*Terra Antartica*
2000, 7(3), 299-310

# Analysis of Downhole Logging Data from CRP-2/2A, Victoria Land Basin, Antarctica: a Multivariate Statistical Approach

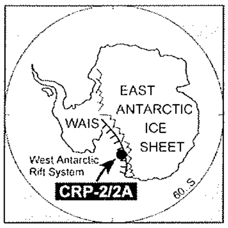
C.J. BÜCKER[1*], R.D. JARRARD[2], T. WONIK[1] & J.D. BRINK[2]

[1]GGA, Joint Geoscientific Research Institute, Stilleweg 2, 30655 Hannover - Germany
[2]Dept. of Geology & Geophysics, University of Utah, 717 WBB, 135 S. 1460 East, Salt Lake City - UT 84112-0111 - USA

**Corresponding author* (c.buecker@gga-hannover.de)

*Received* 30 July 1999; *accepted in revised form* 27 March 2000

**Abstract** - In the northern McMurdo Sound (Ross Sea, Antarctica), the CRP-2/2A drillhole targeted the western margin of the Victoria Land Basin to investigate Neogene to Palaeogene climatic and tectonic history by obtaining continuous core and downhole logs. Well logging of CRP-2/2A has provided a complete and comprehensive dataset of *in situ* geophysical measurements.




This paper describes the evaluation and interpretation of the downhole logging data using multivariate statistical methods. Two major types of multivariate statistical methods were used, each yielding a different perspective: (1) Factor analysis was used as an objective tool for classification of the drilled sequence based on physical and chemical properties. The factor logs are mirroring the basic geological controls (*i.e.*, grain size, porosity, clay mineralogy) behind the measured geophysical properties, thereby making them easier to interpret geologically. (2) Cluster analysis of the logs groups similar downhole geophysical properties into one cluster, delineating individual logging or sedimentological units. These objectively and independently defined units, or statistical electrofacies, are helpful in differentiating lithological and sedimentological characterisations (*e.g.* grain size, provenance).

The multivariate statistical methods of factor and cluster analysis proved to be powerful tools for fast, reliable, and objective characterisation of downhole geophysical properties at CRP-2/2A, resulting in interpretations which are consistent with sedimentological findings.

## INTRODUCTION

### CAPE ROBERTS PROJECT

The main aims of the Cape Roberts Project are to document past variations in Antarctic ice cover and climate and to reconstruct the early uplift history of the nearby Transantarctic Mountains. The Cape Roberts drillholes CRP-1 (Cape Roberts Science Team, 1998) and CRP-2/2A (Cape Roberts Science Team, 1999) are located in a sedimentary basin just seaward of the edge of the present ice sheet, about 20 km offshore Cape Roberts, a small cape *c.* 125 km NE of McMurdo, Ross Island. The time period expected to be sampled by the drillholes (10–50+ Ma) is of interest because present knowledge suggests that it includes the time when Antarctica changed from an ice-free continent to an ice-covered continent. A detailed description of the project and its aims, geological setting, and preliminary results is given by Cape Roberts Science Team (1998, 1999). CRP-2A extended to 624.15 mbsf (metres below sea floor) with an average 95% recovery of Oligocene to Quaternary sediments. Most of the downhole logging tools were be run to the bottom of the hole. Coring and downhole logging of the drillholes are essential prerequisites to achieve the aims of the project.

### DOWNHOLE LOGGING

A detailed description of the downhole logging tools used in CRP-2A and of the downhole logging techniques is given in the Initial Report on CRP-2/2A (Cape Roberts Science Team, 1999). These downhole logs provide a representative record of *in situ* physical properties of formations adjacent to the drillhole. A total of 15 physical and chemical parameters was measured by eight tools. Almost all tools were run over the entire borehole section down to the final depth of 624.15 mbsf. In addition to these conventional borehole measurements, a borehole televiewer (BHTV) was run and a vertical seismic profiling (VSP) experiment was carried out; the results of these are presented by Moos et al. (this volume) and Henrys et al. (this volume), respectively. The temperature profiles are published by Bücker et al. (this volume).

### MULTIVARIATE ANALYSES

Interpretation of the comprehensive suite of logs requires a combination of geophysical and sedimentological perspectives. The abundance of logging data and the demand for a quick, objective, and reliable evaluation and interpretation call for the application of multivariate statistical methods. Multivariate factor analysis is a method of reducing the amount of logging data without losing important information. The result is a set of factor logs that provide a new integrated presentation and are helpful tools for further interpretation. Subsequent cluster analysis of the most significant factor logs is a useful and objective method for identifying and confirming significant downhole log characteristics.

Multivariate statistical analyses are seldom applied to logging data, yet they are an excellent method of handling the large amount of logging data and meeting the demand for a fast, reliable and objective evaluation and interpretation. In this study, the multivariate procedures of factor and cluster analysis are applied to the CRP-2A logging measurements. Factor analysis is used in order to rescale and reduce the original dataset and for deriving a deeper insight into the interrelated rock properties and background processes. Cluster analysis is used to „block" log data and define sedimentological characteristics such as grain size and provenance changes as objectively as possible, which is particularly important in depth intervals with core loss.

## DOWNHOLE LOGGING DATA AND QUALITY

An almost complete set of downhole logging data was recorded in hole CRP-2A using tools from the Institute for Joint Geoscientific Research (GGA, Germany). Although these tools differ slightly from commercial wireline logging tools (*e.g.*, Schlumberger), they are based on the same physical principles and produce comparable results. A full description of the principles and measurements performed by the logging tools is given by Cape Roberts Science Team (1999). The data are considered to be of generally good quality. This is the most complete and comprehensive dataset of *in situ* geophysical measurements ever obtained in Antarctica up to now.

The downhole measurements of physical and chemical properties used in this study consist of the following: spectral gamma ray (GR), thorium, uranium, and potassium contents (Th, U, K), formation bulk density (den), electrical resistivity long spacing (RLong), magnetic susceptibility (sus), sonic velocity (Vp), and neutron porosity (phi). The radius of investigation and vertical resolution of each logging tool depend on the measuring principle and measured property. A summary of these tool responses is given in the CRP-2 Initial Report (Cape Roberts Science Team, 1999).

Figure 1 is a composite plot of all downhole measurements used in this study, together with a simplified lithological profile derived from visual core descriptions (Cape Roberts Science Team, 1999). Integrated in figure 1 are also the continuous, whole-core measurements made in the on-site core lab (Niessen et al., this volume) and the laboratory measurements on core plugs (Brink et al., this volume). The excellent correlation between downhole, whole-core, and core-plug measurements demonstrates that these independently obtained datasets are of overall good quality and are well matched with respect to depth.

### SPECTRAL GAMMA RAY

Spectral Gamma Ray was one of the first measurements after completion of drilling operations (the first measurement was temperature), and the tool was run over the entire borehole depth. Gamma Ray (GR) values together with thorium (Th) and potassium (K) content show a general decrease downhole, whereas uranium (U) varies around a nearly constant value. Th and K have the highest correlation coefficients to GR and thus are the dominant contributors. Often the GR is used as a shale/sand indicator and to estimate shale content (Rider 1996). However, this normally expected distinction between mudstone and sandstone GR values is often subtle within CRP-2, as is evident when comparing the spectral gamma-ray log to lithology (Fig. 1). In the top 300 m of the borehole, there is no obvious correlation between lithology and GR. In contrast, the lower part of the borehole exhibits a generally good correlation, with high values for mudstones and low values for sandstones. This dramatic change in gamma-ray behavior at about 300 mbsf indicates a change in petrofacies and provenance associated with a major unconformity at this depth.

Wherever changes in grain size are reflected in changes in the GR (*e.g.* below 300 mbsf), this tool can be used as a facies indicator. Because of this relationship, the GR log shows a close correlation with the core-based palaeobathymetry curve (Cape Roberts Science Team, 1999).

### DENSITY AND VELOCITY

The downhole logs of velocity and density include gaps due to special borehole conditions (drill string in place at 0-60 and 170–200 mbsf, hole collapse at 444 mbsf) and measurement problems (255-280 mbsf). Differences between log density and core density in the depth interval 170-200 mbsf can also be attributed to drillstring effects. As expected, the core-plug measurements of P-wave velocity are in general slightly higher than the logging data. Usually plugs are taken from intact core sections, resulting in a bias to higher velocities. Average sonic velocity is about 2.0 km/s in the upper part of the borehole, but there is a sharp increase to about 2.7 km/s at 300 mbsf, corresponding presumably to the early/late Oligocene boundary (*ca.* 28-30 Ma). Velocity (as well as density) increases irregularly downhole to about 4.0 km/s at the bottom of the hole. The seismically derived V4/V5 boundary is estimated to correspond with a sharp velocity increase at 440 mbsf (Henrys et al., this volume). The higher velocities below 300 mbsf probably reflect more extensive carbonate cementation (Dietrich, this volume).

### NEUTRON POROSITY

The neutron porosity (phi) measurement does not simply respond to formation porosity; instead, it is a measurement of the total hydrogen content within the bulk rock. Thus, in clay-rich formations, phi records the combined effect of porosity and clay content. As different clay types have different percentages of bound water (Rider, 1996), the neutron porosity measurement gives an integrated mix of information about porosity, clay content, and clay type. Neutron porosity measurements are most accurate in formations with porosities not higher than 40% (Theys, 1991). While CRP-2A neutron porosities are less than 60%, they are considered to result in reliable phi data.

The advantage of having both core and neutron porosity measurements is that comparison of the two can provide

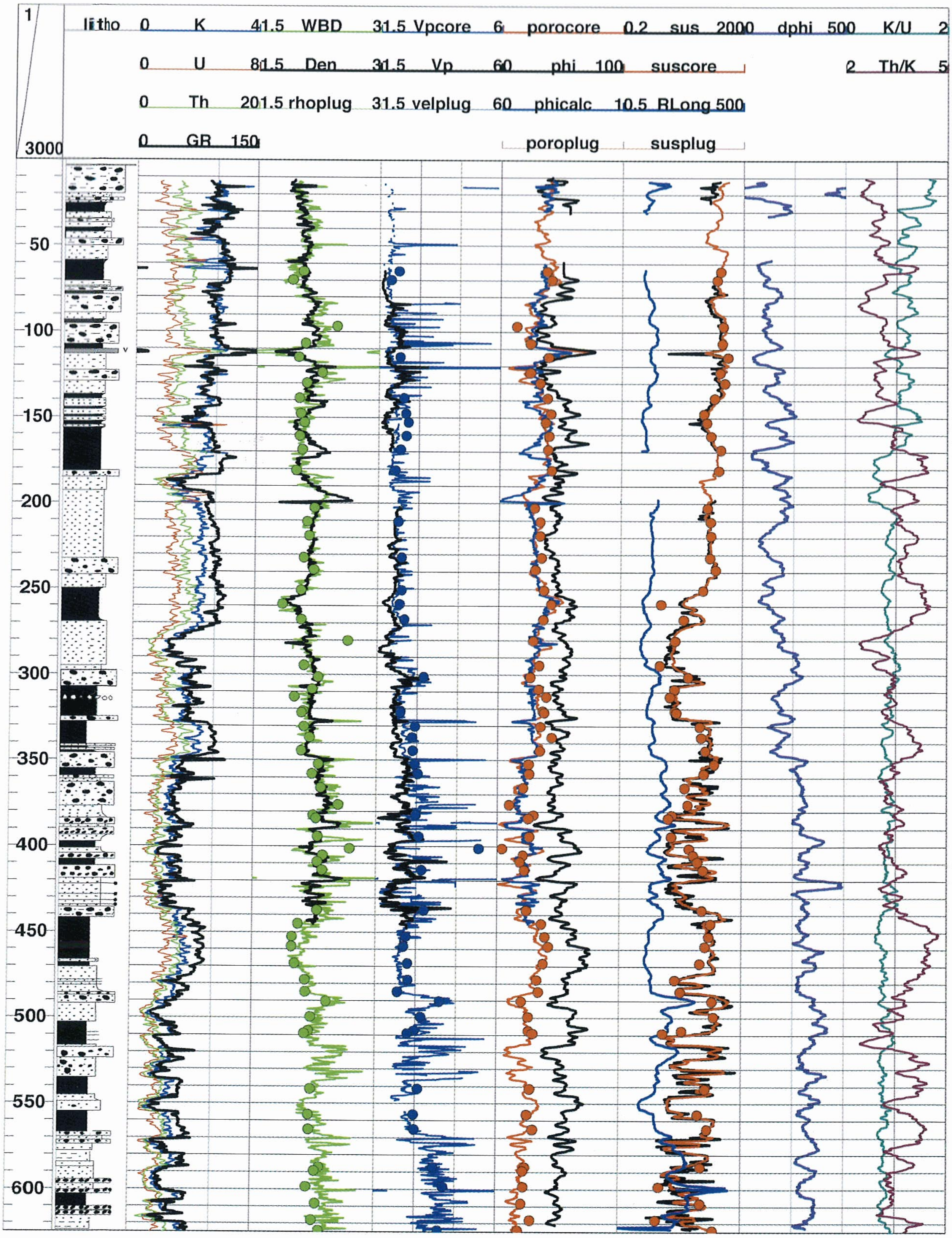



*Fig. 1* - Core derived lithology and downhole logs for hole CRP-2A. From left to right, the following parameters are shown: first column: core derived lithology; second column: potassium (K, 0-4%), uranium (U, 0-8 ppm), thorium (Th, 0-20 ppm), gamma ray (GR, 0-150 API); third column: whole-core density (WBD, 1.5-3.0 g/cm$^3$), density log (Den, 1.5-3.0 g/cm$^3$), core-plug density (rhoplug, 1.5-3.0 g/cm$^3$); fourth column: whole-core P-wave velocity (Vpcore, 1.5-6.0 km/s), velocity log (Vp, 1.5-6.0 km/s), core-plug velocity (velplug, 1.5-6.0 km/s); fifth column: whole-core porosity (porocore, 0-100 p.u.), neutron porosity log (phi, 0-100 p.u.), calculated porosity (phicalc, 0-100 p.u.), core-plug porosity (poroplug, 0-100 p.u.); sixth column (logarithmic scale): susceptibility log (sus, 0.2-200 $10^{-5}$ SI), whole-core susceptibility (suscore, 0.2-200 $10^{-5}$ SI), core-plug susceptibility (susplug, 0.2-200 $10^{-5}$ SI), electrical resistivity log (Rlong, 0.5-50 Ohmm); seventh column: delta porosity (neutron porosity – core porosity) (dphi, 0-50 p.u.); eighth column: potassium/uranium ratio (K/U, 0-2·$10^{4}$), thorium/potassium ratio (Th/K, 2-5·$10^{-4}$). (lithology legend see Barrett et al., this volume).

further insights. The difference between these two measurements is shown as the dphi curve in figure 1. This delta-porosity should isolate information about clay type and clay content, because intergranular porosity measured on cores (using density logging methods) has been removed. The broad pattern of this dphi curve is a downhole increase, pointing to differences in clay type and/or clay content with depth. A step increase in the dphi log can be seen at about 350 mbsf, below this depth the dphi values are constantly higher than in the upper part of the borehole. Whereas neutron porosities are affected both by free water and by different interlayer water contents of clays, the dphi curve is expected to respond mainly to clays. Smectite, for example, has an average interlayer water content of 18-22% (Weaver, 1973), resulting in a neutron porosity value of 0.44 (Rider, 1996), whereas illite has only 8% interlayer water content and a neutron porosity value of 0.30.

For siliciclastic rocks, the responses of the neutron porosity log ($\Phi_{nlog}$) and the density log ($\rho_{log}$), along with the mass balance equation, are given by the following:

$$\rho_{log} = \Phi_{true} * \rho_f + V_{ss} * \rho_{ss} + V_{cl} * \rho_{cl} \quad (1)$$

$$\Phi_{Nlog} = \Phi_{true} * \Phi_{Nf} + V_{ss} * \Phi_{Nss} + V_{cl} * \Phi_{Ncl} \quad (2)$$

$$1 = \Phi_{true} + V_{ss} + V_{cl} \quad (3)$$

with the material properties and indices:
Sandstone matrix: $\rho_{ss} = 2.65$ g/cm$^3$ $\Phi_{Nss} = 0$ d.u.
Clay matrix: $\rho_{cl} = 2.67$ g/cm$^3$ $\Phi_{Ncl} = 0.4$ d.u.
Fluid (sea water): $\rho_f = 1.04$ g/cm$^3$ $\Phi_{Nf} = 1.0$ d.u.
f fluid, cl clay, ss sandstone, N neutron-tool, d.u. decimal units

Equations (1) – (3) can be combined, eliminating the unknown quantities of sandstone content ($V_{ss}$, volume percent) and clay content ($V_{cl}$, volume percent), and solving for true porosity ($\Phi_{true}$) (Western Atlas 1992, Serra 1986):

$$\Phi_{true} = \frac{\Phi_{Ncl} * (\rho_{log} - \rho_{ss}) - \Phi_{Nlog} * (\rho_{cl} - \rho_{ss})}{\Phi_{Ncl} * (\rho_f - \rho_{ss}) - (\rho_{cl} - \rho_{ss})} \quad (4)$$

The true porosity as calculated by equation (4) is shown in figure 1 in the column „phicalc". The high correlation between core-based porosity and this log-based true porosity (R=0.70 with linear regression equation: phicalc = 0.94 + 0.99* porocore) indicates that the calculation procedure given above is accurate. Based on this procedure, the neutron porosity measurement can be used to give information about both true formation porosity and clay content/type.

MAGNETIC SUSCEPTIBILITY

The magnetic susceptibility tool was run in the open-hole intervals 12-25, 63-170, and 200-624 mbsf. Due to the lack of a tool calibration for this new tool, log-based susceptibility was calibrated to whole-core volume magnetic susceptibility ($10^{-5}$ SI) based on linear regression of the logarithm of both measurement suites.

Like GR, magnetic susceptibility can be used as a grain-size indicator. Fine grained sediments usually have much higher concentrations of both magnetic and radiogenic minerals than coarse grained sediments. In general, CRP-2A magnetic susceptibility shows a similar behaviour to spectral gamma ray: correlation to lithology is good in the lower part of the borehole and poor above 300 mbsf. A general change in susceptibility as well as GR behaviour is evident at this depth, with a higher variability for values below 300 mbsf. Thus the magnetic susceptibility and GR logs are most useful as grain-size indicators below 300 mbsf.

ELECTRICAL RESISTIVITY

Electrical resistivity was measured with two depths of investigation: Rlong for the deep depth of investigation, and Rshort for the shallow depth of investigation. In figure 1 only Rlong is shown because Rshort is more affected by drilling activities and borehole wall infiltrations.

Because of the large depth of investigation of this tool, electrical resistivity shows much smaller variations than all the other measurements. Electrical resistivity is mainly responding to formation porosity as the pore fluids are affecting electrical conductivity rather than the rock matrix. In general, electrical resistivity shows a downward increase, reflecting a decrease in porosity and an increase in cementation.

## STATISTICAL METHODS AND THEORETICAL BACKGROUND

A description of the basic „on-ice" data treatment is given in the Initial Report of CRP-2 (Cape Roberts Science Team 1999). In this paper, a detailed procedure of the statistical data processing is described and documented. Excellent reviews of general statistical techniques, their use in geosciences, and examples of their use in borehole geophysics are given by Backhaus et al. (1996), Brown (1998), Bucheb & Evans (1994), Doveton (1994), Davis (1986), Elek (1990), Howarth & Sinding-Larsen (1983), and Rider (1996). The multivariate statistical procedure for the evaluation of the downhole logs is outlined in figure 2 and described below.

DATA PREPARATION, QUALITY CONTROL

The first step consists of filling all gaps in the downhole-measurement dataset, using corresponding whole-core measurements. For the density, porosity, and velocity logs, about 35% were filled up by whole-core measurements (mostly below 445 mbsf), whereas for the susceptibility the share of core measurements is less than 10%. The result is a complete dataset of parameters that will be used for the statistical analyses. Validity of this procedure is verified by the excellent correlation between whole-core and downhole measurements.

The statistical methods employed in this paper require

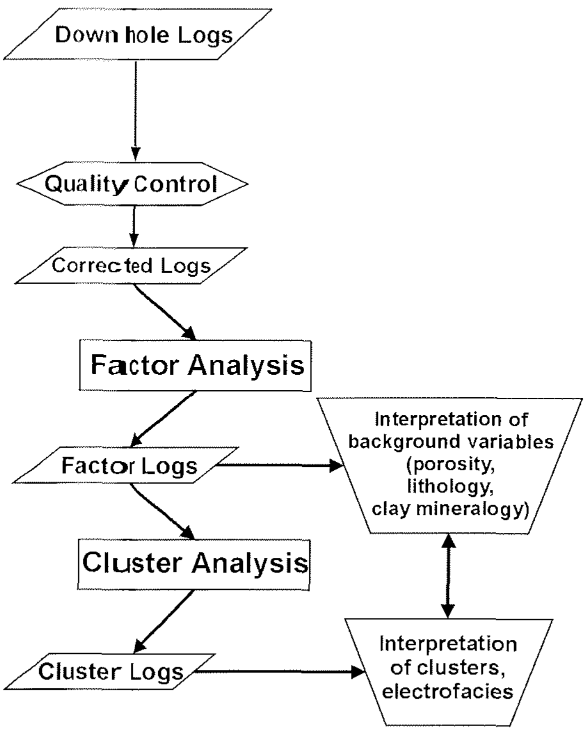



*Fig. 2* - General outline of the multivariate statistical procedure used here for evaluation and interpretation of downhole logs. The quality control includes despiking, standardization, and, if necessary, taking the logarithm of a log. The corrected logs are the input variables for the factor analysis, and the factor logs are the input variables for the cluster analysis, resulting in the clusterlogs. For the final interpretation, the factor logs as well as the clusterlogs are used.

that each observational dataset (*i.e.* geophysical log) be normally distributed. When this is not the case, the observations should be transformed so that they more closely follow a normal distribution. For example, electrical resistivities and magnetic susceptibilities often exhibit log-normal distributions, so application of a logarithmic transform to these logs will yield observations that are more normally distributed. Erroneous values, when they can be clearly identified, must also be omitted from the analysis, but these amounted to less than 1% over the entire borehole. Often, single peaks in the dataset could be attributed to lonestone effects. Fortunately, downhole logging generally provides large, reliable datasets so that this editing procedure has little impact on the analysis.

Finally, the observational data should be standardised prior to the statistical analysis, by subtracting the mean and dividing by the standard deviation. The resulting logs are dimensionless, each with a mean of zero and a standard deviation of 1. This permits an equal-weighting comparison among all the observations, regardless of their original scaling.

The complete dataset, including the computed ratios of potassium/uranium (K/U) and thorium/potassium (Th/K), is shown in figure 1.

## FACTOR ANALYSIS

Factor analysis (FA) is a technique for examining the interrelationships among a set of observations. It is used to derive a subset of uncorrelated variables called factors that adequately explain the variance observed in the original observational dataset (Brown, 1998). Often such analysis reveals structure in the dataset by identifying which observations are most strongly correlated. Interpretation of these correlations contributes to understanding of the properties that are being measured and the underlying processes. A significant advantage of FA is that the number of variables can be dramatically reduced without losing important information. In other words, the dimensionality of the observational dataset can be reduced. Half a dozen or more interrelated variables might be reduced to perhaps two or three factors that account for nearly all the variance in the original dataset. Visualisation of two or three factors is much simpler than visualisation of the entire dataset.

Sometimes FA is confused with principal component analysis (PCA), but there is a significant difference between the two techniques. Strictly speaking, PCA is simply a mathematical manipulation involving the eigenvectors of the covariance or correlation matrix of the observations. Statistical considerations such as probability or hypothesis testing are not included in PCA (Davis, 1986). Often, though, PCA forms the starting point for FA. In FA, a series of assumptions is made regarding the nature of the parent population from which the samples (*i.e.*, observations) are derived. For example, the observations are assumed to follow a normal distribution. Such assumptions provide the rationale for the operations that are performed and the manner in which the results are interpreted (Davis, 1986).

Another way of explaining the difference between FA and PCA lies in the variance of variables (communality) that is analysed. Under FA, attempts are made to estimate and eliminate variance due to error and variance that is unique to each variable (Brown, 1998). Consequently, the FA result concentrates on variables with high communality values (Tabachnick & Fidell, 1989), *i.e.* only the variance that each variable shares with other observed variables is available for analysis and interpretation. In this investigation the FA method is used, because error and unique variances only obscure the picture of underlying processes and structures.

Factors and factor loadings were calculated from the rescaled logging curves using standard R-mode factor analysis procedures (Davis, 1986). A Kaiser Varimax factor rotation (Davis, 1986) is applied because the matrix of factor loadings is often not unique or easily explained. The factor rotation results in a simplification of the factor co-ordinate system. The technique of factor calculation is that of extraction of the eigenvalues and eigenvectors from either the correlation or covariance matrix. With appropriate assumptions, the factor model is simply a linear combination of underlying properties. A factor is taken as being significant for an underlying property if it accounts for a significant amount of variance, or in practical terms, if its eigenvalue is greater than 1. Factors with eigenvalues less than 1 account for less variation than one of the initial variables.

Theoretically, because they are maximally uncorrelated, each factor represents an underlying rock property such as porosity, lithology, grain size, fracture content, water content, or clay type. This is not strictly the case, in reality, since there is obviously no pre-condition that the rock properties will themselves be uncorrelated. Indeed, it is possible to envision highly non-linear interrelations between various rock properties like porosity, lithology, fracture content, fluid content, and clay type. As a first-order interpretation, though, FA provides an objective, rapid, and methodical approach for identifying major features of an observational dataset. Also, since many borehole geophysical tools were initially designed to respond primarily to porosity and lithology, Elek (1990) argued that the first two factors (*i.e.*, the two factors accounting for the highest degree of variance in the observations) derived from FA will also relate directly to porosity and lithology. This is a reasonable generalisation when the interaction between various rock properties is known to be relatively simple.

For the CRP-2 data set, more than 80% of the variance observed in the input variables can be described by the first three factors (Tab. 1). This means that the amount of explained variance is greater than 80% although the number of variables has been reduced from 11 to 3.

### CLUSTER ANALYSIS

After performing FA, statistical electrofacies are defined using cluster analysis. Clustering techniques are generally used for grouping individuals or samples into *a priori* unknown groups. The objective of the cluster analysis is to separate the groups based on measured characteristics with the aim of maximising the distance between groups. Hierarchical clustering methods yield a series of successive agglomerations of data points on the basis of successively coarser partitions. One of the most common methods of complete-linkage hierarchical clustering is the so-called Ward method (Davis 1986), which is also used in this study.

We use the three factor logs that accounted for the greatest amount of variance in the initial data set, rather than the 11 original logs, for the cluster analysis. Prior to applying the cluster analysis, the factor logs are reduced to a 0.5 m depth interval to reduce the number of data points. This step, although not essential, has two advantages. First, the cluster analysis calculations are very time consuming and require a massive amount of computer memory. Reducing the number of data points results in faster calculations. Second, this step was performed in order to get a clusterlog that does not show too many details, i.e. showing a new cluster every few centimetres. A complete-linkage hierarchical cluster analysis using a Euclidean norm („Ward-method“, see Davis, 1986) was performed on the three decimated factors. This allowed the identification of statistical electrofacies, or logging units, with distinct combinations of rock physical and chemical properties (*e.g.*, Serra, 1986). A dendrogram, a tree diagram showing similarity or connectivity between samples and clusters, is used to decide how many clusters are significant and useful. For the CRP-2 site, the number of significant clusters based on multivariate analysis of the three factor logs is 4. Taking into account more clusters would result into a subdivision of these most significant clusters and thus complicating an interpretation.

There are several commercial software packages that can be used to perform all the multivariate statistical methods described above. For this investigation we used WINSTAT 3.1 (Kalmia Software) and MVSP 3.0 (Kovach 1998) on a PC platform under Windows NT 4.0.

## APPLICATION OF MULTIVARIATE STATISTICAL METHODS

### FACTOR LOGS

One of the main advantages of the factor logs is that they are – by definition - independent of each other. This means that the ambiguity of downhole logs is strongly reduced and that they can be interpreted directly in terms of background controlling variables.

For the FA, the downhole logs of figure 1 were taken into account. The shallow resistivity log was not used because it correlates strongly with the deep resistivity log; its inclusion would weight resistivity too heavily compared to the remaining data. Deep resistivity was used rather than shallow resistivity because it is more likely to be representative of the undisturbed sediment away from the borehole.

The results of the factor analysis of the downhole logging data, along with factor eigenvalues and factor loadings, are listed in table 1. The factor logs are plotted in figure 3, together with the lithology column and the multivariate clusterlog. Factor loadings greater than $\pm$ 0.5 are taken as significant, shown in bold in Table 1, and flagged with plus or minus signs at the bottom of Figure 3. A plus sign represents a positive loading whereas a minus sign represents a negative loading of the corresponding variable. Three factors were extracted from the original data set, accounting for 82% of the total variance of the original dataset.

The factor analysis shows that the most discriminating variables are density, porosity, velocity, and gamma ray, each with a factor loading greater than 0.9 (Tab. 2). Gamma ray and susceptibility are mainly related to lithology and grain size (Rider 1996); in this case, GR is particularly related to clay type and clay content. Together with the thorium and potassium contents, and to a lesser extent the uranium content, these variables form the Factor2 log. All factor loadings of Factor2 are positive and greater than 0.8 (except the uranium content with a value 0.66); thus the underlying physical or chemical properties show a good positive correlation. Overlain over the Factor2 log in figure 3 is the silt content derived by Neumann & Ehrmann (this volume). As can be seen, there is a close correlation over the entire section between the experimentally derived grain sizes and the Factor2 log, leading to the conclusion that Factor2 is more or less a high resolution grain size log. A similar, but antithetical relation relates the sand content with Factor2. This means, that grain size is the background process for the physical and chemical properties summarised in Factor2 and is reflecting the lithology.

The ratio K/U, and to a smaller extent the ratio Th/K, is the main loading for Factor1; both factor loadings are greater than 0.75 and showing opposite signs. Thorium/potassium and potassium/uranium ratios often indicate clay type (Rider 1996, Jurado et al. 1997). Thus, Factor1 is probably related to changes in clay type and/or clay content. As illite has the highest potassium content among the different clay types (Rider 1996), borehole sections with high Factor1 values may indicate higher illite concentrations, whereas sections with low Factor1 values may be characteristic of a higher smectite content. But it should be mentioned that a complication for using the ratios Th/K and K/U as clay mineral indicators is that K-rich McMurdo volcanics and K-feldspar are often more abundant in the upper 300 m of the borehole than the clay

*Tab. 1* - Results of the factor analysis of downhole logs from CRP-2A. The number of valid cases reflects the number of data in each borehole log used for the analysis. The upper part of the table presents communalities, eigenvalues and amounts of explained variance. Eigenvalues are assumed to be important if they are greater than or equal to 1; these are indicated with an asterisk here, shown in Figure 3, and included in the subsequent cluster analysis. The lower part of the table gives the factor loadings, the communality, and the total amount of explained variance. Factor loadings greater than 0.5 are shown in bold. The sum of the factor loadings squared is equal to the eigenvalue, which is the variance explained by a factor. Three factors have an eigenvalue greater than 1. The total explained variance due to these three factors is 82%. (An explanation of variables is given in Figure 1, a "c" at the end of a variable name denotes „filled up by core measurements").

**Valid cases: 1216**

**COMMUNALITIES**

| | Communal. estimated | Communal. calcul. |
|---|---|---|
| GR | 1.0 | 0.96 |
| K | 1.0 | 0.96 |
| U | 1.0 | 0.80 |
| Th | 1.0 | 0.90 |
| Th/K | 1.0 | 0.64 |
| K/U | 1.0 | 0.83 |
| WBDc | 1.0 | 0.93 |
| Vpcorec | 1.0 | 0.86 |
| porocorec | 1.0 | 0.93 |
| log(suscorec) | 1.0 | 0.67 |
| log(Rlong) | 1.0 | 0.57 |

**EIGENVALUES:**

| Factor | Eigenvalue | Variance percent | Percentage (cumulative) |
|---|---|---|---|
| *1 | **5.25** | **47.7** | **47.7** |
| *2 | **1.98** | **18.0** | **65.7** |
| *3 | **1.78** | **16.2** | **82.0** |
| 4 | 0.72 | 6.56 | 88.5 |
| 5 | 0.53 | 4.86 | 93.4 |
| 6 | 0.44 | 4.04 | 97.4 |
| 7 | 0.17 | 1.52 | 98.9 |
| 8 | 0.08 | 0.77 | 99.7 |
| 9 | 0.01 | 0.17 | 99.9 |
| 10 | 0.01 | 0.11 | 99.9 |
| 11 | 0.00 | 0.03 | 100.0 |

**VARIMAX FACTORLOADINGS:**

| | Factors 3 | 2 | 1 | Commu-nality |
|---|---|---|---|---|
| WBDc | **0.95** | -0.13 | 0.04 | 0.92 |
| porocorec | **-0.95** | 0.13 | -0.04 | 0.92 |
| Vpcorec | **0.91** | -0.17 | -0.06 | 0.85 |
| log(Rlong) | **0.69** | -0.27 | -0.16 | 0.57 |
| GR | -0.38 | **0.90** | -0.05 | 0.95 |
| Th | -0.35 | **0.87** | -0.13 | 0.89 |
| K | -0.34 | **0.85** | 0.33 | 0.95 |
| log(suscorec) | 0.12 | **0.80** | 0.07 | 0.66 |
| U | -0.28 | **0.66** | **-0.53** | 0.79 |
| K/U | -0.11 | 0.24 | **0.87** | 0.82 |
| Th/K | 0.02 | 0.13 | **-0.78** | 0.64 |
| Sum of squares | 3.61 | 3.59 | 1.82 | 9.02 |
| Percentage of variance | **32.80** | **32.60** | **16.56** | **81.96** |

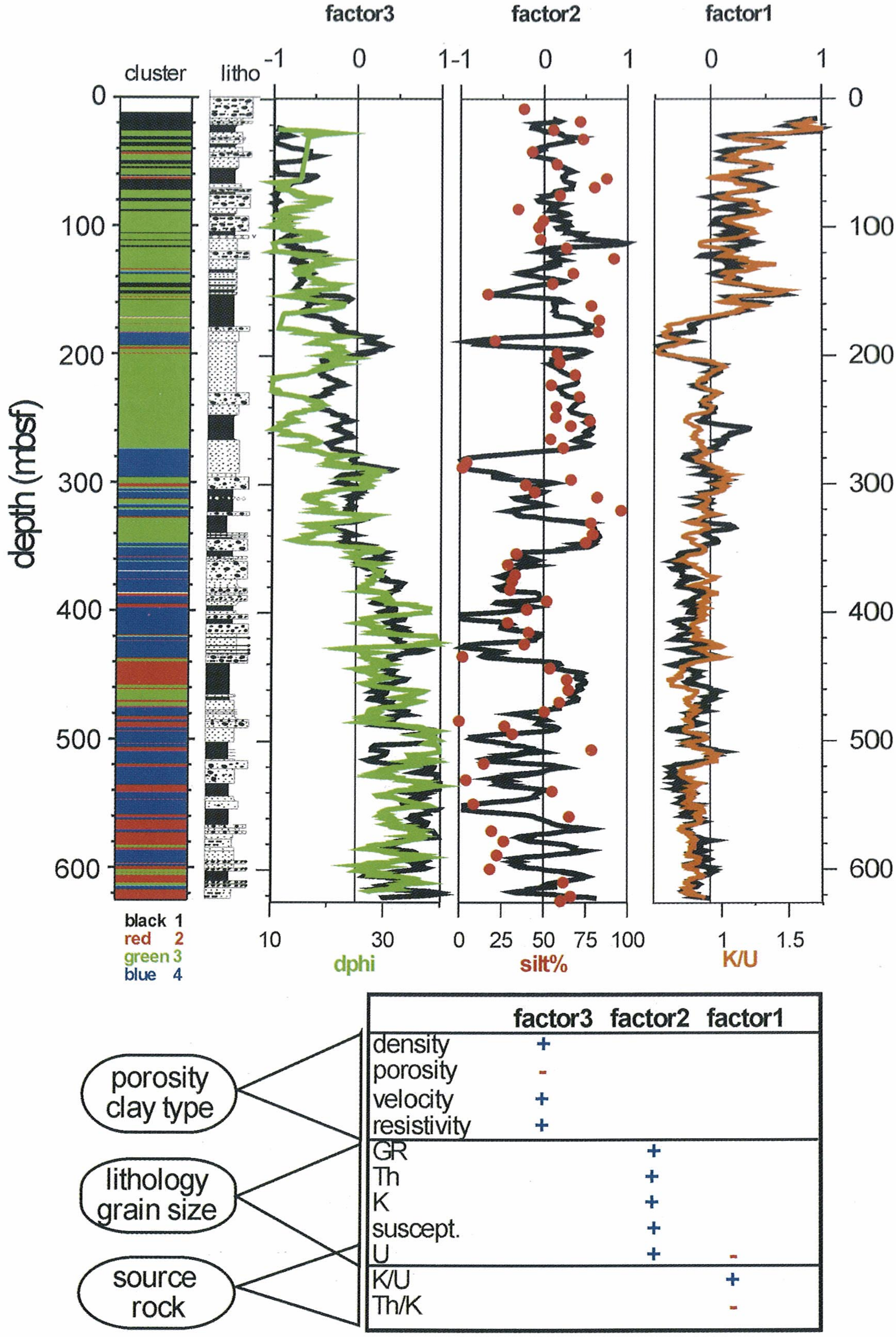



*Fig. 3* - Clusterlog, core-derived lithology, and factor logs. Factor3, Factor2, and Factor1 are the factor logs as derived by factor analysis. Based on these three factor logs, the multivariate cluster log (left column) was calculated. In the bottom part of the diagram, factor loadings with values greater than 0.5 are shown simplified as plus or minus signs. The Factor2 log is mainly related to grain size (high loading of GR, Th, K, susceptibility, and to a lesser extent uranium), whereas the Factor3 log is related to true porosity (high loading of porosity, density, sonic velocity, and resistivity). The green curve overlain over Factor3 shows the difference between core porosity and neutron porosity (dphi) and thus is pointing to differences in clay type/content. The red dots overlain over Factor2 represent measurements of silt content by Ehrmann (this volume). The Factor1 log (loaded by the ratios Th/K and K/U, and negatively to a small extent also by uranium content) is indicative of sediment source. The multivariate clusterlog (left column) shows clear differences above and below 300 mbsf. Above 300 mbsf, green and black colours are dominating, whereas below this depth blue and red colours occur often. As the clusters reflect different physical properties (*cf.* Tab. 2) and the lithology is not changing systematically, this behavior is pointing to a major change in source regions at 300 mbsf.

*Tab. 2* - Back-calculated physical properties of clusters of the multivariate cluster analysis shown in figure 3. Bold numbers represent the average value in each cluster, and numbers in parentheses give the corresponding standard deviation (units see Fig. 1). Each cluster has a different combination of physical properties. For example, clusters 1 and 3 have identical GR values, but they have significantly different K/U and Th/K ratios, pointing to different rock sources. (An explanation of variables is given in Fig. 1).

| cluster # | *GR* | *Th* | *K* | *U* | *sus* | *WBD* | *poro* | *Vp* | *Rlong* | *K/U* | *Th/K* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **1 (black)** | **99**(18) | **7.8**(1.8) | **2.7**(0.4) | **2.0**(0.6) | **234**(92) | **2.09**(0.19) | **36**(13) | **2.13**(0.75) | **2.6**(1.1) | **1.4**(0.3) | **2.9**(0.5) |
| **2 (red)** | **68**(19) | **5.7**(1.7) | **1.4**(0.3) | **2.0**(0.9) | **132**(137) | **2.32**(0.20) | **23**(12) | **3.11**(0.91) | **10.0**(20.1) | **0.8**(0.2) | **3.9**(0.7) |
| **3 (green)** | **98**(11) | **7.9**(1.3) | **2.3**(0.3) | **2.6**(0.5) | **183**(101) | **2.17**(0.15) | **32**(9) | **2.34**(0.46) | **3.3**(3.0) | **0.9**(0.2) | **3.5**(0.6) |
| **4 (blue)** | **50**(13) | **3.8**(1.1) | **1.2**(0.3) | **1.5**(0.5) | **76**(107) | **2.26**(0.14) | **26**(8) | **2.57**(0.53) | **5.1**(5.0) | **0.7**(0.2) | **3.3**(0.6) |

minerals. This may lead to the alternative interpretation that Factor1 is reflecting the source rock region by changing ratios of the radiogenic elements. In the upper part of the borehole down to 300 m, both ratios are negatively correlated (Fig. 1), this antithetical relation is lost below 300 mbsf. This behavior is confirming the assumption, that Factor1 is more related to sediment provenance than to clay type. Thus the background process reflected by Factor1 is the sediment source.The physical properties density and porosity show the highest loadings for Factor3 with values of 0.95. Velocity and resistivity are closely related to porosity and they are also contributing to Factor3. As expected, the signs of the factor loadings for den and phi are opposite. Thus, Factor3 is mainly responding to the porosity of the formation and shows the combined effect of porosity on density, velocity,and electrical resistivuty. Overlain over Factor3 is the dphi-curve, the difference between core porosity and neutron porosity. As mentioned earlier this dphi-curve should reflect porosity-free clay-type/content effects. Compared to the results from Ehrmann on clay minerals (this volume), the dphi curve shows a very similar behavior as the curve for smectite content. With increasing depth, dphi and smectite are increasing, the 350 mbsf depth marks a sharp change in both properties.

## CLUSTERLOGS

The multivariate clusterlog, based on all three factor logs, is shown in color at the left side of figure 3, together with the lithological column. The mean and standard deviation of the physical properties for each cluster are given in table 2. Each cluster represents intervals where the physical and chemical rock properties are presumably similar. Four or six significant clusters could be derived by dendrogram evaluation; for clarity, only the four-cluster solution is shown in figure 2. Each cluster in the cluster log can be seen as a statistically determined electrofacies (or petrofacies) as defined by Serra (1984). This clustering facilitates subdivision of the borehole into logging units that can be compared to lithology, porosity, grain size, or provenance.

As mentioned above, Factor2 mainly reflects grain size and thus lithological changes. For direct comparison of Factor2 with the lithology column, a univariate clusterlog was calculated, based solely on Factor2 (Fig. 4). The backcalculated physical properties within each cluster of this univariate cluster analysis are given in table 3. As can be seen in table 3, Factor2 is clearly differentiating the 4 clusters. The average silt content reflected by the clusters is demonstrated in the box and whisker plot in figure 5. Cluster 4 is characterised by the lowest silt content coinciding with lowest gamma ray properties and also lowest density, velocity, and resistivity values, obviously representing diamictites. The highest silt content is reflected by cluster 3, the physical properties (Tab. 3) show high gamma ray as well as high velocity and susceptibility values, and a high Th/K ratio, pointing to a mudstone. In figure 4, different colours in the univariate clusterlog are directly related to different silt contents and thus to different grain sizes. Coarse grained sections (low silt contents) are graphically enhanced by the horizontal bars, which are based on the blue and black colours of the clusterlog. Agreement between the litho log and clusterlog is best in the lower part of the borehole, below 300 mbsf. In the uppermost section of the borehole (above 150 mbsf), sand and diamictite seem to be overestimated in the lithology log.

## RESULTS AND DISCUSSION

By means of factor and cluster analysis, it was possible to reduce the dimensionality of the CRP-2A downhole logging data without significant loss of information. The resulting set of factor and cluster logs makes subsequent evaluations and interpretations much easier. The analyses resulted in three factor logs. We conclude that Factor3 is a good proxy for true overall porosity, and Factor2 is a good proxy for lithology and grain-size variations. Factor1 contains information primarily related to sediment source and, to a lesser extent, clay type. Clay type and clay content is also reflected by Factor3 because it shows a close correlation to the difference between core porosity and neutron porosity (dphi). As stated earlier, the factor logs should be independent of each other by definition. But obviously the rock properties themselves are not

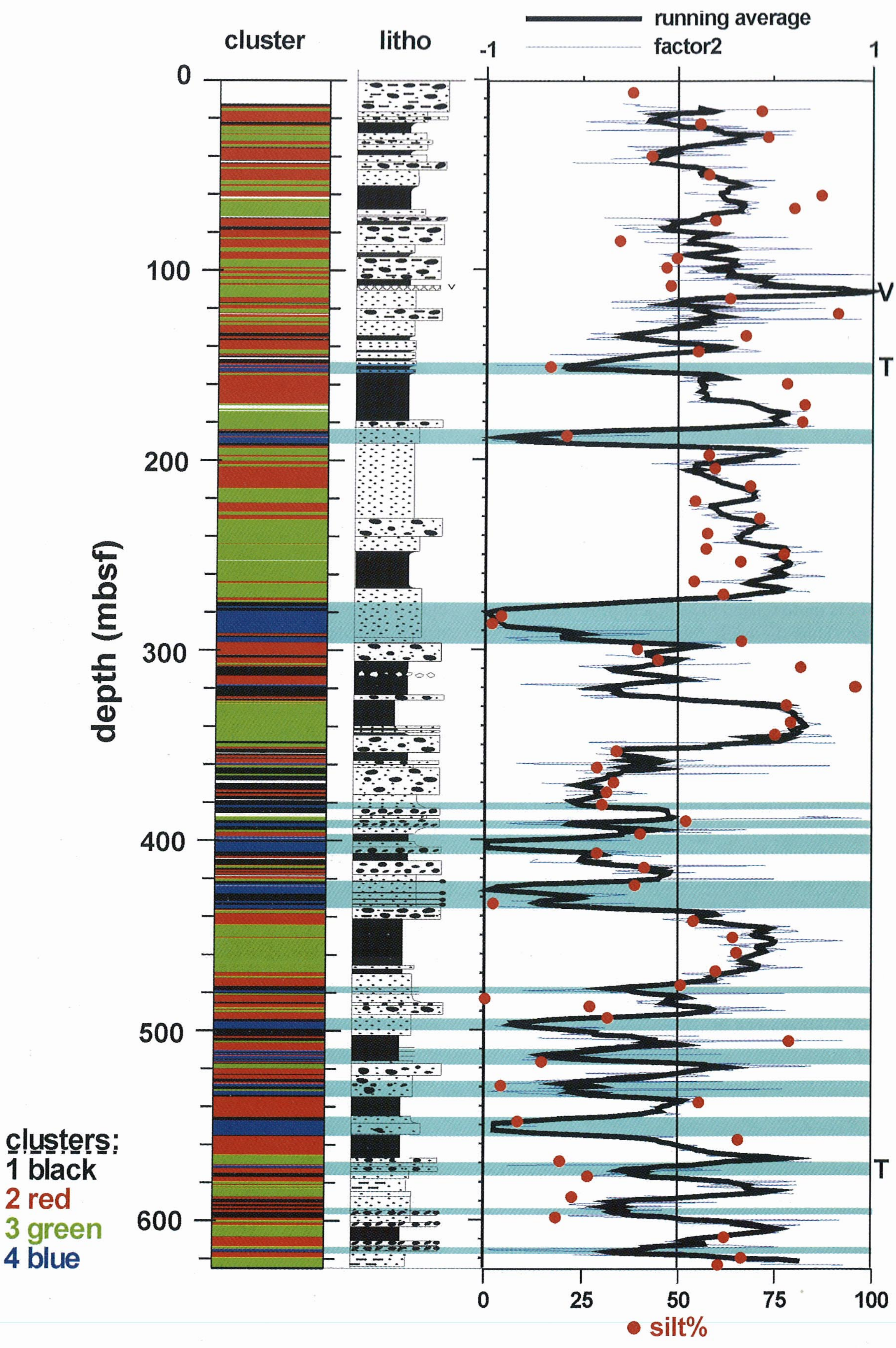



*Fig. 4* - Factor2 log with detailed interpretation. The left column shows a univariate clusterlog that was calculated solely by using the Factor2 log. Note that the colours shown here do not correspond to those shown in figure 3. Because the Factor2 log responds mainly to grain-size changes, this univariate clusterlog represents different grain sizes by different colours. Grain size decreases from blue over black and red to green. Accordingly, blue and black represent diamictites and coarse sandstones, whereas red and green represent siltstones and mudstones. The higher proportion of coarse grained sediments in the lower part of the borehole is optically enhanced by the horizontal bars, which are based on the blue and black colours of the univariate clusterlog. The „V" at the right margin denotes a massive volcanic ash layer, a „T" denotes Temperature anomalies detected by downhole measurements (*cf.* Bücker et al., this volume).

*Tab. 3* - Back-calculated physical properties of clusters of the univariate cluster analysis shown in figure 4. This univariate analysis contrasts with the multivariate analysis of Table 2. Bold numbers represent the average value in each cluster, and numbers in parentheses give the corresponding standard deviation (units see Fig. 1). Each cluster has a different combination of physical properties. Due to the relation between Factor2 and silt content (*cf.* Fig. 4), grain size is decreasing from cluster 4 to clusters 1 and 2 to cluster 3. This means that cluster 4 generally represents diamictites whereas cluster 3 consists mainly of mudstones. (An explanation of variables is given in Fig. 1).

| *cluster #* | *GR* | *Th* | *K* | *U* | *sus* | *WBD* | *poro* | *Vp* | *Rlong* | *K/U* | *Th/K* | *factor2* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1 (black)** | **59**(16) | **4.6**(1.3) | **1.4**(0.5) | **1.6**(0.4) | **103**(122) | **2.25**(0.14) | **27**(8) | **2.64**(0.58) | **5.8**(6.9) | **0.92**(0.28) | **3.3**(0.6) | **-0,4** |
| **2 (red)** | **82**(21) | **6.6**(1.7) | **1.9**(0.6) | **2.1**(0.6) | **179**(116) | **2.25**(0.17) | **27**(10) | **2.63**(0.76) | **6.8**(14.9) | **0.95**(0.32) | **3.5**(0.6) | **0** |
| **3 (green)** | **96**(19) | **8.1**(1.7) | **2.1**(0.5) | **2.6**(0.8) | **179**(128) | **2.20**(0.26) | **30**(16) | **2.74**(0.99) | **6.0**(16.4) | **0.84**(0.28) | **3.9**(0.7) | **0,5** |
| **4 (blue)** | **41**(10) | **3.0**(0.8) | **1.0**(0.3) | **1.2**(0.3) | **39**(43) | **2.15**(0.12) | **33**(17) | **2.20**(0.41) | **3.2**(1.8) | **0.88**(0.25) | **3.0**(0.6) | **-1** |

uncorrelated in this case, resulting in a correlation coefficient of –0.56 between Factor1 and Factor3, which means that these factors are sharing some background information.

The chemical and sedimentological core measurements from CRP-2 can be used to calibrate the results from the statistical evaluations of downhole logs. This step, which enhances quantitative interpretation of the factor logs and clusters, is beyond the scope of the present study; it will be followed up in further studies.

Based on the multivariate clusterlog (Fig. 3), which is taking into account all downhole logs of Figure 1, the hole can be divided into the following 10 logging units or statistical electrofacies (corresponding cluster values are given in table 2):

*Unit 1:* 0 - 40 mbsf, mainly *cluster 1*
*Unit 2:* 40 - 158 mbsf, mainly *cluster 3*, with contributions by cluster 1
*Unit 3:* 158 - 200 mbsf, mainly *cluster 3*, with contributions by clusters 4 and 2
*Unit 4:* 200 - 275 mbsf, mainly *cluster 3*
*Unit 5:* 275 - 327 mbsf, mainly *cluster 4*, with contributions by cluster 3
*Unit 6:* 327 - 350 mbsf, mainly *cluster 3*
*Unit 7:* 350 - 445 mbsf, mainly *cluster 4*, with contributions by cluster 2
*Unit 8:* 445 - 492 mbsf, mainly *cluster 2*, with contributions by clusters 4 and 3
*Unit 9:* 492 - 560 mbsf, mainly *cluster 4*, with contributions by cluster 2
*Unit 10:* 560 - 624 mbsf, mainly *cluster 2*, with contributions by clusters 4 and 3.

This cluster analysis is a helpful tool for reliable, reproducible, and objective definition of logging units. Brink et al. (this volume) determined and described 8 log-based units due to changes in two or more of the downhole logs. Though not clearly visible by the logging unit definition, it must be stated that the 300 mbsf unconformity is a main unconformity which is marked by a suite of parameters showing dramatic changes at this depth. These parameters are the clay mineralogy (Ehrmann, this volume), the dip angle in seismic lines (Henrys et al., this volume), the lithology of clasts (Passchier, this volume), the petrology (Talarico et al., this volume), and the average diatom abundance (Scherer et al., this volume). This overall change in physical, sedimentological, and structural properties, as it is related to a change in logging properties, is summarised in the multivariate cluster log. Thus the multivariate cluster log shows not only changes in lithology and grain size but also changes in provenance and clay type. A further subdivision of the clusters may show more subtle changes.

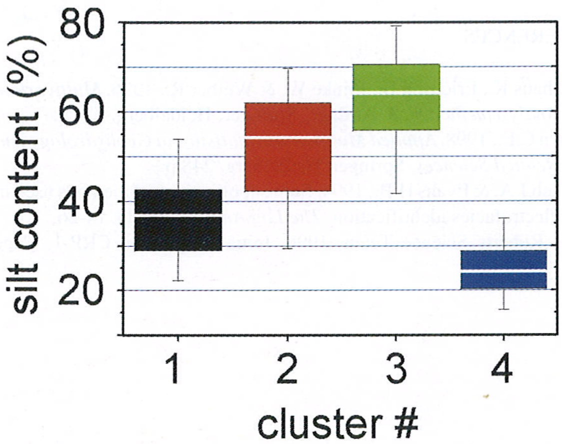



*Fig. 5* - Box-and-whisker plot of silt content according to the four clusters of the univariate cluster analysis shown in figure 4. This plot demonstrates that cluster 4 (blue) represents the lowest silt content and cluster 3 (green) indicates the highest silt content. Sand content exhibits the opposite behavior.

Our statistically derived logging units show a fairly good agreement with the subjectively derived logging units of Brink et al. (this volume), providing an independent confirmation of the overall sedimentary structure of CRP-2A. The good correlation between this multivariate clusterlog and the petrofacies log by Smellie (this volume) confirms that the multivariate clusterlog is reflecting petrofacies and thus provenance changes.

The factor and cluster analyses identify and emphasise different aspects of CRP-2A physical and chemical properties. The factor analysis gives factor logs that provide continuous records of downhole variations in porosity, lithology, grain size, and clay type. The cluster analysis, in contrast, identifies discrete logging units which indicate downhole changes in source region provenance.

ACKNOWLEDGEMENTS

This work was supported by the Alfred Wegener Institut für Polar- und Meeresforschung, (AWI), the Bundesanstalt für Geowissenschaften und Rohstoffe (BGR), and the German Science Foundation DFG (WO 672/1-2). Special thanks go to Peter Barrett and Franz Tessensohn for their constant encouragement and helpful discussions. We would like to thank Frank Niessen (Alfred Wegener Institute, Germany) for putting the dataset of core physical properties at our disposal. The critical reviews by Rob Funnell and Stuart Henrys were very helpful improving the manuscript. Borehole measurements and their logistics are never simple in these „icy" environments, and we would like to thank everyone who helped make these measurements successful: thanks to Pat Cooper, John Alexander, Jim Cowie, and many others at Scott Base and Crary Lab McMurdo.

REFERENCES

Backhaus K., Erichson B., Plinke W. & Weiber R., 1996. *Multivariate Analysemethoden*, **8**. Auflage, Springer, Heidelberg, 416 p.

Brown C.E., 1998. *Applied Multivariate Statistics in Geohydrology and Related Sciences*, Springer, Heidelberg, 248 p.

Bucheb J.A. & Evans H.B., 1994. Some applications of methods used in electrofacies identification. *The Log Analyst*, **35** (1), 14-26.

Cape Roberts Science Team, 1998. Initial Report on CRP-1, Cape Roberts Project, Antarctica. *Terra Antartica*, **5**(1), 187 p.

Cape Roberts Science Team, 1999. Studies from the Cape Roberts Project, Ross Sea, Antarctica, Initial Report on CRP-2/2A. *Terra Antartica*, **6**(1/2), 173 p.

Davis J.C., 1986. *Statistics and Data Analysis in Geology, 2nd edn*. Wiley, New York, 646 pp.

Doveton J.D., 1994. Geologic Log Analysis Using Computer Methods, *Am. Assoc. Petrol. Geol. Computer Applications in Geology* **2**, 169 pp.

Elek I., 1990. Fast porosity estimation by principal component analysis. *Geobyte*, **5**(3), 25-34.

Howarth, R.J. & Sinding-Larsen R., 1983. Multivariate analysis. In: Howarth R.J. (ed.), *Statistics and Data Analysis in Geochemical Prospecting, Vol. 2*, Elsevier, 207-283.

Jurado M.J., Moore J.C. & Goldberg D., 1997. Comparative logging results in clay-rich lithologies on the Barbados Ridge. In: Shipley, T.H., Ogawa Y., Blum P. & Bahr J.M. (eds.). *Proc. ODP Sci. Results*, **156**, College Station TX (Ocean Drilling Program).

Kovach W.L., 1998. *MVSP – A Multivariate Statistical Package for Windows, ver. 3.0*. Kovach Computing Services, Pentraeth, Wales, U.K.

Rider M.H., 1996. *The Geological Interpretation of Well Logs*. Whittles Publishing, Caithness, 280 p.

Serra O., 1986. *Fundamentals of Well-Log Interpretation (Vol. 2): The Interpretation of Logging Data*. Dev. Pet. Sci., 15B, 371 p.

Tabachnick B.G. & Fidell L.S., 1989. *Using Multivariate Statistics*. Harper and Row, San Francisco, 746 p.

Theys P.T., 1991. *Log Data Acquisition and Quality Control*. editions technip, Paris, 330 p.

Weaver C.E. & Pollard L.D., 1973. *The Chemistry of Clay Minerals (Developments in Sedimentology)*. Elsevier, Amsterdam, 289 p.

Western Atlas, 1992. Introduction to Wireline Log Analysis.- Western Atlas International, Inc., Houston, Texas.