

## 16. LITHOLOGIC INTERPRETATION OF DOWNHOLE LOGGING DATA FROM THE CÔTE D'IVOIRE-GHANA TRANSFORM MARGIN: A STATISTICAL APPROACH<sup>1</sup>

Carlos A. Gonçalves<sup>2</sup>

### ABSTRACT

Downhole logging data determined during Ocean Drilling Program Leg 159 on the Côte d'Ivoire-Ghana Transform Margin correlate to a varying degree with lithological and structural features observed in core description. Statistical analysis, such as principal component analysis and non-hierarchical cluster analysis, has been employed to classify these data into lithologically similar groups to produce a statistically derived lithostratigraphy for the holes drilled.

The derived logstratigraphy is in good agreement with the lithostratigraphy defined from the study of core specimens conducted by the Shipboard Scientific Party, but also highlights some apparently anomalous sedimentary regions in the holes, which should be the focus for further examination. These regions, defined here as zones, are related either to lithological or petrophysical variations within the sediments.

### INTRODUCTION

Schlumberger downhole logs are used to characterize the geophysical, geochemical, and structural properties of a drilled sequence. Log data offer advantages over core-based analyses in that they are rapidly collected and represent continuous, in situ measurements of the formation. They can potentially integrate the ground-truth information provided by detailed core analyses with continuous and multivariate log data, resulting in continuous and quantitative records of lithologic sediment variability.

Standard Schlumberger tool strings for Ocean Drilling Program (ODP) operations were used during Leg 159. These include (1) a seismic stratigraphic tool string, (2) a lithoporosity tool string, (3) the Formation Microscanner (FMS), and (4) the Geochemical Logging Tool (GLT) (ODP, 1991; Masle, Lohmann, Clift, et al., 1996). A schematic diagram of these tool strings is shown in Figure 1 and the acronyms for all measurements are present in Table 1.

The GLT was run only in Hole 959D. Gonçalves and Ewert (Chap. 17, this volume) present the preliminary results for the geochemical logging data and their relationship with the sedimentary sequence. Sedimentary and structural analysis using the FMS, which also includes the General Purpose Inclination Tool (GPIT) and the Natural Gamma-Ray Spectrometry Tool (NGT), are given by Basile et al. (Chap. 9, this volume).

In this paper, I will present the principal logging results from ODP Leg 159 in light of the statistical analysis. In formulating the discussion, the primary log curves from the seismic-stratigraphic tool string, and the lithoporosity tool strings are used. Data were badly affected by borehole conditions such as washouts and bridges in two of the holes investigated, and in those cases, the data available for the analysis are reduced. The aggregate statistical results for each hole are compared and a model for the lithostratigraphy development is presented.

### LOCATION AND GEOLOGY OF THE SITES

During ODP Leg 159, four sites were drilled in the Côte d'Ivoire-Ghana Transform Margin (959, 960, 961, and 962; Fig. 2), which is located along the Marginal Ridge between 3° and 4° N and 2°30' and 3°30' W.

The primary objective for drilling the Côte d'Ivoire-Ghana Transform Margin during ODP Leg 159 was to assess the sedimentary and deformation processes along this continental border as a result of the different stages of continental break up and related transform tectonism (Masle, Lohmann, Clift, et al., 1996).

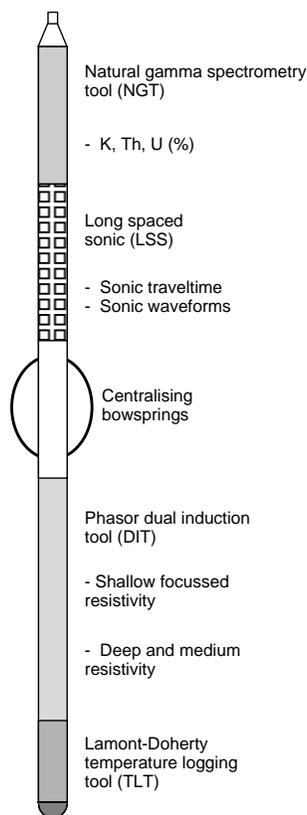
The tectonic deformation and sedimentation on the Côte d'Ivoire-Ghana Transform Margin is best represented in the spatial and stratigraphic variation of sediments cored at Sites 959 and 960. Other important information is provided by cores from Sites 961 and 962, but their contribution to the interpretation of paleoenvironmental settings is comparatively limited. The most striking feature common to all sites is the presence of hiatuses throughout the section. Using the sedimentary data, only short periods of time can be reconstructed on a regional scale. This is mostly because of difficulties in core recovery, removal by erosion or slumping of thick intervals of sediment, and periods of nondeposition. As a result, the statistical log interpretation was focused on the comparatively complete records provided by Sites 959, 960, and 962 to reconstruct lithostratigraphic conditions from these holes. There was no logging in Site 961.

On the basis of the mineralogy observed in core specimens, the sections at each site were subdivided into five main lithostratigraphic units by Shipboard Scientists (Fig. 3). Going downsection, lithologic Unit I, Holocene to lower Miocene, comprises dominantly calcareous sediment with two end-member sediment components: nannofossil ooze (grading downhole into nannofossil chalk) and foraminifer ooze (grading downhole into foraminifer chalk). Lithologic Unit II is comprised of alternating siliceous and calcareous sediment. It was divided into three lithologic subunits based upon the preservation of the siliceous microfossils. The upper lithologic Subunit IIA, comprises interbedded nannofossil chalk, diatomite, and clay, and ranges from early Miocene to early Oligocene in age. A black, middle-early Oligocene-age chert layer occurs below and was designated lithologic Subunit IIB. The sediment in Subunit IIC comprises micrite chalk

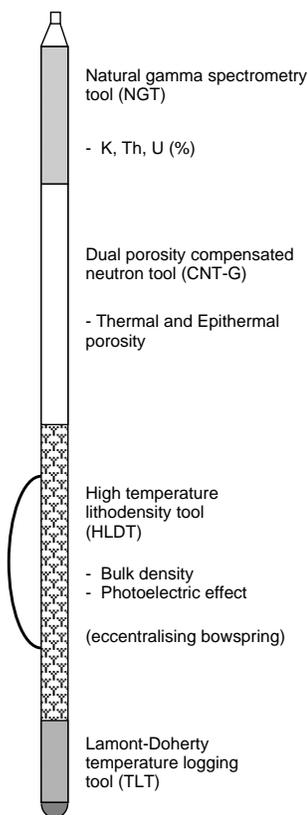
<sup>1</sup>Masle, J., Lohmann, G.P., and Moullade, M. (Eds.), 1998. *Proc. ODP. Sci. Results, 159*: College Station, TX (Ocean Drilling Program).

<sup>2</sup>Laboratório de Engenharia e Exploração de Petróleo (LENEP-UENF), Macaé/RJ, Brazil. carlos@lenep.uenf.br

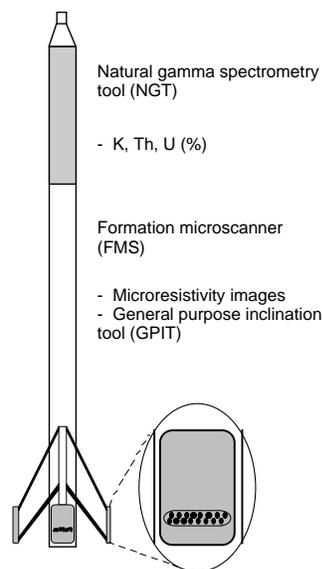
**1) Seismic stratigraphic combination**



**2) Lithoporosity combination**



**3) Formation Microscanner**



**4) Geochemical combination**

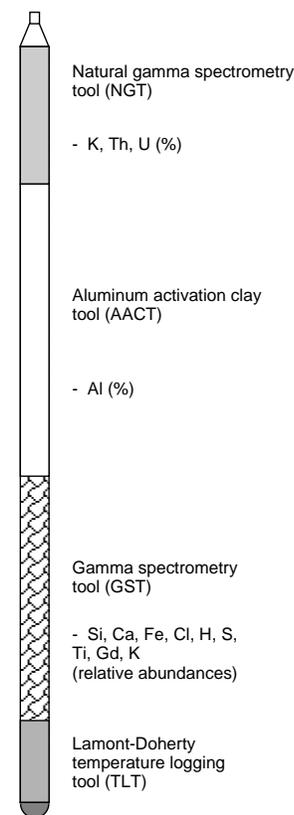


Figure 1. Sketch of the four logging tool combinations used in ODP Leg 159 operations. The seismic stratigraphic and the lithodensity combinations are combined in the standard Quad Combo (after ODP, 1991).

**Table 1. Acronyms for the well-log measurements obtained during Leg 159.**

Acronym	Measurement definition (and tool)
NPHI	Porosity from the CNT-G (Compensated Neutron Tool)
ILD	Deep resistivity from the DITE (Phasor Induction Tool)
ILM	Medium resistivity from the DITE (Phasor Induction Tool)
SFLU	Shallow resistivity from the SFL (Spherically Focused Log)
RHOB	Density from the HLDT (Lithodensity Tool)
PEF	Photoelectric absorption index from the HLDT (Lithodensity Tool)
SGR	Total gamma ray from the NGS (Spectral Gamma-Ray Tool)
THO	Thorium content in ppm from the NGS (Spectral Gamma-Ray Tool)
URA	Uranium content in ppm from the NGS (Spectral Gamma-Ray Tool)
POT	Potassium content in wt% from the NGS (Spectral Gamma-Ray Tool)
VP	Compressional wave velocity from the SDT (Sonic Digital Tool)

and porcellanite and ranges from Oligocene to late Paleocene in age (Masle, Lohmann, Clift, et al., 1996).

Lithologic Unit III is comprised of black claystone, ranges in age from late Paleocene to early Coniacian and has minor black claystone with nannofossils occurring at both the top and base of the unit.

Lithologic Unit IV is comprised of sandy limestone, sandy dolomite, calcareous sandstone, and limestone of early Coniacian to early Turonian ages. Finally, lithologic Unit V is comprised of quartz, sandstone, and silty claystone of late Albian age.

Well log data were obtained for different sections, depending on hole conditions in each of the sites logged. Therefore, not all the units shown in Figure 3 were covered by the logging measurements.

**STATISTICAL DATA ANALYSIS**

**Introduction**

Standard downhole log measurements were performed at intervals of 0.152 m throughout the different boreholes, providing a large data set. Because of the difficulties in handling, characterizing, and representing data of this type, a multivariate statistical approach was selected in this study. Initially, principal component analysis was used to evaluate the overall variability and to reduce the dimensionality of the problem. Subsequently, an iterative non-hierarchical cluster analysis was employed to classify the log responses and to generate a lithostratigraphy.

Summary statistics for the well log data obtained in each hole are presented in Table 2. Note that exceptionally high maximums are observed for total gamma-ray (SGR) and uranium content (URA) in Holes 960A and 960C. They are associated with apatite-rich layers and also correspond to abnormally low concentrations of thorium (THO) and potassium (POT) in Hole 960C. The minimum values (0.00 ppm and 0.00 wt%) actually represent negative values present within the data. This phenomenon is due to corrections during data acquisition and is discussed in Schlumberger (1981). It can also indicate poor borehole conditions. Maximum values of URA content (–10.23 ppm) at Hole 959D are possibly related to clay-rich layers (Masle, Lohmann, Clift, et al., 1996). Porosity (NPHI) values are very high and density (RHOB) values very low in Hole 960A. It is

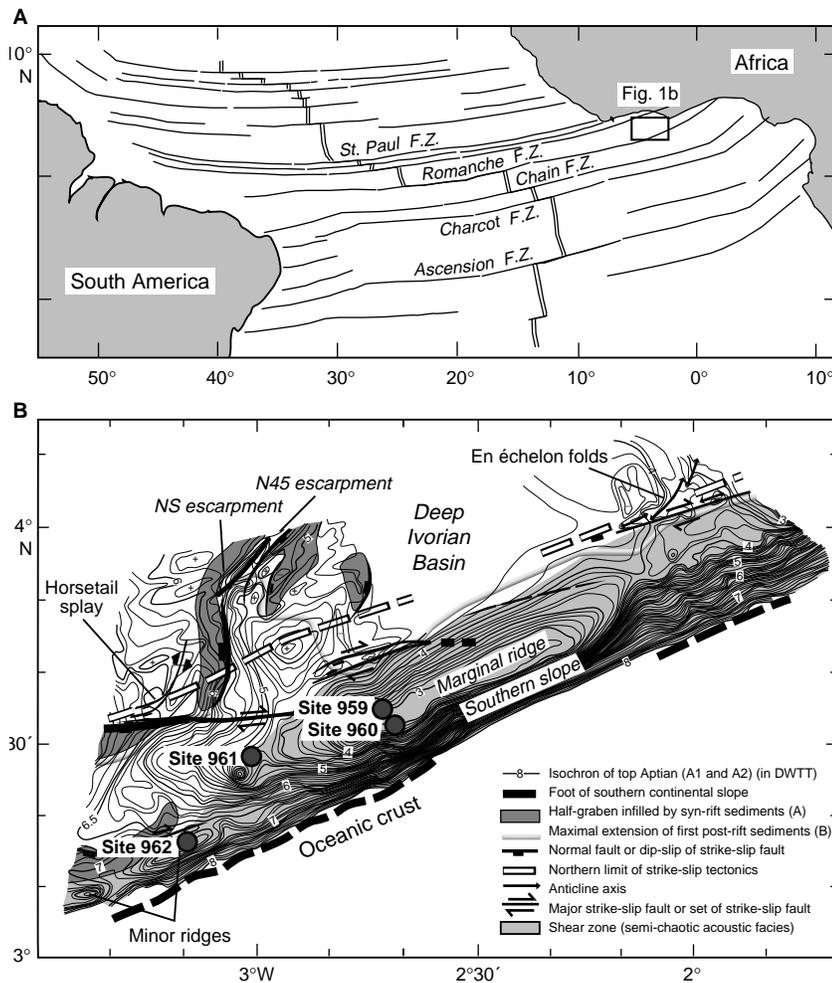


Figure 2. A. Sketch of the Equatorial Atlantic fracture zone and associated continental margins. The box locates figure below. B. Structural sketch of the Côte d'Ivoire-Ghana Marginal Ridge and surrounding area with location of the sites drilled (adapted from Mascle, Lohmann, Clift, et al., 1996).

observed that porosity, density, and velocity values were not obtained for Holes 960C and 962D because of bad hole conditions, which prevented the lithodensity tool from having good contact with the borehole wall.

General relationships within the datasets are summarized in Table 3. The linear correlation matrices for all holes are present. Significant correlations are observed between the resistivity measurements, density, and  $P$ -wave velocity in Hole 959D. Those correlations are mostly governed by NPHI, which shows negative correlations with these measurements. Another important positive correlation is between RHOB and PEF values and between SGR and THO, URA, and POT concentrations. Care should be taken when analyzing RHOB and PEF measurements. Because they are measured from the same source and stand off, poor borehole conditions can incorrectly increase their correlation factor. NPHI, POT, THO, and URA may be related to the presence of clays and their high natural gamma-ray emission, but SGR are dominated by the contributions of POT, THO, and URA within SGR.

Holes 960A and 960C show, in general, similar correlations. Porosity is again inversely correlated to resistivity and density, with PEF and density also showing strong correlation. An interesting aspect to be observed in Hole 960C is the correlation between total gamma-ray and uranium concentration (0.998). As mentioned before, this is because of an apatite-rich layer also observed in Hole 960A. The URA content accounts for almost all SGR concentration. Therefore, negative values in the concentrations of THO and POT show negative correlations with the total gamma-ray.

### Principal Component Analysis

Principal component analysis was conducted on all four holes using the linear correlation matrices shown in Table 3. This was equivalent to performing the analysis on standardized variables, with zero mean and unit standard deviation. This allows all log curves to be considered equally, regardless of their different measurement units. The result for the analysis in each hole is shown in Table 4. It presents the eigenvectors with their correspondent eigenvalues and the percentage contribution. It also shows the loadings for each of the components whose eigenvalues are greater than 1.0. Figure 4 is a graphic representation of the loadings for the different holes studied.

In Hole 959D, the first three components account for 82.80% of the total variability. All the variations (>85%) are included in the resistivity, total gamma-ray, and uranium concentration. Porosity (NPHI) and PEF are the least-constrained elements, with less than 70% of their variability being accounted for in the first three principal components. The first component is strongly correlated with resistivity, density, and  $P$ -wave velocity, and shows negative correlations with porosity. This can be interpreted as being a result of the variations in the physical properties of the rocks, such as water content and grain density. Physical property data measured on board show a continuous decrease downhole in water content (Mascle, Lohmann, Clift, et al., 1996). The second principal component is dominated by SGR, POT, and THO concentrations and was interpreted as reflecting the overall variation of rock composition. It is known from the lithostratigraphy that the interval logged in Hole 959D is dominated by

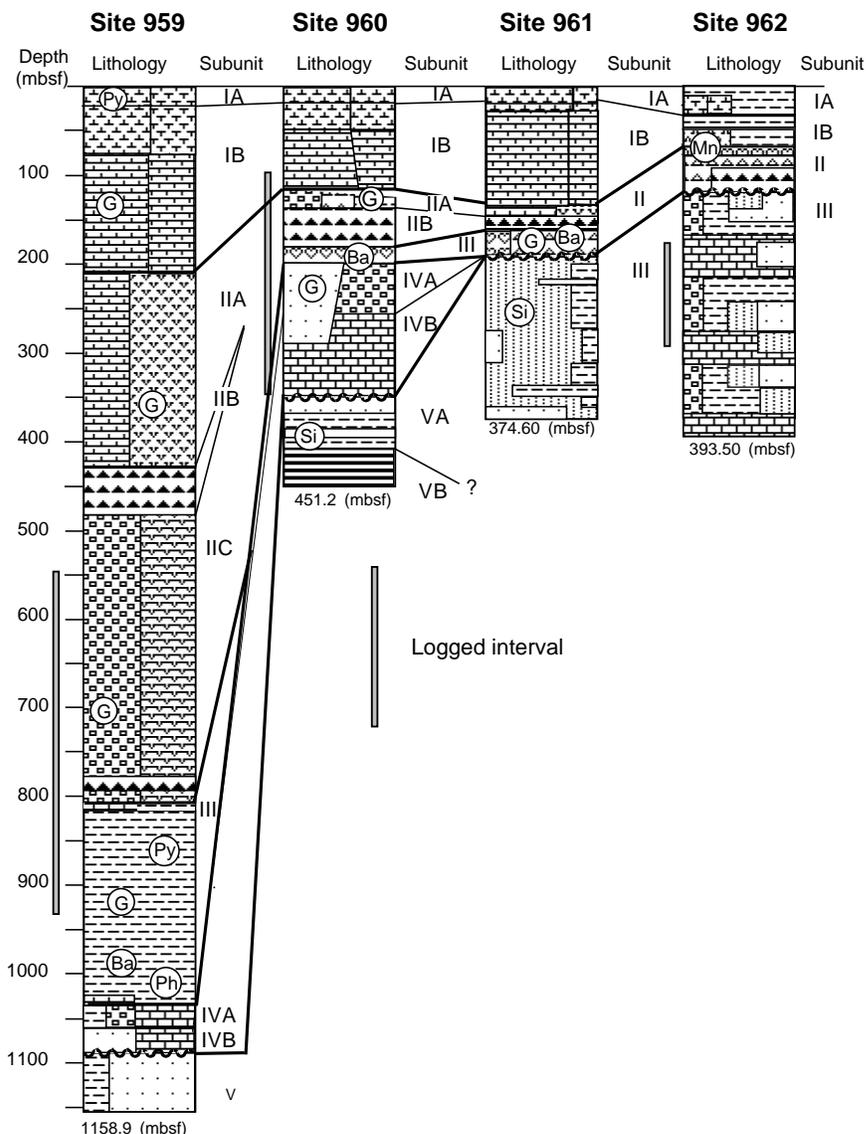


Figure 3. Lithostratigraphic columns for all four sites showing the variations in thickness of the different units between sites. Successive lithostratigraphic units are defined independently for each site as in the IR volume. Therefore, different names can appear for the same unit. See the “Explanatory Notes” chapter in Mascle, Lohmann, Clift, et al. (1996) for key to symbols.

porcellanite with micrite and clay above 770 meters below sea floor (mbsf) and by a black claystone with clay below this depth. The strong correlation of SGR and, to a weaker extent, THO and URA with the presence of clay, together with a negative behavior of resistivity, is suggestive of variations in clay composition. Geochemical well logging studies of this hole (Gonçalves and Ewert, Chap. 17, this volume) show that variations in  $K_2O$ ,  $Al_2O_3$ , and  $FeO$  concentrations may reflect different types of clay present. Particularly important are the peaks in  $K_2O$  and  $Al_2O_3$ , distinguishing clay-rich and carbonate-rich intervals. The third component is dominated by the naturally occurring radioactive element uranium, together with minor positive correlation of resistivity. This component represents 11% of the total variability, but no correlation with any particular lithofacies or petrophysical change was observed.

In Hole 960A, the first three components account for over 76% of the total system variability. Virtually all variations (>80%) in porosity and resistivity are included within the principal components. Note that in Hole 960A, the third component has its eigenvalue less than 1.0. However, it accounts for more than 13% of the system variability and was included in the analysis since it represents the upper sedimentary section of the hole where high porosity (70%–80%) are due to ooze and chalk sediments. Principal components 1 and 2 behave in

much the same way as in Hole 959D. With high positive correlations in resistivity and density and an inverse behavior in porosity, this component represents the variation in the physical properties of the sediments. Principal component 2 reflects a variation in lithology, with the positive correlation in density and negative in porosity and resistivity related to the presence of claystones and sand limestones below 175 mbsf.

Hole 960C presents only two principal components accounting for over 85% of the total variation. Principal component 1 corresponds to almost 50% of the variability and is strongly dominated by total gamma-ray and uranium concentration. This component reflects the high radioactive glauconite claystone with phosphate-bearing beds between 198 and 208 mbsf (Mascle, Lohmann, Clift, et al., 1996). Principal component 2 reflects the variation in resistivity between the clay-rich sediments at the top of the hole and the low porosity sandstone and sand limestones at the bottom. Note, in this hole there is negative correlation of potassium and thorium in both principal components. This is because of the abnormally high concentration of uranium that affects the correct computation of these elements from the downhole measurement tool (Ellis, 1987).

In Hole 962D, the first three principal components account for 83.43% of the total variability. The first component is dominated by

**Table 2. Summary statistics for Holes 959D, 960A, 960C, and 962D.**

Hole	Variable	Mean	Standard Deviation (σ)	Minimum	Maximum
959D*	NPHI	40.927	3.853	20.713	55.814
	ILD	0.951	0.158	0.632	1.792
	ILM	0.986	0.160	0.657	1.880
	SFLU	1.000	0.200	0.583	2.172
	RHOB	1.806	0.153	1.361	2.114
	PEF	2.239	0.388	1.444	3.520
	SGR	32.134	14.961	12.547	111.870
	THO	3.030	1.708	0.498	8.797
	URA	1.800	1.169	0.000	10.233
	POT	0.345	0.198	0.000	1.053
	VP	1.966	0.165	1.597	3.283
960A†	NPHI	53.562	8.264	27.050	82.253
	ILD4	4.003	5.651	0.493	47.281
	ILM	3.885	5.413	0.409	86.094
	SFLU	4.850	13.458	0.355	209.680
	RHOB	1.558	0.089	1.240	1.706
	PEF	2.425	0.502	1.341	3.484
	SGR	74.425	113.171	14.167	1174.100
960C**	ILD	3.696	2.774	0.581	20.184
	ILM	3.635	2.262	0.594	13.008
	SFLU	4.504	5.079	0.552	44.685
	SGR	48.777	112.398	3.082	869.930
	THO	0.723	1.045	0.000	4.766
	URA	5.031	13.267	0.175	103.120
	POT	0.104	0.264	0.000	1.127
962D††	ILD	3.487	0.465	2.640	5.030
	ILM	3.538	1.075	1.990	17.522
	SFLU	3.007	0.504	1.955	5.088
	SGR	52.087	8.457	32.932	75.642
	THO	3.347	1.032	0.318	6.060
	URA	2.722	0.851	0.402	5.739
	POT	0.968	0.293	0.315	1.655

Note: Abbreviations as in Table 1. Negative values are presented in the dataset. \*Number of log curves = 11, number of samples = 2552. † Number of log curves = 7, number of samples = 1413. \*\*Number of log curves = 7, number of samples = 1119. †† Number of log curves = 7, number of samples = 672.

resistivity together with a strong negative correlation to total gamma ray. This component is interpreted as variations in porosity and density, reflected by the presence of fluids in the pores and, hence, variations in the resistivity measurements. The second and third components are tentatively interpreted as lithologic variations along the interval logged. High POT and THO correlation in the second component and total gamma ray in the third component would distinguish claystone and clayey siltstone from calcite-cemented sandstone and quartz limestone.

Scores for the principal components in each hole were calculated from the component loadings and were normalized to yield standardized scores, with zero mean and unit standard deviation. These scores are used in further processing with the non-hierarchical cluster analysis.

### Non-Hierarchical Cluster Analysis

Non-hierarchical cluster analysis is employed here to characterize the well logging responses and to define any lithostratigraphic zonation that might be identified from the datasets. The analysis is performed using two different sets of data: first, the original log curves for each hole and second, the principal component scores discussed in the previous section.

Non-hierarchical cluster analysis has been available for some years, and its use in geological science is well known (Le Maitre, 1982; Davis, 1986; Shepperd et al., 1987). The method proved useful for interpreting different ODP well logging data (Brewer et al., 1990; Gonçalves, 1995). The objective is to split the data set into groups of samples that have similar characteristics.

The basic concept of this method is to select some initial partition of the samples and then modify the cluster membership to obtain an improved partition. This initial partition may either be one based upon groups or a set of seed points (or centroids) around which such groups may be formed. The clusters are separated by precise linear boundaries (Fig. 5). These are the loci of points equidistant from two given points in a straight line perpendicular to the line joining the two points. In a higher dimensional space, the boundaries become segments of a hyperplane.

In this work, the non-parametric technique is applied. It uses the nearest (Euclidean) centroid as the basis of classification rule. One of the major problems is the selection of the number of groups. This selection seems to be obvious in the two dimensional example of Figure 5; however, it is not so simple in a real multidimensional situation, where there is no prior knowledge of the data.

The approach used to solve the problem in this work involves running the analysis for a range of groups, and then choosing the *optimum* number of groups. The allocation of each point to a specific group is performed using the sum of the squared Euclidean distance from each data point to the centroid of the group to which it has been allocated. Logging data frequently represent gradational changes between lithofacies. The number of groups chosen is, therefore, not optimal, but represents convenient divisions within the data whose boundaries are gradational (Shepperd et al., 1987). On the other hand, subtle changes within the same lithofacies can lead to the classification of heterogeneities. More details of the technique in a well-logging context are given by Gonçalves (1995). In addition to group membership for each log response, the analysis yields the centroid composition for each group and the position of each group relative to the other groups. The centroid composition represents the mean of each group and a geological interpretation based on this can be used to make inferences about the nature of all points allocated to that group. Initial seed points in all cases studied were chosen using a method from Ball and Hall (1967), where samples are selected as seeds in such a way that every seed is at least at a given distance from other seeds.

Another important parameter yielded by the non-hierarchical cluster analysis is the Delta value. This value provides a measure of the separating power of an individual log curve for the various combinations of groups. They allow the importance of each log curve in group discrimination to be assessed. Two parameters are used to achieve this; the separation between the standardized group centroids and the standard deviations of the standardized data contained within each group. These were combined in such a way to give a value,  $\delta$ , which provides an indication of the relative importance of the individual log curves. A major difficulty is in deciding the relative weighting that should be given to the two parameters. Shepperd (1987) suggested that

$$\delta_i = [ABS(C_{ij} - C_{ik})]/S_p^2,$$

where  $\delta_i$  is the value of  $\delta$  for the  $i^{th}$  log curve,  $C_{ij}$  is the standardized value of the  $i^{th}$  log curve at the centroid of the  $j^{th}$  group,  $C_{ik}$  is the standardized value of  $i^{th}$  log curve at the centroid of the  $k^{th}$  group, and  $S_p^2$  is the pooled standard deviation of the  $j^{th}$  and  $k^{th}$  groups for the standardized values of that log curve, where

$$S_p^2 = (SP_j + SP_k)/(n_j + n_k - 2),$$

$SP_x$  is the sum of the squares and cross-products of the  $x^{th}$  group and  $n_x$  is the number of samples in the  $x^{th}$  group. The resulting values of  $\delta$  then describe the separating power of individual log curves for the various combinations of groups. The use of  $\delta$  in the assessment of the importance of individual log curves as well as in distinguishing group combinations is described in the analysis of each hole.

**Table 3. Linear correlation matrices for Holes 959D, 960A, 960C, and 962D.**

Principal Component Analysis (PCA)	NPHI	ILD	ILM	SFLU	RHOB	PEF	SGR	THO	URA	POT	VP
<b>Hole 959D</b>											
NPHI	1.0000	-0.5533	-0.5181	-0.4630	-0.2627	0.0304	0.4154	0.4473	0.2514	0.2999	-0.5357
ILD	—	1.0000	0.9776	0.9024	0.5597	0.2144	0.0063	-0.0058	-0.0408	0.1617	0.8011
ILM	—	—	1.0000	0.9102	0.5723	0.2351	0.0193	0.0031	-0.0279	0.1643	0.7803
SFLU	—	—	—	1.0000	0.6872	0.3364	0.0734	0.1267	-0.0535	0.2695	0.7883
RHOB	—	—	—	—	1.0000	0.6007	0.3423	0.4494	0.0824	0.4887	0.5629
PEF	—	—	—	—	—	1.0000	0.1779	0.3599	-0.0514	0.3254	0.1486
SGR	—	—	—	—	—	—	1.0000	0.7436	0.8155	0.6497	-0.0384
THO	—	—	—	—	—	—	—	1.0000	0.2380	0.8416	0.0295
URA	—	—	—	—	—	—	—	—	1.0000	0.1390	-0.1456
POT	—	—	—	—	—	—	—	—	—	1.0000	0.2236
VP	—	—	—	—	—	—	—	—	—	—	1.0000
<b>Hole 960A</b>											
NPHI	1.0000	-0.2458	-0.1900	-0.1950	-0.3116	-0.1239	0.1282				
ILD	—	1.0000	0.8031	0.6562	0.2326	0.3724	-0.1745				
ILM	—	—	1.0000	0.6433	0.2503	0.3867	-0.1705				
SFLU	—	—	—	1.0000	0.0981	0.2276	-0.1023				
RHOB	—	—	—	—	1.0000	0.6789	-0.2703				
PEF	—	—	—	—	—	1.0000	-0.3001				
SGR	—	—	—	—	—	—	1.0000				
<b>Hole 960C</b>											
ILD		1.0000	0.9189	0.8479			-0.2746	-0.1944	-0.2603	0.0857	
ILM		—	1.0000	0.7943			-0.3295	-0.2527	-0.3115	0.0961	
SFLU		—	—	1.0000			-0.2027	-0.1176	-0.1938	0.0871	
SGR		—	—	—			1.0000	-0.3892	0.9989	-0.7776	
THO		—	—	—			—	1.0000	-0.4295	0.6934	
URA		—	—	—			—	—	1.0000	-0.8037	
POT		—	—	—			—	—	—	1.0000	
<b>Hole 962D</b>											
ILD		1.0000	0.5221	0.7408			-0.4382	-0.0358	-0.2284	-0.3691	
ILM		—	1.0000	0.3828			-0.1896	-0.0149	-0.0318	-0.2605	
SFLU		—	—	1.0000			-0.3067	0.0369	-0.1956	-0.2478	
SGR		—	—	—			1.0000	0.4250	0.5404	0.5750	
THO		—	—	—			—	1.0000	-0.4005	0.6175	
URA		—	—	—			—	—	1.0000	-0.3102	
POT		—	—	—			—	—	—	1.0000	

Note: Abbreviations as in Table 1.

### Non-Hierarchical Cluster Analysis Using The Original Log Curves

The first analysis was performed using the original log curves in all four holes, following standardization to a zero mean and unit standard deviation. Initial testing for the number of groups resulted in four, four, five, and three groups being chosen for a full analysis in Holes 959D, 960A, 960C, and 962D, respectively. Table 5 contains the distances between the group centroids for the analysis in each hole.

In Hole 959D, Groups 1 and 2 are very closely related while Groups 3 and 4 present a progressive separation from the first two. Differences between groups are mainly given by *P*-wave velocity and resistivity measurements and, to a lesser extent, to total gamma-ray and thorium content, as shown by the delta values in Table 6. Group 2 is marked by low velocity and low resistivity in the interval between 550 and 740 mbsf (Fig. 6). Low values of total gamma ray around 30 API are also characteristic of this group. It represents the porcellanite (Unit IIC) with clay at the top and micrite at the bottom of the interval. A sharp increase in velocity and increases in resistivity and density are associated with Group 4 between 740 and 775 mbsf. This interval corresponds to a strong seismic reflector and was identified as an overall increase in carbonate content within Unit IIC changing from porcellanite with clay to porcellanite with micrite. Group 3 is characterized between 775 and 845 mbsf. A slight decrease in velocity and resistivity and a continuous increase in carbonate content are present. This interval represents the porcellanite nanofossil chalk. The gradational increase in Ca content continues downhole across the sedimentary boundary of Units IIC and III (812 mbsf) defined on board by the Scientific Party. From 845 mbsf there is a sharp increase in total gamma-ray and thorium responses because of an increase in content of clays. Group 1 is correlated to the black claystones and black siltstones (Unit III) between 845 and 930 mbsf.

Thin layers of Groups 2 and 3 are sparsely observed and correspond to local increases in Ca content within Group 1.

Hole 960A presents Groups 3 and 4 closely related with Groups 1 and 2, but showing a large separation from the latter (Table 5). Table 6 shows the delta values for all log curves used in the analysis. It is observed that total gamma-ray, and to a lesser extent density, porosity, and resistivity, are the log curves responsible for the main differences. Four groups were defined in this hole. Group 1 is characterized by very high total gamma-ray responses and low density and resistivity values (Fig. 7). Total gamma-ray reaches 250 API in certain intervals, reflecting the massive presence of clays. This interval is correlated with the black claystones and the palygorskite claystones between 105 and 200 mbsf with sporadic micrite layers interbedded. The occurrence of a 5 m layer of Group 3 at 155–160 mbsf is defined as an occurrence of barite nodules within the claystone. Groups 2 and 4 appear as very thin layers at 195 mbsf and 205 mbsf and between 270 and 278 mbsf, respectively. These groups do not correspond to any particular sedimentary or petrophysical variation in the data set. High peaks in resistivity and PEF log curves are tentatively correlated to these groups, but very tight holes at those depths can also be the source of this classification. Finally, Group 3 is characterized as low porosity, high density, high resistivity (between 10 and 20  $\Omega$ m) and low gamma-ray radiation. It is strongly correlated with quartz-sand limestone, packstones, and silty sandstone below 200 mbsf. Variations in mineralogy leading to variations in physical properties such as porosity and density are observed in this interval, but do not reflect variations in group classification.

Hole 960C was drilled just a few meters from Hole 960A; therefore, it was expected to give similar log responses. Despite bad hole conditions, which prevented porosity and density data from being obtained, five groups were obtained from the non-hierarchical cluster analysis using only resistivity and total gamma-ray logs. Group 1 is dominated by relatively high (100–200 API) total gamma-ray, very

**Table 4. Eigenvalues and their percentage contributions and principal component loadings for Holes 959D, 960A, 960C, and 962D.**

e-vector	eigenvalues	%	cum.%	p. c. loadings matrix for three components				
				component	1	2	3	%Acc
<b>Hole 959D</b>								
1	4.614	41.95	41.95	NPHI	-0.48	0.66	-0.12	69.20
2	3.246	29.51	71.46	ILD	0.91	-0.24	0.18	92.58
3	1.247	11.34	82.80	ILM	0.91	-0.22	0.17	91.32
4	0.772	7.02	89.83	SFLU	0.94	-0.11	0.04	90.31
5	0.449	4.08	93.91	RHOB	0.80	0.29	-0.18	77.43
6	0.237	2.16	96.07	PEF	0.43	0.31	-0.55	59.34
7	0.199	1.81	97.88	SGR	0.18	0.90	0.38	99.11
8	0.125	1.13	99.02	THO	0.25	0.87	-0.22	87.08
9	0.08	0.79	99.81	URA	-0.01	0.56	0.75	89.60
10	0.02	0.18	100.00	POT	0.40	0.74	-0.22	77.49
11	0.00	0.00	100.00	VP	0.84	-0.23	0.06	77.52
<b>960A</b>								
1	3.008	42.97	42.97	NPHI	-0.40	-0.16	0.88	97.31
2	1.424	20.35	63.32	ILD	0.84	-0.36	0.03	84.24
3	0.914	13.06	76.38	ILM	0.83	-0.35	0.10	83.61
4	0.811	11.59	87.98	SFLU	0.70	-0.50	0.00	75.07
5	0.378	5.40	93.38	RHOB	0.57	0.67	-0.04	78.87
6	0.270	3.85	97.24	PEF	0.67	0.50	0.27	78.88
7	0.193	2.75	100.00	SGR	-0.38	-0.42	-0.20	36.72
<b>960C</b>								
1	3.421	48.88	48.88	ILD	—	-0.65	0.70	92.85
2	2.545	36.36	85.25	ILM	—	-0.66	0.69	92.45
3	0.636	9.08	94.33	SFLU	—	-0.59	0.67	80.94
4	0.178	2.54	96.88	SGR	—	0.86	0.35	88.34
5	0.148	2.11	99.00	THO	—	-0.35	-0.70	62.99
6	0.069	0.99	100.00	URA	—	0.87	0.38	90.81
7	0.000	0.000	100.00	POT	—	-0.74	-0.57	88.36
<b>962D</b>								
1	2.790	39.86	39.86	ILD	0.82	0.29	0.25	83.78
2	1.848	26.40	66.26	ILM	0.58	0.15	0.49	60.52
3	1.202	17.17	83.43	SFLU	0.71	0.35	0.30	72.24
4	0.621	8.87	92.31	SGR	-0.76	0.02	0.63	98.89
5	0.315	4.50	96.82	THO	-0.37	0.81	0.19	84.90
6	0.222	3.17	100.00	URA	-0.22	-0.75	0.60	98.41
7	0.000	0.00	100.00	POT	-0.69	0.61	0.02	85.32

Note: — = no data. Abbreviations as in Table 1. p.c. = principal component.

high uranium content (up to 15 ppm), and low resistivity (Fig. 8). It is related to the palygorskite claystone (Unit III) between 175 and 200 mbsf. Different layers define the interval between 200 and 211 mbsf with the presence of Groups 2, 3, and 4. Group 3 is dominated by abnormally high (up to 800 API) gamma-ray emission and very high uranium content (<40 ppm). Glauconite claystone with phosphate nodules and fish debris (hardground) are the source of this high radioactive layer. It extends for 8 meters, and is shouldered by thin layers of Group 2, which was tentatively correlated with an increase in potassium above and below the glauconite claystone. Table 5 shows that Group 2 is highly associated with Group 3 while showing high separation with other groups. Between 211 and 218 mbsf, a high resistivity and low gamma-ray unit is defined as Group 5. As shown in Table 5, Group 5 is very distinct from the others, and is also present as thin layers down the hole at 290, 300, and 330 mbsf. It is probably correlated to the abundant quartz and feldspar (up to 40%) uppermost part of Unit IVB. The variation in siliciclastic material is gradual downhole with a decrease in resistivity and continuous low gamma-ray profiles. This characterizes Group 4, which correlates with Unit IVB, defined by quartz-sand limestone, grainstone, and silty sandstones. Note that Group 4 in Hole 960C correlates with Group 3 in Hole 960A.

In Hole 962D, all three groups defined by non-hierarchical analysis appear closely related (Table 5). The interval between 190 and 290 mbsf is characterized by minor changes between clay-rich sandstone, calcite-cemented sandstone, and quartz-sand limestone (Unit III), leading to minor changes in resistivity and gamma-ray values. Group 1 is dominated by high resistivity (5 Ωm) and low gamma-ray (40 API) values below 267 mbsf (Fig. 9). It is related to calcite-cemented sandstone lithofacies. Group 3 shows a slight decrease in resistivity and a substantial increase (up to 65 API) in gamma-ray and thorium contents. It is possibly correlated to clay-rich sandstone at different thickness intervals between 200 and 255 mbsf. Changes in

borehole diameter at those intervals can also suggest presence of clays, which are more resistant to borehole washouts. Low resistivity and low gamma-ray values are shown by Group 2 and are possibly related to quartz-sand limestone.

One may directly compare the derived zones (groups) with the lithostratigraphy established by the shipboard scientists. A lithostratigraphic units log obtained from cores is presented in Figs. 6–9. From the examination of the groups, different log-stratigraphic zones have been identified on the basis of the dominance of individual groups or group combinations.

In Hole 959D, Zone 1 is characterized by Group 2 response at the top of the interval and correlated with Subunit IIC. Zone 2 is dominated by ~65% of Group 3 and ~35% of Group 4 and would define a different lithostratigraphic unit between 740 and 840 mbsf. Table 5 shows that there is a reasonably strong correlation between these two groups. Zone 3 is given by Group 1 solely and is correlated with the black claystone and siltstone of Unit III. Hole 960A also presents three different zones. Zone 1, at the top of the interval, is mostly dominated by Group 1 with a thin layer of Group 3. This zone would include four lithostratigraphic units (IIA, IIB, III, and IVA) defined by the Shipboard Scientific Party. Zone 2 is a thin (~5 m) layer defined by Group 2 with very high gamma-ray content and not individualized in the lithostratigraphy section. Subunit IVB from the lithostratigraphic column appears characterized as Zone 3, including mostly Group 3 and a small layer of Group 4. Hole 960C provides better correlation between the zones from group classification and the lithostratigraphic section. It also shows, as expected, strong correlation with zones obtained from Hole 960A. Zone 1 is defined by Group 1 between 175 and 200 mbsf and corresponds to Unit III. Zone 2, as defined in Hole 960A, is characterized by a thin layer (~6 m) of Group 3 with high gamma-ray and uranium measurements. Zone 3 is defined in its majority by Group 4 with a few small layers of Group 5 and is correlated with Subunit IVB. The last hole, 962D, shows only

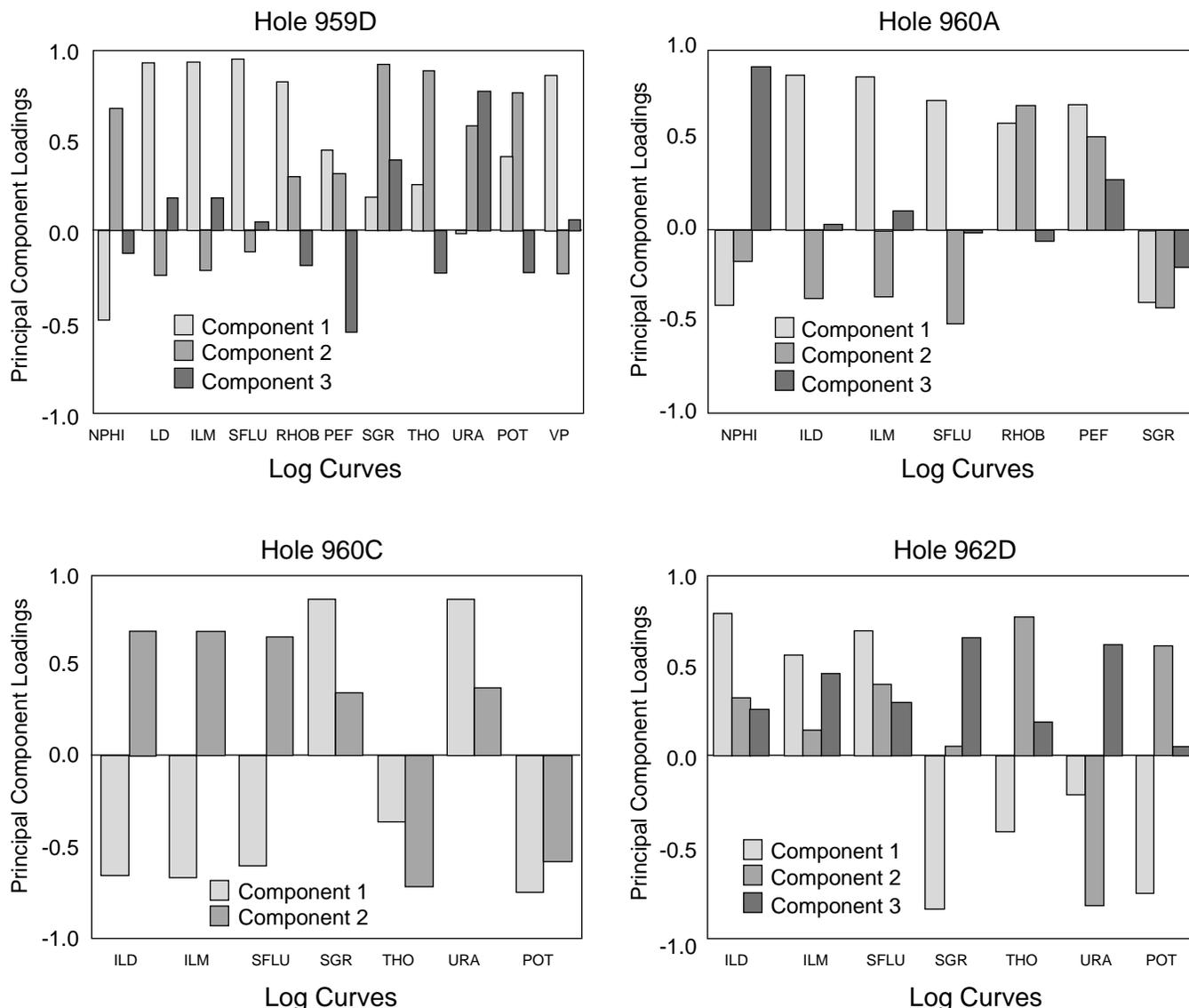


Figure 4. Principal component loadings for Holes 959D, 960A, 960C, and 962D.

two different zones obtained from the groups. Zone 1 would include an interbedded sequence of Groups 2 and 3 from 190 to 265 mbsf, while Zone 2 is defined by Group 1 at the bottom of the interval. Both zones were characterized as one single unit (Unit III) in the lithostratigraphy.

### Non-Hierarchical Cluster Analysis Using The Principal Component Scores

This analysis was performed using the component scores extracted from the principal component analysis. Statistically, this model is supposed to perform better than the one before because the principal component transformation effectively removes the intercorrelation from the variables. Therefore, it is expected that clustering with these data would produce additional information. Initial testing for the number of groups resulted in four, five, five, and four groups being chosen for a full analysis in Holes 959D, 960A, 960C, and 962D, respectively. Table 7 contains the distances between the group centroids for the analysis in each hole.

The first aspect observed in this model is that all groups in all holes are more closely related with each other than in the previous model using the original log curves. Another important aspect is that for some of the holes (960A and 962D) different numbers of groups were generated. As Component 1 contains the greatest variability of the system, and this is the case for the analysis in all holes (Table 4), it has more influence in the separation between the groups. To a lesser extent—and depending on the delta values (Table 8)—components 2 and 3 also influence the separation.

Figures 10–13 show the group logs extracted from the analysis with the principal components. Again, the units from the lithostratigraphic classification and the logstratigraphic zones are shown as well as the first two principal components. Overall, the zones derived from the component scores are similar to those obtained from the original log curves. In Hole 959D, three zones were defined. They correspond to the same zones observed in Figure 6. However, it was noticed that Groups 3 and 4 are inverted here, and that Group 3 now extends to 890 mbsf. This new zonation fails to pick out the top of the black claystone (Unit III) boundary at 840 mbsf. Consequently, Zone

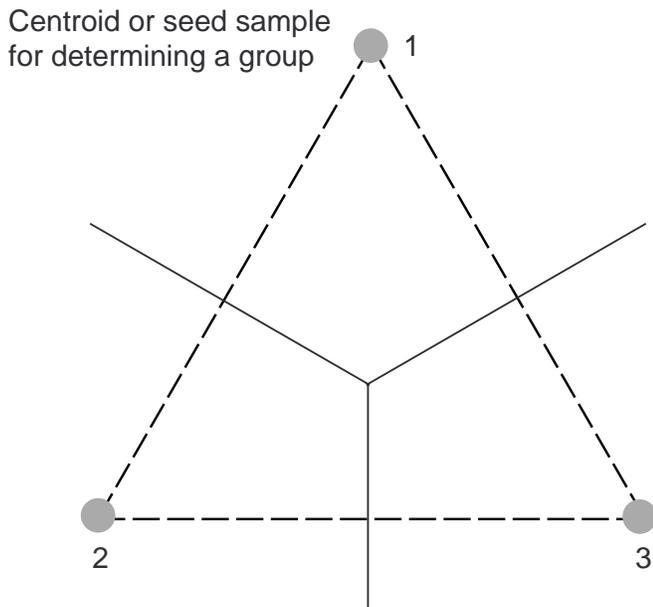


Figure 5. Piece-wise linear boundary separation in non-hierarchical cluster analysis.

Table 5. Distance between group centroids (non-hierarchical cluster analysis for original log curves) for Holes 959D, 960A, 960C, 962D.

Group	1	2	3	4	5
Hole 959D					
1	0.0000	0.1003	0.9276	7.2380	
2	—	0.0000	0.4230	5.7294	
3	—	—	0.0000	3.1106	
4	—	—	—	0.0000	
Hole 960A					
1	0.0000	83.1825	194.4535	243.8435	
2	—	0.0000	26.4464	54.3700	
3	—	—	0.0000	6.7853	
4	—	—	—	0.0000	
Hole 960C					
1	0.0000	104.7304	126.3888	7.6383	36936.5352
2	—	0.0000	1.0305	167.7876	40974.5039
3	—	—	0.0000	194.7777	41383.3594
4	—	—	—	0.0000	35902.7461
5	—	—	—	—	0.0000
Hole 962D					
1	0.0000	0.1003	0.9276		
2	—	0.0000	0.4230		
3	—	—	0.0000		

3 (Group 1) was reduced to just 40 m at the bottom of hole. Hole 960A now shows five different zones. Zone 1 extends from 105 mbsf to 165 mbsf, correlating with both Subunits IIA and IIB from the lithostratigraphic column and consisting of Group 1 with a few intervals of Group 4. Negative scores of components 1 and 2 are observed in this zone. Zone 2 is defined between 165 and 180 mbsf and was associated with the low values (relative to Zone 1) of gamma-ray. It is correlated to Unit III and Subunit IVA. Zone 3 is defined by the extremely high gamma-ray values and sharp decrease in both component scores between 190 and 200 mbsf. It corresponds to Zone 2 in the analysis with the original log curves. From 200 mbsf downhole, two zones are defined. Zone 4 (200–265 mbsf) is given by another increase in the scores for both principal components, and Zone 5 highlights the high-resistivity, calcite-cemented sandstone at the bottom of Subunit IVB.

Table 6. Delta values: analysis of the importance of the log curves in separate groups (analysis using the original log curves).

Log curves	Hole 959D	Hole 960A	Hole 960C	Hole 962D
URA	1.11	10.81	0.68	
PEF	1.16	2.10		
POT	1.47	5.13	1.85	
NPHI	1.48	1.05		
RHOB	1.88	1.44		
SGR	2.11	6.24	10.79	1.41
THO	2.20	4.16	1.43	
ILM	2.27	2.93	3.13	0.91
ILD	2.36	4.56	2.67	1.98
SFLU	2.44	2.14	2.21	1.60
VP	2.80			

Note: Abbreviations as in Table 1.

Hole 960C also shows a more detailed zonation compared to the previous analysis. Zones 1 and 2 appear at the same depths as before, correlating with Unit III and Subunit IVA. However, Subunit IVB is now subdivided into three different zones: Zone 3 (between 210 and 230 mbsf) including Groups 4 and 5, Zone 4 (230–270 mbsf) including Groups 1 and 4, and Zone 5 (between 270 and 345 mbsf) predominantly dominated by Group 4 with sparse layers of Group 5. These last two zones are related to the silty sandstone and calcite-cemented sandstone, respectively. Note the similarity between Zones 3 and 5 in Hole 960C. In Hole 962D the generation of four groups also allowed for a more detailed zonation of the sequence. Zone 1 corresponds to the top 7 m of the hole and is defined by Group 1 only. A separation between the first two components is observed. Zone 2 consists of interbedded layers of Groups 2 and 3 and extends from 197 to 255 mbsf. Zone 3 is very similar to Zone 1, consisting only of Group 1 between 255 and 267 mbsf and possibly related to clay-rich sandstones in Unit III. Finally, Zone 4 is characterized by Group 4 with thin layers of Group 1 at the top of the interval between 267 and 290 mbsf. The variety in lithology described within Unit III (claystone with quartz, siltstone with micrite, calcite-cemented sandstone, quartz sandy siltstone, and limestone; Masclé, Lohmann, Clift, et al., 1996) allows the diversity in the log-stratigraphic zonation.

### CONCLUSION

The zones derived from non-hierarchical cluster analysis, using both the original log curves and the principal component scores, are in good agreement with the lithostratigraphic sequence defined by the Shipboard Scientific Party. Distinction between the models is evident because the analysis using the principal components removes the intercorrelation from the variables, allowing clustering to produce a more detailed response with additional information and in some cases a better interpretation of the zones.

Log curves respond to different lithologic types and different petrophysical characteristics in the formation. When geologic units are defined with different lithologies, it is reasonable to observe a different characterization from the logs for that unit. This is observed in Holes 960A, 960C, and particularly 962D, where a wide range of different lithofacies were defined as Unit III. In part, this study set out to assess the use of techniques for characterizing and summarizing downhole logging data. The results presented show that the approach used can provide a meaningful log-stratigraphic subdivision of a borehole without the aid of any a priori knowledge. Downhole logging can provide an improved interpretation of the units described by core analysis because of the petrophysical characteristics of the formations, which are generally described in logging data. Despite bad hole conditions in some sections of the holes, log data from Leg 159 was particularly valuable because large petrophysical and lithologi-

cal variations occur within the units, and the derived log zones were capable of identifying them.

### ACKNOWLEDGMENTS

The author would like to thank Dr. Peter Harvey at Leicester University Borehole Research for the comments in Geostats, and LENEP/UENF for providing the facilities to finish this paper.

### REFERENCES

Ball, G.H., and Hall, D.J., 1967. A clustering technique for summarising multivariate data. *Behaviour Science*, 12:153–155.

Brewer, T.S., Lovell, M.A., Harvey, P.K., Pelling, R., Atkin, B.P., and Adamson, A., 1990. Preliminary geochemical results from DSDP/ODP Hole 504B: a comparison of core and log data. In Hurst, A., Lovell, M.A., and Morton, A.C. (Eds.), *Geological Applications of Wireline Logs*. Geol. Soc. Spec. Publ. London, 48:195–202.

Davis, J.C., 1986. *Statistics and Data Analysis in Geology* (2nd ed.): New York (Wiley).

Ellis, D.V., 1987. *Well Logging for Earth Scientists*: New York (Elsevier).

Gonçalves, C.A., 1995. Characterisation of formation heterogeneity [Ph.D. dissert.]. Univ. of Leicester, UK.

Le Maitre, R.W., 1982. *Numerical Petrology: Statistical Interpretation of Numerical Data*: Amsterdam (Elsevier).

Masche, J., Lohmann, G.P., Clift, P.D., et al., 1996. *Proc. ODP, Init. Repts.*, 159: College Station, TX (Ocean Drilling Program).

Ocean Drilling Program, 1991. *Wireline Logging Manual* (Vols. 1–3): New York (LDEO-BRG).

Schlumberger, 1981. Natural gamma-ray spectrometry. Schlumberger.

Shepperd, A., Harvey, P.K., and Leake, R.C., 1987. The geochemistry of residual soils as an aid to geological mapping: a statistical approach. *J. Geochem. Expl.*, 29:317–1.

Date of initial receipt: 21 August 1996

Date of acceptance: 23 April 1997

Ms 159SR-031

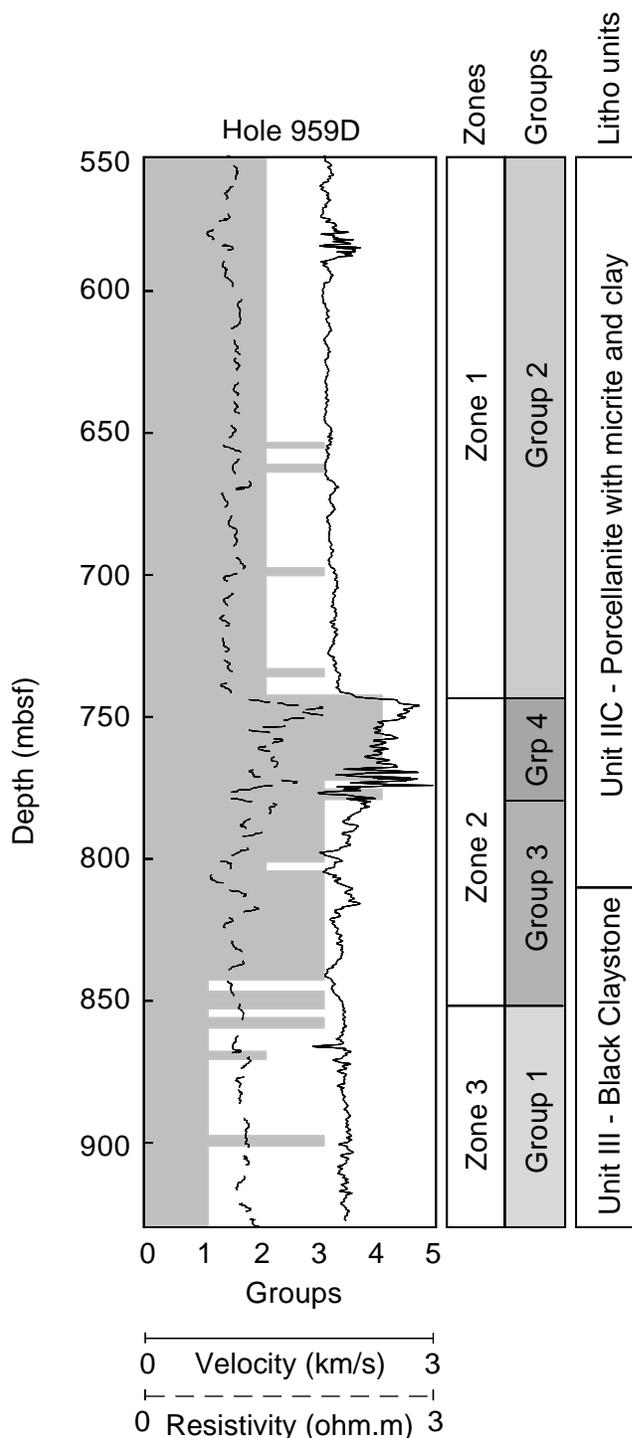


Figure 6. Group logs of Hole 959D for non-hierarchical cluster analysis using the original log curves. Log curves that are of highest importance in defining the groups are also shown.

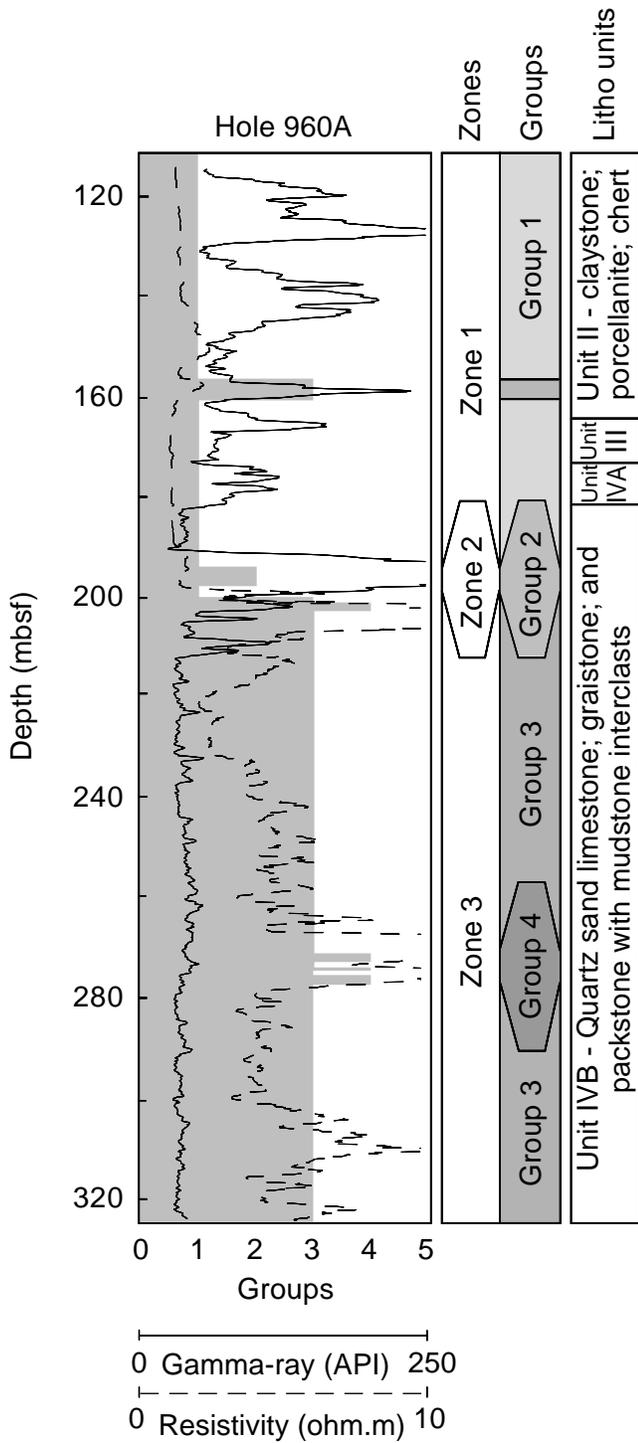


Figure 7. Group logs of Hole 960A for non-hierarchical cluster analysis using the original log curves. Log curves that are of highest importance in defining the groups are also shown.

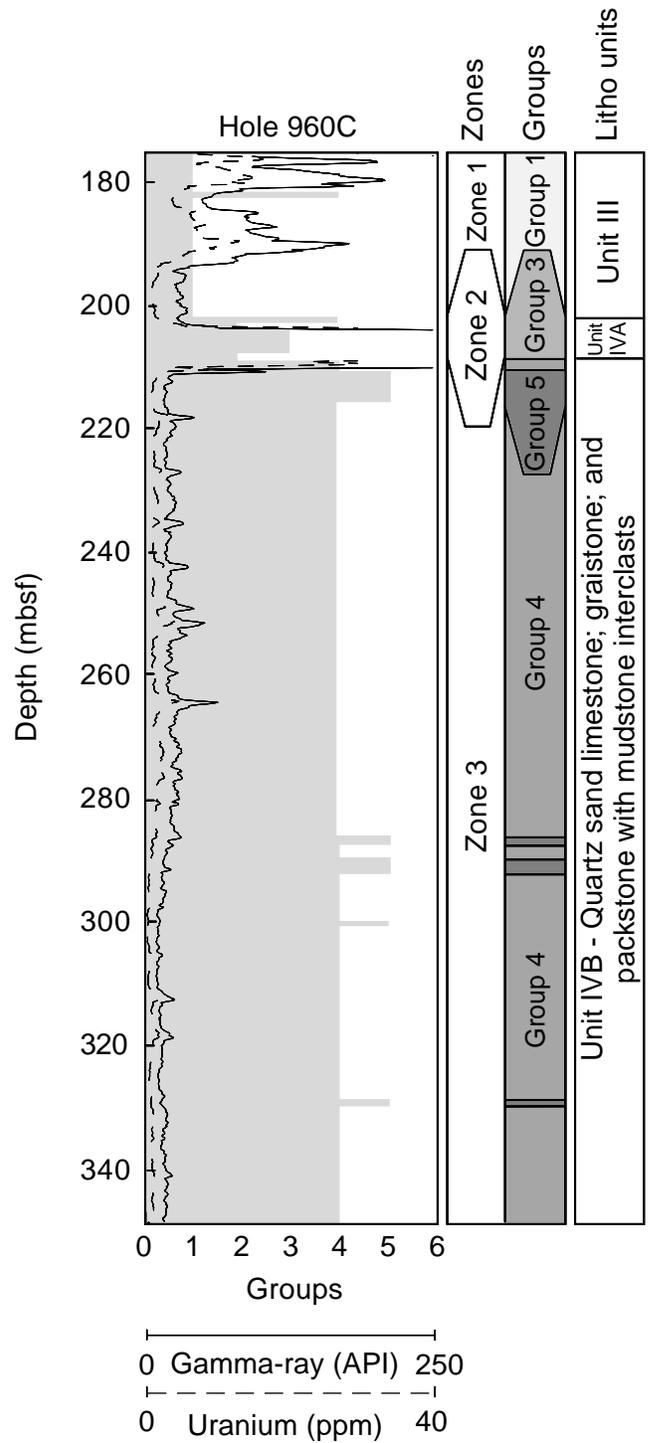


Figure 8. Group logs of Hole 960C for non-hierarchical cluster analysis using the original log curves. Log curves that are of highest importance in defining the groups are also shown.

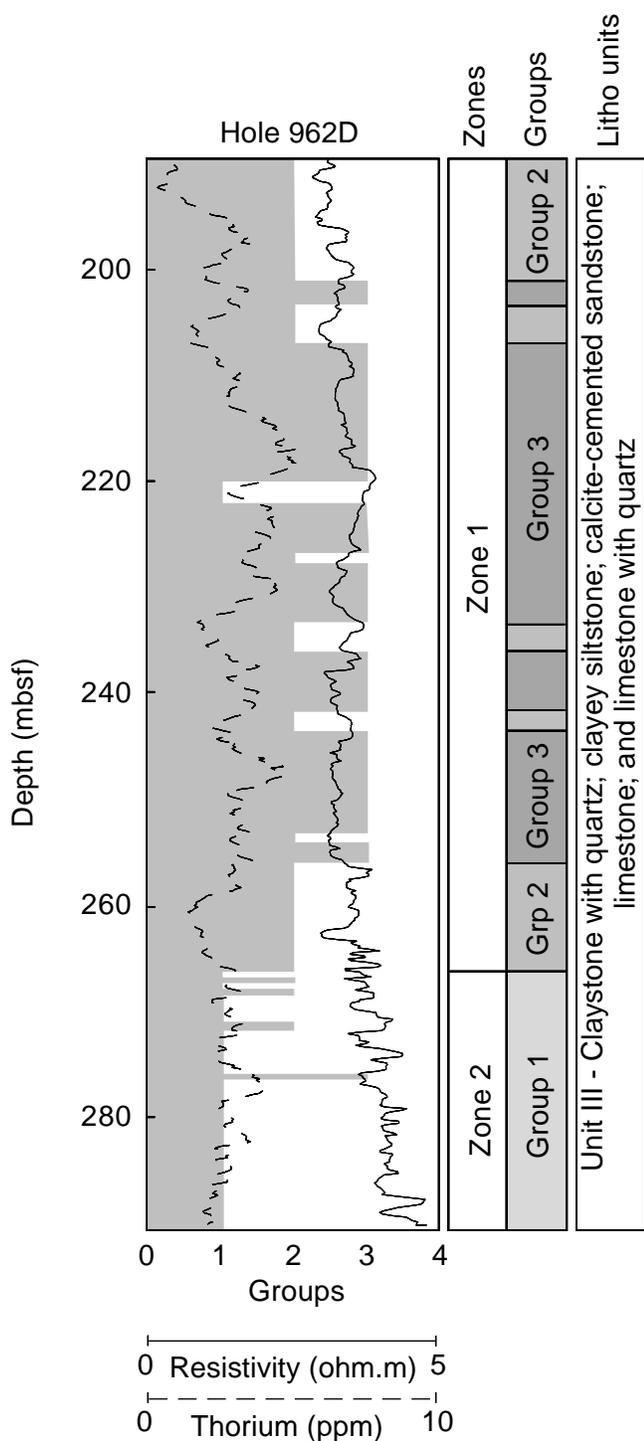


Figure 9. Group logs of Hole 962D for non-hierarchical cluster analysis using the original log curves. Log curves that are of highest importance in defining the groups are also shown.

Table 7. Matrix of distances between group centroids.

Group	1	2	3	4	5
Hole 959D					
1	0.0000	0.1219	0.1269	0.3427	
2	—	0.0000	0.0435	0.1437	
3	—	—	0.0000	0.0715	
4	—	—	—	0.0000	
Hole 960A					
1	0.0000	0.0236	0.1788	0.5510	63.0485
2	—	0.0000	0.0972	0.4081	62.9160
3	—	—	0.0000	0.1070	63.9058
4	—	—	—	0.0000	65.1769
5	—	—	—	—	0.0000
Hole 960C					
1	0.0000	0.0023	0.0043	0.0234	15.6573
2	—	0.0000	0.0005	0.0292	15.9559
3	—	—	0.0000	0.0273	16.1344
4	—	—	—	0.0000	16.2000
5	—	—	—	—	0.0000
Hole 962D					
1	0.0000	0.0092	0.0512	0.2185	
2	—	0.0000	0.0180	0.1522	
3	—	—	0.0000	0.0828	
4	—	—	—	0.0000	

Table 8. Delta values: measure of the importance of the log curves in separate groups (analysis using the principal component scores).

Principal Components	Hole 959D	Hole 960A	Hole 960C	Hole 962D
1	2.95	6.99	8.63	2.11
2	2.78	5.40	4.22	1.70
3	1.65	1.16		1.52

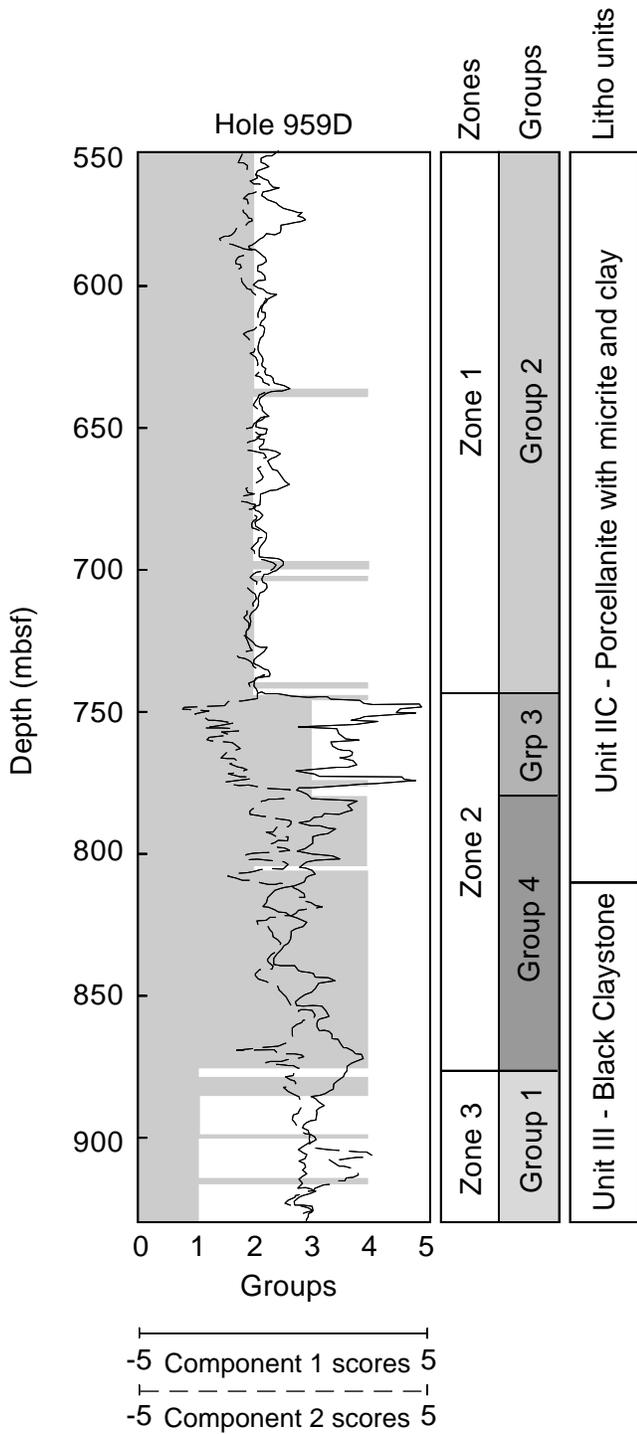


Figure 10. Group logs of Hole 959D for non-hierarchical cluster analysis using principal component scores. Principal Components 1 and 2 are also shown.

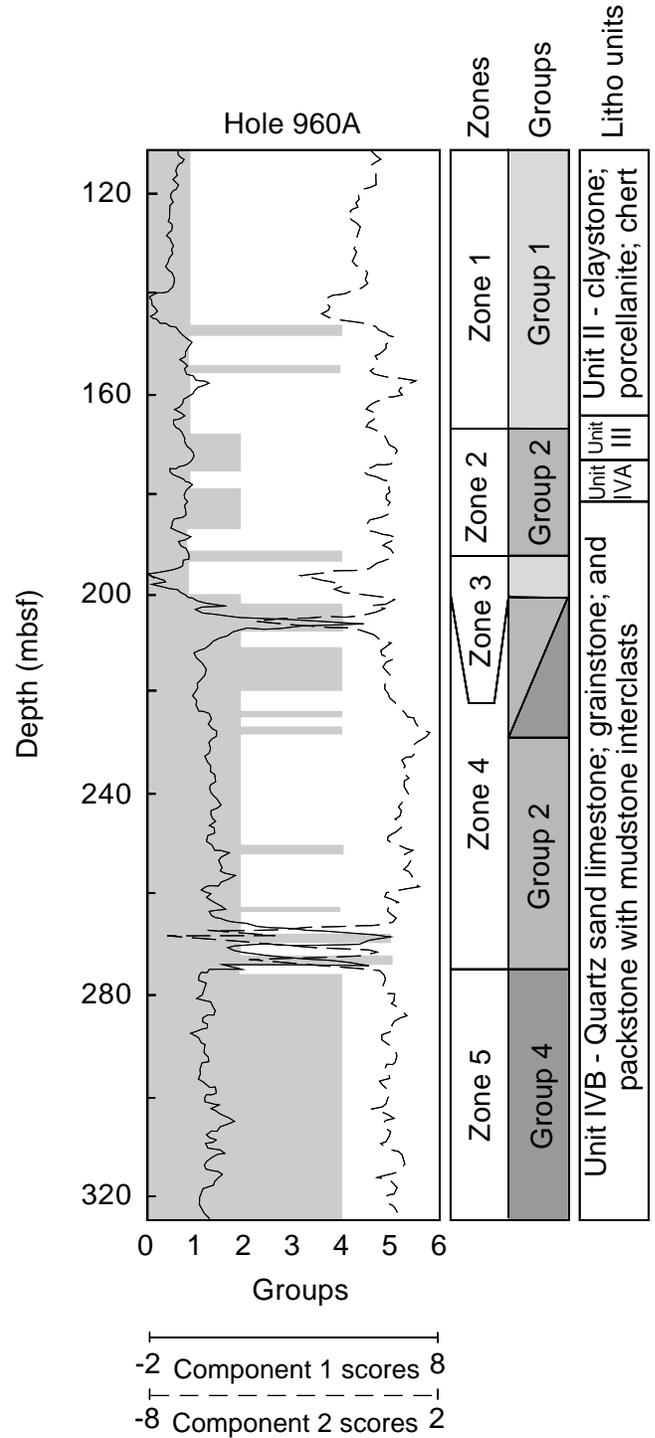


Figure 11. Group logs of Hole 960A for non-hierarchical cluster analysis using principal component scores. Principal Components 1 and 2 are also shown.

