APPLICATION OF COMPUTER DATA ANALYSIS TECHNIQUES TO DIGITIZED WELL LOG INTERPRETATION OF DEVONIAN CARBONATES

B.M. MCVICAR

this article begins on the next page ☞

PETROLEUM SOCIETY OF CIM

PAPER NO. 6801

THIS IS A PREPRINT — SUBJECT TO CORRECTION

APPLICATION OF COMPUTER DATA ANALYSIS TECHNIQUES TO DIGITIZED WELL LOG INTERPRETATION OF DEVONIAN CARBONATES

By

B. M. McVicar, P. Eng., Member CIM
Allison-Marshall Development Co. Ltd., Calgary, Alberta

This paper is to be presented at the 19th Annual Technical Meeting of the Petroleum Society of CIM in Calgary, May 7-10, 1968. Discussion of this paper is invited. Such discussion may be presented at the 19th Annual Meeting and will be considered for publication in CIM journals if filed in writing with the Technical Program Chairman prior to the conclusion of the meeting.

ABSTRACT

A reduction in well log computing costs and an exchange of techniques between professions is accelerating the use of computers in interpretation. A new computer program for Fast Fourier Transformation allows economic application of several processes previously too cumbersome. The digital approach to seismic analysis provides capability in that branch of the oil industry directly applicable to log analysis.

These techniques permit rapid and economic data screening and new opportunities for classification and correlation. Significant correlation of large amounts of data with references such as core analysis can be made. These give important local guidance as to the contribution or lack of contribution of various logs to a desired end result. Many of the tools are simple and straightforward. They allow recalibration of raw input data, the elimination of extraneous data and the smoothing and enhancing of data. This permits easier recognition of trends and correlation. Correlation and classification techniques taken from such sciences as biomedicine and psychology can be applied directly to the determination of lithology. This is improved through the use of Linear Programming. Once estimates of lithology are available from logs, several geological mapping techniques can be exploited.

A major problem in processing log data is the initial preparation, calibration, normalization and editing. Once the data is in computer compatible form and exonerated, many convenient manipulations can be made. These are best presented in graphic form. Visual analysis of graphs allows interaction with the computer for improvement of the processing.

The paper covers general techniques which can be applied to western Canadian prospects, and have been applied to Devonian carbonate sequences. The complexity of Devonian carbonates requires aids such as Time Series Analysis and Multivariate Analysis. Statistical procedures now available help dramatically in the choice of log suites and derivation of porosities and lithologies. At the same time, the impersonal contact that the analyst then has with the raw data allows a variety of errors and distortions to prevail unless precautions are taken.

INTRODUCTION

During 1968 a price breakthrough in well log digital processing started a renaissance of computer log analysis. Libraries of economical digitized logs are being organized. Large-scale computer facilities, with their economies and conveniences, are increasingly available. These factors are combining to encourage computer log analysis and fresh general techniques made attractive and necessary by the wealth of data.

The manual approach to data analysis retains some advantages. This is still the subject of experimentation. Project "MAC" at the Massachusetts Institute of Technology is an example. "MAC" has advanced from "Multiple Access Computer" to "Machine Aided Cognition". It has been quickly discovered that while a computer could be taught to perceive and recognize, the programming task is huge. The eye, together with the mind, can perform certain graphic analysis with relative ease. For example, when the human views a cross-plot of data points he can quickly appreciate: smoothness, averaging, circularity, cyclicity, clusters, and patterns. A promising combination of man and machine calls for an iterative exchange of graphic data. The machine produces a picture for human analysis, then is guided into its next process. This will be a part of log analysis. At the moment it can be done by plotter or printer interface. In due course graphic data input-output terminals will permit real time turnarounds, fast enough to keep the analyst's attention. This will include cursor, light pen and keyboard inputs and cathode ray outputs.

In the meantime, particularly in the case of large batch process jobs, an important step in log analysis may be downgraded. This is the analyst's general survey of the log for linearity, scaling, base line shifts, missing data, noise, drift, depth errors, heading errors and so on. Unless the analyst pre-edits and post-edits raw data, it may be too raw on commencement of processing.

There are a variety of tools available for interim data screening. These are made necessary by the automated approach and, at the same time, they provide a significant new aid. Certain precautions must be taken on a pre-edit before burying input data out of sight, out of mind. Some checks can be made during a computer edit process.

Techniques of data screening tie in with two other general fields: Multivariate Analysis and Random Data Analysis. These admit that an individual sample may be from a universe of a large number of different samples. The differences are judged by one or more of hundreds of physical and chemical properties. When one attempts to classify samples with an infinity of variations by measuring only a half-dozen of an infinity of properties, it is impossible to make an exact decision. The prediction should carry with it an estimate of probability. If this is done, the probabilities can be carried through to final decision-making such as bid price of crown land sales. The approaches are known as stochastic or random. They benefit from the simultaneous measurements of several properties over a sequence of time or depth samples.

The prime result of "Statistical Analysis" has been to provide a description or summary of data qualified as to probability and accuracy. Where one characteristic depends predictably upon another characteristic, it can be described by what is known as a deterministic formula. Resulting errors are primarily those of measurement or observation. Where, on the other hand, a characteristic cannot be exactly predicted by another characteristic and it takes on a statistical distribution of behaviour, the field of Multivariate Analysis applies. It involves the measurement of a number of characteristics, none of which need individually be exactly related. Through a series of mathematical manipulations, groups of the characteristics are correlated so as to predict the behaviour of other single or multiple characteristics within certain bounds of probability. This is facilitated through the measurement of these characteristics on large number of same-population samples. Analysis thus considers not only several criteria but also several cases. It automatically combines portions of individual criteria, by weighting techniques, with others so as to determine underlying trends in data that may not be directly reported by any individual characteristic. These trends are sorted out so as to be relatively independent of each other. Consequently, a suite of one dozen logging curves might be reduced to three or four underlying causative influences. These themselves can be correlated from depth to depth or from area to area.

Moving in a similar direction but more quantitatively oriented is Time Series Analysis. A time series is an ordered series of observations taken at different times. The definition can be stretched to apply to well logs in that depth is analagous to time. This introduces inaccuracies but it has yielded useful results. The range of Time Series Analysis goes from predictable behaviour such as a sine wave to unpredictable behaviour such as static on a radio. The latter case falls under the definition of Random Data Analysis. (1) Random means that no formula can exactly predict future behaviour. Nevertheless, useful signals or information can be buried in random data. If the signal can be extracted in the presence of noise, it can lead to correlations and classifications. The methods which have been devised to do this in Time Series Analysis can be used in log analysis. These include smoothing, enhancing and filtering techniques. In addition, they include the analysis of signals into their component parts by various types of plots. These plots in turn facilitate well to well correlation. Any set of data which can be described as a function of time also can be described as a function of frequency. When this is done, economies of computation can be made and new types of patterns become available for Pattern Recognition.

Taken together, these approaches permit new methods of well to well correlations of zones. They allow classification of individual zones into parent types. While the results have a chance of error, they provide an aid to the geologist or reservoir engineer in his decision-making problem. In the Devonian, for example, several dozen distinct types of lithology may exist. Rock texture may manifest several types. The presence of vugs can have a severe effect upon the response of a Sonic log. The Sonic log reports first arrival time which may be influenced primarily by the rock matrix with its intergranular porosity. Porosity contributed by vugs can therefore be overlooked. Virtually all logs are influenced strongly by lithology, sometimes to a greater extent than they are by porosity. A competent interpretation must take into account all of these variables. This can be done by making best use of all the parameters available. This in turn is best done by making use of multi-parameter manipulation which has been brought to a good standard by other sciences.

MEASUREMENT

The concept of a "system", input to it and output from it underlies techniques of this paper. Comparison of input with output can give a diagnostic system response even in the presence of noise. A study of present-day outputs and the system can lead to inferences of the original input "signal" now perhaps buried in noise.

In the beginning there was a system, and it was called Earth. Over geologic time it was subjected to many external and internal influences. Each region was subjected to cycles of deposition, erosion, diastrophism and compaction. The depth cycles cannot today be exactly interpreted as a determinable function of time. All of the above processes interacted on a non-linear, if not random and noisy, basis. Nevertheless, the geological history does represent a cyclic measure of system response over geolgic time. (Figure 1) Every few million years a major event occurred. These are the events which concern the seismologist. Minor events occurred at lesser intervals of time. They can be measured by the well log. Picture layover correlation has to do with the alignment of major events. This can correspond with the seismogram. Much greater resolution is required by the geologist and reservoir engineer in his day to day correlations within zones. Such resolution is available from logs and core analysis. Normal layover technique depends mostly on amplitude patterns.

The earth system which has resulted can now be tested for response during logging time by a variety of energy generators and sensors. Its spontaneous energies can also be measured. Geometric and geochemical qualities can be logged. These constitute a simultaneous measurement of properties such as is required by Random Data Analysis and Multivariate Analysis. (Figure 2) The energy or signal of the input generator is known or predicted; it passes through a system (earth) which may or may not be linear. Aside from inherent non-linearity the system can provide geometric effects such as the influence of bed thickness. It is altered as it leaves the tool and as it or its biproduct returns to the tool by the environment of measurement; this includes borehole effect, temperature and so on. The signal is further altered by the measurement process, tool geometry, telemetry to surface, surface instrumentation, and by surface graphic output or digital output. Consequently the delivered signal is not fully a representation of system response. Several noisy new criteria or characteristics have entered and must be accounted for. These are amenable to suppression or enhancement by modern statistical techniques.

The act of measuring the output of the system constitutes the observation of a physical process. (Figure 3) Where a physical process is random in nature, such as is the deposition of lithologic units, the ability to measure several outputs is important. This is the observation of an ensemble of characteristics. An ensemble can be two or more predictors. In psychological, geochemical and other studies, a relatively large number of characteristics are used. It behooves the log analyst to use as many curve types as possible for both Random Data Analysis and Multivariate Analysis. Ensemble measurements are necessary for non-stationary processes and for non-ergodic stationary processes. Non-stationary data implies the variation of statistical properties with time. Non-ergodic implies the variation of statistical properties between different parameters. In these cases, where individual parameter statistics are not diagnostic, ensemble properties can be more successful.

The analysis and classification techniques which will be described later make use of a study of several cases or samples, together with the study of several simultaneously measured characteristics or predictors. (Figure 4) Classification of rocks and fluid content within rocks involves an infinity of sample types. These can be categorized into subsets. Rock type could be quickly broken down into carbonate versus sandstone. The infinity of fluid distributions could be broken down to: above transition, within transition, and below transition. On a wildcat basis, the number of lithologies and fluid types which might be encountered is extremely large. In a given well, in a given major subsection, a subset of the samples is encountered. At a given depth, an individual sample is found. Through log measurement, these samples are described by perhaps an individual property such as a Gamma Ray measurement or a subset of properties such as a dozen curves on a well-logged well. This dozen measurements is a small subset of the actual abundance of properties which the samples possess. This infinity of properties is not measurable and available. Therefore the individual sample cannot be precisely classified. An individual property taken alone cannot classify the individual sample. Any gross classification which would be made on the basis of an individual property would be subject to gross error due to the influence of other properties. Practical classification technique uses a subset of properties and delivers only an estimate of classification. While the probabilities may be good, it is impossible to classify exactly any individual well sample as to its lithology, or fluid content, or porosity. What must be done is to acknowledge the probabilistic nature of interpretations and to improve the probabilities through the determining of all significant influences. This is an every day continuing battle full of local surprises.

DATA SCREENING

Before Data Analysis, Time Series Analysis and Multivariate Analysis can be applied, raw data must be carefully screened. This is done mentally by the log analyst as he views hard copy logs by eye. If this process is not continued in digital processing either by preliminary correction of data or examination of data trends through computer tools, errors will persist through Data Analysis to the end result. (3) An example would be a drifting Gamma Ray or a shifted Density log. This might not be noticed by the digitizer or by normal procedure up until the delivery of an erroneous lithology, porosity, or saturation.

Pre-editing is a manual examination process done in cooperation with a digitizer. Some suggested checks are given in Figure 5. Stipulations should also be made as to both relative and absolute accuracy. A suggested table is given in Figure 6. A post-digitizing edit can be made with a series of histogram, power spectral, cross-plot, transformation, normalizing, comparison, correlation, detrending and screening operations. These allow shifting, pruning and adjustment, and also provide a tape of normalized logs useful for curve to curve and well to well correlation. For gross analysis, the data should be screened for missing data, outliers, noise spikes, and so on. Unless this is done, the mean which is derived from the distribution of data can be adversely affected. A series of spectra, histograms and so on is shown in Figure 7. In addition, a simple cross-plot schematic is shown in Figure 8. The simple cross-plot of one parameter against another is an important preliminary method of adjusting data. It shows both nonlinearities and base line shifts. It shows clusters which might otherwise be misused in calibration. A cross-plot, for example of Density with Neutron, can segregate lithologies and the important effect of fluid content. An overlay of regional or normal response for various lithologies and fluids can suggest whether the data needs alteration

Normalization is a transformation of a variable to a dimensionless quantity so that it can be better correlated with other variables. This is usually done by subtracting the mean and dividing by Standard Deviation. Standard Deviation is an expression of variance about the mean. The result is to produce a series with zero mean and amplitude close to $\pm$ 1. Normalizing in log parlance can be extended to include the subtraction of and division by a variety of variables to remove their influence.

A histogram is simply a graph of the frequency distribution with which certain amplitude values occur. Histograms of curves can yield patterns which are correlatable from well to well. In addition, they permit recalibration of the curves themselves. A histogram also permits determination of a mode which can be used as a decision value to create binary logs. (4,6) Binary logs are simply two-valued representations of well logs. They are, in themselves, easily calculated and correlated. The shapes of smooth histograms can be evaluated in detail for normality, skew, tails, outliers and general geometry.

When the logs are transformed into a frequency domain representation, various frequency spectra can be plotted. A spectrum is the graph of a frequency domain characteristic of a series plotted against individual frequencies. The Power Spectra gives an idea of the intensity of the signal level at a given frequency. It can be calculated in several ways. The Power Spectra is the Fourier Transform of the Auto-Correlation function. The Auto-Correlation function itself can be calculated through a lagging process. This lagging process in the case of short records can introduce errors. The Auto Correlation function is a cross-correlation of a series with itself. Cross-correlation can be achieved between records in a variety of ways. One is to align the records then, at any given time (depth) calculate the product of the two amplitudes. This can be done after a normalization. All points are computed and the sum of the results is obtained. The records are given a relative shift and the process is repeated. The correlation coefficients are plotted versus lag and a maximum corresponds to best correlation lag. There are other methods of obtaining the Power Spectra however which can give better results in the case of short records. In the transforming process, an economical technique known as the Fast Fourier Transform is now available. This becomes attractive in the case of long records. At or above the calculation of some 1000 points, the Fast Fourier Transform reduces computer time drastically.

A special type of series is the Fourier case. (5) It is made up of a summed trigonometric series of sine and cosine functions. These can be made to duplicate virtually any shape of complex periodic wave. Conversely it may be said that a complex wave may be broken down into several pure sine and cosine wave components of certain frequencies, amplitudes and phase. (Phase relates to the time shift between a point on the cycle of one wave and the corresponding point on a reference wave). The complex wave can therefore be fully described in another way: by a spectra of the power in each component versus the frequency of that component. This turns out to the Fourier Transform of the Auto-Correlation function.

"Described in another way" defines the word transform. A log of Neutron counts is a transform. Porosity is a transform of formation factor, a square root is a transform. A complex transform is the expression of a time series in a frequency form.

The Fast Fourier Transform (8) computes transforms of a Fourier series in a succession of steps. By breaking the series into separate series, overall computation can be substantially reduced, and accuracy is improved.

In Devonian carbonates which are often difficult to correlate, the pattern provided by Power Spectra will assist the geologist and reservoir engineer in deciding whether a rock unit is correlative or not. (9) Various modifications of the graphic presentation are possible. These include taking the square root of a Power Spectra which becomes the Amplitude Spectra, or taking the logarithm of the Power Spectra. (10) In the case of the histogram, a "rootogram" may be computed. This compares the square root of one frequency diagram with the square root of a reference diagram. An advantage of the Power Spectrum in well to well correlation is that as beds thicken or thin, the spectrum maintains its pattern but merely moves up frequency or down frequency.

An unknown quantity is the phase spectrum. Conceivably, this may offer some information in correlation or analysis, and it is a simply-obtained biproduct of Spectral Analysis. The various spectra offer means, not only of correlation, but also of calibration of logs.

DATA ANALYSIS

Tukey and others (22) have distinguished between Statistical Analysis and Data Analysis as follows: Statistical Analysis is the summarizing of data; Data Analysis is the exposure and contemplation of differences. Figure 9 shows broad categories of data types. Parent types are Parametric which implies quantitative figures, and Non-Parametric which implies states or qualities. Both types of data are available in log analysis but the quantitative approach is being taken first for computer analysis. Non-Parametric Statistical Analysis would lend itself to the creation of binary logs which could then enter quantitative processes. Figure 9 shows the various types of data ranging from periodic and deterministic waves through to completely random or non-deterministic data. Well logs provide the most difficult case: that of random, non-stationary, time varying mean and time varying mean squared. Proper analysis of this case demands normalizing, trend removal, filtering, sharpening and so on to permit correlation and classification.

The actual procedure of analysis shown in Figure 10 consists of approaching the data with a flexible conceptual model, the regression of various variables against other variables, and the regression of sets of variables against other sets. If, when these data sets are plotted against each other, they do not provide a smooth logical relationship, then the data requires screening, or the variables require transformation and combination, or a new model is required. The curve fitting process can automatically take into account the contributions of different variables and discard those which do not add new information. A residual is the deviation or difference of an individual point from the smooth summary function that has been chosen as the model. The basis of Data Analysis is to study these residuals and explain them through finding relationships between residuals and other variables. Once this is done, the model can be readjusted to take into account the effect of other variables. This is best done by an interaction between computer output and the analyst through graphic interaction. The model or relationship is then simplified. An example of this procedure is the multiple lithology cross-plot. The raw data constitutes measurements made by Sonic, Density and Neutron logs. Variables are reformed, for example, the Neutron is converted within the tool from a logarithmic response to a linear porosity response. Within a given area, the contribution of information of each device can be evaluated and tested and the log can be either included in the development of the model or rejected.

The residuals are a biproduct of Regression Analysis. (Figure 11). This is simply a curve fit of a dependent variable with one or more independent variables. Once the fit is made, the residuals can be determined and plotted. They can be plotted against a variety of variables as shown in the figure. If the plot of residuals shows some type of correlation, this can be built into the model.

Regression Analysis (11) is a common denominator of both Data Analysis and Statistical Analysis. The broad category of Data Analysis would also include Multivariate Analysis in that Multivariate Analysis effects the combination of various parameters into linear relationships. These in turn can be entered into the regression process.

RANDOM DATA ANALYSIS

The procedures which have evolved to cope with data observed in the presence of noise, or data which does not strictly follow a formula, or data which cannot be predicted because it depends on too many other variables, are summarized on Figure 12. The manipulation of data by proper Random Data Analysis yields several plots or patterns previously mentioned. These range from a straight degree of correlation between curves, which is useful in lining up the depth shifts of curves within a well, to a complex coherence function. The coherence function offers some hope of well to well correlation by machine. Like most other operative techniques at the moment for time series, it depends upon the series being stationary. This is not the case over long intervals of well logs. Statistical properties in a carbonate section obviously change drastically from those in an intervening shale section. This change itself represents non-stationarity of data. Sections must therefore either be dealt with separately or must be treated by new techniques for the processing of non-stationary records. Nevertheless, where data is reasonably stationary and major changes of lithology are not being included, coherence offers a correlation aid to the geologist in complex carbonates. In small areas where small sections are used, or in large areas where long records are used, the process of transforming, or filtering, or normalizing can substantially improve correlation.

MULTIVARIATE ANALYSIS

A variate is a variable which takes on a statistical distribution in the absence of sufficient data to specify exact response. When several variates can be measured on several samples, a series of correlation and summarizing manoeuvres can be made. These constitute Multivariate Analysis. (19) They simplify, show relationships, and generate new independent variables to which new meanings can be assigned. A subsection of the general field is called Numerical Taxonomy. It is concerned with classification. Schemes which have been devised in biological classification are quite applicable to geological problem-solving. (17) A prominent tool in the summarizing of taxonometric classification is the dendrogram and phenogram. (Figure 13) . This graphic device shows the results of cluster analysis and other correlation procedures. It classifies different subsets of cases into species and shows the intercase relationship or degree of similarity, and the intergroup similarities together with a similarity to a target well, for example.

The general tools of Multivariate Analysis are shown in Figure 14. (2) The first operation is to determine correlations between various variables. Variables may be teamed as a unit, and the relationships between groups of variables to other groups of variables determined. This is known as Canonical Relationship. An important tool is Discriminant Analysis. This is a mathematical procedure for assigning an unknown sample to one or more parent categories. Principal Components Analysis is an unbiased method of studying the interrelationships between variables of different samples to create new underlying trends or summary variables. These in turn can be related to general characteristics of formations, for example. An extension of Principal Components Analysis is Factor Analysis, which starts with a preconceived model and simplifies the results of the Principal Component Analysis. Although factors are more easily interpreted than principal components, a certain amount of information can be lost. Nevertheless, the factors can be mapped vertically and horizontally as indications of fundamental formation changes.

Classification is achieved frequently by clustering of points as in a two-dimensional graph. The assignment of a sample to one cluster or another is based on a variety of distance and linkage measuring techniques. These classification processes, in making use of an ensemble of well logs, can give a general appreciation of whether a given carbonate rock unit is correlative with another when this is not indicated by picture layover correlation alone.

The bases of similarity and correlation measurements in Multivariate Analysis are "distance" functions and correlation coefficients. "Distance" can be a measurement on a two-dimensional graph between points to determine similarity. The concept can be carried into multidimensional space (16) wherein each case has its own distinctive vector and dissimilarity of vectors can be calculated as the net effect of all measured criteria.

Correlation coefficient is judged by a formula which decides if one variable relates to another for a cluster of cases. In the two-dimensional plot, a horizontal, vertical or circular cluster would indicate no (0) correlation. An oblique ellipse of points would indicate correlation. If all points fell on a straight line, strong correlation (1) would be indicated. The formula for calculating correlation considers sum of products, products of sums, and sums of squares for all pairs of characteristics. The coefficient indicates the dispersion of data and the possibility of achieving a tight linear regression. It can be used to select logging programs and transforms for porosity determination and so on.

Multiple Regression is a technique for fitting a descriptive summary line to data. It can satisfy various criteria for goodness-of-fit. (12) One is the Method of Least Squares which frequently approches maximum likelihood fit. Regression can be linear, non-linear, between two variables and between several variables. A steered regression (stepwise regression) successively and automatically tests and includes or rejects variables and transforms so as to add most relevant new information to the model. Regression in general offers the soundest method of predicting such criteria as porosity given logs as predictors. Steered Regression lets the data solve itself for the use of most meaningful input logs.

The method of determining Principal Components (13) fits an axis in N dimensional space along the data points in the direction that gives least variance. In the two-dimensional case, if an elliptical cloud of data is considered, the principal component would be along the major axis. The minor axis would be variance which would be minimized. The principal component would probably be oblique to the x or y axis and would imply some influence other than x or y. The minor axis would represent the next most important theme, independent of the principal component. The concept can be extended to more dimensions and components. The components should have important fundamental causative meaning as opposed to the variables which may contain redundant resultant information.

In Factor Analysis an interpretation simplification is made by tying back components more recognizably to original variables. This is done by an axis rotation scheme. Two types of Factor Analysis are available, "Q" and "R". Through these, relationships between variables or between cases can be studied. In the latter case, plots or maps of dominant factors can be created to assist in extrapolating, interpolating and determining trends. (figure 15)

LITHOLOGY CROSS-PLOT - LINEAR PROGRAMMING

Once data screening, correlation and classification has been facilitated, an embellishment of matrix algebra permits use of the technique known as Linear Programming. (15) This holds out much hope for logging, production, management and reservoir problems. It is basically the simultaneous solution of a set of equations. At the moment, the equations are preferrably linear. This can be arranged through transforms. The technique does not require that the number of known variables equal or exceed unknowns. Instead, the analyst can stipulate certain constraints, such as porosity is not allowed to be negative and percent gypsum cannot be greater than ten and so on. In addition, an objective function can be defined for minimizing or maximizing certain responses. In logging, these could be "find the optimum solution such that given percent error of various variables, when considered in total, is minimized". The Chaveroo (15) technique uses a basic form of Linear Programming. It employs one slack variable which is used to make up a given total, and it uses non-negativity constraints. At a local level, a large number of rules of thumb could be incorporated as constraints. These can include maxima, minima, ratios, and so on. The solution of simultaneous equations through matrix inversion allows exploitation of three or four-end-member lithology cross-plots. Various combinations of curves are possible in arriving at this. The three-end-member process (14) is merely a closed system of solving percentages of minerals and porosity. One can say that a three-end-member system must add up to one hundred percent of bulk volume. The members might be anhydrite, dolomite and porosity. Plotting the relative amounts on a triangular graphic solution allows classification of particular points into clusters which can have a given symbolism or pattern or colour for later use on maps or graphs. A series of these is shown in Figures 16 and 17.

A recommended procedure for determining the approach to use in Multiple Lithology Cross-Plotting is shown on Figure 18. First a Principal Components and/or Factor Analysis is performed to determine how many basic influences are truly involved. These are then correlated with known stratigraphic content. Individual curves are regressed against core analysis porosity as are combinations of curves. The best combination is chosen for this purpose. The same combination, or a different one, is chosen to reflect lithology. Lithology can then be determined and used as a correction for porosity determination, and then hydrocarbon content can be assessed and used as a correction for both lithology and porosity calculation. The successively refined figures can be allowed to interact on each other during iteration until amendments are brought to a given minimum. It will be found in local areas that certain pairs of logs do the porosity definition chore best, and other combinations define lithology best. When all logs are used in combination, there is a danger that one or another will be given an undue weight. This can be accounted for in a complex Linear Programming system by the introduction of constraints. On the other hand, it may be found that a simpler approach after a local preliminary study is best. Through experimenting with constant factors and functional relationships, the best correlation coefficient can be found by machine processing, which is an advance over previous curve fitting by hand.

LITHOLOGY MAPPING

Once lithology has been estimated, several geological machine mapping techniques are made available. (18, 20) These include the plotting of vertical variability and horizontal variability of rock types. Another recommended measurement is that of entropy, a measure of the degree of mixing of lithologies. Isolith and multilith maps are directly available from the lithology plots. An example is shown in Figure 19. This particular map, courtesy of CDP Computer Data Processors Ltd., is the result of a three-end-member ranking system. This is shown in Figure 20. It was designed specifically for use in the middle Devonian of northern Alberta. The philosophy was that the worst situation was to have one hundred percent anhydrite and the best was to have one hundred percent dolomitic reef. An intermediate situation was to have limestone basinal facies. Through this and similar priority schemes, a contour map can be drawn. An extension beyond the contour scheme through use of colours or patterns is the identification of different levels of the hierarchy as shown.

The "D" Function is a special case of three-end-member approach whereby a mathematical solution of the triangle can be made and plotted. The "G" Function shown on the diagram credits Gibson of CDP who was the source of the middle Devonian ranking scheme.

TREND SURFACE ANALYSIS

Important output is in vertical and horizontal graphs. A variety of these is shown in Figures 13, 19, 21 and 22. Residuals can also be plotted in three dimensions. In this case, for example, for structural mapping, a smooth trend can be fitted to a surface of observed depths. This trend is then presumed to represent regional trend. The actual differences between observed depths within individual wells and the corresponding depths reported by the regional trend constitute residuals. These residuals can then be mapped as high or low deviations from regional trends. They then constitute prospects for drilling. The computation of trend surfaces and the residuals is sensitive to the mathematical and computing processes involved. The resulting residual maps must be interpreted with care. Nevertheless they form one other important graphical output, which can be used not only for structure but also for isopachs and for the plotting of quantitative data such as porosity, water saturation, salinity, shaliness, favourability and so on. An extension of the automatic contour in mapping process is the presentation of three-dimensional simulations. All of these aid the log analyst in appreciating the intensity of various parameters at various geographic locations. The general trends can be used for interpolations of such data as water resistivity, general porosity limits, general lithology and so on. In the middle Devonian, for example, perturbing factors include gas, pyrite, pyrobitumen, graphite, sulphur, evaporites and sodium chloride areas. Where these can be mapped, pitfalls of log analysis can be avoided.

CANADIAN DEVONIAN

The large proportion of hydrocarbon reserves contained in the western Canadian Devonian system requires an expertise in log analysis for this formation. The general makeup of this carbonate requires application of statistical techniques such as have been described herein. Firstly it represents a heterogeneous mixture of lithologies. These range throughout the spectrum of limestone types through varying degrees of dolomitization through the myriad rock types which have been catalogued geologically within the major fields thus far discovered. Anhydrite is present frequently and the nemesis of Neutron logging: gypsum. A similar problem for the Sonic log is secondary porosity in the form of fractures and vugs. Through the use of multiple lithology and porosity plotting, this secondary porosity can now be recognized to a degree, although it constitutes one more unknown. An interim technique interpreting the Sonic log was to bend the response upward to account for vugs and report higher bulk porosity. This has now been abandoned at least in theoretical circles. The preferred technique is to name the porosity derived from the Sonic as being primary porosity, and the difference between this porosity and cross-plot porosity as being secondary porosity.

The ratio of secondary porosity to total porosity is called vug index. This general philosophy can be built into a linear program for a given area, although it is subject to some interference from the presence of gas. The effect of gas in carbonates, such as the Devonian, is a matter of debate. Various authors state that at depth within a consolidated formation, the effect of gas and its associated pressure is minimal. On the other hand, examples can be produced which seem to suggest that gas does have some effect. The result is an unduly optimistic porosity. The Sonic nevertheless survives within the Devonian as being one of the best porosity tools.

The recent introduction of the Sidewall Neutron log has met with favourable response. Transforming of the Neutron data is done within the instrument to provide a linear porosity response. This is open to question on a local basis. In addition, Neutrons in general are subject to statistical variation and the Sidewall Neutron may be subject to the problem faced by micro devices whereby only one azimuth of the borehole is studied. For one reason or another, the repeat logs obtained on overlapping Neutron logs are not in exact agreement. This lack of repeat can itself alter lithology interpretations. Consequently lithology calculations should take into account a certain degree of error or variance from the logging procedures alone. If this is done, an envelope of probability can be placed on the resulting lithology calculations.

In Multiple Lithology Cross-Plotting, matrix inversion can be significant. When all logs are allowed to remain in the regression equation, one of them may take on greater local significance than it deserves. This is a result of downgrading the weight or contribution of other logs due to a fear of the effect, for example, of gypsum. For this reason, a preliminary analysis of the formation to eliminate the possibility of certain lithologies may result in a better interpretation and a better combination of logs. Otherwise, digitizing error plus instrumentation alone may place undue weight in the final regression formulae. The Density log is quite sensitive to lithologies of a certain type, hence it is a good lithology detector but in unknown situations can lack porosity resolution. Since, on a wildcat, the Devonian can present many unknown situations including metal mineralization and non-shale radioactivity, it is wise to run the full suite of logs as a first approach. Even on field wells where the lithology is under better control, it may be that definition of water contact or water table is of sufficient economic importance that the full suite should be run. Nevertheless, the statistical techniques which have been outlined can be quite decisive in pointing out the weaknesses of certain logs and the lack of contribution to general information which they make after other logs have been taken into account. From this point of view, the analyses can pay for themselves quickly in the simplification of logging programs in a development area.

The digitizing of well logs with its consequent easing of the problem of integrating, averaging and transforming various curves makes the plotting of reef proximity maps (21) practical. Resistivities can be an index either of the presence of calcium or of its replacement by pyrite. Depending on the model chosen, average resistivity or average conductivity or an inverse function of either can be plotted and contoured as a regional indicator of the favourability or presence of reef growth.

The use of the frequency domain instead of the amplitude domain and of filtering techniques can remove distortions which confuse correlation in normal picture layover amplitude processes. The rapid interlayering of the Devonian distorts various logs and creates, through geometric effects alone, a series of difficulties in correlation. This is compounded by the environment during measurement. This can be alleviated through the use of standardization and normalization transforms. The variability of the Devonian can cause permeability sandwiches which can break up normal water to oil gross transition. One can observe a mixture of critical water saturations and actual water saturations which can result in production of water above production of hydrocarbon throughout various sections of pay. To facilitate completion and reservoir computation, a precise appreciation of lithology, porosity and resistivity is required.

MANIPULATION OF DATA WITHIN THE COMPUTER

The efficient compiling and viewing of the various checks necessary to prepare data requires a system of programs. A representative system is shown in Figure 23. This differs from classical well log computer systems in that the emphasis is on data screening rather than on immediate classical log analysis. For this purpose, a subprogram system was chosen which allows the operation of several different types of process on a small amount of data successively. The entire record is held in and manipulated in core. The subprogram system constitutes overlay Number 1 in an overlay system. Subsequent overlays are large programs which are called in to perform Multivariate Analysis, Classical Log Analysis, Facies Determination, Complex Plotting Schemes and so on.

PITFALLS

With the introduction of logging data into machine-compatible form, a variety of new pitfalls is encountered. Some of these are outlined as follows:

. Until the wider availability of on-line cathode ray graphic terminals with log displays, digitized data will disappear from the log analyst's compensating view and carry with it all errors of recording, mechanical processing, and so on. To minimize the effect of these errors, an intensified system of pre-editing and post-editing must be instituted.

. Among the errors of recording are depth shifts. These can be detected by machine processing and automatic adjustments can be made. This improves interpretation substantially. It is sometimes more convenient to effect minor variations in the post-editing phase than in the pre-editing phase. Lags of up to twenty-five feet should be checked.

. Core analysis creates a problem for Spectral

Analysis in that record lengths can be short and there can be data gaps. A technique is needed for data filling and screening, and preliminary averaging.

. Outliers, wild shots, noise and so on contained in logs, such as skipped cycles, noise spikes, negative readings, etc., should be pre-screened so as not to alter means and standard deviations of distributions severely.

. Logs constitute a case of non-stationary data. Present techniques are designed for stationary data. Consequently, records which traverse drastic changes of amplitude should be treated in separate sections or other compensations made.

. Low frequency drifts and trends can mask information at higher frequencies unless they are removed.

. The resulting measurements of logs depend on a wide variety of factors which include environmental effects, lithology effects, hydrocarbon ratios, and porosity type. The seeking of a solution for any one of these answers cannot ignore the other. At medium to higher porosities, the effect of hydrocarbon upon Neutron and Density response must be compensated and it can have a variable effect on the Sonic.

. Processes of averaging or filtering can introduce cyclicities and patterns which are erroneous. Filter operators must be chosen with care.

. Large amplitude excursions can alter means seriously. For this reason certain random data statistical and spectral techniques should avoid extremes of amplitude and frequency diversion. Otherwise correlation is made difficult and inaccurate.

. Matrix inversion has possibilities of instability at higher orders and in large sizes. Simple matrix inversion used in linear programming involving several logs can transmit an undue importance to one log and suppress the contribution of another log in a manner which may not be appropriate in a given locale. This can be improved through astute choice of slack variables.

. Classical assumptions as to the response of various logs to porosity need to be questioned locally. Multiple Regression Analysis with a full suite of statistical output should be used wherever possible. A preliminary or adjunct to this is transgeneration and correlation.

. Presence of vugs can be overlooked by a Sonic log and should be doublechecked by other logs.

. Individual logs can be subject to scaling errors which can be detected or disproved by cross-plotting. Where overlaps are available, these can be useful on comparison to assign weight to be given to a given curve.

. Regression Analysis has not yet reached the stage where it can cope in a standard manner with error in both variables. Regression can be misapplied so that the wrong variable is taken as dependent or independent. Regression of a given cluster of points should not be extrapolated to another subset of points which have not been checked. For example, if a core analysis thoroughly analyzes a low porosity calcium carbonate intercrystalline section of the pay and ignores the rest, the regression line in a small cluster may take on a widely-variable slope. The slope cannot and does not take cognizance of other clusters of data within the same pay zone which have not been cored and which may have an entirely different response due to gas, vugginess or lithology. Consequently regressions must be constrained in one manner or another, or should not be applied to outside parent populations.

. Digitizing interval can be selected so as to cause aliasing. (7) This in turn can allow the transmission of aliased frequencies into the frequency pattern. Distortions of spectra result. Significance of this has not yet been fully determined. More experience is needed in the filtering process as to what low frequencies to eliminate and what high frequencies to eliminate on account of record length, possibility of noise and so on.

. Within the Devonian, one may expect several of a large number of lithologies to appear from place to place. In addition, a variability of porosity and permeability, fracturing and vugs may be expected. Furthermore a layering of water transition and contact can be present.

. It is impossible to define exactly each individual sample using only a subset of a few parameters. One should therefore mentally, at least, ascribe a probability to the answers, and use as many parameters as are available.

CONCLUSIONS

1. Digital log analysis has suddenly become more economical and rewarding.

2. Libraries of digital logs are now available.

3. Computer costs have steadily reduced and new techniques such as the Fast Fourier Algorithm have caused a dramatic reduction in cost of certain scientific processing.

4. Scientists in other fields have devised excellent methods of coping with multiple parameter solutions and with Random Data Analysis. These are applicable to well logging.

5. A biproduct of the above techniques is an improved process of classification and correlation.

6. Noisy and confusing logs can be filtered, thereby permitting easier correlation. This can be enhanced with graphic output such as variable density and variable area plotting.

7. Featureless logs can be enhanced by sharpening filters and derivative logs, etc.

8. Curves can be calibrated through quick comparison methods. Logs can be improved and made more correlatable and more amenable to regression by the use of transforms of various types.

9. To define properly the large number of possible rock types encountered in the Devonian, a large number of parameters is essential. This is even more the case when definition of fluid content is required.

10. All logs are subject to some effect by fluid type, fluid content, mechanical makeup of the rock, lithology of the rock, porosity amount and type and so on. These effects must be assessed on a local basis by statistical means in order to give each log its proper local weight.

11. The examination of residuals as well as the determination of summarizing curves offers an important method of improving the model to which logs are fit.

12. Proper Regression Analysis is a powerful tool and can be done easily by machine once logs and core analyses are digitized.

13. Random Data Analysis techniques only recently popularized can be used almost in entirety for the processing of well logs. They provide several biproduct plots and indices which can be used for classification and correlation.

14. Numerical Taxonomy and Multivariate Analysis procedures used in the Life Sciences are directly applicable to log analysis, and are particularly appropriate for the Devonian.

15. The use of the above techniques permits an improved lithology determination from logs. This lithology determination in turn opens up possibility of plotting a variety of geological indices vertically and on maps.

16. The large number of qualitative and quantitative cross-checks of data content, nature and accuracy requires a manipulative system of programs.

17. Techniques used presently by life scientists and the seismic industry can be integrated into a generally new system of log analysis. This is used together with classical log analysis to provide a wide variety of reservoir and lithologic answers which can be mapped and plotted vertically. The resulting information can be integrated by machine with geophysical data and fed into a geologic model. (Figure 24) The geologic model can then be coordinated with the reservoir model to yield economic data for drilling and bidding strategies.

REFERENCES

1. Bendat and Piersol; "The Measurement and Analysis of Random Data", John Wiley & Sons, New York, 1966

2. Cooley and Lohnes; "Multivariate Procedures for the Behavioural Sciences", John Wiley and Sons, New York, 1962

3. Connolly, E.T.; "Standardization of Large Volumes of Log Data for Computer Analysis", Second Formation Evaluation Symposium, C.W.L.S., Calgary, May 6, 1968

4. Jeffries, F.S.; "Computer Correlation of Wireline Log Parameters and Core Analysis Parameters", The Log Analyst, Vol. VII, No. 3, September - October, 1966

5. Harbaugh, J.W.; "Geologic Application of Single and Double Fourier Analysis", Michigan University Engineering Summer Conference, May 9-20, 1966

6. Jeffries, F.S.; "A Self-Taught Computer Program To Interpret Micrologs", The Log Analyst, Vol. VII, No. 3, September-Ocotber, 1966

7. Runge, R.J. and Powell, N.J.; "The Effect of Sampling Rate on Sonic Log Span Adjustment", S.P.W.L.A. Seventh Annual Logging Symposium Trans.

8. Gentleman, W.M. and Sande, G.; "Fast Fourier Transforms - For Fun and Profit", A.F.I.P.S. Conference Proceedings, Vol. 29, 1966 Fall Joint Computer Conference

9. Canard, C.R.; "Spectral Analysis of Well Logs for Lithologic Subsurface Correlation", Masters Thesis, Dept. of Geology, U. of Kansas, Oct. 1966

10. Bogert, R.P., Healy, M.J. & Tukey, J.W.; "The Quefrency Alanysis of Time Series for Echoes: Cepstrum, Pseudo Autocovariance, Cross-Cepstrum and Saphe-Cracking", "Time Series Analysis", Rosenblatt, J. Wiley & Sons, New York, 1963

11. Draper, N. & Smith, H.; "Applied Regression Analysis", J. Wiley & Sons, New York, 1966

12. Haseldin, G.M.; "Use of Error Ratios in Least Squares Fitting of Data", C.W.L.S. Second Formation Evaluation Symp., Calgary, May 6, 1968

13. McCammon, R.B.; "Principal Component Analysis and Its Application in Large-Scale Correlation Studies", Jour. of Geol. Vol. 74, No.5, Part 2, Sep.66

14. (a) Savre, W.C.; "Determination of a More Accurate Porosity and Mineral Composition in Complex Lithologies with the Use of the Sonic, Neutron and Density Surveys", Jour. of Pet. Geol. Sept. 1963

(b) Burke, J.A., Curtis, M.R. & Cox J.T.; "Computer Processing of Log Data Enables Better Production in Chaveroo Field", 41st Annual Fall Mtg. SPE-AIME, October 2-5, Dallas, Texas

15. Loomba, N.P.; "Linear Programming", McGraw-Hill, New York, 1964

16. Kendall, M.G.; "A Course in the Geometry of n Dimensions", Charles Griffin & Co. London 1961

17. Sokal & Sneath; "Principles of Numerical Taxonomy" W.H. Freeman & Co., San Francisco, 1963

18. Computer Contributions, Kansas State Geological Survey, U of Kansas, Lawrence, Kansas

19. B.M.D. Biomedical Computer Programs, Health Sciences Computing Facility, Dept. of Preventive Medicine & Public Health, School of Medicine, University of California, Los Angeles

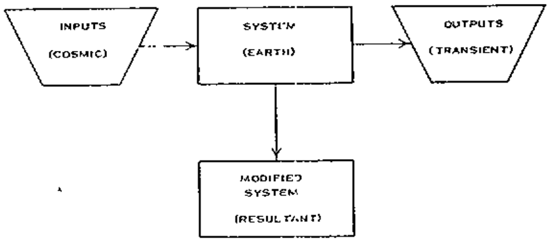
20. Forgotson, J.M.R.; "Review and Classification

and Quantitative Mapping Techniques", Bulletin of the A.A.P.G., Vol. 44, No. 1, January, 1960

21. McCrossan, R.G.; "Resistivity Mapping and Petrophysical Study of Upper Devonian Inter-Reef Calcareous Shales of Central Alberta, Canada", A.A.P.G. Bulletin, Vol. 45, No. 4, 1961

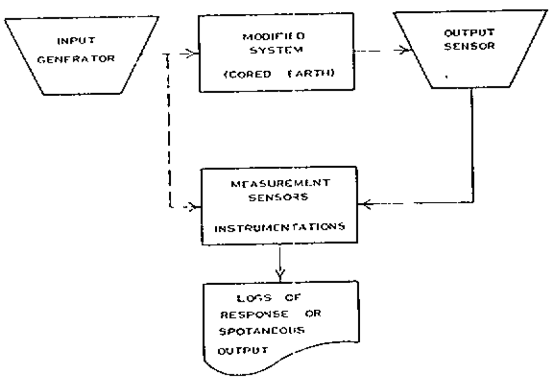
22. "Data Analysis and Statistics: An Expository Overview", 1966 Fall Joint Computer Conference A.F.J.P.S., Spartan Books, Washington, D.C.

SYSTEM RESPONSE



GEOLOGICAL TIME
(FIG 1)

SYSTEM RESPONSE



LOGGING TIME
(FIG 2)

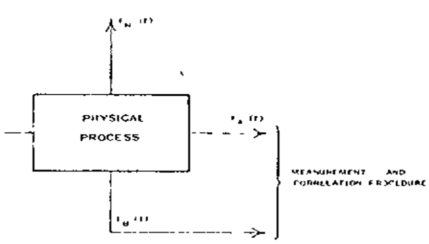


FIGURE 3

QUANTITATIVE ANALYSIS AND
QUALITATIVE CLASSIFICATION

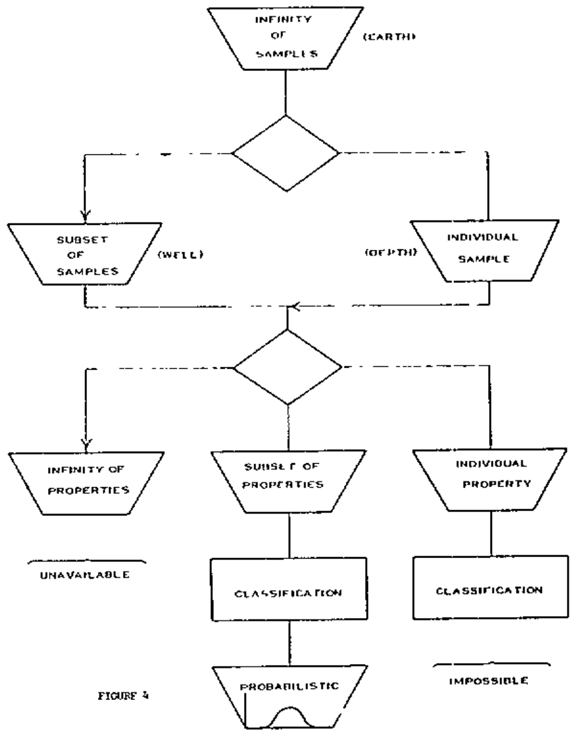


FIGURE 4

APPROVALS:	ALLISON-MARSHALL DEVELOPMENT COMPANY LTD.	SECTION B - 2
		PAGE 1
		REVISION DATE
VOLUME: III	WELL LOG DIGITIZING RESEARCH PROJECT	SUPERSEDES

SUBJECT: PRE-EDITING CONSIDERATIONS ©

Before a log is digitized, it should be pre-edited. Depending on what allowances are automatically made in the digitizing process or post digitizing edit programme, the following considerations apply:

1) Test alphabetic data for reasonableness.
2) Test numeric data (e.g. scales, depths, scale changes) for reasonableness.
3) Test individual film segments against each other for reasonableness.
4) Check depth grid for continuity, accuracy, jams, stretches, etc.
5) Check horizontal grid for accuracy, scale, etc.
6) Check each curve for continuity, static, superimposed signals, drift, zero top and bottom.
7) Check maxima, minima, means, shales, etc. for regional reasonableness.
8) Corelate each curve of a log for depth shifts.
9) Corelate each log of a well for depth shifts.
10) Delete curves and logs not wanted.
11) Delete runs not wanted.
12) Delete overlap and repeat intervals not wanted.
13) Upgrade curve manually where faulty (noise, skipped cycles, error, etc.)
14) Check presence and accuracy of offscale curves.
15) Check accuracy of reciprocation, function forming, etc. for nonlinear curves.
16) Check behaviour below pickup and above casing shoe.
17) Check against offset well response.
18) Is there circuit saturation or galvanometer plateau?
19) Are there geometric defects (e.g. Delaware gradients)
20) Are there negative readings, what do they imply?
21) Is there deformation indicating instrumentation defects?
22) Is there agreement between overlaps, runs, repeats?
23) Is behaviour in shale reasonable?
24) Are monitor curves satisfactory?
25) Is there undesirable tracking or mirror image between curves(e.g. SP-Res, GR-NL)
26) Where curves are to be multiplexed, corelate to a common depth point.
27) If there is a cost or hardware limitation, delete unimportant zones.
28) Indicate by red pencil on the hard copy what the digitizer is to do, clarify curves.
29) Put digitizing instructions in writing (frequency, IRG procedure, input scales, output scales for echo plot, shrinkage procedure, offscale and scale change procedures, averaging procedure, tape identification procedure, exact intervals, acceptable depth and amplitude tolerance, etc.)

FIGURE 5

Tuntative Guide Lines for Amplitudes (for your consideration)

Curve Type	Maximum Relative Error Tolerance (log units)	Maximum Absolute Error Tolerance (Cursor Distance)	
		Persistent	Transient
SP	1%	.1"	.5"
16"	5%	"	"
40" Induction	5%	"	"
Conductivity	5%	"	"
Laterolog	5%	"	"
Sonic	2%	"	"
Density	1%	"	"
Gamma	8%	"	"
Neutron	2%	"	"
Caliper	2%	"	"
Microcurves	5%	"	"

FIGURE 6

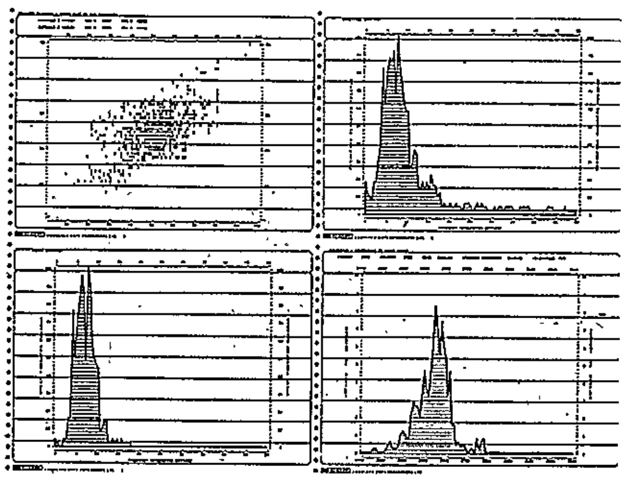


FIGURE 7

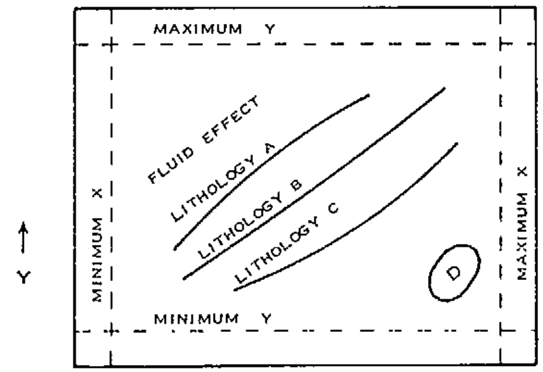


FIGURE 8

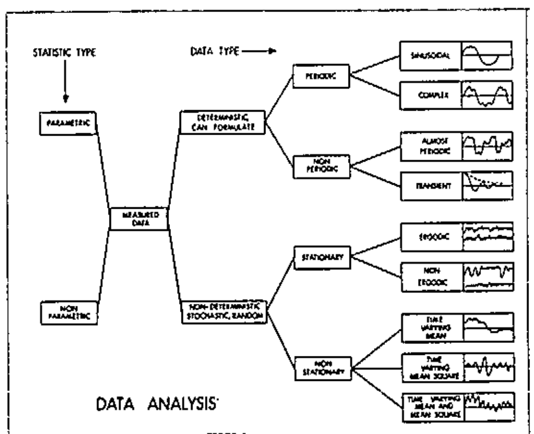


FIGURE 9

DATA ANALYSIS II

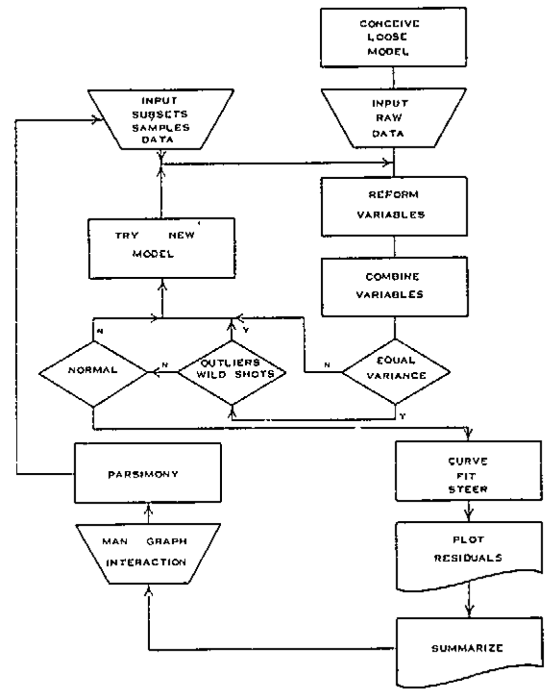


FIGURE 10

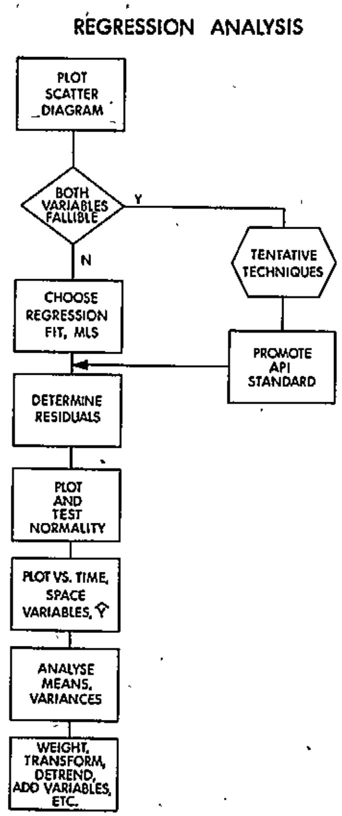


FIGURE 11

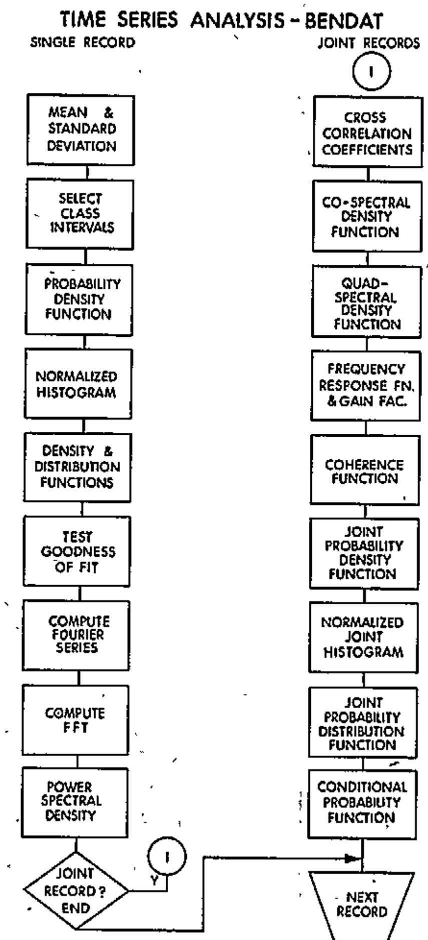


FIGURE 12

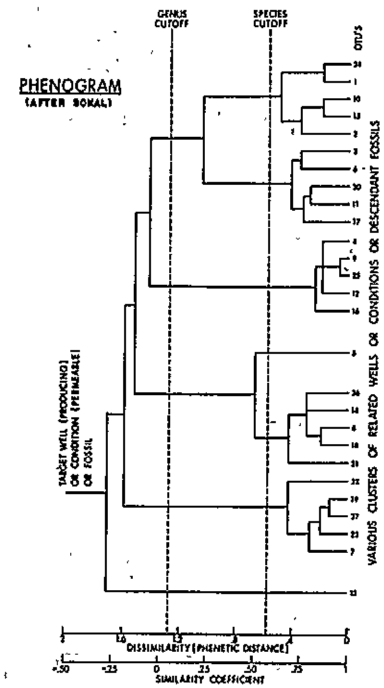


FIGURE 13

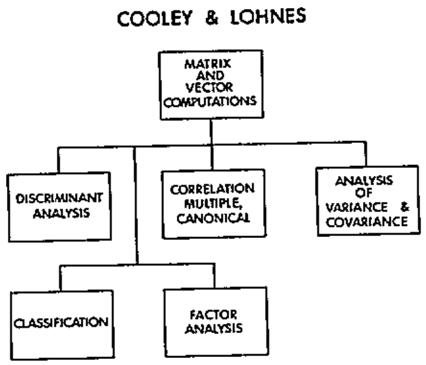
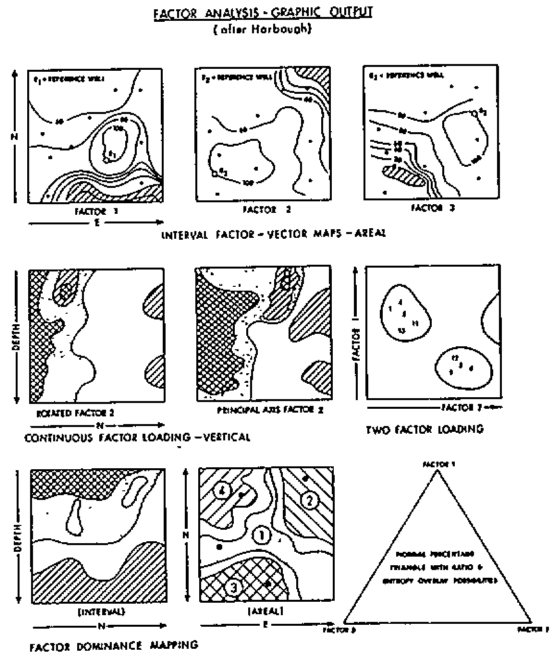


FIGURE 14



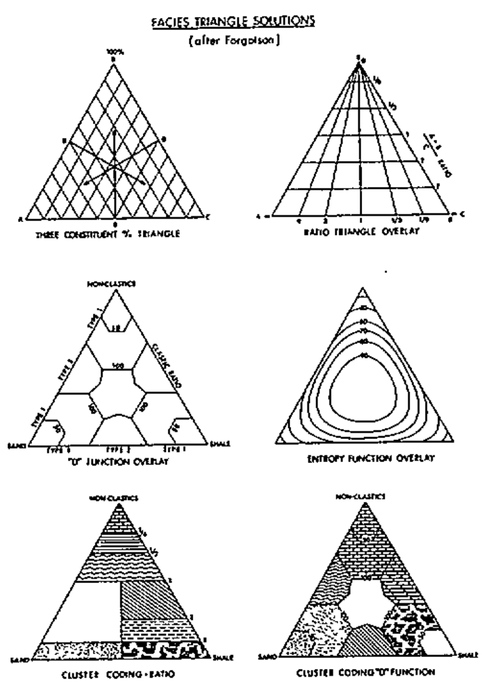


FIGURE 16

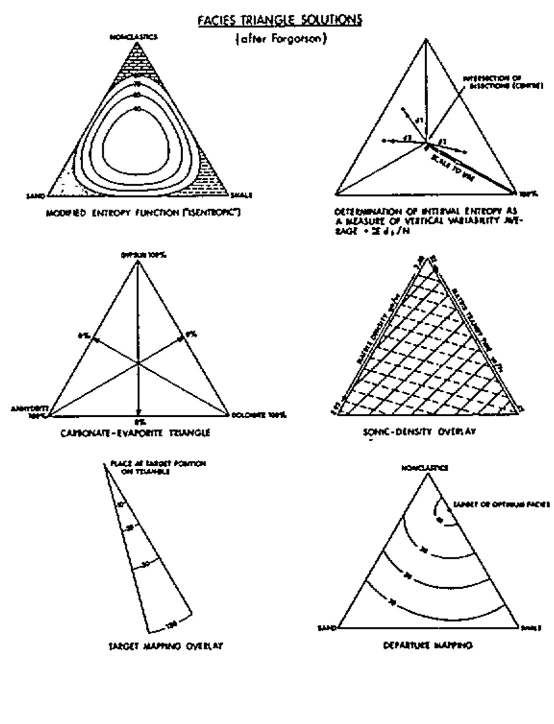


FIGURE 17

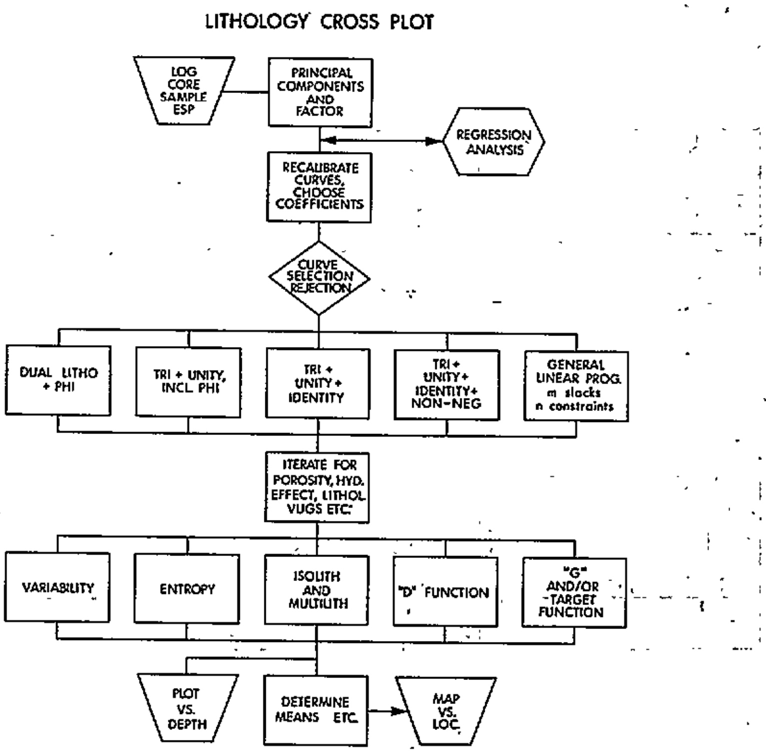


FIGURE 18

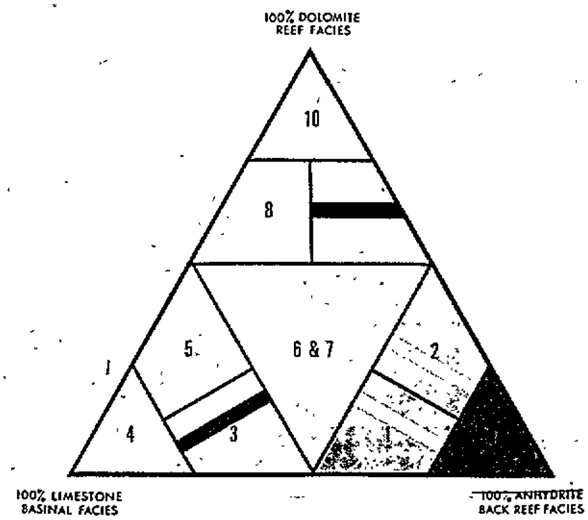


FIGURE 20

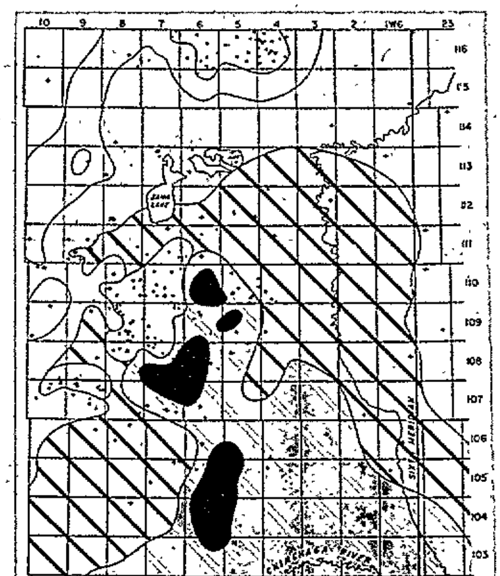


FIGURE 19

HYPOTHETICAL VERTICAL LOGS OF FACIES VARIATIONS
(continuous).

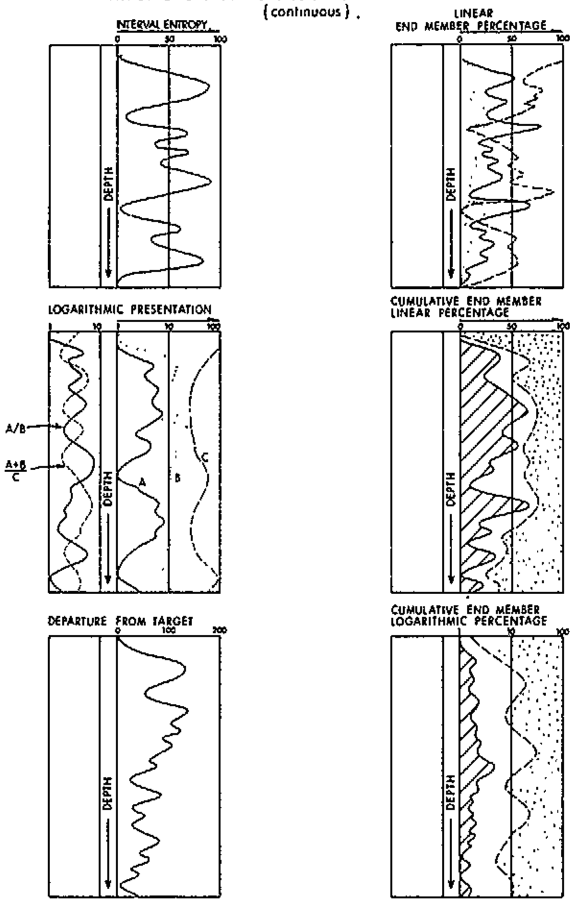


FIGURE 21

HYPOTHETICAL VERTICAL LOGS OF MULTIVARIATE ANALYSES ETC.

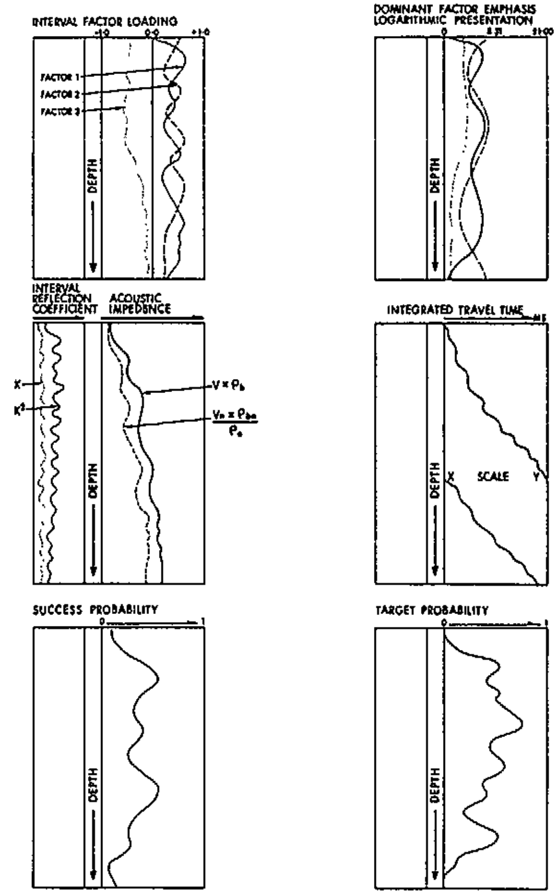


FIGURE 22

AMDEVCO SUBPROGRAM SYSTEM

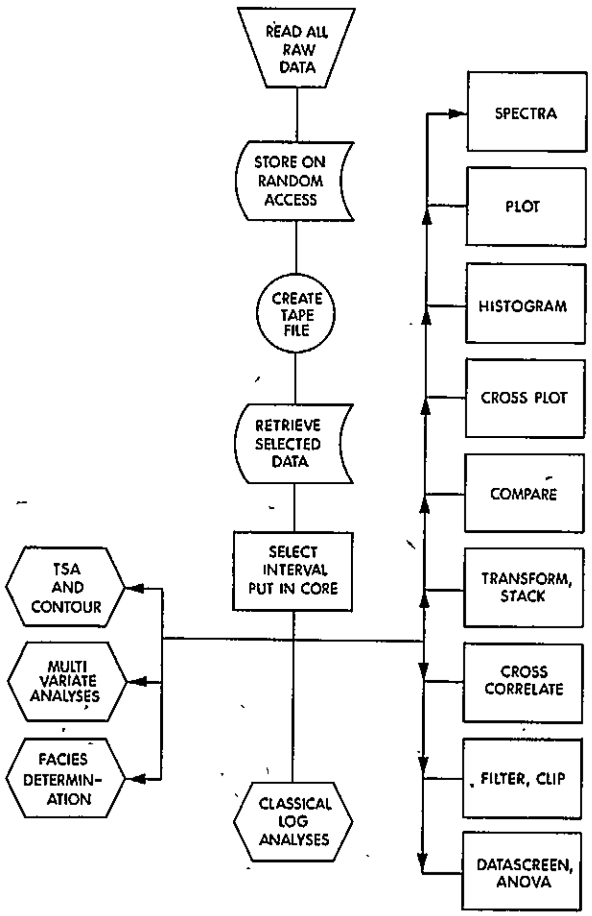


FIGURE 23

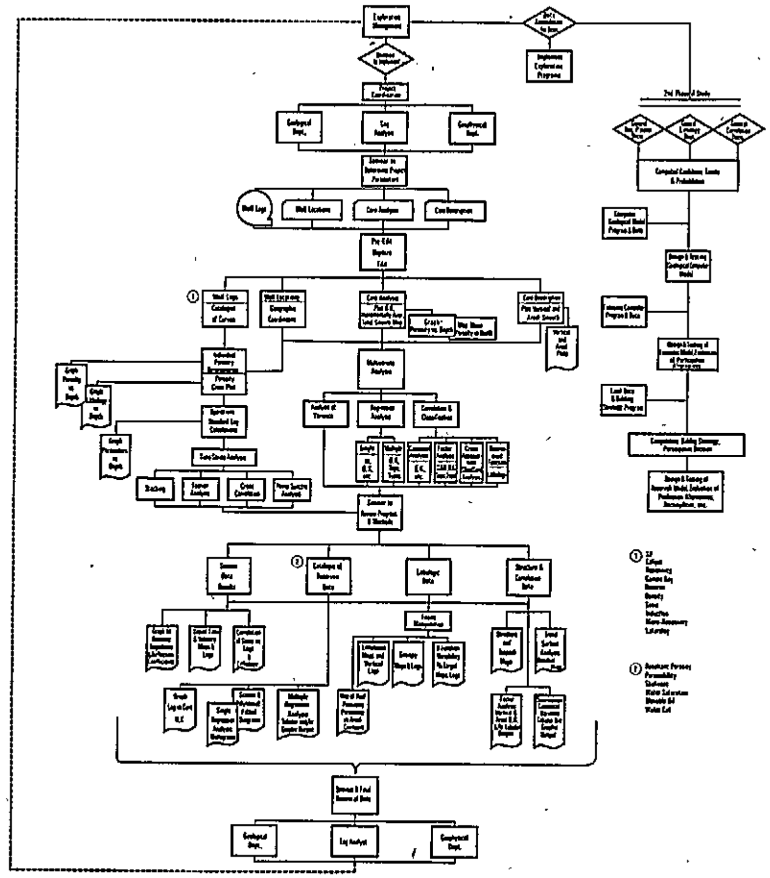


FIGURE 24