물리탐사
Vol. 1, No. 3, 1998, p. 170~175

Automatic Electrofacies Classification from Well Logs Using Multivariate Statistical Techniques

Lim, Jong-Se[1], Kim, Jungwhan[2] and Kang, Joo-Myung[3]

다변량 통계 기법을 이용한 물리검층 자료로부터의 암석물리학상 결정

임종세 · 김정환 · 강주명

Abstract : A systematic methodology is developed for the prediction of the lithology using electrofacies classification from wireline log data. Multivariate statistical techniques are adopted to segment well log measurements and group the segments into electrofacies types. To consider corresponding contribution of each log and reduce the computational dimension, multivariate logs are transformed into a single variable through principal components analysis. Resultant principal components logs are segmented using the statistical zonation method to enhance the quality and efficiency of the interpreted results. Hierarchical cluster analysis is then used to group the segments into electrofacies. Optimal number of groups is determined on the basis of the ratio of within-group variance to total variance and core data. This technique is applied to the wells in the Korea Continental Shelf. The results of field application demonstrate that the prediction of lithology based on the electrofacies classification works well with reliability to the core and cutting data. This methodology for electrofacies determination can be used to define reservoir characterization which is helpful to the reservoir management.

요 약 : 이 연구는 다변량 통계 기법을 이용한 물리검층 자료로부터의 암석물리학상 결정으로 암상을 예측하는 것이다. 기술 통계 분석으로 물리검층 자료의 특성을 파악하고 주성분 분석에 의한 다변량 검층 자료들의 상관도 분석을 통해 변수들을 변환시켜 새로운 변수인 주성분을 구하고 변수들의 차원을 축소한다. 통계적 방법에 의한 주성분 검층 자료의 구획에 의한 효율적 자료 축소와 계산의 효율성을 높여 양질의 해석결과를 얻을 수 있다. 구획된 주성분 검층 자료로부터 계보적 군집 분석에 의해 암석물리학상을 결정한다. 최적 암석물리학상의 수는 전체 변동과 군집내의 변동사이의 비와 코어자료 등에 의해 비교 결정된다. 이 연구에서 개발된 암석물리학상 결정법을 국내대륙붕 물리검층자료에 적용한 결과 결정된 암석물리학상은 시추 코어 및 시추 암편 분석에 의한 암상 구분화와 잘 일치하였다. 이러한 연구는 저류층 특성인자의 신뢰성 있고 정량적인 평가로 유전 개발 및 생산 계획 시 유용한 도구로 활용될 수 있을 것이다.

Keywords : Electrofacies, Multivariate statistical technique, Wireline log data

Introduction

As the variation of petrophysical properties often corresponds to lithologic variation, lithology determination plays an important role for reservoir characterization. Subsurface lithology is traditionally determined from both core and cutting analyses. Cores are generally not continuous and do not provide complete descriptions of formations crossed by a well. Cuttings always have some uncertainties in depth and it can be difficult to restore the components and thickness of lithologic column. As a result, the lithology based on those data is not sufficiently accurate and precise for the quantitative use. On the other hand, well logs have the advantage of providing a continuous record over the entire well and can be obtained in conditions where coring is impossible. Therefore the integration of core and well log data can give a good lithologic description of the formations.

An electrofacies is defined as "the set of log responses which characterizes a bed and permits it to be distinguished from the others" (Serra and Abbott, 1982). The electrofacies derived by selecting, weighting, and combing well log data can be used as an indicator of lithology. Once good correlations between electrofacies and core analysis are established on a local basis, significant geological information can be extracted from well logs alone.

*1998년 10월 30일 접수
1) 서울대학교 공학연구소(Research Institute of Engineering Science, Seoul National University)
2) 한국석유공사 기술실(Technical Department, Korea National Oil Company)
3) 서울대학교 지구환경시스템공학부(Sch. of Urban, Civil & Geosystem Eng., Seoul National University)

In this study, a systematic technique has been developed for the electrofacies determination from well logs using multivariate statistical analysis, and applied to the wells in the Korea Continental Shelf.

Electrofacies Classification

Wolff and Pelissier-Combescure (1982) and Moline and Bahr (1995) developed the multivariate statistical procedures to determine the electrofacies. In this study, to enhance the quality and efficiency of the interpreted results, the zonation of principal components logs is constructed prior to clustering. The segmented principal logs instead of the original logs are used for cluster analysis (Lim *et al.*, 1997). The electrofacies determination procedure used in this study is summarized by the flow chart in Fig. 1.

Descriptive Statistical Analysis

Data can be revealed their features when they are adequately organized. The univariate tools can be used to describe the distribution of individual variable. Histogram is known as one of the most common and useful presentation and organization of data for univariate description. The importance features of most histogram can be captured by the summary statistics: measures of location, measures of spread, and measures of shape (Isaaks and Srivastava, 1989). In addition, the relationships and dependencies among variables are important feature of data when we analyze a multivariate data set like well logs. The correlation coefficient quantifies a relationship between two well logs which are closely related to each other.

Principal Components Analysis

To consider the contribution of each log and reduce the computational dimension, multivariate logs are combined into a single variable through principal components analysis (Doveton, 1994). If m numbers of log responses from a sequence of zones are plotted as points in a space with mutually orthogonal axes, they form a cloud in m-dimensional space. Principal components are the eigenvectors of this cloud, which are computed to locate the major axes in order of importance. The orientations of the principal components are computed from either the covariance or correlation matrix of the log data. The correlation matrix is more commonly chosen because most logs are recorded in radically different units. In order to avoid artificial and undue weighting by any of the logs, the original data should be standard normalized by subtracting the mean and dividing by the standard deviation. Total variance of the original set of m variables is the sum of their separate variances. Because principal components logs are ordered,

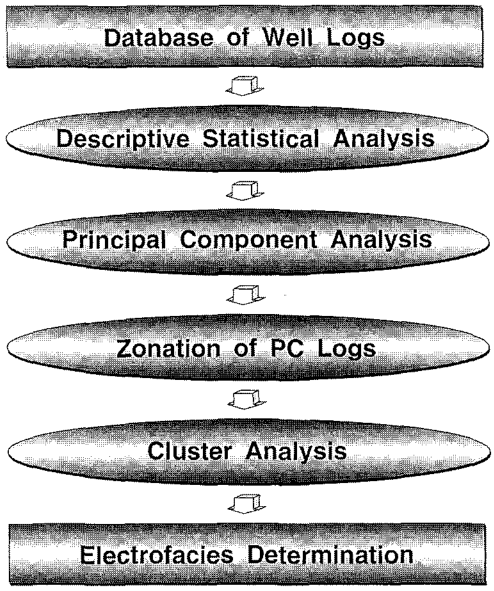


Fig. 1. Flow chart of the electrofacies determination.

the last principal components logs usually contain very little information and may be dropped from further analysis without any significant effect. This can reduce the complexity of cluster analysis.

Consider a hypothetical crossplot of density and neutron porosity readings (Fig. 2(a)). Since the two logs are recorded in different units, it is appropriate to standardize them. Standardization converts covariance matrix to correlation matrix. The correlation matrix is described geometrically by an ellipse (Fig. 2(b)). The eigenvectors of the correlation matrix locate the major and minor axis of the ellipse. These eigenvectors are called the principal components of the correlation matrix (Fig. 2(c)).

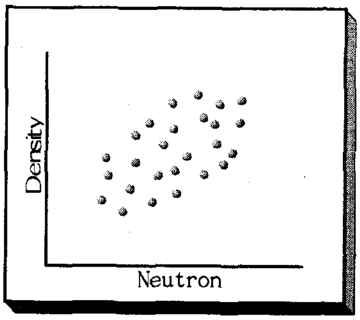
Zonation of Well Log

Well log data are so vast that the data compression is strongly required before such data are analyzed and interpreted. The data compression would enhance the quality and efficiency of the interpreted results. A number of methods have been derived to condense the well logs data into zones. In this study, the concept of analysis of variance (ANOVA) is applied for the zonation of well log data (Dharmawardhana and Keller, 1985). An optimal subdivision is attained if zones are established such that the within-zone variance is minimized and the between-zone variance is maximized. Since an observation is assigned to only one of the zones, the single factor fixed effect model is selected which is formulated as (1).

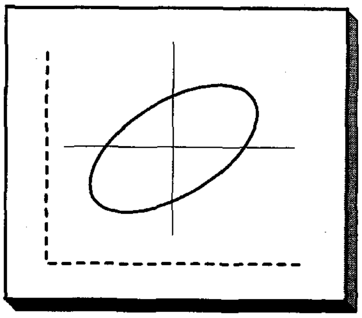
$$X_{ij} = \overline{X}_{..} + a_i + e_{ij} \quad (1)$$

where,

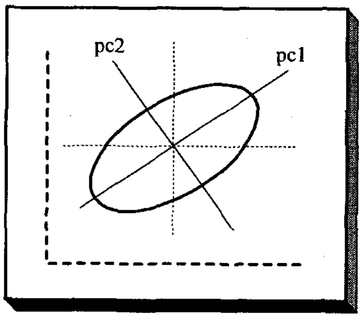
X_{ij} = jth observation of the ith zone



(a) Hypothetical neutron-density crossplot

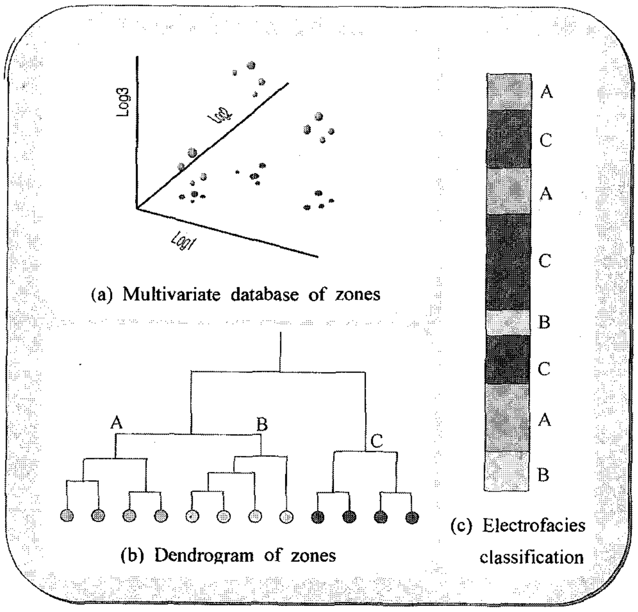


(b) Correlation matrix ellipse superimposed on standardized neutron-density axes



(c) Location of ellipse principal axes by eigenvectors

Fig. 2. Schematic diagram sequence of principal component analysis.



(a) Multivariate database of zones

(b) Dendrogram of zones

(c) Electrofacies classification

Fig. 3. Schematic diagram sequence of cluster analysis of log data.

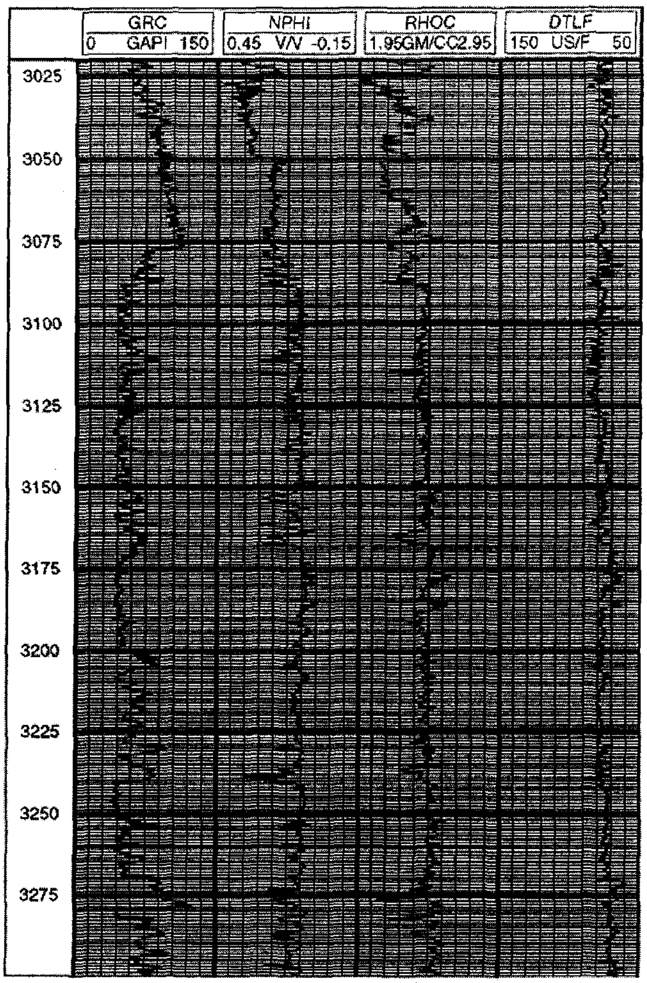


Fig. 4. Measured logs of G-1 used in the investigation.

$\overline{X}_{..}$ = grand mean of the well logs
a_i = effect of the ith zone
e_{ij} = random error associated with each observation

If the means of sub-zones are significantly different, the variance of the combined zones reflecting both the zone effect and the random error will be larger than the variance of the separate groups reflecting the random error alone. Thus, the method is based on the comparison of two independently computed estimates of variance.

$$MSTR = \frac{1}{(r-1)}\left[\sum_{i=1}^{r} n_i(\overline{X}_{i\cdot} - \overline{X}_{..})^2\right]$$

$$MSE = \frac{1}{(N-r)}\sum_{i=1}^{r}\sum_{j=1}^{n_i}\left[X_{ij} - \overline{X}_{i\cdot}\right]^2 \qquad (2)$$

where,

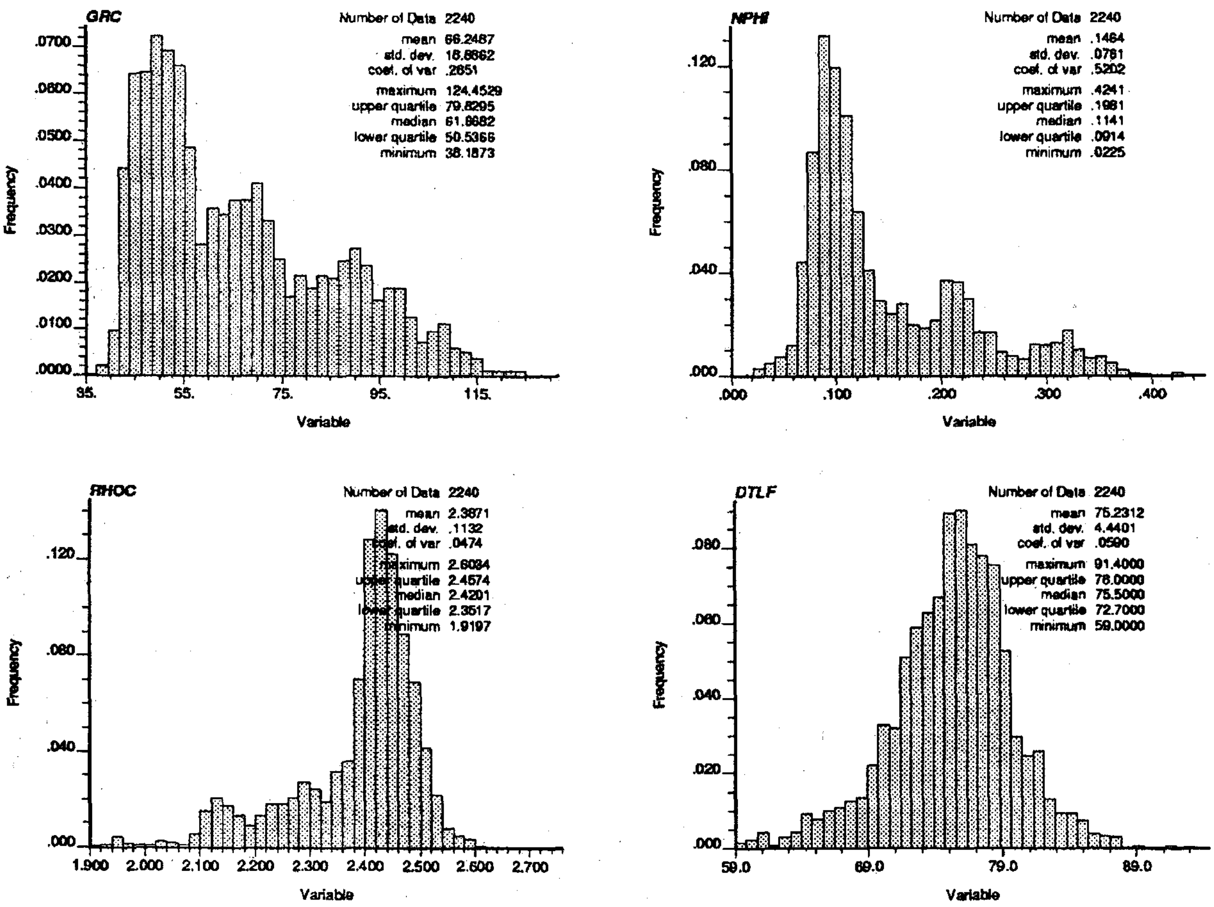


Fig. 5. Descriptive statistics for well logs of G-1.

r = number of zone
n_i = number of observation in the ith zone
$\overline{X}_{i\cdot}$ = mean of the parameter in the ith zone
N = total number of observation

In dividing the well log into physically meaningful occurring zones, the criterion used is to minimize *MSE* and maximize *MSTR*.

Cluster Analysis

This multivariate statistical technique is to group similar objects and to distinguish them from other dissimilar objects on the basis of their measured characteristics. The most common class of clustering methods used in geology is that of hierarchical analysis (Moline and Bahr, 1995: Lim *et al.*, 1997). The basic steps are shown in Fig. 3 as an illustration of the clustering of hypothetical zones based on their log responses. First, a database of attribute measurements is compiled on the basis of the collective treatment of the attributes. The clustering algorithm is applied to the similarity matrix as an iterative process. The pairs of objects with highest similarities are merged, the matrix is re-computed, and the procedure repeats. Ultimately all the objects will be linked together as a hierarchy, which is most commonly shown as a dendrogram. At this point, the objects are in one giant cluster. A certain decision must now be made concerning where to cut the tree diagram into branches that coincide with distinctive groupings. The choice may be based either on visual inspection, a mathematical criterion that appears to reveal a natural breaking point, or some measure that can be used to check potential clusters against some external standard.

Applications

For filed application, the techniques developed in this study is applied to G-1 well in the Korea Continental Shelf. The logs selected for the electrofacies determination were gamma-ray log (GR), sonic log (DT), density log (RHOB), and neutron log (NPHI). Fig. 4 shows the selected well logs of G-1 for the investigation. These logs were chosen based on the descriptive statistical results and the quality of well log data (Fig. 5). The principal components were computed from the correlation matrix of log measurements. The first two principal components collectively accounted for 80% of the total variability. To enhance the quality and efficiency of the interpreted

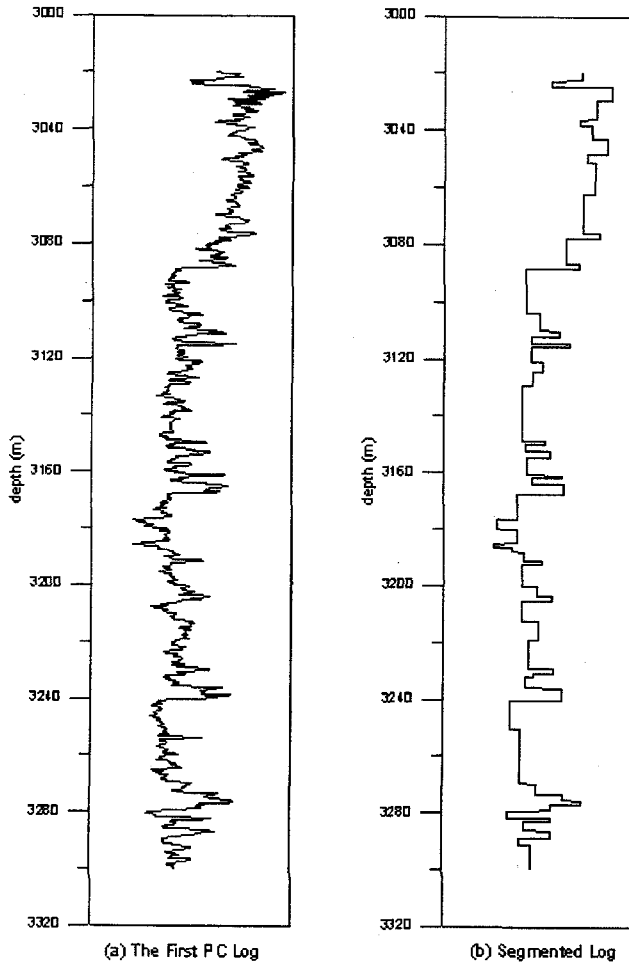


Fig. 6. Zonation of the first PC log by statistical method.

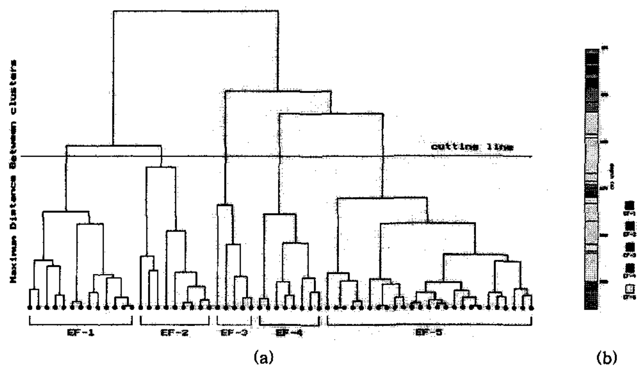


Fig. 7. Results of hierarchical clustering analysis.
(a) Dendrogram of log segments, (b) Electrofacies classification

results, the zonation of principal components logs were constructed by statistical method prior to clustering. The zone boundaries were determined such that the variance of the within zones was minimized and the variance of the between zones was maximized. The segmented first principal component log and the zone boundaries are shown in Fig. 6. Hierarchical clustering was used to group the segmented principal components logs instead of the

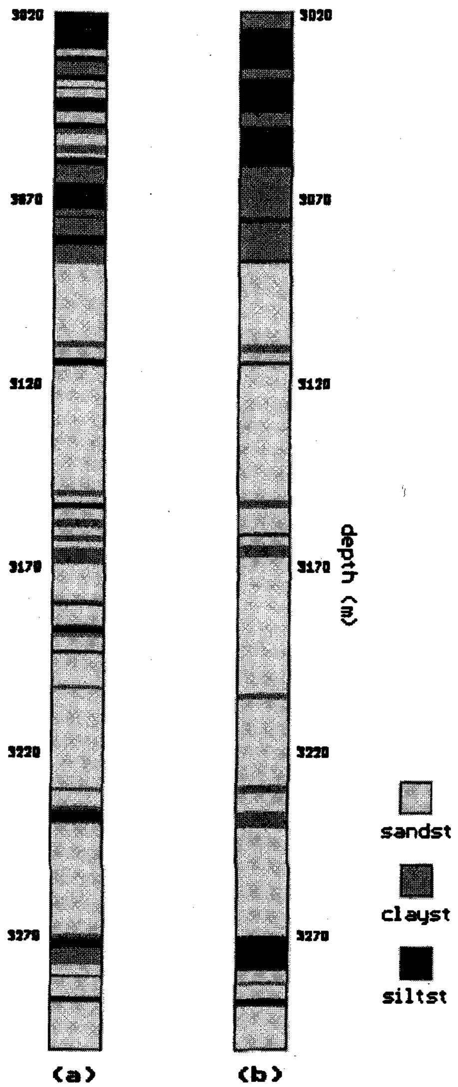


Fig. 8. Comparison of electrofacies from logs with mud-log description from cuttings of G-1 (a) mud log, (b) electrofacies

original logs with similarity based on their multivariate means. The resulting clustered hierarchy is shown in the dendrogram on Fig. 7(a). Each terminal branch represented an individual log segment. The five groups of electrofacies were determined on the basis of the ratio of the within-group variance to the total variance and core standard (Fig. 7(b)). Fig. 8 shows the match between a lithologic description from electrofacies and a mud-log description. The results of field application demonstrate that the lithology predicted based on the electrofacies classification matches well with the core and the cutting data.

Conclusions

A systematic technique has been developed for electrofacies determination from well logs using multivariate statistical analysis, and has been applied to the wells in the Korea Continental Shelf. The results of field application demonstrate that the lithology predicted based on the electrofacies classification matches well with the core and cutting data. This technique can enhance the quality and efficiency of the electrofacies classification. This electrofacies interpretation can be used to define reservoir characteristics that are valuable for reservoir management.

Acknowledgments

This research was funded by Korea Petroleum Development Corporation. The support and permission to publish this work are gratefully acknowledged.

References

1. Dharmawardhana, H. P. K. and Keller, G. V., 1985, Statistical Method for the Determination of Zone Boundaries Using Well Log Data: paper SPE 14470 presented at SPE Annual Conference and Exhibition, Las Vegas, NV, Sept. 22-25.
2. Doveton, J. H., 1994, Geologic Log Analysis Using Computer Methods : AAPG, Tulsa, OK.
3. Isaaks, E. H. and Srivastava, R. M., 1989, Applied Geostatistics : Oxford University Press, New York, NY.
4. Lim, J.-S., Kang, J. M. and Kim, J., 1997, Multivariate Statistical Analysis for Automatic Electrofacies Determination with Well Log Measurements: paper SPE 38028 presented at SPE Asia Pacific Oil & Gas Conference and Exhibition, Kuala Lumpur, Malaysia, April 14-16.
5. Moline, G. R. and Bahr, J. M., 1995, Estimating Spatial Distributions of Heterogeneous Subsurface Characteristics by Regionalized Classification of Electrofacies: Mathematical Geology, **27**(1), 3-22.
6. Serra, O. and Abbott, H. T., 1982, The Contribution of Logging Data to Sedimentology and Stratigraphy : SPE Journal (Feb.), 117-131.
7. Wolff, M. and Pelissier-Combescure, J., 1982, FACIOLOG-Automatic Electrofacies Determination: paper FF presented at SPWLA Annual Logging Symposium, Corpus Cristi, TX, July 6-9.