استفاده از روشهای آماری چند متغیره برای کشف و تفسیر بی‌هنجاریهای نفتی در منطقه یاکولا در چین

نوشته : دکتر محمد رضا حسین نژاد* و دکترحمید رضا رمضی**

Application of Multivariate Data Analysis to Determine Geochemical Hydrocarbon Anomaly in Yakela, China

By: Dr. M. R. Hosseinnejad* & Dr. H. R. Ramazi **

چکیده

بسیاری از روشهای آمار کلاسیک درتحلیلها ی عددی و در بررسی پدیده‌های مختلف طبیعی به‌کار می‌روند ولی متأسفانه بسیاری از این روشهای آماری در تعبیر و تفسیر داده‌های فضایی (مکانی) بویژه در علوم زمین کاربرد ندارند. این بررسی، به‌منظور اثبات توانایی روشهای آماری چند متغیره و استفاده از آنها در تفسیر داده‌های فضایی یک منطقه صورت پذیرفته است.

منطقه یاکولا در شمال باختر کشور چین یکی از مناطق نفتی شناخته شده است. در این منطقه، عملیات اکتشافی و تحقیقاتی زیادی صورت گرفته و به همین دلیل حجم زیادی از اطلاعات، موجود است. این اطلاعات با استفاده از روشهای آماری مفید چند متغیره از قبیل کریجینگ ، میانگین متحرک ، آنالیز فاکتوری، آنالیز خوشه‌ای و فیلترینگ فوریه مورد تجزیه و تحلیل قرار گرفت و نتایج آن منجر به تأیید چهار بی‌هنجاری در حال بهره برداری و دو بی‌هنجاری جدید شد که پس از آن، وجود نفت و گاز در آن دو بی هنجاری جدید، با انجام عملیات حفاری به اثبات رسید.

کلید واژه‌ها: روشهای چندمتغیره، کریجینگ، فیلترینگ فوریه، میانگین متحرک، بی‌هنجاری نفتی

Abstract

Many statistical tools are useful in developing qualitative insights into a wide variety of natural phenomena; many others can be used to develop quantitative answers to specific questions. Unfortunately most classical statistical methods make no use of the spatial information in the earth science data. A large number of useful or potentially useful multivariate techniques are available, e.g., kriging, moving average, factor analysis, cluster analysis, and fourrier filtering as well as an increasingly large number of them are being applied in geologic and geochemical investigations. Yakla is a typical and important oil and gas region in north Tarim basin in China. Therefore, a lot of exploration work has been already carried out in this area. Samples collected from the area were analyzed for their hydrocarbon mass fractions and multivariate techniques combined with other data helped the authors to identify the hydrocarbon anomaly in Yakela.

Keywords: Multivariate techniques, Kriging, Fourrier filtering, Moving average, Hydrocarbon anomaly.

Introduction

Statistical methods are a way of recognizing anomalous populations in surface geochemical data. However, not all-surface geochemical methods utilize statistics in analysis. Soil-gas and major and minor elements data are usually subjected to evaluation by histogram, means and standard deviations, ratios, filtering methods, transformations and, to lesser extent, chi-square distribution and multivariant analysis. Iodine, helium, microbial, and Eh/pH data evaluations are usually restricted to histogram, mean and standard deviations, and filtering methods. These types of data can be entered as variables in a transformation function involving the previously mentioned gases and elements. Radiometric data analysis varies, but usually point data are evaluated like iodine or helium data. Continuous radiometric profile data are typically evaluated based on the slope of the line and visual examination (Le Schack, 1997; Thomas, 1995).

Statistical Analysis and spatial variation

Several statistical techniques are useful in the evaluation of surface and near-surface geochemical data. The numerous statistical methods available to the reader are in a sense limitless in manipulating data for evaluation. However, the present level of understanding of surface geochemical exploration for petroleum precludes using several of the statistical methods and instead focuses on a few simple methods that aid interpretation. Because there may be multiple variable from a surface geochemical survey , analysis typically does not extend beyond (1) simple profiles; (2) use of means, modes, and standard deviations; and (3) contouring of the data. Usually, the raw data from the lab analysis indicate the presence or absence of an anomaly without statistical manipulation. Implementing statistical analysis requires obtaining sufficient data to identify that part (if any) of the population related to micro seepage and then determining how best to conclude confidently that it is indeed anomalous and is the art of the population being explored. Investigations typically use simple statistics. Some have developed proprietary equation, that is, transformations on the form of complex functions, ratios, and filtering, that are supposed to identify data that help to isolate the truly anomalous from false indications and background. Actual practice suggests, however, that these manipulations are not universally applicable. Extensive transformation can skew or stretch a normal population data set into appearing anomalous when it is not, which can lead to misinterpretation.

Statistical methods, in any science, are used to determine the presence in a data set of separate, not always distinctive, populations that are not clearly discernible by visual evaluation. The interpretation of data in petroleum or other natural resources seeks populations that are termed anomalous indicating the presence of the commodity being sought. In general usage, an anomaly is something that departs significantly from the norm. In surface geochemistry, anomaly means a deviation amount of hydrocarbons, iodine, microbes, radioactive minerals, trace and major metals, and helium in the soil.Thus, when the anomaly is incorporated into model, it is no longer anomalous but part of the normal population. Anomalous may therefore be a misnomer in surface geochemistry. Macro seepage and micro seepage are clearly part of the hydrodynamic system in any basin, to a degree that depends on the amount of petroleum present. Prolific basins and areas, such as Monterey, the North Sea, Iran, Saudi Arabia, Kuwait, Illinois, and the gulf Coast, contain tremendous amounts of hydrocarbons today as they did in the past. Numerous geochemical expression anomalies did and do exit at the surface in these areas, and technically they are the part of the normal geochemical soil framework. While it may be true that word anomaly is incorrectly applied in surface geochemistry, this definition is embedded in the usage of the profession: *An anomaly represents that an anamaly is the top or low a population can be normal , lognormal or any other distribution pattern , that is not in the trend of population.*

The statistical methods most appropriate for geochemical interpretation will be presented here.

1. Recognition of multiple populations (histograms)
2. Methods used to depict variations in the collected data (probability plots)
3. How to establish background and normal responses and recognize changes in those population that do not necessarily imply the presence of anomalies (means, odes, and standard deviations)
5. Transformation
6. Use of pattern recognition where classical statistical methods fail
7. Multivariate analysis including factor and cluster analysis, kriging and moving average methods.

Statistical analysis of surface geochemical data is used for two reasons: (1) to interpret a data set that exhibits multiple population and complex pattern that are difficult to correlate with geologic and geophysical data, and (2) to evaluate small data sets. Some have thought that statistical manipulation of data, as applied to a smal data set, is really done to enhance and generate an interpretation for erroneous data. A third form of evaluation i.e. pattern recognation has not been widely used in petroleum exploration. In large databases, pattern recognition is more applicable than either simple or complex statistics, which may or may not provide a supporting role. The reader should not misinterpret the last group of statments to mean that statistics do not have a place or are not really useful in surface geochemistry.

Computer contouring is a similar artificial procedure, even though it can provide an acceptable interpretation with many types of geologic data. Unlike structural and isopach mapping, geochemical values typically demonstrate wider variation over short distances and may not be suited for interpretation with present algorithms. But Fractal and fuzzy logic analyses have good and efficient applications in surface geochemistry (such as fuzzy application in geostatistics and fractal in rastar process).

Because of failing simple statistical methods in Yakla and Yakla-Luntai sophisticated methods were used such as Kriging, furrier filtering, moving average, cluster and factor analysis.

The purpose of surface geochemistry in petroleum exploration is to discover abnormal geochemical patterns or anomalies, related to hydrocarbon seepage and from which to infer the presence of a petroleum accumulation at depth. Determining that a geochemical anomaly is related to petroleum seepage requires more than just unusually high (or low) values of gases, liquids, solids, or components in the sampled media. The geochemical anomaly is defined as a departure from values that are considered to be normal background variations or nonpetroleum seepage populations in geochemical setting.

Numerous geochemical anomalies" may be unrelated to petroleum seepage, and a primary goal of statistical analysis is to identify anomalies resulting from other causes. Success in geochemical interpretation depends on the objective use of statistics.

The goal of statistics is to provide a fundamental starting point from which to evaluate geochemical data with other types of nonrelated techniques. Statistics will fail if it must provide answers when the data are insufficient or cannot support geologic and geophysical models.

Populations and Distributions

The composition of natural material exhibits a distribution that can be useful in determining whether there are different groups or populations present. All types of geochemical data have a range of values that can be expected in any sampling program, and it is necessary to decide what fraction, if any, of the samples represents the anomalous population sought. The classic example of a population distribution is in the form of a histogram. Histograms were visually analyzed, and then the data were contoured based on the artificial determination of frequency intervals. Data skewed to the right are termed a positive skewness. It is the more common histogram form in petroleum exploration. Data skewed to the left are called a negative skewness. It is not a common form and, when present, can result from data that are more representatives of the anomalous population rather than of the background values or of a normal population with relatively low values.

Figure 1 is distribution diagrams for C1, C2. C3, iC4, nC4, iC5, nC5 and uf365 concentration in 195 soil samples before data preprocessing in Yakla oil field. The samples are divided into class intervals, and the numbers of samples in each class are plotted on a bar graph. Frequently, geochemical data with anomalous population are skewed to the right, have more than one peak, and can be termed multimodal. Therefore we can not establish the bell-shaped populations.In some cases negative skewness is more important than positive one. For example , when we look leached zone in nature , negative skewness is much important. So that , due to our subjects , both negative and positive skewnees are important.

Most geochemical data exhibit some form of lognormal distribution. The lognormal transformation is the only common form present in geology. The angular transformation of bimodal data and the square -root transformation of Poisson data for discrete distributions are rare and generally not helpful in practice (Bjorklund & Gustavasson, 1987; Kaiser, 1958; Owen et al ., 1991).

Lognormal distributions are useful only when the variance is high. Small variances typically are not justifying their use.one of the other transformation non-normal distribution is power transformation which is called cox&box. This transformation is very effective and useful in most types of unnormal sistribution.

A more elegant way of analyzing distribution is to use a cumulative frequency or probability plot. The geochemical data from a survey are sorted into increasing orders to determine the range of data. They are then plotted on a special scale called a cumulative percentage versus concentration. The cumulative percentage scales represents the number of samples with values less than given value converted to percentage of the total number of samples in the survey. The formula for a lognormal frequency distribution is:

$$f(w) = \frac{1}{w\beta\sqrt{2\pi}} \exp[-1/2B(Lnw-\alpha)^2]$$

$$f(w) = \frac{1}{w\beta\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{m(w)-\alpha}{\beta}\right)^2\right]$$

These equations are from Koch and Link (1971). If the value of B^2 *(variance)* is small, so is the skewness, and the distribution will be normal.

Figure 3 is the cumulative percentage plot for soil-gas propane from Yakla. The data are plotted as the cumulative percentage of the propane concentration in the soil gas. The diagram indicates that there isn't a reasonably good straight line for the bulk of data. The sample line in this concentration range is likely to have a lognormal range of background value for the area of the survey. The upper group of samples shows an upward deflection, which may be an indication of an anomalous group of samples, an analytical problem, or some other. This upper group of samples is the data to which explorationists are likely to direct their attention. This kind of diagram can provide an unbiased way of setting a boundary between anomalous and background samples.

Transformations

A transformation is a function applied to an observation or data set that subsequently defines a new set of data. Transformations has been applied in surface geochemistry for petroleum exploration more as a filter than as a way to get more definitive results (Xie-Xuejing, 1992). Transformations are typically applied to nonlinear or lognormal distributed data. A common transformation is to take the log of a data set (lognormal distribution) that is skewed, and the result is a new data set. Expanding on this, many acquired data sets can be difficult to interpret. For example, soil-gas data across a survey may not show anomalous values for all the various C1 through C6 hydrocarbons in any group or specific samples. For a problem like this, each sample varies in term of which hydrocarbons are anomalous and may not be comparable between sample sites. A solution has been the transformation. The transformation function is not limited to soil-gas data, nor is it restricted to using only the data acquired with the same method. Hydrocarbon data can be entered into a function that can include iodine, radiometrics, major and minor elements, and microbial values. Therefore, the idea of a transformation is to create a function that brings several parameters together to yield a resulting number that can be placed at the sample location and be simply contoured or can be identified as an anomaly or background. There are three commonly used forms of transformations: (1) ratioing of one hydrocarbon to another hydrocarbons, (C1/C2) (2) smoothing the data for any particular variable, and (3) bringing together a variety of variables that are unrelated and converting them into a new variable easier to work with. (C2+)

Data Filtering

A filter is a mathematical operator that changes a time series into another time series having some desired form. Various forms of filtering have been employed to manipulate data in order to eliminate noise and to determine the anomalous areas from the nonanomalous areas. In tasks of geochemical exploration of petroleum, noise can be isolated anomalous or nonanomalous values that cause disrupting presence. Noise can obscure and confuse analysis by including additional area with the true anomalous area, or it can create a false anomaly. Noise that is anomalous or that seems to be anomalous can be inadvertently and subsequently included in the interpretation. Noise usually has nothing to do with the anomaly being defined. The source of the noise can be in sample contamination, changes in soil chemistry, and minor accumulation of petroleum seeping of present on the surface. Filtering of noise can take a variety of forms from determining background and anomalous values to a more mathematical manipulation.

Moving Average

The moving average is a typical method employed for traverse data and the general equation is

$F_t = [A_t + A_{(t-1)} + - - + A_{(t-(1+n))}] / n$

Figure 4 shows the plotted raw C1 data in Yakla. To filter data every three are added together and divided by 3: F_t becomes the new value for A_t. The resulting data are recontoured or new profiles are presented (Figs. 5 & 6).

Two-dimensional moving averages are an extension of data-smoothing techniques. Generally, a variable must be estimated at a series of point on a grid, or values are to be assigned to successive adjacent squares or rectangles on a map. The data on which the estimates are based are scattered through the map area, and may or may not lie on a grid. Z pattern, analogous to the smoothing interval of time –trend analysis, is centered on the first point to be estimated. All data points within the pattern, perhaps a square or circle, are weighted in same manner and used to estimate the center of point. In the simplest type of moving average, the mean of all observations within the pattern is applied to the estimated point. The pattern is then moved to the next grid intersection and the process is repeated. When one row or column of the grid has been evaluated, the next row or column is taken, until the map area has been completely covered. Any moving average is an expression of the general model $Y_{ij} = \sum_{k=1}^{n} W_k Y_k$

That is, an estimated grid value Y_{ij} is based on the weighted sum of *n* adjacent observation Y_k. The nature of the weighted function varies from one moving average scheme to another.

Most moving average techniques consider distance from the estimated point to the adjacent points in same manner. In the contouring program, the distances to *n* nearest points are measured directly, and each point is weighted accordingly. Methods analogous to one-dimensional smoothing equation require that the data be located upon a grid. Then, the spatial relation between Y_{ij} and each value of Y within the moving average interval is known. In this case, weights remain constant for equivalent points of Y_k as the moving average surface estimated successive value of Y_{ij}. The end two stations are eliminated in this process. The data were plotted on the map and contoured (Figures 8 ,9 and 10). Comparing maps indicates that the structure properties of geochemical field has become clearer and data filtering has achieved obvious effect.

Multivariant analysis

Multivariant analysis is usually applied when a large data set is available with a large number of parameters for each sample. This form of analysis is often used when data set are acquired under different sampling conditions, when soil conditions have been ignored, or when different surveys must be compared. Multivariant analysis upplies the matrix algebra. A matrix is a rectangular array of numbers that looks exactly like a table of data. Here we consider the matrix to be a single unit rather than individual entries. The first subscript identifies the row and the second for the column. The numbers may represent sums of observations, terms in a series of simultaneous equations, variances and covariances (the joint variation of two variables around a common mean), or any set of numbers. The matrix needs to be of equal numbers, or there will be left-over elements and the operation cannot be completed.

The multivariate analysis uses groups or clusters of data. Many naturally occurring spatial distributions show a pronounced tendency toward clustering. Most clustered distributions are typically regarded as combination of two or more simpler distributions. One of the distributions describes the pattern of individual points around the center of the cluster. Multivariant analysis attempts to identify the locations of clusters and the members that belong to a heterogeneous data set.

Cluster Analysis -General Purpose

Cluster analysis is used when a model or control survey is not available and thus a comparison cannot be made. This form of analysis performs classification by assigning observations to groups so that group is homogenous and distinct from every other group.

The term cluster analysis (Cheng – Qiuming et al., 1997; Agterberg & Bonham-Carter, 1989; Desheng et al., 1996) actually encompasses a number of different classification algorithms. The general categories of cluster analysis methods, are: joining (tree clustering), two-way joining (block clustering), and k-means clustering. The purpose of this algorithm is to join together objects (e.g., elements) into successively larger clusters, using some measure of similarity or distance. A typical result of this type of clustering is the hierarchical tree. The method calculates the similarities between all the pairs and recalculates the matrix by averaging the similarities which combined observations have with other observations.

Block clustering or two-way joining procedure cluster both variables and cases simultaneously. Two-way joining

is useful in (the relatively rare) circumstances when one expects that both cases and variables will simultaneously contribute to the uncovering of meaningful patterns of clusters.
In general, the k-means method will produce exactly k different clusters of greatest possible distinction. With mutual similarity methods, observations having a common similarity to other observations are grouped together. The data are computed on a matrix of n×n, and those observations having a close intercorrelation will be closed to +1.

Factor Analysis -General Purpose

The main applications of factor analysis techniques are (1) to reduce the number of variables and (2) to detect structure in the relationships between variables, which is to classify variables. Therefore, factor analysis is applied as a data reduction or structure detection method in the case of large data such as geochemical exploration field. This form of analysis is expressed in the form of a pattern that is present in the variance and covariance and in the related similarities between observations. Factor analysis is best applied when it is assumed that parameters are varying across a survey area; the analysis can determine which parameters are varying (Voudouris, 1997). This method is best used with a data set derived from an exploration and model survey. For example, when there are more than two variables, we can think of them as defining a "space," just as two variables defined a plane. Thus, when we have three variables, we could plot a three-dimensional Scatterplots, and, again we could fit a plane through the data. With more than three variables it becomes impossible to illustrate the points in a Scatterplots, however, the logic of rotating the axes so as to maximize the variance of the new factor remains the same.
The application was completed in the following steps (Stark, 1991):
First, the raw data were standardized using! $z_i = \frac{x_i - x_m}{s}$
Where z_i is ith value of the standardized variable z (with a mean of 0 and a standard deviation of 1). X_m is the mean value of variable x and s is the standard deviation. Standardization tends to inflate variables whose variance is small, and reduce the influence of variables whose variance is large. Also, the standardization procedure removes the influence of different, incompatible units of measurement on the data by making them dimensionless. Next, the variances/covariance and correlation coefficients of the variables were computed using !

$$r_{x,y} = \frac{\sum_i (x_i - x_m)(y_i - y_m)}{\left\{\left[\sum_i (x_i - x_m)\right]^2\left[\sum_i (y_i - y_m)\right]^2\right\}^{\frac{1}{2}}}$$

Where x_i and y_i are the ith values of the standardized variables x and y and x_m ,y_m are their respective means. The correlation coefficients ($r_{x,\,y}$) are presented in matrix form (correlation coefficient matrix).Eigenvalues and eigenvectors were calculated for the covariance matrix. The data were transformed into factors and the numbers of factors were selected. Unfortunately, there are no universally agreed upon criteria for the selection of the retained number of factors. There are many suggestions for selecting the optimal number of factors (amount of the cumulative variance, eigenvalues greater than 1,scree plot) (Kaiser, 1958; Davidson, 19947), but the most straightforward solution to this problem is to extract as many factors as the ruling theory demands.
The application of statistical techniques to the geochemical data is a general way to organize the data and summarize their characteristics. In order to study the relationship between variables or indexes used in oil/gas geochemical exploration, factor analysis method was used. The geological interpretation of the factors gives an insight into main processes, which may control the distribution of hydrocarbon parameters.
The data were transformed into factors and the numbers of factors were selected. Five factors have been selected, which explain more than 94% of the total variance (Table 1). Except Hg and Uf365, others variables have high communality. Thus a good description of them has been obtained by using the four-factor model. And lastly, the contribution of each factor at every site (factor scores) was computed. This step was important for the mapping of geographical distribution of each factor.
The result shows that factor ***I*** accounts for 62% of the total variance and C2,C3,iC4,iC5,nC5,C2+ are closely related to one another and they would have a very similar distribution property in space. Factor ***I*** is a major independent factor of oil/gas geochemical exploration anomaly, and is called anomaly properties integration. Factor ***II*** accounts for 14% of the total variance with C1/C2+ as a main associated variable which suggests that C1 is important to the anomaly property integration. Factor ***III*** and ***IV*** similarly account for 8% of the total variance, show that Hg and UF365 are almost independent variables and probably indicate meaningful hydrocarbon migration.
By comparing and analyzing the effectiveness of various data processing methods we defined that universal kriging, moving average and specially fourrier filtering can effectively reveal oil and gas anomalies.
The method to delineate oil and gas anomaly has been directly put forward in accordance with annular characteristics formed by local peak geochemical index concentration. This method can clearly and effectively delineate the weak anomaly under low background and reduce the artificial character of delineated anomaly. At present, the most common method is to delineate peak area with constant, then determine such oil and gas anomaly as annular anomaly, according to the peak area distribution. This delineation method is complex and accompanied with lots of difficulties. For example, if the lower limit of the fixed anomaly is added by 1-0.5 standard deviation, then a weak anomaly under lower background can not be easily reflected. On the other hand, the second delineation is very artificial. Another common method is trend residual error and it also needs two steps delineation.

Conclusion

In the current work, attention has been given to improve methods of interpretation by means of applying Multivariant Data Analysis to some petroleum geochemical exploration data. Specifically, the hard question addressed is multivariate statistical techniques, which is now common in geochemical exploration, providing a significant advantage over intuitively simpler univariate or non-statistical techniques? The use of all these methods is dependent on the quality of data collection and analysis, density of sampling, and number of samples in a data set. Other variables addressed more subjectively are the effects of soil, weather, and topography, which usually can be identified and those samples normalized or discarded. The overuse of statistics to support the presence of an anomaly when biased data lead to this type of conclusion seems, to be a continuing problem. Experience and the use of pattern recognition methods help to minimize this problem.

Table 1- Correlation coefficient between the ten elements.

C1 C2 C3 iC4 nC4 iC5 nC5 Hg C1/C2+ UF365
C1 1.00
1.00 C2 .59
1.00 .99 C3 .52
1.00 .98 .96 iC4 .46
1.00 .94 .94 nC4 .40 .92
1.00 1.00 .94 .94 .92 iC5 .40
1.00 1.00 1.00 .94 .94 .92 nC5 .40
1.00 .04 .04 .04 .06 .06 .05 Hg -.12
1.00 .05 .94 .94 .94 .98 1.00 1.00 C1/C2+ .55
1.00 -.38 -.11 -.48 -.48 -.48 -.46 -.41 -.34 Uf365 .48

Table 2- Factor Loadings (Unrotated) Yakla data principal components

(Marked loadings are > .700000)

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
C1	**.529751**	**-.803780**	-.223739	-.084763	-.041117
C2	**.980826**	-.087356	-.040214	-.016297	-.002774
C3	**.989414**	-.011020	-.025484	-.006213	.018912
iC4	**.984468**	.057294	-.009381	.008378	.004741
nC4	**.865698**	.170238	.066069	.042792	-.417531
iC5	**.958991**	.089685	.029661	.026160	.051846
nC5	**.890451**	-.138772	-.021299	.029180	.283696
Hg	.034744	.358051	-.478995	**-.800273**	.000717
C2+	**.992871**	-.043855	-.028402	-.006083	.014586
C1/C2+	-.498525	**-.816989**	-.215801	-.092090	-.090873
UF365	-.078239	.255822	**-.805269**	**.529083**	-.005530
Expl.Vr	6.894421	1.576420	.983382	.939809	.268084
Prp.Totl	.%62	%14	%8	.%8	.%2

Table 3- Groups of variables in Yakela

Group A	C1
Group B	C2.C3,iC4,nC4,iC5,nC5,C2+, C1/C2+
Group C	Hg, uf365

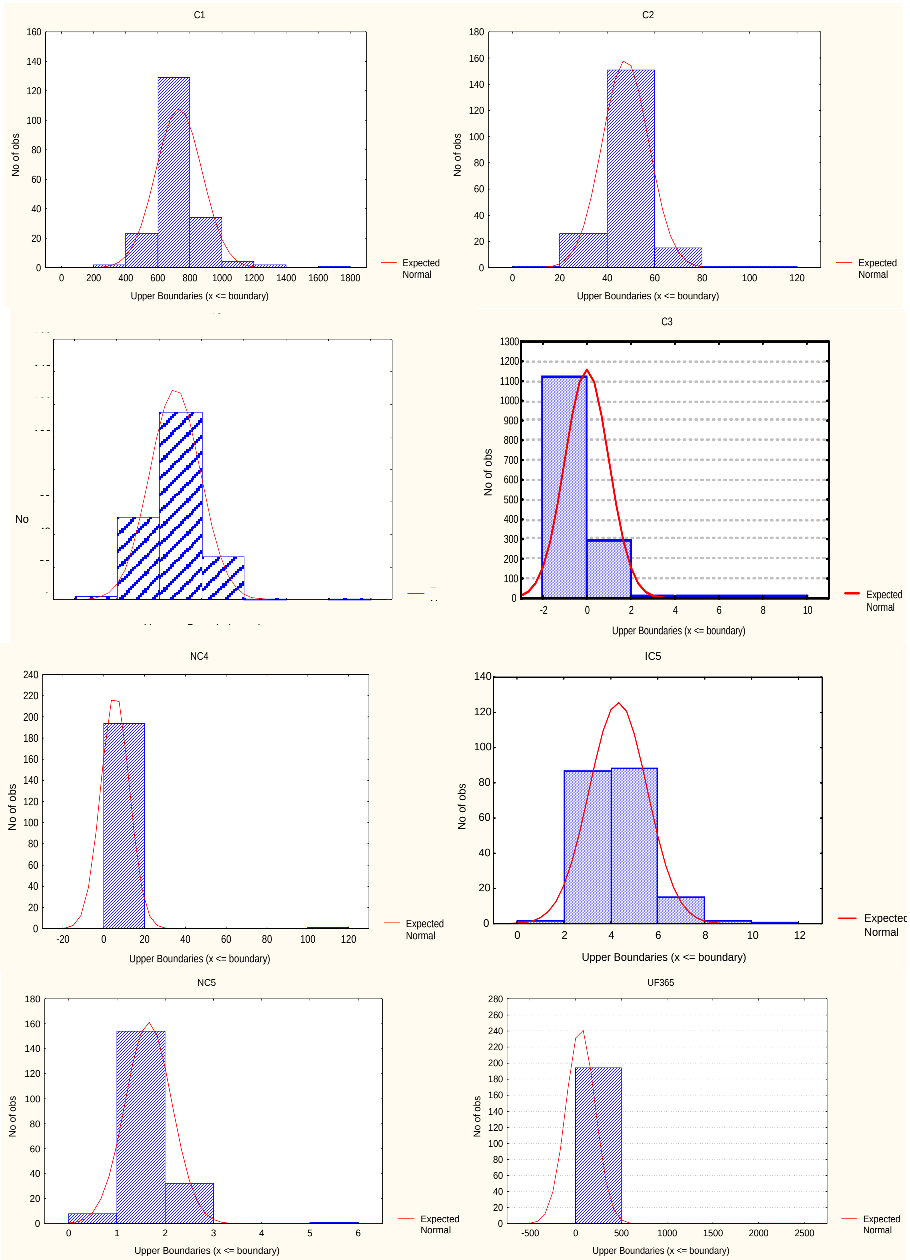

Figure 1- Histograms show the distribution of some indexes before data preprocessing in Yakla

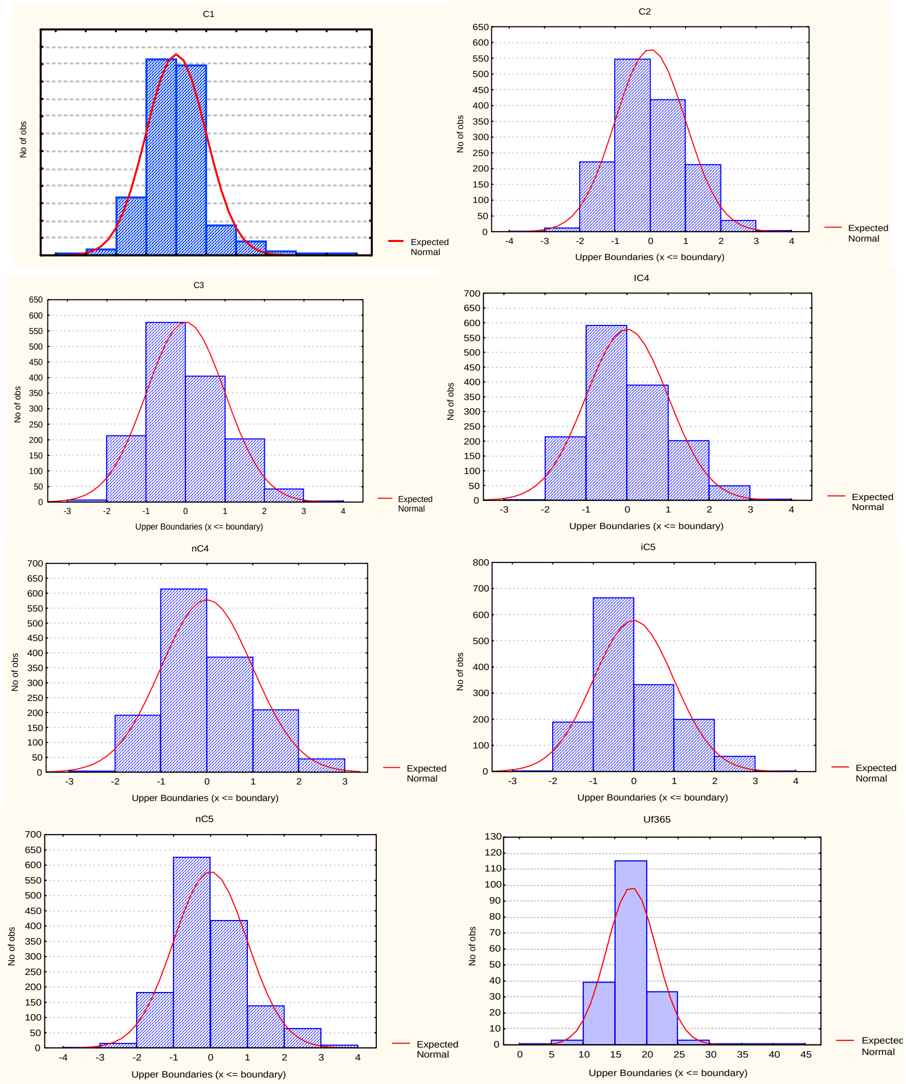

Figure 2- Histograms showing the distribution of data after treatment with data preprocessing in Yakla

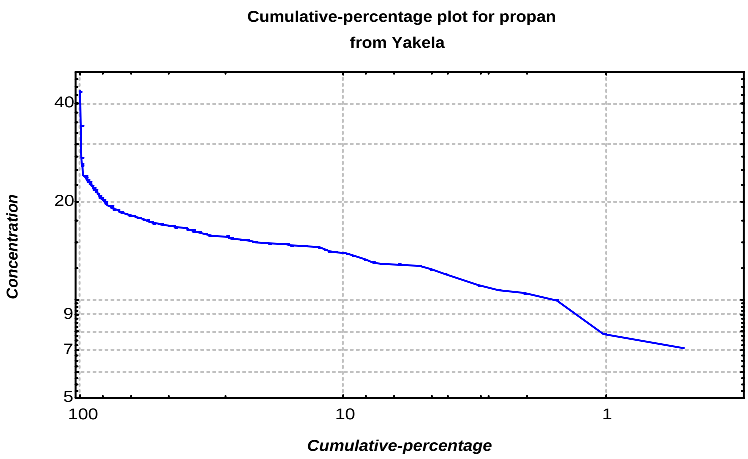

Figure 3- Cumulative-percentage plot for propane in the Yakla.

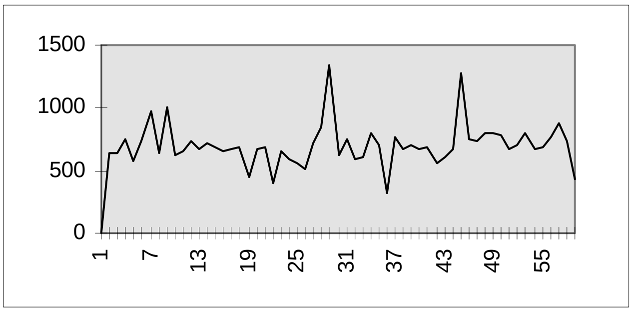

Figure 4 - C1 concentration in Yakla(Raw data)

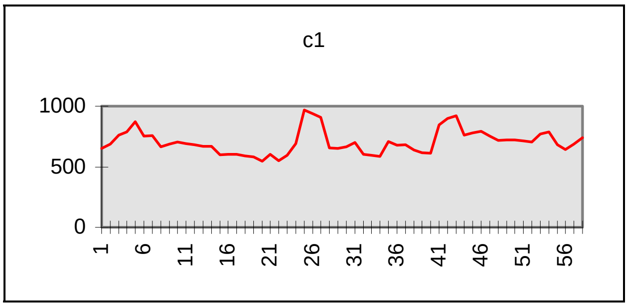

Figure 5 - The same C1 profile as figure 4 but with three-point moving average

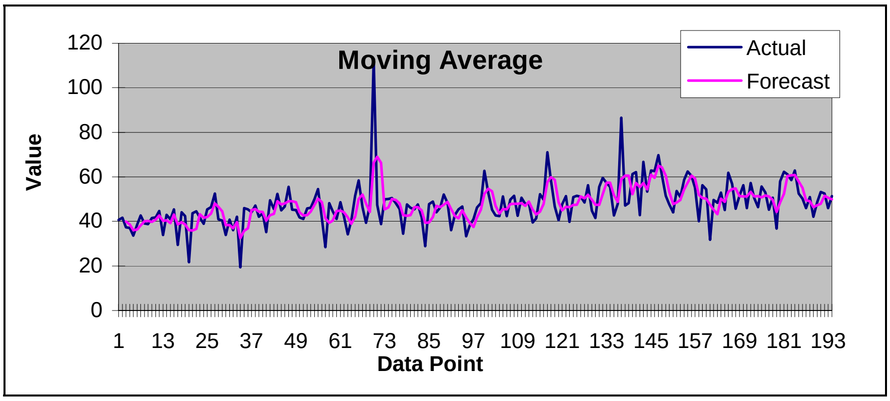

Figure 6- Raw and moving average of C2 data from Yakela

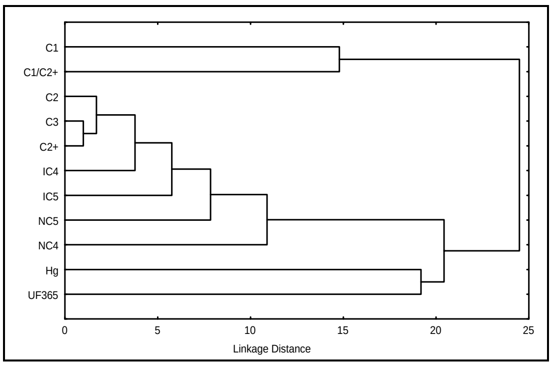

Figure 7- cluster Analysis for variables in Yakla

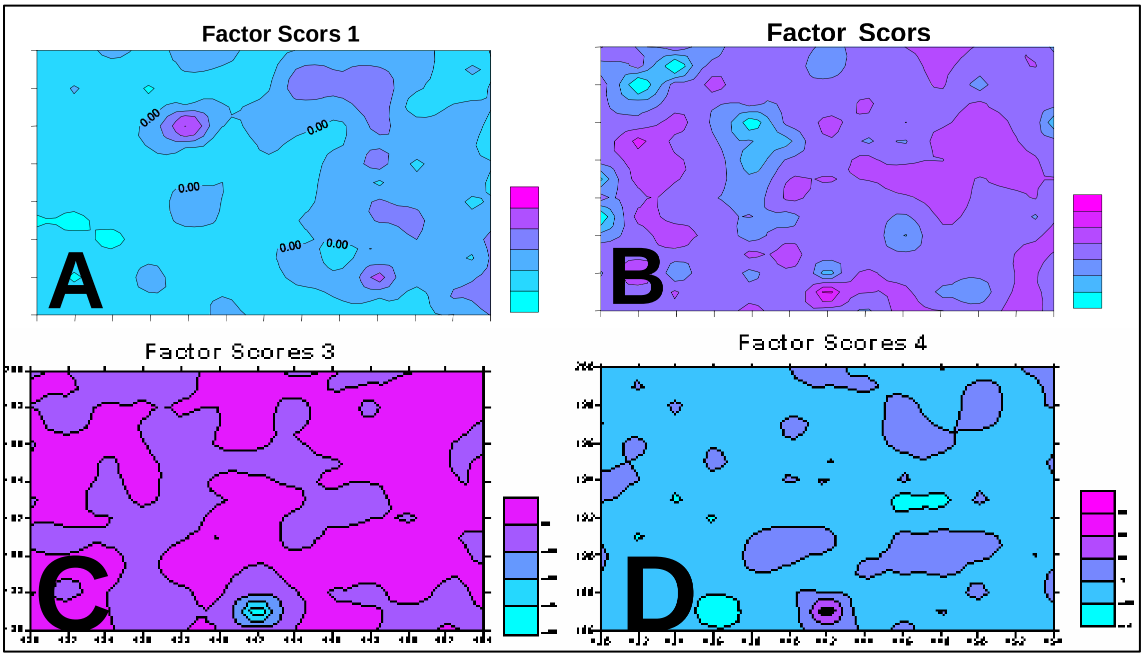

Figure 8- Distribution of factor scores for : A, factor I: B, factor II, C, factor III and D, factor IV.

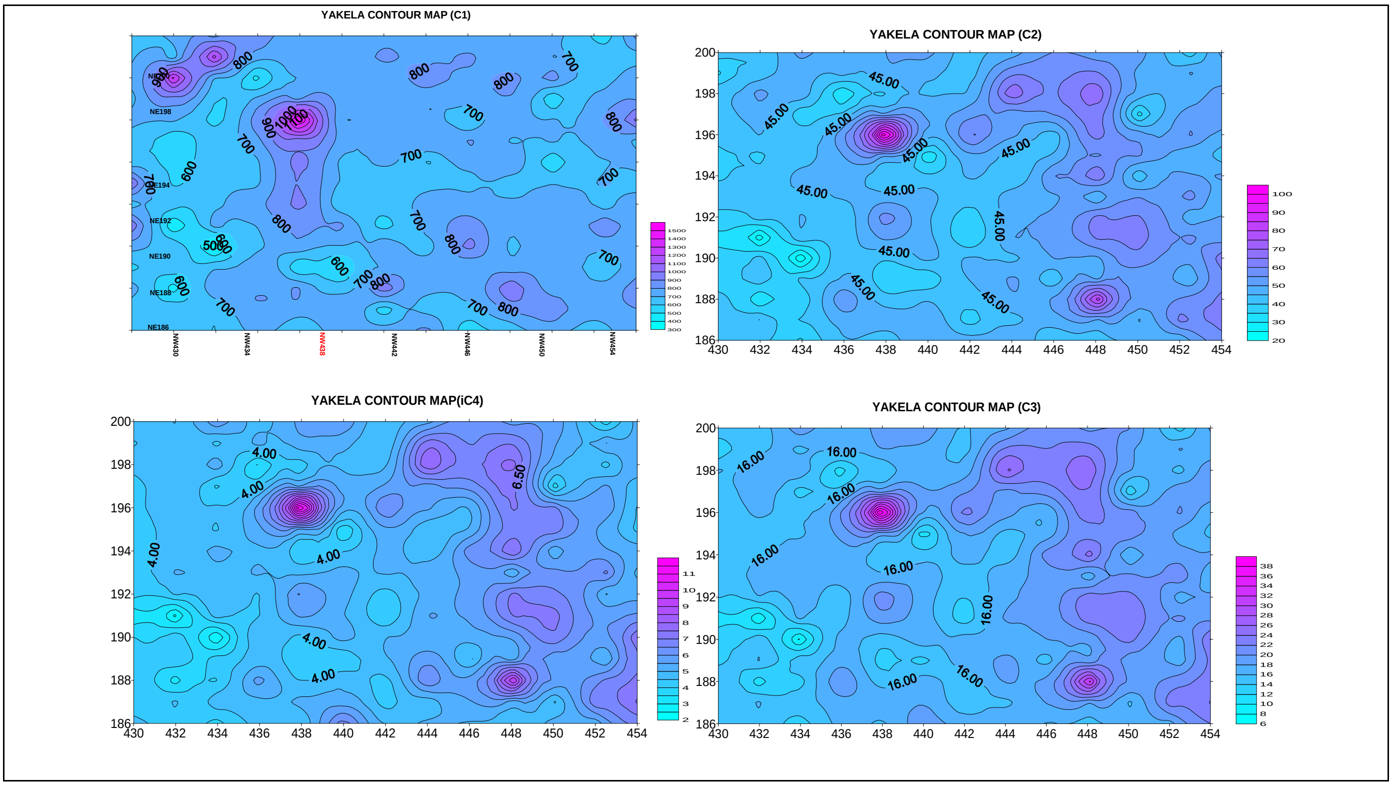

Figure 9- Contour map of hydrocarbons of some data from Yakla, xinjiang, China before filtering

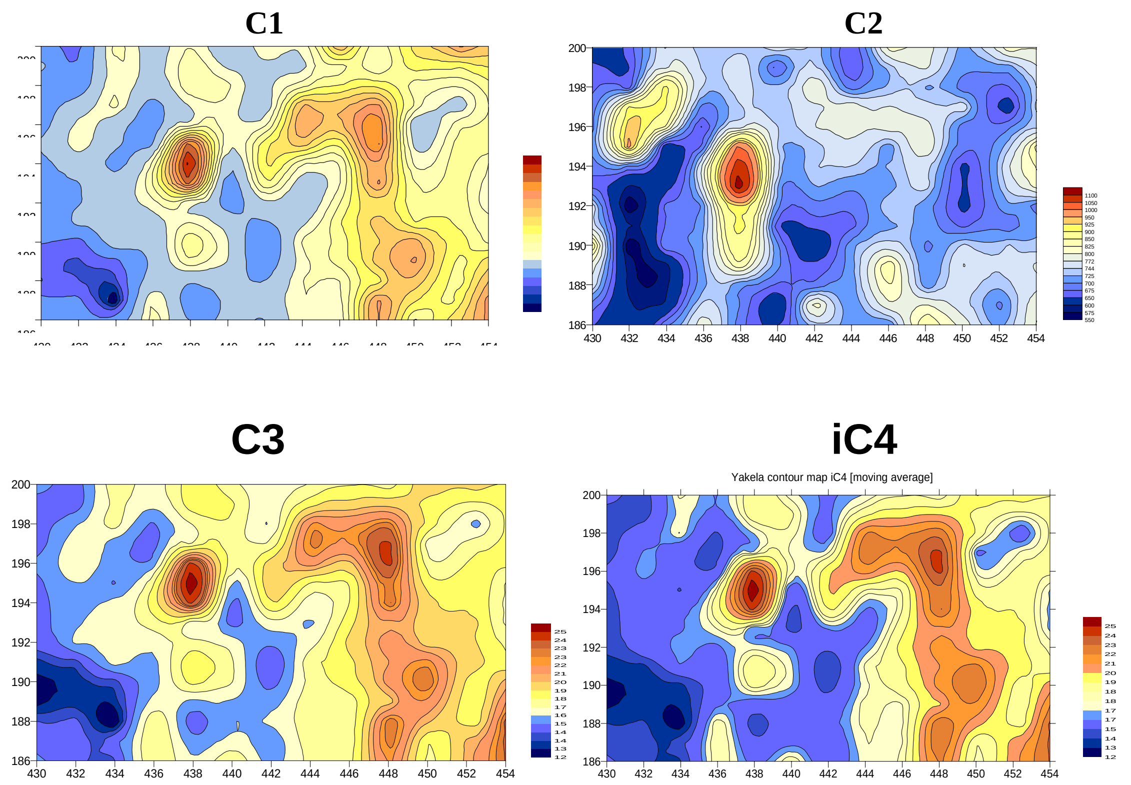

Figure 10- shows three-point moving average for the hydrocarbons data from Figure 8.

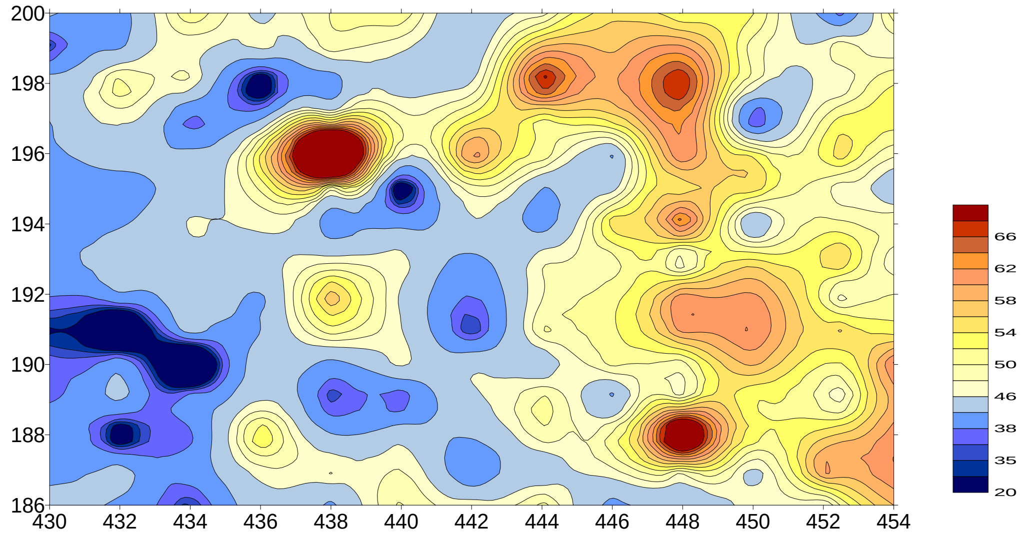

Figure 11- Contour map of hydrocarbons data before data filtering
(Yakla, xinjiang, China)

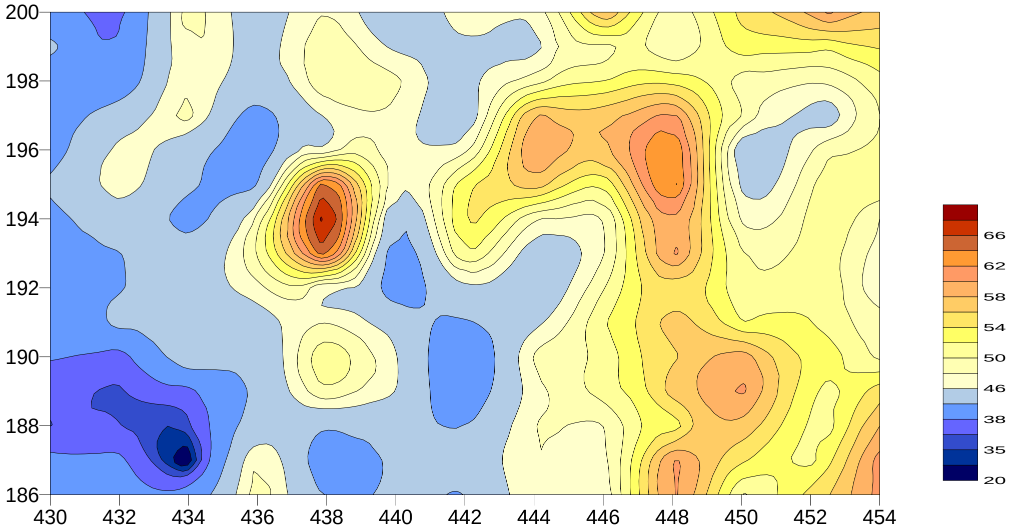

Figure 12- contour map after Furrier filtering on the basis of figure, so structure properties of field seems much clearer.

References

Agterberg,F.P.& Bonham-Carter, G.F., 1989- Statistical Application in the Earth Sciences. Geological Survey of Canada, paper 89-9.

Akaike, H., 1973- Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (Eds.), Second International Symposium on Information Theory. Budapest: Akademiai Kiado.

Akaike, H. ,1983- Information measures and model selection. Bulletin of the International Statistical Institute: Proceedings of the 44th Session, Volume 1. Pages 277-290.

Aldrich, J. H., & Nelson, F. D. ,1984 - Linear probability, logit, and probit models. Beverly Hills, CA: Sage Publications.

Almon, S. ,1965- The distributed lag between capital appropriations and expenditures. Econometrica, 33, 178-196.

Belperio, A.P., Molineux, R., Ivic, D., Sambura, A, Sandercock, N., Daly, S. J ., Morton, J. & Slade, O., 1991- Strategic Framework for Development of a Geosciences Information System. South Australia, Department of Mines & Energy. Report Book 9114.

Bjorklund, A. & Gustavasson, N., 1987- Visualization of Geochemical Data on Maps: New Options. J. Geochmical Exploration, 29:89-103.

Cheng-Qiuming; Bonham-Carter-G-F; Hall-G-E-M., 1997- Bajc-AStatistical study of trace elements in the soluble organic and amorphous Fe-Mn phases of surficial sediments, Sudbury Basin; 1, Multivariate and spatial analysis.Journal of Geochemical Exploration. 59; 1, Pages 27-46.

Cocker-Mark-D., Dyar-Thomas-R., 1993- Geochemistry and hydrogeochemistry of the Oconee River basin, Georgia; the utility of a geographic information system (GIS).Abstracts with Programs - Geological Society of America. 25; 6, Pages 350-351.

Cooke, D., Craven, A. H., & Clarke, G. M., 1982 - Basic statistical computing. London: Edward Arnold.

Cooley, J. W., & Tukey, J. W., 1965 - An algorithm for the machine computation of complex Fourier series. Mathematics of Computation, 19, 297-301.

Cooley, W. W., & Lohnes, P. R., 1971- Multivariate data analysis. New York: Wiley. Dalton, M, G.,and Upehurch, S, G. Interpretation of Hyderochmical facies by Factor Analysis: Ground Water v. 16, no 4. P.228-233.

Darcy,L. and Boston, L.,(editors), 1983- Webster New World Dictionary of Computer Terms: *Simon & Shuster, New York*,282P.

Davidson-Martin-J., 1994- "On the acceptance and rejection of surface geochemical exploration. Oil and Gas Journal. 94; 23, Pages 70-71, 74-76.

Desheng, L, Digang, L., Chengzao, at al. ,1996- Hydrocarbon Accumulations in the Tarim Basin, china,AAPG Bulletin, V,80,No,10 (October 1996). P. 1587-16043

Durance, S. and Korre A., 1996- Statistical and Spatial Assessment of Heavy Metal Contamination Around Lavrio Mine Workings, Greece, imperial College of Science, Technology and Medicine, London, UK 1-6
Ellis-James-M., 1995- Geological and environmental remote sensing for international petroleum operations. AAPG Bulletin. 79; 11, Pages 1727.
Finn, J. D., 1974 - A general model for multivariate analysis. New York: Holt, Rinehart & Winston.
In Enslein K., Ralston, A. and Wilf., H. S. (Eds.), Statistical methods for digital computers. Vol. III. (pp. 203-264). New York: Wiley.
John C., Davis, 1973 - Statistics and Data Analysis in Geology. *Kansas Geological Survey: Wiley & Sons, Inc., New York*
Joreskog, K.G., & Sorbom, D., 1979- Advances in factor analysis and structural equation models. Cambridge, MA: Abt Books.
Haney-Donald-C; Cobb-James-C., 1992- Coal resource investigations for Kentucky; traditional versus coal availability investigations. AAPG Bulletin. 76; 8, Pages 1277.
Hosseinnejad, M. R., Ebrahimi, M., 1998- Application of GIS to analysis and modeling of hydrocarbons anomaly in Yalela,China, jourrnal of Chin university of geosciences, Vol.9, No.2,p. 137-143,August.
Howarth, R. J. , 1983- Statistics and Data Analysis in Geochemical Prospecting *Elsevier scientific publishing company*
Kaiser, ZH. F., 1958- The varimax criteria for analytical rotation in factor analysis: *Psychometrika,v.23. no. 3 p. 187-200*
Koch, G.S., Jr. and Link, R.F., 1970 - Statistics Analysis of Geological Data Vol. 1 *John Wiley and Sons, New York*
Le Schack-Leonard ,A., 1997- Geochemical exploration; 1, Results of magnetic HGI and radiometric surveys in Canada. Oil and Gas Journal. 95; 20, Pages 84-89.
Owen, P. Lavin and Lan Nichol,1991- Application of Some Statistical Techniques to Weathered Rock Geochemical Data from Broken Hill area, New South Wales, Australia 421-451.
Saenz-Guadalupe; Pingitore-Nicholas-E Jr., 1989-"Surface organic geochemical prospecting for hydrocarbons; multivariate analysis." Journal of Geochemical Exploration. 34; 3, Pages 337-349.
Simpson-P-R; Edmunds-W-M; Breward-N; Cook-J-M; Flight-D; Hall-G-E-M;Lister-T-R. , 1993- Geochemical mapping of stream water for environmental studies and mineral exploration in the UK. Journal of Geochemical Exploration. 49; 1-2, Pages 63-88.
Smirnova-I-O; Gololobov-Y-N; Kots-V-G; Prishchepa-Oleg-M; Rousanova-A-A., 1997- "Role of remote sensing in petroleum exploration at Timan-Pechorian Province; from tectonic zonation to investigation of petroleum fields."AAPG Bulletin. 81; 8, Pages 1413.
Stark-P-H ., 1991- Large petroleum data bases; the three I's, integrity, integration, interface. AAPG Bulletin. 75; 3, Pages 676.
Steven A. Tedsco, 1995- Surface Geochemistry in Petroleum Exploration. *Hampan & Hall*
Thomas-John-J; Overeynder-Helen-M; Thomas-Barbara-R., 1995- Indoor radon concentration data; its geographic and geologic distribution, an example from the Capital District, NY.AAPG Bulletin. 79; 9, Pages 1419.
Von-der-Dick-Hans; Callejon-Gimenez-Angel; Bosman-Dane, 1996- "Surface exploration geochemistry; numerical data processing of near surface hydrocarbon gases and application in the Llanos Basin, Venezuela."AAPG Bulletin. 80; 8, Pages 1343-1344.
Voudouris, K,s., Lambrakis, G., Papatheothorou, and Daskalaki,P., 1997- An Application of factor Analysis For the Study of Hydrological Conditions in Plio-pleistocene Aquifers of NW Achaia (New Peloponnesus, Greece) ,Mathematical Geology,Vol. 29,no. 1.
Xie-Xuejing, 1992- "Local and regional surface geochemical exploration for oil and gas. Journal of Geochemical Exploration. 43; 1, Pages 25-42.
Xie-Xuejing; Yang-Binzhong , 1989- Application of multiparametric geochemical methods in the search for oil in the Qinggang region near Daqing oil field. Journal of Geochemical Exploration. 33; 1-3, Pages 203-213..

* دانشگاه علم و صنعت ایران، اراک، ایران
** دانشگاه صنعتی امیر کبیر، تهران، ایران

* Iran University of Science and Technology, Arak , Iran
** Amirkabir Technology of University, Tehran, Iran