CHARACTERISATION OF FORMATION HETEROGENEITY

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

Carlos Augusto Gonçalves
BSc Geology (Rio de Janeiro/Brazil, 1989)
MSc Geophysics (Belém/Brazil, 1991)

October 1995

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346


LEICESTER
UNIVERSITY
LIBRARY

To Patricia and
Juliana

ACKNOWLEDGEMENTS

I would like to express my thanks to the following individuals:

Acknowledgements must first go to my supervisors *Dr Mike Lovell* and *Dr. Peter Harvey*, who have provided input and encouragement throughout. I gratefully acknowledge the funding of my research to the Brazilian Research Council (CNPq) who provided me with grant and scholarship for the last three years. I also thanks the British Council (Brasilia) who provided a scholarship during the academic year 93/94.

Many thanks to *Peter Harvey* for various discussions in 'geostats' and computing and *Mike Lovell* for the inputs in well log data analysis.

To *Dr Andy Green* (CSMA) and *Dr Max Meju* (University of Leicester) for the helpful discussions and suggestions.

To *Dra. Sonia Guerreiro*, *Dr. João Batista*, *Dr. Carlos Dias*, *Dr. Douglas O'Brien* and *Dr. Mike Lovell*, who gave me all the support in transferring this project from the Federal University of Pará (Brasil) to the University of Leicester; *Regina Neves* and *André Luis* at Federal University of Pará for help with the all bureaucratic work and to *Glauter Souza*, *Wilham Eyben* and *Katiane Lima* at CNPq.

Thanks very much to my friends and colleagues at the Department of Geology, University of Leicester, but in particular to the members of Leicester University Borehole Research, *Ceri Williams*, *Gail Williamson*, *Jim Bristow* and *Lee Ewert*. Special thanks to *Ceri Williams* for all the support throughout this project and *Lee Ewert* for the support during Leg 159 onboard of the *JOIDES Resolution*. Thanks also to *Liane Baldock* and *Janette Thompson* for all the clerical support before and during my stay in England.

To *Elvira, Walter, Luiz, Cesar* and *Anna*, my mother, father, brothers and sister respectively, who despite thousands of miles across the Atlantic Ocean, never denied any support and encouragement. To *Elizabeth Vasconcelos* for the great help at home, keeping things going during the last two months.

Finally, I express my deepest sense of gratitude to my wife *Patricia*, not only for her tolerance but also for her constant encouragement, friendship and love.

Carlos Augusto Gonçalves

Characterisation of Formation Heterogeneity

Abstract

The characterisation of formation heterogeneities requires a multidisciplinary study of data acquired using a large number of numerical geophysical and geological measurements and a rigorous evaluation of the precision and accuracy of the data. Another essential aspect of the appraisal of any measurement is the quality assessment and quality control of the data.

In this work multivariate statistical techniques and an Artificial Neural Network (ANN) are used provide lithofacies characterisation and to identify heterogeneities in complex formations as well as to evaluate the boundaries they generate. The precision and accuracy of the data from different sources are very important and are considered here by using sample support in the integration of measurements at different scales. We use examples from two holes of the Ocean Drilling Program and two oilfield holes to show the differences in characterisation obtained with each technique.

Multivariate Statistical Analysis are initially used to group the petrophysical, geophysical and geological parameters extracted from the downhole measurements into distinct geologically definable zones. This technique has the advantage of being quasi-independent of any pre-determined ideas we have about the whole dataset, and has proved very reliable in formation characterisation. Thus the result obtained here is used as a basis for comparison with that obtained from the Neural Network.

Artificial Neural Network is used to characterise the different lithology sequences present in each well. Neural Networks are relatively new tools and have proved very useful in applications where conventional computing methods are inadequate. Another application is the possibility of determining quantitative petrophysical parameters from well logs and core data in uncored intervals.

The results are presented as a comparison between the two techniques. We show that both methods are very encouraging. When comparing the ANN derived petrophysical parameter logs with actual core measurements and other petrophysical parameters prediction techniques we see a good match. Low quality petrophysical measurements can be determined by a mismatch between the responses.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 - Preface

The characterisation of subsurface formations using downhole logging has changed since the improvement of the available technology allows for more detailed geological and petrophysical descriptions of the subsurface sequences. The variety of data from downhole measurements is so large that a knowledge of a wide range of multidisciplinary techniques is necessary not only to give the interpreter help when working with this amount of data but also to prevent their misuse.

Downhole logging can be defined as the technique of using instruments in a borehole to measure continuously the properties (physical and chemical) of subsurface geological formations. The first logs were made in 1927, when Henry Doll used electrical devices in the borehole environment at Pechelbronn, France (Allaud & Martin, 1977). "Electrical coring" was a station-by-station plot of a single formation resistivity measurement used for well-to-well correlation and to localise possible hydrocarbon bearing zones by measuring formation resistivity in porous sedimentary sequences. In the course of attempting to make other formation resistivity measurements, "noise" was repeatedly noted and after sometime was attributed to a spontaneous potential (Ellis, 1987), which was most notable in front of permeable formations. These two first measurements are still performed routinely today and have performed a major role in hydrocarbon discoveries.

In the late 1930s, the first non-electrical measurement techniques were introduced. The nuclear logging tools, basically in the form of natural gamma-ray devices, were immediately considered useful for distinguishing shaly from clean formations. The gamma-ray tool, in a similar form to its predecessor, is still routinely used for this purpose. The spectral gamma tool, which measures the concentration of the radioisotopes present in a formation, was introduced later and is also still in use (Schlumberger, 1982).

Subsequent logging developments, driven primarily by the needs of the petroleum industry, were related predominantly to the analysis of the porous medium. Neutron porosity devices, sonic logging and nuclear magnetic tools were followed by those extending measurements away from the borehole (the borehole gravimeter tool and the vertical seismic profile). These enhanced the process of formation characterisation and the petrophysical description. The introduction of high resolution resistivity-based dipmeter tools guided important structural and sedimentological

interpretations. Recently, the Formation MicroScanner (FMS*) has been developed from these dipmeter tools and gives an improvement in the analysis of the features close to the borehole wall, such as natural and induced fractures and sedimentary structures, since its data are presented as an image of the formation.

One of the recent advances is in nuclear spectroscopy, which has resulted in the ability to determine *in situ* abundances of the major rock forming elements (Chapman et al., 1987). The GLT* (Geochemical Logging Tool) uses natural, activation and prompt neutron capture gamma-ray spectroscopy to measure continuously downhole elements such as Si, Al, Ti, Ca, Fe, K, S, H, Cl and (Mg+Na) as well as the minor elements Gd, Th and U (Hertzog et al., 1989; Lovell et al., 1993 and Bristow, 1993).

Geophysical downhole logging has considerable potential in extending geological knowledge. The technique has been routinely used in hydrocarbon exploration for many years and is the focus of extensive research (Prensky, 1987, 1990-1993). The enormous potential, however, for a wide range of geological and geophysical applications using all the information available has yet to be fully developed. Different techniques have been used for formation characterisation purposes and many examples can be found in the literature. The indiscriminate use of different techniques, however, without a careful study of the data can lead to misinterpretations and errors in the process of formation characterisation. This may be due to the complexity of the dataset rather than the geological formations. This work attempts to show the use of some of these techniques in formation characterisation when a broad range of downhole measurements are considered. The identification of formation heterogeneities through lithofacies and petrophysical characteristic changes is also one of the aspects studied.

1.2 - Aims and objectives of the thesis

The complete characterisation of subsurface formations from the interpretation of well logs and core data is perhaps a fundamental requirement when geologists and geophysicists are involved in the exploitation of energy resources. It generally requires a multidisciplinary study of data acquired using a large number of geophysical and geological measurements.

The principal aim of this research has been to assess the usefulness of well log measurements in the characterisation of formations as well as the identification of heterogeneities. Despite being widely used in hydrocarbon exploration for many years, logging has only recently begun to be employed in other fields and its use for

* Trade mark of Schlumberger

purely scientific purposes has, to date, been limited mostly to the ODP (Ocean Drilling Program) and other continental drilling programs such as Troodos, KTB and Cajon Pass (Lovell, pers. commun.). The broad range of downhole measurements and different geological environments investigated (e.g. sedimentary formations, igneous and metamorphic rocks) allow for the assessment of both the quality and validity of these measurements to determine the most efficient ways of extracting geological information from logging data.

An essential aspect of the appraisal of logging measurements, especially when using new technology such as the GLT, is the quality assessment and quality control of the data. The precision and accuracy of the data from different sources are very important and will be considered here by using sample support in the integration of measurements at different scales.

The basis of the use of a variety of techniques of value in partitioning logging data are presented in this work. From the use of simple conventional cross-plots analysis and the application of multivariate statistics to the use of more complex techniques such as Neural Network, this work discusses the benefits and shows some of the pitfalls of each technique. Data from the Ocean Drilling Program and two oilfield holes are used.

This research concentrates firstly on the use of Multivariate Statistical Analysis to group the petrophysical, geophysical and geological parameters extracted from the downhole measurements into distinct zones. These techniques have the advantage of being quasi-independent of any pre-determined ideas about the dataset. The diversity in terms of geological environments observed in each dataset used in this analysis can also be useful to identify different types of heterogeneities from geological sources.

Secondly, Neural Network is used to characterise the different lithofacies present in each hole investigated. The application of one of the most widely used Neural Networks - the Backpropagation - is shown. Neural Networks are relatively new tools in computing and have been shown to be very useful in applications where conventional computing methods are inadequate. In this case log or core measurements are used to base the training process. The result obtained from the Neural Network for lithofacies identification can then be compared with the results obtained using the Multivariate Statistical Techniques for the same datasets. The computation of different petrophysical parameter logs is another important application of Neural Networks. It is performed directly using the downhole measurements to predict some physical properties which were not possible to be measured through coring.

During the period of the study, additional logging acquisition, interpretation and processing experience was obtained through participation as one of the Logging

Scientists on ODP Leg 159 on the Côte D'Ivoire-Ghana margin. The cruise lasted for two months (January to March/95) and a paper with the preliminary logging results from that work (Gonçalves et al, 1995b) was presented at the I Latin American Geophysical Congress (Rio '95). A copy of this paper is given as Appendix A together with a general description of the Leg. Although that work is not directly concerned with the thesis, it was considered a fantastic experience in terms of research training.

1.3 - Structure of the thesis

The structure and format of the remainder of this thesis is as follows. In Chapter 2 an introduction to the Ocean Drilling Program is given and the well logging tool configurations that this program regularly deploys are described; the logging tools used are state-of-the-art industry tools, but have some modifications for ODP use. Descriptions of the spectroscopy measurements made by the GLT and the images obtained by the FMS tool are given, together with a description of the ODP holes and sites used as datasets in this work. Data quality and log characteristics are also discussed in this chapter.

In chapter 3 a Multivariate Statistical Analysis of the data is presented. Multivariate Statistical Techniques that have been applied to logging data in the course of this work are described. After a description of the theoretical fundamentals, the application of the Principal Component Analysis, Cluster Analysis and Discriminant Analysis used here are shown. In Discriminant Analysis, results from different models are presented and are also used for comparison with results from chapter 4. Finally, a discussion about the results obtained is given together with a summary and some conclusions.

Chapter 4 is based on the application of Neural Networks in formation characterisation. This technique has been used in other fields of science and some work has been done in well log characterisation. A review of what Neural Network is, showing the theoretical fundamentals is introduced. Then a description of the Backpropagation (BEP) Neural Network, its structure and algorithm and its application to different data sets is given. The first application of the BEP Neural Network is on the formation characterisation using well log data from ODP and two oilfield holes. The results are then compared with the ones obtained in chapter 3 with the statistical techniques. The second application is the use of the BEP Neural Network to generate petrophysical logs for uncored intervals. Firstly, Backpropagation Neural Network is used to predict physical property measurements in uncored intervals of ODP holes, using core measurements from the same hole. Then, the application is shown for well-to-well prediction where the Neural Network

is trained with data from one hole and the results are applied in a different hole. As in the previous chapter, a discussion of the results and some conclusions are shown at the end.

Chapter 5 summarises the results obtained from the statistical techniques and the Neural Network both for lithofacies identification and petrophysical parameters prediction. The strengths and weaknesses of all techniques are shown as well as a comparison between the results obtained. Suggestion for further work which can improve the characterisation of heterogeneities are given. Finally, this chapter draws together some discussions and conclusions about the use of those techniques.

1.4 - Author's work and work of others

This section aims to highlight the author's own contribution against the work of others used in this study. It also shows the source of the principal algorithms and lists the main software packages used.

The Leicester University Borehole Research program library provided all the multivariate statistical algorithms used in this study. Some changes in the original code of the algorithms were performed in other to implement the classification given by the techniques. One of these changes is the computation of the Delta values (sections 3.4.2 and 3.4.3) in the Non-Hierarchical Cluster analysis, which produces the importance of each log curve during classification.

The Backpropagation Neural Network algorithm was developed by the University of Genoa (Italy). The main change performed in the original algorithm is related to the input of data which had to be modified to allow its use for petrophysical parameter estimation (section 4.6).

The author's work can be summarised by carrying out all the analysis and interpret the results geologically and in terms of the best models. It also consisted of the interpretation of the datasets in terms of the data reduction and its implications and limitations in the analysis, the evaluation of the performance of each technique and how they could help to improve the process of Formation Characterisation. One of the important contributions in this study is the change of the code and architecture of the Backpropagation Neural Network to adapt it for using in the petrophysical parameter estimation.

Some other software packages were used in this study. They are commonly used for drawing and typing in scientific work. Word Processor, Kaleidagraph, FreeHand, Minicad and EndNote are among them.

Chapter 2

THE OCEAN DRILLING PROGRAM (ODP)

The Ocean Drilling Program (ODP) is the successor to the DSDP (Deep Sea Drilling Project) which was a global investigator of the ocean basins. DSDP began operations in the late 1960's using a 120 meters drillship, *The Glomar Challenger*. The DSDP started with an extensive involvement of international scientists who were invited to participate on drilling cruises. The international interest continued to grow during the 1970's and early 1980's with the formal participation of other nations. The remarkable scientific results obtained and the new questions about the ocean basins it had generated demanded a continuing capability for drilling in the ocean (Rea et al., 1993).

In 1983, the Ocean Drilling Program was organised, international participation was co-ordinated, a new drillship (*The JOIDES Resolution*) was contracted and outfitted, and the first cruise sailed in early 1985. Among the principal objectives is the study of the geological processes that have shaped our planet and modified its environment. The scientific problems being addressed range from the geologic history and structure of continental margins to the study of the ocean crust including ocean history, climate and tectonics. The ODP is the only facility for continuously sampling the ocean basins. It has now operated in all oceans except the ice-covered Artic ocean.

The program operates a specially equipped deep-sea drilling ship, *The JOIDES resolution* (Sedco/BP 471), which contains state-of-the-art laboratories, equipment and computers. The ship is 144 meters long and 21 meters wide. The size and ice-strengthening of the ship allow drilling in high seas and permits a large group of multidisciplinary scientists to interact as part of the scientific party. The derrick towers 64 meters above the sea level and has computer-controlled dynamic-positioning which holds the ship over a specific location while drilling in water depths up to 8200 meters. More than 1000 square meters in the ship are devoted to scientific laboratories and equipment allowing a high capacity of on board petrophysical measurements (Premoli Silva et al., 1993).

Logging is an important part of the program. The ODP provides a full suite of geochemical and geophysical measurements. For each hole investigated, basic oil-industry tools are used: nuclear, sonic and electrical. In addition, Geochemical Logging Tool (GLT) and a borehole televiewer are available. A 12-channel logging tool provides accurate velocity and elastic property measurements as well as sonic waveforms for spectral analysis. Finally a vertical seismic profile can record

reflectors from below the total depth of the hole and the Formation MicroScanner (FMS) provides images of the borehole wall.

Texas A&M University (TAMU) serves as science operator for ODP. The Lamont-Doherty Earth Observatory (LDEO) of Columbia University is responsible for the program's logging operation, including processing the data and providing assistance to scientists for data analysis. It has a sub-contract with Leicester University Borehole Research (University of Leicester - UK) and the Institut Mediterranean de Technologie (IMT - France) which provide planning, data processing, shipboard logging scientists and other services for the cruises. The ODP Data Bank, a repository for geophysical data is also managed by LDEO, whilst TAMU holds a database for core data.

The large variety of geological environments investigated and the use of new technology in downhole measurements and core recovery provide new challenges for the logging interpretation techniques.

2.1 - Well logging in ODP

Well log has been largely used in the oil-industry for many years and has been the subject of much research (see Prensky, 1987, 1990-1993). The use of logging tools in ODP operations is less well known and therefore is introduced in this chapter. The tools described are common to most ODP Legs. A more comprehensive guide on the well logging techniques can be found in ODP (1991).

Although log measurements are still dependent on core-based measurements in some terms, like calibration, the great advantage of the log measurements is that they represent quasi-continuous *in situ* measurements of the formations. After coring is completed, a combination of sensors is lowered downhole to monitor the physical and chemical properties of formations adjacent to the borehole wall (ODP, 1991; Gillis et al., 1993; Premoli Silva et al., 1993).

In ODP, under sea conditions, a heave motion compensator is generally employed to minimise the effect of ship heave on tool position in the borehole. The interpretations of those continuous *in situ* measurements can yield a stratigraphic, lithologic, structural, geophysical and geochemical characterisation of the hole. Six combinations of downhole sensors are available, although four of them are the most commonly used (ODP, 1991):

1) The Schlumberger Seismic Stratigraphic combination, which includes the long-spaced digital sonic tool (LSS-SDT), the natural gamma-ray spectrometry tool (NGT), the phasor dual induction tool (DITE) and the Lamont-Doherty Earth Observatory (LDEO) temperature logging tool (TLT);

2) The Lithoporosity combination, which includes the NGT, the dual porosity compensated neutron tool (CNT-G), the high temperature lithodensity tool (HLDT) and also the TLT;

3) The Formation MicroScanner (FMS) tool string, composed of the general purpose inclinometry tool (GPIT) and the microelectrical scanning tool (MEST), together with the NGT;

4) The Geochemical logging combination consisting of an NGT, an aluminium activation clay tool (AACT) and an induced gamma-ray spectrometry tool (GST), to which also is attached the TLT;

5) The Schlumberger well seismic tool (WST);

6) The dual lateralog tool (DLT) with the NGT and the TLT.

The commonly used four tool strings are schematically shown in Figure 2.1 while Table 2.1 also lists the tools in the above combinations, the main logs they produce, their approximate vertical resolution and range of depth of investigation. Data are typically recorded at vertical intervals of 0.1524 meters, equivalent to 6 inches.

The NGT is run in all combinations to provide a common basis for depth correlation. The DLT is generally used for igneous rocks, where it is better suited than DITE because of the high electrical resistivity of these rocks (Schlumberger, 1989a). Most downhole data become quickly available on board ship, but a few of them like the geochemical log require further processing, which is done onshore after the end of each Leg.

Table 2.1

Tool acronym	Tool name	Method	Measurement	Vertical resolution (cm)	Depth of investigation range (cm)	Log acronym
CALI	Caliper	4 bow arms	Hole diameter			CALI, HD
DITE	Dual Induction tool	Induction	Deep resistivity	200	120-150	ILD, IDPH
		Spherically focused current	Medium resistivity	150	60-80	ILM, IMPH
			Shallow resistivity	75	30-40	SFLU
DLT	Dual lateralog	Focused current	Deep resistivity Shallow resistivity	60 60	> 200 < 100	LLD LLS
LSS-SDT	Long-spaced sonic digital tool	3 travels intervals	Slowness and Full sonic P-wave	61	5-75	LTT1-4 DTLN,DTLF
NGT	Natural gamma-ray tool	Natural gamma-ray	Total gamma-ray U, Th, K	45 45	30-50	SGR, CGR
HLDT	High-temperature lithodensity tool	Gamma-ray scattering	Density Photoelectric effect	45	20-70	RHOB,DRHO PEFL,PEFS
GST	Gamma-ray spectrometry tool	Induced gamma-ray	Ca,Si,Fe,F,H,Cl	45	20-70	CCA,CSI, CFE,CSUL, CHY,CCHL
AACT	Aluminium activation clay tool	Induced gamma-ray	Al	45	20-70	UWAL,ALUM
CNT-G	Compensated neutron tool	Neutrons back-scattering	Porosity	45	20-70	NPHI,TNPH, ENPH
FMS	Formation MicroScanner	Image of the borehole	Resistivity	0.5	25	
MEST	Microelectrical Scanning tool	Caliper	Hole diameter			C1,C2
GPIT	General purpose inclinometry tool	Magnetometer	Magnetic field	4		FX,FY,FZ,FN
		Accelerometer	Acceleration	4		AX,AY,AZ,
			Inclinometry	4		FINC,AZIM, DEVI,PIAZ, RB,HAZI

Table 2.1 - Downhole tools and principal logs table (modified from Gillis et al., 1993).

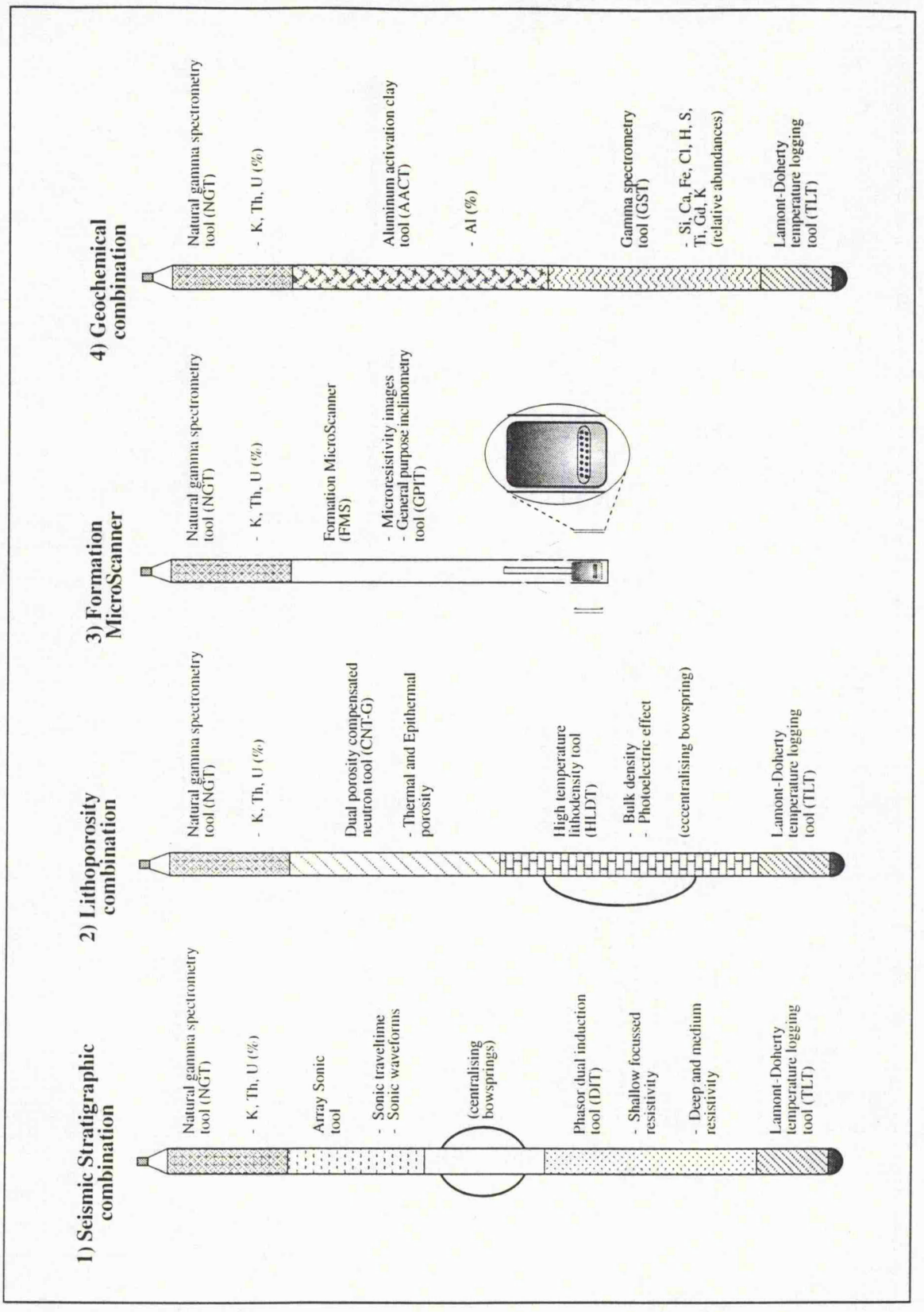


Figure 2.1 - Sketch of the four logging tool combinations used in ODP operations. The Seismic Stratigraphic and the Lithoporosity combinations are usually combined in the standard Quad Combo (after ODP, 1991).

2.1.1 - The ODP tool configuration

2.1.1.1 - The Seismic Stratigraphic combination

The Seismic Stratigraphic combination is designed to measure compressional wave velocity (with shear and tube wave velocity obtained from the full sonic waveform), deep, intermediate and shallow resistivities and the natural radioactivity of the formation in a single logging pass (Rea et al., 1993). It also includes the Lamont-Doherty temperature tool at the bottom of the string.

The DITE provides three measurements of electrical resistivity, all having different radial depths of investigation. The induction devices ("deep" and "medium" resistivities) send high frequency alternating current through transmitter coils, creating a magnetic field that induces a secondary (Foucault) current in the formation (Ellis, 1987). This produces a new signal, proportional to the conductivity of the formation. This signal is then converted to resistivity values. A third device, the spherically focused log (SFL), measures the current necessary to maintain a constant voltage potential across a fixed interval. The vertical resolutions (as shown in Table 2.1) are 2 meters and 1.5 meters for the deep and medium resistivities respectively and approximately 0.75 meters for the SFL. These data are automatically corrected for irregularities in borehole diameter. Porosity, fluid salinity, clay and hydrocarbon content and temperature are important factors influencing electrical resistivity. Some other factors are also important as the concentration of hydrous and metallic minerals and the geometry of the porous space (Bateman, 1985).

The natural gamma spectroscopy tool (NGT) measures the natural gamma radiation of the formation (Lock and Hoyer, 1971). The gamma-ray spectrum is divided into five discrete energy windows and counts are recorded for each window. Most natural gamma-ray are emitted by ^{40}K, U and Th isotopes and their decay series. Near the borehole wall, natural gamma-ray emissions from the formation are measured by a NaI scintillation detector mounted inside the sonde. K, Th and U are generally most abundant in clay minerals; thus this tool is commonly used as an estimator of the clay content of the formation. Silicic volcaniclastic material and K-feldspar-rich rocks can also have high concentration of these elements. Hurst and Milodowsky (1994) also show that some Th contents are related to the presence of detrital heavy minerals, like monazite and zircon, in sandstone. Therefore, the interpretation must be tied to the core lithologies.

The LSS-SDT (long-spaced digital sonic tool) records time required for sound to travel along the borehole wall. It uses two acoustic transmitters and two receivers to measure the time interval for sound waves to travel vertically through the formation near the borehole wall. First arrivals for the individual source-receiver paths are used

to calculate sonic velocities, thus four velocities values are measured at each depth (Schlumberger, 1989a). The vertical resolution of the tool is approximately 61 cm. Only V_p (compressional velocity) is determined aboard the ship, but the full sonic waveform is generally recorded for future processing to determine V_S (shear velocity) and V_t (tube wave velocity). V_p is controlled primarily by density and lithification. Increase in density and increase in lithification cause velocity to increase.

2.1.1.2 - The Lithoporosity combination

The Lithoporosity combination includes the lithodensity (HLDT), the compensated neutron porosity (CNT-G) and the NGT (Figure 2.1). The HLDT uses a gamma-ray source (^{137}Ce of 0.66 MeV) to induce a back-scattered flux of gamma-rays that is measured at fixed distances from the source. The source is mounted in the tool body and an eccentralising spring-loaded caliper arm forces it and a pair of detectors against the borehole wall (Ellis, 1987). The detectors measure the flux of the gamma-rays to determine the formation density (RHOB) and the photoelectric effect (PEF). Attenuation of these induced gamma-rays is mainly caused by Comptom scattering and, therefore, controlled by the density of electrons in the formation. Formation density is extrapolated by assuming that the atomic weight of most rock-forming elements is approximately twice the atomic number. Photoelectric absorption occurs in the energy window below 150 KeV and depends on the energy of the incident gamma-ray, the atomic cross-section and the nature of the atom. The PEF measurement is almost independent of porosity and, therefore, can be used directly as a matrix lithology indicator because it is function of the atomic number. Excessive roughness on the borehole wall allows drilling fluids between the tool and the formation and, consequently, leads to underestimated density measurements (ODP, 1991). The depth of investigation also depends on the density of the rock; the higher the density, the lower the penetration. The vertical resolution of the tool (Table 2.1) is about 0.45 meters.

The CNT-G uses an americium-beryllium source mounted in the sound which emits high energy neutrons (4 MeV) into the formation. These neutrons are scattered by collision with other nuclei (Schlumberger, 1989a). Collisions with heavy atoms do not exchange much energy, but collisions with hydrogen, which has a similar mass, significantly reduce the energy levels of the neutrons. When the neutrons reach a low energy level (less than 0.025 MeV) they are captured and absorbed by atomic nuclei such as hydrogen, chlorine, silicon and boron (ODP, 1991). The tool uses the scattering cross-section as the quantity that describes the rate at which neutrons are slowed. Because the scattering cross-sections of the hydrogen is about 100 times larger than for any other common element in the crust, most energy dissipation is

caused by collision with water molecules. As the thermal neutron flux is affected by hydrogen which is both chemically bound as well as in free water, the device also performs a second measurement using changes in intermediate-energy (epithermal) neutron flux to potentially measure the free water contribution only (Schlumberger, 1982 and Ellis, 1987). The vertical resolution of the tool is about 45 cm (Table 2.1).

2.1.1.3 - The Formation MicroScanner

The Formation MicroScanner (FMS) tool produces high resolution microresistivity images of the borehole wall that can be used for detailed sedimentological or structural interpretations (Ekstron et al., 1986; Pezard and Luthi, 1988) and for determining fracture and breakout orientation (Luthi and Souhaité, 1990 and Schlumberger, 1989b).

The tool (Figure 2.1) consists of sixteen electrodes (called "buttons") on each of four orthogonal pads that are pressed against the borehole wall. The electrodes are spaced about 2.5 mm apart and are arranged in two diagonally offset rows of eight electrodes each. It was originally introduced by Schlumberger in 1986 but a modified sensor was developed for ODP purposes in 1989 because of the restrictions in using it at the narrower-gauge drill pipe (Pezard et al., 1990a and 1990b). The focused current that flows from the buttons is recorded as a series of curves that reflect the microresistivity variations of the formation (Schlumberger, 1989b). It uses a general purpose inclinometry tool (GPIT) which spatially orients the resistivity measurements from accelerometry measurements and from the declination and inclination of Earth's magnetic field vector. On board or shore-based processing converts the measurements into complete, spatially-oriented images of the borehole wall. The FMS string also contains the NGT so that the FMS data may be correlated with the other logs.

The measurement spacing is about 0.5 cm, but coverage is restricted to about 22% of the borehole wall for each pass of the tool, assuming a 24.76 cm in hole diameter.

Applications of the FMS images include detailed correlations of coring and logging depths, orientation of cores, mapping of faults, fractures and formation structures as well as determining strikes and dips of bedding. The FMS can also be used to measure stress in the borehole through breakout delineation. An important limitation of the tool is the restriction of hole diameter to less than 38.10 cm (15 in). Thus no useful information can be obtained from washed-out hole sections.

A detailed study of the FMS tool, including fundamental physics of the measurements, the instrumentation and the processing of the images can be found in Schlumberger (1989b).

2.1.1.4 - The Geochemical logging combination

The geochemical tool string deployed in ODP operations is a combination of the Geochemical logging tool (GLT) and the TLT as shown in Figure 2.1. The GLT includes basically the NGT, a gamma-ray spectrometry tool (GST) and an aluminium activation clay tool (AACT) (Hertzog et al., 1989; Lovell and Anderson, 1989).

The first part of the GLT is the NGT which was described in section 2.1.1.1. The AACT forms the second part of the geochemical logging tool and measures the concentration of Al in the formation by delayed neutron activation (Scott and Smith, 1973). When the natural isotope ^{27}Al absorbs a thermal neutron derived from the 2.5 MeV ^{252}Cf source of the AACT, it forces an unstable ^{28}Al atom having a half-life of about 2 minutes. When this unstable nucleus decays (to ^{28}Si), a gamma-ray having a characteristic energy (1779 KeV) is emitted and subsequently detected by the AACT. Because the AACT simultaneously counts the natural gamma-radiation of the formation, the Al spectrum is determined by subtracting the count rates from the NGT, which is positioned above the AACT in the tool string.

The third part of the GLT is the GST which consists of a pulsed 14 MeV neutron generator and a NaI scintillation detector. Incident neutrons lose energy through inelastic scattering interactions and, on reaching thermal energy levels, are captured by elemental nuclei. Characteristic gamma-rays are emitted upon neutron capture; these gamma and their relative energy levels are recorded by the tool. The 256-channel energy spectrum is deconvolved to determine relative abundances of Ca, Si, Fe, Cl, H and S on board the ship (Bristow, 1993). The post cruise processing of the GST data provide the additional elemental yields of Gd and Ti. The above yields (except Cl and H) are them combined with elements determined from the NGT and the AACT to derive dry weight percentages of the elements Si, Ca, Fe, S, Ti, K and Al in addition to Gd, Th and U (ppm) (Rea et al., 1993). An estimate of (Mg + Na) can be made by using the photoelectric factor from the HLDT (Schweitzer et al., 1988; Hertzog et al., 1989; Lovell et al., 1993).

The post oxide percentages are simply obtained by multiplying the percentage of each element by its associated oxide factor. Discretion is needed in determining which oxide or carbonate values are used, depending on the minerals present and the geological environment being logged. However, as a new technique, it has some potential sources of error. These sources can be split in two main categories, errors associated with data acquisition and errors associated with post acquisition processing. The possible data acquisition errors are numerous but principally concern the quality of the gamma-ray signal which may be degraded by poor counting rates, poor detection resolution and poor signal to noise ratio caused by environmental effects. The post acquisition processing errors are principally attributed to errors in

the spectral inversion of the raw data and errors in the oxide closure model (ODP, 1991). The latter includes errors in the Al and K data, which will, in turn, induce errors into the other derived elements and errors in the oxide closure model caused by the presence of unmeasured elements in the formation and incorrect oxide factor assumptions. Bristow (1993) demonstrates a new method for estimating the statistical uncertainty associated with the GST-derived elements.

Considering the AACT source is 2.35 MeV and the GST is 14 MeV and that the distance that a fast neutron travels through a material is governed by the neutrons initial energy and nuclear properties of that material, it is likely that the distance which a neutron slows down to induce a capture reaction and therefore the distance a neutron flux penetrates into a formation, will vary (Lofts, 1993). Pelling (1992) shows some simulation studies for a sandstone with 40% of porosity. For 14 MeV source neutrons slow to a capture reaction over a distance of 13.8 cm, while a 2.35 MeV source neutrons slow in only 8.1 cm. He also shows examples for a sandstone and a limestone for two different neutron sources and the effect porosity has on both. Porosity and source strength clearly affect the distance travelled by neutrons because of the area and volume of investigation. The AACT does not necessarily have the same depth of investigation as the GST.

2.2 - This section describes the ODP and other data used in this work

2.2.1 - ODP Hole 807C - Leg 130: Ontong Java Plateau

2.2.1.1 - Introduction

The Ontong Java Plateau in the western equatorial Pacific is a broad mid-oceanic submarine plateau striking Northwest and paralleling the Solomon Islands to the south. Its name is taken from an exceptionally large atoll north of these islands. The plateau occupies an extensive area (1000 x 1500 km) and rises to unusually shallow depths in its central region (around 1700 meters). Kroenke (1972) shows that the physiography along the margin is complex, with atolls or seamounts located near the western and south-western edges.

Previous drilling expeditions established the general stratigraphy of the sediment cover (Winterer, Riedel et al., 1971; Andrews and Packham et al., 1975) which consists of a pelagic carbonate deposits of Mesozoic and Cenozoic age, well stratified, more than 1000 meters thick, and overlying a volcanic sequence. The sediments cover a large portion of the plateau.

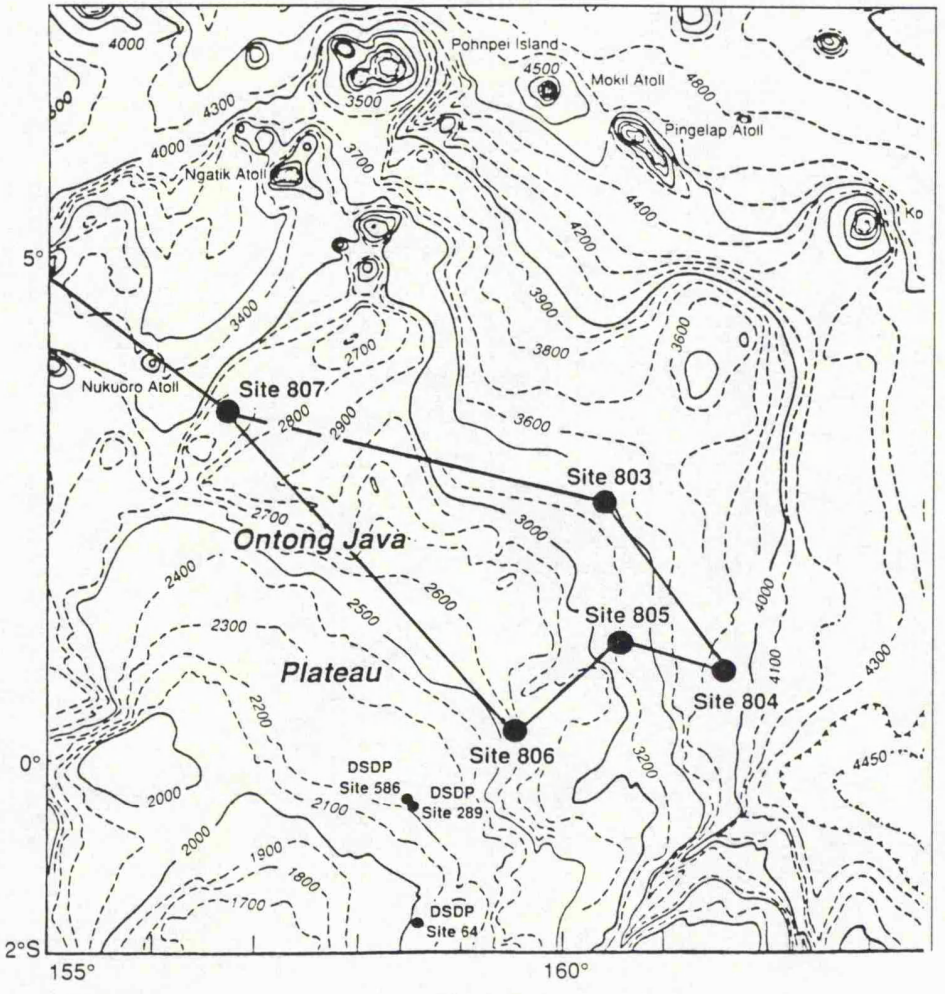


Figure 2.2 - Bathymetry map of the Ontonga Java Plateau showing ODP sites 803-807 during ODP Leg 130.

ODP Hole 807C is located on the northern rim of the Ontong Java Plateau (latitude 3°36.4' N, longitude 156°37.5' E) in 2805 meters of water depth (Figure 2.2). It was the third hole drilled in this location (site 807) - following the 807A and 807B holes - and reached 1528.4 meters below sea floor (mbsf) with 33.8% of total core recovery (Kroenke et al., 1991). It was also successfully logged to the total depth drilled and a complete set of geophysical measurements, including Formation MicroScanner (FMS) images were obtained. The sedimentary sequence identified for the total depth in this hole consists of three main units, as follows:

Unit I (0 - 968.0 mbsf) - composed mainly of Pleistocene to upper/middle Eocene nanofossil ooze and chalk with foraminiferae.

Unit II (968.0 - 1351.4 mbsf) - composed of upper/middle Eocene to upper Campanian limestone, chert and nanofossil chalk with foraminiferae.

Unit III (1351.4 - 1379.7 mbsf) - composed of lower Cenomanian to upper Albian-Aptian claystone, siltstone with radiolarians and limestone.

Unit IV (1379.7 - 1528.4 mbsf) comprises igneous basement and is composed of Albian-Aptian tholeiitic basalts. The units will be described in more detail in the next section.

2.2.1.2 - Lithostratigraphy sequence

For the purpose of this study the interval between 1270 - 1400 mbsf was selected in the ODP Hole 807C. The reason is that it comprises in 130 meters interval three of the four main lithofacies observed in the whole section of the hole. Other reason is that the core recovery in this interval is about 65% which allows a good geological description of the sequences. A complete logging survey was also obtained in this interval and will be described later on.

The sedimentary sequence observed in this interval can be seen in Figure 2.3. It starts at the top with a white and grey limestone sequence (corresponding to the bottom part of unit II) with less than 5% of chert present in the recovered sediments. Some clay-rich clast bearing intervals are observed below 1290 mbsf. Chert is present as nodules or in layers up to 7 cm thick (Kroenke et al., 1991).

The next unit consists of upper Albian to lower Cenomanian claystone and siltstone with varying amounts of radiolarians and Aptian to Albian limestone. The unit can be divided in two subunits (Kroenke et al., 1991) on the basis of claystone and limestone abundances. Subunit IIIA (1351.4 - 1369.7 mbsf) includes claystone, siltstone with radiolarians and sandy siltstone. Siltstone beds are rare at the top of subunit IIIA and become more abundant with depth. The downcore increase in siltstone frequency produces an overall upward fining sequence in subunit IIIA. The contacts between the siltstone and claystone intervals are dominantly gradational, but

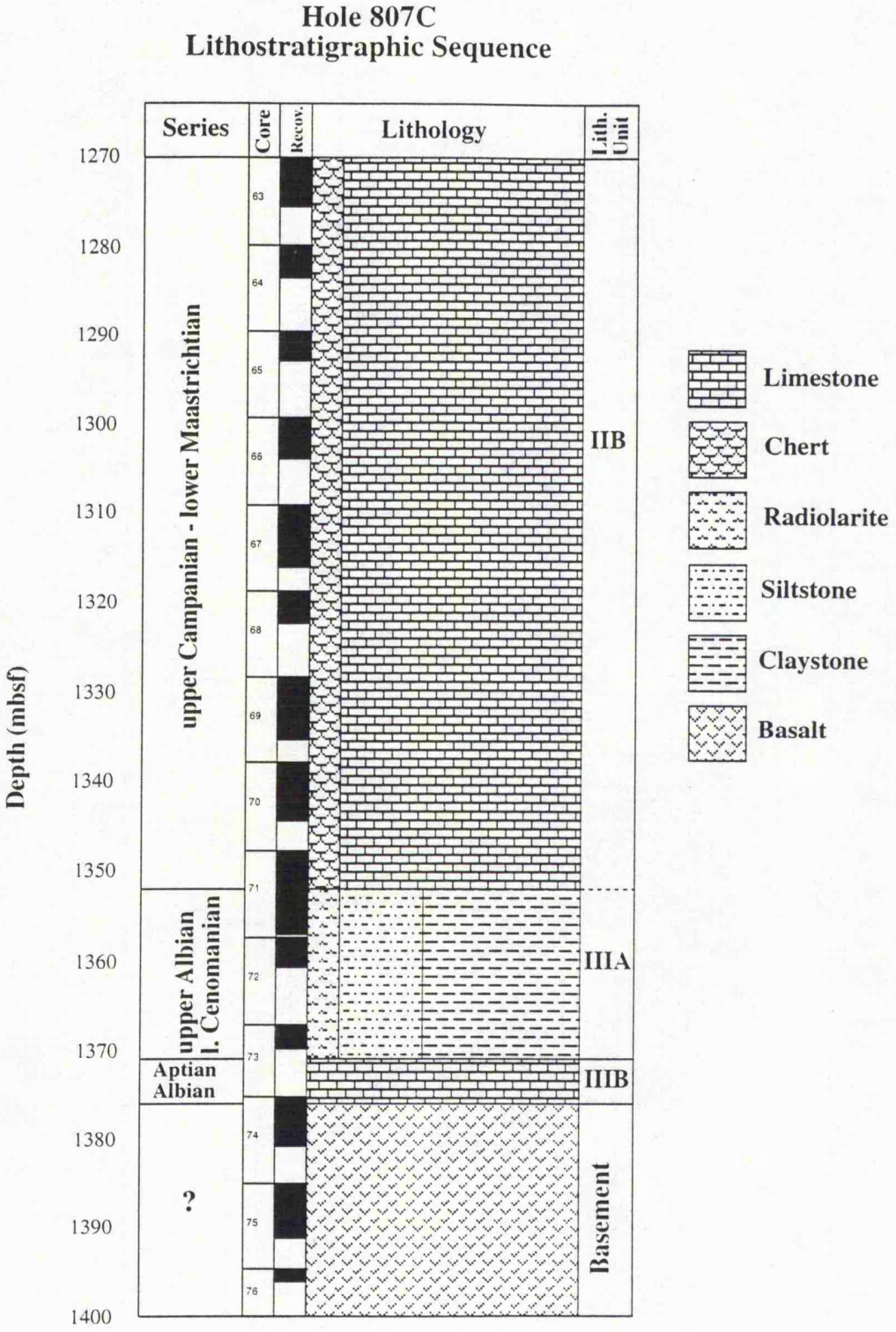


Figure 2.3 - Lithostratigraphic sequence in ODP Hole 807C with core recovery (after Kroenke et al., 1991).

sharp contacts and graded bedding are also present. Wavy contacts are common, producing wavy, lenticular and flaser bedding. The subunit IIIB is composed of bioturbated limestone. It is recovered from 1369.7 - 1375.7 mbsf, but a distinct decrease in drilling rates from 1379.7 mbsf suggests that the latter depth is the contact between limestone and basalt. The limestone is generally grey to light grey with some verticals and fine anostomosing fractures. Microfaults are also abundant.

Volcanic basement starts at 1379.7 mbsf. The basement rocks are predominantly tholeiitic basalts (Albian-Aptian ? in age) consisting of both pillow and thin massive flows. Remarkably, in view of their age, most of rocks are altered only slightly and alteration rarely exceeds a moderate level. The results of a X-Ray Fluorescence (XRF) spectrometer for three samples between 1379.7 and 1400 mbsf can be seen in Table 2.2.

Table 2.2

Core Section Interval (cm) Subunit	**74R - 1 124 - 127 IV**	**75R - 2 61 - 64 IV**	**76R - 1 69 - 71 IV**
Major elements (wt %):			
SiO_2	48.65	48.17	49.06
TiO_2	1.66	1.59	1.61
Al_2O_3	14.73	14.07	14.35
[a]Fe_2O_3	12.54	13.51	13.33
MnO	0.18	0.20	0.18
MgO	6.05	7.26	6.42
CaO	12.43	11.77	11.94
Na_2O	2.45	2.32	2.43
K_2O	0.41	0.35	0.12
P_2O_5	0.14	0.14	0.13
Total	99.91	99.99	99.78
LOI	0.66	0.62	0.20
[b]CIPW norms:			
Quartz			
Orthoclase	2.42	2.07	0.71
Albite	20.73	19.63	20.56
Anorthite	27.99	26.95	27.90
Diopside	27.08	25.16	25.21
Hypersthene	8.65	10.27	15.76
Olivine	5.63	8.45	2.60
Magnetite	2.16	2.33	2.30
Ilmenite	3.15	3.02	3.06
Apatite	0.32	0.32	0.30

[a] All Fe expressed as Fe_2O_3, [b] Fe_2O_3/FeO set at 0.15 and Mg number=(atomic)100 $Mg/Mg + Fe^{2+}$

Table 2.2 - XRF analysis of 3 samples of basalt of ODP Hole 807C (after Berger et al., 1993).

2.2.1.3 - Downhole logging operations

Upon completion of coring at ODP Hole 807C a complete suit of three Schlumberger log ging tools were run for the whole interval. These consisted of the Seismic Stratigraphic combination including NGT/LSS/DITE/TLT; the Lithoporosity combination modified with the inclusion of the AACT instead of CNT-G and also NGT and HLDT; and the Formation MicroScanner including NGT/FMS/GPIT.

The first two runs (the Seismic Stratigraphic combination and the modified Lithoporosity combination) were completed successfully from the base to the top within the interval between 1270 - 1400 mbsf. The third run, the FMS string, which was used with the objective to obtain detailed information about the structure of the sequence, was also logged successfully in the interval. A complete summary of the actual logging operations on ODP Hole 807C is shown in Table 2.3.

Table 2.3

Local day	Local Time	Cumulative hours	Depth (mbrf)	Comments
3/21/90	19:30	0		Start rig up
3/21/90	20:54	1.4		RIH with geophysical tool string
3/21/90	21:59	2.5		At mud line
3/21/90	22:13	2.7	330.4	Start down log: heave compensator on
3/21/90	23:36	4.1	1492.9	Stop downlog
3/21/90	23:39	4.1	1528.3	At bottom of hole; 0.6 m of fill
3/21/90	23:39	4.1	1528.3	start up log ; NGT/LSS/DITE/TLT up at 900 ft/hr
3/22/90	3:58	8.5	348.7	Tool in casing
3/22/90	5:25	9.9		Tool string on deck
3/22/90	6:19	10.8		RIH with NGT/AACT/HLDT
3/22/90	7:53	12.4	1525.8	O n bottom with NGT/AACT/HLDT
3/22/90	8:11	12.7	1467.9	Start up log - 600 ft/hr heave compensator on caliper jammed by debris in lower section. Stop log
3/22/90	8:16	12.8	1527.0	On bottom again Start main up log - 600 ft/hr
3/22/90	14:25	18.9	348.7	Tool string in casing
3/22/90	16:44	21.2		RIH with NGT/FMS/GPIT/TLT
3/22/90	18:06	22.6	0	Passing mud line; pause for TLT tie point
3/22/90	19:14	23.7	1509.1	Set down on bridge Start main up log - 900 ft/hr heave compensator on
3/22/90	21:23	25.9	947.3	Stop main up log hole above too wide for FMS cal going down for repeat
3/22/90	21:50	26.3	1507.8	Start repeat up log from same bridge - 900 ft/hr
3/22/90	23:11	27.7	1095.5	Stop repeat up log
3/23/90	1:20	29.8		NGT/FMS/GPIT/TLT at well head
3/23/90	2:09	30.7		Rigged down from logging runs on ODP Hole 807C

Table 2.3 - Summary of logging operations at ODP Hole 807C (Kroenke et al., 1991).

2.2.1.4 - Log quality

The quality of the log curves can be said to be reasonably good based on the caliper log (Figure 2.4). Tool sticking and borehole washouts are common problems in ODP which typically compromise log quality but these were not apparent in the interval between 1270 - 1400 mbsf in the ODP Hole 807C.

The FMS images appear damaged at intervals where the caliper shows values over 38.10 cm, which is the maximum extent of its spring loaded arms. Sections in the logs affected by washouts may be recognised by intervals in which the FMS tool string twisted rapidly in the hole.

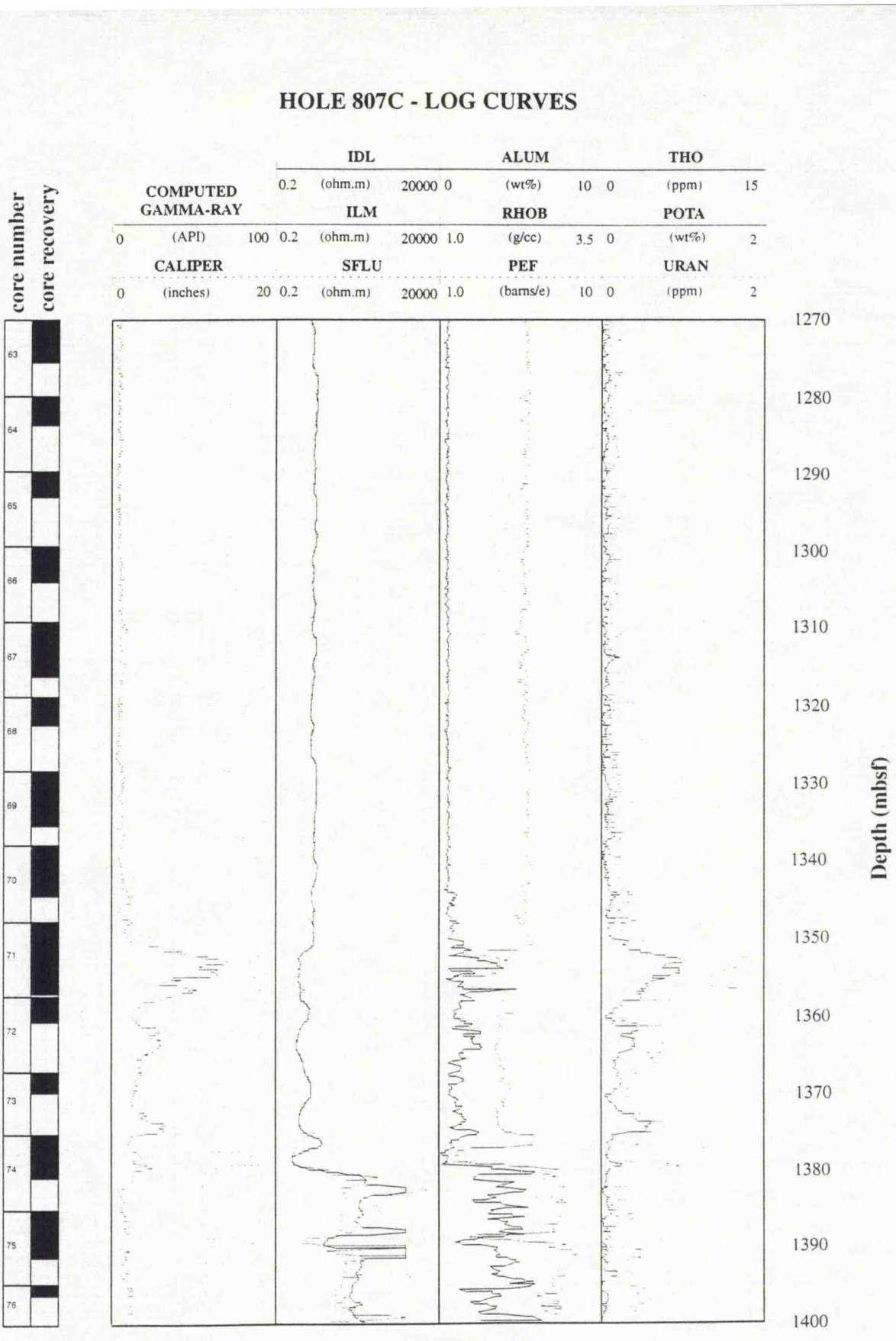


Figure 2.4 - Log curves for ODP Hole 807C.

Depth correlation is very good between all of the logging runs and the lithostratigraphy boundaries obtained from the cores. The claystone/siltstone and limestone boundary at 1351 mbsf appears clearly in the logs as a reduction in density, resistivity and transit time measurements. Local maxima in the profiles are interpreted as limestone stringers within predominantly silt and claystone units.

The boundary of the sediments with basaltic basement is very well defined at 1379.7 mbsf with an increase in density and resistivity and a decrease in transit time.

2.2.1.5 - Log characteristics

Considering the lithological variation observed between 1270 - 1400 mbsf in the ODP Hole 807C, the logging data provide a great deal of information about not only the major changes in lithology but also the more subtle physical and chemical characteristics and cyclicity present in the formations. The high sampling resolution of logging measurements (every 15 cm) allow for the delineation of heterogeneities in the formations which would be more difficult to detect using discrete sampling techniques.

As shown before, the lithologic units observed in the interval between 1270 - 1400 mbsf can be easily identified through the logging data. Decrease in resistivity, bulk density and aluminium concentration and an increase in gamma-ray, transit time, K and Th contents and photoelectric effect show the transition between the limestone and the siltstone/claystone units at 1351 mbsf. At 1379.7 mbsf a decrease in gamma-ray, photoelectric effect and Th and K contents and an increase in resistivity and transit time values show the transition between the sediments and the basement. Any subtle variation in log values within the interval can be explained through the presence of heterogeneities in the formations and their confirmation is attempted through the use of the techniques implemented in chapters 3 and 4.

2.2.2 - ODP Hole 878A - ODP Leg 144: Northwest Pacific Atolls and Guyots

2.2.2.1 - Introduction

The ocean floor of the western Pacific Ocean is covered by numerous scattered seamounts and atolls that cluster together to form large complexes or are roughly organised along a preferential direction to form a chain. Most of the seamounts are guyots that have a flat top and are considered old, drowned atolls (Premoli Silva et al., 1993).

Several authors explain such alignments, like the Hawaiian Chain, as having originated by the progression of a hotspot. Another alignment is illustrated by the

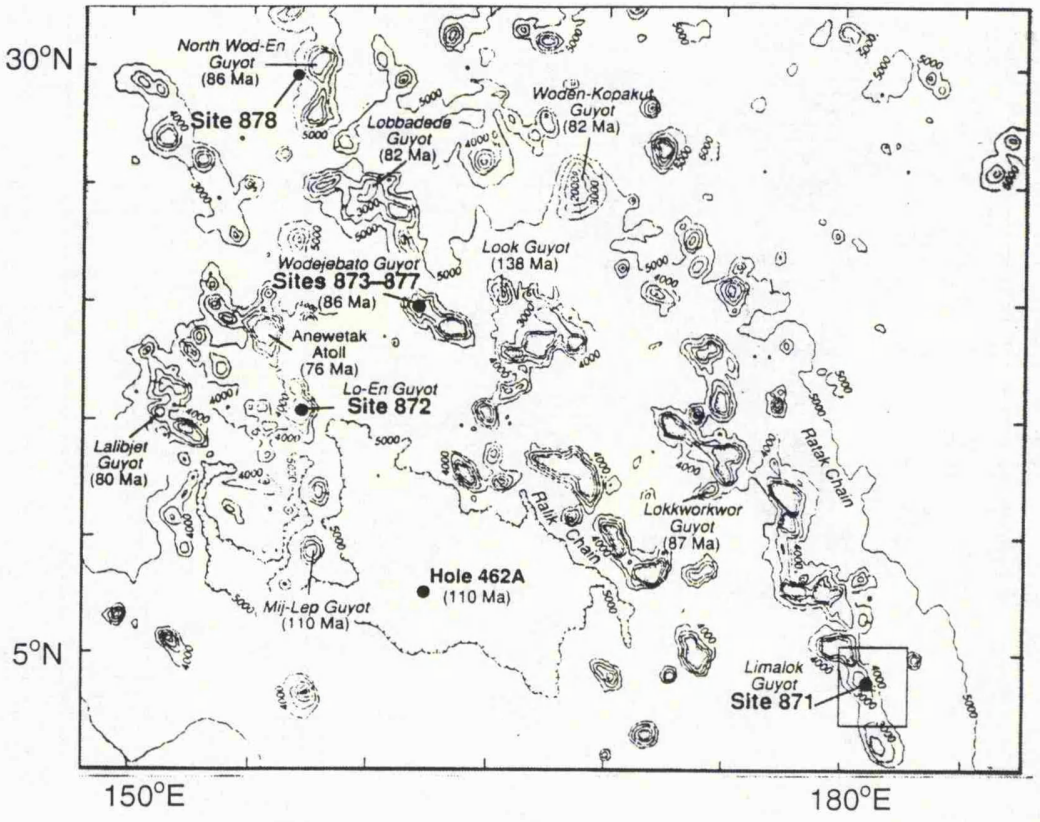


Figure 2.5 - Bathymetry around the Marshall Islands. Contour interval is 1000 m. The locations of ODP sites 871-878 are shown (after Premoli Silva et al., 1993).

Mid-Pacific Mountains, which exhibit a roughly east-west orientation. Other guyots or cluster of guyots may not show any apparent preferred orientation or they may be isolated features. This is the case of MIT (Massachusetts Institute of Technology) Guyot which is the objective of the study by ODP Hole 878A.

The MIT Guyot is an isolated feature in the 18°-28° N guyot band. This guyot is interpreted to be a drowned atoll; on the basis of dredge hauls and geophysical data, the seismic facies were interpreted as a perimeter reef encircling a thick sequence of lagoonal sediments. At the close of the Albian, many of these Barremian-Aptian carbonate platforms and reefs, including the MIT Guyot, as well the younger ones constructed in the Albian, may have been uplifted above sea level (Winterer et al., 1992).

Drilling plans for MIT Guyot included one lagoonal site, 878. ODP Hole 878A is located at 27°19.143' N, 151°53.028' E in a water depth of 1323 meters, on the north-eastern part of MIT Guyot near its southern edge (Figure 2.5). In fact three holes were drilled in site 878. ODP Hole 878A was a multiple re-entry hole and holes 878B and 878C were single-core holes, used for the purpose of recovering the surficial hard ground and pelagic sediments overlying the carbonate platform. During coring at ODP Hole 878A a thin sequence of pelagic sediments, two carbonate platform sequences divided by a volcanic limestone breccia and a portion of the igneous edifice were recovered.

2.2.2.2 - Lithostratigraphic sequence

The complete stratigraphic sequence for ODP Hole 878A can be observed in Table 2.4. Due to the excellent core recovery (± 95%) and the well conditions, the interval between 515 - 600 mbsf was selected to be used in this work. As can be seen from Table 2.4, it comprises the subunit IVC which consists of a breccia predominantly siliciclastic at the top and calcareous at the bottom of the interval (Premoli Silva et al., 1993).

The overall variation in the polymictic breccia within the subunit IVC is as follows: volcanic clasts are dominant in the upper half and decrease in abundance towards the base of the subunit. Some alternation occurs between volcanic-rich and poor horizons throughout the subunit, but carbonate is dominant in the matrix and clasts below 575 mbsf. Premoli Silva et al. (1993) shows that the analysis of bulk carbonate content demonstrates the increase of carbonate in the matrix from the top to the bottom of the unit.

Table 2.4

Unit/Subunit	Depth (mbsf)	Age	Description
Subunit IA	0.0 - 0.95	Pliocene-Pleistocene	Pelitic sediment with manganese nodules
Subunit IB	0.95 - 3.20	Miocene	Pelitic sediment with manganese nodules and chalk fragments
Subunit IIA	3.20 - 67.26	Albian	Wackestone and peloid packstone with recristallized fragments of gastropods
Subunit IIB	67.20 - 86.20	Albian	Wackestone and Packstone
Subunit IIC	86.20 - 202.20	Albian/later Aptian	Wackeston and Packstone with gastropods
Subunit IID	202.20 - 236.07	Albian/later Aptian	Mudstone. Wackestone with gastropods
Subunit IIIA	236.07 - 312.65	Albian/later Aptian	Fine to medium grained grainstone with intervals of rudstone
Subunit IIIB	312.65 - 389.80	Albian/later Aptian	Fine to medium grainstone with *Orbitolina*
Subunit IIIC	389.80 - 399.74	later Aptian	Well lithified foraminifer wackestone and mudstone
Subunit IVA	399.74 - 406.10	later Aptian	Bluish gray clay
Subunit IVB	406.10 - 514.74	later Aptian	Vesicular breccia with volcanic and limestone clasts in a predominantly limestone matrix
Subunit IVC	514.74 - 604.30	Aptian	Vesicular breccia with alternating carbonate-rich and volcanic-rich intervals
Subunit VA	604.03 - 703.00	early Aptian	Peloid foraminifer wackestone. packstone and grainstone
Subunit VB	703.00 - 722.54	early Aptian	Grainstone and rudstone
Subunit VI	722.54 - 907.80	?	Basaltic flows

Table 2.4 - Lithostratigraphic sequence for ODP Hole 878A (after Premoli Silva et al., 1993).

The volcanic component throughout the interval is dominated by basalt clasts. The vesicles in these clasts are most spherical and rarely tubular (Premoli Silva et al., 1993). Some of the clasts are completely altered to clay. Limestone clasts are also present. Several different lithologies are stained, including wackestone, peloid packstone and grainstone. Sometimes breccia matrix fills some voids both in basalt and limestone clasts.

Another characteristic feature of this interval is the steep inclined bedding angle (28° to 62°). The bedding is recognisable as an alter nation of coarser and finer grained intervals on a 2-10 cm scale. Some cross-bedding is apparent at the bottom. Rapid change in the dip orientation within a continuous cylinder and the presence of curved, convoluted and truncated beds strongly suggest that the steep bedding is the result of slumping.

2.2.2.3 - Downhole logging operation

The logging operations in ODP Hole 878A were seriously affected by hole collapse, obstructions and long intervals with washouts. ODP Hole 878A was primarily cored to 910 mbsf of total depth but only 264.94 meters were recovered, leading to an average core recovery of 25.5%. The core recovery was only good within the polymictic breccia (unit IVC), reaching nearly 100% in this interval. It allows for a good sedimentary and structural description of the whole interval and was one of the reasons of the choice of this section for the study.

Upon completion of the drilling, a logging program including the Seismic Stratigraphic combination, the Geochemical tool string, the Lithoporosity combination and the Formation MicroScanner (FMS), was undertaken. The first log to be deployed was the FMS because it had the highest priority for that portion of the hole (Premoli Silva et al., 1993). Even though affected by borehole conditions, a good log was obtained for the sedimentary section, especially for the interval between 515 - 600 mbsf. The next logging attempt was with the Seismic Stratigraphic combination. A good log was recorded all the way to the drill pipe (located at 17 mbsf). The quality of the sonic log, as did happen with the FMS, was degraded by the overgauge hole (> 40 cm) for much of the interval. The ensuing logs - the Geochemical and Lithoporosity combinations - were stopped by a bridge that had formed at the top of the basalt. However, like the FMS, they produced generally good logs throughout the entire sediment section. The summary of the well log data obtained in ODP Hole 878A can be seen in Table 2.5.

2.2.2.4 - Log quality

The quality of the downhole measurements in ODP Hole 878A is reasonably good. Only few intervals appear with large borehole diameter, which affected some of the log curves. In a general sense the logs are valuable for interpreting the sedimentary succession and for identifying possible heterogeneities within the polymictic breccia. The Geochemical, FMS and Lithoporosity combinations were calibrated in terms of depth correlation to the Seismic Stratigraphic run by using the natural gamma-ray tool (NGT) on each string.

2.2.2.5 - Log characteristics

The polymictic breccia sequence (unit IVC) is characterised by high resistivity (6 Ωm increasing downhole to 12 Ωm), increasing density towards the bottom of the sequence (from 2.3 g/cc to 2.5 g/cc), stable natural gamma-ray radiation (20 - 25 API)

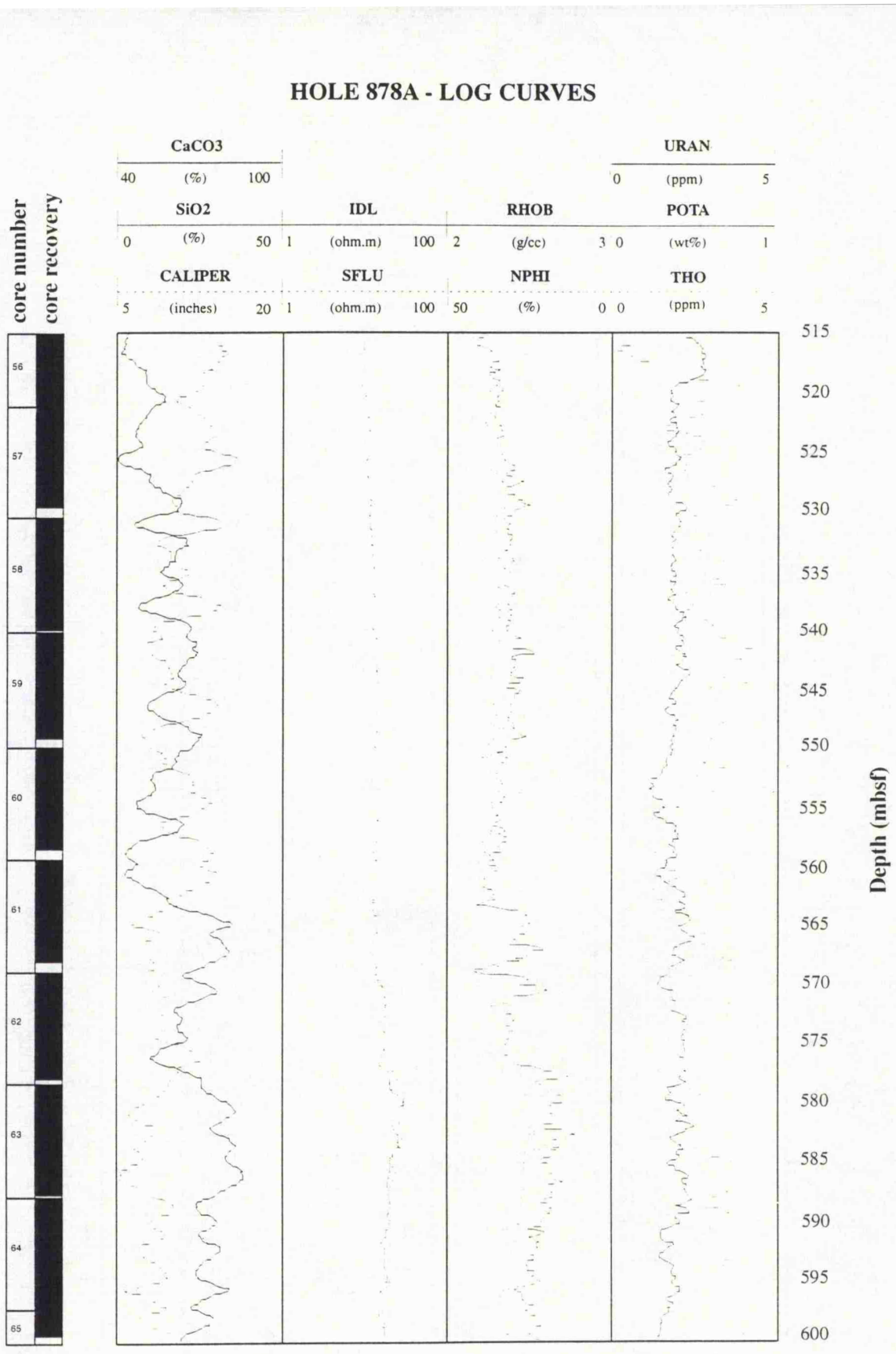


Figure 2.6 - Some of the log curves for ODP Hole 878A.

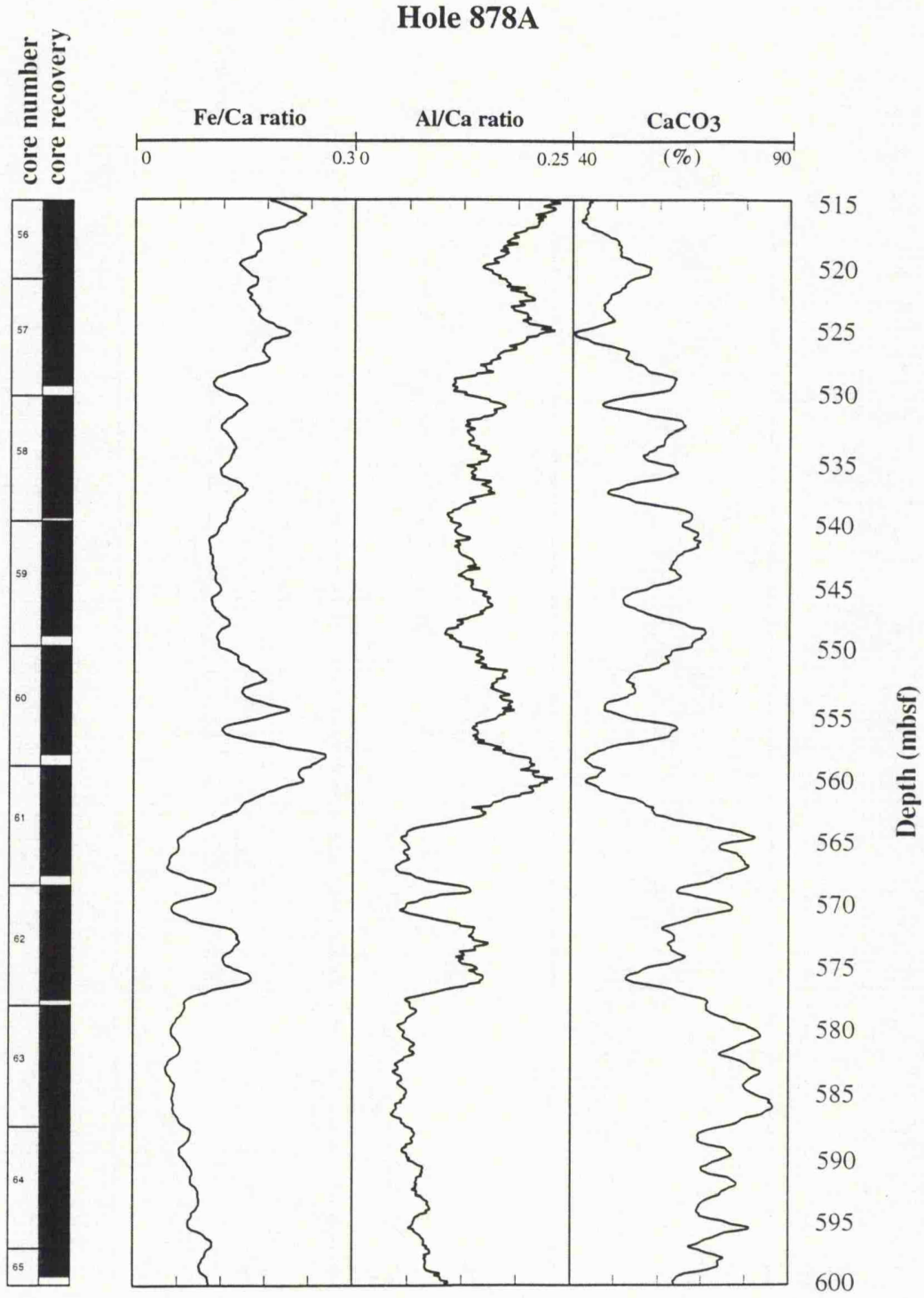


Figure 2.7 - Fe/Ca, Al/Ca ratios and $CaCO_3$ oxide for ODP Hole 878A.

from abundant uranium (1.5 - 2 ppm) and thorium (2 ppm) and high abundance of Fe and Al relative to Ca. This interval also displays progressive downhole increase in Ca content and sonic velocity (Figure 2.6). The Fe/Ca ratio within the breccia interval provides a measure of the relative abundance of volcaniclastic versus carbonate components (Figure 2.7).

The steeply dipping crossbedding and possible slump deposits at the bottom of the interval can be checked through the FMS imagery.

Table 2.5

Log Type	Depth (mbsf)
Natural gamma-ray	17 - 740
Sonic	17 - 740
Neutron porosity (AmBe neutron source)	17 - 740
Lithodensity (Ce gamma-ray source)	17 - 740
Resistivity	17 - 740
* Aluminium clay tool (Ca neutron source)	0 - 740
* Gamma-ray spectrometry	0 - 740

* These logs were recorded in pipe from 0 - 17 mbsf

Table 2.5 - Log summary for ODP Hole 878A (Premoli Silva et al., 1993).

2.2.3 - Oilfield Holes A and B, S.E. Brazil

Oilfield Holes A and B were drilled in an oilfield on SE Brazil. They were logged with gamma-ray (GRAY), porosity (NPHI), density (RHOB), sonic (DT), and resistivity (ILD and ILM) tools. Intervals with similar characteristics were selected in order to allow a reasonable characterisation between both holes.

The stratigraphic sequence for both holes consists of a shale/claystone sequence in which sandstone reservoirs are included. In Hole A, the sand reservoir is a consistent 25 metres thick layer with high resistivity ($\approx$ 100Ωm) and low density (Figure 2.8). In Hole B there is an interbedded sequence of sandstone and shale with the sand layers ranging from 2 to 5 metres thick. Resistivity reaches a maximum of 10Ωm with density values not less than 2.1 g/cc (Figure 2.9). Porosity logs also show differences between the two holes. In Hole A there is a constant value of 30% for all the interval selected for study while Hole B shows more variation in the values, with the sand layers presenting values up to 37%. Gamma-ray variations are typical for shale and sand sequences with values of 80 API and 120 API in shale given probably by the amount of clay in this sequence.

The idea is to use this dataset to provide hole-to-hole lithofacies characterisation using different statistical techniques and then compare the results with other novice techniques such as Neural Networks. Generation of petrophysical parameters will also be considered in this work. Core measurements in Hole A give support not only for a better lithofacies characterisation but also to provide physical

property measurements which will be used as a training basis in the prediction of petrophysical parameters in Hole B using Neural Network.

Hole A
Log Curves

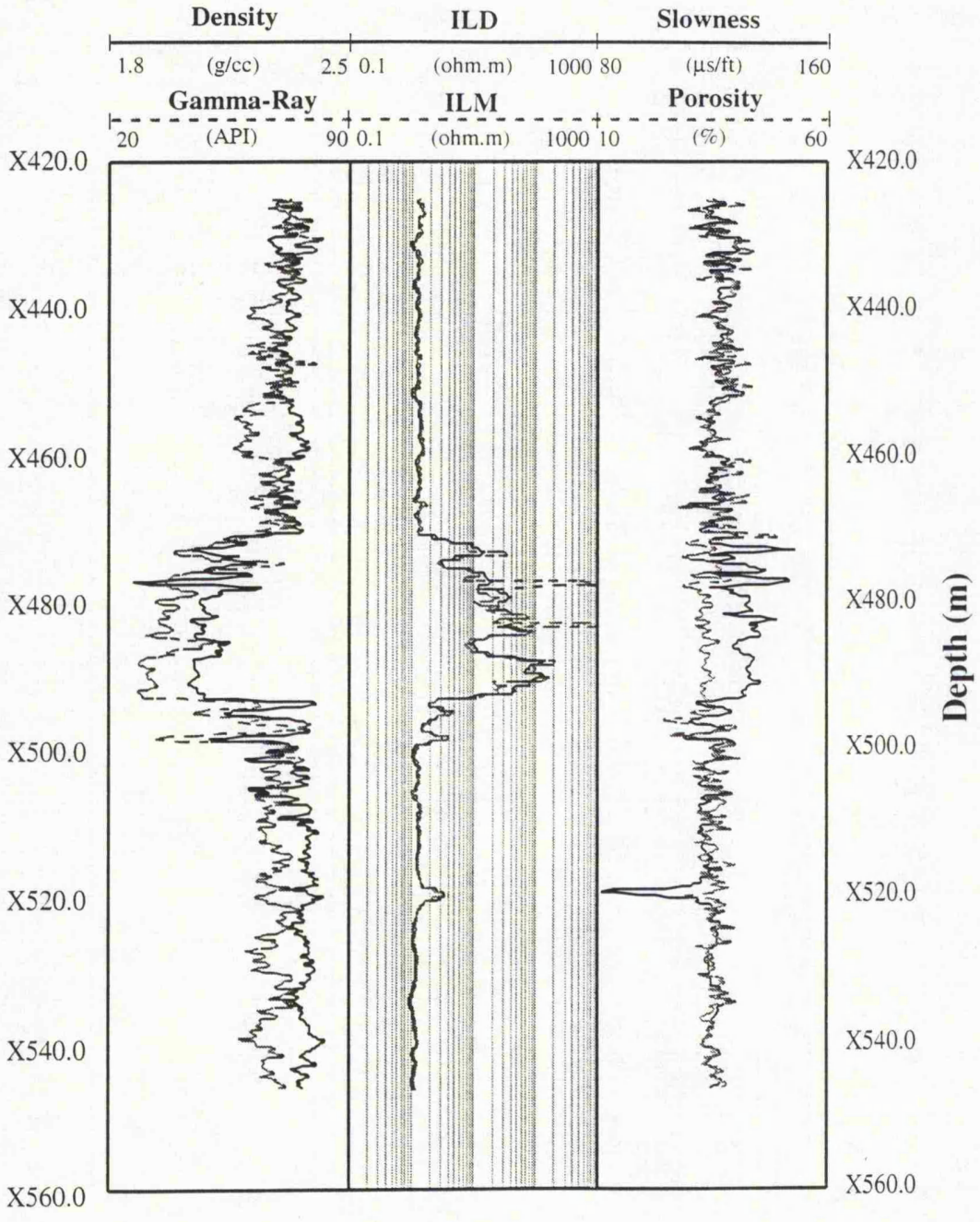


Figure 2.8 - Log curves for Hole A.

Hole B
Log Curves

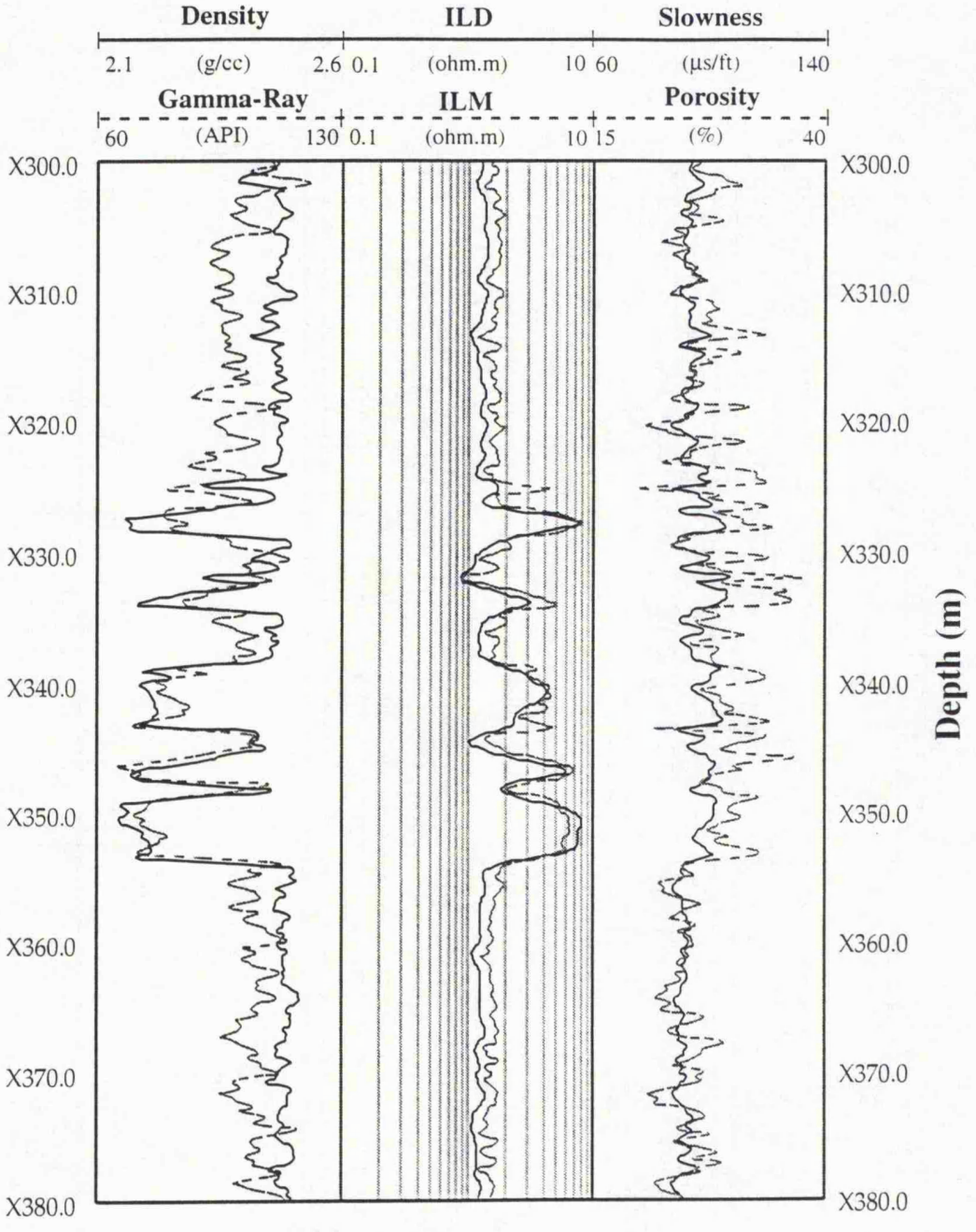


Figure 2.9 - Log curves for Hole B.

Chapter 3

MULTIVARIATE STATISTICAL ANALYSIS OF WELL LOG DATA

3.1 - Introduction

Various techniques applied to lithofacies discrimination using well log data have been described in the literature (Davis, 1986). Most of these techniques are based on statistical analysis of trends within the dataset. Signal theory, pattern recognition and other methods of numerical analysis have been used with relative success.

The scope of a geological investigation may include not only the correct identification of lithofacies, but can also involve other aspects such as diagenetic changes, variation in composition and petrophysical characteristics (e.g. differences in porosity, pore geometry and cementation) (Doveton, 1994). Useful interpretations are still performed from visual examinations of log curves and cross-plots with available core integration support. The numerical and multidimensional format of log data, however, requires quantitative analysis of the data set where facies discrimination takes subtle and complex forms.

Simultaneous analysis of the log curves can show characters hidden in their mutual variation. Therefore, methods which can handle the multidimensional structure of the data are needed. In this work, a selection of multivariate statistical techniques are applied to well log data as aids in the interpretation of geological patterns.

3.2 - Multivariate Statistical Techniques

Multivariate statistical techniques have been used in data analysis to identify or classify a sample, or a group of samples. They include such recognised techniques as Cluster Analysis and Discriminant Analysis with the related techniques of Principal Component Analysis (PCA) and Factor analysis commonly applied to data at a pre-processing stage. It is an effective approach to data interpretation in geology and geophysics. For this reason, and also because of the confusion that has arisen over the use of some of these techniques, it is worth describing some of the concepts and terminology involved at this early stage.

Multivariate statistical techniques are also known as pattern recognition techniques. A pattern consists of a collection of measurements characterising a sample which are considered as an entity for the purpose of subsequent analysis (Sheppard, 1986). The classification imposed by the pattern recognition allocates

entities to initially undefined classes so that the individuals in a class are in some sense close to one another. This method is internally based and does not depend on *a priori* knowledge about the relations between these entities or samples. The samples are free to enter any class such that they emerge in the process of classification. The result is that the entities should be placed into approximately homogeneous groups. These entities may be individual log curves, a number of log curves making up a sample, or classes of samples. Therefore the classes of one classification system may be the entities of another.

Two aspects of the pattern recognition problem are known. The first one is the development of a decision rule whose function is to divide multidimensional measurement space into different decision regions each corresponding to one class. The second one is the implementation of that rule (Young and Calvert, 1974). These two aspects may, for instance in the case of Cluster Analysis, occur simultaneously. This enables the technique to work without *a priori* knowledge, although clearly the existence of a training set allows a user to evaluate the efficiency of the algorithm. The alternative is that it occurs sequentially and therefore can be based on *a priori* knowledge as with Discriminant Analysis.

Sheppard (1986) shows that the successful application of a selected technique is dependent upon careful formulation of the decision rule by the user and on understanding of the assumptions and limitations involved. Without this a successful interpretation of the results may not be made. Careful consideration must be taken in the preparation of the dataset and in the selection and implementation of the methods to be used. The techniques have their origins in many fields of study with a significant amount of literature in diverse areas such as: life science (Simpson, 1961), medical science - including psychiatry (Tryon and Bayley, 1970) and earth science (Le Maitre, 1982 and Davis, 1986). In the case of the oil industry, the term "electrofacies" has been used to describe lithofacies identified on the basis of logging data (Serra and Abbott, 1982; Delfiner et al., 1987; Bucheb and Evans, 1994 and Doveton, 1994). In this work, the author suggests that the term lithofacies is used in replacement of electrofacies because of the use of core support during the analysis.

One of the aims of this work is to find ways of identifying such lithofacies in different geological environments using well log data. With an increasing number of log curves (e.g. density, resistivity, porosity, chemistry, etc.) the amount of information relating to the formation increases, and a more realistic picture of its nature can be produced. The aim of the application of the multivariate statistical techniques considered here is to search for structure and pattern within these data, a knowledge which may help to better understand the data set and simplify their interpretation. Another aim is the verification of when and where the different patterns observed can be interpreted as heterogeneities within the whole dataset.

Of the techniques available, Principal Component Analysis, Cluster Analysis and Discriminant Analysis were found to be particularly useful in the characterisation of log data. These techniques are briefly described in the following sections and their application to well log data of the ODP are also shown.

3.3 - Principal Component Analysis

Principal Component Analysis (PCA) sets out to clarify the structure of a dataset and can, in turn, allow reduction of the system's dimensionality (in this case the number of log curves) and, hence, simplify the interpretation. It is the most commonly used technique for this purpose and has many additional useful properties. The technique is well established and widely used in many fields (Marriot, 1974 and Gnanadesikan, 1977) including both petrological (Le Maitre, 1982) and logging data studies (Doveton, 1986, 1994). In this work only a brief overview of the technique is provided. For a full discussion of the mathematics involved, the reader must refer to a specialist text such as Johnson and Wichern (1982) or Davis (1986).

In logging analysis a dataset of N samples and M measured log curves (variables) can be plotted as points in a space with mutually orthogonal axes forming a M-dimensional cloud of N data points (Figure 3.1). Principal Component Analysis sets out to describe this cloud using a new set of M uncorrelated variables (eigenvectors or principal components) which correspond to the best fitting axes of the cloud. The new axes provide a different framework of reference which is aligned with the natural axes of the cloud, rather than the original log measurements axes. Each of these eigenvectors (the new axes) represents a linear combination of the original variable, with each original variable contributing with some weight (loading) to each eigenvector. The orientation of the principal components are computed from either the variance-covariance or correlation matrix of the dataset. Because most logs are recorded in radically different units the latter operation is preferable. Given the disparate nature of the logging measurements used in this work, the analyses have all been performed using the correlation matrix with Z-standardised scores.

Doveton (1994) shows that the raw cloud is modelled by a single ellipsoid. The ellipsoid orientation and the directions of axis elongation reflect the relationship between the logs. The axis are the eigenvectors which have magnitudes given by their eigenvalues. For instance, consider a hypothetical cross-plot of Log a and Log b (Figure 3.2). Since the two logs are recorded in radically different units, it is appropriate to rescale them in a standardised form, with zero mean and unit standard deviation). This is achieved by relocating the cloud centroid to a new origin. The standardisation converts the variance-covariance matrix to a matrix of correlation coefficients. The correlation matrix is described geometrically by an ellipse. The

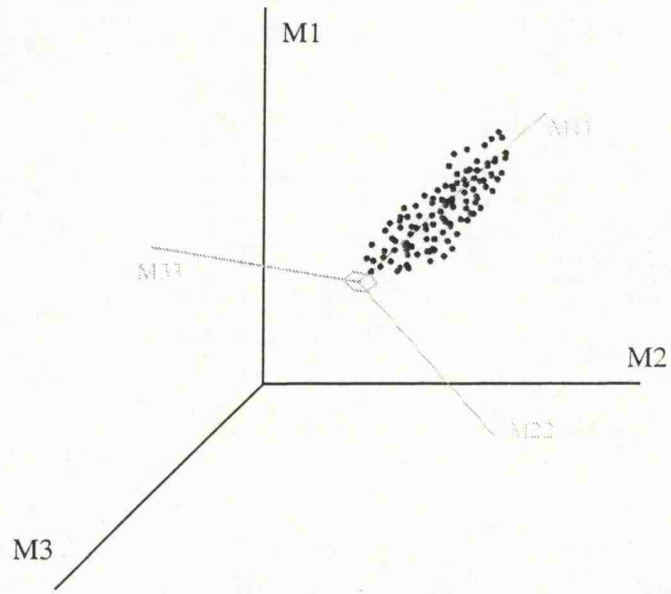


Figure 3.1 - Representation of plotting N samples in a multidimensional space defined by axes M1, M2 and M3, and a new set of M uncorrelated variables M11, M22 and M33 (eigenvectors).

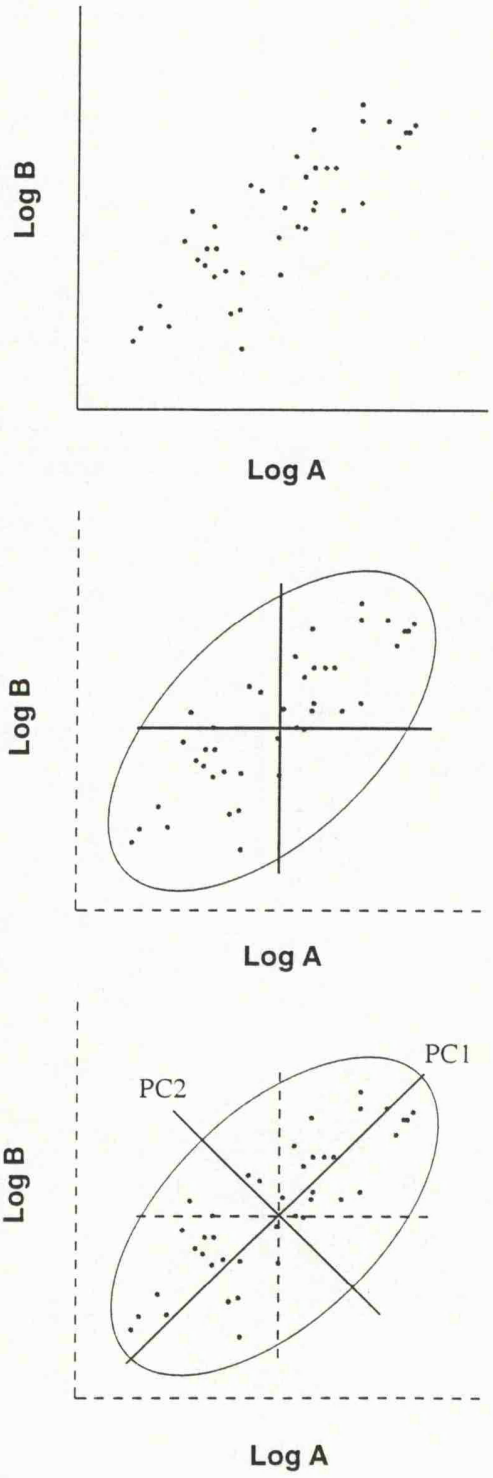


Figure 3.2 - Principal Components I and II of hypothetical data points measured with respect to two log curves presented as dimensionless axes. At the top, Log A-Log B cross-plot; in the middle, the correlation matrix elipse superimposed on standardized Log A-Log B axes; and at the bottom, the location of the pricipal component axes (after Doveton, 1986).

eigenvectors of the correlation matrix locate the major and minor axis of the ellipse or the principal components. The eigenvalues associated with each eigenvectors are the length of each axis. Each eigenvector, extracted from the correlation matrix, not only accounts for the maximum possible amount of residual variance (which has the largest possible eigenvalue) but also preserves the orthogonal relationship between the vectors. Therefore, the first principal component accounts for the maximum variability and the remaining principal components pick up the rest of the variability in ordered allocation.

In the case of a multiple log curve dataset, with all principal components, all the variation present in the initial data set is fully described. In practice, however, many measurements show a significant decrease in the degree of intercorrelation that it is possible to select a number p of eigenvectors (where $p < M$) which describe a substantial portion of the overall variation and use these to represent the dataset (Davis, 1986). This property highlights the amount of information redundancy within the log curves and reflects the dimensionality of the information content as a replacement for the original reference framework. This can be achieved by choosing a cut-off eigenvalue below which the eigenvectors are not considered to be important. In this work this value is usually 1.0 but can be different depending on the importance and how the remaining eigenvectors account for the total variability. The loadings of the p eigenvectors can be used to produce a principal component loading matrix for these p principal components. By considering the loadings on a given component, inferences about the geological processes which that component represents may be made. Using this loading matrix, it is also possible to transform the original data set so that it refers to the principal components only and, thus, produce a set of principal component scores (Pelling, 1992).

The computation of the principal components and their scores is simply a geometrical operation which does not create any new data - it merely provides a new set of variables which are transformations of the original ones and which may help with its interpretation. In this work, the prime aim of the Principal Component Analysis is to evaluate the overall variability and reduce the number of log curves (dimensionality) of the datasets to improve interpretation. The principal component scores can also be used as input data for further processing (such as Cluster and Discriminant Analysis). This operation ensures that the variables are orthogonalised prior to analysis which, since it removes any correlation between variables, should allow more efficient statistical processing. The results of the Principal Component Analysis carried out on ODP well log datasets are presented below.

3.3.1 - Application of Principal Component Analysis to well log data

3.3.1.1 - ODP Hole 807C

The logging dataset used for this analysis consists of 10 log curves as follows: gamma-ray (GRAY); potassium, thorium and uranium concentrations (respectively POTA, THO and URAN); three different resistivity measurements (IDPH, IMPH and SFLU); aluminium content (ALUM); density (RHOB) and photoelectric effect (PEF). The data contain 853 samples and were acquired between 1270 - 1400 mbsf in ODP Hole 807C. Transit time log from the LSS-SDT tool was also available, however, its quality was considered to be insufficient for inclusion in the analysis because of the acquisition of bad data during logging operations.

The physical/chemical properties obtained in ODP Hole 807C show considerable variation, reflecting the different lithofacies present in the interval (Table 3.1). Gamma-ray and Th, K and U concentrations show their highest values for the interval between 1351 - 1379 mbsf due to the presence of clay intervals in the siltstone/claystone sequence (subunit IIIA). The other log curves also show variations in their measurements throughout the interval according to the different lithofacies present (Figure 2.4). The low density values are associated with the siltstone/claystone interval. The presence of a very low density peak is observed in the volcanic sequence and probably represents a fractured and/or altered interval in this unit, which also presents the maximum density values (3.0 g/cc). The PEF values show almost the same variation as the density ones, and again, the maximum values (5.0 barns/electron) are associated with the volcanic sequence. The aluminium content increases downhole from 0.1% in the carbonate sequence to 6.0% in the volcanic sequence.

Table 3.1

Log curves	GRAY	POTA	THO	URAN	IDPH	ALUM	RHOB	PEF	IMPH	SFLU
N. of Samples	853	853	853	853	853	853	853	853	853	853
Standard deviation	10.72	0.31	1.26	0.31	293.76	1.26	0.30	1.06	200.11	395.61
Minimum	2.58	-0.05	-0.31	-0.29	0.64	0.09	1.23	1.36	0.536	0.6559
Maximum	70.82	1.87	7.73	1.95	1950.0	6.27	3.10	6.11	1950.0	9688.9
Range	68.23	1.92	8.04	2.24	1949.4	6.17	1.87	4.74	1949.4	9688.2

Table 3.1 - Summary statistics for the dataset used in ODP Hole 807C.

General relationships in the dataset are summarised in the linear correlation matrix in Table 3.2. As expected a strong positive correlation ($r \cong 0.95$), exists between the GRAY, POTA and THO as well as with URAN ($r = 0.719$) and also between RHOB and PEF ($r = 0.823$). The latter correlation reflects the high PEF values of denser materials (Ellis, 1987), while the three first correlations reflect the high gamma-ray values in the claystone/siltstone sequence due to the high clay content. Strong negative correlations are observed between both RHOB and PEF and GRAY, POTA, THO and URAN. They reflect the low gamma-ray radiation present in the carbonate and volcanic sequences. The highest correlations for ALUM are with RHOB and all resistivity measurements (IDPH, IMPH and SFLU), again reflecting the high ALUM concentration in the volcanic sequence. Most log curves do not show strong correlations with each other. This is the case for IDPH, IMPH and SFLU. Despite the highest resistivity values observed in the volcanic sequence, the correlations between these log curves with RHOB and PEF are not very strong, rarely exceeding $r = 0.20$.

Table 3.2

	GRAY	POTA	THO	URAN	IDPH	IMPH	SFLU	ALUM	RHOB	PEF
GRAY	1.000	0.954	0.945	0.718	-0.050	-0.037	-0.040	0.246	-0.522	-0.602
POTA	-	1.000	0.867	0.574	-0.017	-0.014	-0.037	0.337	-0.413	-0.512
THO	-	-	1.000	0.544	-0.083	-0.065	-0.041	0.148	-0.553	-0.595
URAN	-	-	-	1.000	-0.026	-0.009	-0.024	0.131	-0.436	-0.511
IDPH	-	-	-	-	1.000	0.239	0.316	0.4084	0.341	0.168
IMPH	-	-	-	-	-	1.000	0.092	0.247	0.221	0.094
SFLU	-	-	-	-	-	-	1.000	0.301	0.220	0.126
ALUM	-	-	-	-	-	-	-	1.000	0.405	0.118
RHOB	-	-	-	-	-	-	-	-	1.000	0.822
PEF	-	-	-	-	-	-	-	-	-	1.000

Table 3.2 - Correlation matrix for the dataset used in ODP Hole 807C.

A Principal Component Analysis was carried out on the data using the correlation matrix, which is equivalent to performing the analysis on standardised log curves with zero mean and unit standard deviation. The resultant eigenvectors and their corresponding eigenvalues as well as the percentage contributions for each of the eigenvectors are given in Table 3.3. The first two eigenvectors, which show eigenvalues greater than 1.0, correspond to more than 64% of the total system variability. A principal component loading matrix was calculated using these two eigenvectors (Table 3.4).

Table 3.3

	eigenvalues	% eigenvalues	cum. % eigenvalues
e-vector 1	4.2809	42.80	42.80
e-vector 2	2.1677	21.67	64.48
e-vector 3	0.9138	9.13	73.62
e-vector 4	0.8534	8.53	82.15
e-vector 5	0.6362	6.36	88.52
e-vector 6	0.5209	5.20	93.72
e-vector 7	0.4054	4.05	97.78
e-vector 8	0.1164	1.16	98.94
e-vector 9	0.1052	1.05	100.00
e-vector 10	0.0000	0.00	100.00

Table 3.3 - Eigenvalues and their percentage contributions for each eigenvector in ODP Hole 807C.

Table 3.4

Principal component	**I**	**II**
GRAY	0.9517	0.2380
POTA	0.8733	0.3104
THO	0.9079	0.1504
URAN	0.7393	0.1458
IDPH	-0.1792	0.6829
IMPH	-0.1194	0.4536
SFLU	-0.1329	0.5310
ALUM	0.0687	0.8443
RHOB	-0.7387	0.5002
PEF	-0.7919	0.2315

Table 3.4 - Principal component loading matrix.

The loadings on each of the two components are graphically displayed in Figure 3.3. The principal component I, which accounts for 42.8% of the total system variability, shows strong positive correlations with GRAY, THO, URAN and POTA (r = 0.90), accompanied by strong negative correlation (r = 0.75) with PEF and RHOB. This suggests that this component represents the relation between the high amount of gamma-ray radiation emitted by the rocks and their low densities, characteristic of the claystone/siltstone interval between 1351 - 1369 mbsf.

The principal component II is dominated by a strong positive correlation with ALUM and less strong correlations, with IDPH, IMPH, SFLU and RHOB. All other log curves have low positive correlation. This component accounts for 21.67% of the total system variability and contains, essentially, information concerning the amount

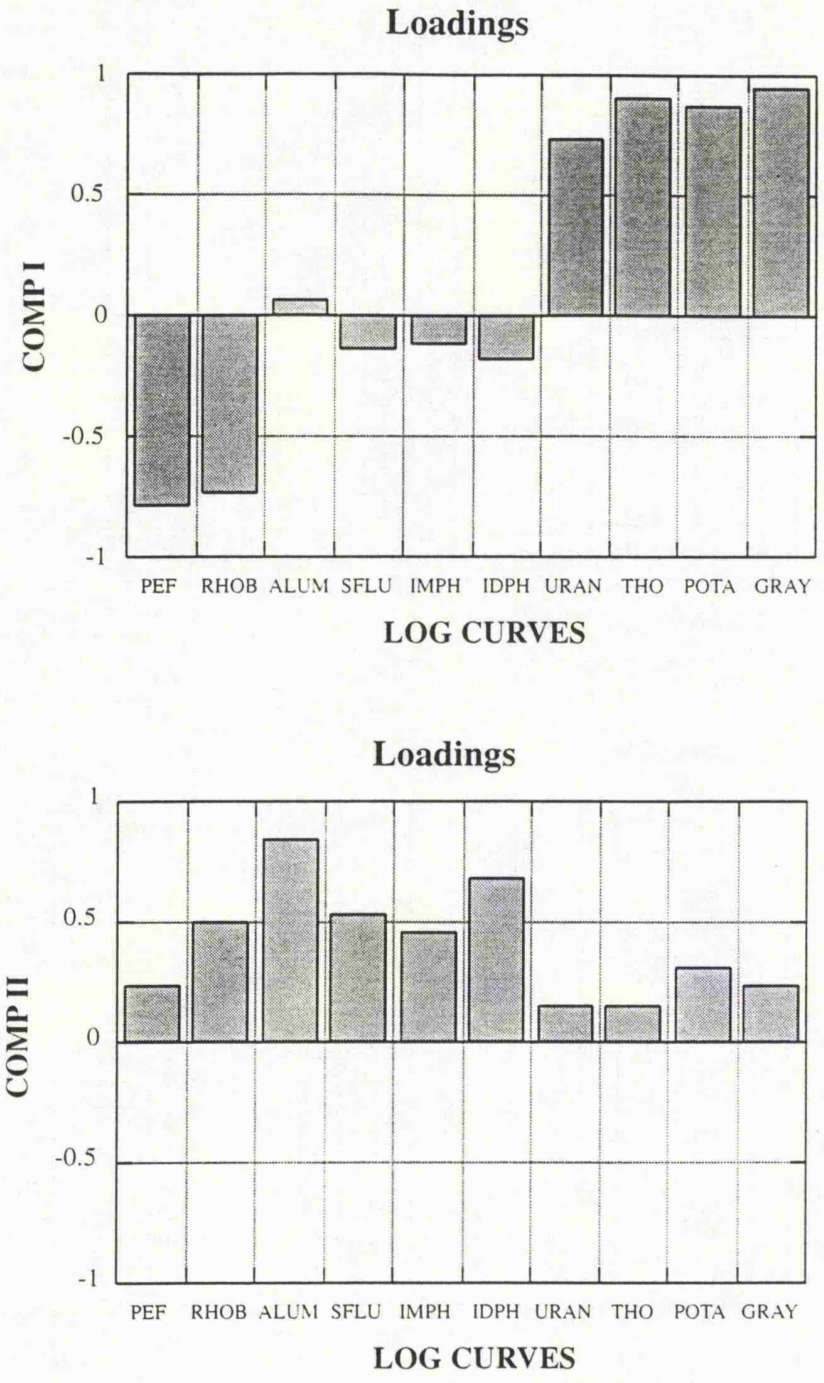


Figure 3.3 - Principal Component loadings for each of the log curves in ODP Hole 807C.

of Al present in the rocks. As shown previously, this is characteristic of the volcanic sequence, which also shows high density and high resistivity values.

The principal component scores along the interval are seen in Figure 3.4. The correspondence between the principal components I and II with the different lithofacies is observed. The principal component scores I show strong positive values between 1351 - 1379 mbsf related to the increase in clay content in that section. Negative values in the upper carbonate sequence and even more negative values within the volcanic sequence are related to low gamma-ray values in these two units. A negative peak is observed in scores I between 1375-1377 mbsf and it is possibly the evidence of the presence of a carbonate layer 2.5 meters thick in this interval. The scores of the principal component I hardly distinguish between the carbonate sequence at the top and the volcanic sequence at the bottom.

In the principal component scores II, strong positive values related to Al content are observed in the volcanic sequence at the bottom of the interval. Some other positive values are observed at the top of the siltstone/claystone sequence and can be explained by the presence of some Al in clays. Apart from the high positive values related to the increase in clay content, the difference between the carbonate and the siltstone/claystone sequence is hardly distinguishable from principal components II.

Figure 3.4 shows that principal component scores I and II nearly track each other in the upper part of the interval. This pattern is also observed on a cross-plot of the scores (Figure 3.5). Negative values for both principal component scores define the carbonate sequence as a very tight cloud. As the variation of both principal component scores within the siltstone/claystone sequence as well as within the volcanic sequence increases, a more spread cloud is obtained for each of these sequences. High principal component scores I and low principal component scores II define the siltstone/claystone sequence while high values in principal component scores II and low values in principal component scores I define the volcanic sequence.

Summarising, the results of the Principal Component Analysis for ODP Hole 807C suggest that the main variation present in the interval studied is that related to the gamma-ray content (high clay content). The presence of high concentrations of aluminium is reflected in the second principal component which also corresponds to high resistivity values. The principal component scores crossplot is very helpful in the determination of lithofacies heterogeneities. An example is the two dashed boxes in Figure 3.5 indicating the presence of clay intervals in the siltstone/claystone sequence. Another example are those high peaks in principal component scores II within the volcanic sequence which are probably related to high concentrations of Al.

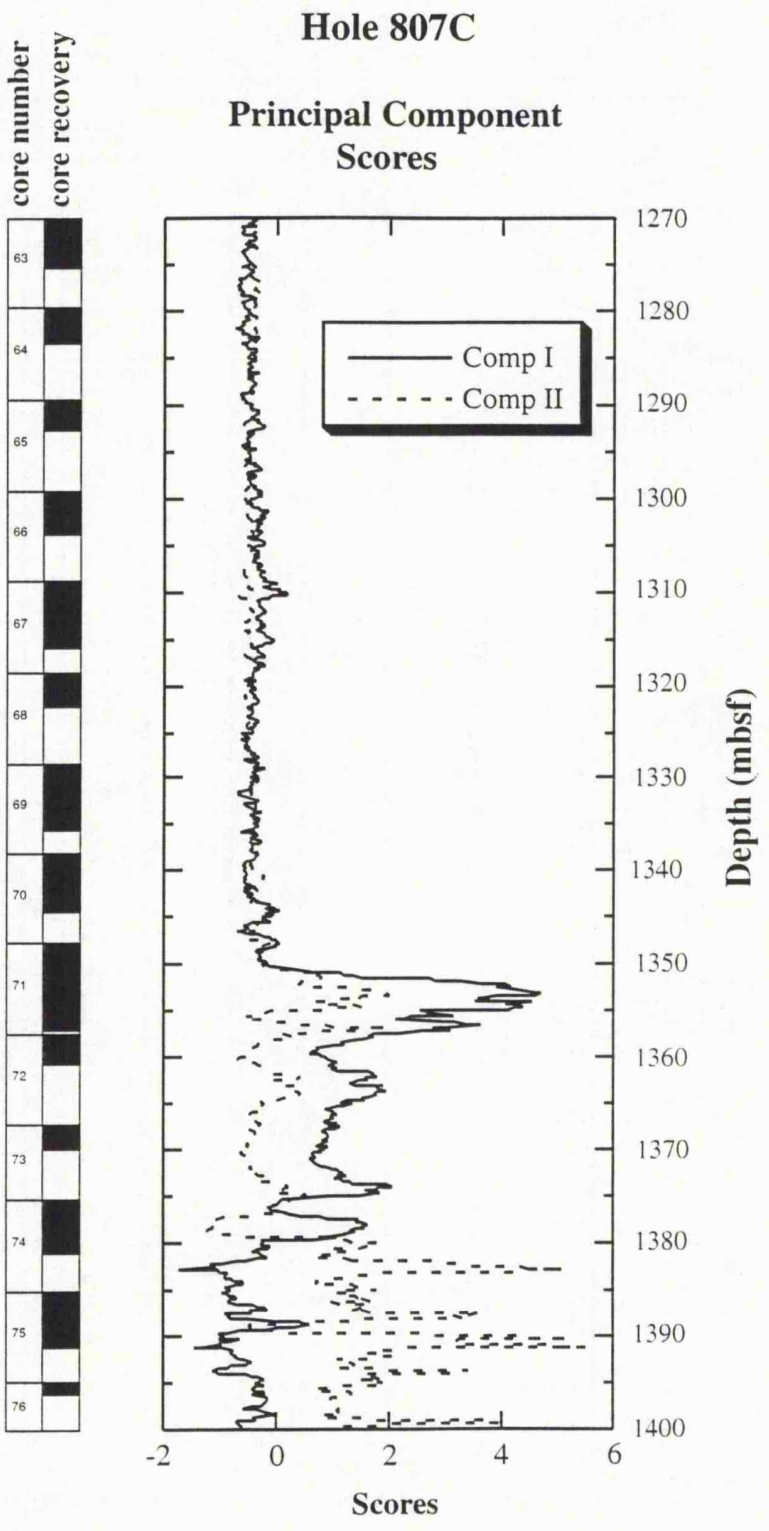


Figure 3.4 - First and second Principal Component scores in ODP Hole 807C.

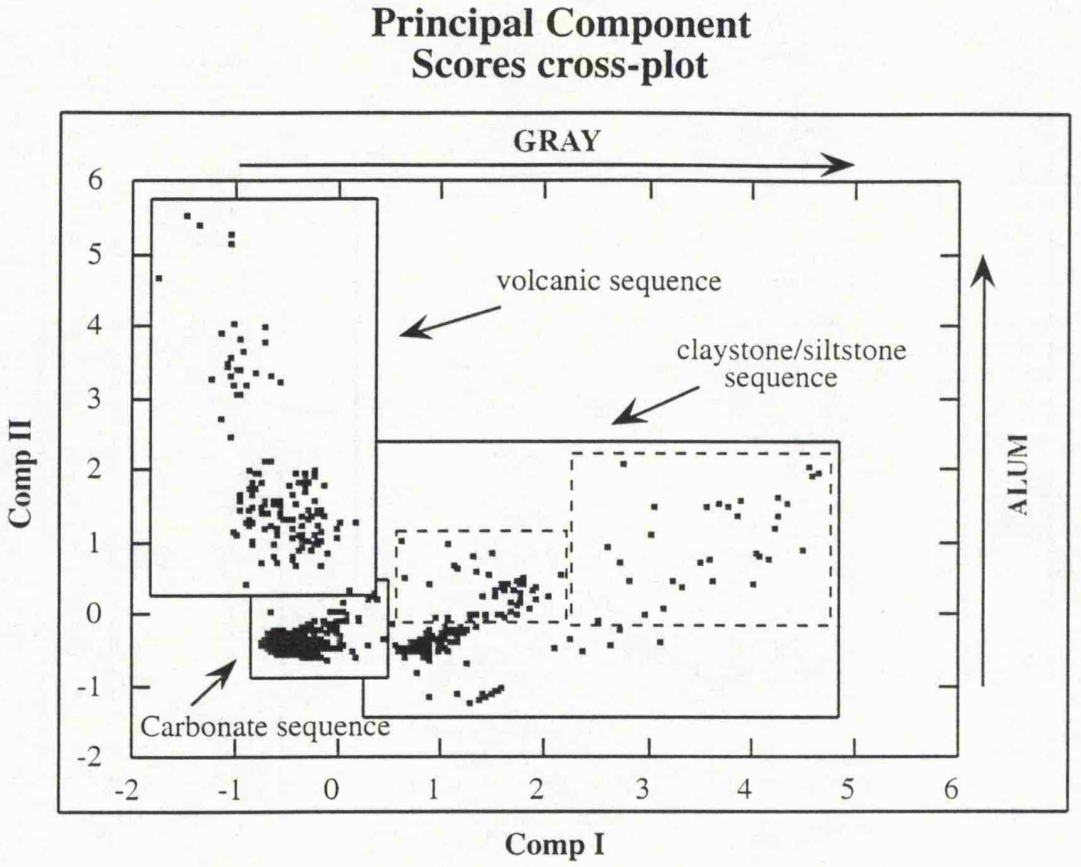


Figure 3.5 - Cross-plot of Principal Component scores I and II computed from original log curves in ODP Hole 807C.

Another interesting aspect to be observed is the contribution of each log curve used in this analysis (Table 3.5). Most of the log curves have a high contribution (>50%) with GRAY, THO and POTA showing high contributions as they are closely related with principal component I. ALUM, RHOB and PEF also appear as high contributors with 71.76%, 79.60% and 68.07% respectively.

Table 3.5

Log curve	% accounted
GRAY	96.24
POTA	85.90
THO	84.68
URAN	56.79
IDPH	49.85
IMPH	22.00
SFLU	29.97
ALUM	71.76
RHOB	79.60
PEF	68.07

Table 3.5 - Percentage contributions for each log curve in Principal Component Analysis of ODP Hole 807C.

3.3.1.2 - ODP Hole 878A

The logging dataset of ODP Hole 878A consists of 13 log curves as follows: porosity (NPHI), density (RHOB), two resistivity measurements (IDPH and SFLU), transit time (DT), thorium, uranium and potassium concentrations (THO, URAN and POTA respectively) and $CaCO_3$, Al_2O_3, SiO_2, FeO and TiO_2 concentrations (CACO3, AL2O3, SIO2, FEO and TIO2 respectively) from the Geochemical tool. As mentioned in chapter II the interval selected for the analysis contains a polymitic breccia (unit IVC - Table 2.4) between 515 and 600 mbsf.

The summary of the physical and chemical log measurements obtained in this hole are shown in Table 3.6. The large variation in most of the log measurements suggests changes in lithofacies. This is confirmed in Figure 2.6. Porosity (NPHI) and density (RHOB) show sharp changes in their values from 563 mbsf down the hole. The variation is different for each curve. When porosity decrease, density increase, reflecting changes not only in the matrix composition but also within the structure of the breccia itself. Other remarkable change observed in the dataset is in $CaCO_3$, SiO_2, Al_2O_3 and FeO concentration. While SiO_2 decrease in concentration from 563 mbsf downhole, $CaCO_3$ increase, denoting the main change in the breccia matrix

Table 3.6

Variables	NPHI	RHOB	IDPH	SFLU	CACO3	SIO2	THO	URAN	POTA	TRATIM	AL2O3	FEO	TIO2
Number of samples	558	558	558	558	558	558	558	558	558	558	558	558	558
Standard deviation	4.92	0.05	3.80	4.64	11.37	7.86	0.68	0.29	0.10	5.76	1.40	1.43	0.45
Minimum	13.19	2.26	9.21	7.77	40.51	1.16	0.12	1.15	0.01	68.58	2.97	2.96	0.58
Maximum	46.05	2.51	30.12	31.15	86.39	36.42	3.42	2.88	0.55	102.49	8.20	9.72	3.14
Range	32.86	0.25	20.91	23.37	45.88	35.26	3.29	1.72	0.54	33.90	5.23	6.76	2.55

Table 3.6 - Summary statistics for each log curve in ODP Hole 878A.

Table 3.7

	NPHI	RHOB	IDPH	SFLU	CACO3	SIO2	THO	URAN	POTA	TRATIM	AL2O3	FEO	TIO2
NPHI	1.000	-0.8547	-07887	-0.7863	-0.7699	0.6372	0.5715	-0.0613	-0.2831	0.8158	0.8247	0.7875	0.7212
RHOB	-	1.0000	0.8149	0.8316	0.7900	-0.6543	-0.5950	0.0268	0.3138	-0.8536	-0.8606	-0.7975	-0.7187
IDPH	-	-	1.0000	0.9688	0.7303	-0.6274	-0.4927	0.0017	0.1159	-0.9071	-0.7558	-0.6961	-0.6377
SFLU	-	-	-	1.0000	0.7306	-0.6319	-0.4937	0.0090	0.1064	-0.9298	-0.7660	-0.6770	-0.6127
CACO3	-	-	-	-	1.0000	-0.9562	-0.6607	0.0473	0.4366	-0.7891	-0.8543	-0.8076	-0.7414
SIO2	-	-	-	-	-	1.0000	0.5477	0.0011	-0.4378	0.6845	0.6899	0.6143	0.5674
THO	-	-	-	-	-	-	1.0000	-0.6028	-0.0937	0.5253	0.6986	0.6800	0.6360
URAN	-	-	-	-	-	-	-	1.0000	-0.3939	0.0411	-0.1064	-0.1194	-0.1533
POTA	-	-	-	-	-	-	-	-	1.0000	-0.2438	-0.2773	-0.3735	-0.3453
TRATIM	-	-	-	-	-	-	-	-	-	1.0000	0.8126	0.7430	0.6511
AL2O3	-	-	-	-	-	-	-	-	-	-	1.0000	0.8614	0.7770
FEO	-	-	-	-	-	-	-	-	-	-	-	1.0000	0.8780
TIO2	-	-	-	-	-	-	-	-	-	-	-	-	1.0000

Table - 3.7 Correlation matrix for log curves in ODP Hole 878A.

composition which changes from a Si based matrix to a Ca based matrix. Changes in the other log curves (AL2O3 and FEO) reflect mainly the variation in presence of volcanic clasts in the matrix.

The relationship observed in the dataset can be summarised by the correlation matrix (Table 3.7). Strong positive correlation can be observed between RHOB, IDPH, SFLU and CACO3 and also between TIO2, FEO, AL2O3, DT, SIO2 and NPHI. Negative strong correlation is seen between CACO3 and SIO2; and also between DT and the resistivity measurements (IDPH and SFLU). The first strong positive correlations reflect basically the high density, high resistivity and high $CaCO_3$ concentration of the Ca based matrix at the bottom of the interval and the other ones reflect mostly the high concentration of TiO_2, FeO, Al_2O_3 and high transit time of the Si supported matrix at the top of the interval.

Using the correlation matrix, a Principal Component Analysis was carried out on the dataset. The resultant eigenvectors are shown in Table 3.8 together with their percentage contribution which represent the total variability of the system. The first three eigenvectors show eigenvalues greater than 1.0 and contain 86.5% of the total system variability. The first principal component (eigenvector 1) accounts for more than 65% with the other two, principal components II and III, accounting for 12.6% and 8.5% respectively.

Table 3.8

	eigenvalues	% eigenvalues	cum. % eigenvalues
e-vector 1	8.4924	65.32	65.32
e-vector 2	1.6477	12.67	78.00
e-vector 3	1.1061	8.50	86.50
e-vector 4	0.5926	4.55	91.06
e-vector 5	0.3004	2.31	93.37
e-vector 6	0.2409	1.85	95.23
e-vector 7	0.1983	1.52	96.75
e-vector 8	0.1250	0.96	97.71
e-vector 9	0.1053	0.80	98.52
e-vector 10	0.0890	0.68	99.21
e-vector 11	0.0774	0.59	99.80
e-vector 12	0.0249	0.19	99.99
e-vector 13	0.0000	0.01	100.0000

Table 3.8 - Eigenvalues and their percentage contribution for each eigenvector in ODP Hole 878A.

Using these three first eigenvectors a principal component loading matrix was constructed (Table 3.9). The loading on each of the three first component are shown in Figure 3.6. The principal component I shows strong positive correlation with TIO2, FEO, AL2O3, DT, THO, SIO2 and NPHI, with negative correlations being observed

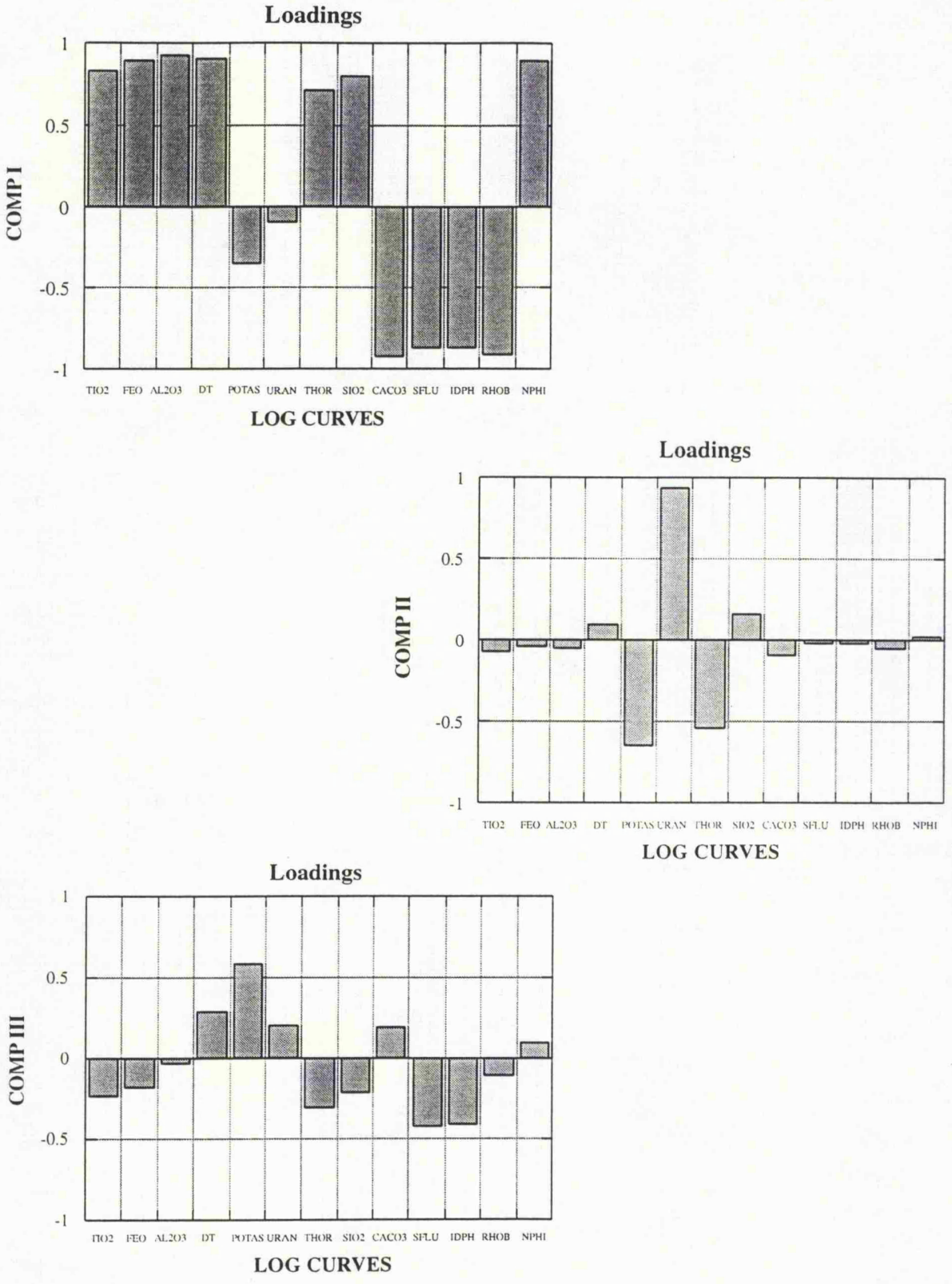


Figure 3.6 - Principal Component loadings for each log curve in ODP Hole 878A.

with CACO3, SFLU, IDPH and RHOB. This suggests that the principal component I responds mainly to changes in matrix composition. The Si based matrix has high porosity and high Ti, Al and Fe content due to the presence of volcanic clasts in the matrix. It also shows low resistivity and density values because of the presence of some clay in the matrix (Premoli Silva et al., 1993). The Ca based matrix shows high values in resistivity and density, and low values in porosity, transit time and Al content, contrasting with the upper part of the interval.

Table 3.9

Log curve	Component I	Component II	Component III
NPHI	0.8955	0.0177	0.0946
RHOB	-0.9186	-0.0489	-0.1078
IDPH	-0.8745	-0.0244	-0.4107
SFLU	-0.8753	-0.0196	-0.4292
CACO3	-0.9256	-0.0902	0.1957
SIO2	0.7990	0.1613	-0.2117
THO	0.7108	-0.5473	-0.3053
URAN	-0.0959	0.9332	0.1986
POTA	-0.3497	-0.6477	0.5879
DT	0.9097	0.0984	0.2883
AL2O3	0.9293	-0.0554	-0.0354
FEO	0.8954	-0.0404	-0.1849
TIO2	0.8312	-0.0741	-0.2293

Table 3.9 - Principal component loading matrix for ODP Hole 878A.

The principal component II shows strong positive correlation with URAN and negative correlation with POTA and THO. It represents 12.6% of the total variability present in the dataset and probably is related to clay concentrations within the Si based matrix or in fractures along the whole interval. The uranium would have concentrated in these fractures through secondary transport processes. However, these assumptions were not confirmed by core description which does not show such fractures.

The third principal component is mainly dominated by a high positive correlation with POTA and negative correlation with both resistivity measurements. It accounts for 8.5% of the system variability and possibly describes a secondary vesicular porosity present in volcanic clasts. Core description describes these vugs as commonly altered and filled with clay and $CaCO_3$ which explains the low but positive correlation of CACO3 and NPHI with this principal component.

The component scores along the whole interval are shown in Figure 3.7. The correspondence of the principal component I with the matrix composition is clearly observed by changes in the scores values. Positive scores in the upper section indicate

the Si based matrix while the negative scores below 563 mbsf represent the Ca based matrix. An interbedded interval can also be observed between 563-577 mbsf where two layers of Si based matrix appear within the Ca based matrix section. The principal component scores II and III do not show clear correlation with depth. Only principal component III show some positive peaks within the Si based matrix section possibly related with the intervals where vesicular porosity would be present.

The cross plot (Figure 3.8) of the principal component scores I and II shows a separation between the two different matrix compositions in principal component I. There is no clear separation of the dataset in principal component II direction. Any other relations using principal component scores II and III are very difficult to be defined.

The results of the Principal Component Analysis carried out on ODP Hole 878A suggest that the principal variation observed in the interval is the different composition in the polymictic breccia matrix (given by principal component I). In fact, this is the only variation that can be clearly interpreted. Other variations are very difficult to be distinguished and the interpretation given above must be checked by the use of other techniques. The variation showed by the other two principal components also need the support of core description. The use of other statistical techniques can also help to better understand what these components represent.

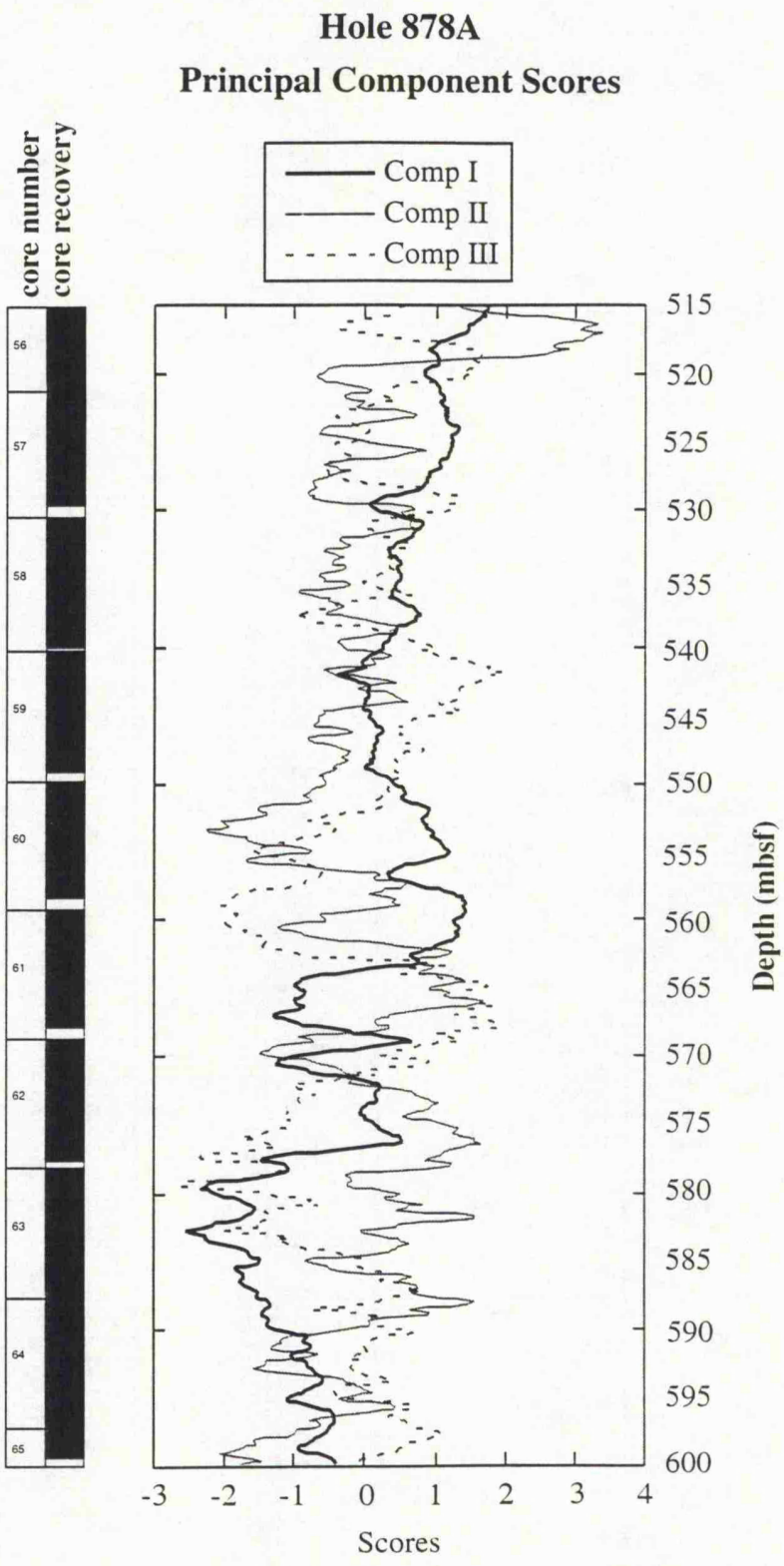


Figure 3.7 - Principal Component scores for ODP Hole 878A.

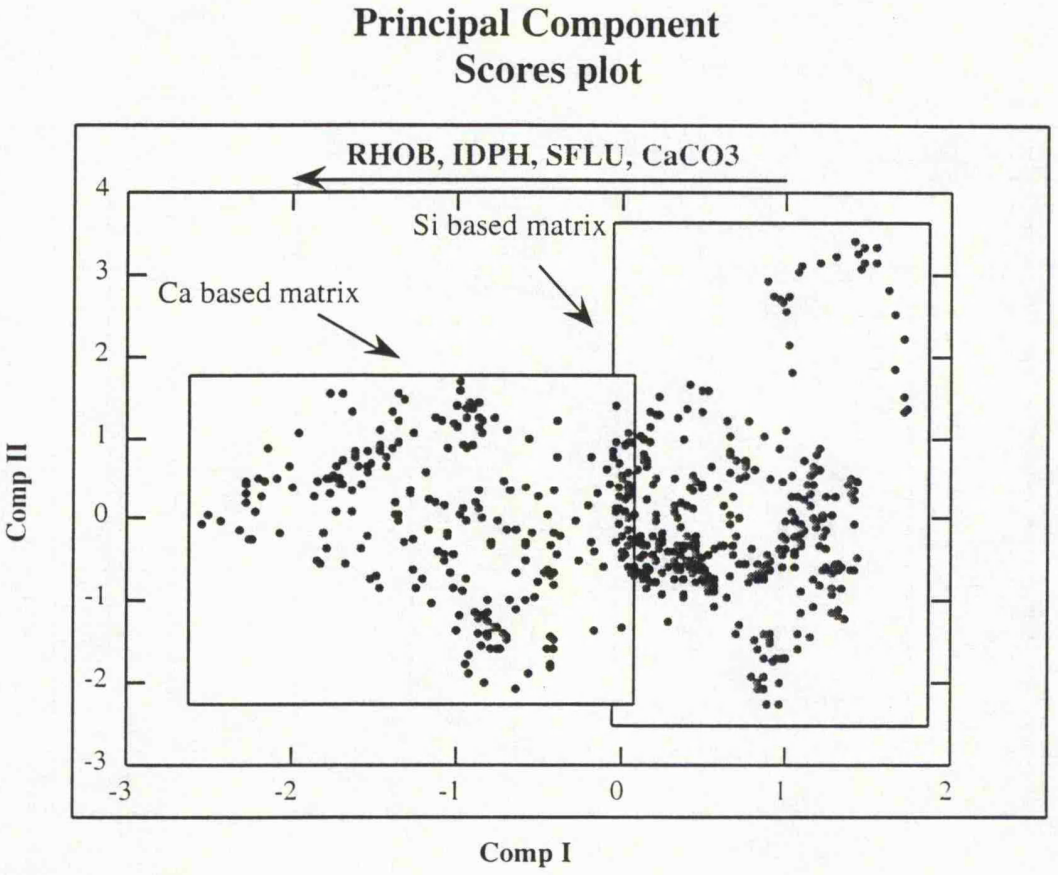


Figure 3.8 - Cross-plot of Principal Component Scores I and II computed from original log surves in ODP Hole 878A.

3.4 - Cluster Analysis

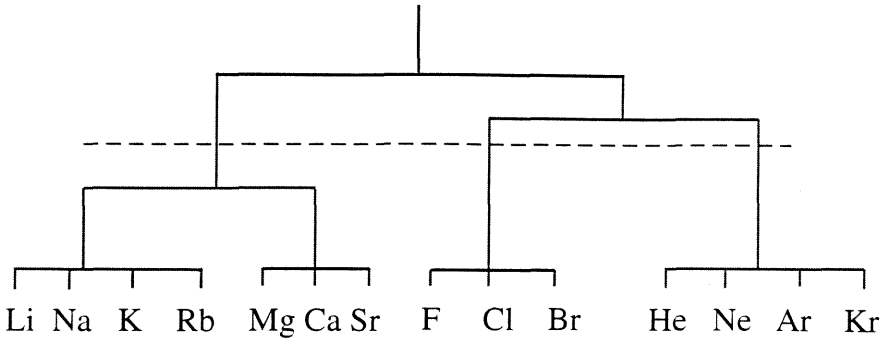
Cluster Analysis is another technique in multivariate statistics that can be defined by the methods which should sort the individual data units or samples into groups such that the degree of "natural association" is high among members of the same group, whilst the clusters themselves are relatively distinct from each other. Obviously the technique selected and the problem under study have a strong influence on the results. Its great advantage over most other techniques used in pattern recognition is that it does not need *a priori* information.

An important aspect in Cluster Analysis is that if the data itself forms well structured clusters that are compact and well separated from each other almost any clustering procedure would provide meaningful and unique cluster. On the other hand, if the distributions of the data resembles a uniform distribution, the resultant clusters may be different for different cluster techniques (Sheppard, 1986).

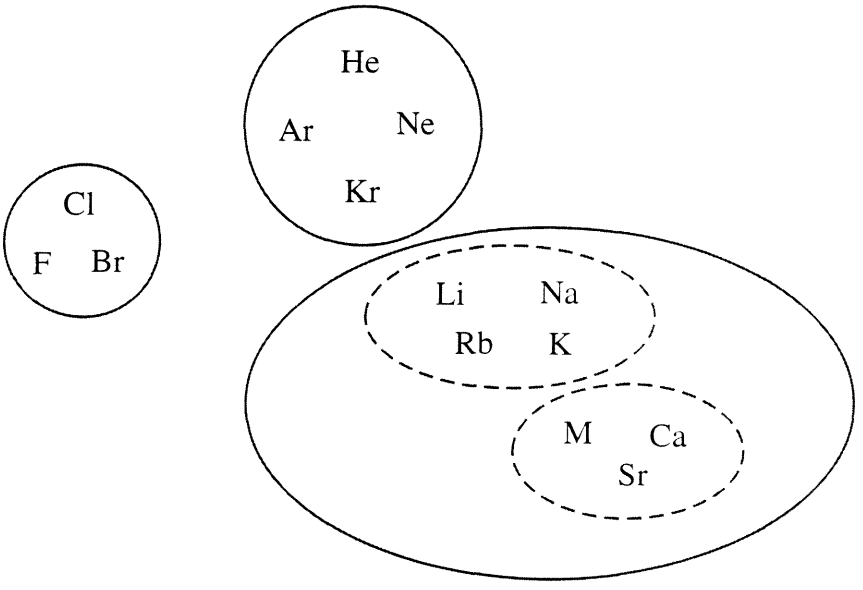
It is recognised (Andenburg, 1973) that in Cluster Analysis two problems are often overlooked. The first is that the data itself may contain no cluster at all. This happens because of the absence of discriminating log curves and also because of a uniform distribution of the points in the measurement space. The second possibility is that the data may contain only one cluster, due mostly to the absence of discriminating log curves associated with a lack of meaningful mutual association among data units.

Theoretically, Cluster Analysis can be used to develop generalisations. Although a set of results should be applied to the samples on which they were based, they may, with appropriate modification, be extended to describe the properties of other samples. It is this approach that is taken for some of the well log data characterisation below.

Two different clustering approaches used in this work are the Hierarchical Cluster Analysis (also called the R-mode) and the Non-Hierarchical Cluster Analysis (Q-mode). In the first one the relationships between the resulting category structures are described in Figure 3.9a. In the Non-Hierarchical Cluster Analysis the data units are divided into groups producing a simple partitioning of the entire data set (Figure 3.9b). In describing which approach may be most suitable for a particular problem it is necessary to be aware of the overall purpose of the study for which the results are required. In comparing the relationships between different sets of data, Hierarchical techniques may well be the most useful. If however, the object of the exercise is only to separate the various groups within a dataset it may be preferable to use Non-Hierarchical techniques (Davis, 1986).



(a)



(b)

Figure 3.9 - Partitioning of a hypothetical data set in (a) Hierarchical and (b) Non-Hierarchical Cluster Analysis

3.4.1 - Hierarchical Cluster Analysis

The Hierarchical methods operate on a similarity matrix which may be based on a variety of measurements. The most commonly used are distance, e.g. the distance between the nearest element in each cluster or the distance between the centroids of each cluster, or even measurements based on correlation. If standardised data are used, which is the case in this work, measurements of distance and correlation coefficients may be directly transformed from one to another. Similarly, distance may be looked on as a measure of dissimilarity.

Davis (1986) shows that in general distance-based measures tend to cluster more successfully using cophenetic correlation than correlation-based measures. They also appear to be less susceptible to changes in clustering method. Although the two measurements tend to give similar results, the distance-based methods are not constrained between +/- 1.0 and therefore may be expected to produce better results if a few of the data units are very dissimilar from one to another. The results are shown as a dendogram or a tree diagram (Figure 3.10). Sheppard (1986) explains in detail the alternative approaches in the construction of the dendogram which will be used in this work.

A serious problem with the Hierarchical procedures is that their theoretical basis is still incomplete. The statistical properties of the various techniques are poorly understood and there are few tests of significance. Therefore there is not the same need as with the other multivariate statistical techniques for a multivariate normal distribution. Multivariate normality does, however, become important if correlations are taken as the measure of similarity.

3.4.2 - Iterative Non-Hierarchical Cluster Analysis (INCA)

Iterative Non-Hierarchical Cluster Analysis (INCA) is a classification technique employed here to characterise the log responses and help define any stratigraphic zonation that might be identified from the data sets. INCA is a simple method, available for a number of years under one name or another, for instance, K-means clustering and relocation analysis (Le Maitre, 1982; Johnson and Wichern, 1982; Sheppard, 1986).

The objective of the technique is to classify (or split) a multivariate data set into groups of samples (log responses) which have similar characteristics. The technique partitions the entire data set of n samples into K discrete groups on the basis of optimisation of a stated mathematical criterion (Doveton, 1986). The number of clusters formed, K, may either have been previously specified or determined during the clustering operation. The technique is not used for the classification of log curves.

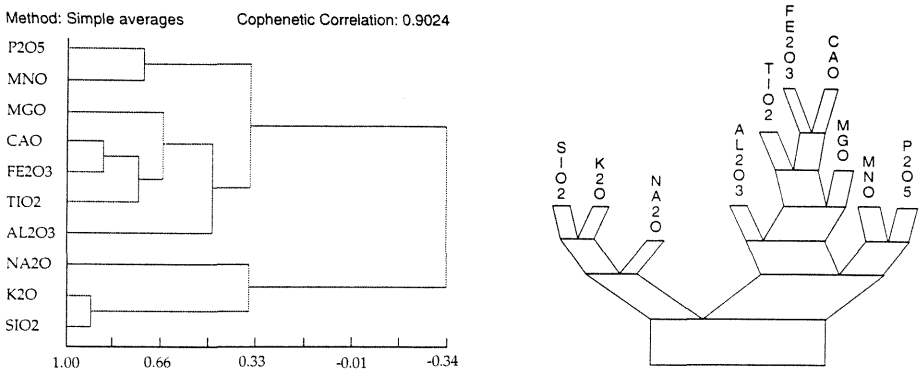


Figure 3.10 - Examples of a dendogram and a KH diagram to show the results on Hierarchical Cluster Analysis.

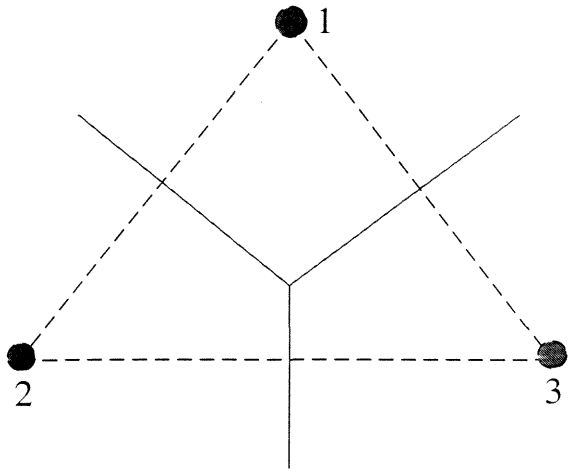


Figure 3.11 - Piecewise linear boundary separation in Non-Hierarchical Cluster Analysis (after Sheppard, 1986).

The basic concept of this method is to select some initial partition of the samples and then modify the cluster membership in order 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 cluster are separated by precise linear boundaries (Figure 3.11). These are the locus 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 (Doveton, 1994).

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 3.11, 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 this problem in this work involves running the analysis for a range of groups, and then choosing the *optimum* number of groups. This is the sum of the squared Euclidean distance from each data point to the centroid to which it has been allocated. Averaging during acquisition and processing means that distinct groups cannot occur in logging data. In addition, these data frequently represent gradational changes between lithofacies. The number of groups chosen is, therefore, not optimal in the true sense but represent convenient divisions within the data whose boundaries are gradational. On the other hand, subtle changes within the same lithofacies can lead to the classification of heterogeneities.

In addition to group membership for each log response, the analysis yields the centroid composition for each group, the position of each group relative to each other group and the "delta values" (δ), which rate the importance of each log curve in terms of group discrimination. 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. Delta values provide a measure of the separating power of 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 standardised group centroids and the standard deviations of the standardised data contained within each group. These were combined in such a way to give a value, δ, which provide 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. Sheppard (1986) suggested that:

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

where δ_i is the value of δ for the i^{th} log curve, C_{ij} is the standardised value of the i^{th} log curve at the centroid of the j^{th} group, C_{ik} is the standardised 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 standardised values of that log curve, where:

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

SP_x is the sums of 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 δ then describe the separating power of individual log curves for the various combinations of groups. Figure 3.11a shows some hypothetical examples for the case of one variable only: a) large separation between group centroids and a small pooled standard deviation, resulting in a very high value of δ; b) small separation between group centroids and a small pooled standard deviation, leading to a moderate to low δ; and large separation between groups and large pooled standard deviation (low δ).

From the values of δ obtained for each log curve in each combination of groups, three parameters are estimated. The first parameter is an overall least to most important log curve classification. This is achieved by calculating the mean of the δ values for each log curve and then sorting then into ascending order. The second parameter is the identification of the worst to best log curves of each group combination. Therefore, it allows the analyst to assess which of the log curves has more weight in the lithofacies discrimination. The δ values of each log curve are comparable as they have all been standardised against standard deviation. The last parameter refers to the overall ease of distinguishing various group combinations. The mean of all δ values for each combination are calculated and then sorted into ascending order. The use of δ in the assessment of the importance of individual log curves as well as in distinguishing group combinations are well described in the analysis of each hole.

Initial seed points were, in all cases studied, chosen by the method of Ball and Hall (1967) where actual samples are selected as seeds. All analyses reported in this work were run until convergence, however, this is not always necessary. INCA analyses are performed on both original and transformed (principal component scores) log curves, and are presented in the later sections of this chapter. Statistically, analysis of principal component scores should be superior since the transform removes any correlation between log curves. This proved particularly pertinent to the datasets used in this work. A complete description of the theory and the mathematics of the technique and also the algorithms used is given by Sheppard (1986) and Davis (1986).

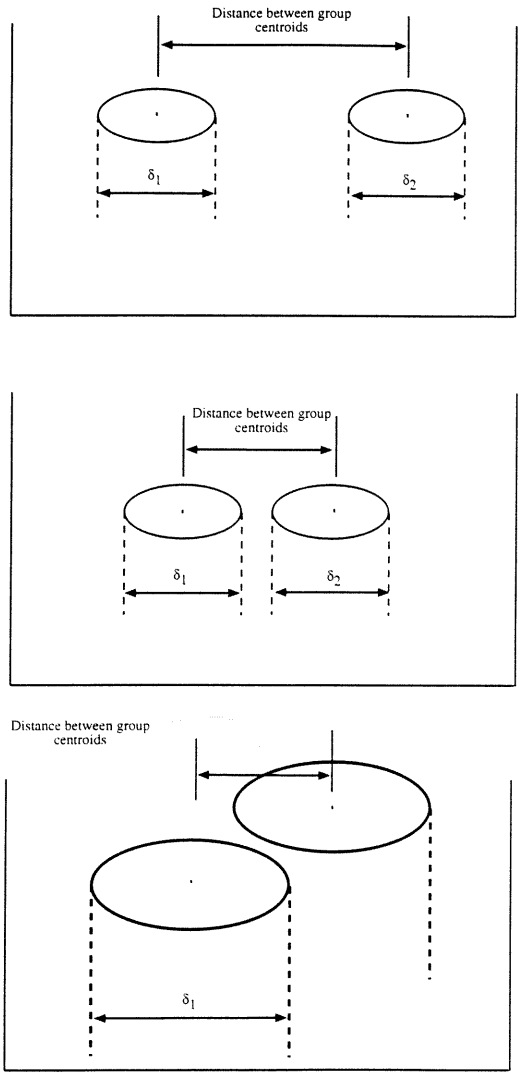


Figure 3.11a - Use of δ values to describe the separating power of individual log curves.

3.4.3 - Application of Cluster Analysis to well log data

3.4.3.1 - ODP Hole 807C

Cluster Analysis was carried out in ODP Hole 807C using the same log curves used for Principal Component Analysis, except for IMPH and SFLU. As discussed before (Table 3.5), these two log curves showed very low contribution in that analysis, therefore, they were removed from the original dataset for the Hierarchical and Non-Hierarchical Cluster Analysis.

Hierarchical Cluster Analysis

The algorithm used in clustering can be viewed in three stages. The first stage is the input of the log data or a similarity matrix that could be computed from the Principal Component Analysis. The second stage is the computation of a similarity/dissimilarity matrix if one was not input directly in stage one. Therefore, the second stage can be optional. The third stage is the clustering of the similarity/dissimilarity matrix using one of a selection of models presented below (Harvey, P.K., pers. communication).

The similarity/dissimilarity matrix support: a) a product moment correlation matrix (similarity), b) a Euclidean distance - not normalised matrix (dissimilarity) or c) a Euclidean distance - Z-standardised (dissimilarity). The clustering algorithms can be computed from the following different models: a) Nearest neighbours, which uses a single linkage and minimum method; b) Further neighbours, which uses complete linkage and maximum method; c) Simple averages, the weighted pair-group method; d) Median, the weighted pair group centroid; e) Group averages, the unweighted pair-group method; f) centroid, the weighted pair-group centroid; and g) Wards method, which uses the minimum group variance. Most research using Hierarchical Cluster Analysis normally experience a variety of similarity measures and clustering techniques. Then a combination that appears to yield the most satisfactory results with their data is used. This, in turn, introduces an element of subjectivity into a process which is supposed to be objective. Because the log data are normalised before processing, the Hierarchical Clustering is conducted using the Euclidean distance - Z-standardised similarity/dissimilarity matrix. After running the algorithm for some of the linkage models listed above, a model with the highest cophenetic correlation is selected for a full analysis.

The results from the Hierarchical Cluster Analysis for ODP Hole 807C are shown in Table 3.10. It presents the Euclidean distance matrix, the results for some of the linkage methods and the detailed result for the method which presents the best

Table 3.10

Hierarchical Cluster Analysis (RCL) 807C.RCL

Euclidean distance matrix.... Z-standardised

	GRAY	POTA	THO	URAN	IDPH	ALUM	RHOB	PEF
GRAY	0.0000	0.3007	0.3317	0.7498	1.4492	1.2277	1.7451	1.7905
POTA		0.0000	0.5147	0.9226	1.4268	1.1512	1.6811	1.7394
THO			0.0000	0.9542	1.4723	1.3047	1.7626	1.7862
URAN				0.0000	1.4331	1.3180	1.6947	1.7386
IDPH					0.0000	1.0878	1.1476	1.2893
ALUM						0.0000	1.0905	1.3277
RHOB							0.0000	0.5952
PEF								0.0000

Model	Cophenetic Correlation	Linkage method
1	0.8731	Nearest Neighbours
2	0.9176	Furthest Neighbours
3	0.9278	Simple Averages
5	0.9258	Group Averages
7	0.6684	Wards Method

Model: .... 3 Simple averages

Cophenetic correlation 0.9278

link no.	log curve retained	log curve deleted	similarity level	linkage order
1	1	2	0.3007	1
2	1	3	0.4232	2
3	7	8	0.5952	3
4	1	4	0.8952	4
5	5	6	1.0878	5
6	5	7	1.2138	6
7	1	5	1.5500	7
				8

Table 3.10 - Hierarchical Cluster Analysis for ODP Hole 807C.

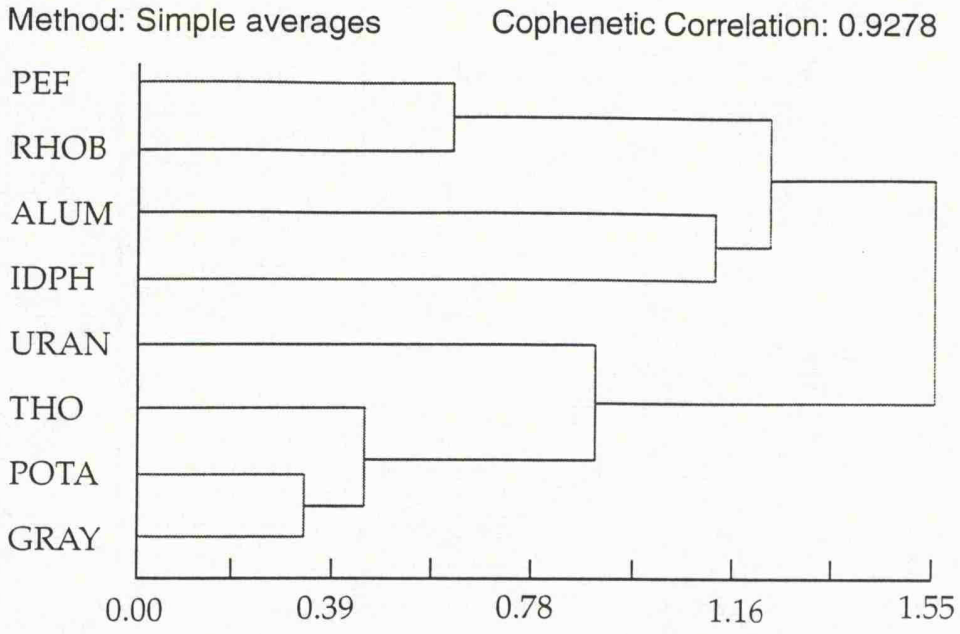


Figure 3.12 - Dendogram for Hierarchical ClusterAnalysis in ODP Hole 807C.

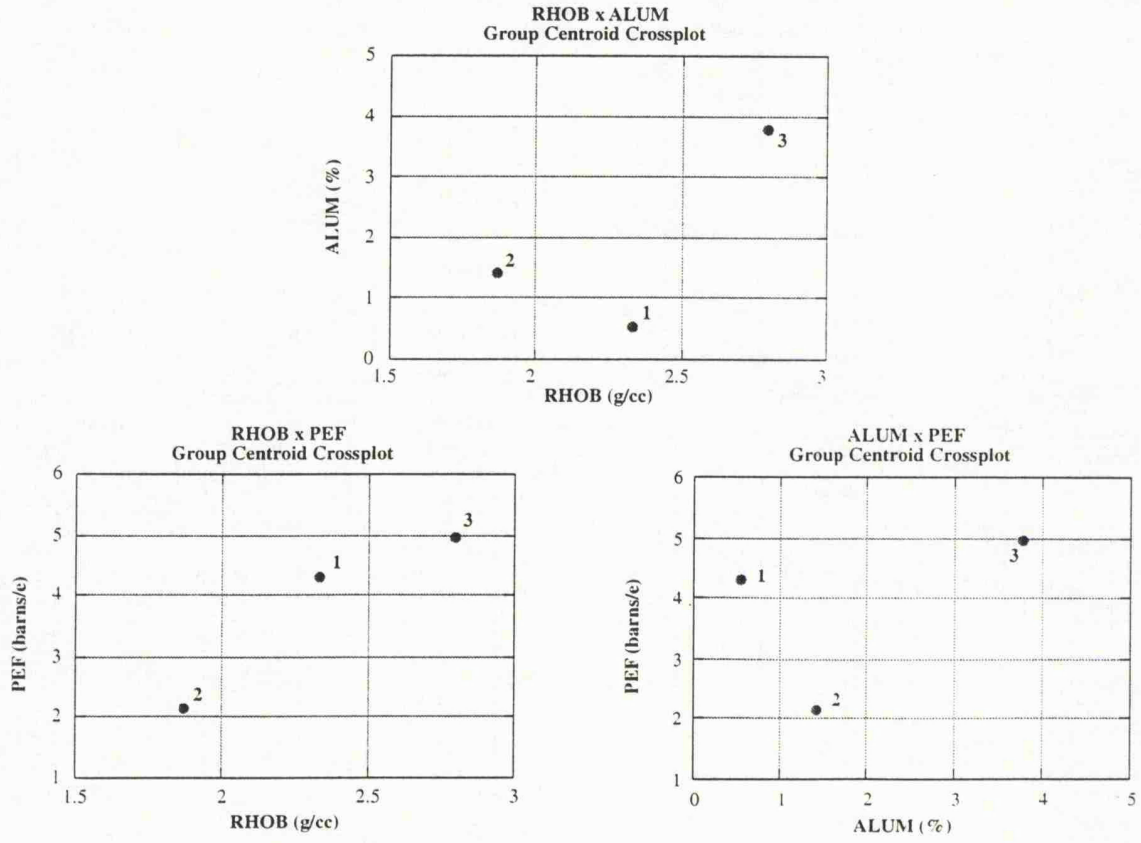


Figure 3.13 - Cross-plots showing the three groups centroids in terms of:
a) RHOB x ALUM, b) RHOB x PEF and c) ALUM x PEF.

cophenetic correlation. The cophenetic correlation, which represents the degree of distortion in the dendogram, are greater than 0.8 for most of the linkage methods presented. Davis (1986) shows that 0.8 is an acceptable value for cophenetic correlation. The pair-group methods, both models 3 and 5, show the best cophenetic correlation values, greater than 0.92. The reason for the minimum difference observed in both methods is because the data were normalised, giving no difference for the weighted or unweighted methods in this case. As the simple averages method shows the greater cophenetic value (0.9278) it was used for the analysis and its result is detailed in Table 3.10.

The dendogram (Figure 3.12) presents the relation between the log curves used in this analysis. It shows strong correlation between GRAY and POTA with approximately 0.30 units of dissimilarity. THO and URAN also appear connected with the previous log curves. On the other side, RHOB and PEF also show strong correlation (0.60 units of dissimilarity). The separation of the log curves in two different groups is clear from the dendogram. The first group is dominated by the nuclear measurements (GRAY, THO, POTA and URAN) which have high correlation. The second one is dominated by RHOB, PEF, IDPH and ALUM. These last two log curves have the largest degree of dissimilarity.

This result is confirmed when it is compared with the first principal component loading, which represents more than 65% of the system variability. For that case the nuclear measurements present strong positive loading. RHOB and PEF, on the other hand, present strong negative loading whilst ALUM and IDPH have their loadings around zero.

The Non-Hierarchical Cluster Analysis

The Non-Hierarchical Cluster Analysis (INCA) was carried out both on the original log curves and, on the principal component scores derived from the Principal Component Analysis and described in section 3.3.1. In both cases the iterative relocation and the Euclidean model were used in the analysis. Z-standardisation was applied to the data and the starting points for clustering were chosen with minimum distance between them.

a) Analysis using the original log curves

Initial testing for the number of groups present in the dataset resulted in three main groups being selected as the *optimum* number for the full analysis. Group I represents the carbonate sequence, Group 2 the siltstone/claystone sequence and Group 3 the volcanic sequence. The centroid compositions of the groups produced are

detailed in Table 3.11. In terms of group discrimination, RHOB, PEF and ALUM were rated the three most important log curves in the analysis based on their averages delta values (Table 3.12). The plots of group centroids for each pair of these log curves are presented in Figure 3.13. As expected, they show good discrimination between the groups formed. The most important log curves for the discrimination between each group are also shown in Table 3.12. In the separation between Groups 1 and 2 and between Groups 2 and 3, PEF and RHOB are the most important elements, which high delta values compared with other log curves. In the separation between Groups 1 and 3, ALUM and RHOB are the most important. It is in fact explained by the great difference in Al concentration between the carbonate sequence (Group 1) and the volcanic sequence (Group 3). Table 3.12 also shows that IDPH and URAN are the log curves with overall least importance in the group identification.

Table 3.11

Logs	GRAY	POTA	THO	URAN	IDPH	ALUM	RHOB	PEF	No. in cluster
Group 1	5.9613	0.1384	0.4124	0.3073	3.5456	0.5444	2.3338	4.2922	550
Group 2	26.2763	0.7027	2.6265	0.7784	1.5609	1.4127	1.8703	2.1543	178
Group 3	9.6006	0.3487	0.4148	0.3597	377.7767	3.7742	2.7948	4.9580	125

Table 3.11 - Centroid composition for each group in Non-Hierarchical Cluster Analysis using log curves.

The group log is shown in Figure 3.14. As can be observed, Group 1 represents the carbonate sequence at the top of the interval and also appears between 1375-1377 mbsf and around 1389 mbsf. In the first case it represents a carbonate interval which occurs at that depth, however, in the second case it seems to be a reflection of a fractured/altered interval with low values of ALUM, RHOB and PEF.

Group 2 represents the siltstone/claystone sequence between 1351-1375 mbsf. The interval between 1377 - 1380 mbsf can be either this lithology or the alteration of the volcanic sequence below. Kroenke et al. (1991) show that the characteristic feature in considering the top of the basement at 1380 mbsf is the strong decreasing in the drill rate penetration at that depth. However, an altered interval could show the same drilling rate as a carbonate sequence. Finally, Group 3 represents the volcanic sequence between 1380 - 1400 mbsf.

Based on the means of the delta values for the separation between each group, Table 3.13 shows the hardest and the simplest combination of groups to distinguish. Groups 1 and 3 and Groups 2 and 3 are the more difficult combinations to distinguish, with means delta values of 2.28 and 2.29 respectively. Groups 1 and 2 appear as the simplest combination to distinguish with mean of delta value equal to 2.43.

Table 3.12

IDPH	Overall least important element	0.85	(Mean of delta values)
URAN		1.09	
THO		1.30	
POTA		1.79	
GRAY		1.95	
PEF		3.62	
ALUM		3.63	
RHOB	Overall most important element	4.46	(Mean of delta values)
IDPH	Worst element for cluster 1 & 2	0.42	(delta)
URAN		1.83	
ALUM		1.93	
THO		2.33	
POTA		2.48	
GRAY		2.70	
RHOB		3.41	
PEF	Best element for cluster 1 & 2	4.35	(delta)
THO	Worst element for cluster s 1 & 3	0.01	(delta)
URAN		0.28	
IDPH		1.27	
PEF		1.46	
GRAY		1.71	
POTA		1.87	
RHOB		5.35	
ALUM	Best element for clusters 1 & 3	6.30	(delta)
IDPH	Worst element for clusters 2 & 3	0.86	(delta)
POTA		1.01	
URAN		1.15	
GRAY		1.44	
THO		1.56	
ALUM		2.65	
RHOB		4.61	
PEF	Best element for clusters 2 & 3	5.05	(delta)

Table 3.12 - Importance of log curves in the group discrimination of ODP Hole 807C.

Table 3.13

Cluster 1 & 3	Hardest combination to distinguish	2.28 (means)
Cluster 2 & 3		2.29 (means)
Cluster 1 & 2	Simplest combination to distinguish	2.43 (means)

Table 3.13 - Hardest and simplest combination to distinguish in ODP Hole 807C.

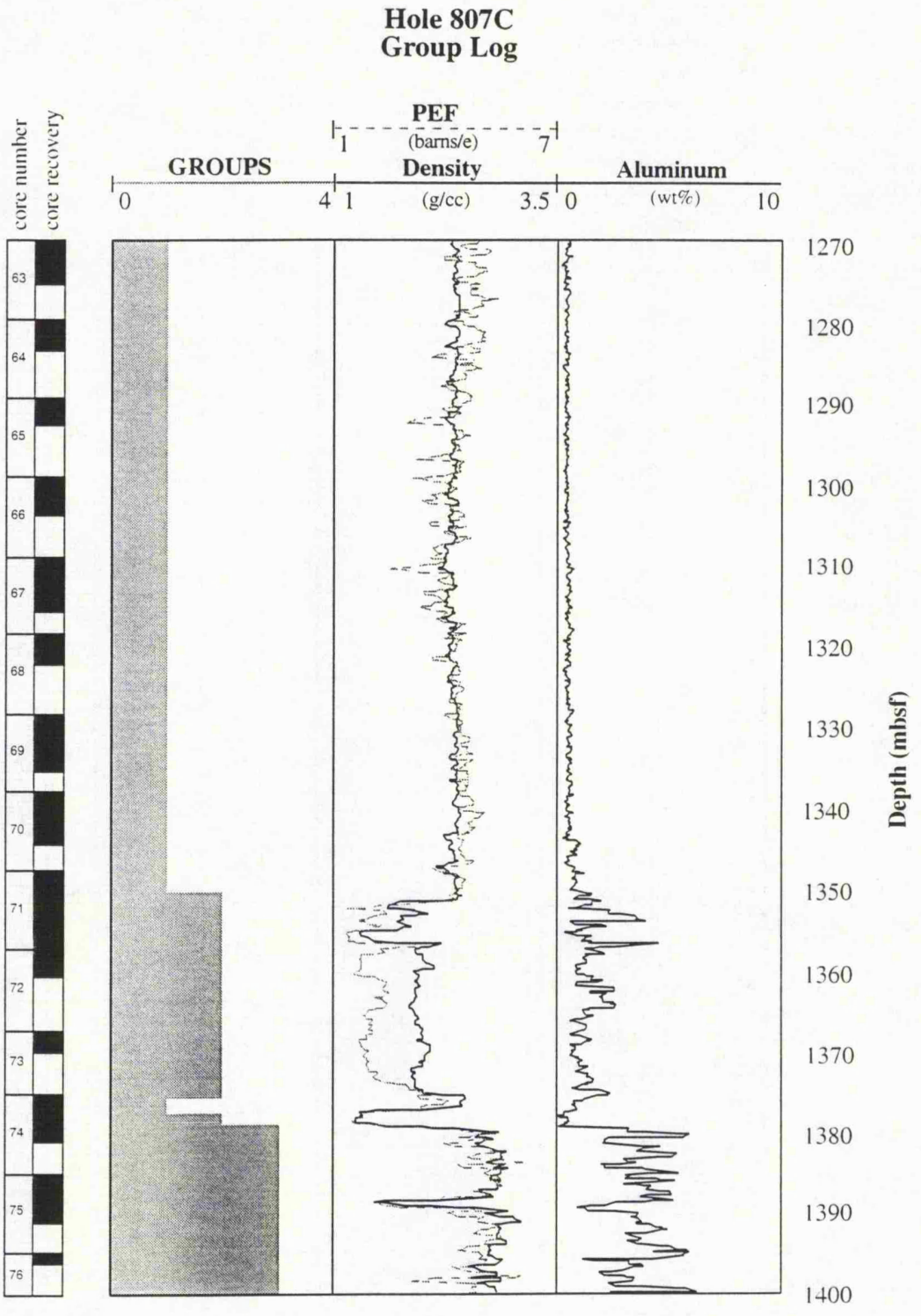


Figure 3.14 - Group log for Non-Hierarchical Cluster Analysis using the original log curves in ODP Hole 807C. ALUM, PEF and RHOB are also plotted for comparison.

b) Analysis using principal component scores

A second Non-Hierarchical Cluster Analysis was performed using the principal component scores calculated using the loading matrix (Table 3.4) produced by the earlier Principal Component Analysis. Initial testing suggested that four groups were the *optimum* number of groups for the full analysis and details of the resultant group centroids are given in Table 3.14.

Table 3.14

	Component I	Component II	No. in cluster
Group 1	-0.4066	-0.4162	551
Group 2	-0.6139	1.9766	119
Group 3	1.1904	-0.2421	151
Group 4	3.6663	0.9577	32

Table 3.14 - Centroid composition for Non-Hierarchical Cluster Analysis (using principal component scores).

Table 3.15

Component 2	Overall least important element	2.79 (Mean of delta value)
Component 1	Overall most important element	7.61 (Mean of delta value)
Component 1	Worst element for clusters 1 & 2	0.98 (delta)
Component 2	Best element for clusters 1 & 2	4.28 (delta)
Component 2	Worst element for clusters 1 & 3	0.67 (delta)
Component 1	Best element for clusters 1 & 3	6.22 (delta)
Component 2	Worst element for clusters 1 & 4	6.07 (delta)
Component 1	Best element for clusters 1 & 4	18.67 (delta)
Component 2	Worst element for clusters 2 & 3	2.44 (delta)
Component 1	Best element for clusters 2 & 3	4.43 (delta)
Component 2	Worst element for clusters 2 & 4	0.87 (delta)
Component 1	Best element for clusters 2 & 4	10.24 (delta)
Component 2	Worst element for clusters 3 & 4	2.44 (delta)
Component 1	Best element for clusters 3 & 4	5.11 (delta)

Table 3.15 - Importance of principal component scores I and II in group discrimination.

In terms of group discrimination, the principal component scores I, which has a mean delta value of 7.61, is the overall most important element (Table 3.15). It also

appears as the most important element for the discrimination between each group, except for Groups 1 and 2, where the principal component scores II is the best element. This can be explained because of principal component II has strong positive loadings (Figure 3.3) in ALUM and IDPH which have their extreme values in the carbonate and volcanic sequences, represented in this analysis by Groups 1 and 2 respectively. Group 3, now, represents the claystone/siltstone sequence (Figure 3.15). One new group (Group 4) was obtained in this analysis and can be observed between 1352 - 1357 mbsf. This group represents the high clay content interval in claystone/siltstone sequence and can be considered as a heterogeneity detected by the analysis with the transformed data.

Table 3.16 shows the hardest and simplest combination of groups to distinguish based on means of delta values. Groups 1 and 2 appear as the hardest combination to distinguish because, as shown before, the great variation in log data between these two groups are given by ALUM and IDPH, strong in principal component II. The simplest combination, and therefore the groups which show the more distinguishable characteristics are Groups 1 and 4 with a mean delta value of 12.37. The analysis with the principal component is very important because it indicates some features not detected using the original log curves. Although the interval between 1352 - 1357 mbsf might be distinguishable through visual analysis of the log curves, it was only detected as a new group in the analysis with the principal component scores, which uses a reduced dataset, therefore simplifying the process.

Table 3.16

Clusters 1 & 2	Hardest combination to distinguish	2.63 (means)
Clusters 2 & 3		3.43 (means)
Clusters 1 & 3		3.44 (means)
Clusters 3 & 4		3.77 (means)
Clusters 2 & 4		5.55 (means)
Clusters 1 & 4	Simplest combination to distinguish	12.37 (means)

Table 3.16 - Hardest and simplest combinations to distinguish in ODP Hole 807C.

Both Non-Hierarchical analysis have been able to identify most of the gross lithological changes which occur within the whole interval. The existence of large variations in some log curves is a major contributor to this success. Given this fact, some of the features could have been identified by visual appraisal if the number of log curves were reduced. However, when the number of log curves involved rises and some of the logs show more subtle changes, the visual approach becomes more complicated and the multivariate analysis is necessary. The principal component

Hole 807C
Groups and Principal Component Scores

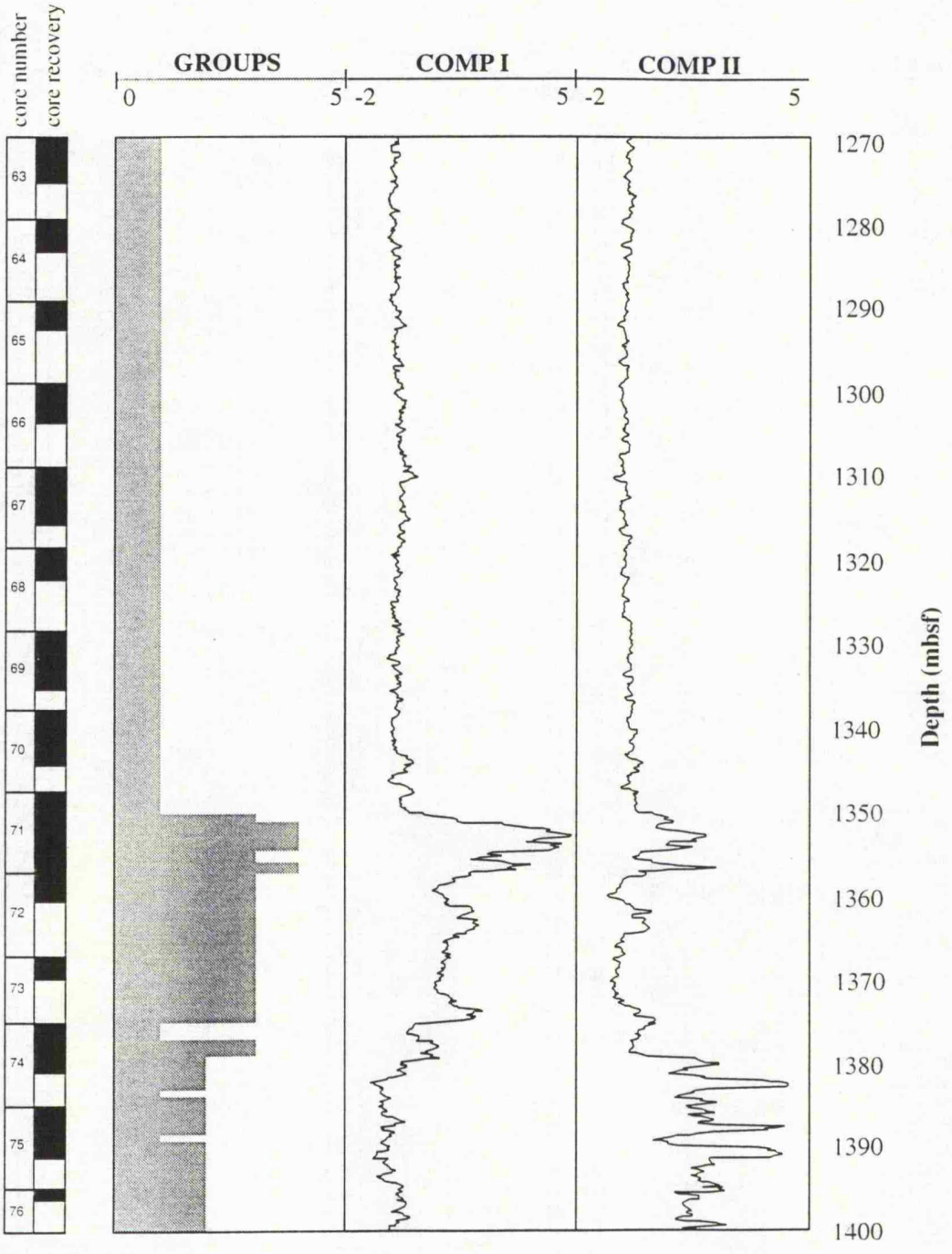


Figure 3.15 - Group log for Non-Hierarchical Cluster Analysis using principal components I and II in ODP Hole 807C.

transform allows analysis (both statistical and visual) to be performed on a reduced dataset and, most important, simplifies the presentation of the logs and allows the detection of other structures or heterogeneities not observed in the analysis with the original log curves.

3.4.3.2 - ODP Hole 878A

In the case of ODP Hole 878A, Cluster Analysis used exactly the same dataset used in the earlier Principal Component Analysis. As the contribution for each log curve used in that technique (Table 3.17) shows that all of them account for more than 70% in the analysis, it was decided to keep all the log curves for the Cluster Analysis.

Table 3.17

Log curve	% accounted
NPHI	81.11
RHOB	85.79
IDPH	93.40
SFLU	95.09
CACO3	90.32
SIO2	70.92
THO	89.80
URAN	91.95
POTA	88.75
DT	92.03
AL2O3	86.79
FEO	83.76
TIO2	74.89

Table 3.17 - Log curve contributions in Principal Component Analysis of ODP Hole 878A.

Hierarchical Cluster Analysis

Using the Euclidean distance - Z-standardised dissimilarity matrix and the Simple averages model, the results for the Hierarchical Cluster Analysis in ODP Hole 878A are shown in Table 3.18. It contains the Euclidean distance matrix, the results for some of the models presented in section 3.3.3.1 and the detailed results for the simple averages model. As can be seen, this model does not show the best cophenetic correlation, but its value (0.9389) is approximately the same as for other models.

The dendogram (Figure 3.16) for this analysis expresses the same variation observed in the first principal component loadings. It shows that SFLU - IDPH, TIO2

- FEO and AL2O3 - NPHI contain the strongest correlation with the minimum dissimilarity values. Two main groups can also be distinguished from the dendogram. The first group consists of CACO3, SFLU, IDPH, RHOB, URAN and POTA, and

Table 3.18

Hierarchical Cluster Analysis (RCL) ... 878A.RCL
Euclidean distance matrix.... Z-standardised

	NPHI	RHOB	IDPH	SFLU	CACO3	SIO2	THO	URAN	POTA	DT	AL2O3	FEO	TIO2
NPHI	0.000	1.926	1.891	1.890	1.881	0.851	0.925	1.456	1.601	0.606	0.592	0.651	0.746
RHOB		0.000	0.608	0.580	0.648	1.818	1.786	1.395	1.171	1.925	1.929	1.896	1.854
IDPH			0.000	0.249	0.734	1.804	1.727	1.413	1.329	1.953	1.873	1.841	1.809
SFLU				0.000	0.734	1.806	1.728	1.407	1.336	1.964	1.879	1.831	1.795
CACO3					0.000	1.978	1.822	1.380	1.061	1.891	1.925	1.901	1.866
SIO2						0.000	0.951	1.413	1.695	0.794	0.787	0.878	0.930
THO							0.000	1.790	1.479	0.974	0.776	0.800	0.853
URAN								0.000	1.669	1.384	1.487	1.496	1.518
POTA									0.000	1.577	1.598	1.657	1.640
DT										0.000	0.612	0.717	0.835
AL2O3											0.000	0.526	0.667
FEO												0.000	0.494
TIO2													0.000

Model	Cophenetic Correlation	Linkage method
1	0.948	Nearest Neighbours
2	0.964	Furthest Neighbours
3	0.963	Simple Averages
5	0.965	Group Averages
7	0.683	Wards Method

Model: ....3 Simple Averages Cophenetic correlation ...0.963

link no.	log curve retained	log curve deleted	similarity level	linkage order
1	3	4	0.249	1
2	12	13	0.494	11
3	1	11	0.592	10
4	2	3	0.594	12
5	1	10	0.609	13
6	2	5	0.691	6
7	1	12	0.712	7
8	1	6	0.855	2
9	1	7	0.910	3
10	2	9	1.156	4
11	2	8	1.530	5
				8

Table 3.18 - Hierarchical Cluster Analysis for ODP Hole 878A.

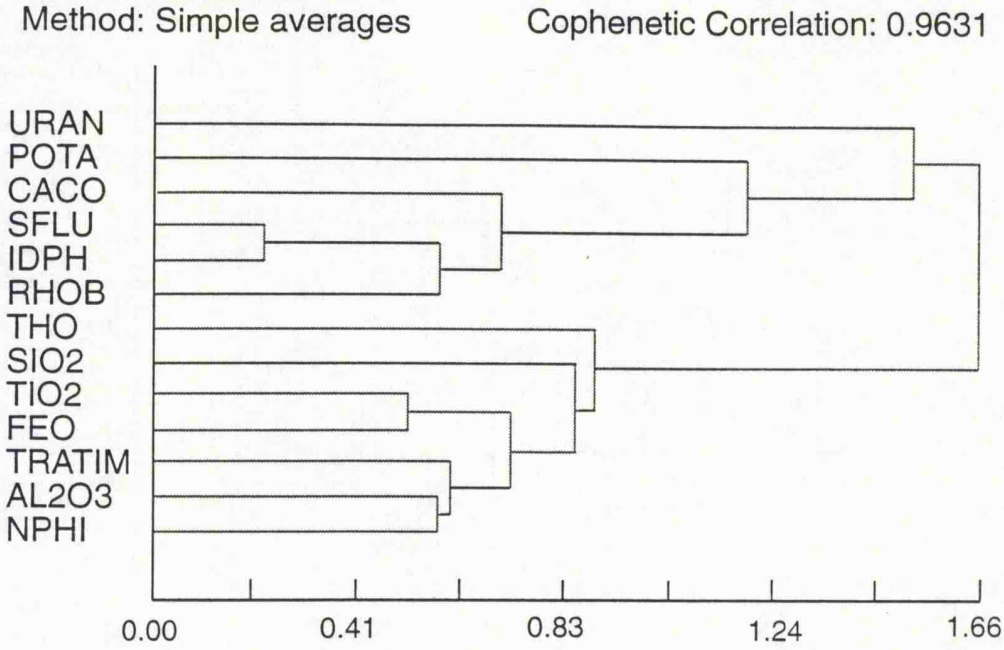


Figure 3.16 - Dendogram for Hierarchical Cluster Analysis in ODP Hole 878A.

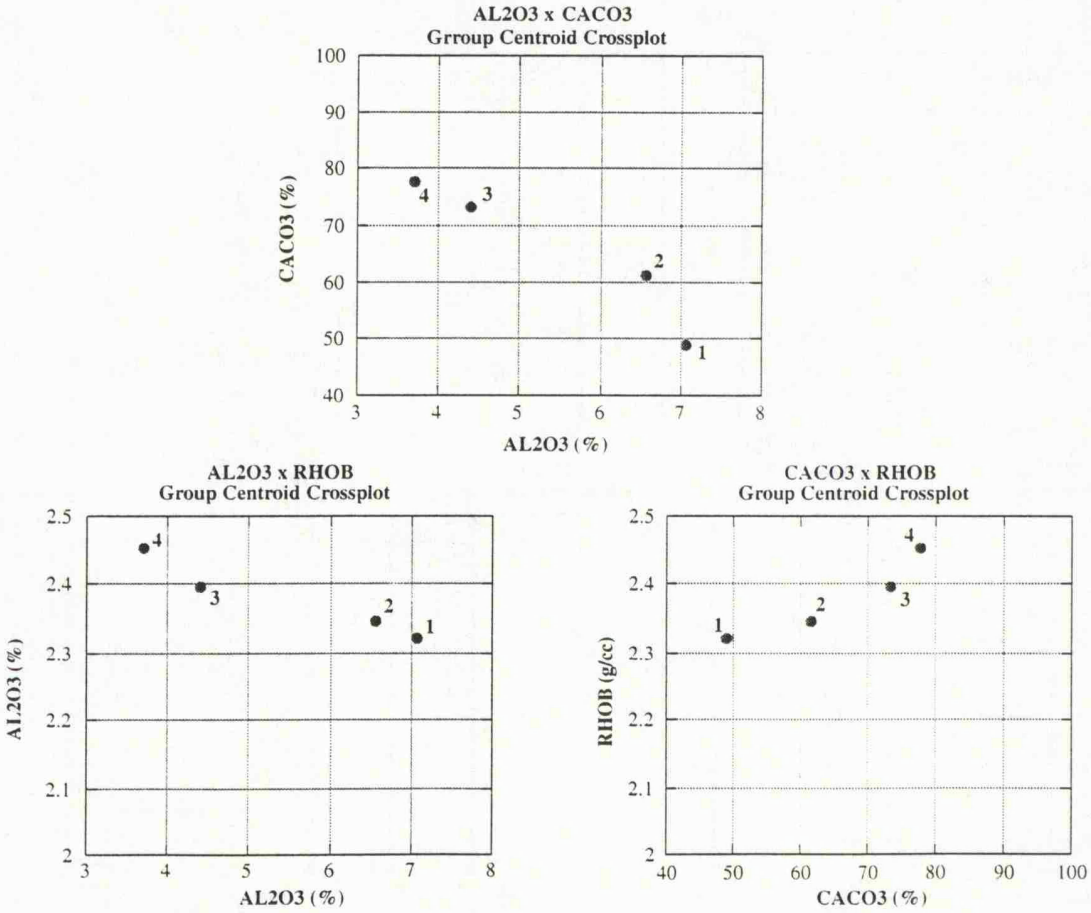


Figure 3.17 - Cross-plots showing the four group centroids in terms of:
a) AL2O3 x CACO3, b) AL2O3 x RHOB and c) CACO3 x RHOB.

shows negative principal component loading values. The second group, with positive loading values, is defined by THO, SIO2, TIO2, FEO, DT, AL2O3 and NPHI. Most of the log curves present dissimilarity values less than 0.20 showing a good correlation. The largest degree of dissimilarity is shown by URAN and POTA which also present the lower first principal component loadings.

Non-Hierarchical Cluster Analysis

The Non-Hierarchical Cluster Analysis was performed in ODP Hole 878A only using the original log curves. The reason for this is that the first principal component is responsible for approximately 65% of the total system variability and the two other components represent less than 20%. Therefore, using only the first principal component scores the results would appear very similar as using the original log curves. Nonetheless, the results for the analysis using the principal component scores I are shown in Table 3.23 and Figure 3.20 at the end of this section in order only to make a comparison between these results and the results using the original log curves. The Euclidean model and the iterative relocation as well as the Z-standardisation were also applied in this analysis.

a) Analysis using the original log curves

Initial testing for the number of groups resulted in four groups being selected as the *optimum* number to use for a full analysis. Table 3.19 shows the centroid composition of the groups produced. AL2O3, CACO3 and RHOB appear as the overall most important log curves in the analysis based on their average delta values (Table 3.20).

URAN and POTA are overall the least important log curves. The most important log curves for the discrimination between each group are also shown in Table 3.20. AL2O3 and CACO3 are the most important elements in the separation between Groups 1-2, 1-3, 1-4 and 2-3 while URAN and THO are the least important log curves. For the other groups separation (2-4 and 3-4) RHOB - AL2O3 and RHOB - SFLU are the most important ones.

Figure 3.17 shows the crossplots of the group centroids for the three overall most important log curves. It is observed that all crossplots show a reasonable separation between the four groups. The separation between the Si based matrix (Groups 1 and 2) and the Ca based matrix (Groups 3 and 4) into distinct zones in the plots can also be observed. Table 3.21 shows the simplest and the hardest combination of groups to distinguish. The lowest values are shown by the combinations of Groups 1-2 and Groups 3-4 with 1.23 and 1.52 means of delta values

respectively. As expected these are the hardest combinations to distinguish. Combinations 1-4 and 2-4 are the simplest combinations to distinguish with means of delta values equal to 4.16 and 3.28 respectively.

Table 3.19

Table of centroid composition

Log curves	Group 1	Group 2	Group 3	Group 4
NPHI	38.328	35.951	31.234	25.980
RHOB	2.322	2.347	2.397	2.454
IDPH	11.908	13.595	16.264	21.936
SFLU	11.440	13.526	16.777	23.849
CACO3	48.935	61.534	73.251	77.691
SIO2	27.243	17.015	11.396	9.667
THO	2.106	1.785	1.195	0.820
URAN	1.988	1.914	1.853	2.092
POTA	0.230	0.360	0.398	0.321
DT	93.152	89.293	83.967	78.491
AL2O3	7.076	6.562	4.402	3.702
FEO	7.333	6.441	4.821	3.901
TIO2	2.156	1.850	1.391	1.127

Table 3.19 - Centroid composition for Non-Hierarchical Cluster Analysis using the original log curves in ODP Hole 878A.

The group log is shown in Figure 3.18. It presents the group distribution over the whole interval in ODP Hole 878A. The upper section in the interval is dominated by Groups 1 and 2 where the principal component scores show positive values. They correspond to the Si based matrix. Where the principal component score values approach zero Group 2 is present while larger score values indicate Group 1. Between 563 - 577 mbsf there is an alternating sequence of Group 2 and Group 3 showing the same variation observed in the principal component scores (Figure 3.19). Group 3 is observed for negative score values and Group 2 appears when the values are positive. Below 577 mbsf the interval is dominated by Groups 3 and 4 as well as by strong negative scores. These are characteristics of the Ca based matrix, with the difference between the two groups given by the content of Ca in the matrix.

Groups 1 and 2 define the Si based matrix at the top of the interval with Group 2 related to the sections where carbonate/volcanic clasts and an increase in Ca content are observed. Groups 3 and 4 define the Ca based matrix with Group 3 showing the sections which have lower Ca content. Some variations in group distribution can be observed at 542 mbsf, which shows a Group 3 classification. It also corresponds to a negative value in the first principal component scores. The relation between the groups formed can also be observed when the clustering is carried out using the

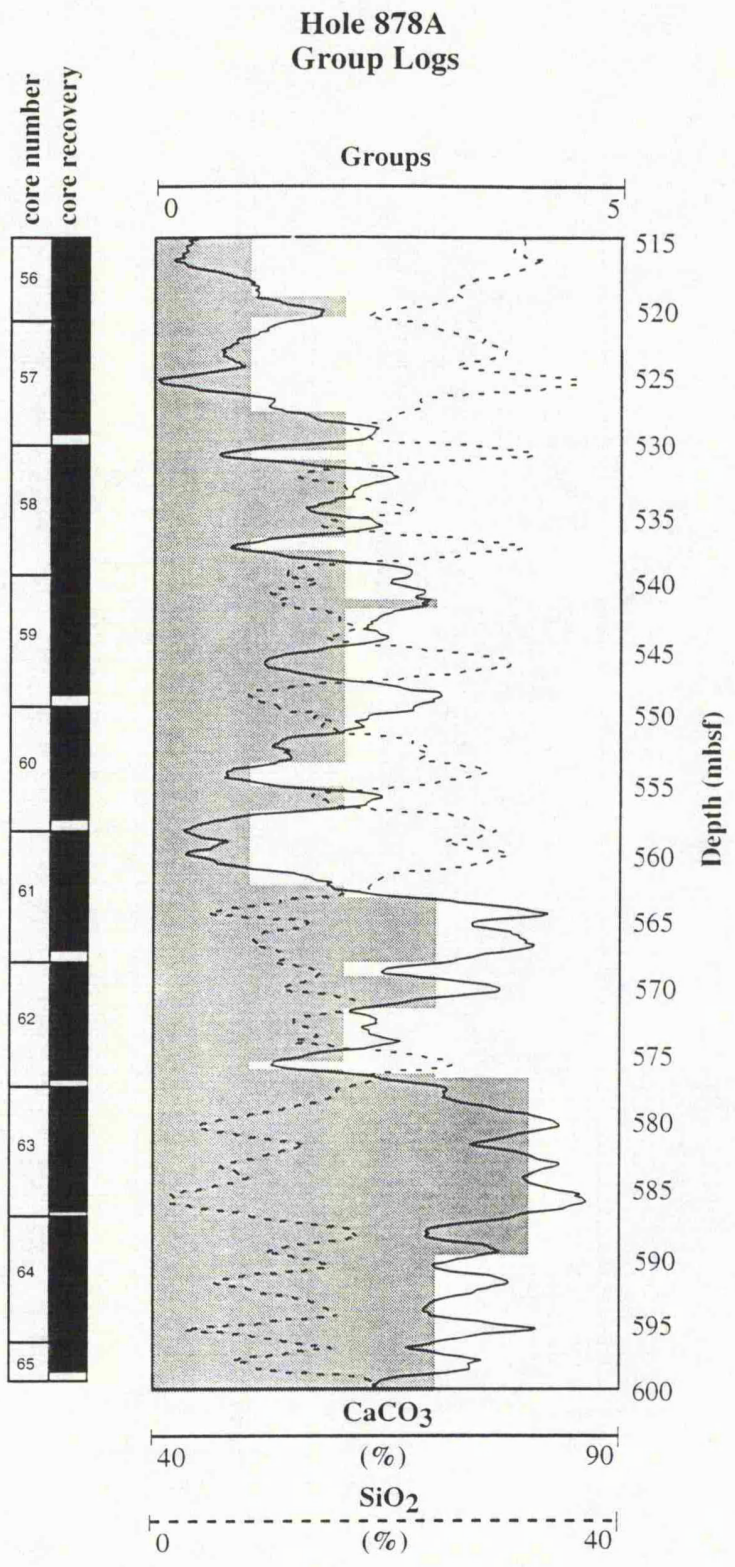


Figure 3.18 - Group log for Non-Hierarchical Cluster Analysis using the original log curves in ODP Hole 878A. $CaCO_3$ and SiO_2 curves are also plotted.

Table 3.20

Summary of delta values

URAN	Overall least important element	0.49	Mean
POTA		1.13	
...		...	
...		...	
CACO3		3.52	
AL2O3	Overall most important element	4.01	Mean
URAN	Worst element for Groups 1 & 2	0.24	Delta
THO		0.58	
...		...	
...		...	
SIO2		2.24	
CACO3	Best element for Groups 1 & 2	2.76	Delta
URAN	Worst element for Groups 1 & 3	0.40	Delta
THO		1.56	
...		...	
...		...	
AL2O3		4.53	
CACO3	Best element for Groups 1 & 3	5.51	Delta
URAN	Worst element for Groups 1 & 4	0.32	Delta
POTA		1.17	
...		...	
...		...	
CACO3		6.18	
AL2O3	Best element for Groups 1 & 4	6.96	Delta
URAN	Worst element for Groups 2 & 3	0.24	Delta
POTA		0.46	
...		...	
...		...	
CACO3		2.49	
AL2O3	Best element for Groups 2 & 3	3.94	Delta
POTA	Worst element for Groups 2 & 4	0.50	Delta
URAN		0.78	
...		...	
...		...	
RHOB		5.12	
AL2O3	Best element for Groups 2 & 4	6.31	Delta
SIO2	Worst element for Groups 3 & 4	0.41	Delta
THO		0.91	
...		...	
...		...	
RHOB		2.65	
SFLU	Best element for Groups 3 & 4	2.97	Delta

Table 3.20 - Importance of log curves in group discrimination in ODP Hole 878A.

groups obtained from the Non-Hierarchical Cluster Analysis. The results for all models are presented in Table 3.22 as well as the detailed description of the Simple

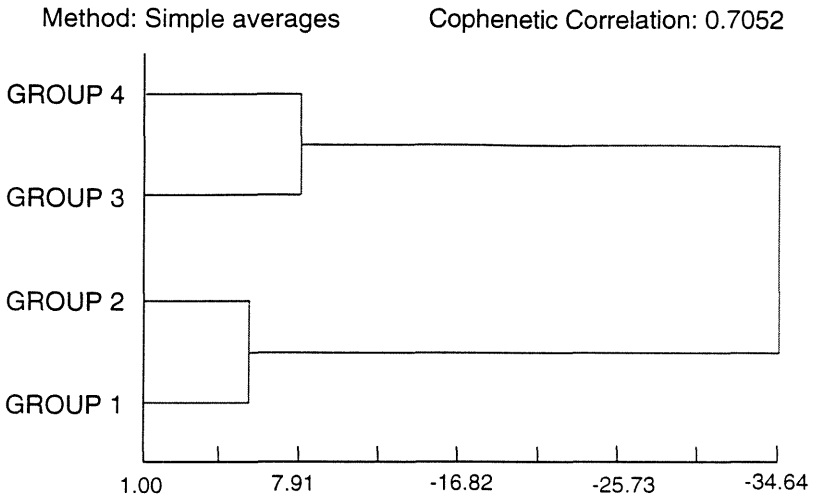


Figure 3.19 - Dendogram for Hierarchical Cluster Analysis using the groups produced by Non-Hierarchical Cluster Analysis in ODP Hole 878A.

Table 3.21

Groups 1 & 2	Hardest combination to distinguish	1.23 (Means)
Groups 3 & 4		1.52 (Means)
Groups 2 & 3		1.75 (Means)
Groups 1 & 3		2.90 (Means)
Groups 2 & 4		3.28 (Means)
Groups 1 & 4	Simplest combination to distinguish	4.16 (Means)

Table 3.21 - Hardest and Simplest combinations to distinguish in Non-Hierarchical Cluster Analysis (ODP Hole 878A).

averages model which present the best cophenetic correlation (0.7052). The dendogram in Figure 3.19 shows that Groups 1 and 2 and Groups 3 and 4 are related with approximately same dissimilarity units. The two sets of groups show a strong dissimilarity determined mainly by the contents of Si and Ca.

Table 3.22

Hierarchical Cluster Analysis (RCL) 878A.RCL
Euclidean distance matrix .... Z-standardised

	Group 1	**Group 2**	**Group 3**	**Group 4**
Group 1	0.000	7.130	32.290	66.209
Group 2		0.000	10.896	37.152
Group 3			0.000	10.615
Group 4				0.000

Model	**Cophenetic Correlation**	**Linkage method**
1	0.436	Nearest Neighbours
2	0.672	Furthest Neighbours
3	0.705	Simple Averages
4	0.699	Median
5	0.705	Group Averages
6	0.699	Centroid
7	0.683	Wards Method

Model: ....3 Simple Averages Cophenetic correlation ...0.705

link no.	**log curve retained**	**log curve deleted**	**similarity level**	**linkage order**
1	1	2	-5.130	1
2	3	4	-8.615	2
3	1	3	-34.637	3
				4

Table 3.22 - Hierarchical Cluster Analysis using groups extracted from Non-Hierarchical analysis (ODP Hole 878A).

Table 3.23

Groups 3 & 4	Hardest combination to distinguish	3.55 (Means)
Groups 1 & 2		3.55 (Means)
Groups 2 & 3		4.55 (Means)
Groups 2 & 4		7.62 (Means)
Groups 1 & 3		7.65 (Means)
Groups 1 & 4	Simplest combination to distinguish	10.06 (Means)

Table 3.23 - Hardest and simplest combinations to distinguish in Non-Hierarchical Cluster Analysis using principal component scores (ODP Hole 878A).

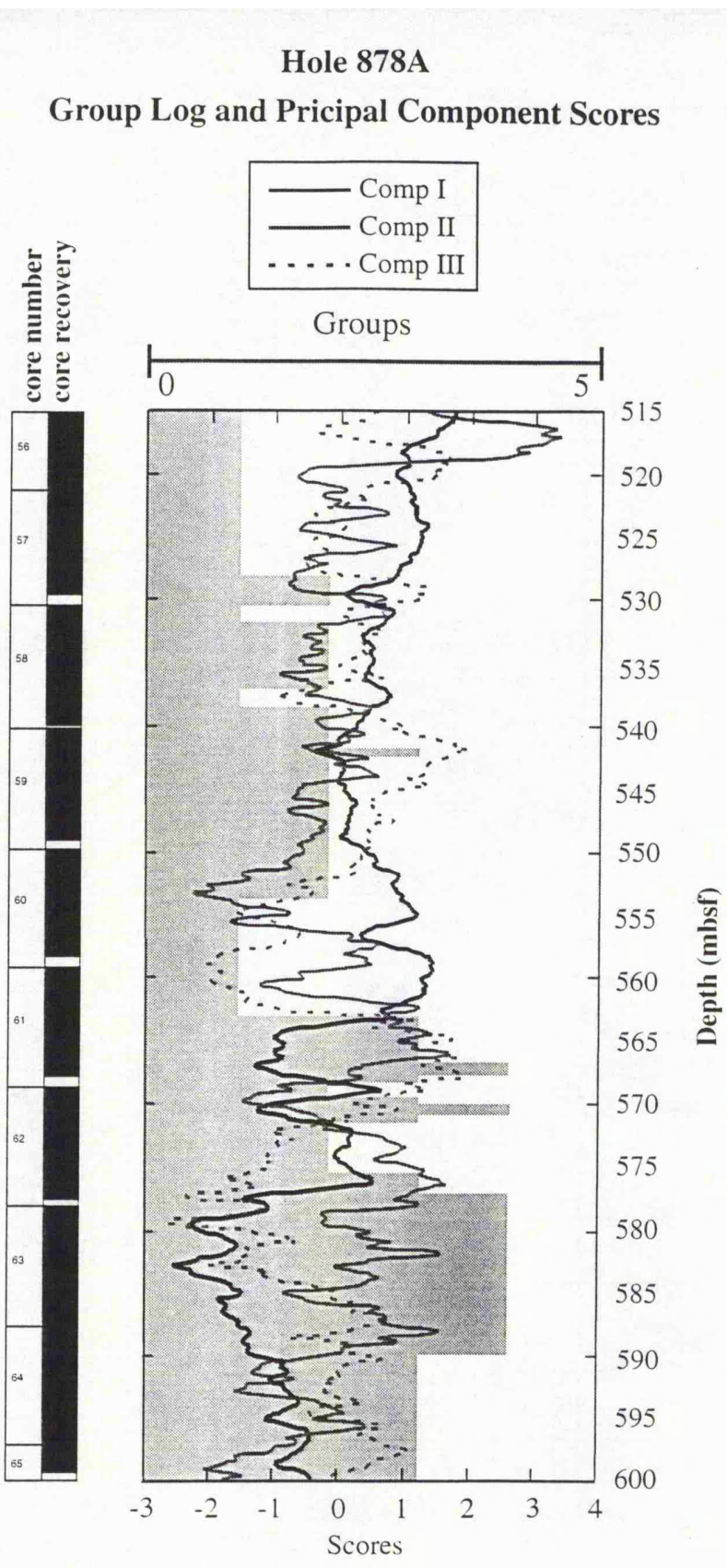


Figure 3.20 - Group log for Non-Hierarchical Cluster Analysis using the Principal Component scores I, II and III.

3.5 - Discriminant Analysis

3.5.1 - Introduction

Another important technique is the Discriminant Analysis. In this case, the approach differs completely from the use of Principal Component Analysis and Cluster Analysis. Lithofacies are no longer defined implicitly from the data structure but are identified *a priori* by core descriptions or other geological information. Therefore, a discriminant function can be developed in order to distinguish separate lithofacies using the log responses of their contained zones. Where the discriminant function effectively distinguish the lithofacies, sections in intervals that are not cored or from where no geological information is available can be assigned to one or other of these lithofacies considering only the log responses. It must be noted, however, that core physical properties are not always accurate representation of formation properties. Some caution must be exercised when using them for such analysis.

In this work, Discriminant Analysis are used to distinguish lithofacies within the same hole and also in an offset hole. Different methods of classification are tested and discussed for each of the examples. The results are then compared with the ones obtained for the Cluster Analysis.

3.5.2 - Principles

Discriminant Analysis is one of the most widely used multivariate techniques in geology because of its relative simplicity and robustness. This technique has been applied with reasonable success in many geological studies including well log analysis (Chaynes, 1964; Davis, 1986; Doveton, 1994).

It is a multivariate statistical technique in which the original log curves are combined in such a way to maximise differences between two pre-determined clusters or groups. Therefore, the need of *a priori* information distinguishes it from the Cluster Analysis described in the last section. The Discriminant Function (DF) computed from the groups previously defined (or their training sets) may then be used to allocate new samples of unknown origin to one of these groups (Doveton, 1986 and1994). The DF transforms a n original set of measurements on a sample or group of samples into discriminant scores. These scores then represent the position of the sample along the line defined by the DF in, for instance, the case of two groups. In the case of more than two groups, Multivariate Discriminant Analysis, in contrast, provides for the simultaneous comparison of several groups in multidimensional space.

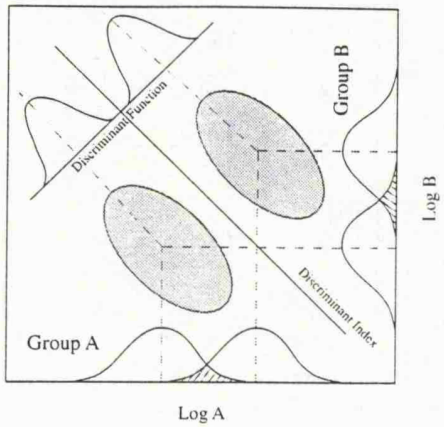


Figure 3.21 - Hypothetical two-dimensional example of the Linear Discriminant Function (LDF) (after Davis, 1986).

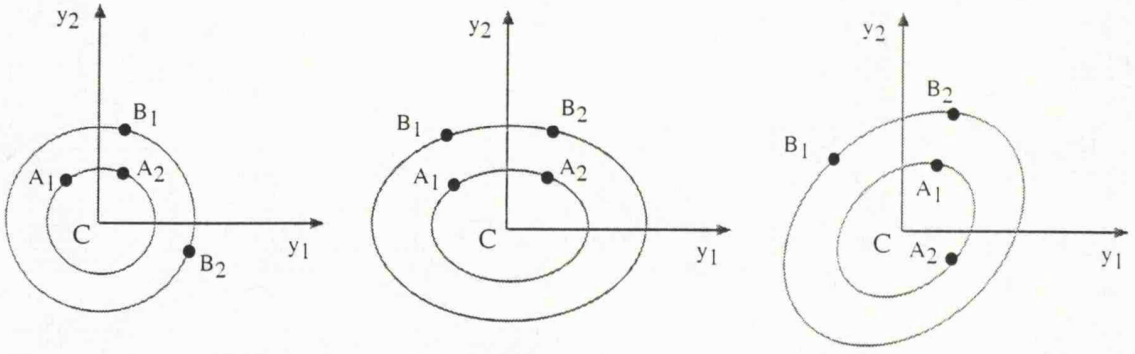


Figure 3.21a - Euclidean measure, Mahalanobis measure and Group Generalised distance measure (after Gnanadesikan, 1977).

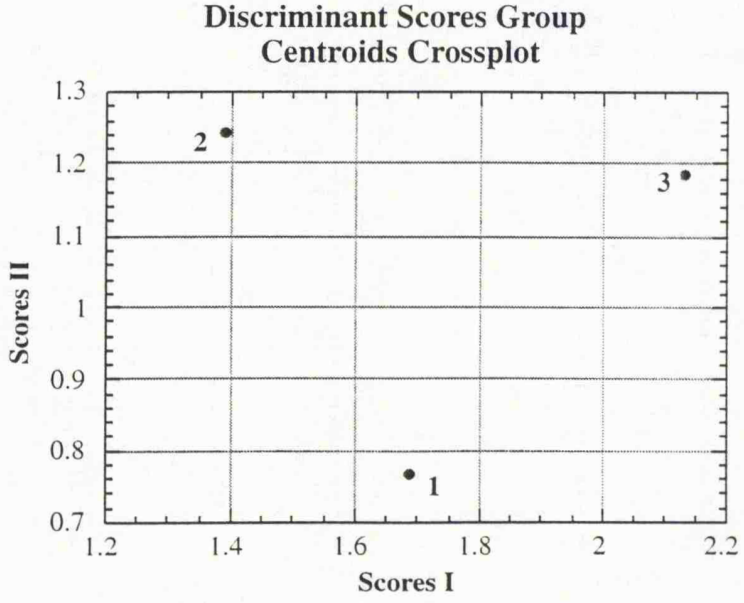


Figure 3.22 - Cross-plot of Discriminant scores showing the group centroids in ODP Hole 807C.

Figure 3.21 diagramatically shows a hypothetical two-dimensional example of the DF. No adequate separation may be made using either log curve Log A or Log B. The groups may however be distinguished by projecting members of the two groups onto the discriminant function line (Davis, 1986). This line is the one orthogonal to the line which best separates the two groups and minimises the degree of misclassification.

The equation which describes the DF has the form:

$$R = \lambda 1.\psi 1 + \lambda 2.\psi 2 + ... + \lambda m.\psi m \qquad (1)$$

where R is the discriminant score in the sample, λm is the coefficient associated with the m^{th} variable and ψm is the mean value of the m^{th} variable.

This can be solved through:

$$[D] = [Sp^2] . [\lambda]$$

where [D] is the column vector of m differences between the means of two groups:

$$Dj = \mathbf{A}j - \mathbf{B}j = ((\Sigma_{1,n} Aij)/n) - ((\Sigma_{1,n} Bij)/n)$$

where $\mathbf{A}j$ is the mean of the j^{th} variable in group A, and $\mathbf{B}j$ is the mean of the j^{th} variable in group B. $[Sp^2]$, which is the m x m matrix of the pooled variances and covariance of the m variables, is described as:

$$[Sp^2] = ([SpA] - [SpB]) / (n_1 + n_2 - 2)$$

with SpA the matrix of sum of squares and cross products of all variables in [A] and n_1 is the number of samples in A. The same is for group B. $[\lambda]$ is the column vector of m coefficients of the discriminant function.

At the end, the equation can be solved by inversion and multiplication where:

$$[\lambda] = [Sp^2]^{-1} [D] .$$

The values of λ may now be inserted for use in the equation 1.

The discriminant index R_0 is the point along the DF line which is exactly half-way between the centre of groups A and B. Its value may be found by the substitution of the midpoints between the two group means into each ψ_m position of the equation 1:

$$(\mathbf{A}j + \mathbf{B}j) / 2 .$$

Similarly the values of R_A and R_B, the discriminant scores describing the centres of groups A and B, can be found in the same way.

The relative contribution of individual variables to the DF may be measured by using the quantity E_j, where:

$$E_j = (\lambda_j \, . \, D_j) / D^2$$

λ_j is the coefficient of DF as described earlier, Dj is the difference between the j^{th} means of the two groups, and D^2 is the Mahalanobis distance, which is a measure of the separation between the two multivariate means expressed in units of the pooled variance (R_A - R_B) (Sheppard, 1986).

Given a space to represent the objects (depth measurements), the fundamental problem in classification is reduced to choosing a metric or a distance measure. For, if such a metric is available, an object which needs to be assigned to one of the groups may be identified to the group which it is closest as judged by the metric. From the point of view of well log analysis, the prescription of a distance function will generally be a trial and error task in which the use of some general techniques needs to be aided by other geological information such as core data (Gnanadesikan, 1977).

One useful general class of squared distance functions is provided by a class of positive semidefinite quadratic forms. Specifically, if $\mathbf{u}'=(u_1, u_2,...,u_p)$ denotes the p-dimensional observation on an object that is to be assigned to one of the prespecified group, then for measuring the squared distance between $\mathbf{u}$ and the centroid of the i^{th} group, one may consider the Mahalanobis distance

$$D^2(i) = (\mathbf{u}-\boldsymbol{y}_1)' \, \mathbb{M} \, (u-\boldsymbol{y}_2),$$

where $\mathbb{M}$ is a positive semidefinite matrix to ensure that $D^2(i) > 0$. $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are the object mean vectors of each group. The object will be assigned to the group for which $D^2(i)$ is smallest as i takes from the values from one through a g number of groups. Different choices of the matrix $\mathbb{M}$ lead to a different metrics, and the class of squared distance functions represented by the equation above is not unduly narrow.

Thus, when $\mathbb{M} = \mathbb{I}$, one obtains the familiar Euclidean squared distance between the unknown sample to be classified and the centroid of the i^{th} group in the p dimensional space of response. Geometrically, as shown in Figure 3.21a for the case when p=2, the use of such measure of squared distance amounts to measuring distances by circles (or spheres when p>2). Points A_1 and A_2 lying on the same circle are considered to be the same distance away from the centre C, while points B_1 and B_2 lying on the outer circle are considered to be farther away from C than are A_1 and A_2.

Another way of accomplishing this would be to use "elliptical" (or ellipsoidal) distance measures as shown in Figure 3.21a. Again A_1 and A_2 are considered to be equidistant from C, while B_1 and B_2 are considered to be farther from C than A_1 and A_2. Algebraically, this measure of squared distance corresponds to specifying $\mathbf{M}$ in the equation to be a diagonal matrix. Still another extension of the distance measure may be made to accommodate intercorrelations among the responses as well as possible differences among their variances. When p=2 and the statistical correlations between y1 and y2 is positive, the right part on Figure 3.21a shows how one may use "elliptical" distance measures by tilting the ellipses so that their major axis is oriented in a direction reflecting the positive correlation. This is called the Generalised squared distance measure.

Of considerable importance to the application of the technique is the fact that prior to its use a suitable training set must be selected which represents each class. Also the inclusion of measurements which poorly distinguish between groups may outweigh the beneficial effects of good discriminators. For this reason they should be removed at the earliest opportunity (Doveton, 1994).

In the analysis of log data, the training groups will be selected from the sections along the whole interval which best represent a lithofacies or geological sequence. It means that most of the log curves of the sections selected should have distinguishable values for the lithofacies they represent. With this purpose the same datasets studied with the previous techniques are used and the results compared at the end of this chapter. Another data set, now from two oilfield holes, are used to demonstrate the efficiency of this technique when applied in an offset well.

3.5.3 - Application of Discriminant Analysis to well log data

3.5.3.1 - ODP Hole 807C

A multigroup discriminant function analysis was used to distinguish the three units present in the interval between 1270-1400 mbsf in ODP Hole 807C. The discrimination is made in terms of their well log responses. The main purposes for the analysis are: to assess whether the distinction was even possible, based only on the log curves; to verify which ones appear to contribute to any such distinction; and if possible, to determine the geological reasons that would account for any lithological or petrophysical distinctions. If the discriminant function was found to be useful then the functions could be used for classification purposes in other parts of the interval or other holes.

The log curves used for discrimination were all curves used in Cluster Analysis (section 3.3.3.1). The training set selected was based on the result of the

Principal Component and the Cluster Analysis. The intervals selected for that are: between 1300 - 1310 mbsf in the carbonate sequence (TR_SET_1), between 1352 - 1358 mbsf in the claystone/siltstone sequence (TR_SET_2) and between 1382 - 1387 mbsf in the volcanic sequence (TR_SET_3). In the choice of the best sections for the training sets, we consider in general short intervals in order to reduce the processing time. We also compare the summary statistics of each interval considered as a training set (Table 3.24).

Table 3.24

Summary statistics for training set 1		
Log curve	**Mean**	**Standard deviation**
GRAY	6.036	1.078
POTA	0.119	0.053
THO	0.310	0.257
URAN	0.405	0.154
IDPH	3.037	0.206
ALUM	0.462	0.080
RHOB	2.293	0.045
PEF	3.953	0.436
Summary statistics for training set 2		
Log curve	**Mean**	**Standard deviation**
GRAY	45.022	15.821
POTA	1.239	0.439
THO	4.892	1.707
URAN	1.072	0.631
IDPH	1.216	0.211
ALUM	1.792	0.967
RHOB	1.960	0.058
PEF	2.109	0.346
Summary statistics for training set 3		
Log curve	**Mean**	**Standard deviation**
GRAY	7.722	2.049
POTA	0.231	0.098
THO	0.407	0.280
URAN	0.352	0.248
IDPH	568.813	801.505
ALUM	3.874	0.878
RHOB	2.821	0.076
PEF	5.203	0.359

Table 3.24 - Summary statistics for the training sets of each of the lithofacies in ODP Hole 807C.

The result for this analysis is shown in Table 3.25. The computed F-test value, based on the Mahalanobis' distance, proved to be highly significant, rejecting the null

hypothesis that there were no appreciable differences between the three groups. The Table also shows the discriminant functions for each log curve and the discriminant score distribution for each group centroid. Both discriminant functions I and II show that RHOB is the most important log curve in this analysis with values equal to 0.68 and 0.54 respectively. Figure 3.22 shows the crossplot of the scores for the group centroids. A good separation between the groups is clearly seen. None of the samples in each group was misclassified and the discrimination can be considered as excellent.

Table 3.25

Multiple Discriminant Analysis

	Discriminant functions	
Log curve	**DF # 1**	**DF # 2**
GRAY	0.038	0.053
POTA	-0.643	-0.656
THO	-0.137	-0.156
URAN	-0.307	-0.476
IDPH	0.000	0.000
ALUM	0.016	0.081
RHOB	0.685	0.540
PEF	0.029	-0.129

	Discriminant scores for the group centroids	
Groups	**DF # 1**	**DF # 2**
Group 1	1.686	0.768
Group 2	1.387	1.244
Group 3	2.133	1.186

Total number of cases 158
Number misclassified 0
Percentage misclassified 0.00

F-ratio for overall discrimination 278.566
Degrees of freedom .....16 and 296

Computed level of probability for F 0.00

Ho: Group centroid = common centroid is rejected at alpha=0.001
The group discrimination is excellent

Table 3.25 - Result for the Discriminant Analysis in ODP Hole 807C.

Hole 807C
Discriminant Scores Log

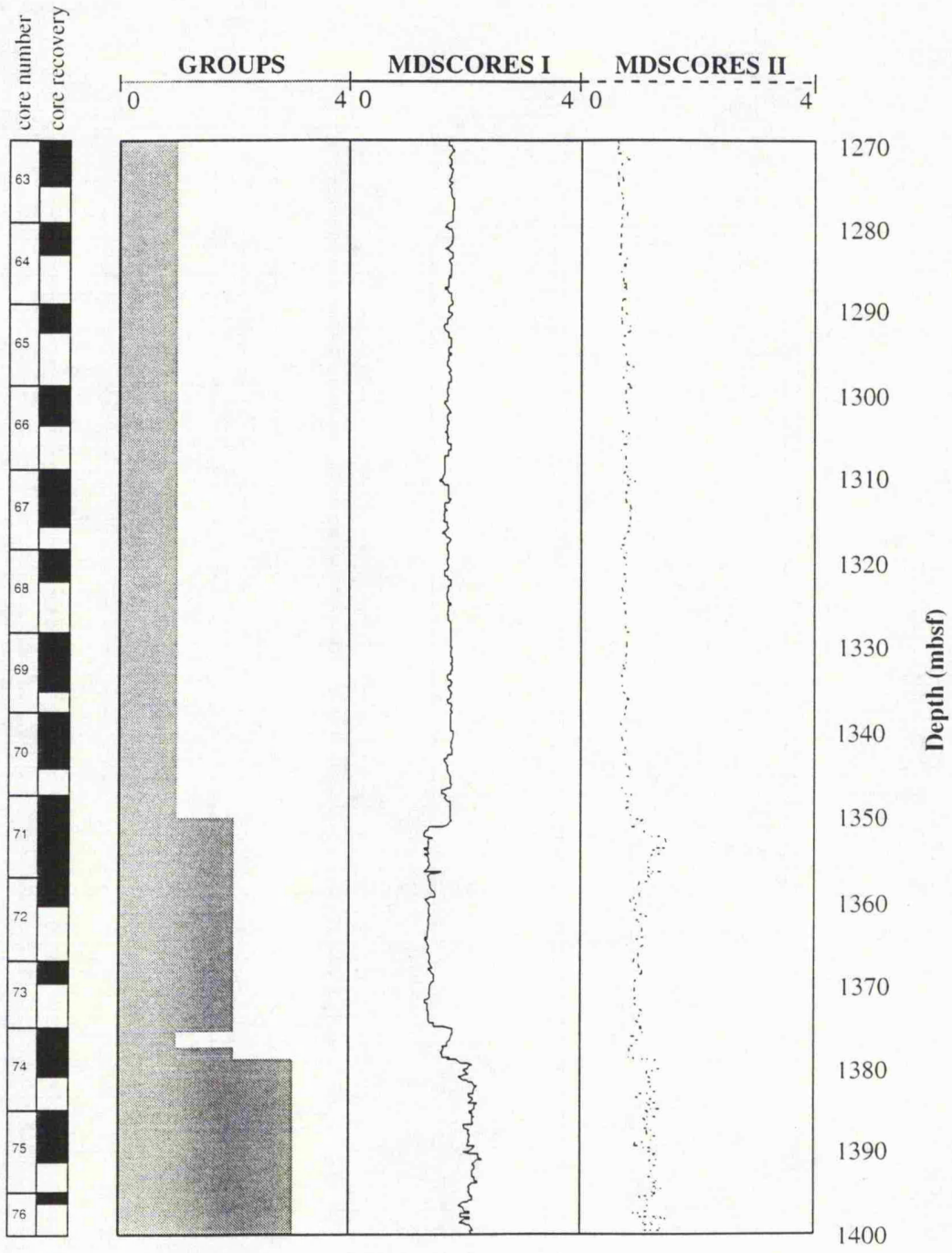


Figure 3.23 - Discriminant scores log curves in ODP Hole 807C. Groups from the Cluster Analysis are also plotted for comparison.

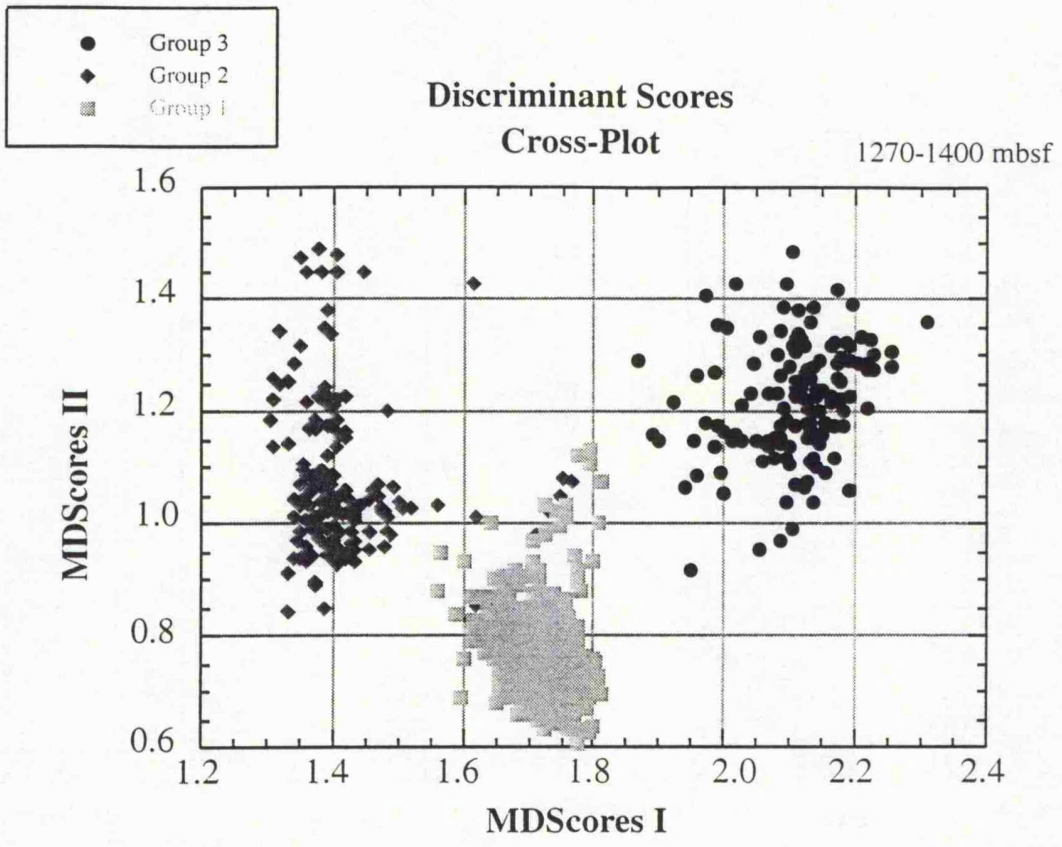


Figure 3.24 - Cross-plot of the discriminant scores in ODP Hole 807C.

After the processing of the discriminant functions using the training sets, we applied them for the whole interval in ODP Hole 807C. The result is not surprising in terms of the Z-score logs, which show an almost perfect separation between the three lithofacies (Figure 3.23). For the scores I, values between 1.6 and 1.8 are associated

Table 3.26

EUCLIDEAN DISTANCE CLASSIFICATION

Classification Summary

Group number	Actual number in group	Number assigned to group	Number correct in group	% correct for group
1	548	694	527	96.17
2	167	134	108	64.67
3	138	25	25	18.12
Totals	853		660	77.37

MAHALANOBIS DISTANCE CLASSIFICATION

Classification Summary

Group number	Actual number in group	Number assigned to group	Number correct in group	% correct for group
1	548	552	517	94.34
2	167	165	134	80.24
3	138	136	135	97.83
Totals	853		786	92.15

GROUP GENERALISED DISTANCE CLASSIFICATION

Classification Summary

Group number	Actual number in group	Number assigned to group	Number correct in group	% correct for group
1	548	491	491	89.60
2	167	225	167	100.00
3	138	137	137	99.28
Totals	853		795	93.20

Table 3.26 - General distance classification using three different models in ODP Hole 807C.

with the carbonate sequence at the top of the interval. Values below 1.5 are associated with the claystone/siltstone sequence and values around and greater than 2.0 are associated to the volcanic sequence at the bottom of the interval. In terms of scores II, it is still possible to see the separation between the three lithofacies. However, the

interval between 1352 - 1357 mbsf in the claystone/siltstone sequence shows values as high as the volcanic sequence. The groups obtained in the Cluster Analysis are also shown in order to make a comparison with the Discriminant Analysis. The general distance classification for the whole data can be seen in Table 3.26. The summary statistics for the three lithofacies and the result for three different classification models are shown. The Euclidean distance models show a relatively high percentage of disagreement with 22.63% of misclassified samples. However, this value is given mostly because of the low percentage of correct samples in Group 3. This is due to the similar values in density between this lithofacies and the carbonate sequence, to where most of the samples were misclassified. The other two models seem to be more efficient in the analysis with a total of 92.15% and 93.20% of correct samples for the total data set.

A plot of the two multigroup discriminant scores can be seen in Figure 3.24. It also shows clearly the separation between the three main lithofacies. It appears much better when compared with the principal component scores. The carbonate sequence appears as a dense cluster in the middle of the graph. The other two lithologies can also be easily detected. However, variations in log values, especially RHOB, within the intervals occupied by these lithofacies lead to a more scattered distribution of these cross-plots.

3.5.3.2 - ODP Hole 878A

In ODP Hole 878A, a Discriminant Analysis was carried out to distinguish between the two different matrix based breccias present within the interval 515 - 600 mbsf. The log measurements used in this case are the same as used in the Principal Components and Cluster Analysis (sections 3.3.1 and 3.4.3 respectively).

The training set selected in this method is also based on the comparison of the summary statistics for each group (Table 3.27) and using the minimum amount of samples to reduce the processing time. The training sets for both facies are as follow: TR_SET_1 (Si based matrix) 532 - 539 mbsf and TR_SET_2 (Ca based matrix) 579 - 582 and 587 - 591 mbsf.

The result for this analysis can be observed in Table 3.28, where the F-test rejected the null hypothesis that there was no great difference between the two lithofacies. The discriminant function shows that RHOB is the most important log curve in this analysis with other log curves such as POTA and TIO2 also being important. The discriminant scores for the group centroid do not show a strong difference in their values but they look enough different to perform the distinction between the lithofacies. As a final result, the group discrimination is also considered excellent.

Table 3.27

Summary statistics for training set 1		
Log curve	**Mean**	**Standard deviation**
NPHI	34.937	1.609
RHOB	2.348	0.006
IDPH	12.652	0.354
SFLU	12.506	0.663
CACO3	59.772	5.241
SIO2	18.720	5.666
THO	2.125	0.257
URAN	1.906	0.123
POTA	0.334	0.039
DT	89.978	1.186
AL2O3	6.466	0.427
FEO	6.487	0.303
TIO2	1.971	0.185
Summary statistics for training set 2		
Log curve	**Mean**	**Standard deviation**
NPHI	26.552	3.830
RHOB	2.446	0.024
IDPH	21.487	3.661
SFLU	23.015	3.824
CACO3	75.896	4.221
SIO2	10.826	3.718
THO	0.847	0.412
URAN	2.028	0.274
POTA	0.353	0.067
DT	79.151	4.130
AL2O3	3.892	0.382
FEO	4.101	0.520
TIO2	1.186	0.269

Table 3.27 - Summary statistics for the training sets of each of the lithofacies in ODP Hole 878A.

The Discriminant Function was then applied for the whole interval. The linear discriminant scores (Figure 3.25) show a reasonable separation between the two lithofacies. The Si based matrix section can be seen to have low score values while the Ca based matrix has a high score value, in general higher than 8.0. The general distance classification for the whole data can be seen in Table 3.29 together with the summary statistics for both lithofacies and three different distance classification models which show very low percentage of misclassified samples. The Euclidean distance model classifies 93.73% of the samples correct while the Mahalanobis

distance and the Group Generalised distance models, as observed in ODP Hole 807C, present better results with 99.28% and 99.46% of correct samples respectively.

Table 3.28

Multiple Discriminant Analysis

Discriminant functions	
Log curves	**DF # 1**
NPHI	0.003
RHOB	0.860
IDPH	-0.020
SFLU	0.019
CACO3	0.061
SIO2	0.058
THO	-0.047
URAN	-0.061
POTA	0.481
DT	0.008
AL2O3	0.008
FEO	-0.016
TIO2	0.116

Discriminant scores for the group centroids	
Groups	**DF # 1**
Group 1	7.838
Group 2	8.347

Total number of cases 112
Number misclassified 0
Percentage misclassified 0.00

F-ratio for overall discrimination 303.447
Degrees of freedom .....13 and 98

Computed level of probability for F 0.00

Ho: Group centroid = common centroid is rejected at alpha=0.001
The group discrimination is excellent.

Table 3.28 - Result for the Discriminant Analysis in ODP Hole 878A.

Table 3.29

EUCLIDEAN DISTANCE CLASSIFICATION

Classification Summary

Group number	Actual number in group	Number assigned to group	Number correct in group	% correct for group
1	365	342	336	92.05
2	193	216	187	96.89
Totals	558		523	93.73

MAHALANOBIS DISTANCE CLASSIFICATION

Classification Summary

Group number	Actual number in group	Number assigned to group	Number correct in group	% correct for group
1	365	361	361	98.90
2	193	197	193	100.00
Totals	558		554	99.28

GROUP GENERALISED DISTANCE CLASSIFICATION

Classification Summary

Group number	Actual number in group	Number assigned to group	Number correct in group	% correct for group
1	365	362	362	99.18
2	193	196	193	100.00
Totals	558		555	99.46

Table 3.29 - General distance classification using three different models in ODP Hole 878A.

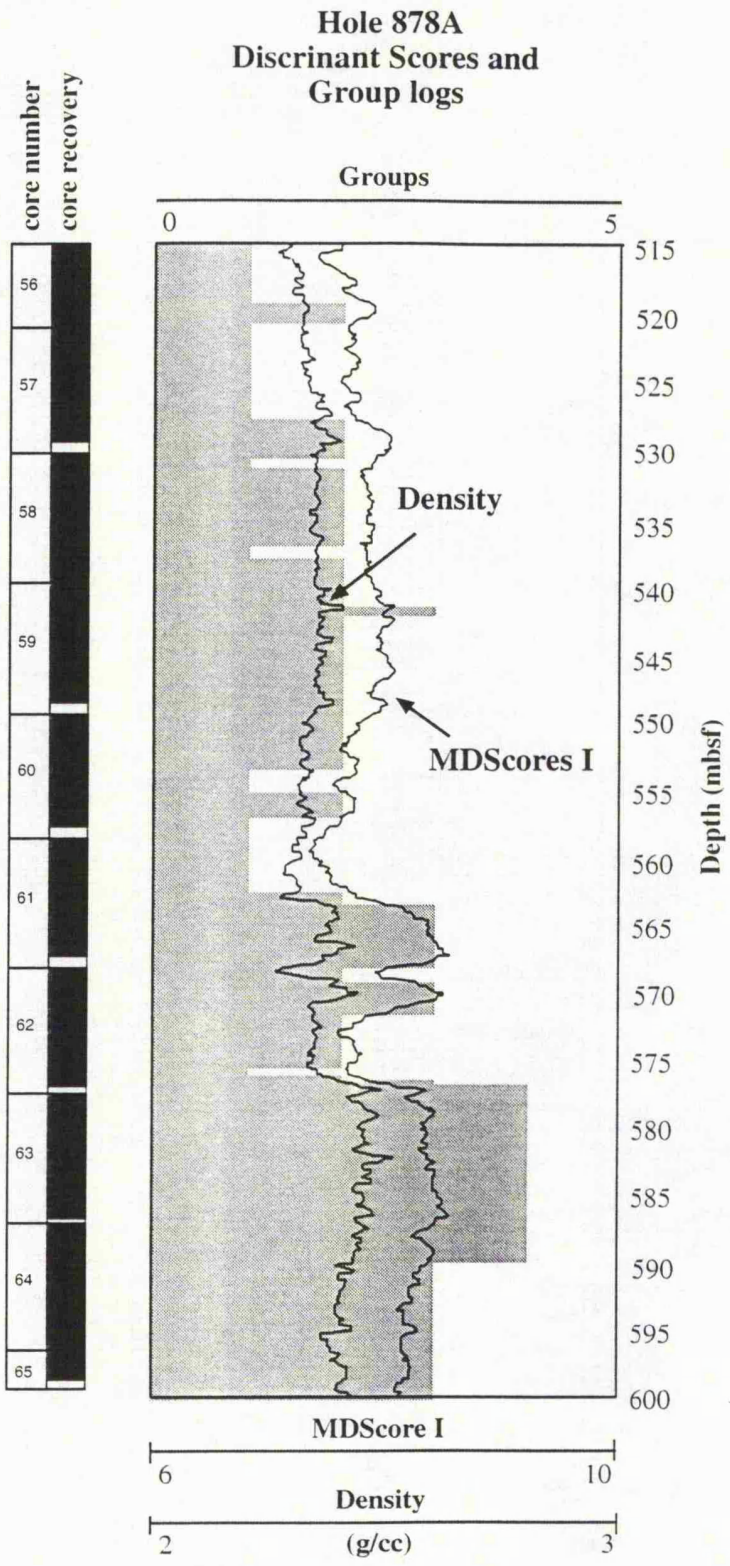


Figure 3.25 - Discriminant scores log for ODP Hole 878A. Groups from Cluster Analysis and the density log are also plotted for comparison.

3.5.3.3 - Oilfield Holes A and B

A multigroup Discriminant Analysis was used in this case to obtain a lithofacies classification in Hole B using the discrimination based in terms of well log responses in the nearby Hole A. The purpose for this analysis is to check whether the lithofacies distinction was possible, based only on the log curves of a nearby hole. The challenge is to obtain a reasonable lithofacies classification and to verify how changes in facies across the holes affect the classification.

The six log curves in Hole A were used for discrimination. The training sets were selected based on the result of the Cluster Analysis (Figure 3.26) which found two groups in this hole. The intervals selected are: between X480 - X490 metres for the reservoir zone (Group 1) and between X450 - X460 and X510 - X515 metres for the rest of the interval (Group 2). Short intervals were considered in order to reduce the processing time. Bearing in mind that changes in facies between Oilfield Holes A and B are always possible to occur, the training sets must have different characteristics to allow good discrimination function to be obtained. The characteristics of each training set are summarised in Table 3.30.

The result for the discrimination function is shown in Table 3.31. The computed F-test value, based on the Mahalanobis' distance, proved to be highly significant, rejecting the null hypothesis about differences between the two groups. The Table also shows the discriminant functions for each log curve and the discriminant score distribution for each group centroid. The discriminant function shows that RHOB is the most important log curve in the separation of both groups with value of 0.999 while the other log curves have values around zero. The two training sets were considered excellent for the separation of the lithofacies observed in Hole A. Presuming that the same lithofacies occur in Hole B, the discriminant function obtained here should work in the nearby hole. However, in order to test how efficiently the discriminants work, they were firstly applied to the whole interval in Hole A and the result in terms of the discriminant score logs is seen in Figure 3.26.

After testing the discriminant functions in Hole A with a reasonable success they were applied to Hole B and the result in terms of the discriminant scores log is observed in Figure 3.27. Multigroup discriminant scores show high values ($\approx$ 2.75) associated with the shale sequence and low values ($\approx$ 2.55) associated with the sandstone reservoir. The discriminant functions obtained from Hole A seem to characterise well the two main lithofacies in Hole B. It is possible to observe the interbedded variation between the sandstone reservoir and the shale between X325 - X354 metres. However, when this result is compared with the one obtained for Cluster Analysis, it is observed that the discriminant function in Hole A does not

complete describe the changes in facies in Hole B. That technique suggested three different groups: Group 3 as the sandstone reservoir and Groups 1 and 2 as the shale/claystone sequence. Group 1 only appears at the upper part of the interval, indicating some variation in facies for that section. Therefore, it appears that the training sets in Hole A were not able to identify this facies variation at the upper section in Hole B, resulting in the same value of discriminant scores for all shale/claystone sequence over the interval.

Table 3.30

Summary statistics for training set 1

Log curve	Mean	Standard deviation
RHOB	2.335	0.028
ILD	1.128	0.106
ILM	1.217	0.117
NPHI	36.712	2.127
GRAY	67.813	4.256
DT	112.957	2.949

Summary statistics for training set 2

Log curve	Mean	Standard deviation
RHOB	2.080	0.028
ILD	32.818	19.646
ILM	86.762	184.559
NPHI	33.672	1.333
GRAY	38.343	5.018
DT	126.535	4.988

Table 3.30 - Summary statistics for the training sets in Hole A.

Table 3.31

Multiple Discriminant Analysis

Discriminant functions

Log curve	DF # 1
RHOB	0.99998
ILD	-0.00012
ILM	0.00000
NPHI	0.00597
GRAY	-0.00089
DT	0.00262

Discriminant scores for the group centroids

Groups	DF # 1
Group 1	2.789
Group 2	2.574

Total number of cases 118
Number misclassified 0
Percentage misclassified 0.00

F-ratio for overall discrimination 511.233
Degrees of freedom ..... 6 and 111

Computed level of probability for F 0.00

Ho: Group centroid = common centroid is rejected at alpha=0.001
The group discrimination is excellent

Table 3.31 - Result for the Discriminant Analysis using the training sets in Hole A.

Hole A

Principal Component Scores
Groups and Discriminant Scores

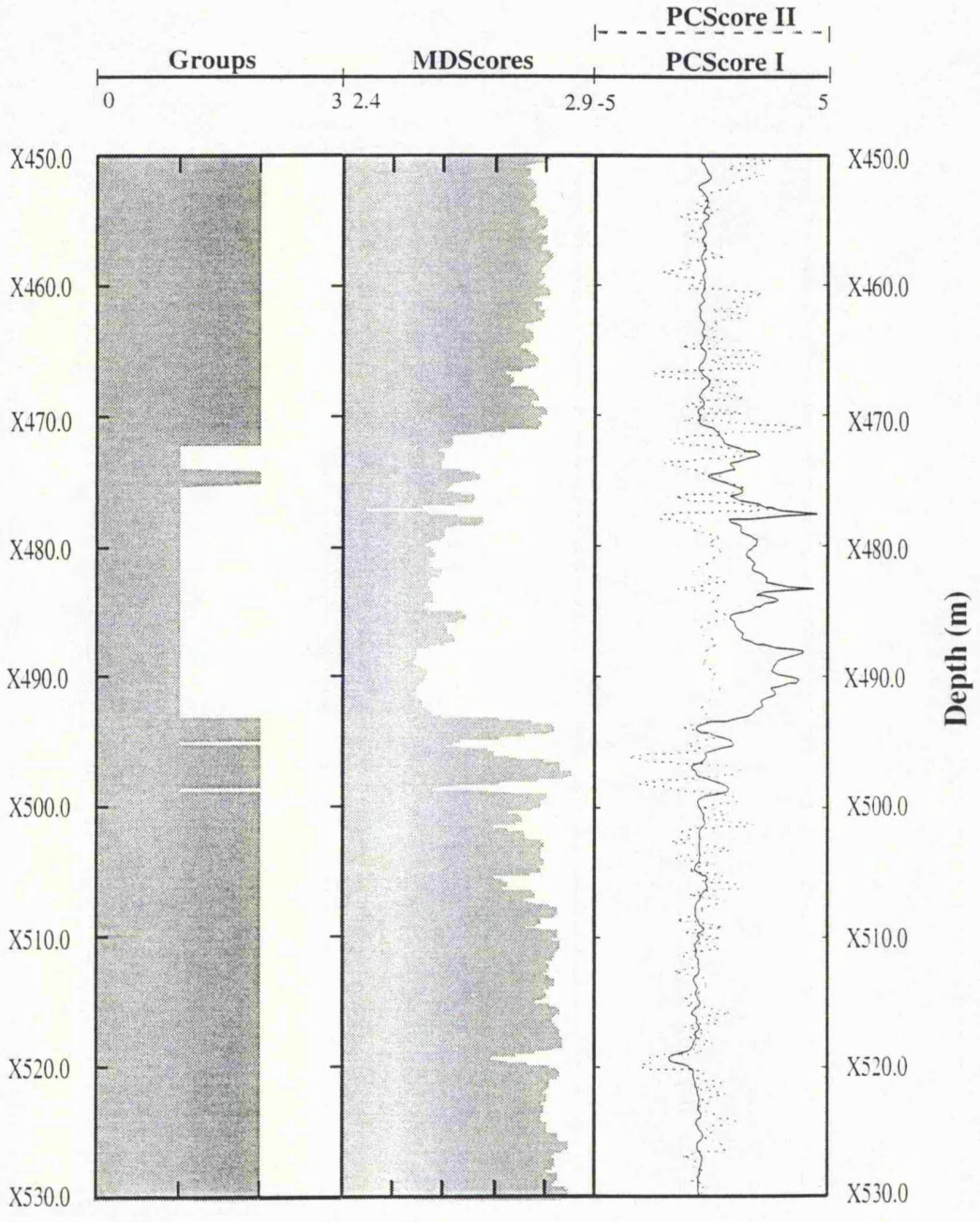


Figure 3.26 - Principal Component Scores, Groups and Discriminant scores for Hole A.

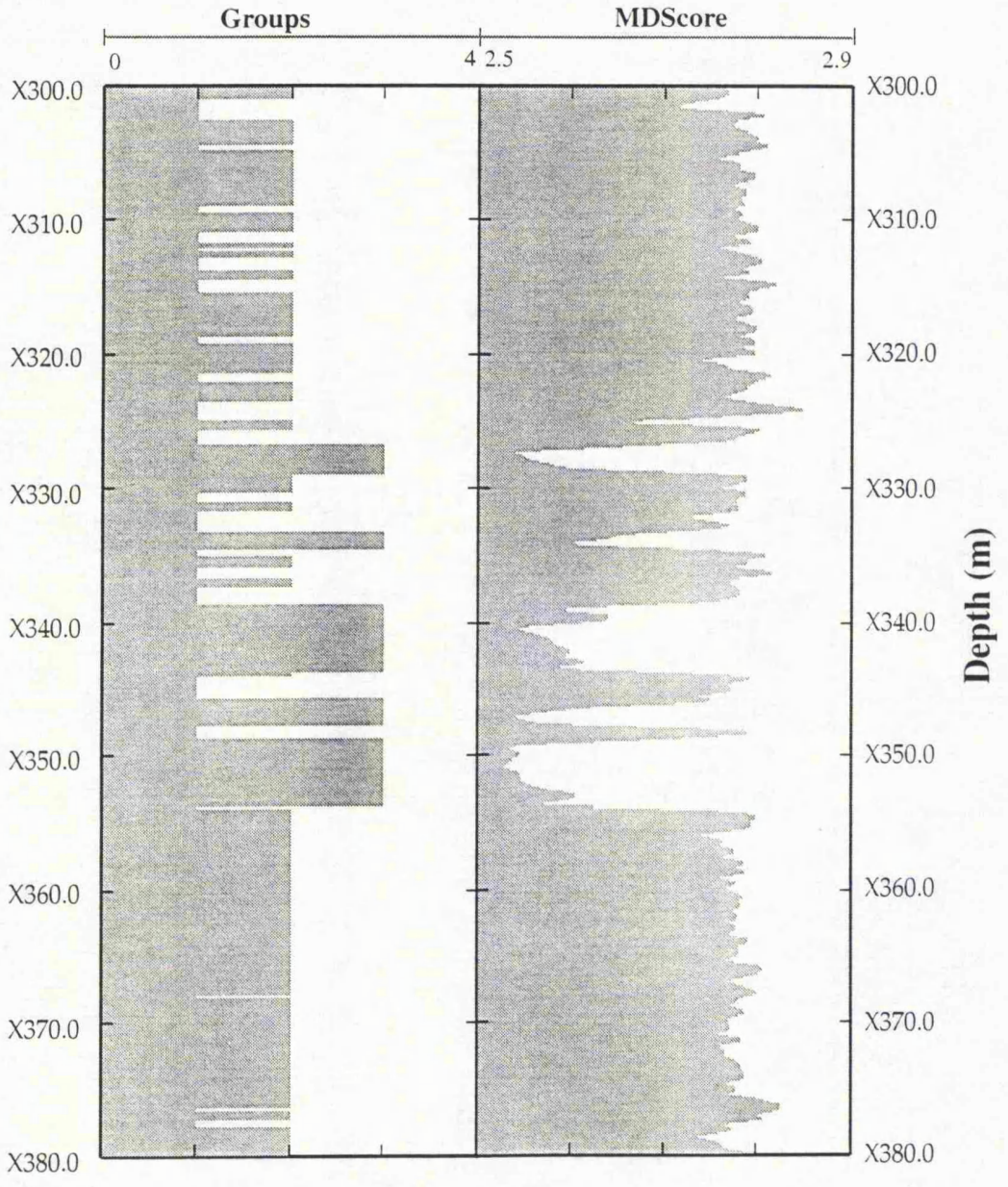


Figure 3.27 - Groups and Discriminant Scores in Hole B.

Table 3.32

EUCLIDEAN DISTANCE CLASSIFICATION

Classification Summary

Group number	Actual number in group	Number assigned to group	Number correct in group	% correct for group
1	100	100	100	100.00
2	422	422	422	100.00
Totals	522		522	100.00

MAHALANOBIS DISTANCE CLASSIFICATION

Classification Summary

Group number	Actual number in group	Number assigned to group	Number correct in group	% correct for group
1	100	100	100	100.00
2	422	422	422	100.00
Totals	522		522	100.00

GROUP GENERALISED DISTANCE CLASSIFICATION

Classification Summary

Group number	Actual number in group	Number assigned to group	Number correct in group	% correct for group
1	100	105	100	100.00
2	422	417	417	98.82
Totals	522		517	99.04

Table 3.32 - General distance classification using three different models in Hole B.

Compared with the other techniques, Discriminant Analysis proved to be useful where a priori information about changes in facies are available. Core data or other geological information can help to define the best training set to be used depending on the objective to be reached. Variation in facies not described in the training sets can lead to misclassifications as observed in Hole B.

3.6 - Summary and Discussion

3.6.1 - Results from ODP Hole 807C

ODP Hole 807C which features three main lithofacies exhibiting both significant physical and chemical contrasts, provided an ideal opportunity in which to investigate the efficiency of the statistical techniques.

Initially, the physical and chemical log data were subjected to the analysis of principal components. The results of this analysis suggest that the principal component I responds to the variation in gamma-ray radiation while the principal component II responds to the high Al concentration. Another aspect of the Principal Component Analysis is the determination of the poor contribution given by the log curves IMPH and SFLU with values lower than 30%.

Cluster Analysis was performed in two steps. The first step was the Hierarchical analysis, where the relations between the log curves were obtained. As observed earlier in the Principal Component Analysis, especially in principal component I, two main groups of log curves can be detected. The second step in the Cluster Analysis was the application of the Non-Hierarchical method. Firstly the analysis was performed on the raw data and succeeded in defining a stratigraphic zonation of three groups which closely corresponds to the lithologic description. The second analysis was carried out based on the two principal component scores, obtained from the earlier Principal Component Analysis, and features an almost identical zonation. However, in this case, only two component scores were used in its construction, illustrating the efficiency of the principal component transform.

The third technique applied to the data set was the Discriminant Analysis. Contrary to the Cluster Analysis, this technique uses *a priori* knowledge about the data set. The result is shown in different ways. The mathematics of the discriminant function generates some form of solution, even when there are no systematic differences between the groups. This contingency can be checked statistically by the F-test in ODP Hole 807C data which is based on the Mahalanobis' distance. The F-test rejected a common centroid between the groups and established the group discrimination as excellent. Mahalanobis' distance was also used to assess the relative contribution of each log curve to discrimination. RHOB was considered the log curve which presented the highest relative contribution. In the earlier Non-Hierarchical analysis using the raw log data, this log curve had already been considered the overall most important with a high value of 4.46 delta mean. Finally, a Table of correct and incorrect classification for the calibration data set and later for the whole data set gave a direct measure of the classification performance

3.6.2 - Results from ODP Hole 878A

ODP Hole 878A consists of a polymictic breccia with variation basically in the matrix composition and in the amount of clasts. Although the physical and chemical log data are more complete than the previous hole, especially concerning the geochemical measurements, the changes in their values are more subtle and therefore hard to detect.

The initial technique applied was the Principal Component Analysis. The result of this analysis suggested that the principal components represent the variation in the matrix composition of the breccia, mostly given by principal component I which accounts alone for 65.3% of the total variability. Once more, it is possible to see the great advantage exerted by the Principal Component Analysis, reducing the dimensionality of the problem without losing the main information present in the original log data.

Cluster Analysis was performed in ODP Hole 878A using the same raw log data. Firstly, the Non-Hierarchical analysis looked for the relation between the variables. In this case, two main groups were also distinguished (Figure 3.16). This result confirms what was observed in the loadings of principal component I, where the variables are well characterised by strong positive and negative loadings.

Secondly, a Non-Hierarchical (Q-mode) analysis was carried out both on the raw log data and in the principal component transformed data. On the first case, a four group discrimination was achieved. Groups 1 and 2 show the variation within the Si based matrix at the top of the interval while Groups 3 and 4 are related with the variation within the Ca based matrix at the bottom. The differences between Groups 1 and 2 and between Groups 3 and 4 were considered to be due to the content of Si and Ca respectively in each zone. The analysis using the principal component scores showed basically the same zonation encountered before, with the advantage of using a reduced amount of data.

3.6.3 - Results from Oilfield Holes A and B

In the case of these two holes, Discriminant Analysis was used to interpret facies changes in Hole B using the discriminant functions obtained in Hole A.

In general, the results can be considered good for lithofacies identification with the discriminant scores showing the two main lithofacies. However, when these results are compared with the Groups obtained from Cluster Analysis, it is clear that the discriminant function from the training sets in Hole A were not capable of interpreting changes in facies spotted by Groups 1 and 2 in Cluster Analysis.

Variations in facies which are not included in the training sets cannot be expected to be identified by the Discriminant Analysis.

Chapter 4

APPLICATION OF NEURAL NETWORK IN WELL LOG DATA

4.1 - Introduction

Different terminologies such as Neural Network, Neural Computing and Neurocomputing are the most used to designate this technique. Other names such as Parallel Distributed Processing and Superpositional Models reflect more the way in which Neural Networks represent concepts as pattern distributed over links between neurons. Another name commonly used is Connectionism, which refers more to the philosophy underlying the neural approach. There are other terms in the literature, but in this work we use Neural Network or Neural Computing simply because they are the most common.

The terminology varies and so do the definitions. Neural Networks are computational systems, either hardware or software, which mimic the computational abilities of biological systems by using large numbers of simple interconnected artificial neurons (Anderson and Rosenfeld, 1988). Other authors like Aleksander and Mortn (1989) go into more detail stating that "Neural Computing is the study of cellular networks that have a natural property for storing experimental knowledge. Such systems bear a resemblance to the brain in the sense that knowledge is acquired through training rather than programming and is retained due to changes in neural functions. The knowledge takes the form of stable states or cycles of states in the operation of the net". He continues: "... A central property of such nets is to recall these states or cycles in response to the presentation of clues."

The problem of such definitions is that they may be useful and accurate but they may seen confusing at first. In general working definitions it is possible to say that Neural Computing is a general activity of getting machines to do brain-like tasks by utilising some of the features of biological neural systems. It has to be said, however, that our artificial Neural Network are very simple and small compared to biological Neural Networks, but they can still perform a remarkable task that conventional computing cannot, or if they do so, complex mathematical models are needed.

So what is the difference between Neural Computing and conventional approaches? Conventional computing relies on the instructions we provide to the machines which then execute a series of operations. Neural Computing relies upon training where the machine learns from experiences. In addition, in conventional

computing the task is explicitly represented whereas the representation is implicit within the links of a Neural Network.

4.2 - What is Artificial Intelligence ?

Today there is a progressively increasing amount of numerical data coming to be interpreted by log analysts. It is natural, therefore, that they look for increasingly more sophisticated methods of analysis, and artificial intelligence can be such a method. It is a potentially very powerful tool which incorporates rudimentary reasoning and pattern recognition features. These techniques, while simple individually, become a powerful system when used together.

A number of books and articles in the use of artificial intelligence examples on logging analysis are available. Most of these show the two paradigms that currently compete as models of the human mind. The first is a symbolic system. Its approach takes thinking to be an abstract chain of reasoning, and its roots are the Cartesian notions of mind/body separation. The second paradigm is based on a system of connections observed by biological studies. In this case, thinking and learning are considered to result from electrochemical changes in the massive parallel neural architecture of the brain (Anderson and Rosenfeld, 1988). This connectionist model is inductive in operation because its learning process is not set by predetermined rules but evolves complex interactions where the generated solution is based on the structure of the data set itself. The symbolic and the connectionist approaches produce two different ways of analysis: the expert system and the Neural Network. Both paradigms have their strengths and weaknesses when applied to problems of log analysis (Crain, 1985, Caudhill, 1991). The choice between them is generally dictated by the form of the petrophysical problem.

In this study it was decided to work with the connectionist approach, which stalled until major conceptual breakthroughs were made in the mid-eighties (Anguita, 1993). The reason for this choice is that the connectionist approach theory is inspired by functional studies of human brains. Obviously, it is not hoped to use in this work a model that is even a feeble approximation of the human brain because of its immense structural complexity that links about a hundred million neurons (MacGregor, 1987). However, the central idea has been put to work for artificial intelligence applications.

Neural Networks differ from expert systems in several fundamental ways. They are trained by being exposed to a large number of input patterns that cause them to learn by experience. Therefore, they operate in contrast to the style of expert system which is directed by an external knowledge base. A Neural Network may be

slower in its ability to work with data, especially when handling a huge amount of training data, but it has the potential for positive results.

4.3 - Neural Networks

In this work the back-error propagation (BEP) Neural Network has been employed for both the characterisation of formation heterogeneities, and also in the construction of petrophysical logs for uncored intervals of a borehole.

4.3.1 - Artificial Neural Networks

An artificial Neural Network is an information-processing system that has certain performance characteristics in common with biological Neural Networks. They have been developed as generalisations of mathematical models of neural biology, based on the following assumptions:

- information is processed at many simple elements called neurons;
- signals are passed between neurons over connecting links;
- each connecting link has an associated weight, which, in a typical Neural Network multiplies the signal transmitted;
- each neuron applies an activation function (usually non-linear) to its input (sum of weighted input signals) to determine its output signal.

A Neural Network is characterised by its architecture (in which the neurons are interconnected), by the method of determining the weights on the connections (the algorithm) and by its activation function (Haykin, 1994). These defined characteristics, which distinguish Neural Networks from other approaches of information processing, are now considered in turn.

A Neural Network consists generally of a large number of processing elements called *neurons*, *cells* or *nodes* (Hetch-Nielsen, 1990). Each neuron is connected with other neurons through connection links, each one with an associated weight. The weights represent the information being used by the neural net to resolve a problem. Each neuron has an internal state, called its *activation function*, which is a function of the input it has received. Typically, a neuron sends its activation as a signal to several other neurons. Even though neurons can only send one signal each time, this signal can be broadcast to several other neurons (Fausett, 1994). Let us consider a neuron Y in the simple example in Figure 4.1, which receives input from neurons X_1, X_2 and X_3. The output signals of these neurons are x_1, x_2 and x_3, respectively. The weights on the connections from X_1, X_2 and X_3, to neuron Y are w_1, w_2 and w_3. Therefore,

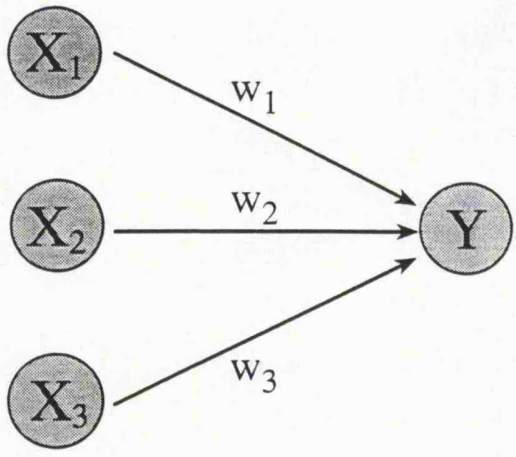


Figure 4.1 - Artificial Neural Network without hidden layer (after Anguita et al., 1993).

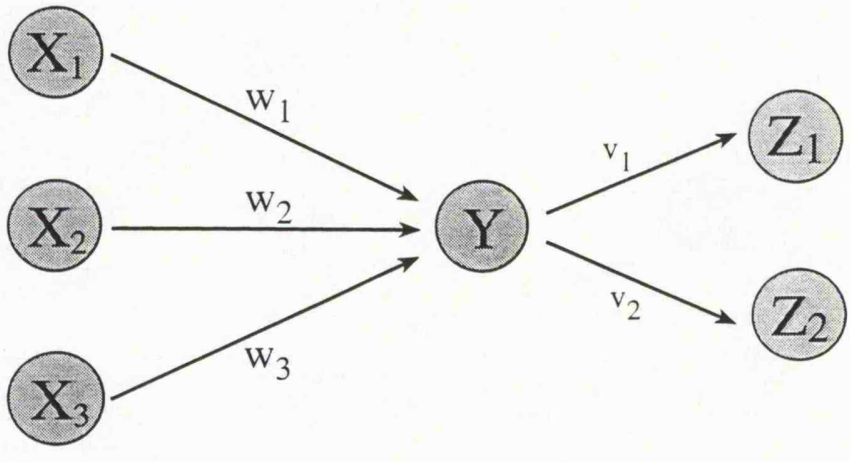


Figure 4.2 - Artificial Neural Network with one hidden layer (after Anguita et al., 1993).

the Neural Network input y_in, to neuron Y is the sum of the weighted signals from neurons X_1, X_2 and X_3.

$$y\_in = w_1x_1 + w_2x_2 + w_3x_3 \quad .$$

The activation y of neuron Y is then given by a function which is the sum of the inputs, y = f(y_in), for instance the sigmoid function,

$$f(x) = 1 / [\ 1 + \exp(-x)\ ] ,$$

or any other activation function. The most commonly used activation functions are described later, in a discussion about the back-error propagation Neural Network.

Now suppose that neuron Y is connected to neurons Z_1 and Z_2 as in Figure 4.2. The weights now between them are v_1 and v_2 respectively. The neuron Y sends its signal to each one of these new neurons. In a typical Neural Network, the activations z_1 and z_2 would depend on inputs from several or even many neurons, not just one, as shown here. Although the example in Figure 4.2 is very simple, the presence of a hidden unit, together with an activation function, gives it the ability to solve many more complicated problems.

4.3.2 - Biological Neural Networks

There is a close analogy between the structure of a biological neuron and the processing unit presented above. In fact, a biological neuron has three types of components that are of particular interests in understanding an artificial neuron: its *dendrites*, *soma* and *axon*. The millions of dendrites in a human brain receive signals from other neurons. The signals are electric impulses that are transmitted across a synaptic gap by means of a chemical process. The action of the chemical transmitter modifies the incoming signal in a manner similar to the action of the weights in an artificial Neural Network (Johnson and Brown, 1988).

The *soma*, or cell body, sums the incoming signals. When sufficient input is received, the cell transmits a signal over its axons to other cells. A generic biological neuron is shown in Figure 4.3, together with axons from two other neurons (from which the illustrated neuron could receive signals) and dendrites for two other neurons (to which the original neuron would send signals). Several key features of processing elements of artificial Neural Networks are suggested by the properties of the biological neuron:

• the processing unit receives many signals;

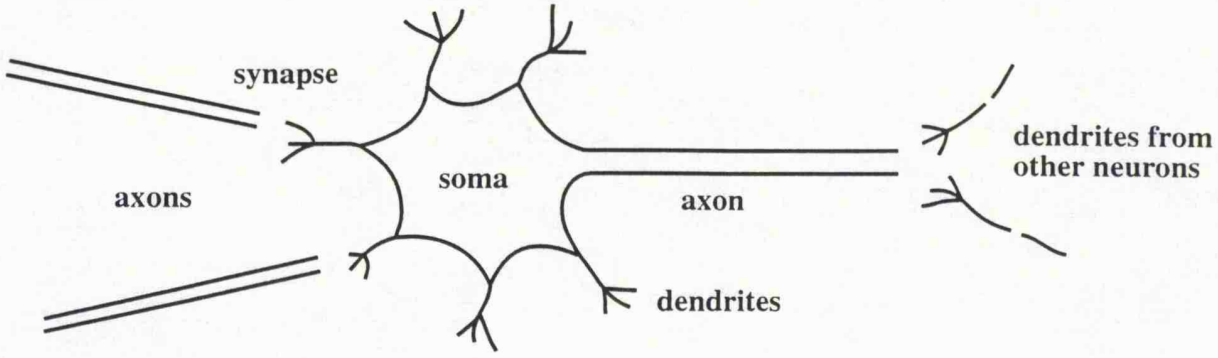


Figure 4.3 - Typical biological neuron (Fausett, 1994).

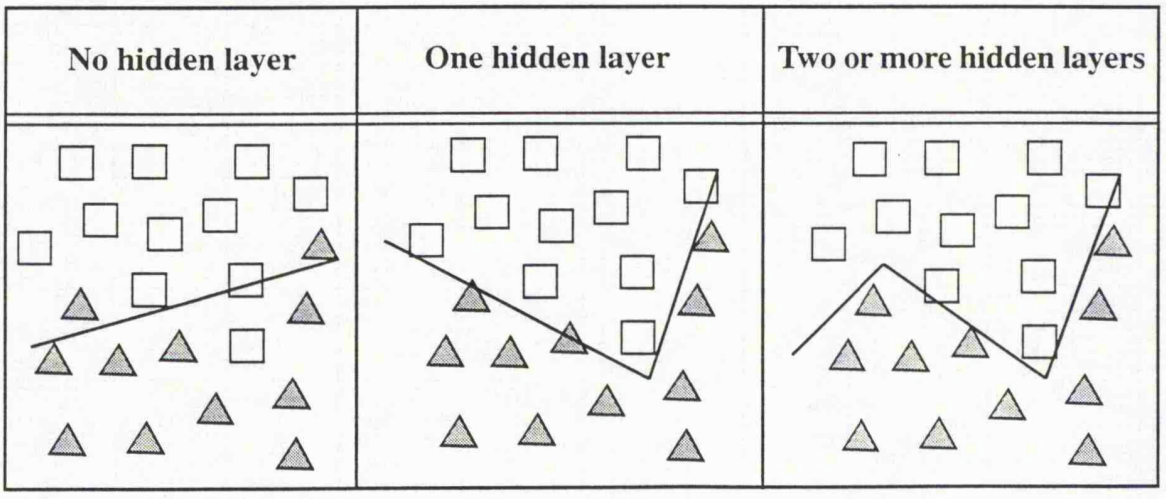


Figure 4.4 - Relationship between the result (partition) given by a Neural Network and the number of hidden layers in its structure (modified from Baldwimet al., 1990).

- the signals may be modified by weights at the receiving synapse;
- the processing unit sums the weighted inputs;
- under appropriate circumstances (sufficient input), the neuron transmits a single output; and
- the output signal of a particular neuron can be transmitted to other neurons.

Some other artificial Neural Network features are also suggested by biological neurons, such as:

- the information processing is local; and
- the memory is distributed:
  - long-term memory resides in the neurons' synapses or weights;
  - short-term memory corresponds to the signals sent by the neurons.

Yet another important characteristic that artificial Neural Networks share with biological neurons is fault tolerance (Fausett, 1994). Biological systems are fault tolerant in two ways. Firstly, humans are capable of recognising many input signals which are different from any other signal they have received before. An example of this is our ability to recognise any object in a picture we have not seen before or to recognise an object after a long period of time. Secondly, we are able to accept any fault in our own nervous system. The human is born with about 100 billion neurons. Most of these are located in the brain and most are not replaced when they die (Johnson and Brown, 1988). Despite the continuous loss of neurons, humans continue to learn. Even in cases of traumatic neural loss, other neurons can be trained to take over the functions of the damaged cells. In a similar manner, artificial Neural Networks can be designed to be insensitive to small damage to the network, and the network can be retrained in cases of significant damages, for instance, loss of data and/or some connections.

4.4 - Back-error propagation Neural Network (BEP)

Single layer Neural Networks were first used in mid 50s. As shown in Figure 4.1, the earliest Neural Networks had the input signals directly broadcast to the output units, with the signals being multiplied by the weights present in these connections and the activation function. However, the limitations of these systems were a significant factor in their decline and lack of interest towards 1970 (Minsky and Papert, 1969). By the mid 80s, an increase was seen in the interest for a new system of Neural Networks, which include one or more hidden layers. In this new approach, called back-error propagation, the final result given by the Neural Network during the

training process is compared with an expected result. The differences between them are retransmitted to the previous layer or layers, together with a change in their weights. The process is repeated until a convergence is achieved with the expected result.

As in the case of most Neural Networks, the aim of the back-error propagation Neural Network is to train the net to achieve a balance between the ability to respond correctly to the input patterns that are used for training (memorisation) and the ability to give reasonable responses to an input that is similar, but not identical, to that used in training (Fausett, 1994).

The training of a back-error propagation Neural Network involves three stages: the feedforward of the input training pattern, the computation and backpropagation of the associated error, and the adjustment of the weights. After training, the application of the net involves only the feedforward process. Even when the training process is slow, a trained net can produce its output very rapidly because there is no propagation of the errors in the opposite direction. Numerous variations of the backpropagation technique have been developed to improve the speed of the training process. In this work, the original algorithm for backpropagation training (Rumelhart et al., 1986) uses an implementation method called Vogl's acceleration (Anguita et al., 1993).

Although a single-layer net is severely limited in the mappings it can learn, a multilayer net, with one or more hidden layers, can learn any continuous mapping to an arbitrary accuracy. More than one and sometimes two layers may be beneficial for some applications, but generally one layer is sufficient. Figure 4.4 gives an idea about the relationship between the number of Neural Network layers and the complexity of partition boundaries that can be generated in a discrimination space. Depending on the number of neurons in the hidden layer, the partition can be improved. The optimum number of neurons depend on the performance of the Neural Network in learning a specific task. In some cases, too many neurons in the hidden layer can produce an overtrained Neural Network, which will not be able to interpret a general input data because it is too sensitive to small changes within the dataset (Baum and Haussler, 1989).

4.4.1 - Architecture and algorithm

A typical back-error propagation Neural Network with one layer of hidden units is shown in Figure 4.5. The output units Y, the hidden layer units Z and the input units X are presented as well as the associated weights w and v. As can be seen, only the feedforward process propagation direction is shown. In the backpropagation stage the signals have the opposite direction.

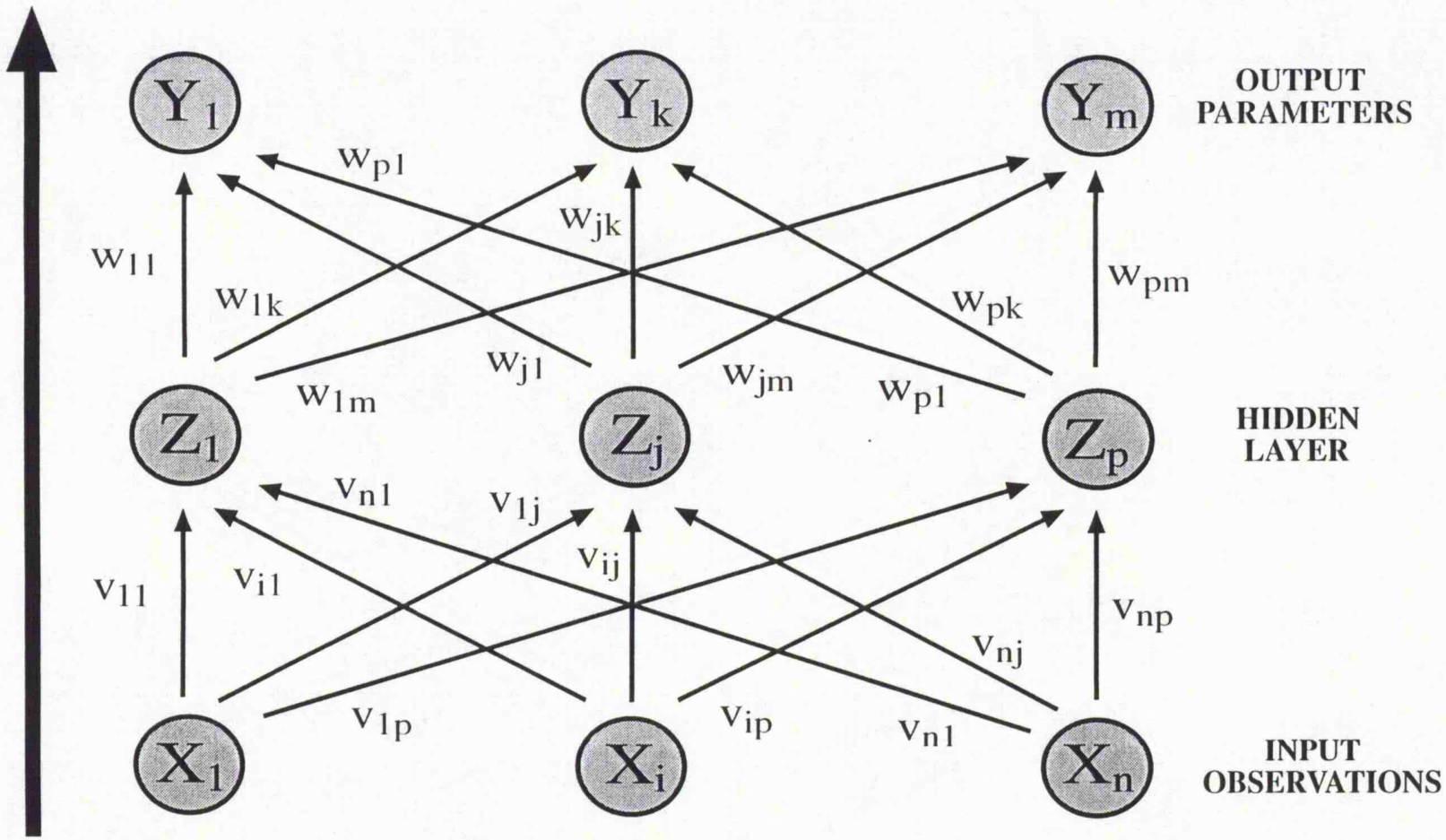


Figure 4.5 - Typical Back-error propagation Neural Network structure with one hidden layer (after Fausett, 1994).

As mentioned earlier, the training process of a back-error propagation Neural Network involves three stages: the feedforward of the input training pattern, the backpropagation of the associated error and the adjustment of the weights. During the first stage, each input unit X_i receives an input signal and broadcasts this signal to each of the hidden layer units $Z_1,..., Z_p$. Each unit then computes its activation function and sends its signal z_j to each output unit. Each output unit Y_k computes its activation y_k to form the response of the net to the previous input pattern.

Yet, during the training process, each output unit compares its computed activation y_k with the expected target value t_k in order to determine the associated error or difference between both. Based on this error or difference, a factor δ_k (k=1,...,m) is computed. This value is used to distribute the errors at the output units back to all units in the previous layer (the hidden layer). It is also used to update the weights between the output layer and the hidden layer. In a similar manner, the factor δ_j (j=1,...,p) is computed for each hidden unit. In fact, it is not necessary to propagate the error back to the input layer, but δ_j is used to update the weights between the hidden layer and the input layer (Haykin, 1994).

After all the δ factors have been determined, the weights for all layers are adjusted simultaneously. The adjustment to the weight w_{jk} is based on the factor δ_k and the activation z_j and the adjustment of the weight v_{ij} is based on the factor δ_j and the activation x_i.

The training and validation algorithm for a back-error propagation Neural Network is described as follows (Fausett, 1994). For the training process we have:

Step 0 - Initialise weights (generally set to small random values).

Step 1 - While stopping condition is false, do steps 2 to 9.

Step 2 - For each training pair input/output, do steps 3 to 8.

Feedforward process:

Step 3 - Each input unit X_i (i=1,..., n) receives an input signal x_i and transmits this signal to all other units in the layer above.

Step 4 - Each hidden unit Z_j (j=1,..., p) sums its weighted input signal through

$$z\_in_j = v_{oj} + \sum_{(1,n)} x_i \, v_{ij} \ ,$$

where v_{oj} is the bias on the hidden unit j (generally set as 1), and then applies its activation function ($z_j = f(z\_in_j)$) to compute its output signal which will be sent to all units in the layer above.

Step 5 - Each output unit Y_k (k=1,..., m) sums its weighted input signals from the hidden layer through

$$y\_in_k = w_{ok} + \sum_{(1,p)} z_j w_{jk} \ ,$$

where w_{ok} is the bias on the output unit k (generally set as 1), and then applies its activation function ($y_k = f(y\_in_k)$) to compute its output signal which will be compared with the expected target value t_k. Backpropagation of the error:

Step 6 - In each output unit Y_k is computed the error which is the difference between the expected value t_k and the output signal y_k as

$$\delta_k = (t_k - y_k)\, f'(y\_in_k) \ ,$$

where δ_k is the error due to the output signal y_k and that will be used to correct the weights w_{jk} through

$$\Delta w_{jk} = \alpha\, \delta_k\, z_j \ ,$$

where α is the learning rate of the Neural Network. This value is also used to correct the bias w_{ok} through

$$\Delta w_{ok} = \alpha\, \delta_k \ .$$

δ_k is then send to the layer below.

Step 7 - Each hidden unit Z_j (j=1,..., p) sums its delta inputs as

$$\delta\_in_j = \sum_{(1,m)} \delta_k w_{jk} \ ,$$

and multiplies this value by the derivative of its activation unit to calculate its error information term. Then,

$$\delta_j = \delta\_in_j \; f'(z\_in_j) \; .$$

The weight correction term, which is used to update v_{ij} later is obtained through

$$\Delta v_{ij} = \alpha \, \delta_j \, x_i \; ,$$

and calculates its bias correction term (used to update v_{oj} later) as

$$\Delta v_{oj} = \alpha \, \delta_j \; .$$

Update weights and biases:

Step 8 - Each output unit Y_k update its bias and weights through

$$w_{jk}(new) = w_{jk}(old) + \Delta w_{jk} \; , \text{ and}$$

each hidden unit Z_j does the same using

$$v_{ij}(new) = v_{ij}(old) + \Delta v_{ij} \; .$$

Step 9 - Test stop condition.

The mathematical basis for the backpropagation algorithm is the optimisation technique known as gradient descent (Battiti, 1992). The gradient of a function (in this case, the function is the error and the variables are the weights of the Neural Network) gives the direction in which the function increases more rapidly; the negative of the gradient gives the direction in which the function decreases more rapidly. Fausett (1994) gives a complete explanation about the derivation of the weight update rules, which clarifies the reason why the weight updates should be done after all of the δ_k and δ_j expression have been calculated, rather than during the backpropagation process.

After training, a backpropagation Neural Network is applied by using only the feed-forward phase of the training algorithm. The application procedure is as follows:

Step 0 - Initialise weights (from training algorithm).

Step 1 - For each input vector, do steps 2 - 4.

Step 2 - For i = 1,..., n: set activation of input unit x_i.

Step 3 - For j = 1,..., p:

$$z\_in_j = v_{oj} + \sum_{(1,n)} x_i \, v_{ij} \; ;$$

$$z_j = f(z\_in_j) \; .$$

Step 4 - For k = 1,...., m:

$$y\_in_k = w_{ok} + \sum_{(1,p)} z_j \, w_{jk} \; ;$$

$$y_k = f(y\_in_k) \; .$$

In the case of a Neural Network with more than one hidden layer, only a few modifications in the algorithm for one hidden layer are necessary.

In this work, we use a implementation of the backpropagation algorithm, called "matrix back propagation" (Anguita, 1993). The purpose of this version is to implement an advanced backpropagation algorithm compared with the original one developed by Rumelhart et al. (1986). The backpropagation algorithm is taken from the classical method developed by Vogl et al. (1991). This is an effective gradient descent method. A modified version of Anguita et al. (1993) for RISC-based workstations and implemented in C was tested and used here.

4.4.2 - Activation function

An activation function for a back-error propagation Neural Network should have a few distinguishable characteristics. It should be continuous, differentiable (as can be seen in steps 6 and 7 of the backpropagation algorithm) and monotonically non-decreasing. Moreover, for computational efficiency, it is desirable that its derivative be easy to compute.

One of the most typical activation functions is the binary sigmoid function, which has a range of (0,1) and is defined in Figure 4.6a as:

$$f_1(x) = 1 \, / \, (\, 1 + \exp(-x) \,) \; ,$$

with

$$f'_1(x) = f_1(x) \, [\, 1 - f_1(x) \,] \; .$$

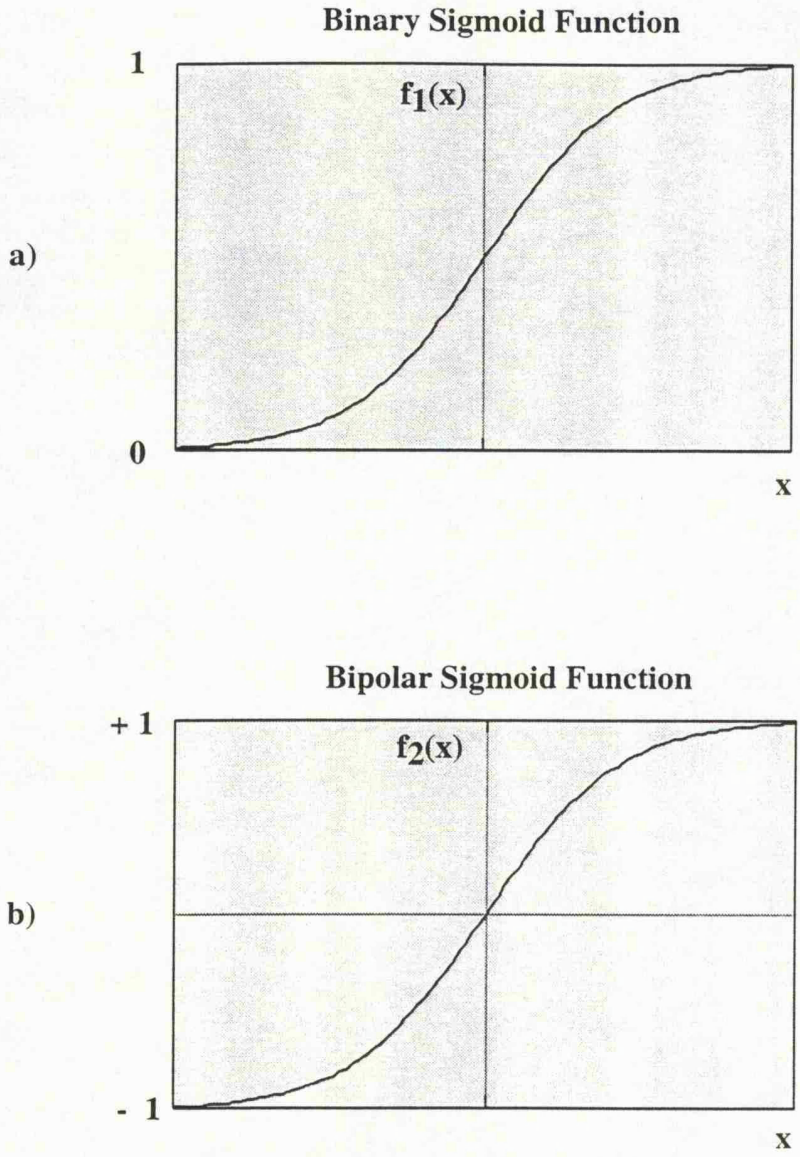


Figure 4.6 - Common activation functions used with Back-error Propagation Neural Network. a) binary sigmoid function and b) bipolar sigmoid function.

Another common activation function is the bipolar sigmoid function (Figure 4.6b), which range from -1 to +1 and is defined as:

$$f_2(x) = (2 / (1 + \exp (-x))) - 1 \ ,$$

with

$$f'_2(x) = (1/2) (1 + f_2(x)) (1 - f_2(x)) \ .$$

4.5 - Application of the Back-error Propagation Neural Network to well log data

There are two principal objectives in the use of the back-error propagation Neural Network in the well log data of the ODP. The first objective is the lithofacies characterisation and identification of formation heterogeneities over specific intervals in different holes already investigated with the multivariate statistical techniques (Chapter 3). The ability of the Neural Network to characterise different geological sequences using both well log and core data was tested here. The results for each hole are compared with results obtained with the statistical techniques used in chapter 3.

The second objective is to test the ability of the Neural Network in determining quantitative physical property estimates from well logs in uncored intervals using core measurements from other intervals as a training basis. In this case the precision and accuracy of the data from different sources are very important in the training of the Neural Network, and are considered here by using sample support in the integration of measurements at different scales. We also show the ability of the Neural Network to construct physical property logs for uncored intervals.

4.5.1 - Data standardization

The aim of standardization or equalisation is to cause each variable or log curve to have a common numerical magnitude. This removes the problem which arises when different units of measurements are used to express element concentrations or physical properties (i.e. weight percent or part per million; resistivity and porosity). It also accommodates log curves which use the same scale of measurement but have different magnitudes. The process of standardization is achieved by dividing all the measurements for a variable by an equalising factor expressed in the same units. This converts all the log curves to a single index of similarity.

The method of homogenising the log curves used in this work is that of reducing to a standard form (i.e. zero mean and unit standard deviation) by:

$$z_{ij} = (x_{ij} - \bar{x}_i) / \sigma_i \ ,$$

where z_{ij} is the standardized value for the i^{th} log curve of the j^{th} measurement, x_{ij} is the i^{th} log curve of the j^{th} measurement, $\bar{x}_i$ is the overall mean of all values of the i^{th} log curve and σ_i is the standard deviation associated with the i^{th} log curve

This process removes the effect of individual magnitude of the various log curves (e.g. 10% of SiO_2 may now be equivalent to 5 ppm of Th or 3 ppm of U). In a large number of experiments which were carried out on both standardized and unstandardized data for which there was *a priori* knowledge, it was found that normalised data tended to give improved results.

4.5.2 - ODP Hole 807C

The interval studied in ODP Hole 807C, the one investigated by multivariate statistical analysis, is between 1270 - 1400 mbsf. The geological sequence, as described before, consists of a carbonate sequence from 1270 to 1351 mbsf (Unit A), followed by a siltstone/claystone sequence to 1380 mbsf (Unit B) and a basic volcanic sequence to the bottom of the hole (Unit C). In addition there is a thin carbonate sequence at the bottom of Unit B, between 1375 - 1377 mbsf. The first objective is to compare the lithofacies characterisation obtained here with the one produced by statistical techniques and also with the classification observed in the core description. The second objective is to produce physical property logs for the uncored sections of the hole using physical property measurements from the cores as training sets.

4.5.2.1 - The lithofacies characterisation

For comparison with multivariate statistical analysis in ODP Hole 807C the set of log curves were used. The data set consists of 10 log curves given by the resistivity tool (ILD and SFLU), the natural gamma tool (GRAY, THOR, POTA and URAN), the lithodensity tool (RHOB and PEF), sonic tool (DT) and the aluminium clay tool (ALUM). The lithofacies characterisation obtained from core description is used here to build the training set. It consists of short intervals which best characterise each lithofacies. The intervals selected are the same as for the Discriminant Analysis (see chapter 3) in order to perform a comparison between the two results.

The training algorithm operates accordingly to the Neural Network structure. The user specifies the desired network size and initial structure, with the number of input and output neurons dictated by the task being solved (Gonçalves et al., 1995a).

The structure of the Neural Network used for ODP Hole 807C has 10 neurons in the input layer, 7 neurons in the hidden layer and 3 neurons in the output layer (Figure 4.7). The input layer corresponds to the number of log curves used in this analysis and the output layer is given by the three different lithofacies to be identified. The number of neurons in the hidden layer is something to be defined by the performance of the Neural Network during the training process. Transformation tasks require exact or nearly exact solutions for use as targets. Very small, if any, errors are allowable. Any variance in input signals that are not due to noise represent true variations in formation characteristics that the network must be able to track. Therefore, the final number of neurons in the hidden layer is defined by the best result obtained in the training process, or the minimum error between the net result and the expected target. Figure 4.8 illustrates the performance of the back-error propagation for training process in ODP Hole 807C using 7 neurons in the hidden layer. It shows the number of iterations needed to obtain a minimum error between the result and the expected target. Other tests were done for different numbers of neurons in the hidden layer. Table 4.1 shows the result in terms of number of iterations and the minimum error achieved by the Neural Network.

Table 4.1

Number of neurons in the hidden layer	Number of iterations	Minimum error achieved
5	150	6.8%
6	212	2.5%
7	100	0.27%
8	327	1.90%
10	430	3.45%

Table 4.1 - Number of iterations and minimum error achieved for different number of neurons in the hidden layer for training process in ODP Hole 807C.

Verifying generalisation is a critical step in Neural Network training. The term generalisation refers to the ability of a trained Neural Network to process data that is not part of the training set. Good generalisation means that the network has learned general principles from a limited training set and that its final structure is most generally correct. Poor generalisation means that the network is not able to discover general data trends from the training set. Critical features are not learned or not learned well enough. Verification of generalisation is important because a network

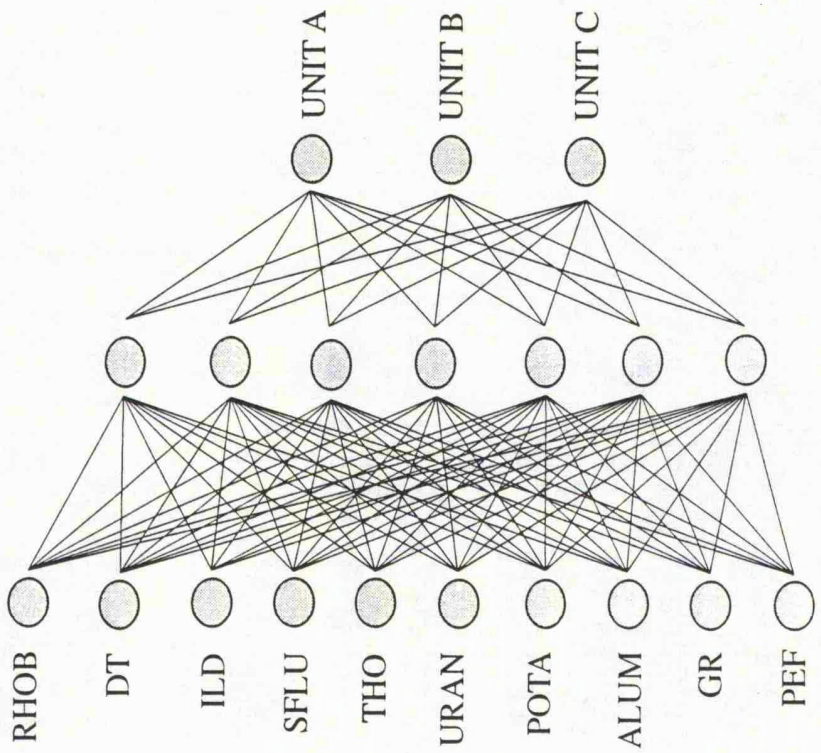


Figure 4.7 - Structure of the Neural Network for lithofacies identification in ODP Hole 807C.

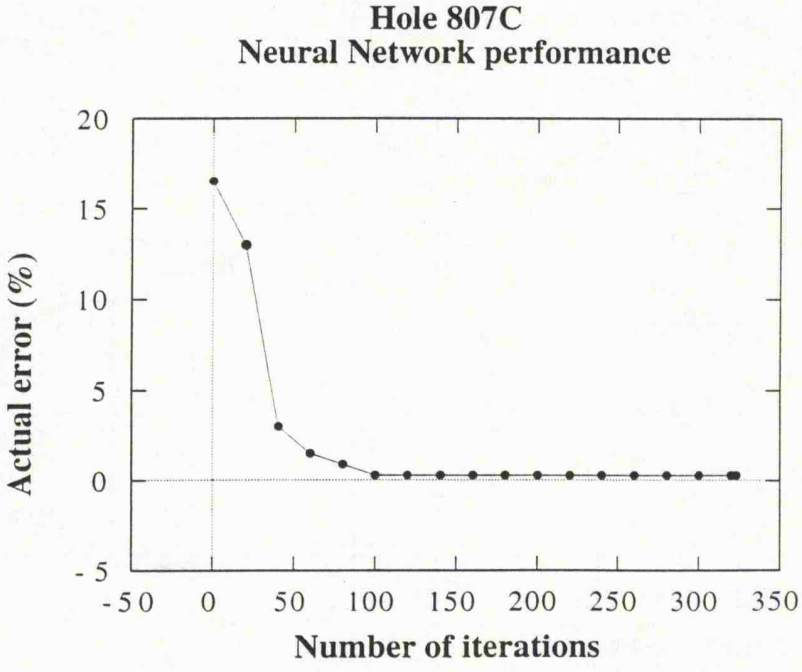


Figure 4.8 - Neural Network performance during training process for lithofacies identification in ODP Hole 807C.

might learn the training set so well that it becomes too sensitive to small features which are normally not part of the routine data set.

The "brainstate" obtained by the training process outlined above is then applied in a feed-forward mode (see section 4.3.1) to the whole data set. It means that the output from all input layer neurons is modified by the connection weights found during training. The result of the generating output process for ODP Hole 807C is shown in Figure 4.9. The three different lithofacies are represented here as Unit A, Unit B and Unit C for the carbonate sequence, the siltstone/claystone sequence and for volcanic sequence respectively. As can be seen, most of the interval studied is in excellent agreement with the core description. The variations observed are generally due to two or even three lithofacies being characterised within the same interval, for instance at 1310-1315 mbsf, 1345-1350 mbsf and 1375-1380 mbsf where the three different lithofacies are defined by the Neural Network. In the first two intervals described above the predominance of Unit A over the other two lithofacies is clear. In the last interval, it seems that the Neural Network was not able to identify the right unit present as all lithofacies seem to be well characterised by the Neural Network. Between 1375-1377 mbsf the result gives the volcanic sequence (Unit C) as the strongest unit present in this interval, with the presence of Unit B and Unit A reduced. Between 1377-1380 mbsf although the Neural Network gives a strong siltstone/claystone characterisation, it also shows strong presence of Unit A (carbonates) for this interval. These misclassifications are the result of similar values in some log curves at these intervals or even bad hole conditions as can be observed in the caliper log (Figure 2.4).

Based on the absolute classification obtained from the core description, it is observed that the Neural Network gives a result with 2.87% of misclassified samples. This figure is given by the number of samples that do not match the classification given by the core description and Group classification. Comparing the classifications obtained here with the back-error propagation Neural Network and the one obtained by Discriminant Analysis in Chapter 3, the Neural Network still gives a better result than the one obtained by the Euclidean distance classification method which gives a 22.67% of misclassified samples.

One way to find out the possible sources of the misclassifications observed for ODP Hole 807C is to run the analysis for different data sets. In this case, we use two criteria to split the data. The first one is based on the property dependence of each log curve. The second criterion is the vertical resolution and depth of investigation of each log curve. In the first one it is looking for whether some physical or chemical properties instead of all data set is responsible for the misclassifications.

In the second criterion, as measurements with different vertical resolutions and also different depth of investigations are being taken, and assuming that both have

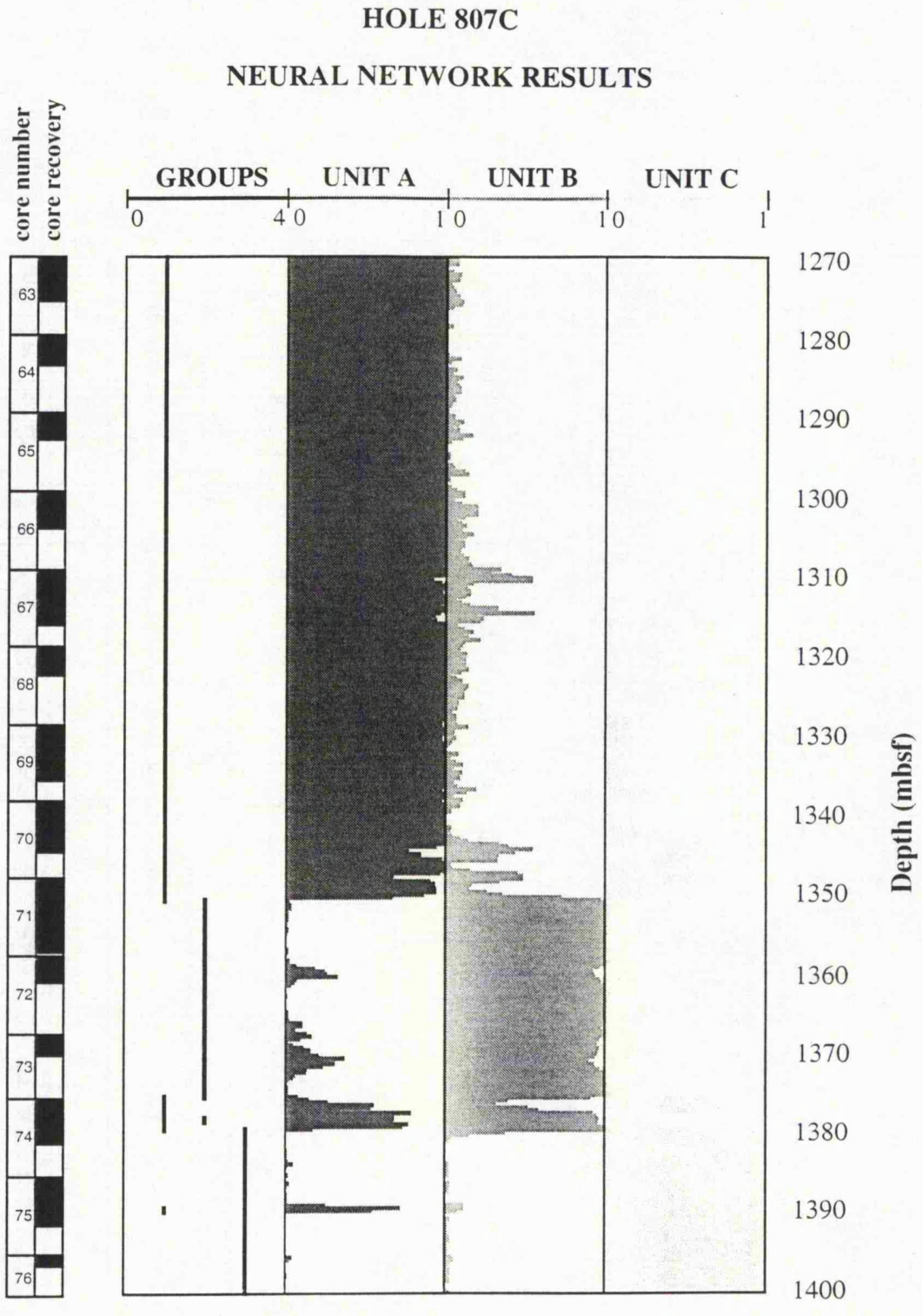


Figure 4.9 - Neural Network results for lithofacies identification in ODP Hole 807C.

generally the same value (Theys, 1990),. it is possible that some intervals are being affected by measurements from nearby intervals. Therefore, the whole data set is split into different groups of log curves depending on their property dependence and vertical resolution characteristics. The analysis for each group is shown below.

Property Dependence analysis

The log curves in ODP Hole 807C were divided into two different groups according to their property dependence. Group I is composed of the log curves which are more affected by the lithology type whilst Group II are defined by the log curves which respond to variations in porosity and fluid type present in the rock. Both groups can be observed in Table 4.2.

Table 4.2

Group I	Group II
GRAY	DT
THOR	ILD
URAN	SFLU
POTA	RHOB
PEF	
ALUM	

Table 4.2 - Property dependence groups for back-error propagation Neural Network analysis in ODP Hole 807C.

The results for the analysis for each group are shown in Figures 4.10 and 4.11. While there is a general agreement between the two models and that obtained with the complete set of logs, some distinct differences are apparent. These differences in terms of misclassification (two or more lithofacies appearing at the same interval) are very strong.

In the analysis of Group I, it is observed that Unit C although only described by core description below 1380 mbsf, appear as a misclassification throughout the upper and middle section of the hole. These misclassifications are due to the presence of four log curves (GRAY, THOR, URAN, PEF) which present almost no difference in the range of their values in Units A and C. The misclassifications observed in the middle section (1352-1380 mbsf) are due mostly to the presence of the same URAN content within Units B and C.

For Group II, a different result in terms of misclassification is observed. As in the analysis with Group I, it is still possible to see a good characterisation of the three

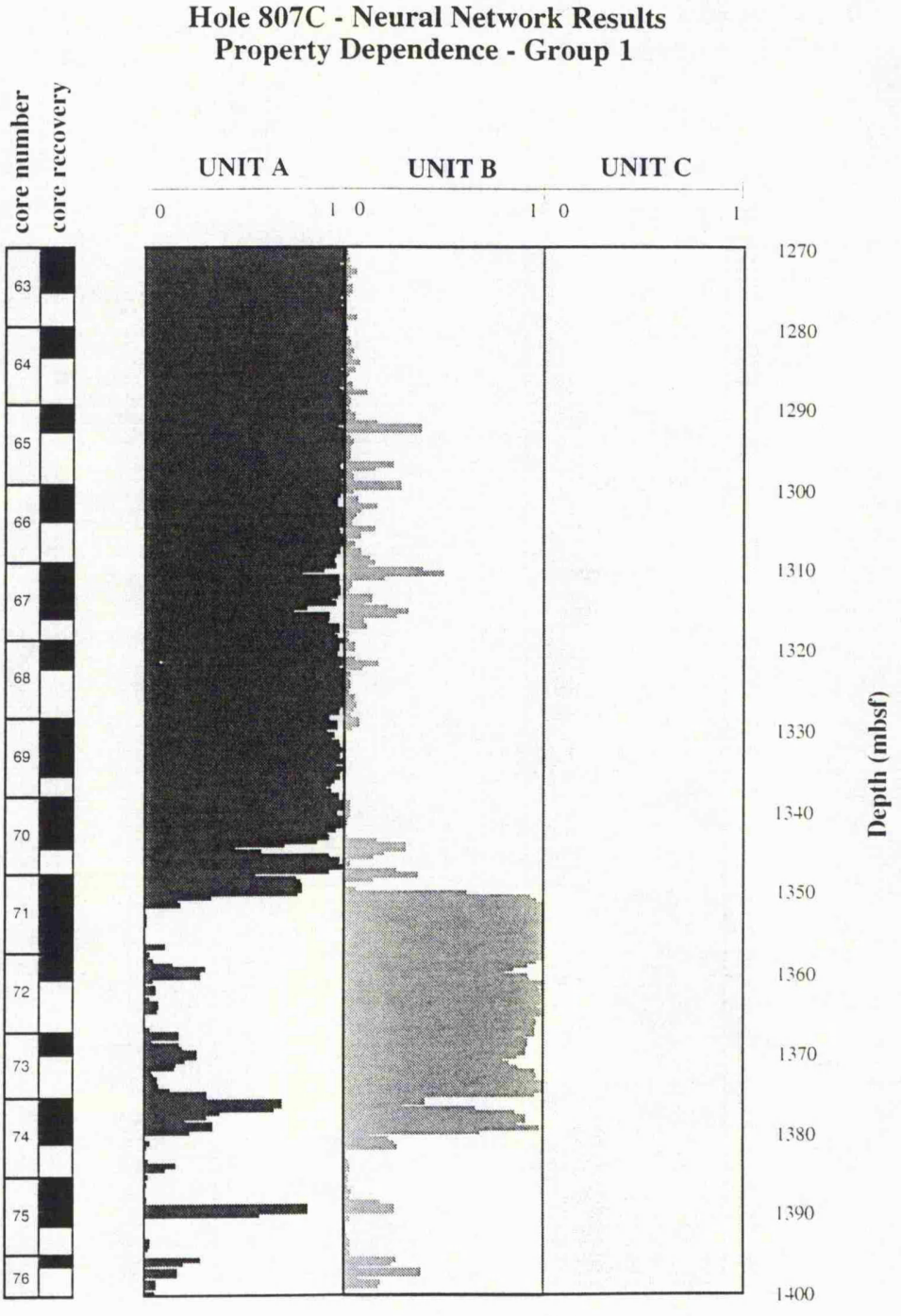


Figure 4.10 - Neural Network results for lithofacies identification using Group 1 log curves from property dependence analysis in ODP Hole 807C.

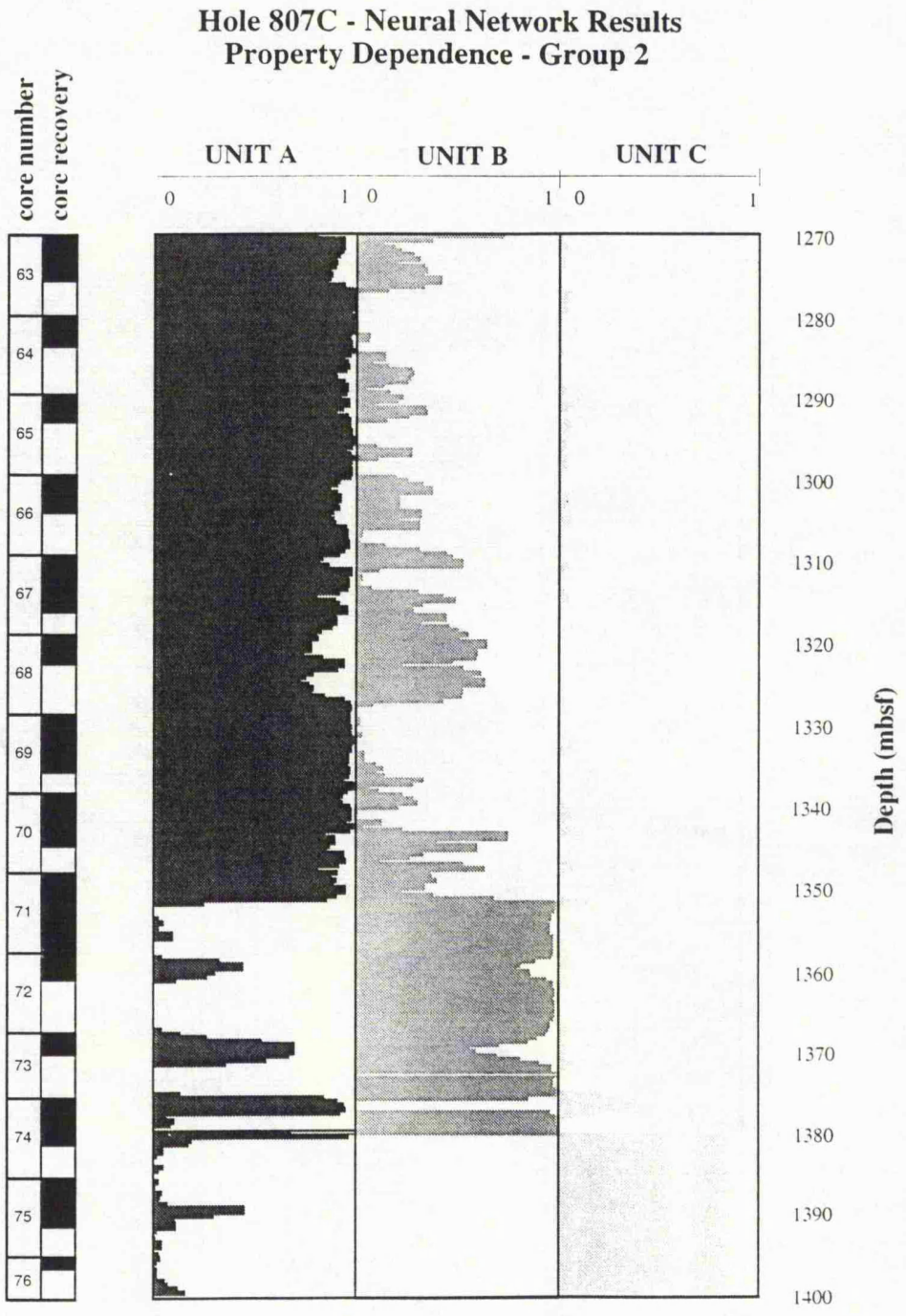


Figure 4.11 - Neural Network results for lithofacies identification using Group 2 log curves from property dependence analysis in ODP Hole 807C.

different units. However, in this case, Unit B appears as a strong misclassification in the upper part of the interval. Here, resistivity measurements (ILD and SFLU) appear to be the main source of these misclassifications.

From these analyses we can start to interpret the misclassifications obtained for the analysis with all of the log curves (Figure 4.9). Variations in porosity within Units A and B (given by resistivity measurements) can possibly be the source of the misclassifications observed within these two units. Small increases in potassium content and also small decreases in photoelectric factor values are the main sources of the misclassifications of Unit C in the upper part of the interval. In the same way, similar gamma-ray values (mainly between 1345-1350 mbsf) can also be a source for those misclassifications.

Vertical Resolution and depth of investigation analysis

The objective here is to see whether if any of the misclassification observed in the analysis with the complete set of log curves can also originate from different vertical resolutions and/or depths of investigation of the log curves. The log curves were divided into four different groups (Table 4.3).

The analysis with the Neural Network was carried out for each group and the results can be observed in Figures 4.12 to 4.15. Again, in general all the results identify well the three lithofacies. However, in each analysis the misclassifications appear differently. A major difference can be observed between Groups I, II and III and Group IV. Although the three first groups represent different vertical resolutions and consequently different depths of investigation, two of them contain a component of the resistivity measurement. They are the main source of all misclassifications observed within Units A and B. The misclassifications are more evident in the results of Groups I and III. In Group II, the presence of ALUM (different range of values within Units A and B) reduces the amplitude of the misclassifications. On the other hand Group IV, which does not contain any resistivity curve, gives a completely different response. In this case, the noise is represented by Unit C being characterised in the upper and middle interval. When we compare the result with the one obtained in Figure 4.10 (Group I of property dependence) we can see in general the same variation. In both cases the difference is that here we have RHOB (density) instead of ALUM. Possibly the close range of density values between Units A and C is responsible for the increase in the misclassifications in this case.

What happens is that instead of vertical resolution and/or depth of investigation, the results are responding more to the property dependence of each log curve combination. In all these examples, the reduction in the number of log curves results in a worst response by the Neural Network, while more misclassifications are

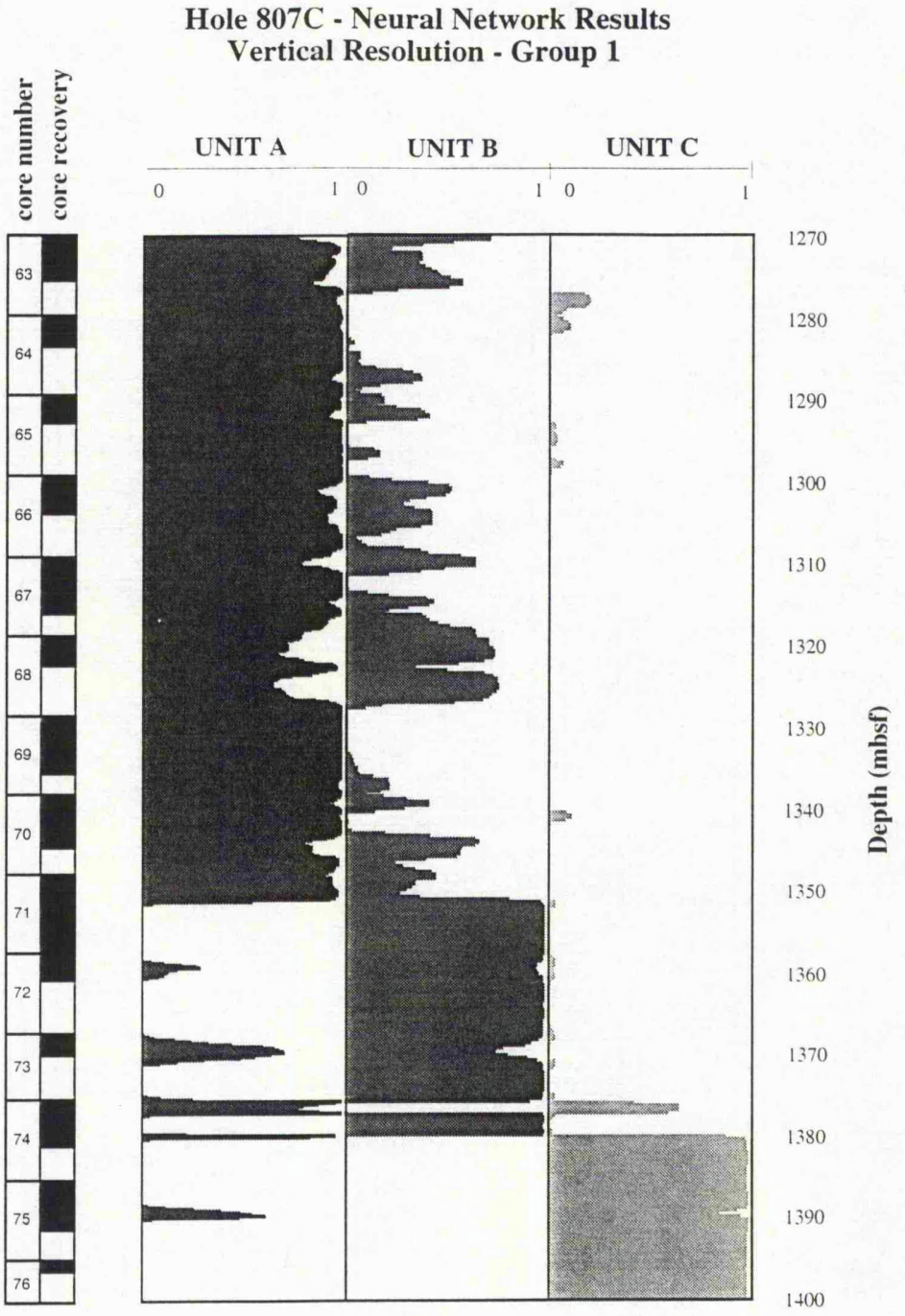


Figure 4.12 - Neural Network results for lithofacies identification using Group 1 log curves from vertical resolution and depth of investigation analysis in ODP Hole 807C.

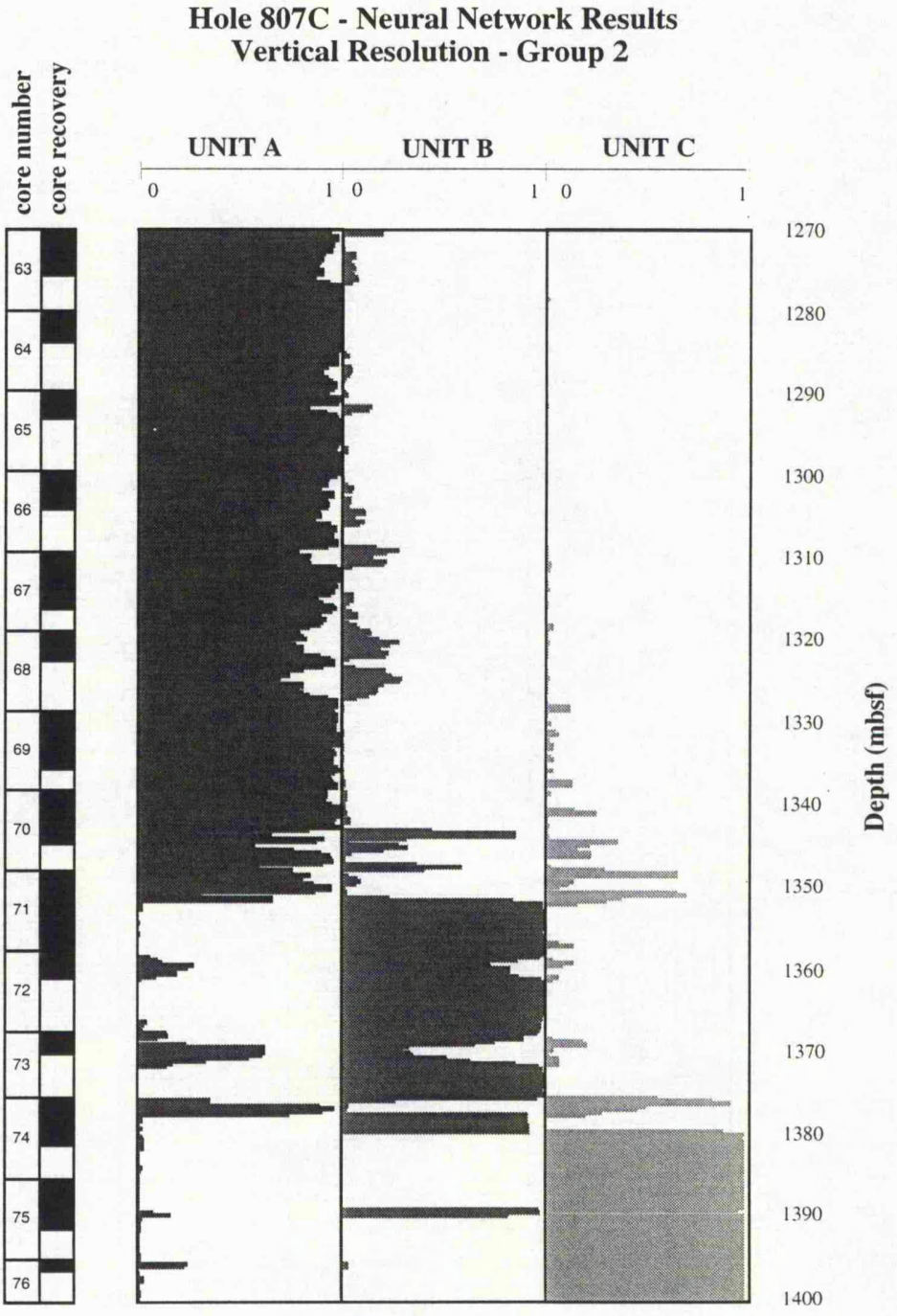


Figure 4.13 - Neural Network result for lithofacies identification using Group 2 log curves from vertical resolution and depth of investigation analysis in ODP Hole 807C.

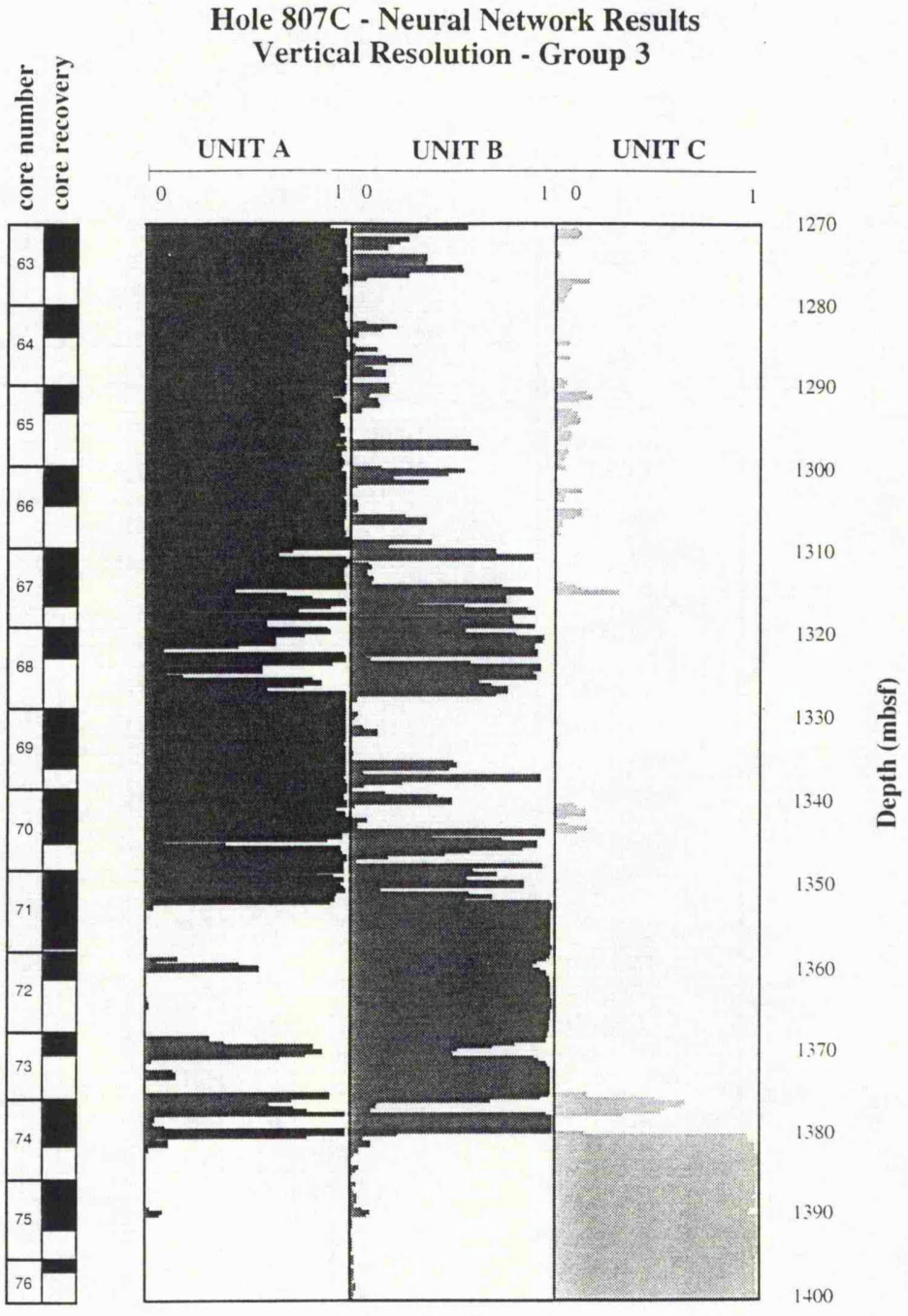


Figure 4.14 - Neural Network result for lithofacies identification using Group 3 log curves from vertical resolution and depth of investigation analysis in ODP Hole 807C.

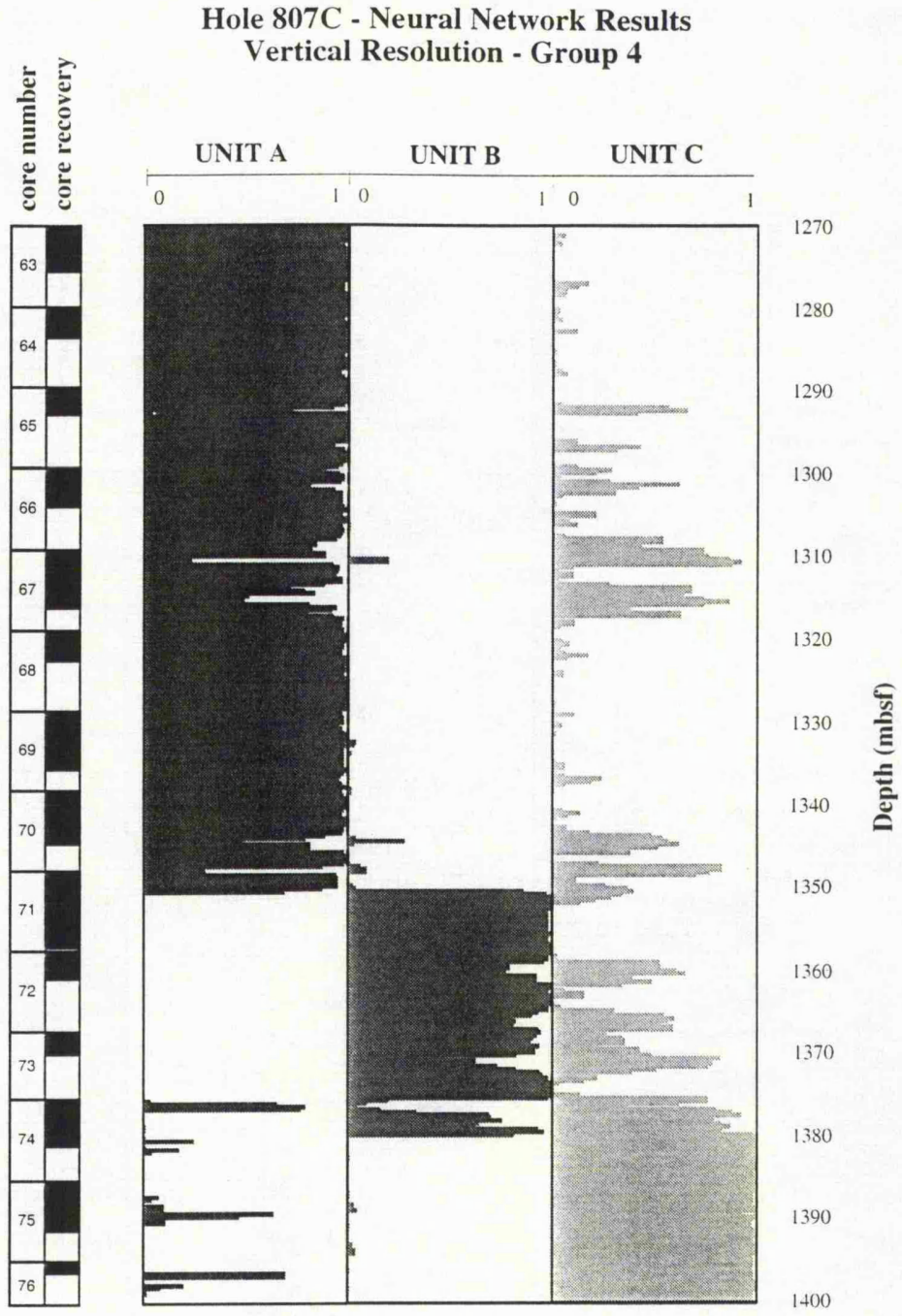


Figure 4.15 - Neural Network result for lithofacies identification using Group 4 log curves from vertical resolution and depth of investigation analysis in ODP Hole 807C.

observed. The results lose quality and resolution although it is still possible to detect the right unit at the right interval.

Table 4.3

Group I	Group II	Group III	Group IV
ILD	ALUM ILM	DT SFLU	GRAY THOR URAN POTA PEF RHOB
	average vertical resolution		
2 meters	1.5 meter	0.75 meter	0.4 meter

Table 4.3 - Vertical resolution groups for back-error propagation Neural Network analysis in ODP Hole 807C.

4.5.3 - ODP Hole 878A

The objectives for the analysis with the back-error propagation Neural Network for ODP Hole 878A included a comparison between the result given by the Neural Network and the result obtained with the statistical techniques (chapter 3). The ability of the Neural Network to produce petrophysical logs for the uncored intervals is also tested. The interval to be investigated is the same as in chapter 3, between 515-600 mbsf. As two different matrix breccias are the main lithofacies within the interval, they are represented in the training set as they were in the statistical analysis.

4.5.3.1 - The lithofacies characterisation

The structure of the Neural Network used for ODP Hole 878A consists of 13 neurons in the input layer, 7 neurons in the hidden layer and 2 neurons in the output layer (Figure 4.16). The number of neurons in the input layer correspond to the 13 log curves used as input and the number of neurons in the output layer correspond to the two different matrix supported breccias observed in this hole. Again, the number of neurons in the hidden layer was dictated by the performance of the neural net during the training process. Figure 4.16a shows the performance of the Neural Network for lithofacies characterisation in ODP Hole 878A. As observed, the number of iterations to obtain the minimum error (400) are greater than in previous example. This is probably due to the variability in the log responses caused by the presence of a more complex lithofacies in this hole.

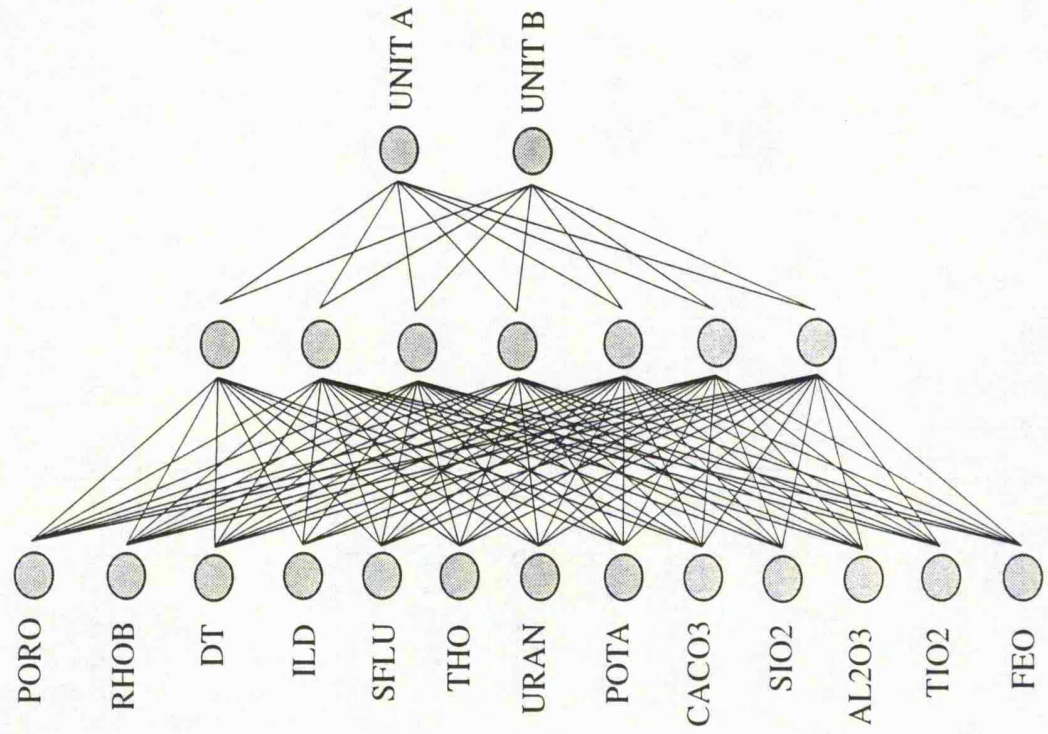


Figure 4.16 - Structure of the Neural Network for lithofacies identification in ODP Hole 878A.

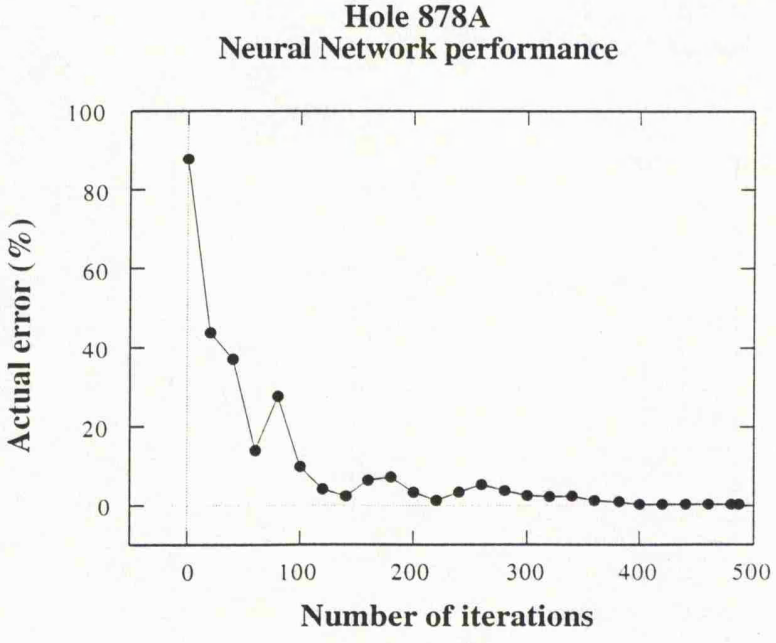


Figure 4.16a - Neural Network performance during the training process for lithofacies identification in ODP Hole 878A.

The weights obtained during the training process were then applied to the whole data set and the result can be observed in Figure 4.17. The two different lithofacies are here represented by Unit A (Si-matrix breccia) and Unit B (Ca-matrix breccia). The interpretation given by the Neural Network corresponds in general with the characterisation observed in the cores for the whole interval. The interval between 563-571 mbsf, where an interbedded sequence of the two different units are observed, was also well characterised.

Along the whole interval there are only two sections that were not well characterised by the net. Although the results show a strong presence of Unit A for the interval between 540-548 mbsf, Unit B is also well represented. The second section is the interval between 571-577 mbsf, which presents the same characteristic described above. Both misclassifications were also observed in the multivariate statistical analysis (chapter 3). The group classification of Cluster Analysis (Figure 3.19) shows Group 3 for the interval between 540-548 mbsf.

Compared with the absolute classification obtained through the excellent core recovery within the whole interval (95%), the Neural Network shows a value of 4.84% of samples that do not agree with that classification. This value is mostly given by the intervals described in the last paragraph. The classification obtained from the Discriminant Analysis using the Euclidean distance classification method gives a value of 6.20% of misclassified samples. Even though the number of misclassified samples are so different, the intervals where they were observed are the same for both techniques. This last example shows how much better a Neural Network model can improve a lithofacies classification in a complex lithological sequence.

As in the previous hole, in order to find out possible sources of the misclassifications, the analysis with the back-error propagation Neural Network was run for different data sets. To do this, the original data set was once more split according to the same criteria used before. The first criterion is the property dependence and the second is the vertical resolution/depth of investigation of each log curve.

Property Dependence analysis

Five different groups were formed accordingly to the property dependence of each log curve in ODP Hole 878A (Table 4.4). Groups I and II are related to variations in rock porosity with Group II also showing influences of fluid type present in the rock. Groups III, IV and V are related with the lithology type.

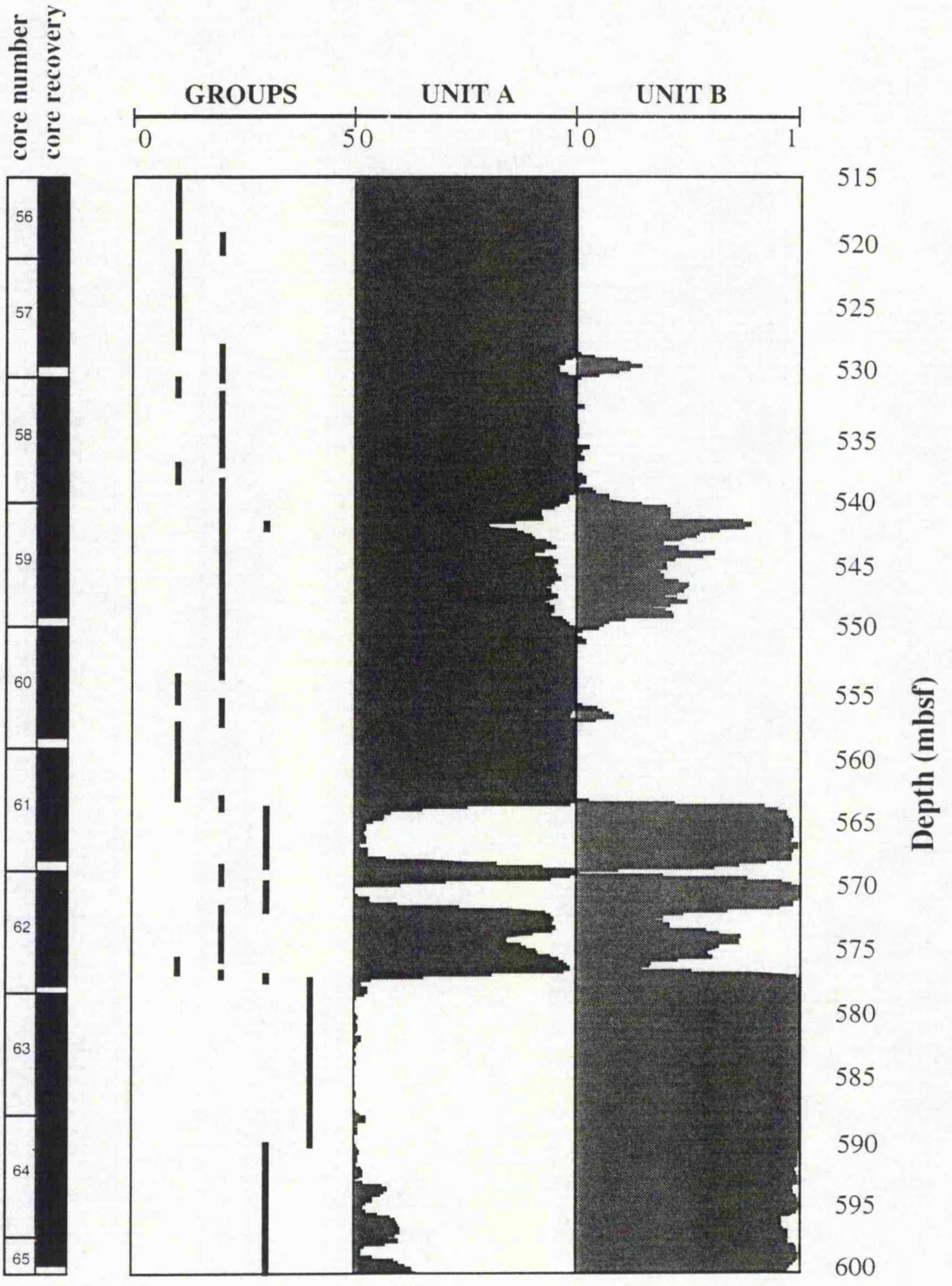


Figure 4.17 - Neural Network results for lithofacies identification in ODP Hole 878A.

Table 4.4

Group I	Group II	Group III	Group IV	Group V
NPHI	DT	THOR	CACO3	AL2O3
RHOB	ILD	URAN	SIO2	TIO2
	SFLU	POTA		FEO

Table 4.4 - Property dependence groups for back-error propagation Neural Network analysis in ODP Hole 878A.

The result for the analysis with each group can be observed in Figures 4.18 to 4.22. As seen, the best result obtained is given by Group IV(Figure 4.21). This is because the main variation in the polymitic breccia is the presence of a Ca or Si matrix. In this case the $CaCO_3$ and SiO_2 oxides show the best fit to the analysis with all set of log curves (Figure 4.17). The other groups show different results, but basically keep the two different units well characterised. The worst result, or the one with most misclassifications, is Group III which shows some mismatches not observed before. This are probably due to the small variation in THOR, URAN and POTA (Figure 3.6) within the interval.

These results confirm the lithologic description of the cores. The main variation is given by the matrix component. The variations in Ca and Si content are the principal sources of the mismatches. Another interesting factor here is that, contrary to the previous examples, the reduction in the number of log curves (when the right ones are used) does not affect the result of the analysis and also shortens the time needed for analysis.

Vertical resolution and depth of investigation analysis

Based on their vertical resolution, the log curves in ODP Hole 878A were divided into three different groups (Table 4.5). The result obtained for each group can be observed in Figures 4.23 to 4.25. The results given by the three different groups show a greater dependence on property rather than to the vertical resolution or depth of investigation of the log curves. Group I (Figure 4.23) shows a response very similar to that obtained for the analysis using all log curves (Figure 4.17) and to the one using $CaCO_3$ and SiO_2 oxides. The other two groups show a slightly different responses, with the mismatches increased or decreased depending on the log curves present.

Although some groups are still able to give a relatively equal response to that obtained with a complete set of log curves, it is clear that the analysis loses resolution. An example is the interval between 563-577 mbsf, where two thin interbedded layers

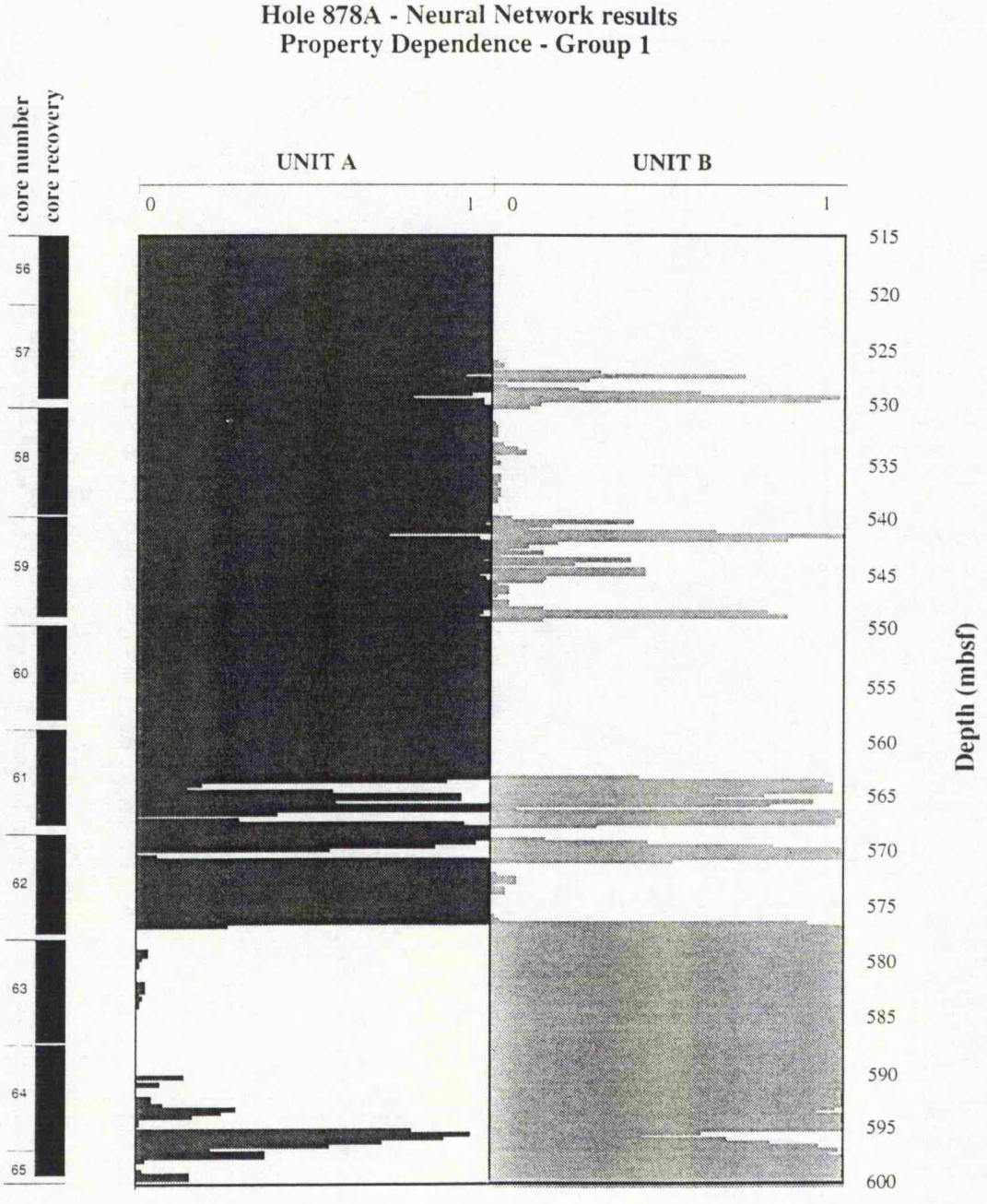


Figure 4.18 - Neural Network results for lithofacies identification using Group 1 log curves from property dependence analysis in ODP Hole 878A.

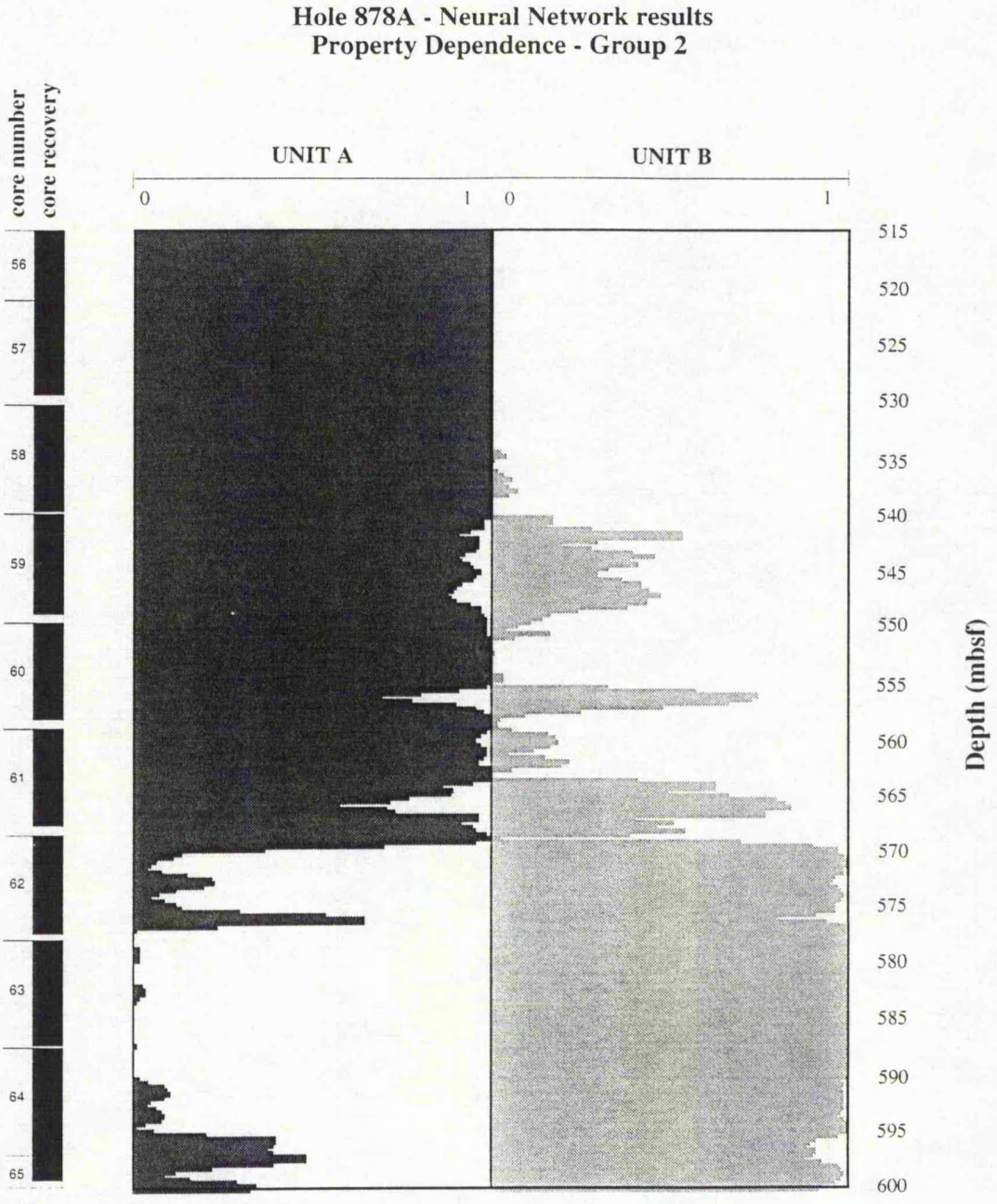


Figure 4.19 - Neural Network results for lithofacies identification using Group 2 log curves from property dependence analysis in ODP Hole 878A.

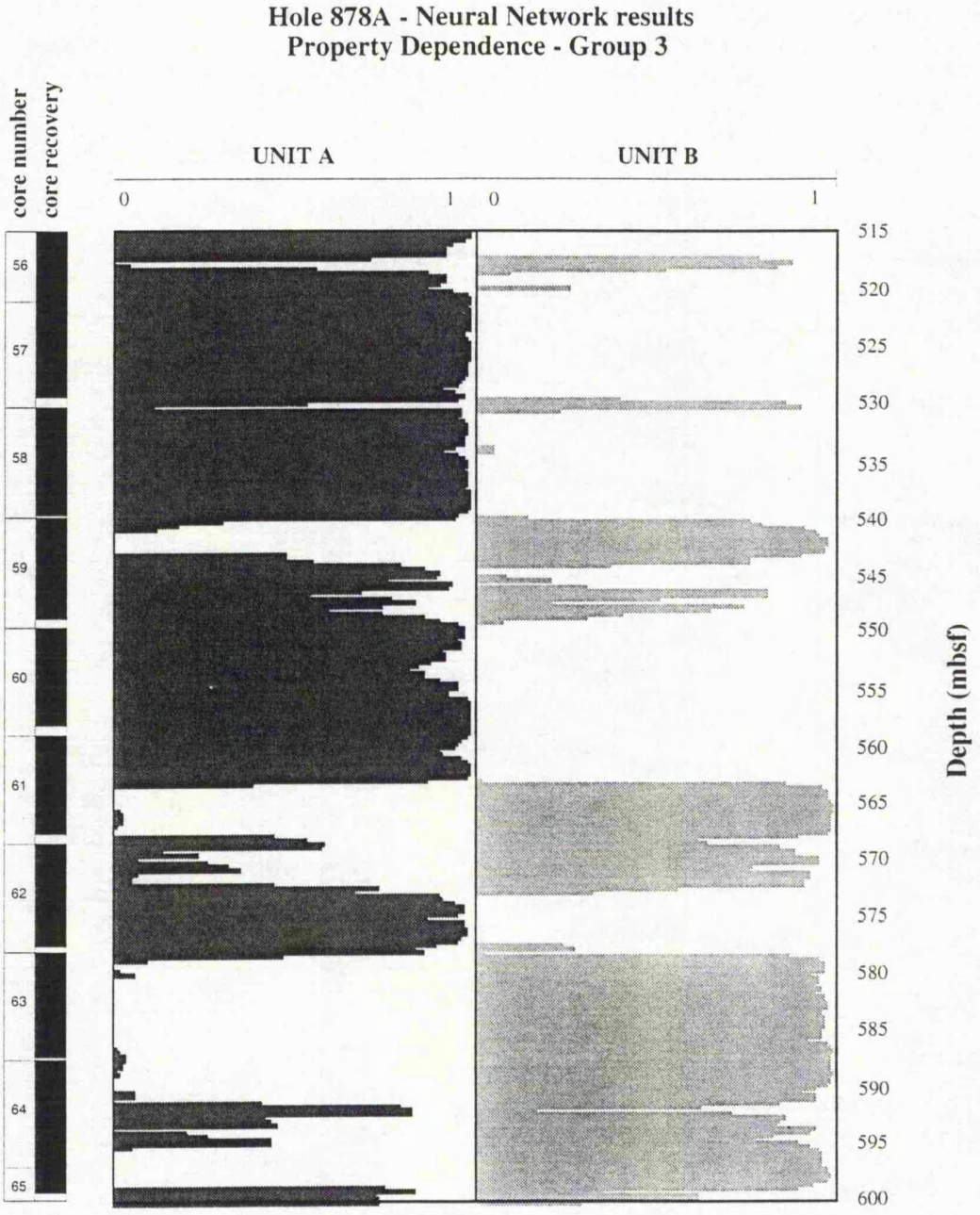


Figure 4.20 - Neural Network results for lithofacies identification using Group 3 log curves from property dependence analysis in ODP Hole 878A.

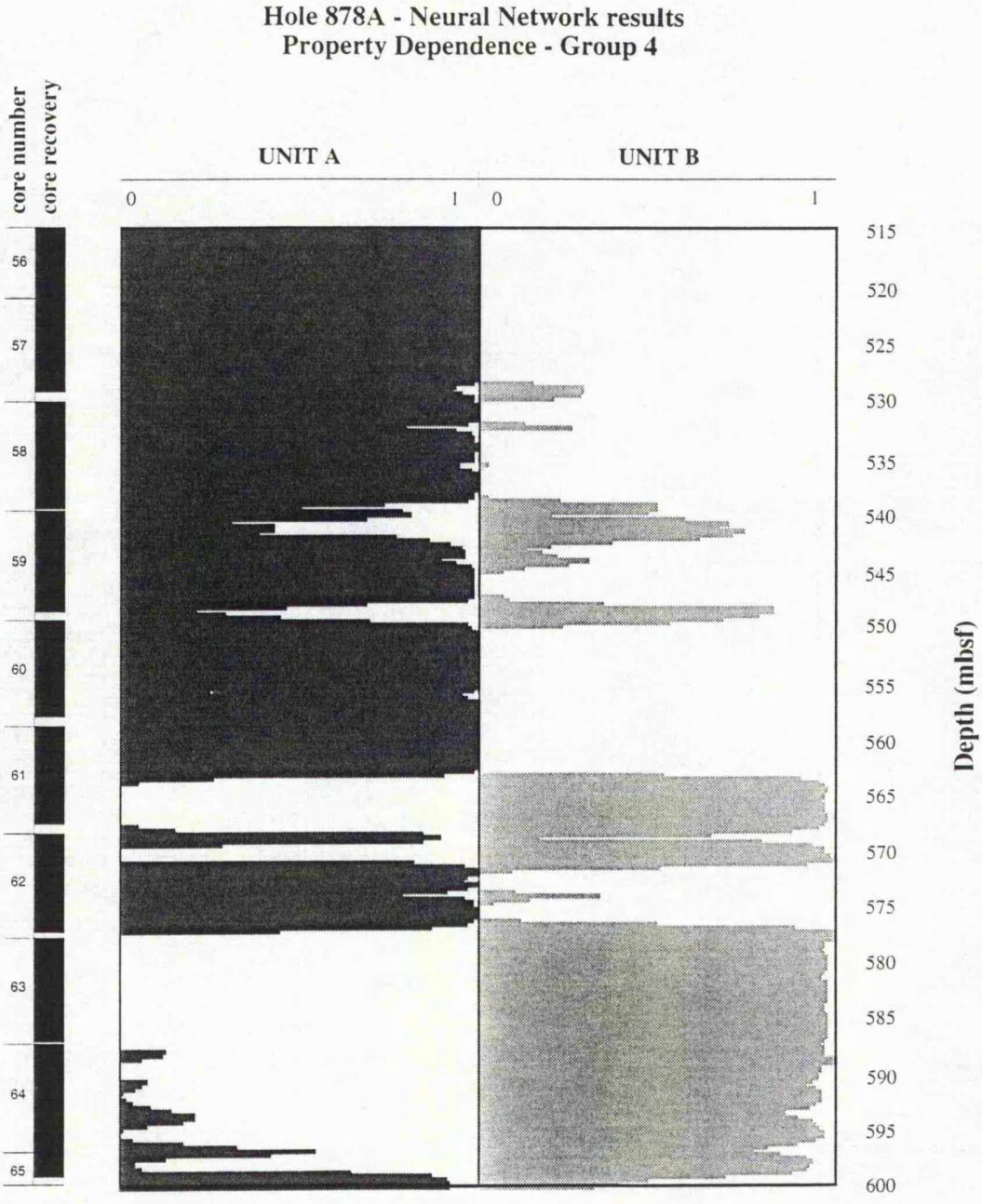


Figure 4.21 - Neural Network results for lithofacies identification using Group 4 log curves from property dependence analysis in ODP Hole 878A.

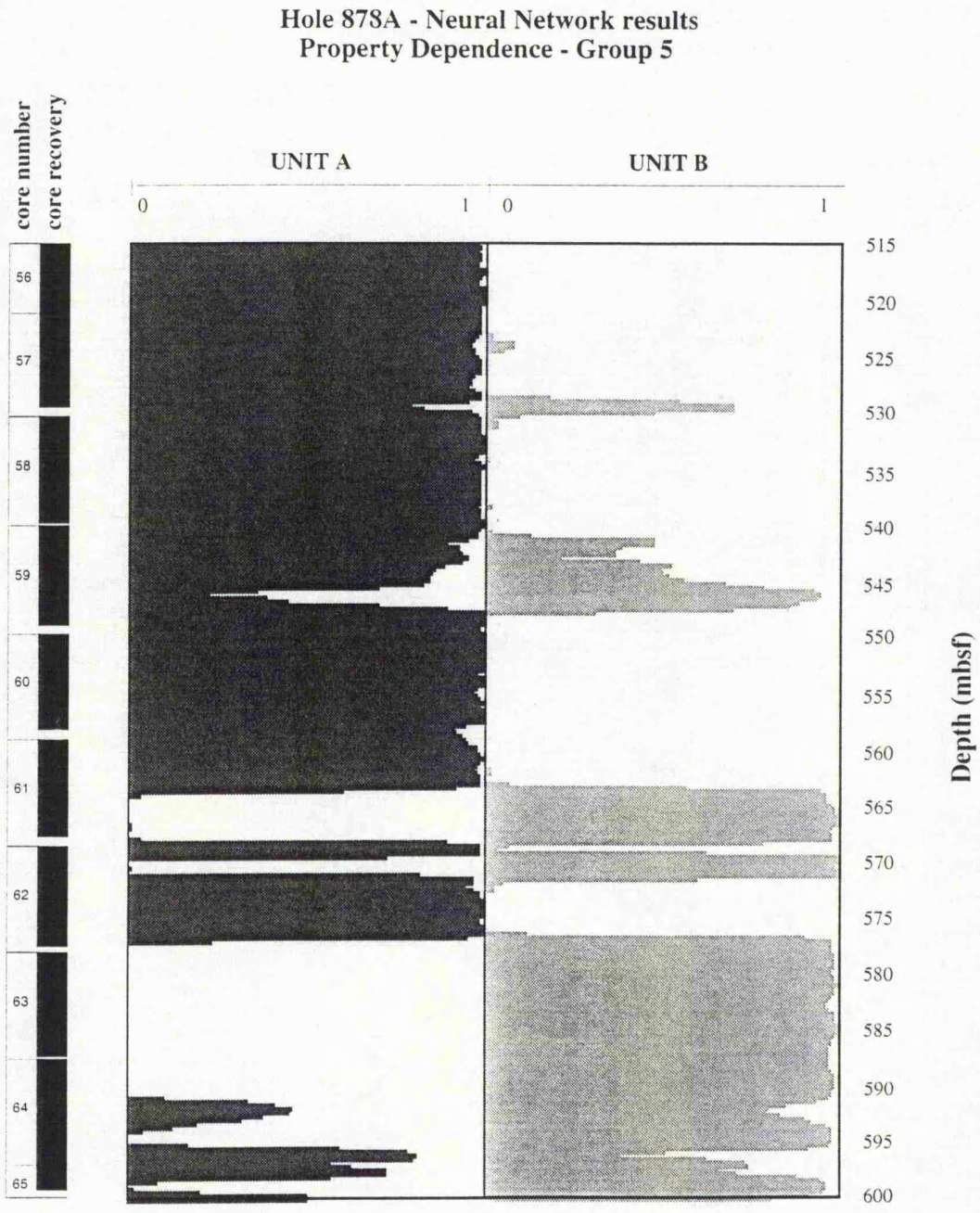


Figure 4.22 - Neural Network results for lithofacies identification using Group 5 log curves from property dependence analysis in ODP Hole 878A.

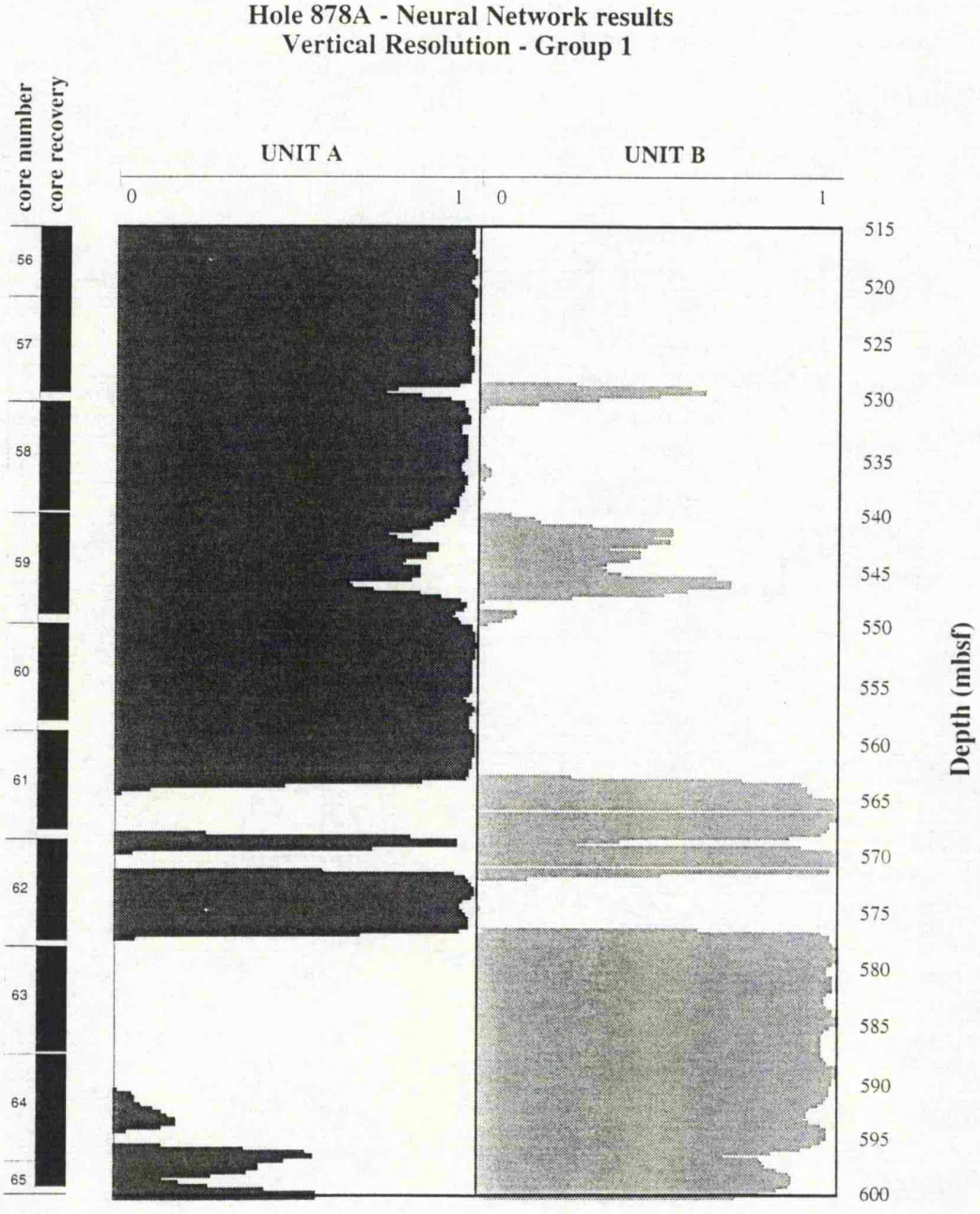


Figure 4.23 - Neural Network results for lithofacies identification using Group 1 log curves from vertical resolution and depth of investigation analysis in ODP Hole 878A.

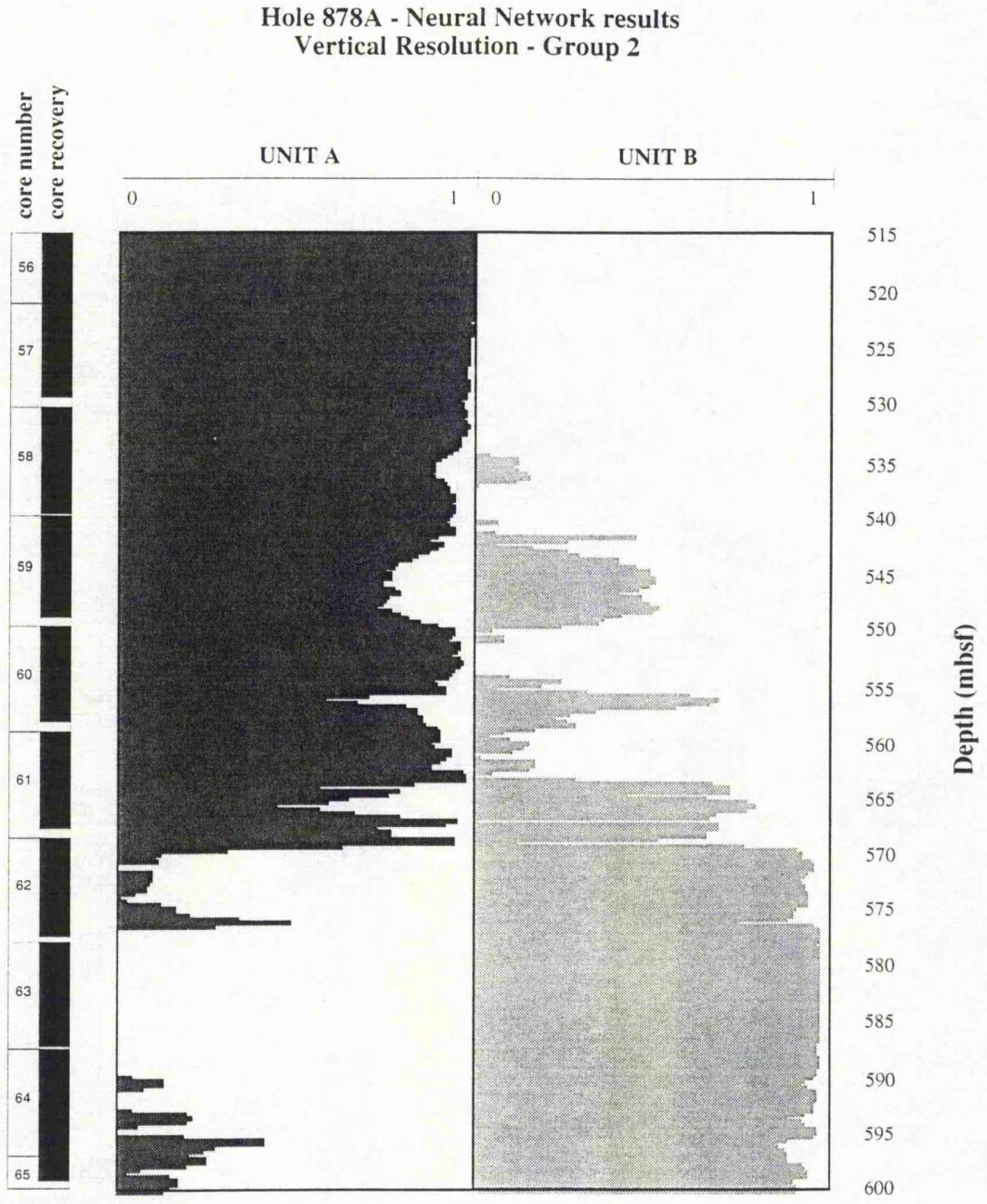


Figure 4.24 - Neural Network results for lithofacies identification using Group 2 log curves from vertical resolution and depth of investigation analysis in ODP Hole 878A.

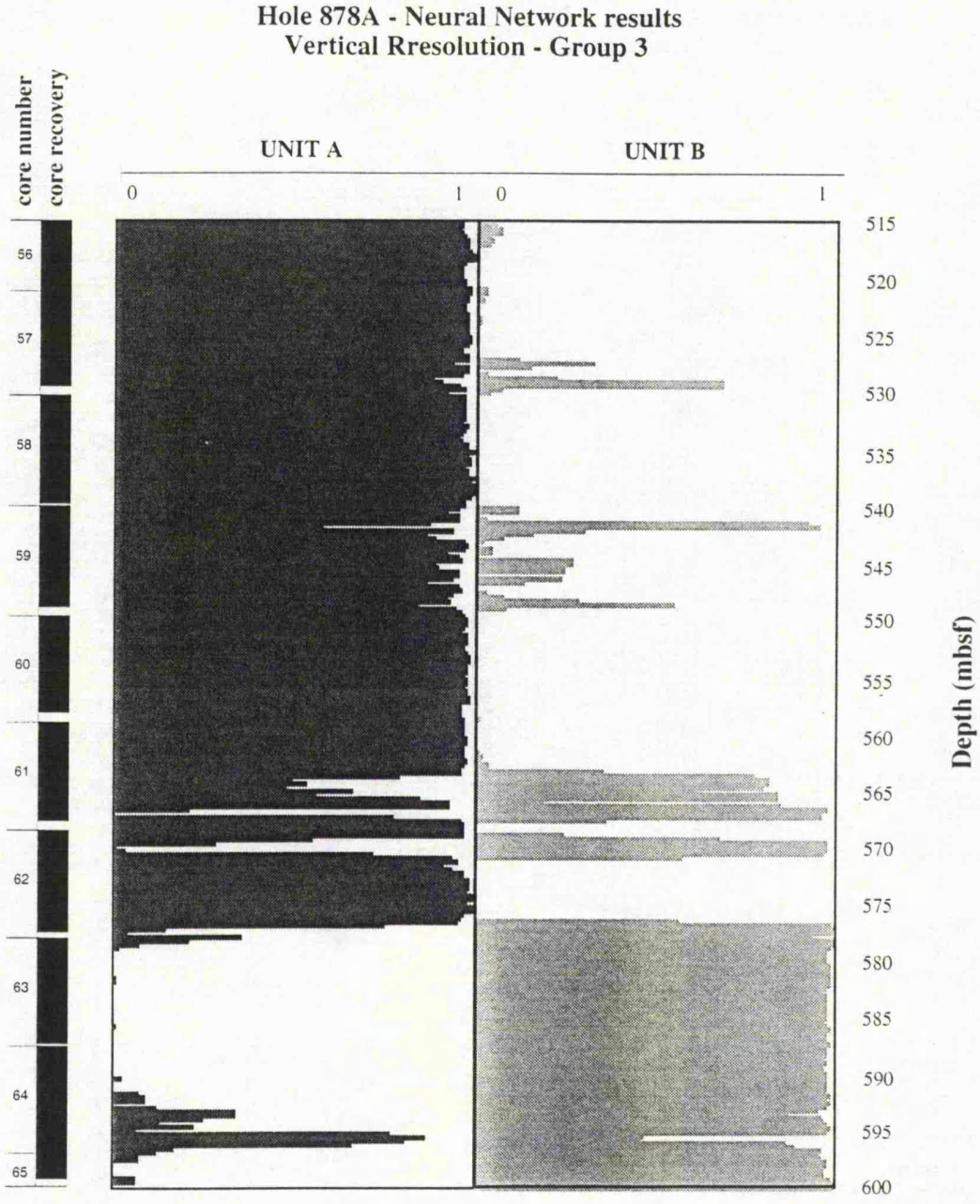


Figure 4.25 - Neural Network results for lithofacies identification using Group 3 log curves from vertical resolution and depth of investigation analysis in ODP Hole 878A.

of Si and Ca matrix are observed. These layers were not well identified in the analysis with the other groups.

Table 4.5

Group I	Group II	Group III
AL2O3	DT	THOR
SIO2	SFLU	URAN
CACO3		POTA
FEO		RHOB
TIO2		
	average vertical resolution	
1.5 meter	0.75 meter	0.4 meter

Table 4.5 - Vertical resolution groups for back-error propagation Neural Network analysis in ODP Hole 878A.

4.5.4 - Oilfield Holes A and B

4.5.4.1 - The lithofacies characterisation

Two oilfield holes from SE Brazil are used here to test the application of the backpropagation Neural Network in identifying lithofacies in an offset hole. The geological sequence in both holes is given by a sandstone reservoir within a shale/claystone sequence. Hole A (Figure 2.8) shows a consistent sandstone reservoir 25 metres thick with high resistivity ($\approx$ 100Ωm) and low density. Hole B is characterised by a sequence of interbedded sandstone and shale between X325 - X354 metres (Figure 2.9). Sand layers within the reservoir zone range from 2 to 5 metres thick. From a quick look analysis of the log curves, the shale/claystone appears as a uniform sequence throughout the interval.

The log curves in Hole A for the same interval selected in the Discriminant Analysis were used to train the backpropagation Neural Network to predict the two lithofacies present. As in the application of Discriminant Analysis, the Groups resulted from the Cluster Analysis were used as a basis for the lithofacies identification. The structure of the Neural Network for Hole A has 6 neurons in the input layer given by the six log curves and two neurons in the output layer given by the two different lithofacies to be identified (Figure 4.26). The number of neurons in the hidden layer was given by the performance of the Neural Network during training. Table 4.6 shows the result in terms of number of iterations and the minimum error achieved by different Neural Network structures compared with the expected target

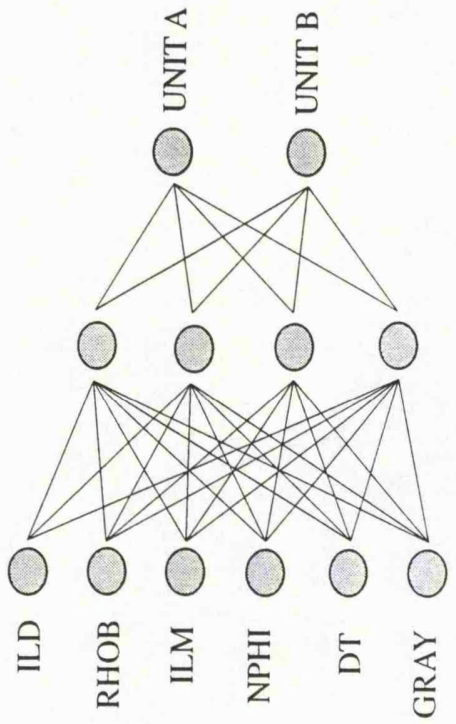


Figure 4.26 - Structure of the Neural Network for lithofacies identification in HoleA.

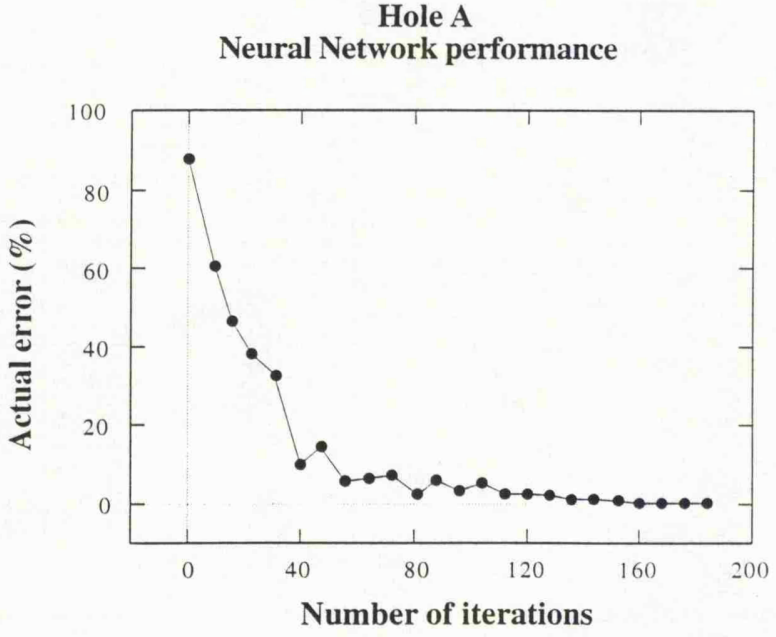


Figure 4.27 - Neural Network performance during the training process for lithofacies identification in Hole A.

Hole B
Neural Network and Log Derived Petrophysical Parameters Curves

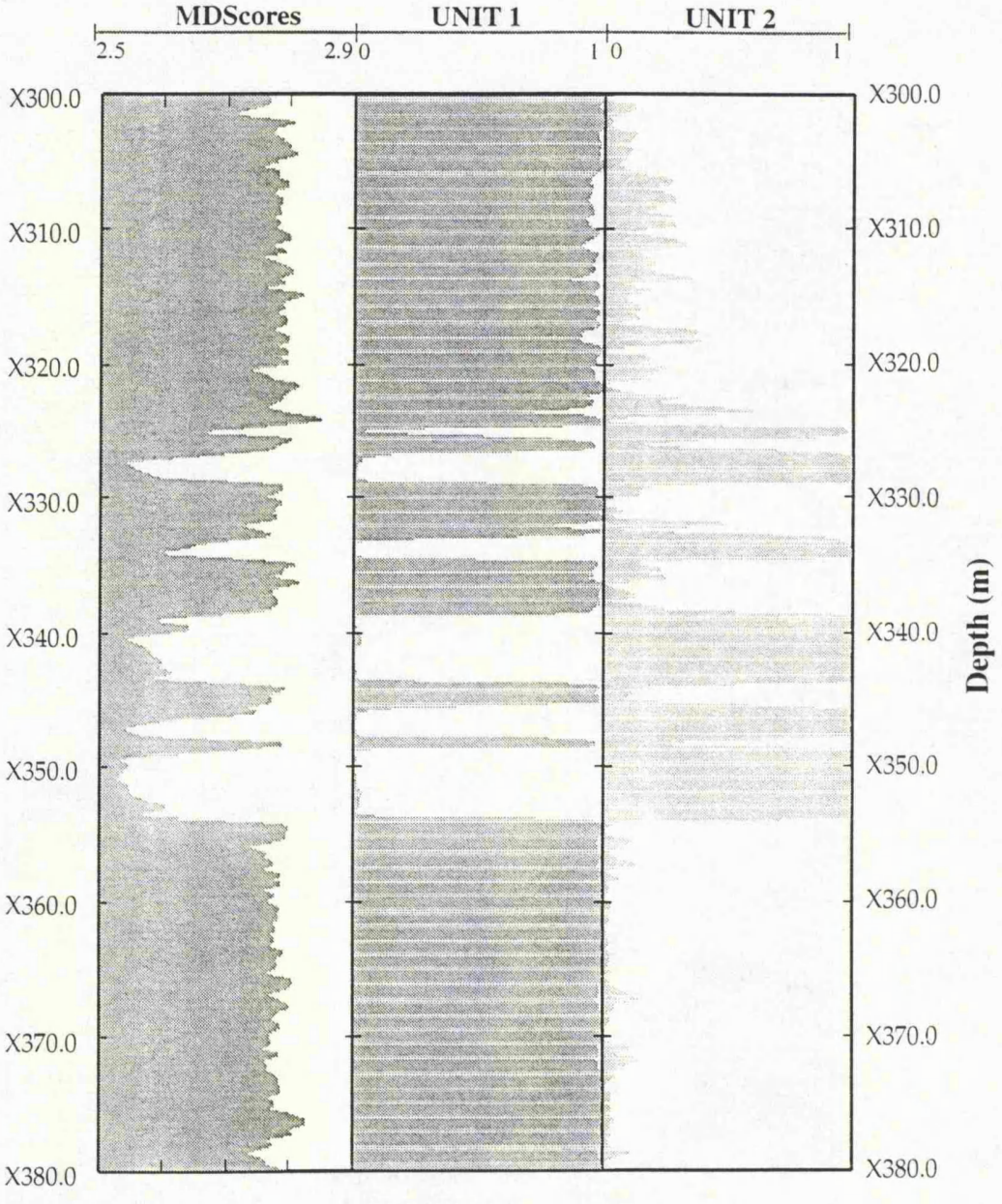


Figure 4.28 - Neural Network predicted lithofacies characterisation and Discriminant Scores in Hole B.

during training. Figure 4.27 illustrates the performance of the Neural Network for the minimum error achieved in Table 4.6.

Table 4.6

Number of neurons in the hidden layer	Number of iterations	Minimum error achieved
3	210	1.85%
4	187	0.38%
5	202	1.96%
6	173	1.86%
7	257	1.36%

Table 4.6 - Number of iterations and minimum error achieved for different number of neurons in the hidden layer during training stage in Hole A.

After training, the final weights learned by the Neural Network from Hole A were applied to predict the lithofacies in Hole B. The result for the Neural Network predicted lithofacies is shown in Figure 4.28. The discriminant scores were also plotted for comparison. Both lithofacies were well defined by the Neural Network. The interbedded sand/shale reservoir zone was also well characterised with all the sand layers being identified. However, the upper part of the interval shows some variation in the characterisation. It is observed that the Neural Network identified both lithofacies between X305 - X324 metres. Although the classification given by the Neural Network shows that the shale/claystone sequence is the strongest for that section, the identification of some sand can be a result of facies changes in the shale/claystone sequence for the upper part of the interval in Hole B, which was not characterised by the discriminant scores. This variation had already been observed in the Groups of the Cluster Analysis (Figure 3.27), where Groups 1 and 2 were generated.

The result produced by the Neural Network can be considered reasonable good, where the main lithofacies were characterised and heterogeneities in one of the lithofacies were also identified.

4.6 - The generation of physical property logs

One of the important tasks for geologists and geophysicists during formation characterisation and evaluation is to provide estimates of the physical properties for the hole being investigated. Where core recovery is extensive throughout the intervals

investigated, physical property values can be obtained for all sections and different lithofacies through direct core measurement. However, where there are gaps in the core recovery or even where some holes in the same area have no core at all, then a method of predicting physical properties must be derived.

There are many techniques available for physical properties prediction. Algorithms using simply the log responses, geochemically derived mineral abundance and of course direct core measurements are among them. However, all generally require the formation to be zoned into intervals of different petrophysical characteristics in order for the methods to come close to providing satisfactory results (Jenner and Baldwin, 1994). Another important aspect to be pointed out is the geographic dependence of these methods. When they are applied elsewhere, they are less likely to be successful.

As in the classical procedure for most Neural Networks, physical properties are estimated here by training the system with measured core data from recovered sections. During training, the objective is to reduce the difference between the core measurements, which are used as target, and the Neural Network result. This is performed by adjusting the connecting weights in the Neural Network (Hetch-Nielsen, 1990). The final weights obtained are then applied to the whole interval for comparing the results between core and Neural Network estimates and also to generate physical property logs.

The structure of the Neural Network used for this procedure (Figure 4.29) is slightly different from that used for lithofacies characterisation. The number of the neurons in the input layer is given by the log curves and the output layer is defined by the physical properties to be estimated. However in this case, for each set of output values of physical properties, there are a few sets of input values used. Bearing in mind that the core measurements were taken from single or discrete points and the log data is an average over their vertical resolution, the use of a "windowing" technique allows for resolution matching as well as compensating for any minor depth matching error. The number of depth windows in the input layer depends on the vertical resolution of the logs used. Since the log measurements were taken at intervals of 0.15 metres and the average vertical resolutions of the logs used are 0.7 and 0.6 metres for the first and second examples respectively, different numbers of depth windows were necessary for each input and each output training pair.

Another important aspect is the reliability of the results. Comparison of generated output with core data can provide some information on how well the Neural Network has been trained. However, this does not provide any information about the quality or accuracy of the results estimated in intervals with no core recovery. In order to give a confident level for the result, the Neural Network is also trained to predict "known" quantities, which are different physical properties measurements previously

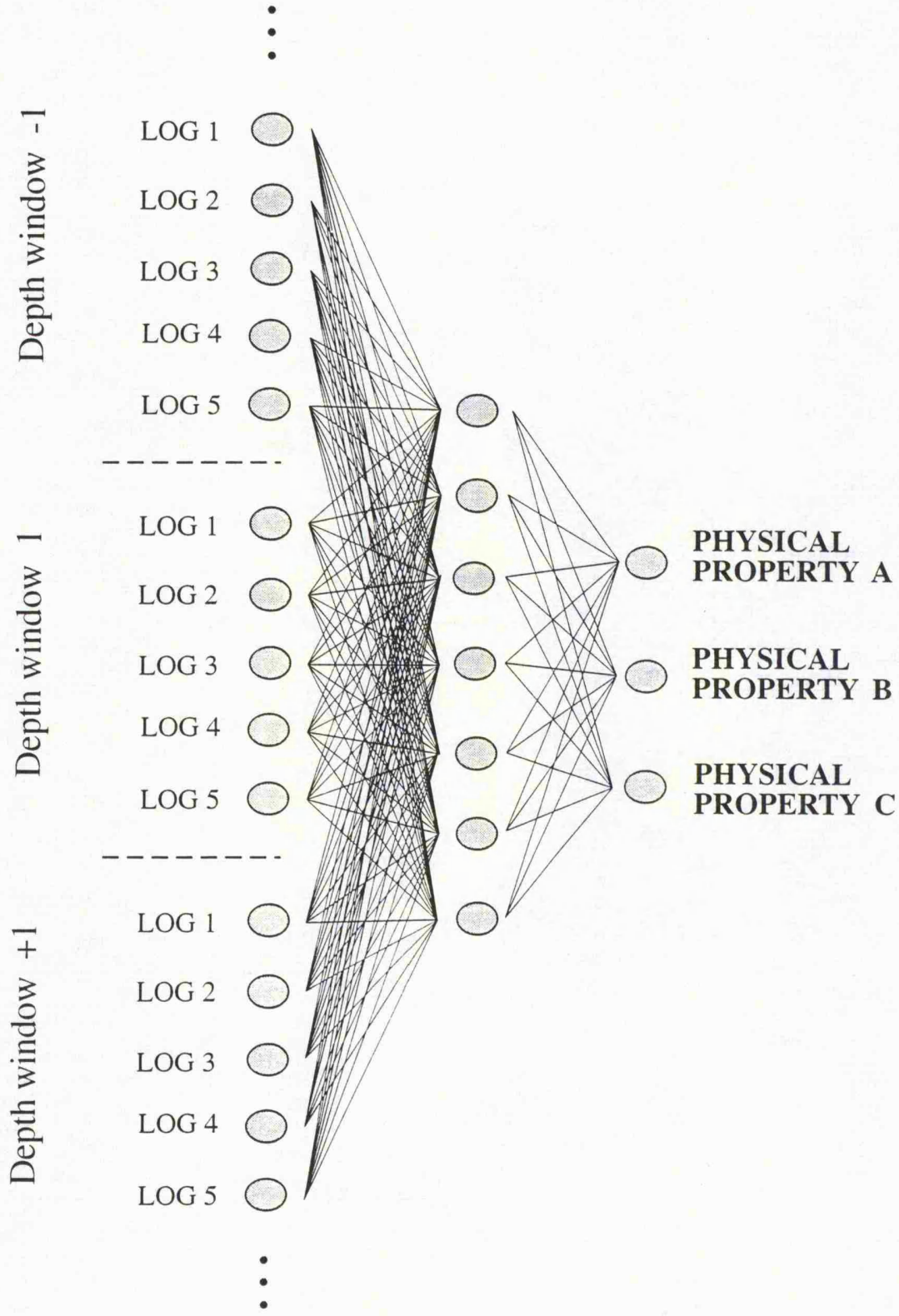


Figure 4.29 - Structure of the Neural Network for physical property and petrophysical parameter estimation showing the "windowing" technique.

obtained with other methods. When intervals where the Neural Network prediction agrees well with the "known" quantities, the network has been trained in a way to correctly identify the meaningful variations in the log data.

Because the Neural Network is being trained to predict physical properties on the strength of log response, the logs must be sensitive to changes in the physical properties. Changes in lithology, resulting in changes in physical properties which are not reflected in the log response will not be correctly predicted by the Neural Network.

4.6.1 - ODP Hole 807C

In ODP Hole 807C a number of physical properties were measured on core samples for the whole interval drilled. Between 1270-1400 mbsf, due to reduced core recovery (65%) complete physical property characterisation was not possible for the whole interval. Measurements of water content, grain density and porosity were taken at different depths depending on core recovery and core conditions. Below 1380 mbsf, due to the poor quality of the cores, no physical property measurements were made (Kroenke et al., 1991). The physical property measurements obtained in ODP Hole 807C are shown in Table 4.7.

The structure of the Neural Network used in this case is given by 4 neurons at the input layer, which represent the four log curves (ILM, SFLU, DT and RHOB) used as input. The output layer has 3 neurons representing the three different physical properties to be estimated. The vertical resolution of the logs range from 0.4 metre (DT and RHOB) to 1.5 metres (ILM). In this case, we used a depth window of five measurements spaced at 0.15 metre in the training process. As mentioned before, this allows for resolution matching and any depth shift.

After training the computed weights were applied to the whole interval and a comparison of the Neural Network predicted physical properties to the actual core measurements is shown in Figures 4.30 to 4.32. The figures show the 1:1 ratio line and the Reduced Major Axis (RMA) solution. This solution is obtained by the fit which intersects the mean points of each variable and which slope is given by $\lambda=s_y/s_x$ (s_y and s_x standard deviations). A quantitative measure of goodness of fit between the two solutions can be obtained by computing the standard normal deviate (Z) and verifying the hypothesis that $\lambda_1=\lambda_2$. Therefore, $Z= (\lambda_1-\lambda_2)/(s_x^2+s_y^2)^{(1/2)}$, where s_x^2 and s_y^2 are the variances of x and y variables respectively. For a 95% of confidence limit, Z should be less than 1.96. In ODP Hole 807C the values for Z are 0.007, 0.007 and 2.42 for water content, porosity and grain density respectively. It shows that only for grain density the hypothesis is not verified.

In general, a good agreement is observed for low porosity, low water content and high grain density. These values are related to the carbonate sequence (Unit A) at the top of the interval which shows less variation in the log measurements. The results are confirmed when Neural Network predicted physical property logs are generated for the whole interval (Figure 4.33). It is possible to observe the good agreement between core and Neural Network data for the top of the interval. Below 1351 mbsf, the Neural Network predicted curves are still able to follow the main trend present in core data, although it does not match all core measurements. One of the reasons for that is the reduced amount of core data used to train the Neural Network for the lithofacies variation within the interval 1351-1379 mbsf (siltstone/claystone sequence - Unit B). The upper part of the interval (carbonate sequence) was trained with a greater number of core measurements.

Table 4.7

ODP Hole 807C - Physical property measurements

Core number/section	Depth (mbsf)	Water Content (%)	Porosity (%)	Grain density (g/cc)
63R - 1	1271.08	9.2	19.8	2.74
63R - 2	1272.87	8.9	19.2	2.73
66R - 1	1301.07	11.9	24.2	2.74
66R - 2	1302.04	12.1	24.3	2.72
66R - 3	1303.64	12.0	24.2	2.73
67R - 1	1310.25	10.4	21.7	2.72
67R - 2	1312.09	9.5	20.1	2.72
67R - 3	1313.23	10.1	21.1	2.71
67R - 4	1314.82	10.5	21.8	2.71
67R - 5	1315.87	11.4	23.3	2.73
68R - 2	1320.88	11.4	23.1	2.70
68R - 3	1322.68	9.6	20.2	2.71
69R - 2	1331.36	9.7	20.5	2.73
69R - 3	1331.83	9.3	19.8	2.72
69R - 4	1334.16	10.0	21.1	2.74
69R - 5	1334.95	9.8	20.8	2.74
70R - 1	1339.09	10.1	21.0	2.71
70R - 2	1340.98	8.6	18.7	2.74
70R - 3	1342.37	10.8	22.4	2.73
70R - 4	1343.42	10.4	20.8	2.58
71R - 1	1348.45	11.0	22.5	2.71
71R - 2	1350.80	11.6	23.6	2.73
72R - 2	1359.62	39.5	48.6	2.44
73R - 1	1368.07	31.3	42.7	2.45
73R - 2	1369.45	21.7	36.1	2.67

Table 4.7 - Some of the Physical property measurements for ODP Hole 807C between 1270-1400 mbsf.

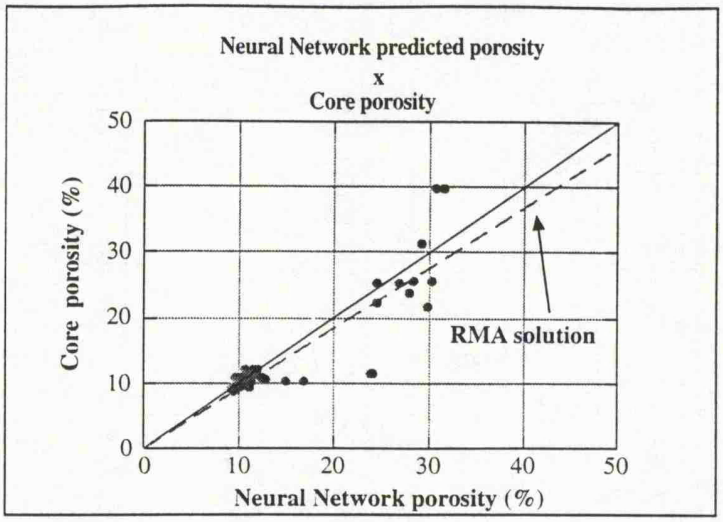


Figure 4.30 - Comparison of Neural Network predicted porosity with core porosity in ODP Hole 807C.

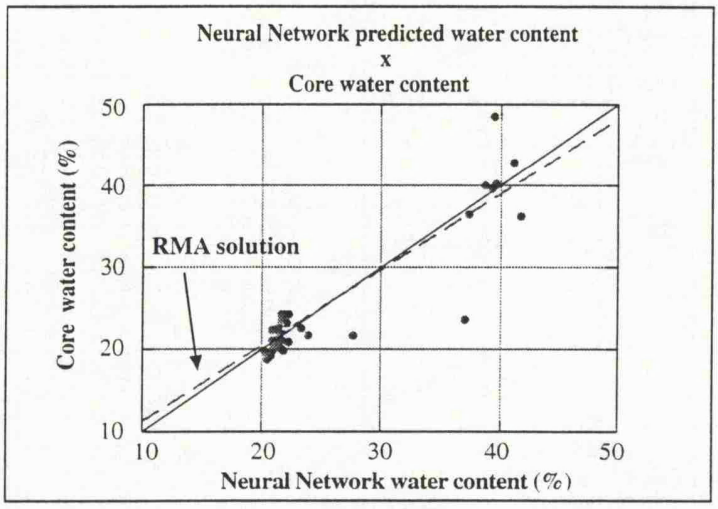


Figure 4.31 - Comparison of Neural Network predicted water content with core water content in ODP Hole 807C.

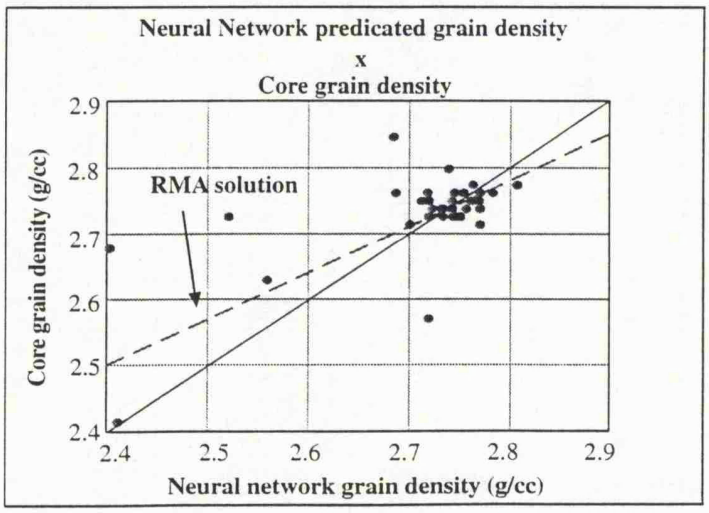


Figure 4.32 - Comparison of Neural Network predicted grain density with core grain density in ODP Hole 807C.

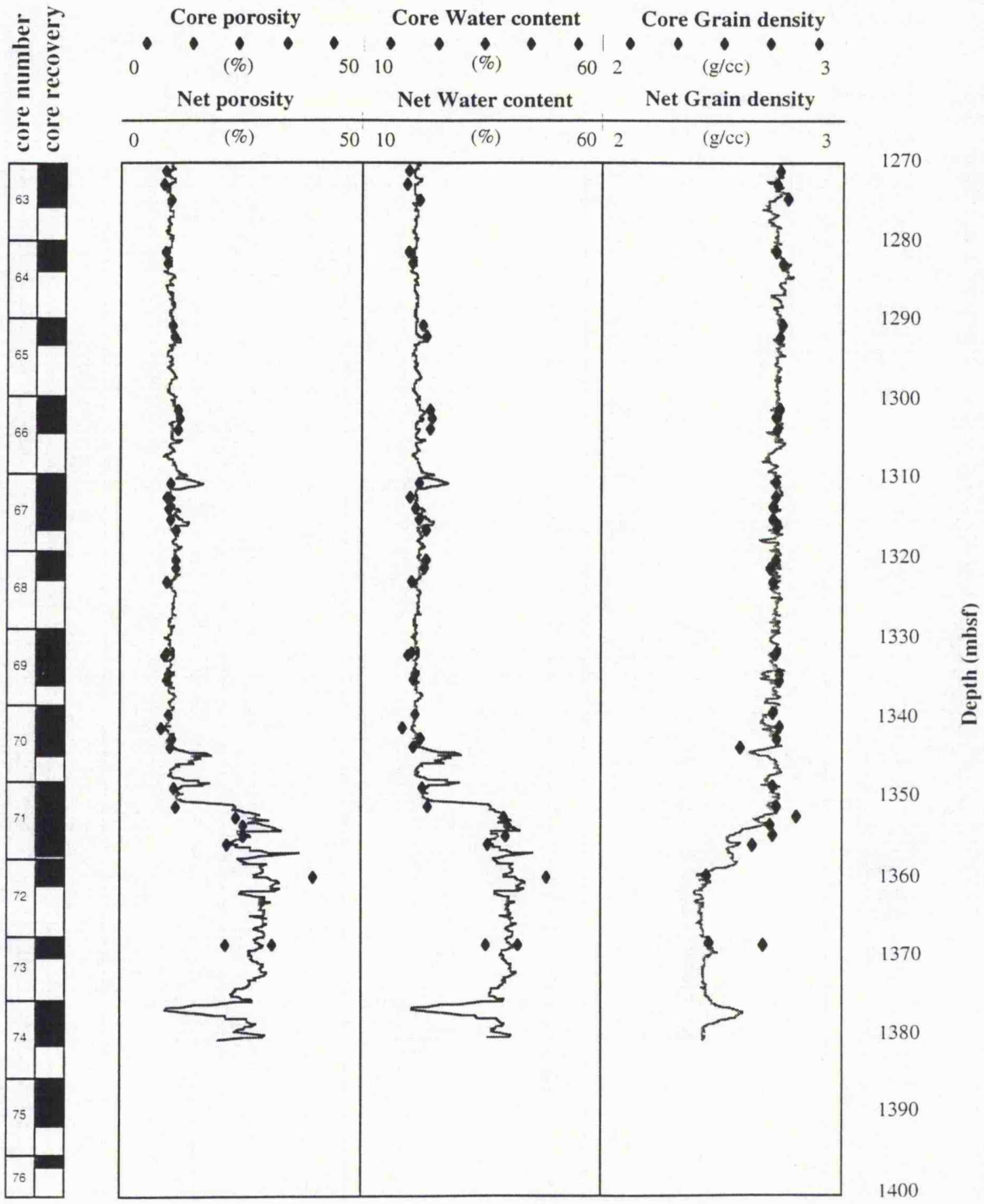


Figure 4.33 - Log curves for Neural Network predicted physical properties and core measurements in ODP Hole 807C.

4.6.2 - ODP Hole 878A

In ODP Hole 878A, grain density, porosity and water content measurements were also taken from cores and are listed in Table 4.8.

The structure of the Neural Network used in this case is the same as used in the previous example. Four neurons are present at the input layer, which represent the four log curves (DT, RHOB, NPHI and SFLU) and three neurons at the output layer represent the three different physical properties to be estimated. The vertical resolution of the logs ranges from 0.4 metre (RHOB) to 0.75 metre (DT and SFLU). In this case, we used a depth window of three measurements spaced at 0.15 metre in the training process was used.

Despite the complex changes in lithofacies (mainly given by Si and Ca content and grain size variation within the polymitic breccia) the good core recovery and density of measurements along the whole interval allow a good correlation between predicted physical properties and core measurements (Figures 4.34 to 4.36). Again, the RMA solutions were obtained and compared with the 1:1 ratio lines. The standard normal deviate show that only for grain density the $\lambda_1=\lambda_2$ hypothesis is not valid. The grain density cross-plot appears to be more scattered probably due to the strong variations in the breccia matrix along the interval. These lithofacies variations within the breccia affect the log curves and consequently the physical property estimation. The Neural Network predicted physical property logs for the whole interval are observed in Figure 4.37. The main variation in the physical property characteristics are represented. As expected because of their close relation, porosity and water content show nearly the same variations along the interval, following in general the variations observed in core measurements and lithofacies changes.

4.6.3 - Oilfield Holes A and B

The objective here is to obtain predicted petrophysical parameters using backpropagation Neural Network in an uncored hole using the well log curves and petrophysical parameters measured in a cored hole.

Core measurements were obtained in Hole A for clay volume (VCL). Effective porosity (PHIE) and water saturation (SW) (Figure 4.38). Using the log curves in Hole A, the Neural Network was trained to predict petrophysical parameters from core data. After training, the weights learned by the Neural Network for that hole were applied to predict the same petrophysical parameters in Hole B, where no core measurements were available.

The results for the Neural Network predicted petrophysical parameters log curves are shown in Figure 4.39. In general, a good agreement between Neural Network predicted and log derived VCL, PHIE and SW are observed. In the upper section of the interval, however, there is a clear mismatch between both. The

Table 4.8

ODP Hole 878A - Physical property measurements

Core number/section	Depth (mbsf)	Grain density (g/cc)	Porosity (%)	Water content (%)
56R - 3	515.58	2.83	31.6	16.9
56R - 5	518.46	2.75	26.6	13.1
56R - 3	519.20	2.79	26.6	13.1
57R - 1	521.57	2.80	25.3	12.2
57R - 3	524.42	2.79	25.7	12.4
57R - 5	527.71	2.74	24.1	11.6
58R - 1	531.60	2.75	25.4	12.1
58R - 3	534.30	2.74	24.1	11.4
58R - 5	536.48	2.74	25.5	12.4
58R - 7	538.87	2.77	29.3	14.2
58R - 7	539.35	2.77	24.0	11.3
59R - 1	540.33	2.79	26.1	12.0
59R - 3	543.04	2.80	25.1	11.6
59R - 4	544.86	2.83	24.5	10.9
59R - 5	545.67	2.78	26.1	12.2
59R - 6	546.78	2.79	26.7	12.6
59R - 7	548.13	2.79	26.1	12.0
60R - 1	550.44	2.76	25.3	11.9
60R - 1	551.07	2.75	28.6	13.9
60R - 2	551.77	2.85	26.2	11.8
60R - 3	553.67	2.76	24.6	11.4
60R - 4	554.94	2.77	25.2	11.5
60R - 5	555. 67	2.74	22.0	10.1
61R - 2	560.84	2.79	25.0	11.8
61R - 4	564.63	2.79	24.8	11.9
62R - 1	568.88	2.72	27.8	13.8
62R - 2	570.59	2.74	20.9	9.6
62R - 4	573.26	2.80	25.5	12.1
62R - 7	577.43	2.72	17.5	7.5
63R - 1	578.87	2.74	21.0	9.5
63R - 1	579.41	2.73	20.9	9.4
63R - 6	585.36	2.74	20.1	8.9
64R - 1	587.73	2.76	18.7	8.1
64R - 2	589.68	2.73	16.7	7.2
64R - 5	594.24	2.79	23.1	10.7
65R - 1	597.23	2.81	24.9	11.6
65R - 1	597.61	2.82	24.2	11.2

Table 4.8 - Physical property measurements for ODP Hole 878A between 515 - 600 mbsf.

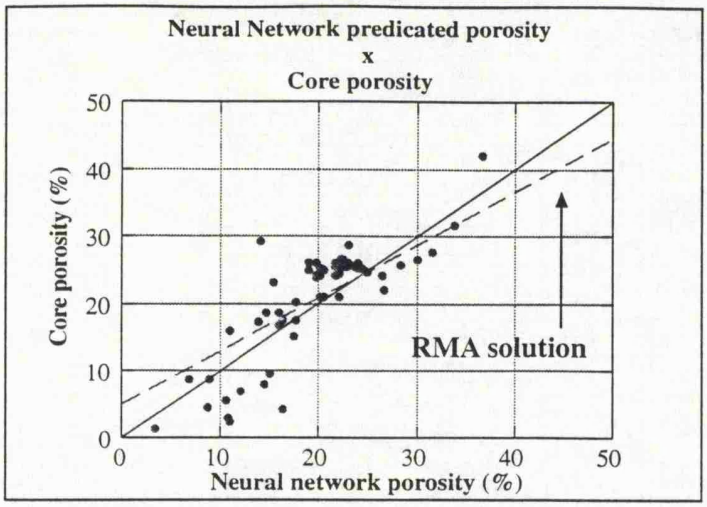


Figure 4.34 - Comparison of Neural Network predicted porosity with core porosity in ODP Hole 878A.

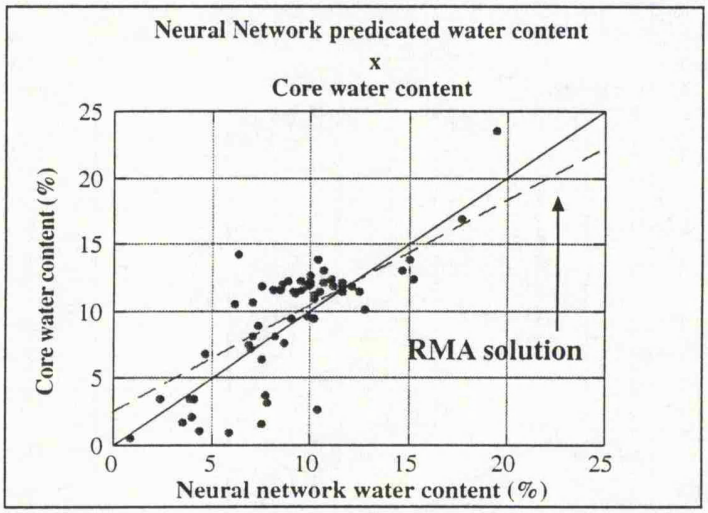


Figure 4.35 - Comparison of Neural Network predicted water content with core water content in ODP Hole 878A.

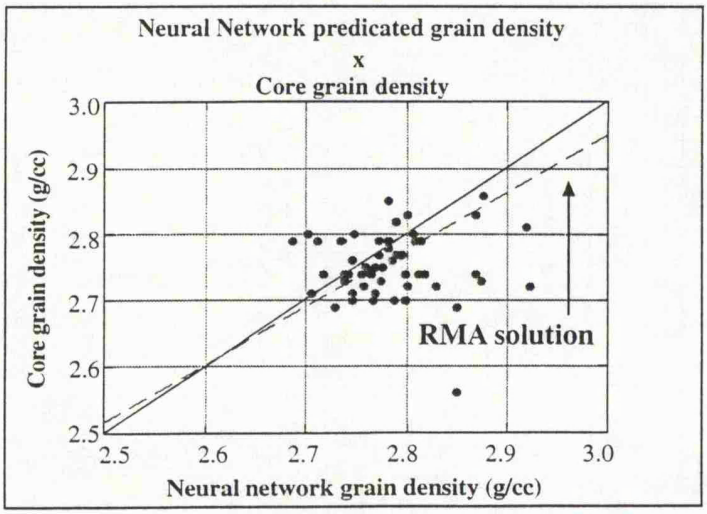


Figure 4.36 - Comparison of Neural Network predicted grain density with core grain density in ODP Hole 878A.

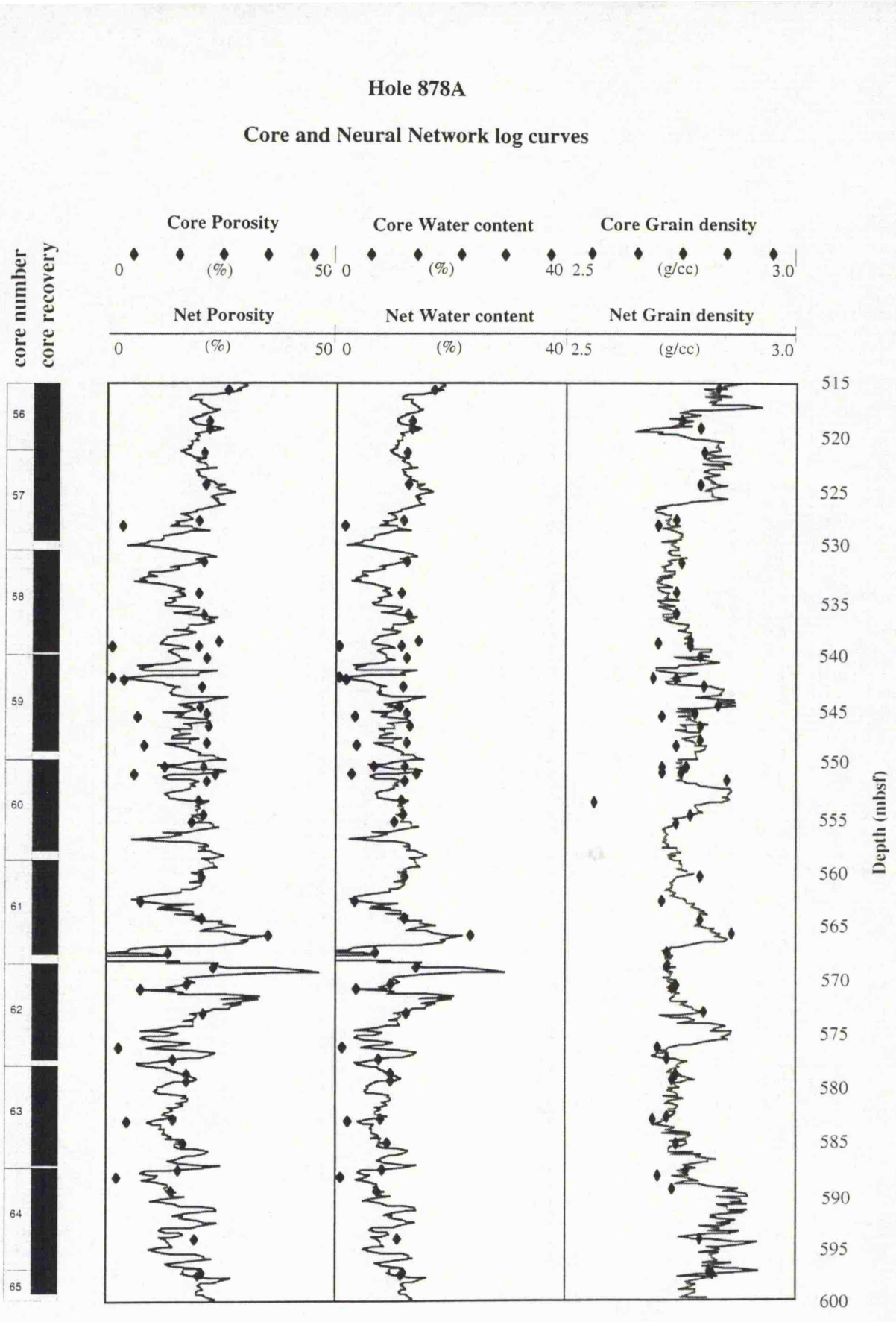


Figure 4.37 - Log curves for Neural Network predicted physical properties and core measurements in ODP Hole 878A.

mismatch is observed for all three parameters, suggesting that it could be caused by changes in facies between the holes which affected the log curves in Hole B and that have not been trained by the Neural Network in Hole A.

It shows the same problem already observed in the lithofacies classification of the same hole (section 4.5.4). Here, despite the considerable amount of core data in Hole A to train the Neural Network, those do not entirely describe the changes in petrophysical characteristics in the upper part of Hole B.

For the other parts of the interval, there is a good match between both results. Figure 4.40 shows the cross-plot of the Neural Network and log derived petrophysical parameters within the reservoir zone (between X325 - X354 metres) in Hole B. Here, the standard normal deviate for the three petrophysical parameters show that in all of them the $\lambda_1=\lambda_2$ hypothesis is valid. Despite changes in the reservoir characteristics, a good match between both results is observed.

In this case, hole-to-hole prediction of petrophysical parameters seems to work well, producing reasonable results. But again, lateral changes of facies which are not trained by the Neural Network require skill in finding the best net structure to provide good results.

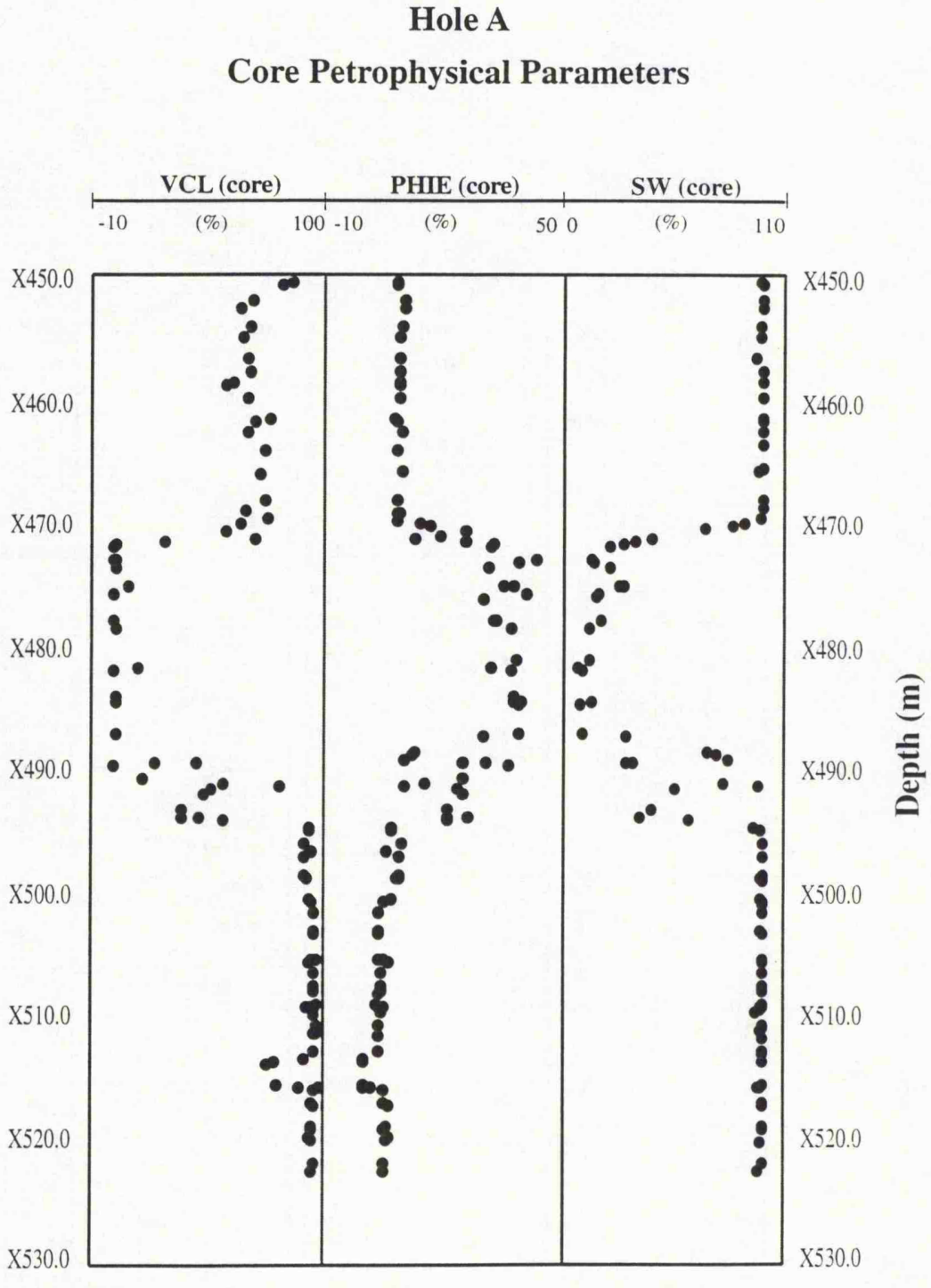


Figure 4.38 - Petrophysical parameters measured from cores in Hole A.

Hole B
Neural Network and Log Derived Petrophysical Parameters Curves

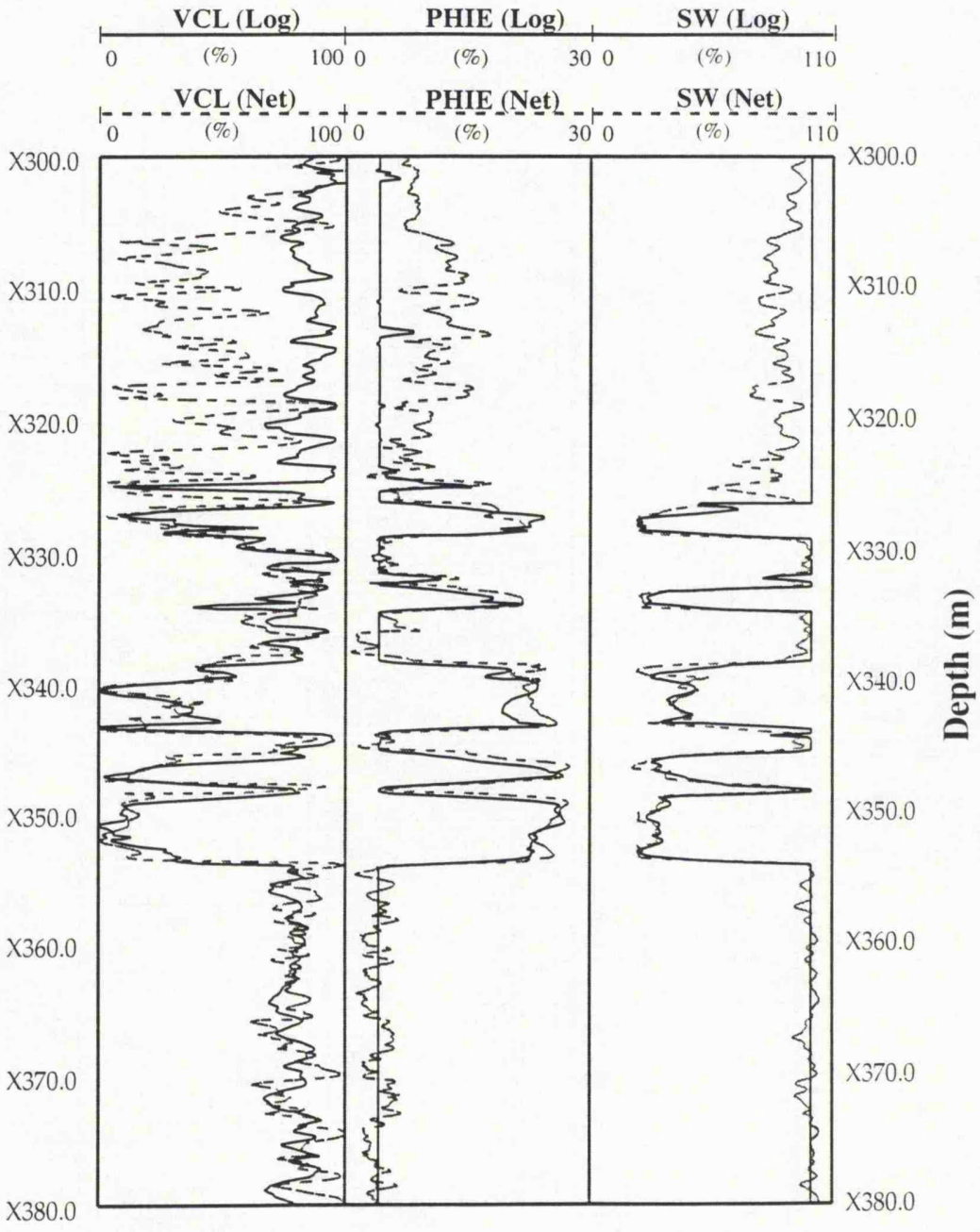
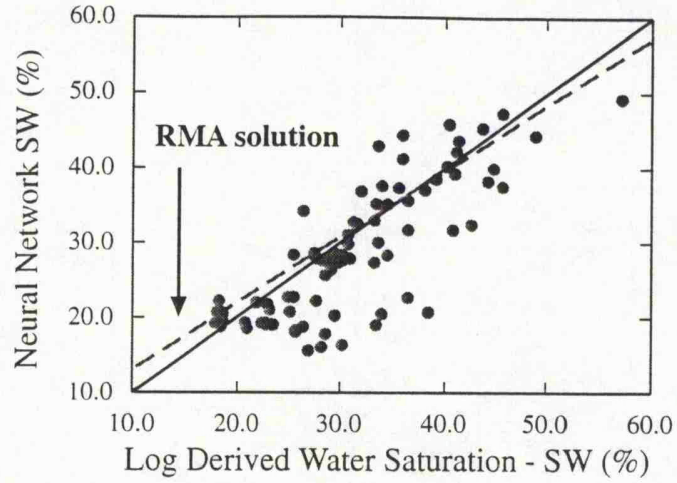
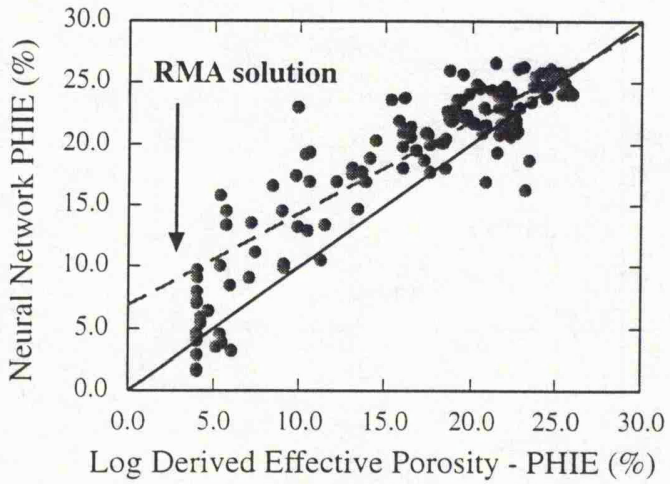


Figure 4.39 - Neural Network predicted petrophysical parameters and log derived petrophysical parameters in Hole B.

Hole B





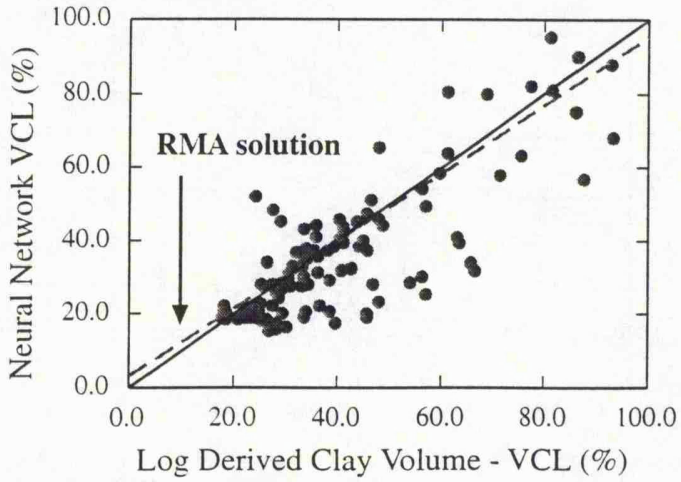


Figure 4.40 - Cross-plots of the Neural Network predicted petrophysical parameters and log derived petrophysical parameters for the reservoir zone in Hole B.

4.7 - Summary and Discussion

The Neural Network has been shown to give good results both for the lithofacies characterisation and for physical property estimation. The results for lithofacies characterisation from well logs were compared with other supervised classification methods; in this case statistical Discriminant Analysis. In terms of physical property estimation, Neurocomputing can also be considered a powerful tool for petrophysical evaluation. The results for the three examples in this work also show that petrophysical logs can be obtained for uncored intervals with a reasonably high degree of success.

Based on results obtained from two different holes from the Ocean Drilling Program, lithofacies characterisation using the Neural Network seems to work better than the other techniques using a large set of log curves is used. When the number of log curves is decreased, the interpretation given by the Neural Network may lose quality and resolution. In this case and also for a reduced number of input parameters (log curves) there are two ways to keep quality in the results. The first one is to select a set of log curves which best represent the lithologic variations observed. This is not a trivial task, since many times only a combination of log curves will best represent the variations. The second option is to change the structure of the Neural Network. In this case, an increase in the number of neurons in the hidden layer or even an increase in the number of hidden layers may help to produce a better partition of the data. Although both choices will lead to a longer training time, the results are likely to improve. In this work we used the first choice as an attempt to keep the computing time as low as possible. It has also been demonstrated that Neural Networks provide good results for hole-to-hole classification. In the example using Oilfield Holes A and B, despite some lateral variation in facies between the holes, the results obtained are reasonably good. Mismatches in the result are expected when these facies variations are strong.

The Neural Network-based model also performed better than the statistical Discriminant Analysis in lithofacies characterisation. Despite its longer training time, the Neural Network model can provide significantly improved performance without any pre-determined formulas or ideas about the data set. While the Discriminant Analysis in ODP Hole 807C gives a misclassification of 22.67% of the total samples, the Neural Network model only gives a misclassification of 2.87%. For ODP Hole 878A, where complex changes in lithofacies are present, again the Neural Network shows an improved result with only 6.20% of misclassified samples. In comparison, the Discriminant Analysis shows 22.87%. The same training sets were used in both techniques. In the case of Oilfield Holes A and B, it is observed that Neural Network

results were able to pick up lateral changes in facies within the shale/claystone sequence at the top of the interval, although these variations in facies were not trained by the net. The same variations were not observed in the discriminant scores, showing that Neural Network had some advantage in the classification.

The physical property estimation based on the quantitative approach using Neural Networks also presents good results, even in complex lithological sequences such as in ODP Hole 878A. An important factor here is the use of a "windowing" technique, which allows not only a more realistic training set for the Neural Network, but also avoids any minor depth matching error between core and log data. The results for both Holes 807C and 878A have shown a very good match between actual core measurements and the Neural Network predicted physical properties. The physical property logs generated by the Neural Network also seem to follow the main lithological variations along both intervals. Again, ODP Hole 878A shows some degree of variation in the generated physical property logs due mostly to the complex lithologies present. In the case of Oilfield Holes A and B, the prediction of petrophysical parameters in an offset hole also showed good results. Clay volume, effective porosity and water saturation were estimated in Hole A using core measurements from Hole A. Variations in facies at the top of the interval, as observed in the lithofacies characterisation, affected the petrophysical characteristics of the section. Therefore, a mismatch between Neural Network and log derived petrophysical parameters is observed for this part of the interval. For the rest of the interval the results produced by Neural Network show a good match with the log derived ones.

As seen in this work, training a Neural Network is a heavy computational problem which really requires a RISC or SPARC workstation or possibly a high specification PC. The use of different Neural Network-based models, such as fuzzy Artmap (Wong et al., 1995) can provide a competitive amount of time/effort required to solve a complex well log problem, but the performance and quality of the results hardly compare with the ones generated by the back-propagation model.

Comparison of Neural Network results with core data and other techniques to predict lithofacies and petrophysical parameters is very important to check the validity of this technique. In this work, the Neural Network proved to give better results and also showed a good match with actual core measurements.

Chapter 5

SUMMARY AND CONCLUSIONS

5.1 - Introduction

Well log data offer an opportunity to interpret geological sequences in the intervals of a borehole where core is not recovered. For many years well log measurements have been used to this end both scientifically and commercially. Recently, the increase in the amount and diversity of log data acquired has offered the chance of enhancing the geological information which can be extracted by well logs. However, the large amount of measurements normally obtained from downhole logging, is somewhat difficult to manage. This study concerns the problem of assessing such large datasets. One of the aims was to investigate the applicability of statistical analysis of well log data, as an aid to lithology determination and heterogeneity characterisation. Another objective was to test the applicability of artificial intelligence through the use of a Neural Network, both in lithofacies determination and heterogeneity characterisation as well as in the prediction of petrophysical parameters. A summary of the results of this work for each of the techniques used is presented below.

5.2 - Statistical analysis

5.2.1 - Principal Component Analysis

Principal Component Analysis sets out to determine the structure of the dataset and improve interpretation through reducing the number of variables to those which represent most of the total variability of the system. Principal Component Analysis has been shown to be very useful in the case of large well logging datasets. The application of this technique in ODP Holes 807C and 878A showed that the number of log curves can be reduced to a smaller number of components (Principal Components) which represent the main variability of the system. In ODP Hole 807C, the initial 10 log curves were reduced to 2 principal components corresponding to 64.5% of the total system variability. Component I was shown to represent variations in the natural gamma-ray radiation of the formations. Low density was another important factor represented by this component. The loadings of the original log curves in each of the principal component gives an idea about the interrelationship of the new variables.

Principal component II represented the aluminium content in the formations which is evident in the high peaks of principal component scores in the volcanic sequence. Scores of both components show a good partition of the entire dataset (Figure 3.5) where the three different lithofacies can be determined. Another important aspect is the characterisation of some heterogeneities, given mainly by the variations in clay content within the siltstone/claystone sequence and high resistivity peaks in the volcanic sequence.

In ODP Hole 878A, the original 13 log curves were reduced to 3 principal components representing 86.5% of the total system variability. Component I responds mainly to variations in the breccia matrix composition. The principal component scores show positive values for the Si based matrix while negative values are observed for the Ca based matrix. Components II and III represent 12.7% and 8.7% respectively. The significance of each component was not clearly defined, although positive loadings in potassium and negative loadings in resistivity for principal component III suggest that this component can be related to secondary vesicular porosity present in the upper part of the interval. Only the first principal component could differentiate between both lithofacies based on the Si and Ca variation over the breccia. An important aspect here was the identification of the interbedded interval between 563-577 mbsf, which was defined by principal component I.

5.2.2 - Cluster Analysis

Cluster Analysis was used in this work in a way that permitted the dataset to be divided into different groups which show a high association among the members or samples of each group while the groups themselves are distinct from each other. The analysis was performed in two different ways. Firstly, it described the relationship between the log curves (Hierarchical method) of each dataset. Secondly, it described how the samples (depth measurements of each log curve) are associated, how they form the different groups, and which of them related to lithofacies variations (Non-Hierarchical method).

In ODP Hole 807C, the hierarchical analysis produced a dendogram (Figure 3.12) which shows the correlation of the log curves. It was observed that gamma-ray, potassium, density and photoelectric effect presented the highest correlations. Another graphical representation of the dendogram is illustrated in Figure 5.1, where cross-plots of the principal log curves are shown. It is observed that the correlation shown in the dendogram match the ones observed in the cross-plots. The log curves also formed two main distinct groups which are similar to the loadings observed for the first principal component I. In other words it means that an increase, for instance, in PEF,

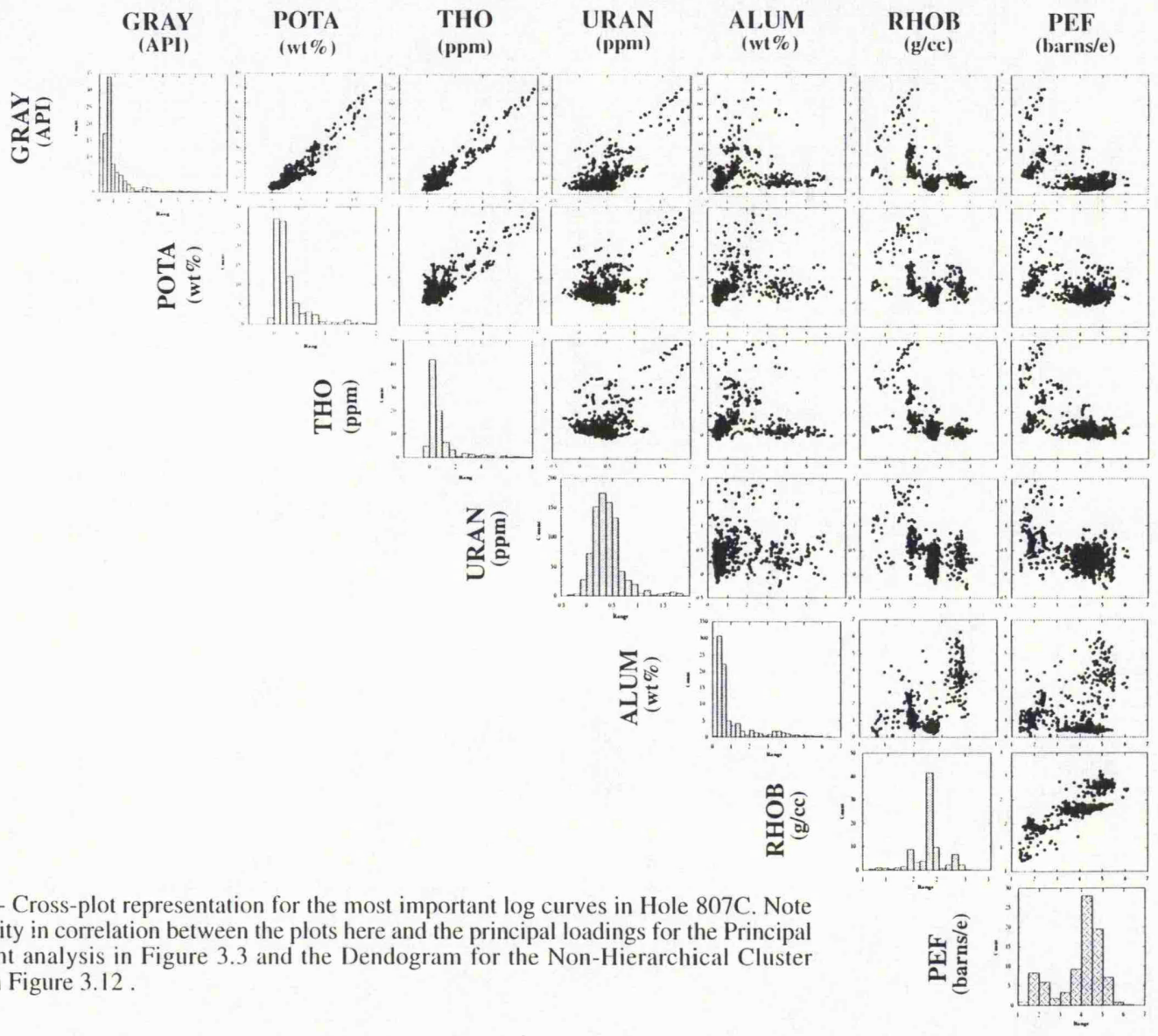


Figure 5.1 - Cross-plot representation for the most important log curves in Hole 807C. Note the similarity in correlation between the plots here and the principal loadings for the Principal Component analysis in Figure 3.3 and the Dendogram for the Non-Hierarchical Cluster analysis in Figure 3.12 .

RHOB and IDPH in the siltstone/claystone sequence is accompanied by a decrease in GRAY , POTA and URAN

The Non-Hierarchical Cluster Analysis for ODP Hole 807C was performed using firstly the original log curves and then the principal components. In the first case, the original dataset was divided into three different groups, each one representing one of the lithofacies to be characterised. The analysis also showed that RHOB, PEF and ALUM were the most important log curves to distinguish between the groups (lithofacies). As observed in Figure 3.14, the analysis using the original log curves was able to identify the major three lithofacies, however, no information about heterogeneity was obtained. When using the principal component scores as an input, there was a clear improvement in the results with not only the main lithofacies being identified but also the characterisation of heterogeneities being observed. The key features picked out in the second analysis were: a) the inclusion of an extra group (Group 4) in the analysis with the principal component scores compared to the analysis with the original log data. This new group is related to the high clay content present between 1352 - 1358 mbsf in the claystone/siltstone sequence; b) The 5 meters thick carbonate layer between 1371 - 1375 mbsf which was considered thicker based on the drill rate data. It was observed that when the analysis was performed using only the "essential" information contained in the dataset (the case of the principal component scores), improved results were obtained with the characterisation of heterogeneities not observed in the analysis using the original log curves.

In ODP Hole 878A, Hierarchical Cluster Analysis showed that the principal correlations are between the resistivity measurements (IDPH and SFLU) and between TIO2 and FEO. Figure 5.2 shows the cross-plot of the principal log curves with the correlations observed in the dendogram of the Hierarchical analysis and in the correlation matrix (Table 3.18). Again, two main groups can be observed. The Non-Hierarchical Cluster Analysis was carried out using both the original log curves and the principal component scores. Apart from two or three minor intervals, there is not an evident difference between the groups resulting from each analysis. This might be because the three principal components represent more than 85% of the total variability contained in the dataset. The important point here is that both analyses were able to show variations within the two different breccia matrices. The variations were related to the Si and Ca content in each of the lithofacies. Therefore, for the top of the interval, Groups 1 and 2 in both analyses are related to the Si content while Groups 3 and 4 are related to the Ca content at the bottom of the interval. A Hierarchical analysis using the Groups obtained in the Non-Hierarchical analysis was performed to illustrate how the groups were correlated (Figure 3.19).

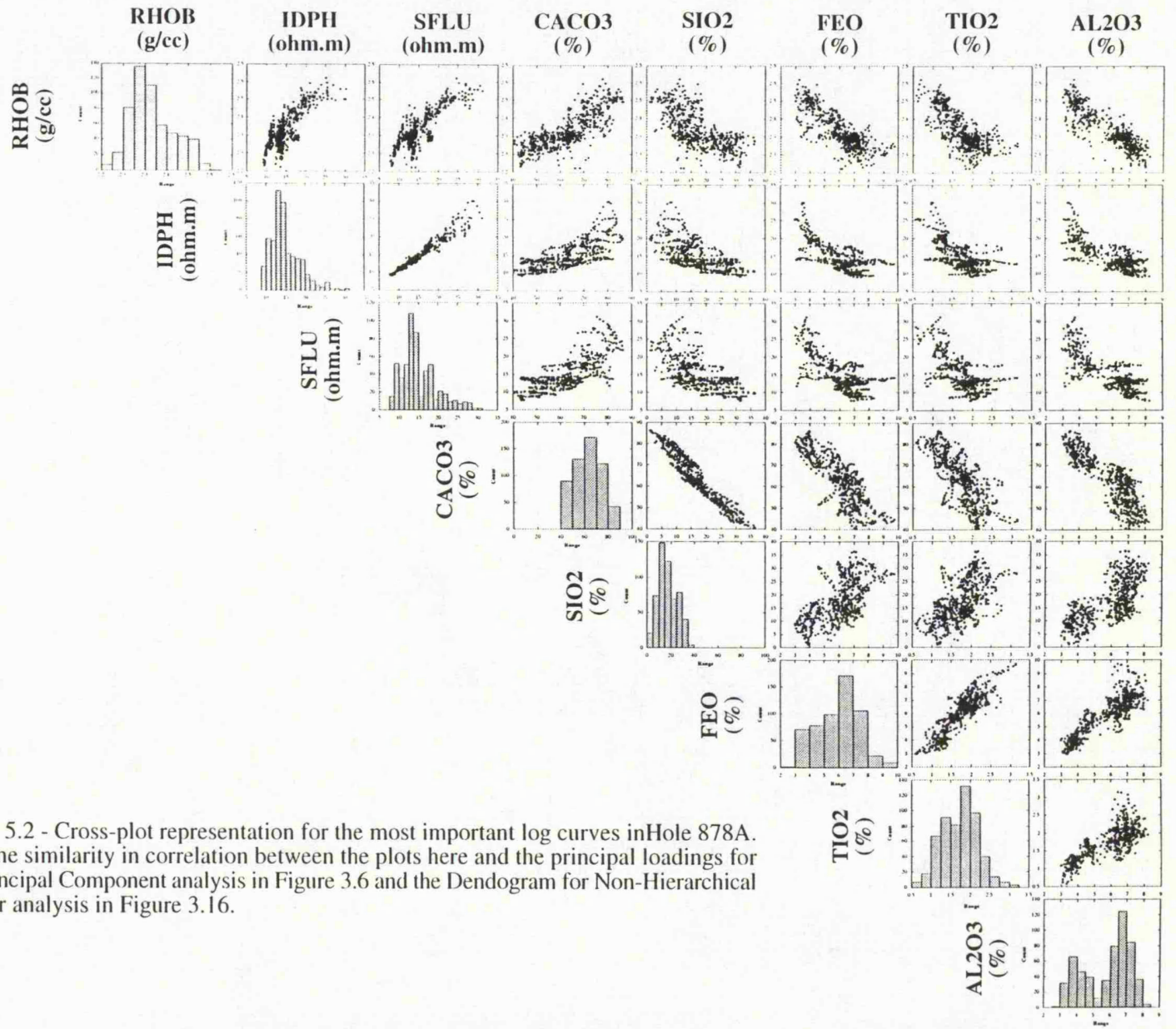


Figure 5.2 - Cross-plot representation for the most important log curves inHole 878A. Note the similarity in correlation between the plots here and the principal loadings for the Principal Component analysis in Figure 3.6 and the Dendogram for Non-Hierarchical Cluster analysis in Figure 3.16.

5.2.3 - Discriminant Analysis

In the Discriminant Analysis an *a priori* knowledge about the dataset was used to construct a discriminant function which was able to identify and distinguish lithofacies using the well log responses. This technique was used not only to produce lithofacies characterisation in the two holes examined with the previous principal component and Cluster Analysis, but also in the two oilfield holes where the "knowledge" from one hole was applied to identify lithofacies in the other hole.

The result in terms of two discriminant scores for ODP Hole 807C shows a good separation between the three lithofacies present (Figure 3.24). The discriminant function obtained from short sets of data were able to identify the lithofacies however they do not confirm the heterogeneities observed with the two previous techniques.

In ODP Hole 878A the discriminant scores respond mainly to the variations in the Si and Ca content within the breccia and again were not able to identify any heterogeneity present within the dataset. Three different distance classification methods were tested for both datasets and showed high percentages of corrected samples classified for each lithofacies. In ODP Hole 807C, only the Euclidean distance classification method shows a high percentage of misclassified samples (22.63%) due to the similar values in density for lithofacies 1 and 3 (Table 3.26). In ODP Hole 878A, all three classification models show high percentages (> 90%) of corrected samples, indicating that the discrimination function would have worked well.

An important step in the Discriminant Analysis was performed with the lithofacies discrimination of Hole B using the discriminant function obtained from Hole A. The results in terms of discriminant scores showed reasonable discrimination between the two main lithofacies including the interbedded sequence between 563 - 577 mbsf. The discriminant scores do not show any major variation within each lithofacies, suggesting that no lithological heterogeneities were present, or at least, no variation was detected by the discriminant function in Hole A. The three distance classification models show "excellent" sample classification ($\approx$ 100%). However, a comparison with the groups from the Cluster Analysis show that variations do really occur in the upper part of the interval.

5.3 - Neural Network

In a different way from the statistical techniques described above, a Neural Network was used to provide not only lithofacies discrimination but also petrophysical parameters estimation for the three datasets. While the statistical

techniques were based on mathematical models, where the tasks are explicitly represented, the Neural Network relies upon training where the machine learns from experience. Therefore, a training set which describe the main variations within the data is needed.

In the lithofacies characterisation of ODP Holes 807C and 878A, the results presented by the Backpropagation Neural Network were able to show not only a good description of the lithofacies but also formation heterogeneities which were detected in the Cluster Analysis. One of the examples is the interval between 540 - 550 mbsf in ODP Hole 878A. The identification of the sources of these heterogeneities was attempted through the application of the Neural Network exploiting the physical/chemical property dependence and vertical resolution of the log curves. These analyses were able to respond to some of the questions concerning the source of the heterogeneities, but in general, failed to produce reliable information about the sources. In general, it was demonstrated that a reduction of the number of the log curves tends to produce a result of low quality. When applied to Oilfield Holes A and B, there was a very good match between the Neural Network responses and the lithofacies characterisation. The important point here is the characterisation of formation heterogeneities for the upper part of the interval in Hole B which were not described in the training sets of Hole A. The results confirm what was observed in the Group classification (Cluster Analysis) where Groups 1 and 2 were identified for the upper part of Hole B.

The second objective of the application of the Backpropagation Neural Network is in predicting petrophysical parameters from well log data in uncored intervals, or in case of Oilfield Holes A and B, in uncored holes. Cored sections were used as a basis to provide the Neural Network with enough training. The results for ODP Holes 807C and 878A show a reasonable match between the predicted Neural Network physical properties and the core measurements. The Neural Network predicted petrophysical logs also show a good result, following the main variation observed in the core data. When applied to an offset hole, the Neural Network predicted petrophysical parameter also seem to work very well, although some changes in the petrophysical characteristics not described in the training sets lead to a mismatch between the Neural Network results and the log derived petrophysical parameters. These mismatches are the result of the heterogeneities observed during the lithofacies characterisation process.

5.4 - Conclusions

Investigation of the performance of statistical techniques and Neural Network in well log data was the central theme of this research work and, it has been

demonstrated that for the data used here, these techniques are clearly of much use as a lithofacies discriminator and heterogeneities indicator, providing information which could not have been obtained by simply having a visual look at the log curves. This is in a good agreement with much previous work using these techniques.

It was realised at an early stage in this work that the task of producing lithofacies discrimination and heterogeneities characterisation using a large set of well log data could most easily be accomplished by the use of a variety of multivariate statistical techniques and also other non-linear new techniques such as the Neural Networks. Among other applications, prediction of petrophysical parameters was also included in the use of this technique.

Principal Component Analysis helped to get a better understanding about the structure of the data set and the importance of each original log curve. In addition, the transform of the original log curves into principal components, reducing the amount of data to be analysed, was a major step for future processing.

A powerful Cluster Analysis has been described in both the Hierarchical and Non-Hierarchical models. The advantage over some of the other multivariate statistical techniques currently being used is that there is no requirement for any *a priori* information, it is simple to use and it can handle large amounts of data (e.g. logging data). The use of the K-means algorithm in the Non-Hierarchical analysis makes this technique a powerful identification tool, capable of identifying lithological groups. The use of this technique together with the principal component scores can produce improved results in the interpretation as shown in ODP Hole 807C. Due to the huge amount of memory required, it is not possible to perform a Hierarchical analysis (R mode) using the depth measurements.

Discriminant Analysis was also considered to work well in terms of lithofacies discrimination, but showed poor results in terms of heterogeneities characterisation. Despite the use of *a priori* knowledge, when facies changes present within the formations are not described in the training sets the discriminant functions are unlikely to resolve these variations. The use of different distance classification models shows that the fundamental problem in classification is related to choosing an appropriate distance measure.

Neural Network proved to work well both in lithofacies discrimination and in prediction of petrophysical parameters. Variations in lithofacies, considered here as formation heterogeneities, were well characterised both in single holes and inter-hole correlation. The results for the three holes used in this work are shown in Table 5.1. In two of the cases, Neural Network provided superior results to that based on statistical Discriminant Analysis. In ODP Hole 807C, the Neural Network performed much better than Discriminant Analysis, however it took a long training stage to achieve this result. In ODP Hole 878A, again Neural Network produced a better result. The low

difference in the number of misclassified samples between the two techniques are due to the high complexity of the breccia. Finally, in the inter-hole correlation, Discriminant Analysis performed better than the Neural Network. In this case, the variations observed in the Neural Network results for the upper part of the interval in Hole B was considered as a heterogeneity present in that lithofacies. Note that the response of the Neural Network (Figure 4.28) shows that UNIT I is more strongely represented in that interval.

From the case examples, it has been proved that Neural Network is a powerful technique in pattern recognition which suffers from some practical problems. Some of the practical difficulties are that before training a pre-specified network topology is required. It is known that too few layers and neurons will not learn satisfactorily, and too many of these elements will overtrain the network (Dayhoff, 1990). Therefore, a careful monitoring of training is required. Another problem is related to the incrementing of knowledge. Sometimes it is very difficult, or even impossible, to adapt the Neural Network to an additional input and it is necessary to re-train the net.

Based on the results shown it was considered that the Backpropagation Neural Network performed in general better than some statistical analysis (e.g. Discriminant Analysis) in lithofacies characterisation. This fact is probably related to the non-linearity and non-parametric properties of the method.

Table 5.1

	ODP Hole 807C	**ODP Hole 878A**	**Holes A & B**
Discriminant Analysis (Euclidean Distance Model)	22.63	6.27	0
Backpropagation Neural Network	2.87	4.84	3.23

Table 5.1 - Lithofacies classification results (measured in % of misclassified samples).

The Backpropagation Neural Network has been shown to give good results for petrophysical parameters estimation. It can also be considered a powerful tool for petrophysical evaluation in inter-hole interpretation as long as lithofacies and petrophysical characteristic changes can be trained. The results presented here showed

that petrophysical logs can be obtained for uncored intervals with a reasonable degree of success. An important factor for its success was the use of a "windowing" technique, which allowed a more realistic training set for the Neural Network. The results for all examples have shown good matches between actual core measurements and the Neural Network predicted physical properties. The physical property logs generated by the Neural Network match the main lithofacies variation present within the intervals. Again, ODP Hole 878A shows some degree of variation in the generated physical property logs due mostly to the complex changes in lithofacies. The Neural Network presented here also holds great promise in helping inter-hole prediction of petrophysical parameters. The results obtained with the examples show good correlation with other log derived values even in the presence of lateral changes of facies which generally request some sort of skill and trial and error procedures in producing good results. However, it is recognised that a complete comparison which includes other techniques of predicting petrophysical parameters should be performed in order to obtain a better idea on how efficiently this technique works.

5.5 - Recommendations

It is considered that future research in this area should include a more careful study of the Discriminant Analysis, including a detailed analysis of the group discriminant classification and its different models, which can help better understand how this technique works and also improve its results. The fact that the technique uses *a priori* knowledge about the data helps the integration between log and geology.

Finally, another aspect in which the research should be focused from now is in the use of different structures for the Backpropagation Neural Network, which could produce better results and consume less computing time. Most of the work so far has shown that other Neural Networks which do not use training as a major step in the processing, can produce a quicker response (Anderson et al., 1990). The quality of the results, however, is uncertain. Therefore, the use of a training technique which will allow a refinement in the process of trial and error in selecting its better structure is the main direction in the future works.

REFERENCES

Aleksander, I. and Morton, H., 1989, An introduction to Neural Computing. London: Chapman & Hall.

Andenburg, M.R., 1973, Cluster analysis for applications. Academic Press, New York.

Anderson, J.A. and Rosenfeld, E., 1988, Neurocomputing, foundation of research. Cambridge, MA. MIT Press.

Anderson, J.A., Pellionisz, A. and Rosenfeld, E., 1990, Neurocomputing 2: Directions for Research. Cambridge, MA: MIT Press.

Andrade, A.J.N., Luthi, S. and Guerra, C.E., 1995, Determinação de interfaces em perfis através da rede neuronal de Hopfield. Anais do IV Congresso Internacional da Sociedade Brasileira de Geofísica. SBGf, Volume II, pp 771-774.

Anguita, D., 1993, MBP - Matrix back propagation, v. 1.1 - An efficient implementation of the BP algorithm. DIBE - University of Genova, 26 pp.

Anguita, D., Parodi, G. and Zunino, R., 1993, Speed improvement of the back-propagation on the current-generation workstations. WCNN '93, July 11-15 1993, Portland, USA, pp. 165-168.

Allaud, L. and Martin, M., 1977, Schlumberger, the history of a technique. John Wiley and Sons, New York, 333 pp.

Andrews, J.E., Packham, G.H., et al., 1975, *Init. Repts. DSDP,* 30: Washington (U.S. Govt. Printing Office).

Baldwin, J.L., Bateman, R.M. and Wheatley, C.L., 1990, Application of Neural Network to the problem of mineral identification from well logs. The Log Analyst, 3, pp. 279-293.

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

Bateman, R.M., 1985, Open-hole log analysis and formation evaluation. D. Riedel Publishing Co., Boston, 647 pp.

Battiti, R., 1992, First and Second-order methods for learning: between steepest descent and Newton's method. Neural Computation n.4, pp. 141-146.

Baum, E.B. and Haussler, D., 1989, What size nets give valid generation? Neural Computation, 1(1):151-160.

Berger, W.H., Kroenke, L.W., Mayer, L.A. et al., 1993, *Proc. ODP, Sci. Results,* 130: College Station, TX.

Bristow, J.F., 1993, Physical and chemical characteristics of rocks from downhole measurements. PhD thesis, University of Nottingham, 125 pp.

Bucheb, J.A. and Evans, H.B., 1994, Some applications of methods used in eletrofacies identification. The Log Analyst, v. 35, n. 1, pp. 14-26.

Chapman, S., Colson, J.L., Flaum, C., Hertzog, R.C., Pirie, G., Scott, H., Everett, R., Herron, M.M., Schweitzer, J., Lavigne, J., Quirein, J. and Wendlant, R., 1987, The emergence of geochemical well log. The Technical Review, 35(2), pp. 27-35.

Chayes, F., 1964, A petrographic distinction between Cenozoic volcanics in and around ocean. J. Geophysical Research, vol. 69, pp. 1573 - 1588.

Crain, E.R., 1985, A primer on artificial intelligence and expert systems in the petroleum industry. Canadian Well Logging Society Journal, v. 14, pp. 17.

Caudhill, M., 1991, Neural network training tips and techniques. AI Expert, v. 6, n. 1, pp. 56-61.

Davis, J.C., 1986, Statistics and data analysis in geology. J. Wiley & Sons, New York, 646 pp.

Dayhoff, J.E., 1990, Neural Network architectures: an introduction Van Nortrand Reinhold, New York.

Delfiner, P.C., Peyret, O. and Serra, O., 1987, Automated determination of lithology from well logs. SPE Formation Evaluation, 2(3): 303-310.

Doveton, J.H., 1986, Log analysis of subsurface geology. John Wiley & Sons, New York, 273 pp.

Doveton, J.H., 1994, Geologic log analysis using computing methods. AAPG Computer Applications in Geology, n. 2, 169 pp.

Ekstron, M.P., Daham, C.A., Chen, M-Y., Lloyd, P.M. and Rossi, D.J., 1986, Formation imaging with microelectrical scanning arrays. *Trans.* of SPWLA 27th Ann. Logging Symposium, paper BB.

Ellis, D.V., 1987, Well logging for earth scientists. Elsevier, New York, 532 pp.

Fausett, L., 1994, Fundamentals of neural networks: architectures, algorithms and applications. Prentice Hall, New York, 461 pp.

Gillis, K., Mével, C., Allan, J. et al., 1993, *Proc. ODP, Init. Repts.,* 147/148: College Station, TX (Ocean Drilling Program).

Gnanadesikan, R., 1977, Methods for statistical data analysis of multivariate observations. Wiley and Sons, New York, 311 pp.

Gonçalves, C.A., Harvey, P.K. and Lovell, M.A., 1995a, Application of a mutilayer neural network and statistical techniques in formation characterisation. SPWLA 36th Annual Logging Symposium Transactions, Paris, France, paper FF.

Gonçalves, C.A., Ewert, L and ODP Leg 159 Scientific Party, 1995b, Transform Margin Developments: Preliminary logging results from ODP Leg 159 on the Cote-D'Ivoire-Ghana Margin. I Latin American Geophysical Congress Transactions, Volume II, pp 487-489.

Hall, J. and Scandella, L, 1995, Estimation of critical formation evaluation parameters using techniques of neurocomputing. SPWLA 36th Annual Logging Symposium, Paris, paper PPP.

Harvey, P. K., & Lovell, M. A. (1989). Basaltic litho-stratigraphy in Ocean Drilling Program Hole 504B. Nuclear Geophysics, 3(2), 87-96.

Harvey, P. K., Lovell, M. A., Pezard, P. A., Taylor, R. N., & Leg 126 Shipboard Scientific Party, 1992, Geochemical characterisation of volcanoclastic sediments in the Izu-Bonin Arc from downhole measurements. Transactions 5th. Circum-Pacific Energy and Mineral Resources Conference (1990).

Haykin, S., 1994, Neural Networks: a comprehensive foundation. MacMillan Press. New York, 696 pp.

Hertzog, R., Colson, L., Seeman, B., O'Brien, M., Scott, H., McKeon, D., Wraight, P., Grau, J., Schweitzer, J. and Herron, M., 1989, Geochemical logging with spectrometry tools. SPE Formation Evaluation Symposium, 4 (2), pp. 153-162.

Hetch-Nielsen, R., 1990, Neurocomputing. Reading, MA. Addison-Wesley

Hurst, A. and Milodowski, T., 1994, Characterisation of clays in sandstone: Thorium content and spectral log data. *Trans.* of 16th European SPWLA Formation Evaluation Symposium, Aberdeen, Scotland, paper S.

Jenner, R. and Baldwin, J.R., 1994, Application of an artificial neural network an improved permeability profile. SPWLA 16th European Formation Evaluation Symposium, Aberdeen, Scotland, paper M.

Johnson, R.C., and Brown, C., 1988, Cognizers: Neural networks and machines that think. John Wiley & Sons, New York.

Johnson, R.P. and Wichern, D.W., 1982, Applied multivariate statistical analysis. Prentice-Hall, New Jersey, 594 pp.

Kroenke, L.W., 1972, Geology of the Ontong Java plateau. Hawaii Inst. Geophys., paper HIG72-5.

Kroenke, L.W., Berger, W.H., Janecek, T.R., et al., 1991, *Proc. ODP, Init. Repts.*, 130: College Station, TX (Ocean Drilling Program).

Le Maitre, R.W., 1982, Numerical Petrology: statistical interpretation of numerical data. Elsevier, Amsterdan, 281 pp.

Lock, G.A. and Hoyer, W.A., 1971, Natural gamma-ray spectral logging. The Log Analyst, n. 12, pp. 3-9.

Lofts, J.C., 1993, Integrated geochemical-geophysical studies of sedimentary reservoir rocks. PhD thesis, University of Leicester, 122 pp.

Lovell, M.A. and Anderson, R., 1989, When downhole logging turns to geochemistry. Geology Today, Sept-Oct, pp. 164-166.

Lovell, M.A., Harvey, P.K. and Lofts, J., 1993, Geochemical Logging. III International Congress of the Brazilian Geophysical Society, Vol II, pp. 964-968.

Luthi, S.M., and Sohaité, P., 1990, Fracture apertures from electrical borehole scans. Geophysics, v. 55, n. 7, pp. 821-833.

Mac Gregor, R.J., 1987, Neural and brain modelling. Academic Press, San Diego, USA.

Marriot, F.H.C., 1974, The interpretation of multiple observations. Academic Press (London), 117 pp.

Minsky, M.L. and Papert, S.A., 1969, Perceptrons, Cambridge, MA: MIT Press.

Ocean Drilling Program (ODP), 1991, ODP wireline logging manual, volumes 1-3. ODP (LDGO, Palidases, New York).

Pelling, R., 1992, Integrated geophysical-geochemical studies of the oceanic crust. PhD thesis, University of Nottingham, 147 pp.

Pelling, R., Harvey, P. K., & Lovell, M. A. (1989). Statistical analysis of Geochemical Logging Data from ODP Boreholes in Oceanic Basement. Proceedings of the Third MLGS International Symposium on Borehole Geophysics for Minerals, Geotechnical and Groundwater Applications, Las Vegas, Nevada., 1, 349-383.

Pezard, P.A. and Luthi, S.M., 1988, Borehole electrical images in the basement of the Cajon Pass scientific drill hole, California; fracture identification and tectonic implication. Geophysical Research Letters, n. 15, pp. 1017-1020.

Pezard, P.A., Lovell, M.A. and ODP Leg 126 Shipboard Scientific Party, 1990, Downhole images: electrical scanning reveals the nature of subsurface oceanic crust. EOS, n. 71, pp. 709.

Pezard, P.A., Hiscott, R.N., Lovell, M.A., Collella, A. and Malinverno, A., 1990b, Evolution of the Izu-Bonin Intraoceanic Forearc basin, Western Pacific, from cores to FMS images. *in* Geological Applications of Wireline Logs II, edited by A. Hurst, C.M. Griffith and P.W. Worthington. Geological Society Special Publication no. 65, London, pp.43-70.

Premoli Silva, I., Haggerty, J., Rack, F., et al., 1993, *Proc. ODP, Init. Repts.,* 144: College Station, (Ocean Drilling Program).

Prensky, S.E., 1987, Geological applications of well logs: an introductory bibliography and survey of the well logging literature (through Sept. 1986) The Log Analyst 28(1), pp. 71-107.

Prensky, S.E., 1990, Bibliography of the well logging applications, annual update, October 1, 1989-September 1, 1990. The Log Analyst, 31(6), pp. 395-424.

Prensky, S.E., 1991, Bibliography of the well logging applications, annual update, September 1, 1990-October 1, 1991. The Log Analyst, 32(6), pp. 693-737.

Prensky, S.E., 1992, Bibliography of the well logging applications, annual update, September 30, 1991-October 1, 1992. The Log Analyst, 33(6), pp. 520-558.

Prensky, S.E., 1993, Bibliography of the well logging applications, annual update, October 1, 1992- September 30, 1993. The Log Analyst, 34(6), pp. 37-87.

Rea, D.K., Bason, I.A., Janecek, T.R., Palmer-Julson, A. et al., 1993, *Proc. ODP, Init. Repts.,* 145: College Station, TX (Ocean Drilling Program).

Rumelhart, D.E., McClelland, J. and the PDP research group, 1986, Parallel distributed processing. Volume 1, The MIT Press.

Schlumberger, 1982, Natural gamma-ray spectrometry: essentials of NGS interpretation. Schlumberger, Houston, Texas, 70 pp.

Schlumberger, 1989a, Log interpretation principles and applications. Houston, TX.: Schlumberger Educational Services, 150 pp.

Schlumberger, 1989b, Formation MicroScanner image interpretation. Houston, TX: Schlumberger Educational Service, 117 pp.

Schweitzer, J.S., Manente, R.A. and Hertzog, R.C., 1988, Elemental concentration from gamma-ray spectroscopy logs. Nuclear Geophysics, n. 2, pp. 175-181.

Scott, H.D. and Smith, M.P., 1973, The Aluminium Activation Clay Log. The Log Analyst, n. 14, pp. 3-12.

Serra, O. and Abbott, H.T., 1982, The contribution of logging data to sedimentology and stratigraphy. Soc. Pet. Eng. Journal, v. 22, n.1, pp. 117-131.

Sheppard, A., 1986, The geochemistry and evolution of Lizard Complex, Cornwall. PhD thesis, University of Nottingham, 251 pp.

Simpson, G.G., 1961, Principles of animal taxonomy. Columbia University Press, New York.

Theys, P., 1990, Log data acquisition and quality control. Editions Technip, Paris, 330 pp.

Tryon, R.C. and Bayley, D.E., 1970, Cluster Analysis. MacGraw-Hill, New York.

Vogl, T.P., Mangis, J.K., Rigler, A.K., Zink, Z.T. and Alkon, D.L., 1991, Accelerating the convergence of the back-propagation method. Biological Cybernetics, n. 59, pp. 257-263.

Winterer, E.L., Riedel, W.R., et al., 1971, *Init. Repts. DSDP,* 7: Washington (U.S. Govt. Printing Office).

Winterer, E.L., Duncan, R.A., McNutt, M.K., Natland, J.H., Premoli Silva, I., Sager, W.W., Sliter, W.V., van Waasbergen, R.J. and Wolfe, C.J., 1992, Cretaceous Guyots in the northwest Pacific: an overview of their lithology and geophysics. In Pringle, M.S., Sager, W.W., Sliter, W.V. and Stein, S. (editor), The Mesozoic Pacific. American Geophysical Union, Geophysical Monography Series.

Wong, P. M., Gedeon, T.D. and Taggart, I.J., 1995, The use of fuzzy ARTMAP for lithofacies classification: A comparison study. SPWLA 36th Annual Logging Symposium, Paris, paper U.

Young, T.Y. and Calvert, T.W., 1974, Classification, estimation and pattern recognition. Elsevier, Amsterdan.

APPENDIX A

ODP LEG 159 - CÔTE D'IVOIRE-GHANA TRANSFORM MARGIN

From 4th January to 1st March 1995, ODP Leg 159 drilled four sites on the west African Margin, along the Romanche fracture zone. The four sites drilled are located on continental crust on a ridge that runs close to the ocean-continent boundary. Drilling the Côte D'Ivoire-Ghana transform continental margin had both tectonic and paleoceanographic objectives. The main tectonic objective was to better understand the evolution of transform continental margins. Transform faults represent the third major category of plate boundaries, but are less well understood than either convergent and divergent examples and have never been investigated by scientific drilling. Among the paleoceanographic objectives was to document changes in deep and intermediate waters passing through the eastern equatorial Atlantic and the changing geometry of the Eastern Atlantic basins since continental break-up in the Early Cretaceous times.

The Lamont-Doherty Borehole Research Group, in conjunction with Schlumberger Well Logging Services, provides the geophysical well logging aboard the master drillship *Joides Resolution*. Designed for use in hydrocarbon exploration, the logging tools used have been adapted to meet ODP requirements, primarily the reduction of tool diameter to allow insertion into the 3.8 inches (9.65 cm) drill-string bore. Downhole logs are used to characterise the geophysical, geochemical and structural properties of the sequences drilled. Log data offer advantages over core-based analysis in that they are rapidly collected and represent continuous, in situ measurements of the formations. Four different Schlumberger tool combinations were used in Leg 159: a) a Seismic Stratigraphic tool string, b) a Lithoporosity tool string, c) the Formation MicroScanner (FMS) and d) the Geochemical tool string (see Chapter 2).

The role of the Logging Scientist is to promote acquisition, processing and interpretation of the well log data acquired. Soon after the data are collected, they have to be depth shifted and correlated between the different runs using the Natural Gamma-Ray Tool (NGT). This was a particularly difficult task during Leg 159 where rapidly varying lithology in zones with poor core recovery sometimes impaired the multilog analyses. The Sonic log was also corrected for cycle skipping and was an important tool in the refinement of the seismic horizons. A pre-process was performed onboard with the FMS images. Although borehole conditions were far from the ideal for this tool, FMS images were very important because of the structural objectives of the Leg. After processing, the interpretation of the well log data was performed in a multidisciplinary way which involved other members of the Scientific Party. The well log data on this Leg were of particular importance due to the low core recovery obtained in most of the holes drilled.

Even though the participation on ODP Leg 159 had no connection with the work presented in this Thesis, it was considered of the great importance in the training of well log analysis and formation evaluation. The experience acquired in the processing of different logging data gives an extra input to interpretation, allowing the Logging Scientist to better understand the mechanisms which generate the log responses.

Transform Margin Development: Preliminary logging results from ODP Leg 159 on the Côte D'Ivorie-Ghana margin

Gonçalves, C.A.[1,2], *Ewert, E.*[1], *Basile, C.*[3] *and ODP Leg 159 Scientific Party*[†]

[1] Borehole Research, Department of Geology, University of Leicester, U.K.
[2] On leave from NPGP/Federal University of Pará - Brazil.
[3] Université J. Fourier, France

SUMMARY

ODP Leg 159 studied the Côte d'Ivoire-Ghana transform margin in the eastern equatorial Atlantic during January and February, 1995. Thirteen holes were drilled at four sites, recovering sediments from Late Albian to Pleistocene age. The margin has been created by major transform motion between the African and South America plates and is one of the best known examples of a former transform boundary between continental and oceanic crust (Mascle and Blarez, 1987).

Logging data were collected at three of the sites, and despite bad hole conditions, logging runs were obtained in four holes. This paper shows the preliminary logging results which cover the full sedimentary sequence cored. Together with the standard tools used by the oil industry, the Geochemical Logging Tool (GLT™) and the Formation MicroScanner (FMS™) tool were also used.

INTRODUCTION

From 4th January to 1st March 1995, the master drillship of the Ocean Drilling Program (ODP), Joides Resolution, drilled 13 holes on the west African margin, offshore Ghana (Figure 1). The principal objectives of the cruise were the development of the transform margin, its nature, structure and deformation history, and the relationship between sedimentation and deformation. A detailed study of the Cenozoic paleoceanographic and climatic history of the central eastern equatorial Atlantic were also among the main objectives.

The four sites drilled, 959, 960, 961 and 962, are located on continental crust on a ridge that runs close to the continent/ocean transition (Figure 2) (Mascle, Lohmann, Clift et al., in press). Three of these sites were logged with the Quad Combo, the FMS tool and the Geochemical tool string. The log curves were used to match the sedimentary cover with the FMS and GLT logs being used for structural and geochemical detail respectively. In addition to these tools, the Lamont-Doherty Earth Observatory temperature tool (TLT) was run with the Quad Combo tool string.

LOG QUALITY AND DATA PROCESSING

Logging runs were carried out in four holes in sites 959, 960 and 962, in water depths of 2102, 2050 and 4667 meters. Site 961 was not logged due to hole stability problems and the hole was abandoned after 390 meters. A complete combination of the three tool strings was run in Hole 959D between 395 - 1081 meters below sea floor. Log quality is generally good throughout the interval despite intermittent borehole washout, especially near the top, where caliper measurements range from 14.5" to 17".

Holes 960A and 960C were logged at Site 960. Poor hole conditions in both holes prevented deployment of all three tool strings. Hole 960A was logged only with the Quad Combo tool string due to extreme hole diameter (greater then 17" for most of the interval). In Hole 960C, although the hole conditions were no better than in 960A, the FMS tool string was run due to the importance of collecting data on the structural development of the transform margin. In this hole, only a reduced combination of the Quad Combo tool string was run, consisting of the Natural Gamma-ray, the Phasor induction and the Array Sonic tools. Site 962 was also affected by poor hole conditions and once more only a reduced combination of the Quad Combo tool string and the FMS were run. Data from the Quad Combo and FMS tool strings underwent preliminary processing onboard. Data processing for the Quad tool string consisted of removing "cycle skipping" effects from the sonic data and

Transform Margin Development: Preliminary logging results from ODP Leg 159 on the Côte D'Ivorie-Ghana margin

Gonçalves, C.A.[1,2], *Ewert, E.*[1], *Basile, C.*[3] *and ODP Leg 159 Scientific Party*[†]

[1] Borehole Research, Department of Geology, University of Leicester, U.K.
[2] On leave from NPGP/Federal University of Pará - Brazil.
[3] Université J. Fourier, France

depth-shifting all the log curves to match the core data. Preliminary shipboard processing of the FMS data was performed using Schlumberger's Geoframe software, but complete processing will be done post cruise. Geochemical data needs to undergo extensive onshore processing to convert the relative concentrations of Si, Ca, Fe, S, H and Cl and wet weight percentages of K, Th, U and Al to dry weight percentages and to determine the Gd and Ti (Mascle, Lohmann, Clift et al., in press).

LOGGING RESULTS

The natural gamma-ray, density, porosity, velocity and resistivity logs provide confirmation of most lithologic units derived from core description in Hole 959D (Figure 3). A sequence from Subunit IIA (nannofossil ooze chalk) to Subunit IVA (calcareous sandstone) was covered by the logs (Mascle, Lohmann, Clift et al., in press).

At Site 960, the natural gamma-ray, resistivity and caliper logs (Hole 960A only) and the density and porosity logs (Holes 960A and 960C), show trends which correlate well with most of the lithologic unit boundaries derived from core description (Figure 4).

At Site 962, the logged interval does not cross any lithologic boundaries observed in the core description and fall entirely within Unit III (Figure 5), which includes claystones, siltstones, sandstone and also limestones.

The log intervals for the three sites correlate well with each other and the lithologic boundaries from the core description can be identified in all holes. Variations in the geochemical log observed in the non-processed data also enable future study of the variations of clay types observed in the cores.

FMS DATA

In situ structural measurements have been performed using the Formation MicroScanner tool between 547 and 936 mbsf in Hole 959D and between 173 and 354 mbsf in Hole 960C. The logged intervals cover a porcelanite unit (IIC) and the upper part of a black claystones (III) in Hole 959D, and a limestone unit (IVB) in Hole 960C (Mascle, Lohmann, Clift et al., in press)

In Hole 959D, bedding is dipping NW, with an increase from 5° to 14° with depth (Figure 6), as anticipated from the pre-cruise seismic lines (Basile et al., 1993). More surprisingly, the bedding is dipping NE in Hole 960C, and there is no evidence of increasing dips with depth (Figure 7).

At both sites, dips and azimuths of the bedding exhibit important variations around their average values at decimeter to meter scales. These variations may have been induced by fan-shaped deposits, and an associated rotation axis can be retrieved from successive bedding measurements.

At both sites, the rotation axes are mainly trending WNW to NNE (N300 to N30). In Hole 960C, the WNW rotation axes are mostly due to cross-bedding dipping towards NNE. In Hole 959D, the fan shape cannot be explained by sedimentary features, but is probably related to sliding of lithified sediments along listric faults.

In both holes, the scattering of the rotation axes is believed to be related to time variations with the strike of the slope, and may reflect the interferences between uplift of the Marginal Ridge and subsidence of the Deep Ivorian Basin.

CONCLUSION

Logging data collected during ODP Leg 159, although adversely affected by a combination of bridges and borehole wash out.

Transform Margin Development: Preliminary logging results from ODP Leg 159 on the Côte D'Ivorie-Ghana margin

Gonçalves, C.A.[1,2], *Ewert, E.*[1], *Basile, C.*[3] *and ODP Leg 159 Scientific Party*[†]

[1] Borehole Research, Department of Geology, University of Leicester, U.K.
[2] On leave from NPGP/Federal University of Pará - Brazil.
[3] Université J. Fourier, France

are nonetheless informative and encourage further analysis of the fully processed data. Onboard interpretation allowed good correlation with the main lithologic sequences described by the core data. Pre-processed FMS measurements provided important clues about sedimentary structures and the structural history of the continental margin.

REFERENCES

Basile, C., Mascle, J., Popoff, M., Bouillin, J.P. and Mascle G., 1993, The Côte d'Ivoire-Ghana transform margin: a marginal ridge structure deduced from seismic data. Tectonophysics, 222:1-19.

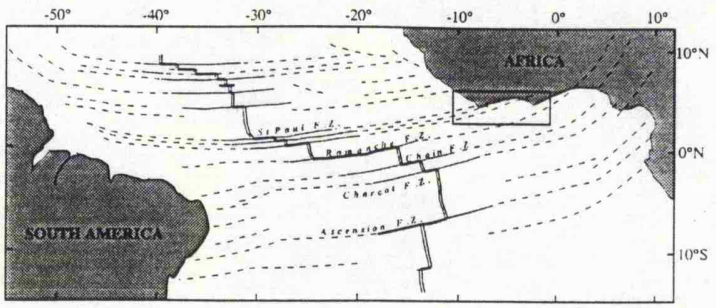
Mascle, J. and Blarez, E., 1987, Evidence for transform margin evolution from the Côte-d'Ivoire-Ghana transform margin. Nature, 326:378-381.

Mascle, J., Lohmann, P., Clift, P. and ODP Leg 159 Scientific Party, ODP Volume 159, part A, in press.

[†] ODP Leg 159 Scientific Party:—

Jean Mascle, GEMCO, Villefranche-sur-Mer, France, Co-chief
G.P. Lohmann, Woods Hole Oceanographic Institution, USA, Co-chief
Peter Clift, Texas A&M University, ODP Staff Scientist
Thomas Akamaluk, Ghana Geological Survey
Simon Allerton, Oxford University, UK
Maria Ask, Royal Institute of Technology, Stockholm, Sweden
Enriqueta C. Barrera, University of Michigan, USA
Eric Barton, Lousville, Colorado
Christophe Basile, Université J. Fourier, France
Jean-Pierre Bellier, Université Pierre et Marie de Curie, France
Jean Benkhelil, Université de Perpignan, France
Emmanuel Brantuoh, Ghana Geological Survey
Rosemary Edwards, Institute of Oceanographic Sciences, UK
Eileen Ewert, University of Leicester, UK
Carlos Goncalves, University of Leicester, UK
Aleksandra G. Janik, University of Miami, USA
Mary Anne Holmes, University of Nebraska-Lincoln, USA
Ken-Ichiro Hisada, University of Tsukuba, Japan
Kyger C. Lohmann, University of Michigan, USA
Sumito Morita, Ocean Research Institute, Japan
Carlos A. Mortera-Gutierrez, Texas A&M University, USA
Richard D. Norris, Woods Hole Oceanographic Institution, USA
Francisca E. Oboh, School of Mines and Metallurgy, Rolla, USA
Thomas Pletsch, Universität Tübingen, Germany
Elizabeth a. Picket, Edinburgh University, UK
Greg Ravizza, Woods Hole Oceanographic Institution, USA
Samir Shafik, Australia Geological Survey Organization, Australia
Im Chul Shin, University of Nebraska-Lincoln, USA
Kari O. Strand, University of Oulu, Finland
Thomas Wagner, University of Bremen, Germany
David Watkins, University of Nebraska-Lincoln, USA

™ Trade mark of Schlumberger



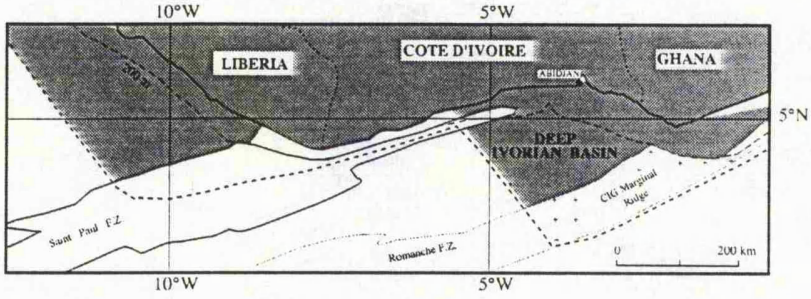


Figure 1 - Location and area surveyed by ODP Leg 159.

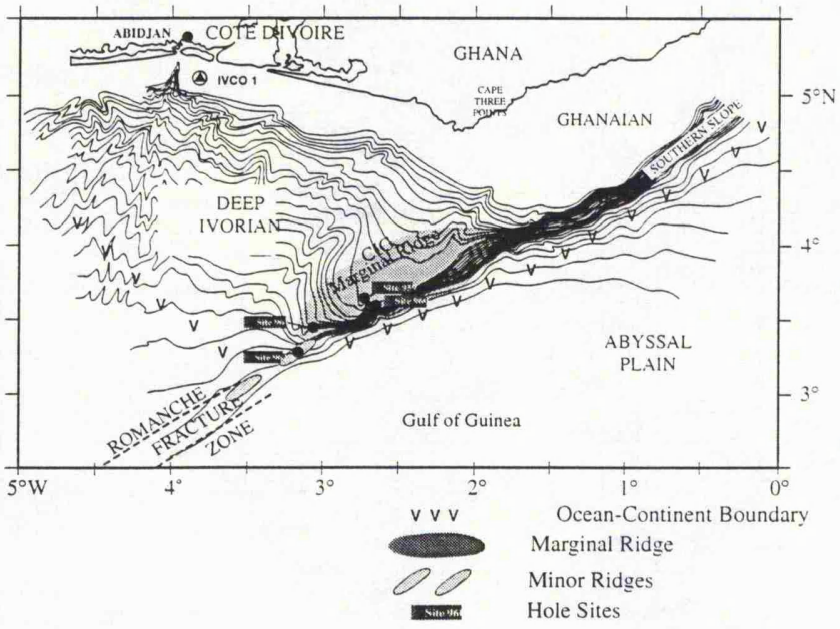


Figure 2 - Location of the four sites drilled during ODP Leg 159.

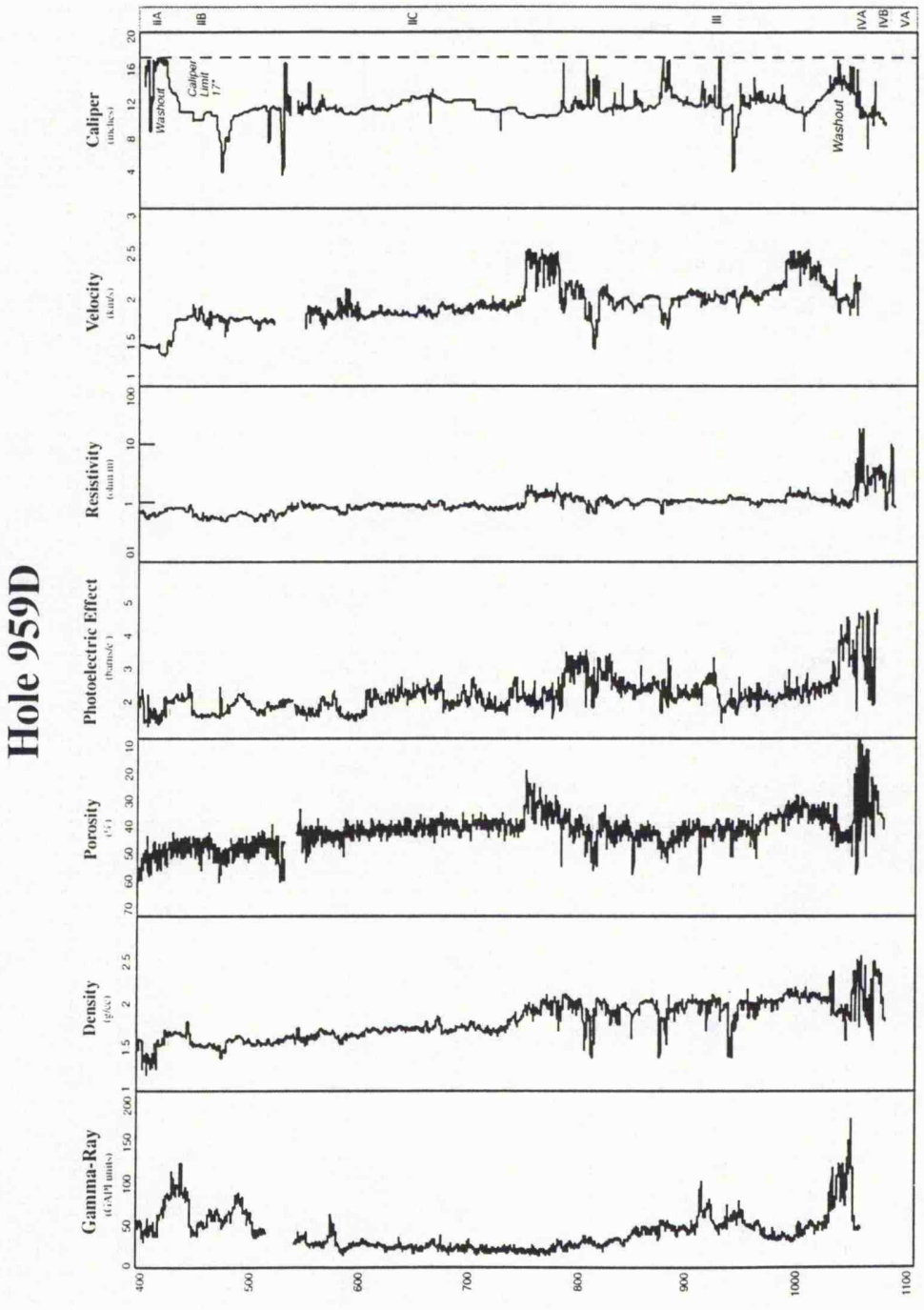


Figure 3 -Well log curves for Hole 959D.

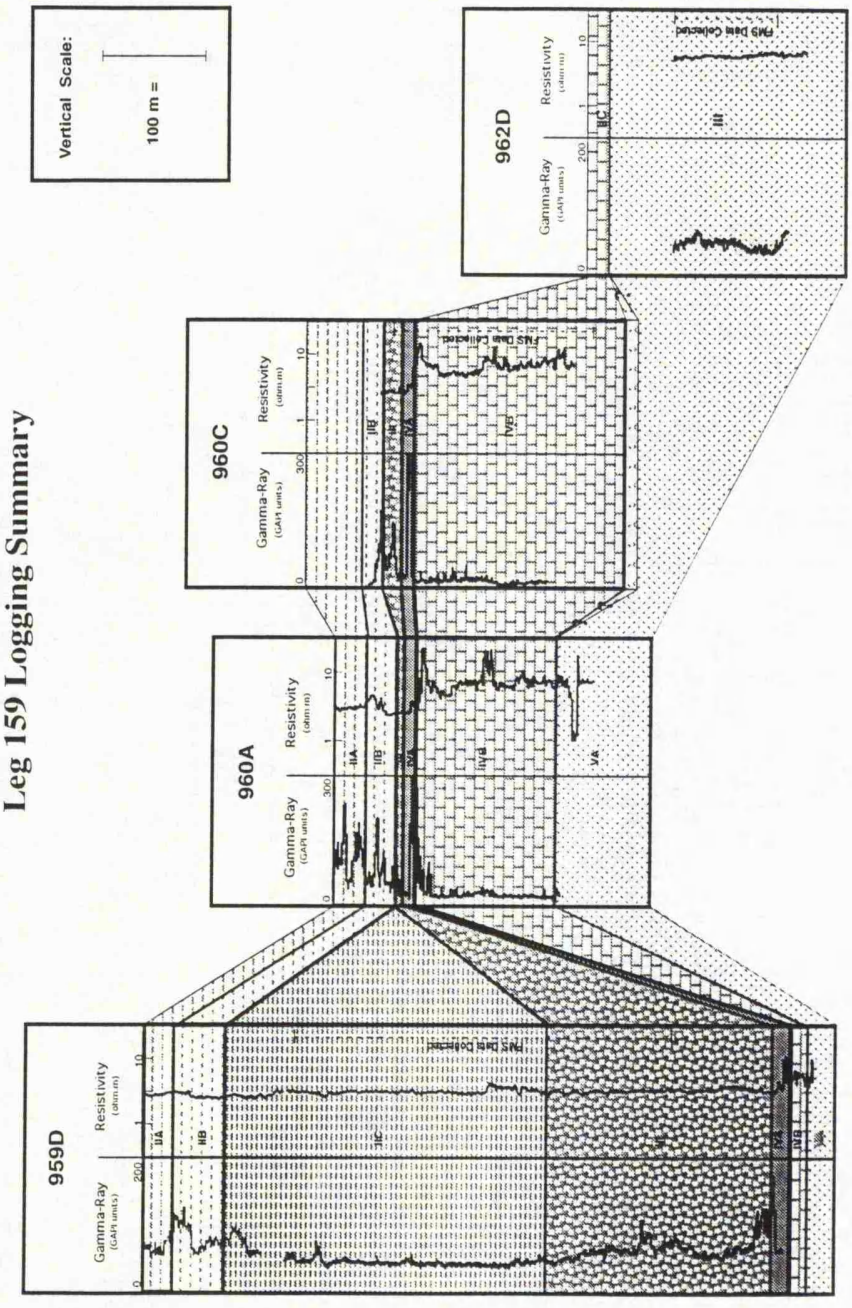


Figure 4 - Correlation between the four sites drilled during Leg 159 using gamma-ray and resistivity logs.

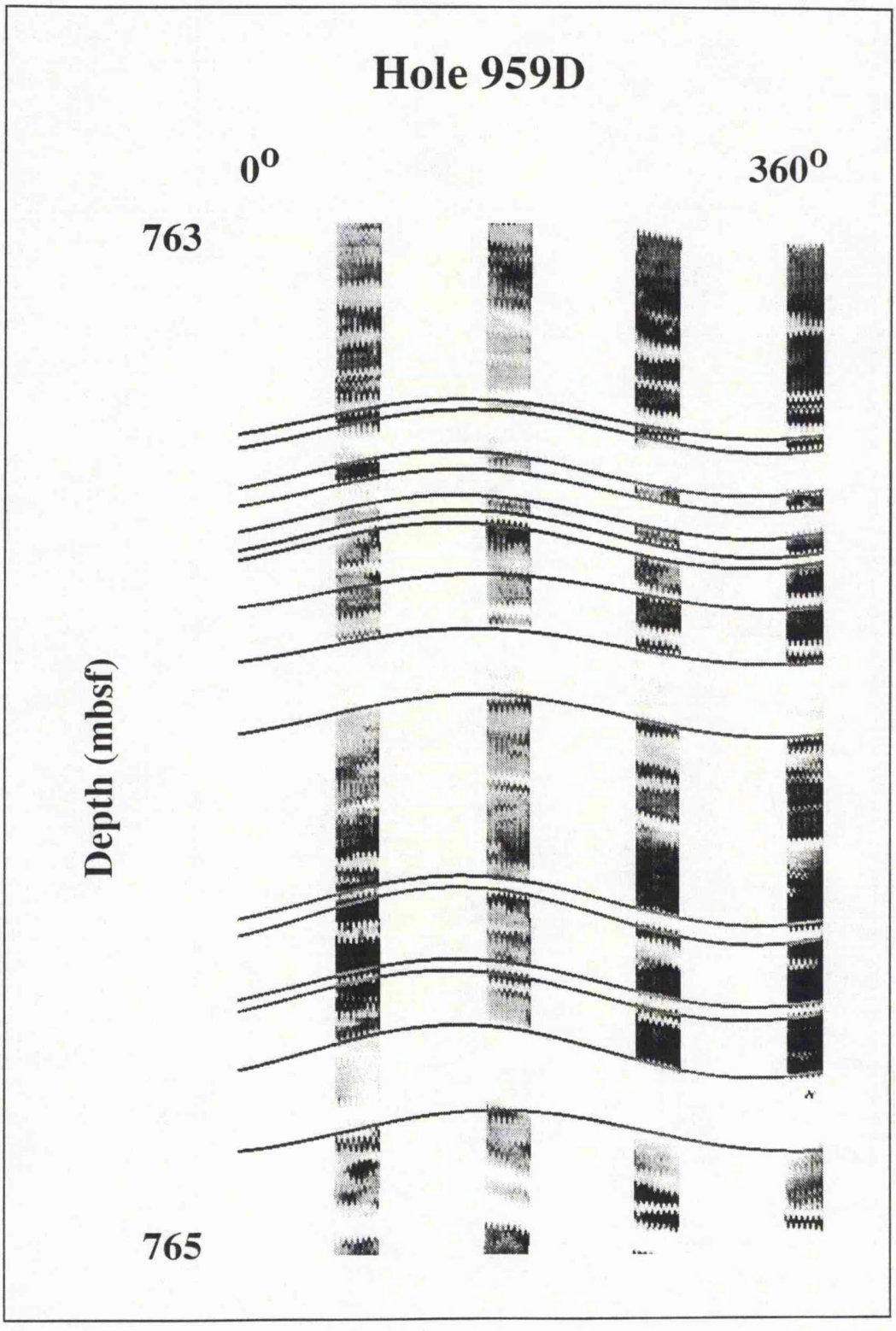


Figure 5 - FMS image for a short section in Hole 959D.

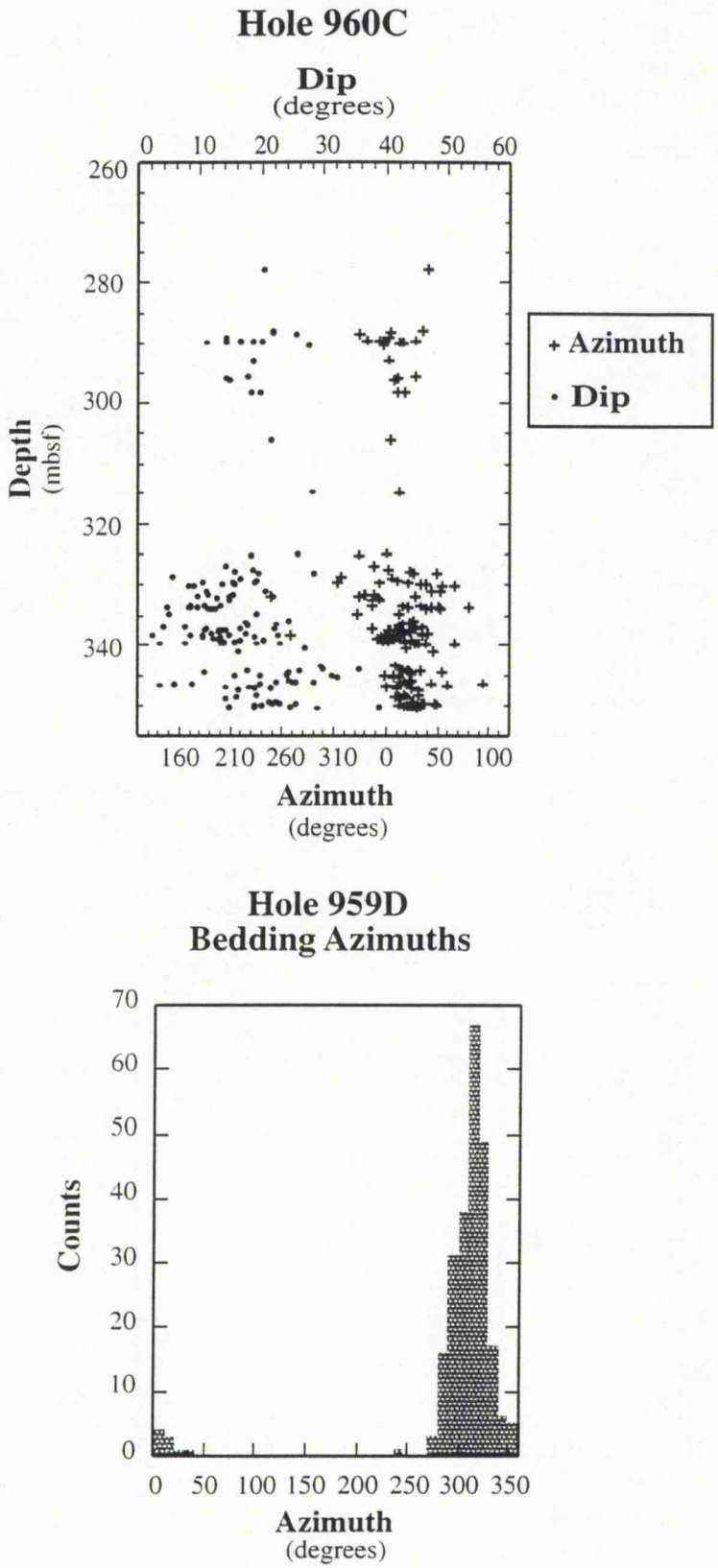


Figure 6 - Bedding and azimuths from FMS images in ODP Holes 959D and 960A

APPENDIX B

APPLICATION OF A MULTILAYER NEURAL NETWORK AND STATISTICAL TECHNIQUES IN FORMATION CHARACTERISATION

Gonçalves, C.A.[†‡]; Harvey, P.K.[†] and Lovell, M.A.[†]
[†] Borehole Research, Dept. of Geology, University of Leicester, UK.
[‡] On leave from NPGP/Federal University of Pará - Brazil.

ABSTRACT

The characterisation of formation heterogeneities requires a multidisciplinary study of data acquired using a large number of numerical geophysical and geological measurements and a rigorous evaluation of the precision and accuracy of the data. Another essential aspect of the appraisal of any measurement is the quality assessment and quality control of the data.

In this work multivariate statistical techniques and an artificial Neural Network (ANN) are used to identify heterogeneities in complex formations and to evaluate the boundaries they generate. The precision and accuracy of the data from different sources are very important and are considered here by using sample support in the integration of measurements at different scales. We use examples from two wells to show the differences in characterisation obtained with each technique.

Multivariate statistical analyses are initially used to group the petrophysical, geophysical and geological parameters extracted from the downhole measurements into distinct geologically definable zones. This technique has the advantage of being quasi-independent of any pre-determined ideas we have about the whole data set, and has shown to be very reliable in formation characterisation. Thus the result obtained here is a useful basis for comparison with that obtained from the Neural Network.

An artificial Neural Network is used to characterise the different lithology sequences present in each well. Neural Networks are relatively new tools and have proved very useful in applications where conventional computing methods are inadequate. Another application is the possibility of determining quantitative petrophysical parameters from well logs and core data in uncored intervals. We also show the ability of the Neural Network to construct petrophysical logs.

The results are presented as a comparison between the two techniques. We show that both methods are very encouraging. When comparing the ANN derived petrophysical parameter logs with actual core measurements we see a good match. Low quality petrophysical measurements can be determined by a mismatch between the responses.

INTRODUCTION

The characterisation of formation heterogeneities and the determination of petrophysical parameters are part of a key issue in reservoir evaluation using well logs. Because of the complexity of the different factors which influence the log responses, and also because the increasing amount of downhole measurements employed, a large number of techniques have been used to establish an adequate interpretation. Quantitative petrophysical measurements are also influenced, and most of the time it is difficult to determine any theoretical or empirical formula for an accurate reservoir analysis.

There are many techniques available for reservoir characterisation, giving different degrees of success in their responses. This paper compares the response of some statistical techniques, which have been used with reasonable success in formation characterisation, with the results obtained from an artificial Neural Network, using data from the Ocean Drilling Program (ODP).

When core recovery is good along the interval to be investigated, formation characterisation and petrophysical parameters can easily be obtained. However, where the recovery is poor or even where there is no recovery, a different method of interpretation and prediction of petrophysical parameters must be derived. The artificial Neural Network shows how reliable values for petrophysical parameters can be obtained from logging data even when the core recovery is poor.

MULTIVARIATE STATISTICAL TECHNIQUES

Multivariate statistical techniques have been used in well log analysis to identify or classify lithofacies (Davis, 1986; Doveton, 1986; Bucheb and Evans, 1994). They include such recognised techniques as Cluster Analysis and Discriminant Analysis with the related techniques of principal component analysis and factor analysis commonly applied to data as a pre-processing stage. They provide an effective approach to data interpretation in geology and geophysics.

The multivariate statistical techniques are also known as pattern recognition. A pattern consists of a collection of measurements characterising a sample which are considered as an entity for the purpose of subsequent analysis. The classification imposed by the

pattern recognition allocates entities to initially undefined classes so that the individuals in a class are in some sense close to one another. It is internally based in as much as it does not depend on *a priori* knowledge about the relations between these entities or samples (Sheppard, 1986). The samples are free to enter any class such that they emerge in the process of classification. The result is that the entities should be placed into approximately homogeneous groups.

The successful application of a selected technique is dependent upon careful formulation of the decision rule by the analyst and on understanding of the assumptions and limitations involved. Without this, a successful interpretation of the results may not be made. Careful considerations must be taken in the preparation of the data set and in the selection and implementation of the methods to be used. In the case of the oil industry, the term "electrofacies" has coined to describe lithofacies identified on the basis of logging data (Serra & Abbott, 1982; Delfiner et al., 1987; and Doveton, 1994).

One of the aims of this work is to find ways of identifying such "electrofacies" in different geological environments observed in ODP data. With an increasing number of log curves, the amount of information relating to the formation increases, and a more realistic picture of its nature can be produced. The aim of the application of the multivariate statistical techniques considered here is to search for structure and pattern within these data, a knowledge which may help to simplify their interpretation. Another aim is to obtain a reliable result to compare with the results obtained from core recovery and from Neural Network. Of the techniques available, Cluster Analysis and Discriminant Analysis were found to be particularly useful in the characterisation of logging data.

ARTIFICIAL NEURAL NETWORKS

An artificial Neural Network is an information-processing system that has certain performance characteristics in common with biological Neural Networks or the human brain. They have been developed as generalisations of mathematical models of neural biology, based on the following assumptions:

- information processing occurs at many simple elements called neurons;
- signals are passed between neurons over connection links;
- each connection link has an associated weight which, in a typical neural net, multiplies the signal transmitted;
- each neuron applies an activation function (usually non-linear) to its net input (sum of weighted input signals) to determine its output signal.

A Neural Network is characterised by its architecture (the way in which the neurons - the key processing unit - are interconnected), by the method of determining the weights on the connections (algorithm) and by its activation function. Neural Networks can be distinguished from other approaches to information processing by both how and when they are used. Let us therefore consider some defining characteristics of Neural Networks.

A neural net consists of a large number of processing elements called *neurons*, *cells* or *nodes*. Each neuron is connected with other neurons through connection links, each one with an associated weight. The weights represent the information being used by the neural net to resolve a problem. Each neuron has an internal state, called its *activation function*, which is a function of the input it has received. Typically, a neuron sends its activation as a signal to other neurons. Even though neurons can only send one signal each time, this signal can be broadcast to several other neurons (Fausett, 1994).

In this work we specifically use the "back-error propagation" (BEP) Neural Network. The proposal is not only the application of the BEP Neural Network in the characterisation of formation heterogeneity, but also to test its usefulness in the construction of petrophysical logs.

Back-error propagtion (BEP) Neural Network

As in the case with most Neural Networks, the aim of the back-error propagation Neural Network is to train the net to achieve a balance between the ability to respond correctly to the input patterns that are used for training (memorisation) and the ability to give reasonable responses to a input that is similar, but not identical (generalisation), to that used in training (Fausett, 1994).

The training of a back-error propagation Neural Network involves three stages: the feedforward of the input training pattern, the computation and Backpropagation of the associated error, and the adjustment of the weights. After training, the application of the net involves only the feedforward process.

A typical back-error propagation Neural Network with one layer of hidden units is shown in Figure 1. The output units Y, the hidden layer units Z and the input units X are presented as well as the associated weights w and v. As can be seen, only the feedforward process propagation direction is shown. In the backpropagation stage the signals have the opposite direction. During the first stage, each input unit X_i receives an input signal and broadcasts this signal to the each of the hidden units $Z_1, ...Z_p$. Each hidden unit then computes its activation function and sends its signal z_j to each output unit. Each output unit Y_k computes its activation y_k to form the response of the net to the previous input pattern. During the training process, each output unit compares its computed activation y_k with the expected target value t_k in order to determine the associated error or difference between both. Based on this error or difference, the factor δ_k (k=1.....m) is computed.

This value is used to distribute the errors at output units back to all units in the previous layer (the hidden layer). It is also used to update the weights between the output layer and the hidden layer. In a similar manner, the factor δ_j (j=1,...,p) is computed for each hidden unit.

After all the δ factors have been determined, the weights for all layers are adjusted simultaneously. The adjustment to the weight w_{jk} is based on the factor δ_k and the activation z_j. The adjustment of the weight v_{ij} is based on the factor δ_j and the activation x_i.

Even if the training process is slow, a trained net can produce its output very rapidly because there is no propagation of the errors in the opposite direction. Numerous variations of backpropagation method have been developed to improve the speed of the training process. In this work, the original algorithm for a backpropagation training (Rumelhart et al., 1986) uses an implementation method called Vogl's acceleration (Anguita et al., 1993).

FIELD EXAMPLES

Data from two holes in the Ocean Drilling Program were selected to be tested with these techniques. Hole 807C is located in the northern rim of the Ontong Java Plateau in the western equatorial pacific ocean (Kroenke et al., 1991). The interval between 1270 - 1400 mbsf (meters below sea floor) was selected as the interval to be investigated. It shows a reasonable core recovery (65%) and consists of three main lithofacies defined as follows: an upper unit composed mainly of limestone, chert and chalk; a middle unit consisting of claystone and siltstone; and a lower unit which comprises igneous basement composed of tholeiitic basalts. The quality of the log curves obtained is generally good (Figure 2) and their precision and accuracy were considered through sample support and core analysis.

Considering the lithology variation observed between 1270 - 1400 mbsf in Hole 807C, the logging data provides a great deal of information about not only the major changes in the different lithofacies, but also in the more subtle physical and chemical characteristics cyclically present in the formations.

The second example is from Hole 878A, located in the northeastern part of the MIT Guyot in the western pacific ocean (Premoli Silva et al., 1993). It was also selected due its excellent core recovery (95%) and well conditions. The interval to be investigated is located between 515 - 600 mbsf and consists basically of a polymitic breccia predominantly siliciclastic matrix at the top and calcareous at the bottom. The overall variation in the polymitic breccia is given by a predominance of volcanic clasts in the upper half of the interval with a decrease in abundance toward the bottom. Carbonate is dominant in the matrix and clasts below 577 mbsf but some alternation between volcanic and carbonate matrix occurs between 563 and 577 mbsf. In a general sense the logging data set (Figure 3) for this hole has good quality and seems to be valuable for interpreting the sedimentary succession and for identifying possible heterogeneities.

Multivariate statistical approach

After a pre-processing analysis with principal components, a Cluster Analysis and a Discriminant Analysis were carried out in the data from both holes. The result of these techniques applying to the data sets is described in the following sections.

Hole 807C

In Hole 807C 10 log curves were used as input data. They are: gamma-ray (GR); potassium, thorium and uranium contents (POTA, THO and URAN respectively); three different resistivity measurements (ILD, ILM and SFLU); aluminium content (ALUM), density (RHOB) and photoelectric effect (PEF).

In the Cluster Analysis using the Non-hierarchical or Q-mode model, initial testing for the number of groups present in the data set resulted in three being selected as an *optimum* number (Figure 4). As can be seen, they match almost perfectly with the lithofacies identification obtained from the core recovery. Group 1 represents the carbonate section at the top of the interval and also appears between 1375 - 1380 mbsf. Group 2 represents the claystone/siltstone lithofacies sequence present between 1351-1370 mbsf, and finally, Group 3 is defined by the tholeiitic basalt sequence from 1380 mbsf towards the bottom of the interval.

In terms of group discrimination, RHOB, ALUM and PEF were rated the most important curves based on their average delta values (Table 1), and Groups 1 and 3 and Groups 2 and 3 are the most difficult combinations to distinguish (Table 2).

Based on the classification obtained by the core recovery and also regarding the results from the Cluster Analysis, a Discriminant Analysis was used to distinguish the three units present in Hole 807C. A training set composed by short intervals of each lithofacie was selected based on the core recovery. After the processing of the discriminant functions using the training sets, those were applied for the whole interval. The result, not surprisingly in terms of the discriminant scores, also show a reasonable separation for the three lithofacies (Figure 4). The groups obtained in the Cluster Analysis and the core description are also shown in order to make a comparison with the discriminant scores. Other way to show how well the discriminant scores identify each lithofacie is looking to the cross-plot of scores I and II (Figure 5).

A general distance classification was also performed for the whole data. In this case the Euclidean distance model is used. It shows a relatively high percentage of disagreement (22.63%) between its

classification and the core description. However, this value is given mostly because of the low percentage of corrected samples in lithofacie 3 (basalts). This is due to the similar values in RHOB between this lithofacie and the carbonates (lithofacie 1), to where most of the samples were misclassified. The result was expected since the Cluster Analysis had already pointed out Groups 1 and 3 as the hardest combination to distinguish.

Hole 878A

The logging data set for Hole 878A consists of 13 log curves: porosity (NPHI); density (RHOB); two resistivity measurements (IDL and SFLU); transit time (DT); thorium, uranium and potassium contents (THO, URAN and POTA respectively); and $CaCO_3$, Al_2O_3, SiO_2, FeO and TiO_2 concentrations.

As in previous example, an initial Cluster Analysis resulted in four different groups being selected as the *optimum* number to define the data set (Figure 6). When these results are compared with core description and Cluster Analysis, Groups 1 and 2 seem to represent the siliciclastic matrix breccia with the difference between them related to the content of Si present in the matrix. Groups 3 and 4 seem to represent the carbonate matrix breccia with the difference between them also related to the Ca content.

Al_2O_3, $CaCO_3$ and RHOB appear as the most important variables for group discrimination (Table 3). The simplest and hardest group combination to distinguish are shown in Table 4. As expected, combinations 1-2 and 3-4, which represent the siliciclastic and the carbonate matrix respectively, are the hardest ones to distinguish. On the other hand, combination 1-4 is the simplest one since it represents the extremes in the matrix composition.

Due to the fact that only two major lithofacies were distinguished by core recovery, a linear Discriminant Analysis was carried out. The training sets consisting of short intervals for each lithology were based on the core description. The discriminant scores obtained by the discriminant function for the whole interval can also be observed in Figure 6, where the groups from the Cluster Analysis are also shown. The linear discriminant scores show a reasonable separation between the two main lithofacies. However, it was not able to pick the differences in the content of Si and Ca within both major units obtained by the Cluster Analysis.

A general distance classification was also performed using the Euclidean Distance model. Here, it shows a result with 93.73% of correct samples.

The Neural Network approach

The second part of this paper concerns the application of a back-error propagation Neural Network to the same data set. Firstly, the Neural Network is used to obtain a lithofacies identification and the results are compared with the ones obtained by core recovery and by the multivariate statistical approach. Secondly, the Neural Network is used to generate petrophysical parameter logs, based on the measurements obtained in cored intervals.

As showed before, the training process is done in three stages. The first stage is the training of the neural net with samples which best represent the lithofacies/petrophysical parameters to be characterised/computed. The input layer of the net is given by the number of log curves used. Bearing in mind the log data is an average over their resolution, the depth matching error must be minimum to avoid wrong input data set. This is performed using sample support from core data. In the case of the petrophysical parameters, a windowing technique (Jenner and Baldwin, 1994) is used to compensate any minor depth shift. In this case, more than one depth unit is used to train the net (Figure 7).

The second stage is the computation and backpropagation of the errors, the difference between the output value and the expected (target) value. This error will be used in a third stage to change the connection weights and then to enhance the Neural Network result.

In order to give more confidence to the results, the net is also trained to provide values of known quantities, instead of only giving the unknown information. This is applied in the computation of the petrophysical logs.

Results

The structure of the net used for lithofacies identification in Hole 807C has 10 neurons in the input layers. They correspond to the number of log curves used. The hidden layer has 7 neurons and the output layer 3 neurons which represent the three different lithofacies observed. In this case and also for Hole 878A, we tried to use the same training interval used in disciminant analysis in order to make a comparison between the results. After the training process, the weights computed from the training set were applied in a feedforward process to the whole interval. The response given by the net for Hole 807C can be observed in Figure 8. Unit A, Unit B and Unit C represent the carbonate section at the top, the siltstone/claystone sequence in the middle and the volcanic sequence at the bottom of the interval respectively. As can be seen, all three lithofacies appear well defined by the net. There are some sections where the result given by the net is a mixture between two and even three lithofacies. This is mostly due to similar log curve characteristics in those sections. However, it is possible to define the major lithofacie present in all this sections. The only interval where the net was not able to give a reasonable response is between 1375-1380 mbsf. Although Unit B appears to be stronger than the other lithofacies, they are also quite well represented in this interval. The group classification from the Cluster

Analysis is also shown in order to compare the results. For the above mentioned interval, it shows Group 1 (Unit A) as the strongest lithofacie in this interval. When compared with core description, the net results misclassify 2.87% of the samples.

The structure of the net for Hole 878A has 13 neurons in the input layers corresponding to the 13 log curves used in the classification. The hidden layer has 8 neurons and the output layer has 2 neurons corresponding to the two different matrix composition of the polymitic breccia. Usually, the number of nodes in the hidden layer is given by the performance of the net during the training process. Sometimes more than one step in the training process is necessary before we get the appropriate number of nodes in the hidden layer.

The result for the feedforward process (generalisation) is shown in Figure 9. Unit A and Unit B represent the two different polymitic breccia matrix. As in Hole 807C, here the two main lithofacies also appear well defined by the neural net response. Only one interval, between 540 - 550 mbsf, appears with its result as a mix between the two lithofacies. The interval 563 - 577 mbsf which is defined as an interbedded interval between the two different lithofacies also appears well characterised. Both the thin layers of Si and Ca matrix between 568 - 571 mbsf were picked up by the net. When compared with the core description, the net result shows 4.84% of misclassified samples. The groups from Cluster Analysis also show good agreement with the net results.

Comparison of the BEP Neural Network predicted porosity and water content values to the actual core porosity and water content measurements were made for both holes. Figures 10a and 10b show both petrophysical parameters measurements from Neural Network and core for Holes 807C and 878A.

In case of Hole 807C, a good agreement between both measurements can be observed especially for the upper interval (carbonates - Unit A), represented by low porosity and low water content. For the intermediate unit (siltstone/claystone - Unit B), the results seem to be a little more scattered although a reasonable agreement can also be seen. In Hole 878A, due to the complex lithology determined by the breccias, the results for both petrophysical parameters appear more scattered.

These results are also reflected when petrophysical logs (Figures 11 and 12) are generated for the whole interval in both holes. As can be seen, Hole 807C shows a better agreement between the generated log and the core data. For Hole 878A the variation obtained in the core measurements are reflected in the neural net result, which exhibits more variation than in the previous example.

CONCLUSION

The lithofacies characterisation given by the statistical techniques and the Neural Network has shown excellent results. In both Hole 807C and Hole 878A the discrepancies between the results obtained by each technique and the classification obtained by the core description generally do not exceed 7%. The Neural Network shows better results for both holes with 2.87% and 4.84% of misclassified samples for Holes 807C and 878A respectively, while the statistical Discriminant analysis shows 22.63% and 6.27% for the same holes. The difference could be due to greater complexity in the lithofacies of Hole 878A. The differences in Ca and Si contents can only be observed in some of the log curves used for the analysis.

The ability of the Neural Network to construct petrophysical logs was also tested with reasonable success. The water content and porosity values obtained for each well show good quality when compared with the core measurements. In Hole 878A, although the net results are still coherent with the core data, a minor discrepancy between the results is observed. The greater discrepancy for the bottom part of the interval in Hole 878A is due to the poor amount of core measurements used to train the net. It serves to illustrate the importance of using a reasonable quantity of measurements in the training process. The Neural Network can also be used to determine any other petrophysical parameter as long as it is provided with a correct training set.

REFERENCES

- Anguita, D., Parodi, G. and Zunino, R., 1993, Speed improvement of the Back-propagation on the current generation workstation. WCNN '93, July 11-15 1993, Portland, USA, pp 165-168.
- Bucheb, J.A. and Evans, H.B., 1994, Some applications of methods used in electrofacies identification. The Log Analyst, v.35, n. 1, p. 14-26.
- Davis, J.C., 1986, Statistics and data analysis in geology. J.Wiley and sons, New York.
- Delfiner, P.C., Peyret, O. and Serra, O., 1987, Automated determination of lithology from well logs. SPE Formation evaluation, 2(3): 303-310.
- Doveton, J.H., 1986, Log analysis of subsurface geology. John Wiley and sons, New York, 273 pp.
- Doveton, J.H., 1994, Geologic log analysis using computing methods. AAPG Computer Applications in Geology, n.2, 169 pp.
- Fausett, L., 1994, Fundamentals of Neural Networks: architectures, algorithms and applications. Prentice Hall, New York.
- Jenner, R. and Baldwin, J., 1994, Application of an artificial Neural Network to obtain an improved permeability profile. SPWLA 16th European Formation

Evaluation Symposium. Aberdeen, Scotland, 11-13 October, paper M.
- Kroenke, L.W., Berger, W.H., Janecek, T.R. et al., 1991, *Proc. ODP Initial Reports,* 130: College Station, TX.
- Premoli Silva, I., Haggerty, J., Rack, F. et al., 1993, *Proc. ODP Initial Reports,* 144: College Station, TX.
- Rumelhart, D.E., McClelland, J. and the PDP research group, 1986a, Parallel distributed processing, vol. 1, the MIT Press.
- Serra, O. and Abbott, H.T., 1982, The contribution of logging data to sedimentology and stratigraphy. Soc. Pet. Eng. Journal, v. 22, n. 1, p. 117-131.
- Sheppard, A., 1986, The geochemistry and evolution of the Lizard complex, Cornwall. PhD thesis, University of Nottingham, UK, 251 pp.

ABOUT THE AUTHORS

Carlos A. Gonçalves graduated from Federal University of Rio de Janeiro, Brazil, in 1989 with a BSc degree in Geology. He obtained his MSc in Geophysics from Federal University of Para, Brazil, in 1991. He is now a PhD student at University of Leicester, UK, as part of a split scheme between University of Leicester and Federal University of Para. His research interests include well log analysis and computer modelling. He is sponsored by CNPq - Brazilian Research Council.

Peter Harvey obtained a BSc degree in Geology from Bristol University in 1964 and a PhD in the same university in 1967. From 1967 to 1989 he was a lecturer/senior lecturer in Geology at the University of Nottingham, with an interval during 1971/1972 at the Geological Survey of Kansas. He transferred to the University of Leicester in 1989 where his research is concerned with nuclear and electrical logging and computer modelling.

Mike Lovell obtained a BSc degree in Geological Geophysics from the University of Reading in 1977. He gained a MSc in Marine Geothecnics and in 1984 was awarded a PhD for research into the physical properties of sediments, both from University of Wales. In 1984 he joined the University of Nottingham as a lecturer in Geology, and in 1989 transferred to the University of Leicester. In 1992-1993 he was a Visiting Professor in Borehole Geophysics at the Universidade Federal do Para in Belem, Brasil.

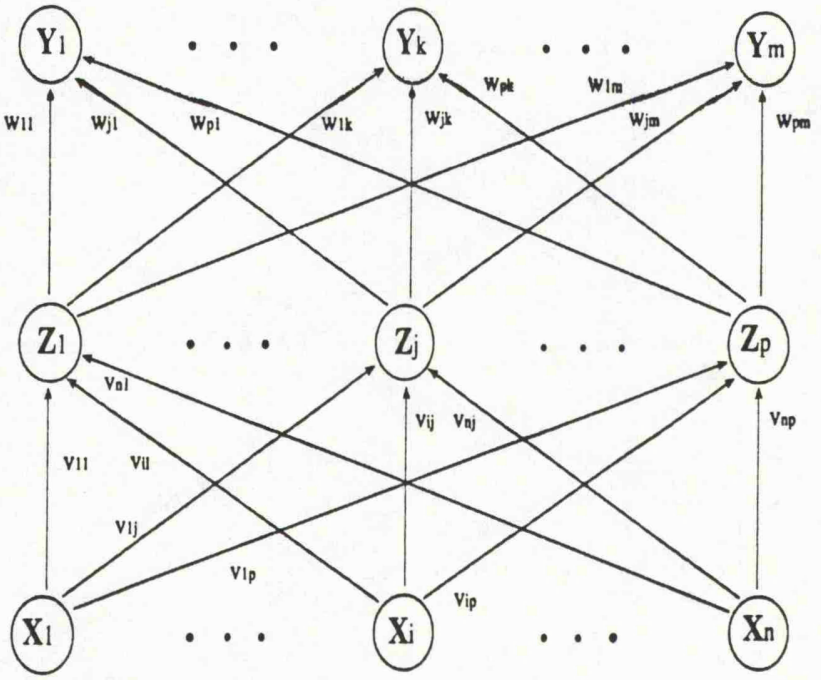


Figure 1 - Typical back-error propagation neural network with one hidden layer.

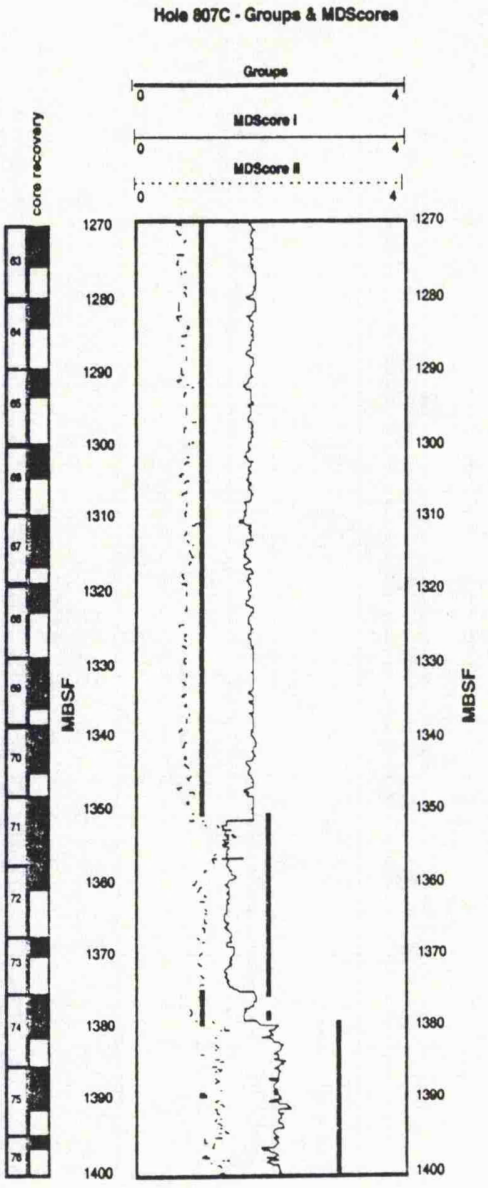


Figure 2 - Group classification and multidiscriminant scores for Hole 807C.

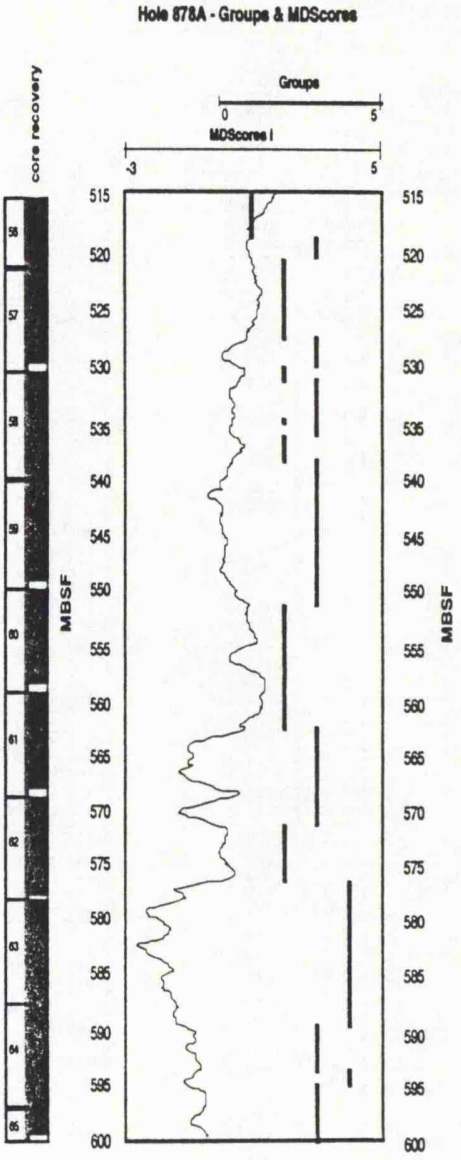


Figure 3 - Group classification and multidiscriminant scores for Hole 878A.

FF

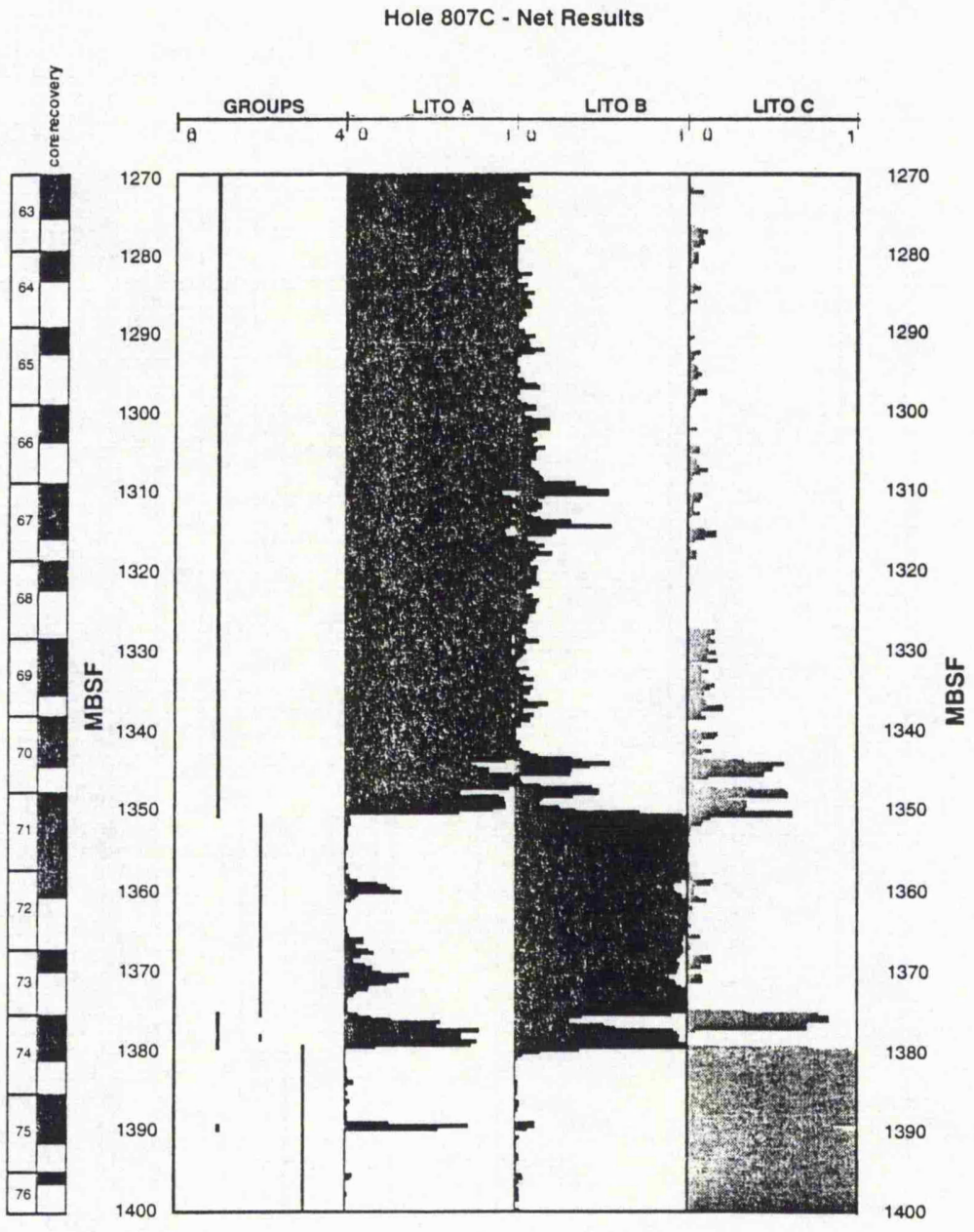


Figure 4 - Result of back-error propagation neural network for lithofacies identification in Hole 807C.

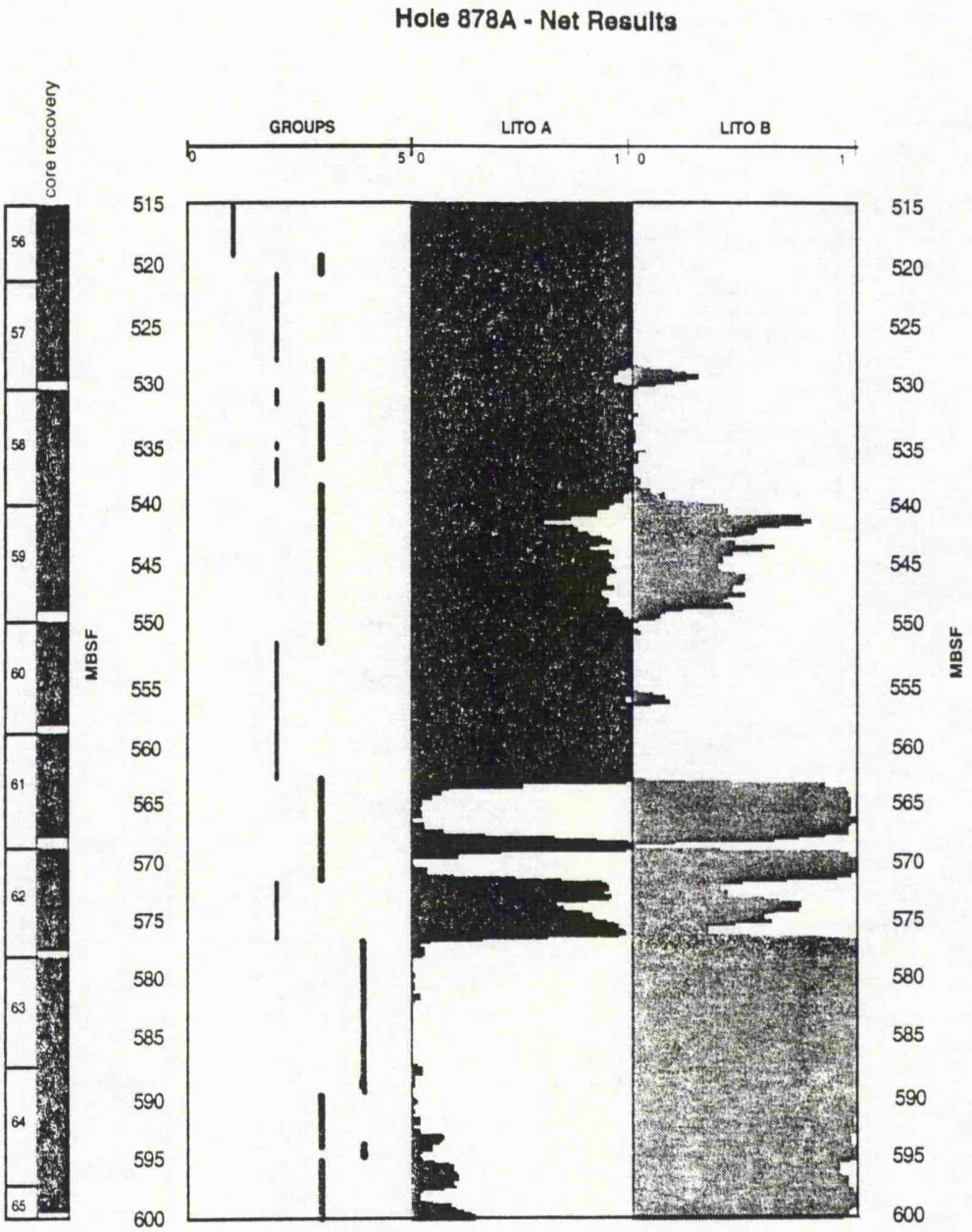


Figure 5 - Result of back-error propagation neural network for lithofacies identification in Hole 878A.

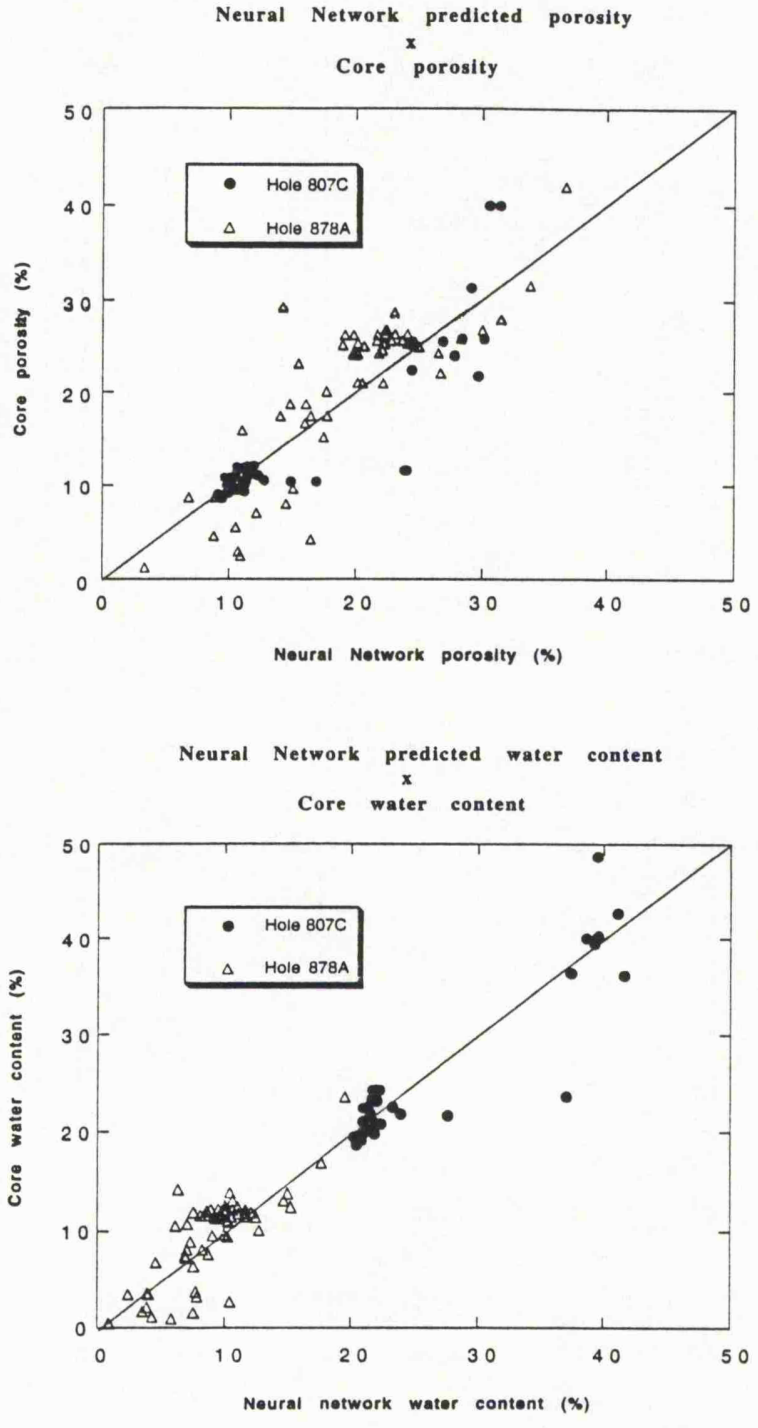


Figure 6 - Cross-plot of predicted neural network and core porosity (top) and water content (bottom) for Holes 807C and 878A.

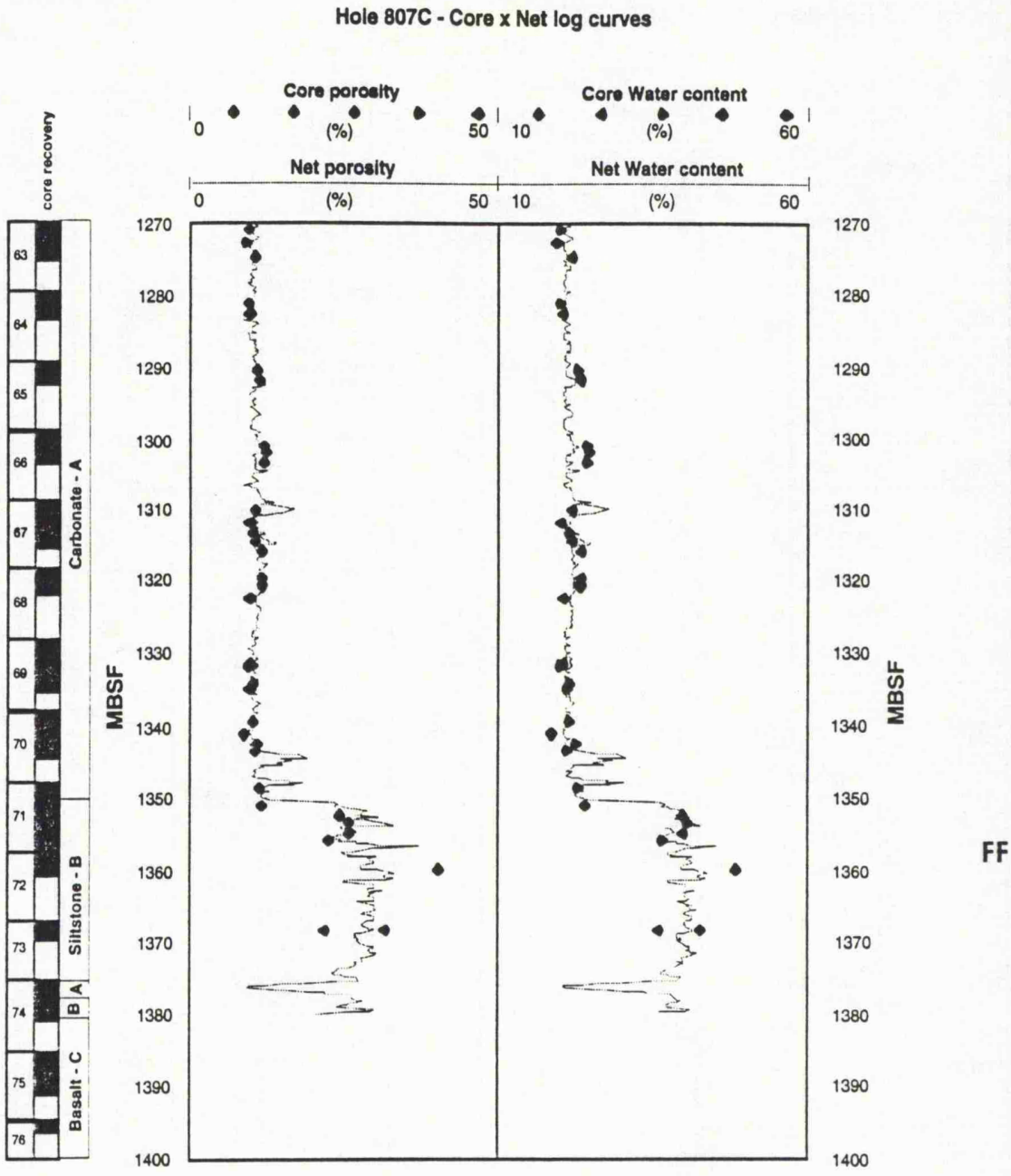


Figure 7 - Neural network generated petrophysical parameters log for Hole 807C.

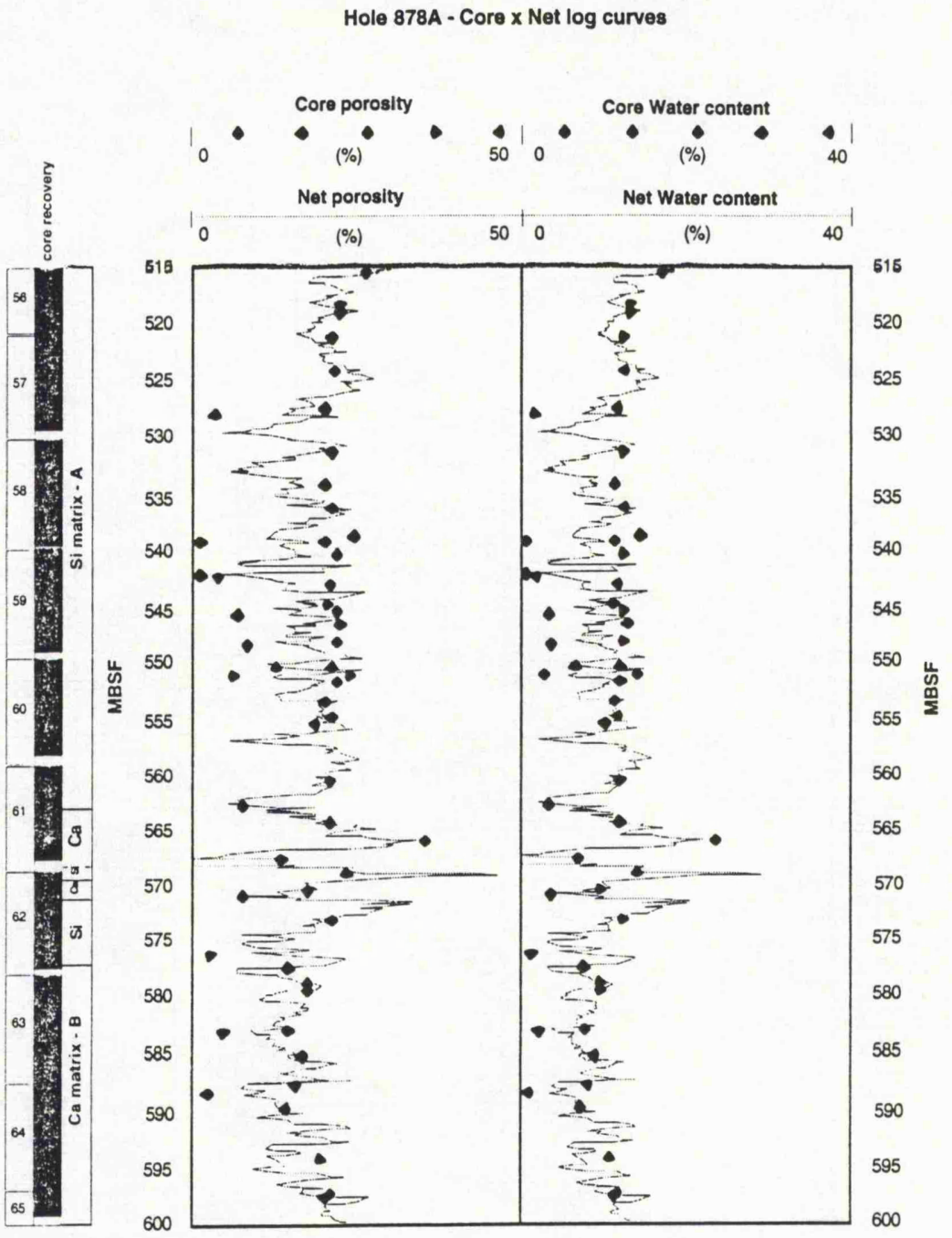


Figure 8 - Neural Network generated petrophysical parameters log for Hole 878A.