direction are referred to as being autocorrelated. That is, the value of one residual is related to the value of the adjacent residual. A similar set of data points and a regression which fits essentially as well as the previous is shown in Figure 9. However, there are large areas of autocorrelated residuals represented by the plus and minus signs. In the two figures, the line of regression accounts for about the same amount of variation in Y. However, it is apparent in the second example that the line is not a good representation of the data because of the presence of large areas of autocorrelated residuals. This is a simplification of the second criterion used to judge the models. In the trend surfaces, models containing trigonometric terms definitely are superior to polynomial models in reducing the size of autocorrelated residuals, although the two alternatives produced almost identical goodnesses-of-fit.

Both second-degree models produce fits which are significant statistically, but a more realistic picture of variation of organic carbon is produced using more complex models. Plate 3 shows the trend surface formed by an equation containing cubed terms. The color convention is the same as in previous diagrams. A significant difference between this and previous models is that carbon concentration drops both to the east and to the west. The zone of high organic carbon content moves eastward at higher stratigraphic

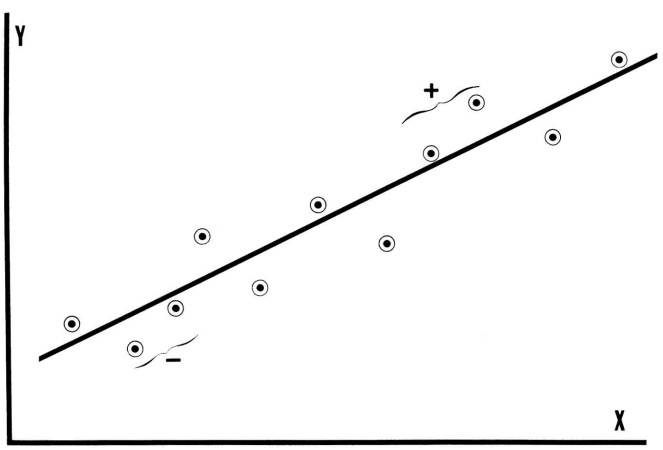


Figure 8. - Regression of Y and X with small autocorrelated areas.

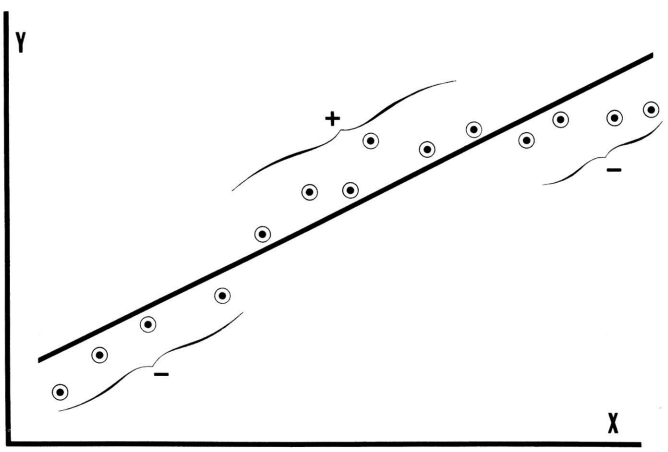
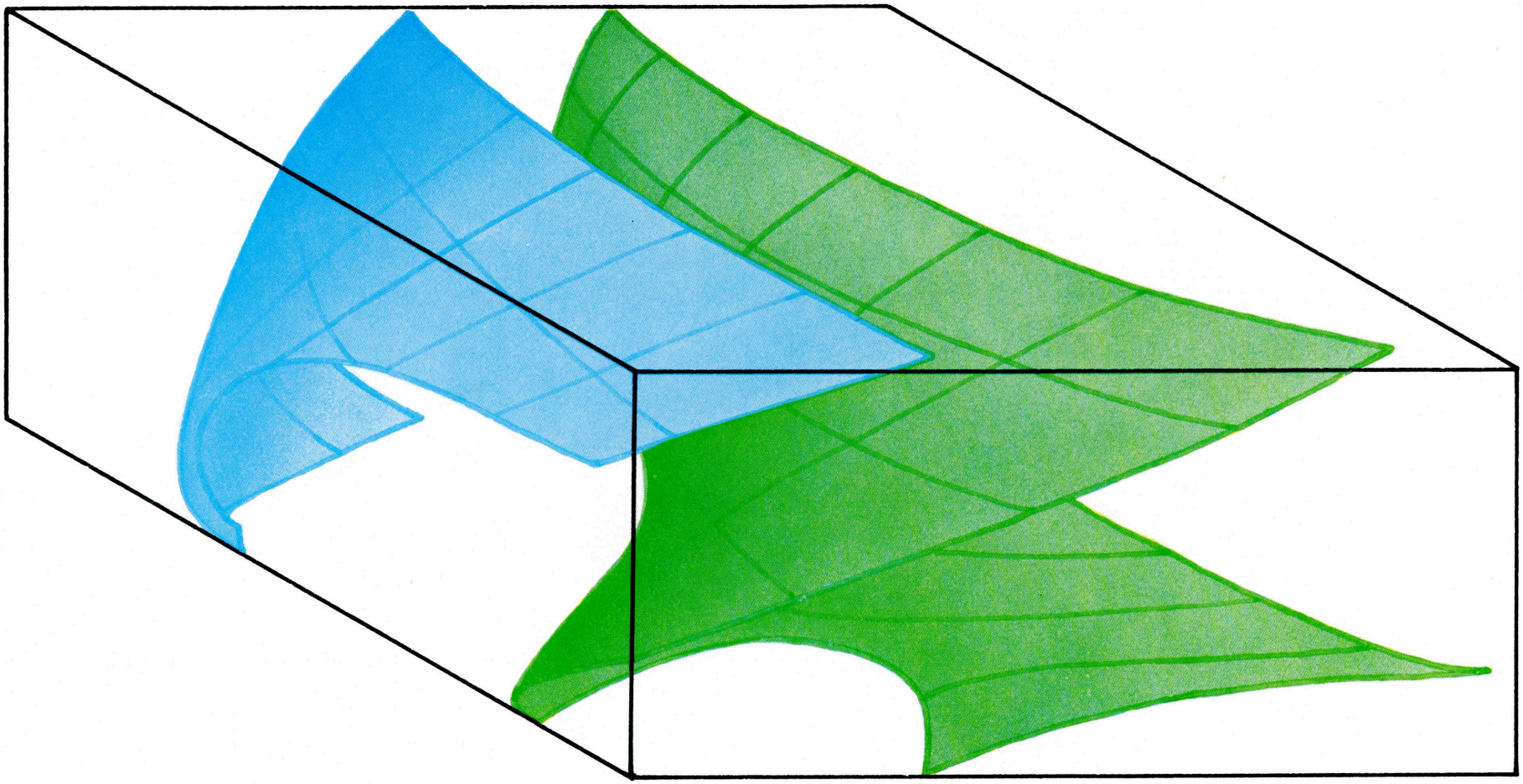
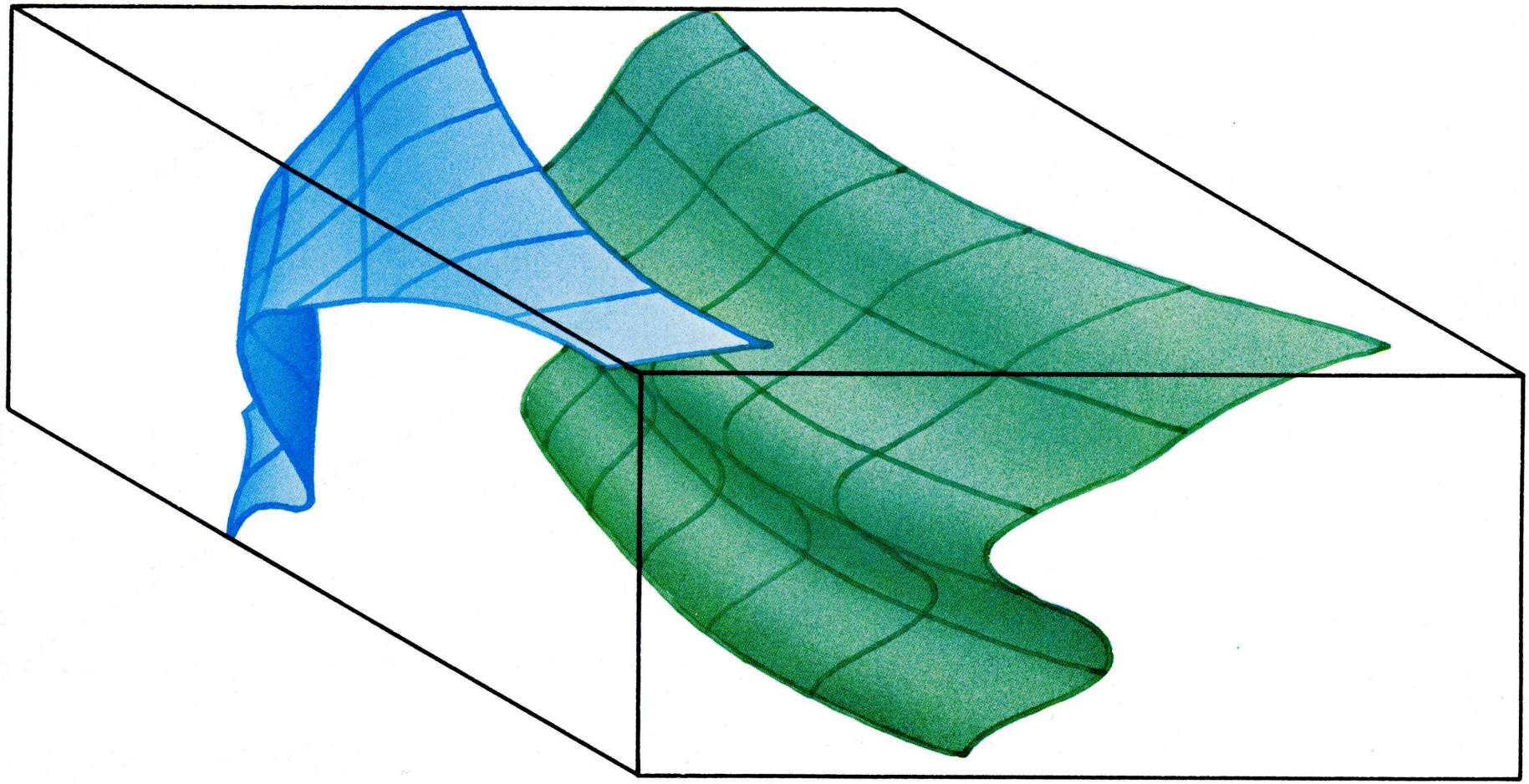


Figure 9. - Regression of Y and X with large autocorrelated areas.

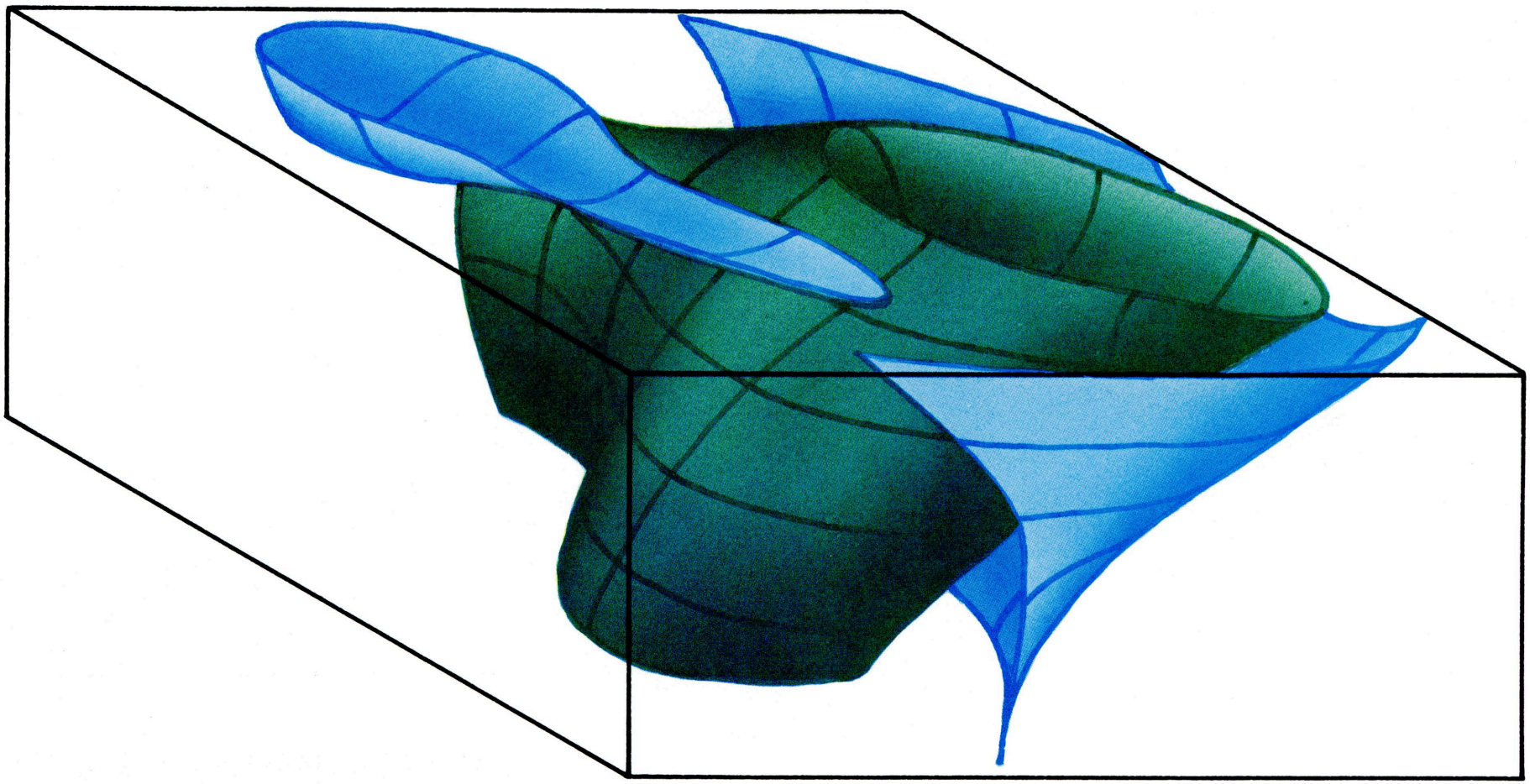
2nd Degree Polynomial



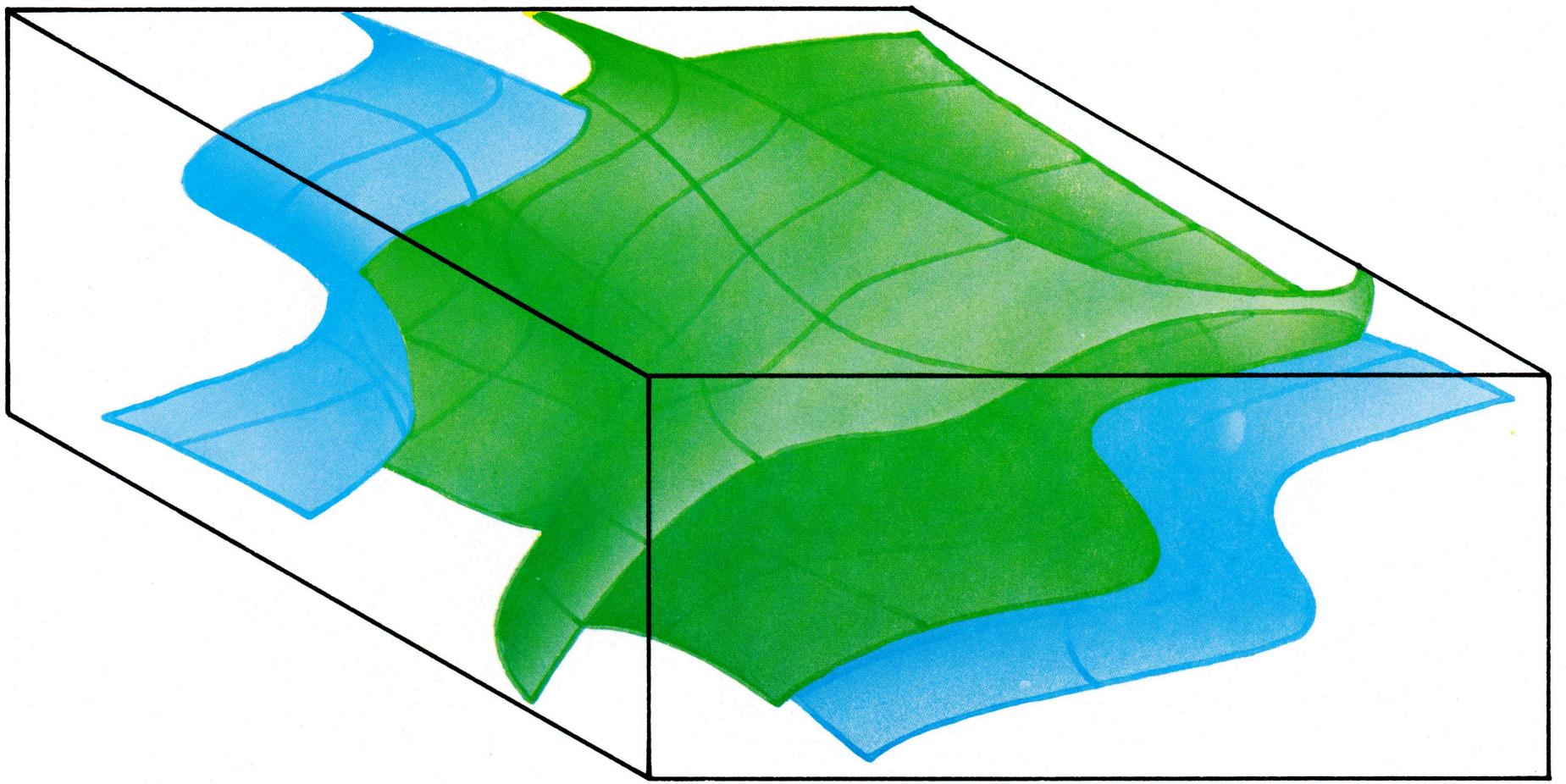
2nd Degree Polynomial - Fourier



3rd Degree Polynomial



3rd DEGREE POLYNOMIAL – FOURIER



intervals. Although not shown on this diagram, a zone of extremely high organic carbon content, above three percent, occurs within the cylinderlike part of the green envelope.

The model represents a significant improvement over simpler equations but large autocorrelated residuals remain. In particular, there is a large volume of autocorrelated negative residuals within the western margin of the cylindrical envelope. A trend surface created with a combination of cubic terms and vertical trigonometric terms is shown in Plate 4. The cyclic pattern produced by trigonometric functions appears as undulations within the envelope. Relative goodness-of-fit of this representation is approximately the same as for the previous trend surface. However, the large area of negative residuals which was present in the other model has been removed. Note that some of the contour sheets shown here are incomplete. The blue sheet which appears on the left, for example, has been truncated to show the form of the high carbon envelope behind it. In general, this sheet parallels the two-percent carbon layer.

More complex models were tested but did not prove more efficient than this one. Addition of terms did not reduce deviations significantly from the trend nor reduce significantly the volume of autocorrelated residuals. It was concluded that this model was about as efficient as possible without going to an extremely complex equation.

If we look carefully at this model some features which can be related to the geologic nature of the shale unit can be seen. For one, the area of low carbon values in northwestern Wyoming is related to the initial transgression of the Frontier Sandstone. This is reflected in relatively coarser grained zones within the shale. The coarser zones are low in organic carbon. Undulations in the western margin of the high carbon zone are a response to silty tongues in the shale which apparently are extensions of coarser clastic wedges from the west. The drop of organic carbon values to the east corresponds to the gradual change of the Mowry Shale to more typical marine shale and also to proximity of coarse shoreline facies such as the "J" sandstone.

From other lines of evidence, the Mowry is known to be time transgressive and the siliceous, organic-rich facies climbs section toward the east. West of the high carbon zone, the Mowry becomes increasingly silty. East of the zone, it becomes more typical of normal marine shales. Both organic carbon and silica content in the unit seem related and best developed in locations which seemingly were controlled by distance to clastic tongues. In general, the zone of favorably hydrocarbon production migrated eastward in response to spreading of the terrestrial facies of the Frontier Sandstone from the west.

SUMMARY

Coincidence of the high organic carbon zone with Lower Cretaceous oil fields in Wyoming is too striking to be completely fortuitous. The Mowry Shale may or may not have been the source of hydrocarbons which are now found as petroleum in these reservoirs. However, the significance of this study does not hinge just on the possibility that the Mowry provided the oil. Patterns within the shale unit indicate that optimum zones of hydrocarbon production seemingly were controlled by a distance from the Cretaceous shoreline, and migrated in response to clastic influxes. In most respects, the Mowry Shale is not different significantly from other Cretaceous shales in the Rocky Mountains. If organic material in the Mowry is concentrated in response to distance from shore, this may provide a guide in the exploration for high carbon source beds elsewhere. It is possible that similar patterns, perhaps less well expressed because of lithologic differences, exist in other Lower Cretaceous shales such as the Thermopolis. If this is correct, local sources of organic hydrocarbons may be present near reservoirs which now are producing so prolifically.

Furthermore, this model may provide a general guide to source rock exploration. It suggests that zones of high organic carbon production are related to stratigraphically equivalent tongues of clastic material which spill into the sedimentary basin. Mapping these clastic wedges is considerably easier than tracing organic carbon zones. If a simple relationship can be found for distance between these clastic tongues and the zone of high organic production, an exploration tool of significant value may result.

REFERENCES

Burk, C.A., 1956, Subsurface stratigraphy of the pre-Niobrara formations in the Wind River Basin, central Wyoming, in Wyoming stratigraphy: Wyoming Geol. Assoc., Casper, Wyoming, p. 23-33.

Davis, J.C., 1963, Origin of the Mowry Shale: Contributions to Geology, Univ. of Wyoming, v. 2, p. 135-146.

Draper, N.R., and Smith, H., 1966, Applied regression analysis: John Wiley & Sons, New York, 407 p.

Esler, J.E., Smith, P.F., and Davis, J.C., 1968, KWIKR8, a FORTRAN IV program for multiple regression and geologic trend analysis: Kansas Geol. Survey Computer Contr. 28, 31 p.

Harbaugh, J.W., 1964, A computer method for four-variable trend analysis illustrated by a study of oil-gravity variations in southeastern Kansas: Kansas Geol. Survey Bull. 171, 58 p.

Harbaugh, J.W., and Merriam, D.F., 1968, Computer applications in stratigraphic analysis: John Wiley & Sons, New York, 282 p.

Haun, J.D., and Barlow, J.A., Jr., 1962, Lower Cretaceous stratigraphy of Wyoming, in Symposium on Early Cretaceous rocks of Wyoming: Wyoming Geol. Assoc., Casper, Wyoming, p. 15-22.

James, W.R., 1966, FORTRAN IV program using double Fourier series for surface fitting or irregularly spaced data: Kansas Geol. Survey Computer Contr. 5, 19 p.

Merriam, D.F., and Harbaugh, J.W., 1963, Computer helps map oil structures: The Oil and Gas Jour., v. 61, no. 47, p. 158-159, 161-163.

Merriam, D.F., and Harbaugh, J.W., 1964, Trend-surface analysis of regional and residual components of geologic structure in Kansas: Kansas Geol. Survey Sp. Dist. Publ. 11, 27 p.

Mills, N.K., 1956, Subsurface stratigraphy of the pre-Niobrara formations in the Bighorn Basin, Wyoming, in Wyoming stratigraphy: Wyoming Geol. Assoc., Casper, Wyoming, p. 9-22.

Peikert, E.W., 1965, Model for three-dimensional mineralogical variation in granitic plutons based on the Glen Alpine Stock, Sierra Nevada, California: Geol. Soc. America Bull., v. 76, no. 3, p. 331-348.

Schrayer, G.J., and Zarrella, W.M., 1963, Organic geochemistry of shales - I. Distribution of organic matter in the siliceous Mowry shale of Wyoming: Geochemica et Cosmochimica Acta, v. 27, p. 1033-1046.

Schrayer, G.J., and Zarrella, W.M., 1966, Organic geochemistry of shales - II. Distribution of extractable organic matter in the siliceous Mowry shale of Wyoming: Geochemica et Cosmochimica Acta, v. 30, p. 415-434.

Smith, J.W., and Harbaugh, J.W., 1966, Stratigraphic and geographic variation of shale-oil specific gravity from Colorado's Green River Formation: U.S. Bur. Mines Rept. Inv. 6883, 11 p.

PRACTICAL COMPUTER USAGE FOR SUBSURFACE GEOLOGISTS

by

James O. Lewis[1/]

ABSTRACT

The "success ratio" of the experienced subsurface geologist can be improved by implementing proven subsurface exploration methods with utilization of the computer. Techniques for proper utilization need to be developed by the experienced subsurface geologist thoroughly familiar with computer programs used in solving exploration problems.

Output from the computer is not the end result, but is the beginning point for the exploration geologist. The "geology" of an area can be displayed in a form acceptable and familiar to the experienced geologist. The amount and quality of the displayed information will give the geologist more information, in an objective form, than has ever been practically available previously. This information, interpreted by the experienced geologist, will result in a higher quality of "decision-making" than has been possible.

Use of a computer will not enable the reduction of an exploration staff; but used properly will increase the need for experienced geologists and increase the "success ratio".

[1/]Consultant, Houston, Texas

8. Prediction of the performance of a solution gas drive reservoir by Muskat's Equation, by Apolonio Baca, 1967 . . . . . $1.00
9. FORTRAN IV program for mathematical simulation of marine sedimentation with IBM 7040 or 7094 computers, by J.W. Harbaugh and W.J. Wahlstedt, 1967 . . . . . $1.00
10. Three-dimensional response surface program in FORTRAN II for the IBM 1620 computer, by R.J. Sampson and J.C. Davis, 1967 . . . . . $0.75
11. FORTRAN IV program for vector trend analyses of directional data, by W.T. Fox, 1967 . $1.00
12. Computer applications in the earth sciences: Colloquium on trend analysis, edited by D.F. Merriam and N.C. Cocke, 1967 . . . . . $1.00
13. FORTRAN IV computer programs for Markov chain experiments in geology, by W.C. Krumbein, 1967 . . . . . $1.00
14. FORTRAN IV programs to determine surface roughness in topography for the CDC 3400 computer, by R. D. Hobson, 1967 . . . . . $1.00
15. FORTRAN II program for progressive linear fit of surfaces on a quadratic base using an IBM 1620 computer, by A.J. Cole, C. Jordan, and D.F. Merriam, 1967 . . . . $1.00
16. FORTRAN IV program for the GE 625 to compute the power spectrum of geological surfaces, by J.E. Esler and F.W. Preston, 1967 . . . . . $0.75
17. FORTRAN IV program for Q-mode cluster analysis of nonquantitative data using IBM 7090/7094 computers, by G.F. Bonham-Carter, 1967 . . . . . $1.00
18. Computer applications in the earth sciences: Colloquium on time-series analysis, edited by D.F. Merriam, 1967 . . . . . $1.00
19. FORTRAN II time-trend package for the IBM 1620 computer, by J.C. Davis and R.J. Sampson, 1967 . . . . . $1.00
20. Computer programs for multivariate analysis in geology, edited by D.F. Merriam, 1968 . $1.00
21. FORTRAN IV program for computation and display of principal components, by W.J. Wahlstedt and J.C. Davis, 1968 . . . . . $1.00
22. Computer applications in the earth sciences: Colloquium on simulation, edited by D.F. Merriam and N.C. Cocke, 1968 . . . . . $1.00
23. Computer programs for automatic contouring, by D.B. McIntyre, D.D. Pollard, and R. Smith, 1968 . . . . . $1.50
24. Mathematical model and FORTRAN IV program for computer simulation of deltaic sedimentation, by G.F. Bonham-Carter and A.J. Sutherland, 1968 . . . . . $1.00
25. FORTRAN IV CDC 6400 computer program for analysis of subsurface fold geometry, by E.H.T. Whitten, 1968 . . . . . $1.00
26. FORTRAN IV computer program for simulation of transgression and regression with continuous-time Markov models, by W.C. Krumbein, 1968 . . . . . $1.00
27. Stepwise regression and nonpolynomial models in trend analysis, by A.T. Miesch and J.J. Connor, 1968 . . . . . $1.00
28. KWIKR8 a FORTRAN IV program for multiple regression and geologic trend analysis, by J.E. Esler, P.F. Smith, and J.C. Davis, 1968 . . . . . $1.00
29. FORTRAN IV program for harmonic trend analysis using double Fourier series and regularly gridded data for the GE 625 computer, by J.W. Harbaugh and M.J. Sackin, 1968 . $1.00
30. Sampling a geological population (workshop on experiment in sampling), by J.C. Griffiths and C.W. Ondrick, 1968 . . . . . $1.00
31. Multivariate procedures and FORTRAN IV program for evaluation and improvement of classifications, by Ferruh Demirmen, 1969 . . . . . $1.00
32. FORTRAN IV programs for canonical correlation and canonical trend-surface analysis, by P.J. Lee, 1969 . . . . . $1.00
33. FORTRAN IV program for construction of Pi diagrams with the Univac 1108 computer, by Jeffrey Warner, 1969 . . . . . $1.00
34. FORTRAN IV program for nonlinear estimation, by R.B. McCammon, 1969 . . . . $0.75
35. FORTRAN IV computer program for fitting observed count data to discrete distribution models of binomial, Poisson and negative binomial, by C.W. Ondrick and J.C. Griffiths, 1969 . . . . . $0.75
36. GRAFPAC, graphic output subroutines for the GE 635 computer, by F.J. Rohlf, 1969 . $1.00
37. An iterative approach to the fitting of trend surfaces, by A.J. Cole $1.00
38. FORTRAN II programs for 8 methods of cluster analysis (CLUSTAN I), by Davis Wishart . . $1.50
39. FORTRAN IV program for the generalized statistical distance and analysis of covariance matrices for the CDC 3600 computer, by R.A. Reyment, Hans-Ake Ramden, and W.J. Wahlstedt $1.00
40. Symposium on computer applications in petroleum exploration, edited by D.F. Merriam . $1.00