*JSUT **32**(2), 2006, p. 69-75*

Application of Multivariate Cluster Analysis in Logfacies Determination and Reservoir Zonation, Case Study of Marun Field, South of Iran

V. Tavakoli, A. Amini*
School of Geology, University College of Science, University of Tehran, PO Box: 14155-6455, Tehran, Iran
***Corresponding author, e-mail: ahamini@khayam.ut.ac.ir**
(*received: 31/1/2005 ; accepted: 29/5/2005*)

Abstract
Reservoir zonation is the most important job in oil industry which dominantly relies on the major properties of rocks. Fundamental properties of rocks are usually understood by their detailed description in the field (lithofacies analysis) and laboratory (petrofacies analysis). The facies (lithofacies and petrofacies) determination in most subsurface studies is impractical, due to lack of cores and cuttings. In such situations, where the wire line logs are the only data available, the logfacies or electrofacies are determined instead. Using multivariate cluster analysis, in this study practiced the logfacies determination on the Asmari Formation of Marun Field by programming a new algorithm with MATLAB software. The study is carried out on the logs of a well in which the Asmari Formation is fully cored. Determined logfacies are correlated with lithofacies defined on the cores and thin sections. Reliability of the procedure in logfacies determination is examined and an appropriate cut off level is defined for the cluster analysis. Result from this study is prepared to use in logfacies determination of the adjacent wells with no cores and cuttings.

Keywords: Logfacies, Multivariate cluster analysis, Reservoir zonation, Marun Oil Field, South Iran.

1. Introduction
Well logs are principal sources of subsurface geological information. They provide significant information on mineralogical composition, texture, sedimentary structures and petrophysical properties such as porosity and permeability. By compiling data from various well logs, one can discriminate sedimentary units with comparable log characteristics. The sedimentary units which defined on this basis and characterized from wire line logs are known as electrofacies or logfacies in the literatures (Serra 1986). The term logfacies is preferred here, because some non-electrical logs are also involved in its description. In this regard the logfacies is applied for a set of log responses which characterizes a bed/strata and allows it to be distinguished from others (Serra 1986). Logfacies analysis is the most important tool in petroleum industry, sedimentological and depositional environment study of the bearing rocks, especially where wire line logs are only reliable data available (Serra 1986).

Logfacies analysis can be carried out manually or automatically using mathematical techniques. In manual method, the logfacies is determined mainly based on the shape of log curves. This method is applied on the set of logs which are depth matched and illustrated beside each other. In automatic method, all log responses are involved. Several mathematical techniques have been introduced to automate the task of logfacies identification. All of the techniques are derived from same data grouping and dissociating of groups from statistical point of view.

Multivariate cluster analysis (as the best method of data grouping) is one of the most accurate and affective methods in oil bearing reservoir zonation. The method is applied on both detrital and carbonate rocks (Gill *et al*. 1993). This method gets more support by improvements in algorithms and statistics. Proper combination of logs and appropriate algorithm will increase the accuracy, reliability and effect of the method. With distinct combination of logs and cutoff level, types and amount of interested logfacies can be attained. In this method, any geological parameter described from other sources such as cores and thin sections can be related to wells with comparable facies. The accuracy and reliability of defined logfacies can be examined in wells from which suitable cores are available. Results from such a comparison provide a fundamental base for study of wells with poor core and cutting data.

This paper considers a new algorithm of this method based on advanced statistics techniques for logfacies determination of a well in Marun Field in which Asmari Formation is fully cored. The objectives of this study are to find a foundation for zonation of Asmari Formation in other wells for which there is no core data available and interpretation of depositional environment and correlation of wells throughout the field.

Clustering Procedure

Similar facies may have different log responses due to diverse factors that affect the logs. Since using statistical methods and procedures are mandatory, in clustering procedure data are grouped with minimum distance and maximum homogeneity. It is obvious that distinct geological parameters can be related to a group of data, as logfacies, which be used by geologists for further interpretation. For this calculation, all log readings are considered as "observations" and the used logs as the "values of the observations".

There are several ways to compute the distance between objects. The "*City Block metric*" distance is used here in from the MATLAB software, because more accurate results are obtained with this procedure. If we have "*i*" readings and relate "*j*" logs to them, the distance between x_r and x_s (d_{rs}) is calculated using following equation:

$$d_{rs} = \sum_{j=1}^{j} \left| x_{rj} - x_{sj} \right| \qquad (1)$$

In which:
x_r: rth reading
x_s: sth reading
d_{rs}: distance between x_r, x_s
j: number of logs
x_{rj}: jth log at x_r reading
x_{sj}: jth log at x_s reading.

The distance is computed between pairs of objects in *i*-by-*j* matrix X, which is treated as *i* vectors of size *j*. For a dataset made up of *i* objects, there are $(i-1)*i/2$ pairs. The output, *Y*, is a vector of length $(i-1)*i/2$, containing the distance information. The distances are arranged in the order of (1,2), (1,3), ..., (1,*i*), (2,3), ..., (2,*i*), ..., ..., (*i*-1,*i*). In other words, the value of *Y* as calculated distance is:

$$Y = \text{d}(1,1),\text{d}(1,2),\text{d}(1,3),\ldots,\text{d}(1,i),\text{d}(2,3),\text{d}(2,4),\ldots,\text{d}(i-1,i) \qquad (2)$$

In cluster analysis smallest distances are connected together to build a pair. Usually the number of logfacies are less than the number of readings, since then, pairs of vectors (pairs of log readings) are linked together to build a cluster (logfacies). The lower-rank clusters are linked together to build higher-rank types. This procedure continues until a single cluster (representing the whole data) is built.

There are various methods to link two clusters. In some of which the minimum distance of the cluster components are utilized to link them. In other words, "*p*" and "*q*" clusters link to each other if "i_1" reading from "*p*" cluster has the minimum distance with "i_2" reading from "*q*" cluster. Maximum distances, as other optional method, can be used in the same way. It is obvious that some information may be discarded in such procedures. In this paper the *average distance method* is used in which the average distance between all pairs of objects in two clusters is used. For example, the average distance between clusters "*p*" and "*q*" is achieved using following equation:

$$d(p,q) = d(\bar{x}_p, \bar{x}_q) \qquad (3)$$

In which:

$$\bar{x}_p = \frac{1}{n_p} \sum_{i=1}^{np} x_{pi} \qquad (4)$$

Here:
n_p: number of log readings in cluster *p*.
x_{pi}: *i* reading in cluster *p*.

The $\bar{x}_q$ value is calculated in the similar way.

The resultant distances are arranged in a new matrix (*Z*) which is (*i*-1) by 3 with distance information in the third column.

The *cophenetic correlation coefficient* is used to examine the reliability of applied method (Scott et al., 2004). The coefficient indicates how really the data fit into the structure suggested by the classification. Cophenetic correlation coefficient can be achieved using following equation (Everitte, 1993):

$$C = \frac{\sum_{i<j} (Y_{ij} - y)(Z_{ij} - z)}{\sqrt{\sum_{i<j} (Y_{ij} - y)^2 \sum_{i<j} (Z_{ij} - z)^2}} \qquad (5)$$

In which:
Y_{ij}: distance between two objects in *Y* matrix
Z_{ij}: distance between two linked groups in *Z* matrix
y and *z* are the average values of *Y* and *Z*, respectively.

For a high-quality solution, the magnitude of this value should be close to 1. This value can be used to compare alternative cluster solutions obtained from different algorithms. The cophenetic correlation coefficient for the present study (Marun Field) is determined 0.82, which reflects high quality respond of the applied method.

Due to numerous readings of the logs, analysis of clusters and logfacies achieved from resultant numbers is very complicated, if not impossible. To overcome the problem a "*dendrogram*" is normally used (Zhou *et al*. 2002, Samir *et al.* 2003, Scott *et al*. 2004). A dendrogram consists of numerous U-shaped lines connecting objects in a hierarchical tree. The Lines connect two clusters with minimum distance and the procedure continues until all the groups come together as a single cluster (Fig. 1). The distances between clusters are usually observed in horizontal (x) axis.

On the basis of desired accuracy, a cutoff level is defined and each cluster on the dendrogram represents a logfacies. The cutoff level can be determined at any distances of the dendrogram. Different distances will only change the number of logfacies determined. Smaller distances and consequent higher number of logfacies will lead to more detail description of the studied formation. In practice, disruption among clusters which reflect difference of various groups has greater influence on data analysis. On this basis, similar groups with minimum difference in distance are connected together as a cluster and disruption occurs among the dissimilar groups. It is important to notice that rocks have too many properties and in this method only those properties recorded on the logs are analyzed. Due to such a limitations and high sensitivity of the logs, slight differences in correlation of logfacies and those defined from the cores (lithofacies) are inevitable. Utilizing the logs which are mostly affected from lithology (e.g. GR, CGR, LDL) will significantly reduce such differences.

Application to Marun Oil Field

Multivariate cluster analysis is used with assist of softwares such as CONISS on both detrital and carbonates rocks in similar studies (Gill *et al*. 1993). In present study, a new algorithm based on a foresaid discussion is programmed with MATLAB software to demonstrate the ability of the method in logfacies determination. Data from a well with full core recovery of the Asmari Formation in Marun Oil Field is used in this study. This was for correlation of determined logfacies with those of cores and thin sections (Fig. 2).

The Asmari Formation is subdivided into layers and sub-layers in previous studies (Gholipour & Haggi 1989, Intera 1992). The classification by Intera Company researchers (Intera 1992) is used here because it has a lithological base and it almost covers other classifications. Lithological characteristics of the layers, sub-layers and their boundary surfaces are represented in Table 1. Four routine logs, GR, CGR, Sonic and LDL, that are available in nearly all wells of the field, have been used here. Lithology logs constructed from cores, graphic well logs, and thin sections, have been used for the method validation. The output logfacies are expected to correlate with that of Intera at a proper cutoff level. All logs are digitized at a vertical interval of 15 cm. To minimize the effects of scales and units of log variables, logs are standardized by subtracting of the mean from each reading and dividing the result in to the standard deviation (equation 6).

$$Z = \{V\text{-mean}\ (V)\}/\delta \qquad (6)$$

In which:
V: column of log readings.
δ: standard deviation.

According to the mentioned procedures, the data was analyzed. The used logs, facies classification and corresponding dendrogram are shown in Fig. 1. Moving from left to right in the resultant dendrogram, number of groups (logfacies) become less and less until one general logfacies remains. Horizontal and vertical axes represent the distance and depth, respectively. Since the number of logfacies is definite, this process is terminated in a distinctive distance which is termed "cutoff level" in the literature. For correlation with Intera classification the 0.8 distance is determined in the Asmari Formation.

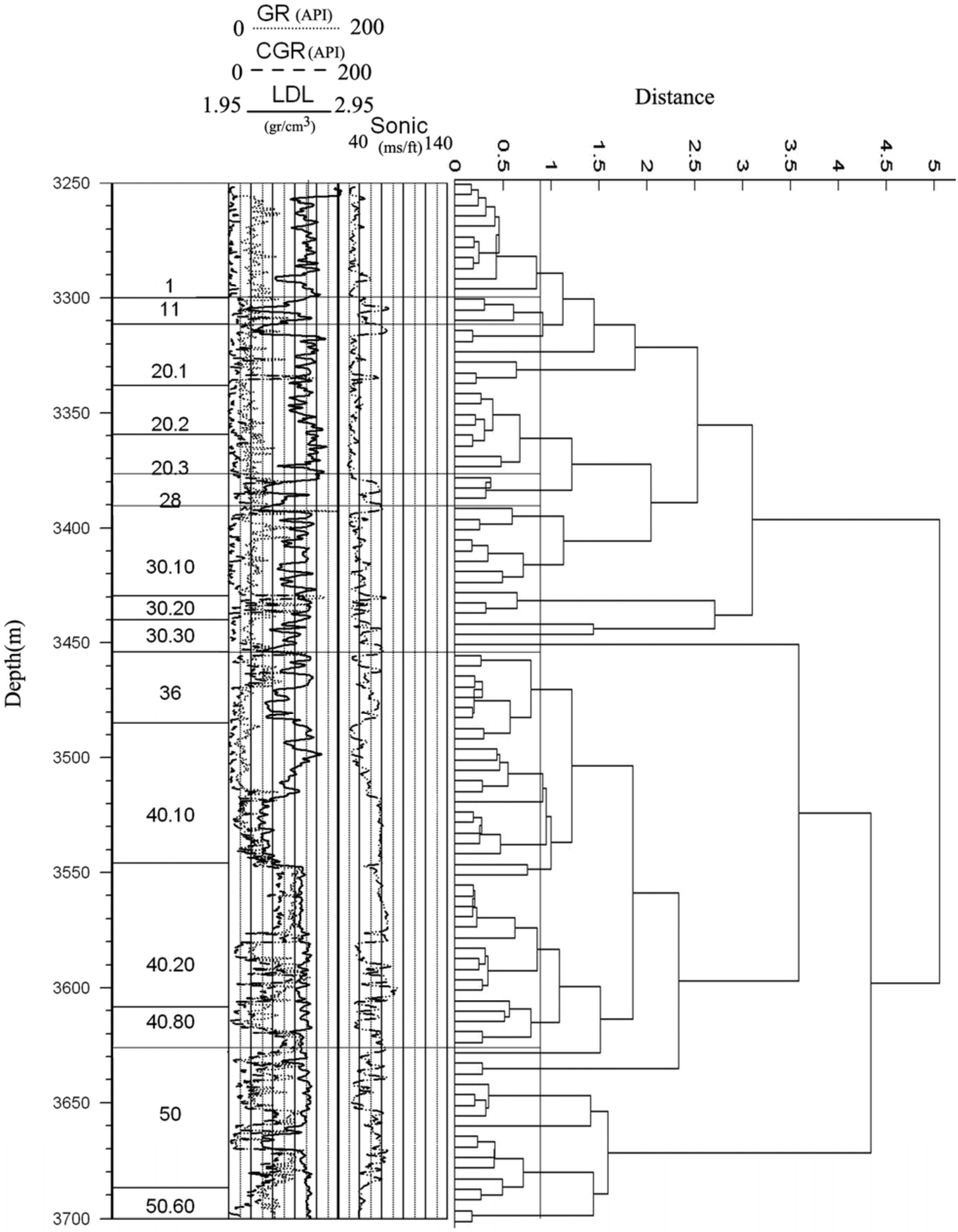

Fig. 1- Used logs, resultant dendrogram and logfacies correlation. Cutoff level is shown with a vertical line on the dendrogram.

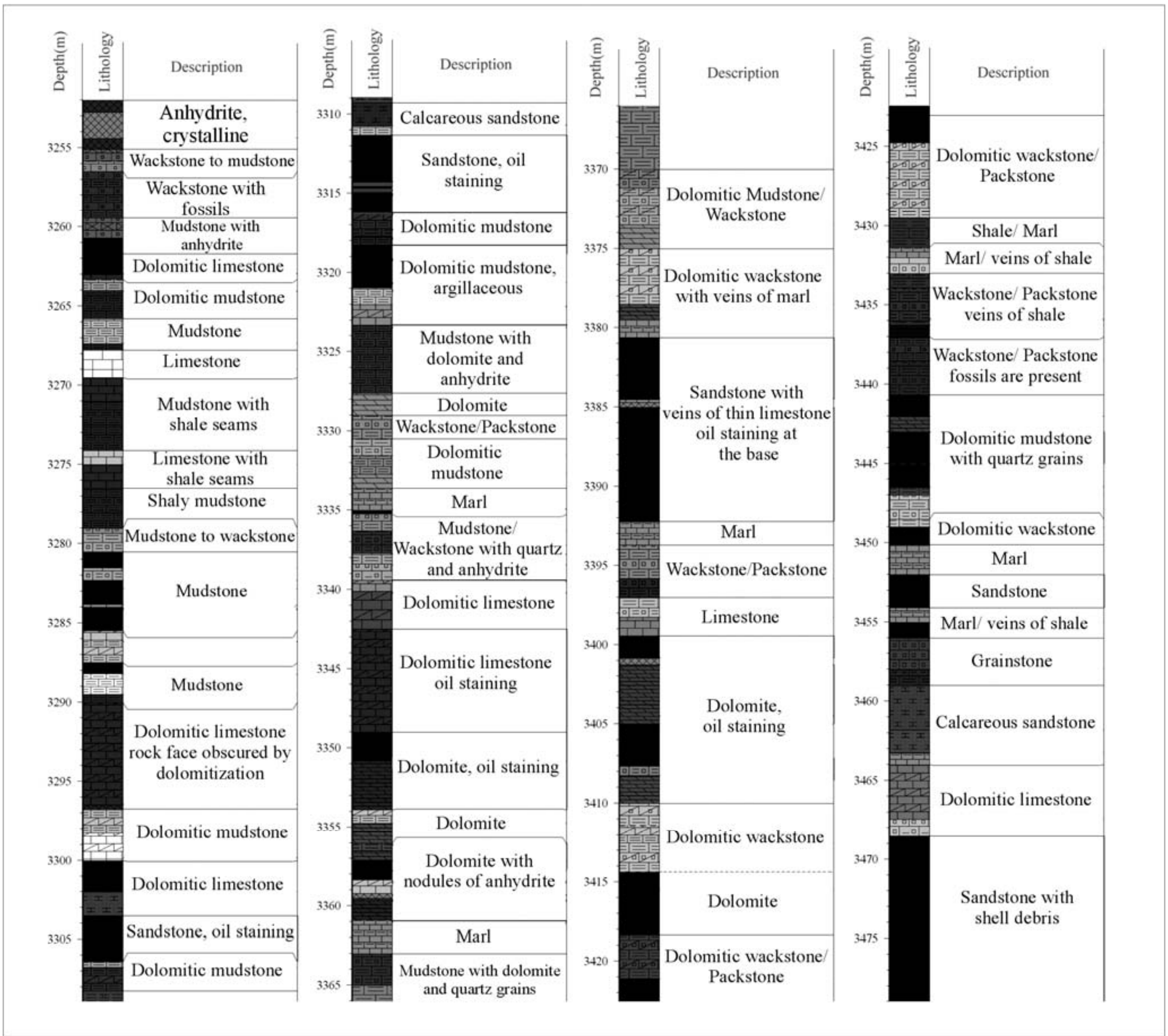

Fig. 2- Lithological log of the studied well prepared based on the data from cores and thin sections.

Tab. 1- Zonation of the Asmari Formation in the studied well, based on Intera (1992).

Zones	ELV (m)	description
1	3258	Mudstone, anhydrite in places. Dolomitization at the base
11	3300	Interbedd limestone and sandy limestone
20/1	3311.5	Dominantly mudstone to packstone, dolomitized one after the other. Some shale and sandy layers at the top
20/2	3338.2	Dolomite with thin layer of limestone at the top (rich in interaclasts)
20/3	3359.5	Mudstone to dolomitized wackstone with intercalations of marls. Anhydrite in places
28	3378	Sandstone with thin layers of marl
30/10	3391	Alternation of dolomite and dolomitic wackstone. Marl layers at the top of the zone
30/20	3429.5	Alternation of shale and wackstone
30/30	3440	Dolomite, thin layers of wackstone in the middle and sandstone at the base
36	3454	Sandstone, dominantly calcite cemented. Thin layers of shale and marl
40/10	3485	Boundstone, packstone to grainstone with detrital quartz. Detrital constituent increases downward
40/20	3546	Shale at the top and alternation of shale and limestone at the base
40/80	3608/5	Calcareous shale and marl, sandstone layer at the base
50	3628	Shale with intercalations of marl, limestone intercalations at the base
50/60	3687	Mudstone to marly packstone

In the Upper and Middle Asmari (upper than sub-zone 40.10) 11 logfacies in this cutoff level were recognized, 8 of which significantly correlated with those of Intera classification. There are 3 extra zones, which are normal in this type of study. This is due to the lithology diversity of each zone in the Intera classification and higher accuracy of this technique in comparison to those used by Intera (Intera 1992). By using lower cutoff level, all logfacies boundaries will correlate with those of lithological boundaries but greater numbers of logfacies are defined. This can be expected by detail inspection of lithology diversity in the Intera zonation.

Due to lack of detailed zonation in the lower Asmari (except sub-zones 40.80 and 50.60) and for reliability of the used techniques, the lower Asmari is analyzed with higher accuracy (Fig.3). The cutoff level is selected on 0.5, on which basis 14 logfacies were recognized in this part of the formation. These logfacies are called F1 to F12 and 40.80 and 50.60, the last two are preferred with their old used names (Table 2). In this part of the formation, all determined logfacies correlated with their lithological counterparts, which reflects the high accuracy of the used procedures.

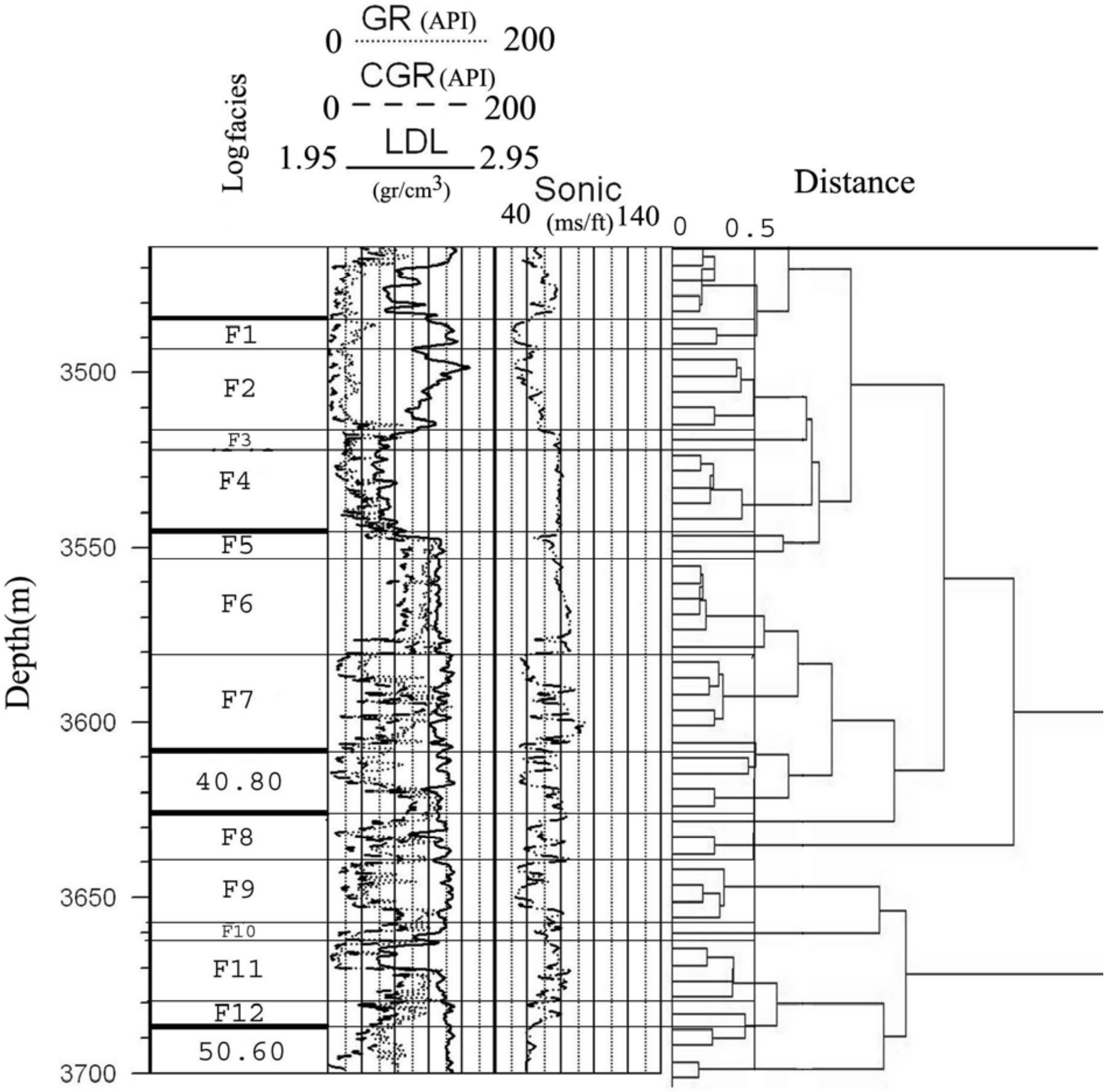

Fig. 3- Wire line logs, logfacies and resultant dendrogram in lower Asmari of the studied well.

Tab. 2- Major characteristics of resultant logfacies in the lower Asmari.

Logfacies	Interval	Description
F1	3485- 3492	Grainstone to packstone. High CGR and low Sonic
F2	3492- 3516.5	Wackstone to dolomitized packstone. High LDL and low sonic
F3	3516.5- 3522	Alternation of sandstone and shale. High GR
F4	3522- 3545	Grain supported sandstone with intercalations of shale. Fairly low GR, CGR and LDL
F5	3545- 3552	Shale. High GR, LDL and CGR
F6	3552- 3580.5	Shale with intercalations of marl. All logs are fairly high
F7	3580.5- 3608	Interbedded shale, grainstone and packstone. High LDL and Sonic
F8	3626- 3639	Alternation of shale, marl, calcite cemented sandstone and limestone. Fairly low sonic and alternation of low and high GR and CGR
F9	3639- 3657	Regular alternation of shale and calcareous shale. High LDL, low GR, CGR and Sonic
F10	3657- 3662.5	Calcareous marl. Low CGR and GR
F11	3662- 3680	Alternation of shale, marl and sandstone. High LDL, CGR and GR
F12	3680- 3686.5	Alternation of shale and limestone. High LDL and very low CGR

Conclusions

The reliability of multivariate cluster analysis in assist with a MATLAB based software for logfacies determination is documented in this study. As showed, a great amount of data can be obtained by an ordinary program. On the basis of different cutoff levels, desirable number of logfacies for any given formation can be achieved. Correlation of determined logfacies with those defined from cores and cuttings is essential to check the reliability of used methods and to define a meaningful cut off level for wells from which no cores or cuttings are available. Flexibility of this technique in using different combination of logs culminates to desirable logfacies definition. Performance of the established program with normal personal computers enables the users to obtain huge volume of data from limited logs. Absolute mathematical formulas decrease the possible blunders in to a minimum level. The logfacies defined in the studied well of the Marun Oil Filed can be confidently used for any geological investigation of the formation through out the field, such as in reservoir zonation and depositional environment analysis.

Acknowledgements

This work financially supported by National Iranian South Oil Company (NISOC) and University of Tehran. The data for Marun Oil Field are provided by NISOC, geology department. The assistance of Mr. Amiri-Bakhtiar, head of geology department (NISOC) and Mr. Haghparast, industrial consultant for the project, are highly acknowledged.

References

Everitt B.S. 1993: Cluster analysis, John Wiley and Sons, New York- Toronto, 3rd edition.

Gill D., Shomrony A., Fligelman H. 1993: Numerical zonation of log suites and logfacies recognition by multivariate clustering. *AAPG Bull.* **77(10):** 1781-1791.

Gholipour A.M., Haggi A. 1989: Geological study of layers 4 and 5 of the Asmari Formation. NISOC, report No 4135.

Intera ECL Petroleum Technologies Ltd. 1992: Marun Field study, Phase 2, Reservoir characterization, Geophysics. Intera Petroleum Production Division, Calgary, Canada.

Samir A.M., Abdou H.F., Zazou S.M., El-Menhawey W.H. 2003: Cluster analysis of recent benthic foraminifera from the northwestern Mediterranean coast of Egypt. Rev. Micropaleont. **46:** 111-130.

Scott G.H., King P.R., Crundwell M.P. 2004: Recognition and interpretation of depositional units in a late Neogene progradational shelf margin complex, Taranaki Basin, New Zealand: foraminiferal data compared with seismic facies and wireline logs. *Sediment. Geol.* **164:** 55-74.

Serra O. 1986: Fundamentals of well log interpretation. The interpretation of logging data, Elsevier, Amsterdam, Vol. 2.

Zhou W., Maerz N.H. 2002: Implementation of multivariate clustering methods for characterizing discontinuities data from scanlines and oriented boreholes. Comp. & Geosci. **28:** 827-839.