Geophys. J. Int. (1999) 136, 477–491

Inferring the lithology of borehole rocks by applying neural network classifiers to downhole logs: an example from the Ocean Drilling Program

D. Benaouda,[1] G. Wadge,[1] R. B. Whitmarsh,[2] R. G. Rothwell[2] and C. MacLeod[3]

[1] *Environmental Systems Science Centre, University of Reading, Harry Pitt Building,* PO Box 238, *Reading,* RG6 6AL, *UK.*
E-mail: gw@mail.nerc-essc.ac.uk
[2] *Southampton Oceanography Centre, University of Southampton, European Way, Southampton,* SO14 3ZH, *UK*
[3] *Department of Earth Sciences, University of Wales College of Cardiff,* PO Box 914, *Cardiff,* CF1 3YE, *UK*

Accepted 1998 September 4. Received 1998 June 15; in original form 1997 August 12

SUMMARY

In boreholes with partial or no core recovery, interpretations of lithology in the remainder of the hole are routinely attempted using data from downhole geophysical sensors. We present a practical neural net-based technique that greatly enhances lithological interpretation in holes with partial core recovery by using downhole data to train classifiers to give a global classification scheme for those parts of the borehole for which no core was retrieved. We describe the system and its underlying methods of data exploration, selection and classification, and present a typical example of the system in use. Although the technique is equally applicable to oil industry boreholes, we apply it here to an Ocean Drilling Program (ODP) borehole (Hole 792*E*, Izu-Bonin forearc, a mixture of volcaniclastic sandstones, conglomerates and claystones). The quantitative benefits of quality-control measures and different subsampling strategies are shown. Direct comparisons between a number of discriminant analysis methods and the use of neural networks with back-propagation of error are presented. The neural networks perform better than the discriminant analysis techniques both in terms of performance rates with test data sets (2–3 per cent better) and in qualitative correlation with non-depth-matched core. We illustrate with the Hole 792*E* data how vital it is to have a system that permits the number and membership of training classes to be changed as analysis proceeds. The initial classification for Hole 792*E* evolved from a five-class to a three-class and then to a four-class scheme with resultant classification performance rates for the back-propagation neural network method of 83, 84 and 93 per cent respectively.

Key words: artificial intelligence, borehole geophysics, drill cores, ocean drilling, sediments, statistical methods.

INTRODUCTION

In boreholes with little or no recovery of core it is possible to infer the nature of the rocks surrounding the hole using log data from downhole geophysical sensors. These log measurements correspond both to parts of the hole from which core has been recovered and to those parts of the hole from which there is no core recovery. By using these measurements as surrogates for the missing core, we can fill in, at least partly, the missing information and improve our understanding of the geological stratigraphy of the hole, one of the goals of petrophysical well log interpretation for many years.

The Ocean Drilling Program (ODP) is a long-term, international scientific endeavour to explore the floors of the world's oceans by drilling. The principal source of information comes from core recovered from the drilled hole. Often core recovery is only partial (less than 50 per cent) and consequently the scientific returns are reduced. Generally, a far greater diversity of lithologies and borehole environments are met in the ODP than in petroleum reservoir boreholes. This requires an approach that is more flexible than simple matching to a library of log responses. In this paper we describe how to address this problem using supervised statistical classification techniques, training the classifiers on those parts of the hole with both logs and core and then applying the classifiers to those parts of the hole lacking core. Supervised classification techniques are used routinely in geology (Davis 1986) and in downhole logging (Doveton 1986, 1994). Here we use both discriminant analysis and neural network techniques. The similarities of the two approaches have been analysed by

Gallinari *et al.* (1991) and Cheng & Titterington (1994). Neural network techniques have been applied successfully to logging problems (e.g. Busch *et al.* 1987; Baldwin *et al.* 1990; Rogers *et al.* 1992; Wong *et al.* 1995) including ODP data (Goncalves 1995; Chang *et al.* 1997). One of the principal aims is to show fully comparable results from both discriminant analysis and neural network techniques but our approach here is from the practical perspective of how to solve the problem, specific to the ODP context. Our methods and how they are implemented with regard to problem definition and data selection are described in detail. Another of our aims is to demonstrate the need to refine the choice of classes based on an analysis of the data. We then apply these techniques to a data set from ODP Hole 792*E* drilled in 1989 in the northwest Pacific Ocean in which we use about 50 per cent of the hole for training and testing the networks and describe the results of applying our classification techniques to the rest of the hole.

METHODS AND IMPLEMENTATION

We use a generic method to solve the problem of lithological classification of ODP holes from integrated core-log data. We break down the method into three steps: data exploration, data selection and classification. The purpose of data exploration is to understand the limits and dimensionality of the log and core data and to assess whether the problem of assigning uncored parts of the hole to the assumed classes is reasonable. We use principal components analysis and *K*-means cluster analysis to illustrate this. Data selection involves the application of quality control to both the log and the core data and the partitioning of log data for the training and testing parts of the supervised classification process. We then use two classification techniques, discriminant analysis and feed-forward neural networks, to produce the results after training and testing.

We implemented these methods on a custom-designed computer system (Wadge *et al.* 1998). Although the methods below are described sequentially, the system is flexible enough to allow the user to refine iteratively the definition of the problem and the selection of the logs, classes and data.

Statistical data exploration

The aim of statistical data exploration is to understand the structure of the data in terms of the variance properties of the logs and the number of lithological classes those data might support.

Principal components analysis (PCA)

In most sets of downhole logs, measurements made by different sensors are correlated with each other, sometimes highly so. Thus one approach is to transform the original set of data variables into a smaller set that are mutually uncorrelated. Consequently, the usual goal of PCA is to eliminate redundancies from a data set of correlated variables. It may therefore be worthwhile to pre-process the data set by PCA before using other techniques such as clustering or classification. One important decision is how many principal components should be retained in such an analysis in order to provide an adequate representation of the data set. The most widely used criterion is that only those principal components whose associated eigenvalues are either greater than 1 (Kaiser 1960) or greater than 0.7 (Jolliffe 1972) are retained. Although the first few principal components may provide a very useful summary of a data set, all the original variables are needed to compute their scores. Jolliffe (1972, 1973) has suggested a way to select a subset of variables based on PCA that contains all the information from all sets of the original data. Basically, we select p variables (or logs in our case) that are associated with each of the first p components that have the largest absolute value of the coefficient.

Cluster analysis (CA)

If we have no geological knowledge to guide a classification scheme (as when core recovery is low) then we want to know how many 'natural' classes are represented in the log data. Two such clustering strategies are available: *hierarchical clustering analysis*, which uses an iterative clustering algorithm based on similarity or dissimilarity measurements between pairs of samples, and *non-hierarchical clustering analysis*. In our case-study we have used a *K-means clustering algorithm* of the latter type, which is described by Hartigan (1975) and Hartigan & Wong (1979).

The K-means clustering algorithm groups n samples measured on m variables into K ($K \geq 2$) clusters in such a way that the within-cluster sum of squares is minimized, based on the Euclidean distance. Thus, the data need to be scaled since the variables are required to be mutually uncorrelated (for instance, PCA may be applied). The algorithm searches iteratively for a K-partition with a locally optimal within-cluster sum of squares by displacing samples from one cluster to another. One should choose initial seeds (initial cluster centres) that are sufficiently good so that few iterations are required (Milligan 1980). However, we have used the initial cluster centres which were suggested by Hartigan & Wong (1979). When the algorithm has converged, the coordinates of the cluster centres, the cluster membership, and residual sums of squares [RSS(K)] for the solution involving K clusters can be analysed. Hence, the sum-of-squares-error function decreases when the number of clusters K is increased. If, for instance, a data structure with n samples has well-separated $\hat{K}$ clusters, one would want to see that the above criterion function decreases sharply until $K = \hat{K}$, and decreases slowly when $K > \hat{K}$ until it reaches zero at $K = n$. More formally, we have used an approximate *Fisher* test (F-ratio) of significance for the appropriate number of clusters (Sparks 1973), which expresses how well a given K-cluster description fits the data. To do this, we re-run the algorithm for the solution involving $(K+1)$ clusters. Hence, the F-ratio test of the null hypothesis H_0: *the solution for* $K+1$ *clusters provides no better fit than the solution for* K *clusters*. The pseudo-F-ratio statistic is computed as

$$F_{K+1,K} = \frac{\mathrm{RSS}(K) - \mathrm{RSS}(K+1)}{\mathrm{RSS}(K+1)} \Big/ \left\{ \left[\frac{n-K}{n-(K+1)} \sqrt[m]{\left(\frac{K+1}{K}\right)^2} \right] - 1 \right\}, \tag{1}$$

with $\tau_1 = m$ and $\tau_2 = m[n-(K+1)]$ degrees of freedom, and where the null hypothesis H_0 is rejected if the F-ratio statistic exceeds the tabled value $F_{\tau_1,\tau_2,\alpha}$ at some level of significance α.

Data selection

Knowing the dimensionality (m) of the log data and the number of classes (K) to which they might be allocated helps the geologist to pose the correct classification problem. Equally important is the need to choose the best data to solve that problem. The vertical interval of the borehole, the number and type of logs and the classes must be selected. The vertical interval is usually constrained by the availability of data or by the geologically defined limit of the classification problem (for example, the boundary between the sedimentary succession and igneous basement). Wadge *et al.* (1998) showed that with ODP logs, neural network classifiers give generally better performance as the number of logs increases. Beyond an initial judicious choice of logs it was found not to be worth attempting to reduce the number of logs or replacing them with principal components to improve computing time. The quality of the log data is important. Isolated outlier values or data from sections of the hole where there is some reason to suspect spurious values (such as a greatly enlarged borehole) are searched for and removed. How many and which classes to use is largely a matter for expert judgement after exploration of log data which has been matched by depth to the cores. Many of the individual cores are shorter than a 9.7 m cored interval and therefore contain an incomplete record of the stratigraphy penetrated. In such cores we cannot assign the recovered material to an absolute depth interval with any accuracy and hence match it with the digital log values. Statistically, the best estimate of the position should be given by *Euler's Beta* distribution (Agrinier & Agrinier 1994) but in practice core tends to be recovered preferentially from the upper part of the cored interval. Hence, for neural network training and testing purposes, we only use cores that have greater than 90 per cent recovery and then linearly normalize (stretch) their depth assignments to 100 per cent to match the logs.

The matched core-log data sets are split into two populations, one for training the classifiers and one for testing their performance. Two ways of doing this are explored. In the first we take all available samples of each class and assign them alternately in sequence down the hole to arrive at two whole-class-population sample sets of equal size. In the second we take the size of the smallest population of any class and choose the same number of samples (appropriately evenly spaced) from all the other classes. This subset is then split into two smallest-class-population sample sets of equal size. The potential benefit of the second approach is that it should avoid any bias caused by the larger population classes when training the neural network; the danger is that the complete data distribution of the larger classes may become under-represented.

Classification

There are two main types of classifier suitable for our current task of assigning lithological classes to the ODP data: discriminant analysis and the feed-forward neural network. They are both supervised classifiers and we now describe their nature and how they can be applied.

Discriminant analysis (DA)

The task of discriminant analysis is to find the class, among those currently available, that most closely matches each unclassified sample. Thus the discriminant analysis method requires a prior knowledge of classes, whereas clustering constructs such a classification. In the simple case with two classes (Fisher 1936) involving measurements on two variables, the purpose of discriminant analysis is to find the line through the parameter space which maximizes the separation between the centres of the two classes; any new samples can then be assigned to their most likely class by finding their projections onto this line, the discriminant function, and the projections are called discriminant function scores, as illustrated in Fig. 1. Many data sets have more than two classes. The two-class discriminant-function analysis can still be used by analysing separately each pair of classes, but the number of pairs increases with the number of classes found in the data. Multiple discriminant analysis was developed by Fisher (1938), Rao (1952) and Rao & Slater (1949), providing a simultaneous comparison of several classes. There are two main groups of methods, *geometrical* and *probabilistic*, that define the classes and allocate new samples to those classes, respectively.

Geometrical methods. These consist of computing discriminant functions, which are linear combinations of variables that best separate the classes. In a K-class problem, $(K-1)$ discriminant functions are necessary to separate the classes. As for principal components analysis, fewer than $(K-1)$ discriminant functions may suffice if they represent a higher proportion of the total interclass variability. Subsequently, the multiple discriminant analysis aims for a reduction of multiple measurements to one or more weighted combinations providing maximum average separation among the classes. Thus, the solution of the multiple discriminant-function problem consists of maximizing a certain cost function or a criterion function J; for example, Devijver & Kittler (1982) have proposed

$$J(A)=\frac{|A^{t}S_{B}A|}{|A^{t}S_{W}A|}, \tag{2}$$

where $|S|$ denotes the determinant of a matrix S. S_B and S_W are respectively the between- and within-classes covariance matrices.

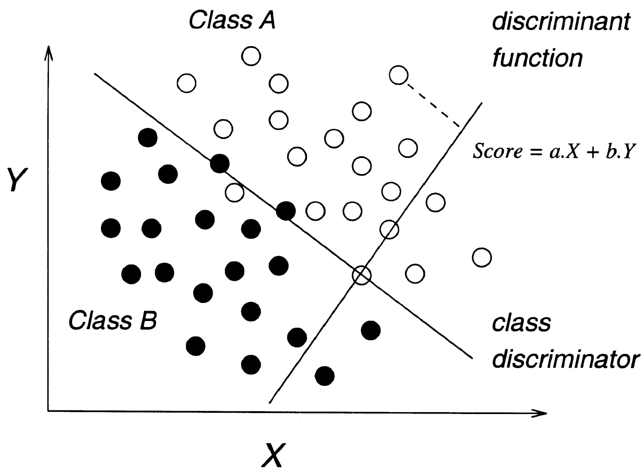

Figure 1. Graph of discriminant function analysis applied to two-variable data (X, Y). Each new sample projects perpendicularly onto the discriminant function line.

From Healy (1986), the solution is given by solving the eigenvalue problem as

$$(S_B - \lambda S_W)a = (S_W^{-1} S_B - \lambda I)a = 0\,. \tag{3}$$

Again, similar to principal components analysis, one may wish to determine how many significant discriminant functions with associated variances $(\lambda_1, \ldots, \lambda_p)$; $p = \min\{K-1, m\}$ there are. This is done by using a statistical *chi square* test of the null hypothesis $H_0: \mu_1 = \mu_2 = \ldots = \mu_K$ (where μ_i is the m-variate mean of the ith class).

Probabilistic methods. Here we are concerned with decision methods for assigning each new sample to one of the several classes with the minimum probability of error. Let us suppose $(p_1 = n_1/n, \ldots, p_K = n_K/n)$ are respectively *a priori* probabilities (where n_i is the number of elements in the ith class and n is the total number of elements in the data) and the probability distribution of a sample $x = (x_1, \ldots, x_m)$ is known for each class j $(j = 1, \ldots, K)$, denoted as $f_j(x)$. The probability of x belonging to the jth class is given by *Bayes' rule* as

$$P(j|x) = \frac{p_j f_j(x)}{\sum_{i=1}^{K} p_i f_i(x)}\,. \tag{4}$$

Thus, *Bayes' rule* is used to assign the sample x to a class that has the largest posterior probability, greater than a certain threshold value, computed by eq. (4). It remains then to know or estimate probability densities $f_j(.)$; $j = 1, \ldots, K$. Either a *parametric method* based on a multivariate normal distribution can be used, or a *non-parametric method* based on the observed samples themselves can be used.

There are a variety of distributions and kernels that could be chosen for our problem. In the case of the parametric method we have used a multivariate normal distribution with an equal variance–covariance matrix (referred to as normal linear) and an unequal one (referred to as normal quadratic). In the other case, the nonparametric method, we have used *normal and Epanechnikov kernels* to estimate non-parametric density (Duda & Hart 1973; Silverman 1986). We have used equal and unequal bandwidth samplings for each kernel.

Feed-forward neural network

Artificial neural networks are computer algorithms whose aim is to model the operation of the human neuron. A wide range of networks have been developed for applications such as classification, pattern recognition and optimization problems (Haykin 1994; Bishop 1995). For classification problems the *back-propagation learning algorithm* is the most popular (LeCun 1985; Rumelhart *et al.* 1986). Geological applications, particularly in downhole log analysis have shown good results (see Busch *et al.* 1987; Baldwin *et al.* 1990; Rogers *et al.* 1992; Wong *et al.* 1995; Goncalves 1995; Wadge *et al.* 1998; Chang *et al.* 1997).

The important features of a neural network are (a) the basic computing elements (called neurons or nodes), (b) the network structure describing the connections between the computing elements, and (c) the training algorithm used to adjust or adapt the internode connection weights for solving a particular problem. Subsequently, the neural computing network learns to generalize from previous examples to new ones. For instance, in a pattern recognition task the network's response is insensitive to noisy or distorted patterns. The neural network is a set of computing elements organized into layers as illustrated in Fig. 2(a). The computing elements shown in the middle layer are called *hidden nodes*. There is no connection between nodes in the same layer. The input layer constitutes the input data for the nodes in the second layer (e.g. first hidden layer). The output data of the second layer are inputs to the third layer, and so on. The output (final) constitutes the overall response of the network to a given data in the input (first) layer. Thus, this type of neural network structure is called a *feed-forward network*. Each computing element, as shown in Fig. 2(b), calculates its net input by combining the values of all other computing elements to which it is connected. In other words, it multiplies the input values (X_i) issued from other computing elements by the corresponding connection weights (W_i), sums these values, and subtracts φ, the node's local threshold, from this sum. Finally, this weighted summation is modified by a transfer function f.

Different types of transfer function are used. In the threshold function type, which was used in the earliest neuron model (McCulloch & Pitts 1943), the output Y takes a value of 1 when its weighted summation exceeds a certain local threshold φ

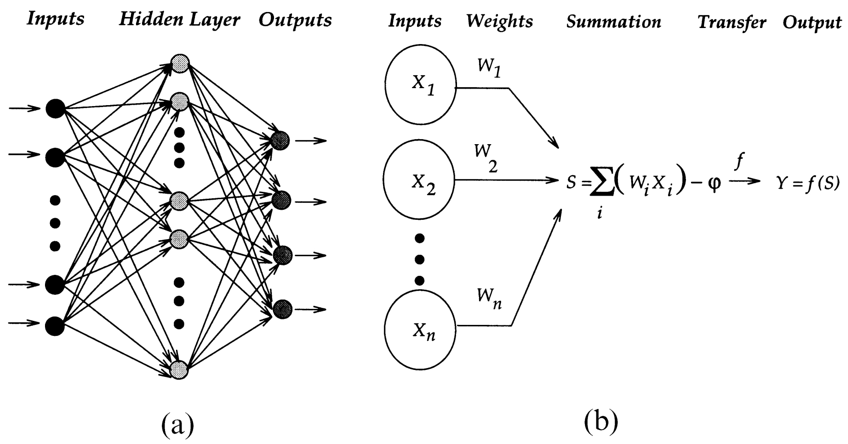

Figure 2. A multilayer neural network structure with one layer of hidden nodes (a) and an output computation of a single computing element (b).

and 0 otherwise. Sigmoid functions f expressed as

$$f(x)=\begin{cases}\frac{1}{1+e^{-x}} & \text{(sigmoid function)}\\ \frac{1-e^{-x}}{1+e^{-x}} & \text{(hyperbolic tangent function)}\end{cases} \tag{5}$$

are also popular.

The crucial problem is to identify a training rule that enables the neural network to fulfil a particular task correctly, that is to modify and adapt the connection weights in response to given input data. There are a variety of training algorithms in use today. The most widely used algorithm for training multi-layer feed-forward neural networks in a supervised manner is the *back-propagation algorithm*, which is based on the error-correction rule, and is the type we use. The network produces a set of outputs by propagating the input from the first layer to the last layer, then the difference (error) between this actual response and the desired response is propagated backwards to adjust the weights so that the error is minimized. The *learning coefficients* of the learning rule have a significant impact on the neural network performance.

Let $x_i^{(l)}$ be a current output value of ith node in layer (l), $w_{ij}^{(l)}$ be a connection weight joining the jth node in layer ($l-1$) to the ith node in layer (l), and $S_i^{(l)}$ be the weighted summation of inputs to the ith node in layer (l).

As described above, a back-propagation element transfers its input as

$$x_i^{(l)}=f(S_i^{(l)})\,, \tag{6}$$

where f is a transfer function.

Let E be the network's global error function. The local error at the ith node in layer l ($e_i^{(l)}$) is computed as

$$e_i^{(l)}=-\frac{\partial E}{\partial S_i^{(l)}}=f'(S_i^{(l)})\sum_k e_k^{(l+1)}w_{ki}^{(l+1)}\,. \tag{7}$$

The goal of the learning process is to minimize the network's global error function by modifying the connection weights by Δw_{ij} based on the knowledge of local error at each node. This is done using a *gradient descent rule* as follows:

$$\Delta w_{ij}^{(l)}=-\mathrm{lcoef}\left(\frac{\partial E}{\partial w_{ij}^{(l)}}\right)=\mathrm{lcoef}\,e_i^{(l)}x_j^{(l-1)}\,, \tag{8}$$

where lcoef is a learning coefficient.

As the divergent behaviour might occur during the minimization process, different variations on the standard algorithm (above) have been proposed, for instance another term called the *'momentum coefficient'* is added so that a portion of the previous delta weight (Δw_{ij}) is fed through the current delta weight as expressed by

$$\Delta w_{ij}^{(l)}=\mathrm{lcoef}\,e_i^{(l)}x_j^{(l-1)}+\mathrm{momentum}\;\Delta w_{ij}^{(l)}\,. \tag{9}$$

The global error function we use here is given by

$$E=\frac{1}{2}\sum_k (d_k-o_k)^2\,, \tag{10}$$

where d_k and o_k are respectively the desired and the actual output at each node in the output layer. Hence, E defines the global error of the network of a particular (I, D), where I is an input vector associated with a vector class D, usually represented as 1 of N code [e.g. $D=(0, \ldots, 0, 1, 0 \ldots 0)$, where only the node corresponding to the desired class has a value of 1; the others have values of 0]. Pre-scaling the raw data (input/output) before being presented to the neural network being used is vital. For instance, when the sigmoid (the hyperbolic tangent) transfer function is used, the input/output data should be suitably scaled between 0 and 1 (-1 and 1) by a simple linear transformation using minimum and maximum values of the corresponding data in order to avoid saturation.

There is no theoretical limit to the number of hidden layers. However, in practice generally a maximum of three hidden layers are required to solve a complex classification problem. Here we have used a three-layer network structure (one input, one hidden and one output layer) with either 15 or 20 hidden nodes in the hidden layer, which is the optimized network architecture we found for our classification problem involving 15 logs and three–five classes. The resulting *neural network* is $-N_iN_hN_o$, where N_i, N_h, and N_o are the number of nodes respectively in the input layer (corresponding to the number of logs), in the hidden layer and in the output layer (corresponding to the number of classes), e.g. 15154 = 15 logs, 15 hidden nodes and four classes network structure.

In our study we have used two variants of the learning algorithm that are best suited to our classification problem in terms of network performance rate: the Normalised Cumulative Delta (Norm-Cum-Delta) and the Extended Delta-Bar-Delta (EDBD) learning rules. Using the latter rule, several thousand iterations are needed to reach a convergent state, with the *learning* and *momentum rates* being adjusted through the iterations. The Norm-Cum-Delta can be a lot faster in terms of convergence time, because the *learning coefficient* and *momentum rates* can be constant. In practice the *learning coefficients* are changed manually in order to obtain optimum results.

We have used the NeuralWare software (Neuralware Professional II Plus; NeuralWare 1993) to design the neural networks (e.g. networks' structure and learning rules). These can be exported as C code programs to be accessed from the graphical user interface (PV-WAVE) of our system for training. As we may generally want to retrain the networks on new data sets, the issue of optimizing the *learning rates* to improve the classification rates is important. Hence, the EDBD rule is more suitable for our system to solve our problem of optimization, although the convergence process is slow, as mentioned above.

APPLICATION TO ODP HOLE 792*E*

Defining the problem

ODP Hole 792*E* was drilled in the forearc sedimentary basin of the Izu-Bonin arc south of Japan in 1989 (Taylor *et al.* 1990), with the aim of improving the understanding of the Neogene evolution of the nearby volcanic arc. The sediments recovered consisted of vitric sands and silts, pumiceous and scoriaceous gravels and conglomerates and claystones. During periods when the arc was relatively inactive, sedimentation was dominated by pelagic clays lithifying to claystone. Deposition of coarser-grained volcanic sandstones and siltstones was as a result of turbidity currents and reflects active construction of nearby volcanoes and their erosion. The geochemical logs from Hole 792*E* were analysed by Pratson *et al.* (1992) and Lovell *et al.* (1992). Also, Pezard *et al.* (1992a,b) and Hiscott *et al.*

(1992) analysed data from the resistivity imaging tool, the Formation Microscanner (FMS), that was deployed in this hole. Interpretation of these data relevant to this work are discussed later.

In this paper we analyse a lower part of the drilled section from 482–732 m below sea floor (mbsf), which is early Oligocene in age. Most of these rocks accumulated at a high rate, perhaps in less than 1 million years (Taylor *et al.* 1990). 79 per cent of this 250 m interval was recovered as core. Our criterion that cores to be matched with the logs for training purposes should represent at least 90 per cent of the cored interval is a form of quality control on the core data and reduces the total core available for this purpose to 50 per cent (13 out of 26 cores).

Choice of lithological classes

The way that the natural variability of rocks is divided into lithological classes that are of use to the geologist does not necessarily correspond to the natural subdivisions of the log data space based on physical measurements. Lithological classification is largely done on the basis of grain size and composition, whereas geophysical logs are largely controlled by the proportion and connectivity of the pore spaces and the fluids within them. Although there are guidelines (Mazzullo *et al.* 1987) for classifying the rocks encountered in the ODP, the actual assignments of sediment type are often made by more than one sedimentologist on a free-form descriptive paper record (VCD, visual core description form). Hence there may be a strong element of subjective expert knowledge in the initial classification.

We have converted the shipboard lithological description of the cores (VCD) for the 482–732 mbsf interval into a digital assignment of lithological class every 0.15 m down the hole, corresponding to the sampling interval of the logs. Five classes were used in the initial scheme: claystone, silty–sandy claystone, muddy–silty sandstone, siltstone–sandstone–ash and conglomerate–gravel. We have also used two other schemes, one three-class and one four-class. As we shall show, the very few depth-matched samples representing the claystone (8) and muddy–silty sandstone (34) classes and the lack of distinction between them prompted us to merge them with the silty–sandy claystone class to produce a single claystone class in the three-class scheme (claystone, sandstone, conglomerate). The four-class scheme was derived from the three-class scheme by splitting the conglomerate class into two: conglomerate1 and conglomerate2 based on clast type. The ability to modify the class assignments to specific depth intervals based on the data proved to be a vital component of the analysis.

Choice of logs

Four separate runs were made down Hole 792*E* using resistivity, lithodensity, sonic, natural gamma and geochemistry sensors. The sensors can generally resolve vertical differences over distances of about 0.4–0.6 m, except for the three resistivity sensors, whose vertical resolutions are 2, 1.5 and 0.8 m respectively. Thus beds thinner than these values will give 'mixed' signals from more than one rock type at the data sampling interval of 0.15 m. From these sensors, 17 parameters were considered for use in classification: spectral gamma, computed gamma, radioactive potassium, thorium and uranium, deep, medium and shallow resistivity, density, photoelectric effect, sonic velocity and the oxide contents of calcium, silicon, iron, titanium, potassium and aluminium. Unfortunately, no density and photoelectric effect measurements are available below 550 mbsf. Hence we use only those 15 logs that are available for the whole interval 482–732 mbsf.

Where the hole has been widened by collapse of the walls or because of some other artefact of the drilling process, the physical properties recorded by the downhole sensors can have spurious values. Standard log processing techniques do, however, correct to some extent for the effects on logs. Three additional logs—caliper, density-correction and geochemical factor logs—can be used to detect places where such spurious values in the other logs may occur. Threshold values on these quality-control logs (caliper <11.8 inches, density correction <0.1 gm cm^{-3} and geochemical factor <800) have been set in this way by examining the data and using experience. Application of these thresholds removes about 3 per cent of the data that occur as outliers. The cores that achieved greater than 90 per cent recovery and are used for training and testing of the classifiers are 37, 38, 40, 42, 43, 46, 48, 49, 50, 56, 57, 60 and 62.

Principal components analysis

The results of PCA on the 15 logs for the 482–732 mbsf interval are shown in Tables 1 and 2. Six components have eigenvalues greater than 0.7 (Jollife 1972) and the first nine components each contain greater than 1 per cent of the original variance. The first six components are dominated by the following logs: 1 = natural radioactivity (particularly potassium) and resistivity; 2 = sonic velocity; 3 = silicon/calcium/iron oxides; 4 = radioactive thorium/uranium; 5 = aluminium oxide; 6 = titanium oxide. The application of the quality control has only a marginal effect on the PCA results. The dominant influence of variations in natural radioactivity,

Table 1. Eigenvalues of principal components and their percentage of contributions in terms of variance with respect to the total of the variances of the original variables, after application of quality-control measures. The fourth column shows the cummulative percentage, which accounts for the total variation of the data set by the principal components.

Principal components	Eigenvalues	Individual % variance	Cummulative % variance
1	5.118	34.12	34.12
2	2.891	19.28	53.39
3	2.625	17.50	70.90
4	1.232	8.22	79.11
5	1.112	7.41	86.52
6	0.915	6.10	92.62
7	0.542	3.61	96.23
8	0.279	1.86	98.09
9	0.213	1.42	99.51
10	0.045	0.30	99.81
11	0.022	0.15	99.96
12	0.005	0.03	99.99
13	0.001	0.00	100.0
14	0.000	0.00	100.0
15	0.000	0.00	100.0

Table 2. Principal component loading matrix for the first six principal components of Table 1 whose eigenvalues are greater than 0.7 (Jolliffe 1972).

Logs	PCA(1)	PCA(2)	PCA(3)	PCA(4)	PCA(5)	PCA(6)
SGR	0.842	−0.526	0.093	0.007	−0.016	0.030
POTA	0.871	−0.450	0.080	0.145	0.017	0.013
THOR	−0.079	−0.448	0.075	−0.714	−0.175	0.096
URAN	−0.272	0.361	0.127	0.738	−0.119	0.055
ILD	0.799	0.492	−0.275	−0.086	−0.049	−0.058
ILM	0.805	0.493	−0.275	−0.081	−0.055	−0.055
SFLU	0.784	0.485	−0.271	−0.081	−0.062	−0.061
VMEAN	0.436	0.740	−0.268	−0.154	−0.136	0.010
CaO	−0.279	−0.283	−0.855	0.073	0.034	−0.060
SiO_2	0.111	0.271	0.896	−0.079	−0.178	−0.259
FeO	−0.085	−0.302	−0.742	0.103	−0.369	0.272
TiO_2	0.090	0.256	0.410	−0.025	−0.280	0.794
K_2O	0.818	−0.512	0.138	0.147	0.071	−0.019
Al_2O_3	0.175	0.197	−0.029	−0.063	0.885	0.337
Eigv.	5.118	2.891	2.625	1.232	1.112	0.915

particularly potassium, is typical of many ODP holes. Thorium, usually indicative of a volcanic source, has a negative input to principal component 4. The three resistivity logs are highly correlated.

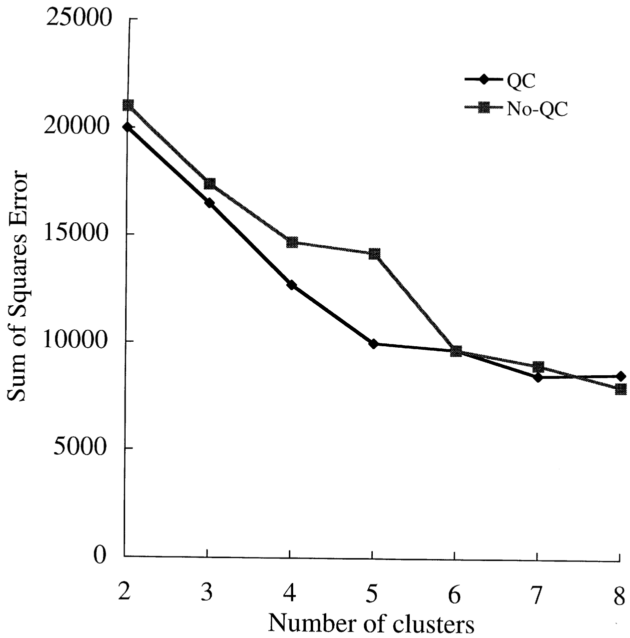

Figure 3. Graph of the sum-of-squares-error function against the number of clusters for Borehole 792*E* logging data where the quality control is applied and is not applied.

Cluster analysis (CA)

From the PCA we infer six significant dimensions to the data. A K-means unsupervised cluster analysis should help us to understand these data in terms of distinct classes. Using a sum-of-squares-error criterion in a pseudo-Fisher test of significance we find that the data seem to support five clusters (Table 3, Fig. 3). Interestingly, the data without quality control achieve the null hypothesis at only four clusters (Table 3), although this looks like a false result in Fig. 3.

Fig. 4 shows the first two principal component values (representing over 50 per cent of the variance) plotted according to the core-based classes in the five-class scheme (Fig. 4, upper) and the K-means clustering (Fig. 4, lower). Clusters 4 and 5 of the K-means analysis are quite distinct but the other three clusters are less so, more suggestive of a continuum of values. The two low-population classes (claystone and muddy–silty sandstone) in the core-based scheme are not sufficiently distinct to be separable from the other classes, and the conglomerate–gravel class has two distinctly separate populations (Fig. 4, upper). Hence the original core-based five-class scheme would not be a good basis to classify the uncored sections of Hole 792*E*. Instead, two other classification schemes have been created. In the first we merge the claystone and muddy–silty sandstone with the silty claystone class to produce a three-class scheme. In the second we take this scheme and split the conglomerate–gravel class into two classes. Inspection of the core descriptions for those conglomerate intervals that largely correspond to cluster 4 in Fig. 4 (lower) show that they are rocks dominated by claystone clasts, rather than clasts of volcanic rock. This then forms the geological rationale for splitting the conglomerate class into conglomerate1 (volcanic

Table 3. A pseudo-Fisher test of significance to decide whether the K_2-cluster is an appropriate solution given the K_1-cluster solution, in other words testing the null hypothesis H_0 *'the solution for K_2 clusters provides no better fit than the solution for K_1 clusters'*. The test suggests that the Borehole 792*E* logging data (depth 482–732 mbsf) have five clusters where the quality control is applied and four clusters where the quality control is not applied.

Clusters K_1	Clusters K_2	Degrees of freedom τ_1	Degrees of freedom τ_2	F-ratio F_{K_1,K_2}	Tabled values $F_{\tau_1,\tau_2,\alpha=0.05}$	Tabled values $F_{\tau_1,\tau_2,\alpha=0.01}$	Test result of H_0
Quality control is applied							
2	3	15	11610	3.60	1.67	2.04	rejected
3	4	15	11595	7.10	1.67	2.04	rejected
4	5	15	11580	8.65	1.67	2.04	rejected
5	6	15	11565	1.50	1.67	2.04	accepted
Quality control is not applied							
2	3	15	12270	3.60	1.67	2.04	rejected
3	4	15	12255	4.00	1.67	2.04	rejected
4	5	15	12240	1.44	1.67	2.04	accepted

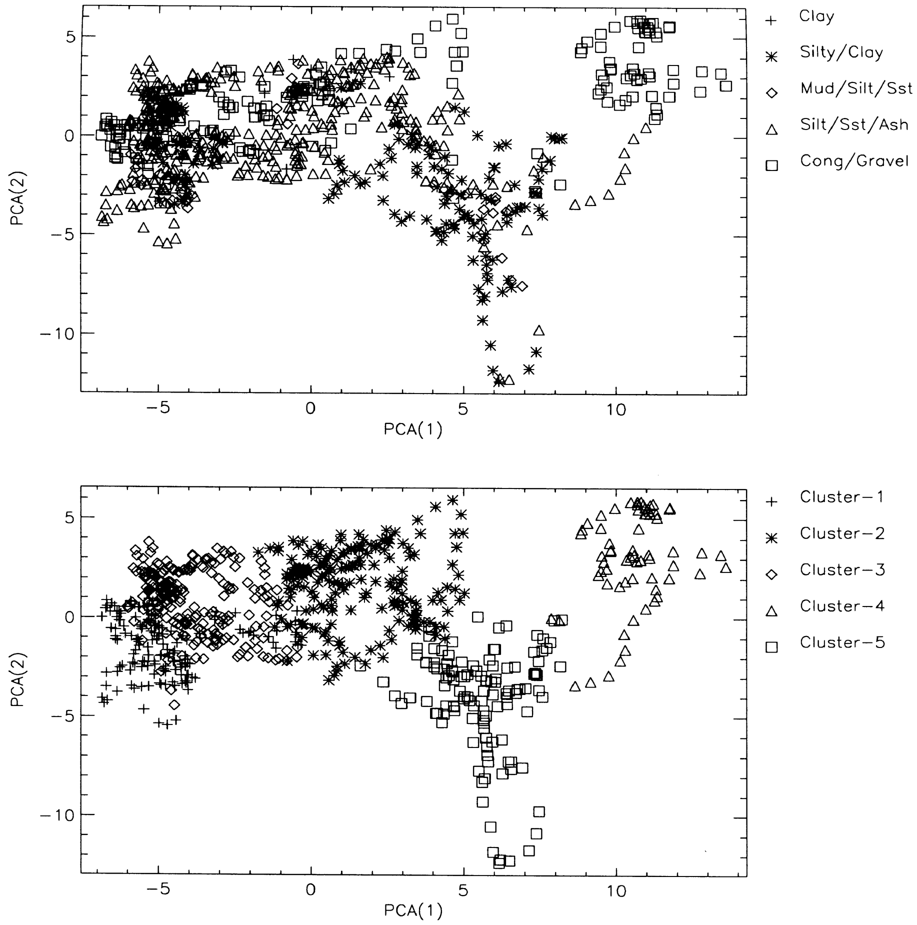

Figure 4. Cross-plots of the first two principal components for Borehole 792*E* logging data (upper) for the five-class core classification, and (lower) the five-cluster *K*-means classification.

clasts) and conglomerate2 (claystone clasts) in the four-class scheme (Fig. 5, upper). The sandstone and conglomerate1 classes are still not well separated in this plot, nor in the equivalent four-cluster plot (Fig. 5, lower).

Supervised classification

We now have three core-based classification schemes for Hole 792*E* matched to log data sets in both quality-controlled and raw forms. For each we create two sets of training/testing data using the smallest-class-population and whole-class-population methods of data selection. Copies of these data sets are available from the second author. Each data set is then used in both the back-propagation neural network (BpNN) and discriminant analysis (DA) classifiers to give a matrix of $3\times2\times2\times2$ results.

Results

The relative performance, measured as a percentage of test samples classified as belonging with 70 per cent certainty to the correct class, of the various discriminant analysis classifier types on the data, varied greatly. The normal-kernel classifier was consistently the best (Table 4). The first two discriminant function scores for the three classification schemes are shown in Figs 6, 7 and 8. The four-class scheme is apparently the most successful. This is confirmed by the performance figures: three-class = 81.4 per cent, four-class = 91.2 per cent, five class = 81.2 per cent. However, the DA performance was consistently bettered by the BpNN, typically by 2–3 per cent: three-class = 84.2 per cent, four-class = 93.3 per cent,

Table 4. Classification performance results of the four-class scheme of discriminant analysis (DA) and neural network techniques at the 0.7 threshold. The values between parentheses are for the Borehole 792*E* logging data where the quality control is not applied; for all other values the quality control is applied. The minimum-class-population method is used for producing training and testing data sets.

	Training rate (%)	Testing rate (%)
Normal linear DA	65.24	65.85
Equal variance	(69.51)	(67.07)
Normal quad DA	87.20	79.88
Unequal variance	(90.24)	(78.66)
Normal kernel DA	98.78	83.54
Equal bandwidth	(96.34)	(80.50)
Normal kernel DA	98.78	82.32
Unequal bandwidth	(98.78)	(82.32)
Epanechnikov kernel DA	75.61	70.12
Equal bandwidth	(90.00)	(66.46)
Epanechnikov kernel DA	91.46	65.85
Unequal bandwidth	(96.34)	(54.88)
Neural network 15154	100(100)	**85.00(86.59)**

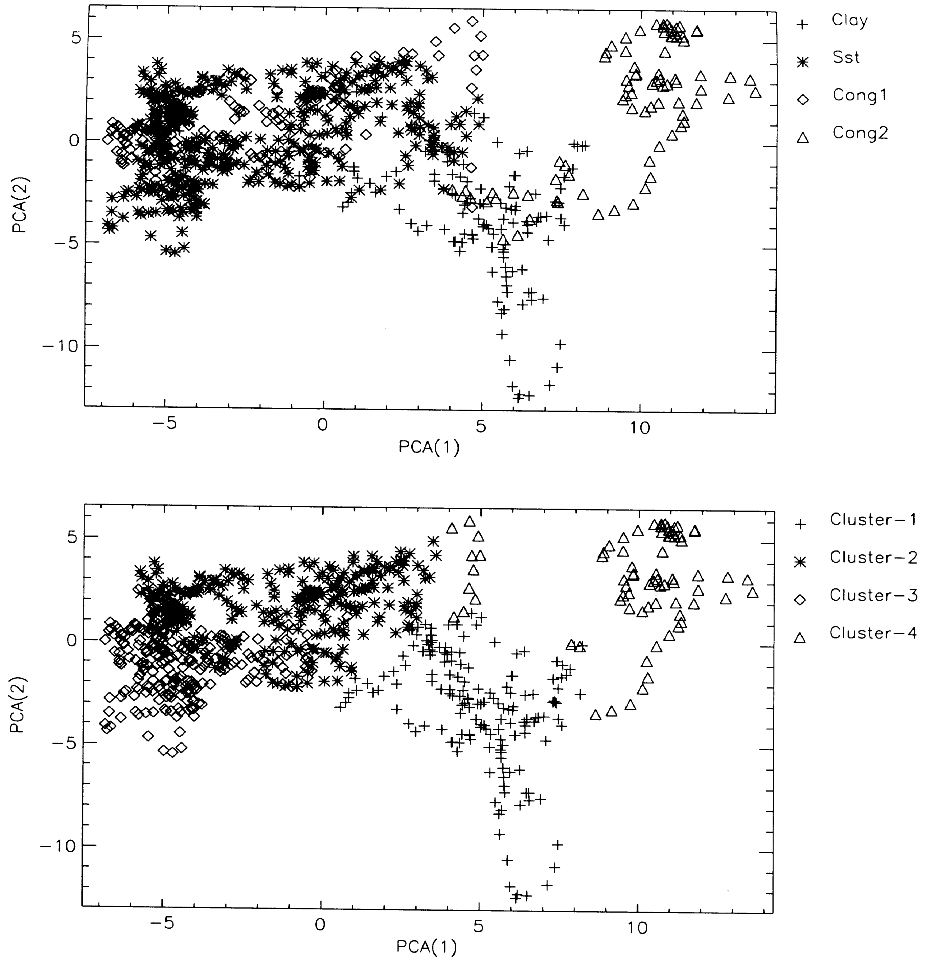

Figure 5. Cross-plots of the first two principal components for Borehole 792*E* logging data (upper) for the four-class core classification, and (lower) the four-cluster *K*-means classification.

five-class = 83.0 per cent (Tables 6, 7 and 8). Comparison of results from the whole-class-population and the smallest-class-population data sets show markedly better performance by the whole-class- population data set by 5–7 per cent (Tables 5 and 6). The effect of this can best be appreciated by a graphical representation of the results (Figs 9 and 10). The most obvious difference is that there is a general change in class assignments to the sandstone class (largest population) in the whole-class-population case from the other classes (particularly the conglomerate1 class) in the smallest-class-population case. This supports the idea that the distribution of the sandstone class was being under-represented in the smallest-class-population case. The extra improvement in the classification performance brought about by the quality-control measures is only slight (0–1 per cent) in both the DA and BpNN classifiers (Tables 5 and 6).

The graphical illustration of the superior performance of the BpNN over the DA for this hole can be seen in the

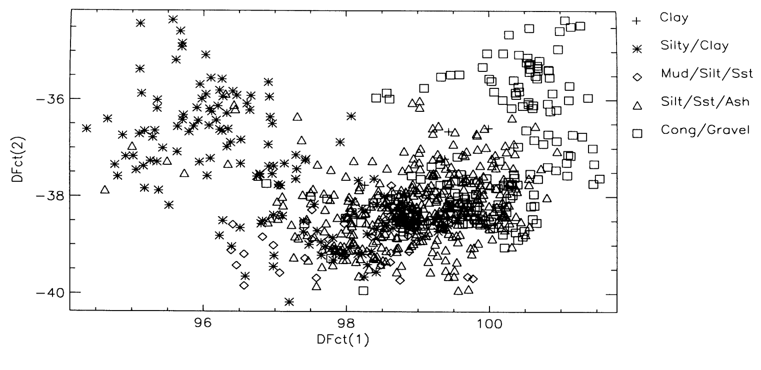

Figure 6. Cross-plot of discriminant-function scores computed from the ODP Borehole 792*E* logs using the five-class core-based classification. According to the chi-square test four discriminant functions are needed to represent the five-class population.

Table 5. Confusion matrix of the four-class scheme of discriminant analysis of Borehole 792*E* logging data using normal kernel density estimates with unequal bandwidth at the 0.7 threshold. The values between parentheses are for the data where the quality control is not applied; for all other values the quality control is applied. The minimum-class-population method is used for producing training and testing data sets.

Probabilistic classification	Training data set Manual classification Clay	Sst	Cong1	Cong2	Total
Clay	41(41)	0(0)	0(0)	0(0)	41(41)
Sst	0(0)	40(40)	0(0)	0(0)	40(40)
Cong1	0(0)	0(0)	40(40)	0(0)	40(40)
Cong2	0(0)	0(0)	0(0)	41(41)	41(41)
Other	0(0)	1(1)	1(1)	0(0)	2(2)
Total	41(41)	41(41)	41(41)	41(41)	164(164)
Rate (%)	100.0(100.0)	97.56(97.56)	97.56(97.56)	100.0(100.0)	98.78(98.78)

Probabilistic classification	Testing data set Manual classification Clay	Sst	Cong1	Cong2	Total
Clay	35(38)	0(7)	1(1)	0(3)	36(49)
Sst	4(1)	25(22)	2(1)	0(0)	31(24)
Cong1	0(1)	7(8)	37(37)	0(0)	44(46)
Cong2	0(0)	0(0)	0(0)	40(38)	40(38)
Other	2(1)	9(4)	1(2)	1(0)	13(7)
Total	41(41)	41(41)	41(41)	41(41)	164(164)
Rate (%)	85.37(92.68)	60.98(53.67)	90.24(90.24)	97.56(92.68)	83.54(82.32)

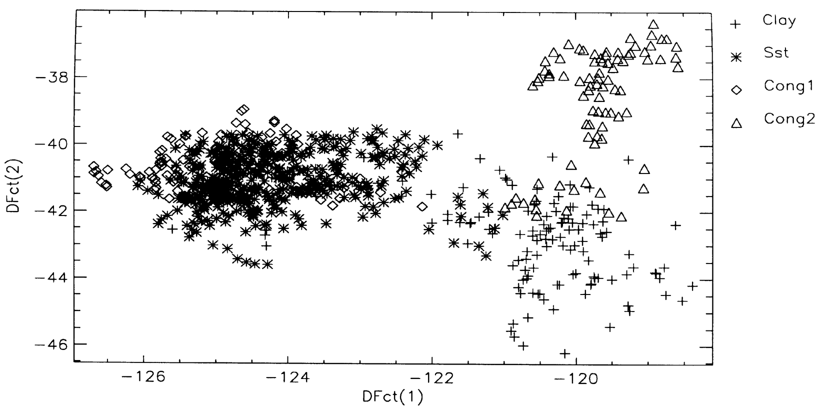

Figure 7. Cross-plot of discriminant-function scores computed from the ODP Borehole 792*E* logs using the four-class core-based classification. According to the chi-square test, three discriminant functions are needed to represent the four-class population.

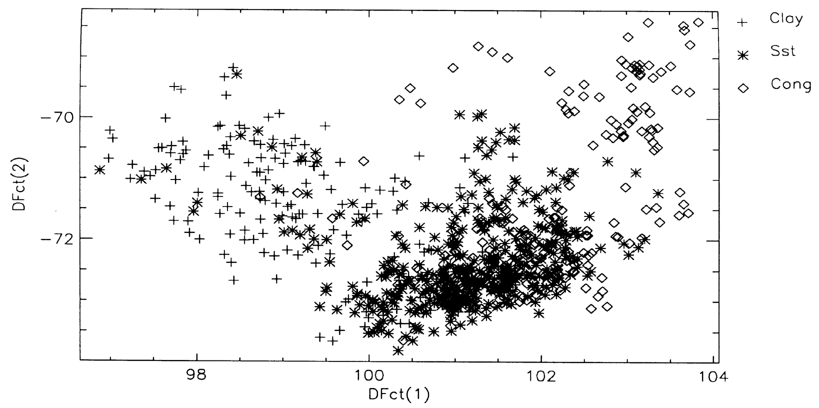

Figure 8. Cross-plot of discriminant-function scores computed from the ODP Borehole 792*E* logs using the three-class core-based classification. According to the chi-square test, two discriminant functions are needed to represent the three-class population.

Table 6. Confusion matrix of the four-class scheme of discriminant analysis of Borehole 792*E* logging data using normal kernel density estimates with unequal bandwidth at the 0.7 threshold. The values between parentheses are for the data where the quality control is not applied; for all other values the quality control is applied. The whole class population is used for selecting training and testing data sets.

Probabilistic classification	Training data set, Manual classification: Clay	Sst	Cong1	Cong2	Total
Clay	65(66)	0(0)	0(0)	0(0)	65(66)
Sst	0(0)	215(231)	0(2)	0(0)	215(233)
Cong1	0(0)	0(0)	54(53)	0(0)	54(53)
Cong2	0(0)	0(0)	0(0)	41(41)	41(41)
Other	2(3)	3(2)	8(11)	0(0)	13(16)
Total	67(69)	218(233)	62(66)	41(41)	388(409)
Rate (%)	97.01	98.62	87.10	100.0	96.65
	(95.65)	(99.14)	(80.30)	100.0	(95.60)

Probabilistic classification	Testing data set, Manual classification: Clay	Sst	Cong1	Cong2	Total
Clay	61(62)	4(2)	0(1)	1(1)	66(66)
Sst	3(3)	205(221)	4(7)	0(0)	212(231)
Cong1	0(0)	0(0)	48(46)	0(0)	48(46)
Cong2	0(0)	0(0)	0(0)	40(40)	40(40)
Other	3(4)	9(10)	10(12)	0(0)	22(26)
Total	67(69)	218(233)	62(66)	41(41)	388(409)
Rate (%)	91.04	94.04	77.42	97.56	91.23
	(89.86)	(94.85)	(69.70)	(97.56)	(90.22)
	Neural network 15154 results				
Rate (%)	91.04	95.41	83.87	100.0	93.30
	(88.61)	(96.34)	(79.97)	(100.0)	(93.00)

classification of the conglomerate1 in cores 48/49 (Fig. 10). However, it is the classification results from the cores not used in the training and testing [cores 39 (recovery 72 per cent), 41(71 per cent), 44(68 per cent), 45(59 per cent), 47(51 per cent), 51(50 per cent), 52(83 per cent), 53(0 per cent), 54(87 per cent), 55(37 per cent), 58(54 per cent), 59(85 per cent), 61(64 per cent)] that are of primary interest. For all cores but 53 we have partial core recovery and can attempt some qualitative matching of the classification results with the described cores (see below and Fig. 10). Material recovered at the top of each core is likely to be closest to its 'correct' position downhole. In addition, we can compare these results with the FMS-derived log results of Hiscott *et al.* (1992, Fig. 4). These give a spatially detailed (about 2 cm resolution) log based on an assumed correlation between grain size and resistivity, and show a high degree of correlation with the resistivity logs.

Core 39. This is classed as an intercalation of claystone and sandstone by both classifiers. The BpNN appears to underestimate the claystone in favour of sandstone.

Core 41. The BpNN is more successful in picking up the conglomerate1 at the top of the core.

Core 44. Both classifiers assign the conglomerate at the top of this core as conglomerate1 not conglomerate2, with largely sandstone below.

Core 45. Both classifiers see a substantial claystone horizon at the top of this core that is not represented in the recovered core, which is an intercalation of sandstone, conglomerate1 and claystone. The two classifiers differ in the lower part, with the BpNN assigning mainly conglomerate2 in continuity with core 46 below, whilst the DA has mainly sandstone ending abruptly at the top of core 46. The continuity of the BpNN result is apparently confirmed in the FMS log, which shows a gradational increase of grain size into core 46.

Core 47. The recovery from this core is of conglomerate1 and conglomerate2. This fits much better with the BpNN assignments for the upper part of this interval whilst the DA has no conglomerate1 class. Both classifiers agree closely with a claystone, sandstone, conglomerate1 sequence at the bottom of the interval. The FMS log agrees closely with the BpNN result.

Core 51. A submetre-scale intercalation of sandstone and conglomerate1 in the material recovered from this core is mainly assigned to sandstone in both classifiers, although the BpNN appears to do better at assigning more conglomerate1, a trend which is continued down the core. The 3-m-thick conglomerate bed near the bottom of core 51 is clear in the FMS and BpNN logs but is missing from the DA result.

Core 52. Neither classifier matches well the recovery, which is a submetre-scale mixture of claystone and sandstone. Both underestimate the claystone.

Table 7. Confusion matrix of the three-class scheme of discriminant analysis of Borehole 792*E* logging data, where quality control is applied, using normal kernel density estimates with unequal bandwidth at the 0.7 threshold. The whole class population is used for selecting training and testing data sets. The equivalent neural network results are shown.

Probabilistic	Training data set				Testing data set			
	Manual classification				Manual classification			
classification	Clay	Sst	Cong	Total	Clay	Sst	Cong	Total
Clay	78	0	0	78	66	11	0	77
Sst	0	195	0	195	9	170	2	181
Cong	0	0	84	84	0	2	79	81
Other	12	10	8	30	15	22	11	48
Total	90	205	92	387	90	205	92	387
Rate (%)	86.67	95.12	91.30	92.25	73.33	82.93	85.89	81.40
	Back-propagation neural network 15153 results							
Rate (%)	100.0	100.0	100.0	100.0	66.67	94.63	78.26	84.24

Table 8. Confusion matrix of the five-class scheme of discriminant analysis of Borehole 792*E* logging data, where quality control is applied, using normal kernel density estimates with unequal bandwidth at the 0.7 threshold. The whole class population is used for selecting training and testing data sets. The equivalent neural network results are shown.

	Training data set					
Probabilistic	Manual classification					
classification	Clay	Silt/clay	Mud/silt/sst	Silt/sst/ash	Cong/gravel	Total
Clay	4	0	0	0	0	4
Silt/clay	0	67	0	0	0	67
Mud/silt/sst	0	0	12	0	0	12
Silt/sst/ash	0	0	0	192	0	192
Cong/gravel	0	0	0	0	87	87
Other	0	5	5	8	8	26
Total	4	72	17	200	95	
Rate (%)	100.0	93.06	70.59	96.00	91.58	93.30
	Testing data set					
Probabilistic	Manual classification					
classification	Clay	Silt/clay	Mud/silt/sst	Silt/sst/ash	Cong/gravel	Total
Clay	0	0	0	0	0	0
Silt/clay	0	57	5	10	1	73
Mud/silt/sst	0	0	1	0	0	1
Silt/sst/ash	2	5	7	176	3	193
Cong/gravel	0	0	0	0	81	81
Other	2	10	4	14	10	40
Total	4	72	17	200	95	388
Rate (%)	0.00	79.17	5.88	88.00	85.26	81.19
	Back-propagation neural network 15205 results					
Rate (%)	25.00	83.33	47.06	88.00	81.05	83.00

Core 53. There was no recovery here. The DA largely assigned claystone to this interval. The BpNN assigned claystone and conglomerate2. The FMS log shows a finely bedded intercalation of claystone and fine sandstones. The DA performs better on textural grounds in this case than the BpNN.

Core 54. A high recovery rate of sandstone and thinner claystone beds is moderately well matched by the BpNN assignments, although some sandstone is misassigned to claystone. This tendency is exaggerated in the DA results.

Core 55. Claystone above sandstone is the pattern of the BpNN results, which is the same as that of the recovered rocks.

Core 58. A low recovery of dominantly claystone with minor sandstone is not well matched by the classifiers which assign sandstone, exclusively in the case of the DA.

Core 59. A similar pattern to core 58, although with much better recovery.

Core 61. The BpNN appears to match well the sandstone-above-conglomerate1 pattern from the 64 per cent recovered rocks. The DA misses the conglomerate1 horizon altogether.

The above observations indicate that the BpNN does significantly better at recognizing conglomerate1 than the DA (e.g. cores 41, 47, 51, 61) and possibly at recognizing conglomerate2. Both classifiers perform less well, as expected, when the beds are thin ($< \sim 0.5$ m), close to the approximate resolution of the sensors. This is particularly apparent in cores 52 and 53. Both classifiers also show evidence of difficulties in discriminating between some sandstone and claystone, particularly in the lower part of the hole (e.g. 52, 58, 59). These cores contain samples that were classified as muddy–silty sandstone in the original classification, thus reinstating such a class, intermediate between sandstone and claystone, may be a useful next step. We do not pursue this here but point out again the value of being able to adapt the classification scheme.

DISCUSSION AND CONCLUSIONS

The geological significance of the classification results from Hole 792*E* are discussed in preliminary form in Wadge *et al.* (1998). At the level of individual cores we are convinced that useful interpretative information can be derived from the classified results. Our results show that within this 250 m interval there are five major depositional sequences with boundaries at about 515, 590, 660 and 715 mbsf. The third boundary was not recognized previously, but its existence becomes apparent by classifying the poorly sampled interval between 570 and 665 mbsf. Above 515 mbsf the sequence is a bimodal combination of claystone and sandstone, whilst below, down to 590 mbsf, upward-fining sandstones and conglomerates dominate. The boundary at 590 mbsf also marks the beginning of a downhole increase in smectite concentration and magnetic susceptibility (Taylor *et al.* 1990). The interval from 590 to 600 mbsf is also a series of sandstone and conglomerate1 in its upper half but is dominated by claystone and conglomerate2 in its lower half. Below this depth there is a return to sandstone lithologies. The boundary at 715 mbsf marks a change to claystone/conglomerate2 rocks, and multichannel seismic data suggest this corresponds to a major unconformity.

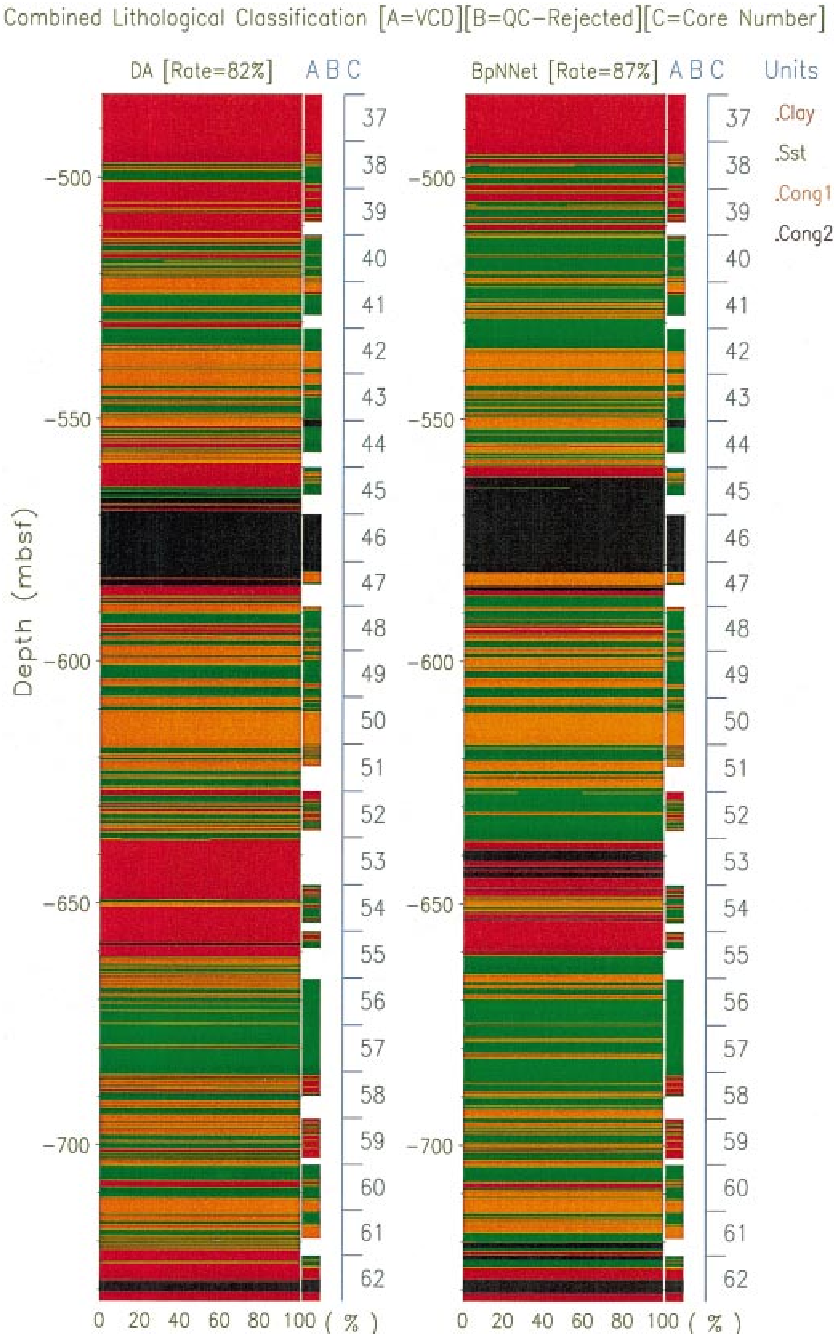

Figure 9. Graphical display of lithologies for the four-class scheme (claystone, sandstone, conglomerate1 and conglomerate2) for each sample of the hole, using the two classifiers discriminant analysis (DA) and back-propagation neural network (BpNNet) for the Borehole 792*E* logging data. The results displayed are based on the dominant lithology (> 50 per cent). The performance rate of the applied approach is evaluated from an independent test data set. The minimum-class-population method is used for producing training and testing data sets. No quality control is applied. Column A shows the class assignments based on the shipboard description (VCD) of recovered core. Column B shows the intervals of data rejected by the quality-control exercise. Cores with only partial recovery (39, 41, 44, 45, 47, 51–55, 58, 59, 61) are shown with their class assignments 'hung' from the top of their depth intervals.

In their study of the geochemical well logs of Hole 792*E*, Lovell *et al.* (1992) noted the general lack of broad correlations between the geochemistry and the geological units. However, when they performed a cluster analysis (INCA) on seven logs (silicon, aluminium, titanium, iron, calcium and potassium oxides and sulphur) they found five cluster groups in the 482–732 mbsf interval (Lovell *et al.* 1992, Fig. 5), the same number as identified by our *K*-means analysis using more logs. Most of these groups were widely distributed except a high-potash and high-silica group that was restricted to the 650–670 mbsf interval. Although we do not have such a lithological type in our results, we do note that the depth (about 660 mbsf) corresponds to one of the depositional sequence boundaries identified by us. We see little other obvious correlation between these geochemical groupings and our log results.

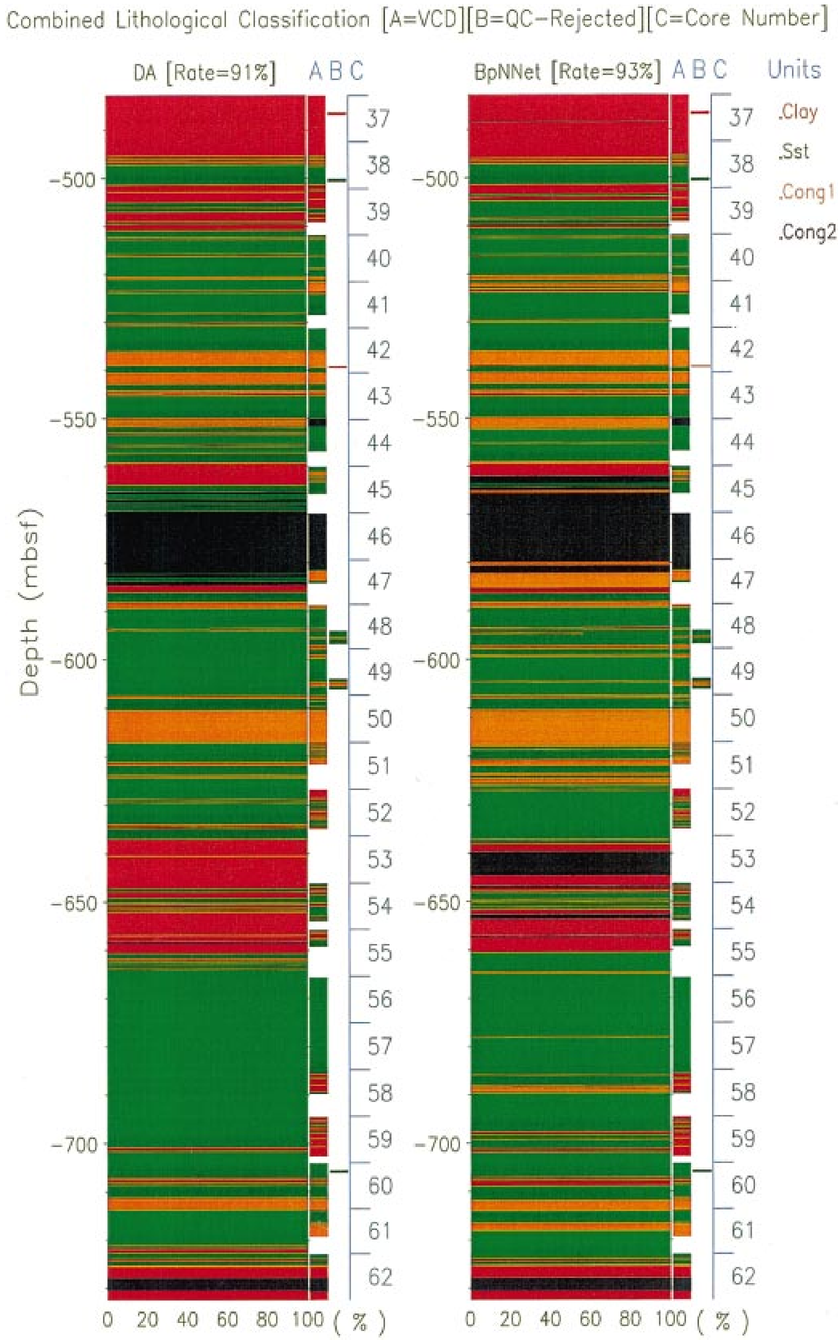

Figure 10. As for Fig. 9 but for the whole-class-population method for training and testing the classifiers with quality control applied to the log data.

In our discussion of results from cores with poor recovery we used the FMS-derived log of Hiscott *et al.* (1992) and Pezard *et al.* (1992b) as an independent validatory source of information. Alternatively, the FMS data could be incorporated into the data set used for classification. This would potentially improve the textural accuracy of the classification, particularly with regard to bed boundaries. Although the classes used in our four-class result have textural labels, the data that underpin the classification process obviously depend on a combination of the geochemical, textural and fluid properties of the rocks. The optimum classification scheme in any case depends on how the geologist seeks to interpret the sequence.

Our classification with four classes illustrates what we believe to be a sound method for deriving a complete lithological log from downhole logs and partial core. The neural network classifier has proved itself superior to the best of the discriminant analysis classifiers tested, both in testing performance rates and in most, although not all, FMS-validated qualitative analysis of low-recovery cores. We have also demonstrated that for selecting data with which to train and test the classifiers, the whole-class-population method is superior to the smallest-class-population method. Quality-control exclusion of log-value outliers gave only a minor improvement on classification performance, although it would be greater for noisier data.

Exploratory data analysis using principal components analysis and K-means clustering showed that the log data could sustain up to five lithological classes. We argue strongly that the approach to lithological classification for ODP holes must be flexible, recognizing both the subjectivity of the original shipboard record and the need to modify the classification subsequently as more is learnt about the character of the rocks in individual holes.

ACKNOWLEDGMENTS

This work was funded by NERC grants GST/02/0993 and F60/G6/12/02. We would like to thank the Borehole Research Group at Lamont-Doherty Observatory and the ODP staff at Texas & AM University for supplying data. Comments from an anonymous reviewer helped us to improve the quality of the paper. We also wish to thank everyone who contributed to this project either directly or indirectly, particularly Drs Christopher Godsalve, David Pearson and Kevin Hodges who helped us to type this manuscript in LaTeX. Finally, we wish to thank Dr Howard Grubb of the Department of Applied Statistics of the University of Reading for his fruitful discussions and Mrs Jane Brookling, our secretary, for her support.

REFERENCES

Agrinier, P. & Agrinier, B., 1994. A propos de la connaissance de la profondeur a laquelle vos echantillons sont collectes dans les forages, *Comptes Rendus de l'Academie Sciences de Paris,* **318,** serie II, 1615–1622.

Baldwin, J.L., Bateman, A.R.M. & Wheatley, C.L., 1990. Application of neural networks to the problem of mineral identification from well-logs, *The Log Analyst,* **3,** 279–293.

Bishop, C.M., 1995. *Neural Networks for Pattern Recognition,* Clarendon Press, Oxford.

Busch, J.M., Fortney, W.G. & Berry, L.N., 1987. Determination of lithology from well logs by statistical analysis, *SPE Formation Evaluation,* **2,** 412–418.

Chang, H.C., Chen, H.C. & Fang, J.H., 1997. Lithology determination from well logs with fuzzy association memory neural network, *IEEE Trans. Geoscience and Remote Sensing,* **35,** 773–780.

Cheng, B. & Titterington, D.M., 1994. Neural networks: a review from a statistical perspective, *Statistical Sci.,* **9,** 2–54.

Davis, J.C., 1986. *Statistics and Data Analysis in Geology,* John Wiley, New York.

Devijver, P. & Kittler, J., 1982. *Statistical Pattern Recognition,* Prentice Hall International, Englewood Cliffs, NJ.

Doveton, J.H., 1986. Log analysis of subsurface geology-concepts and computer methods, *John Wiley & Sons, New York,* 273 p.

Doveton, J.H., 1994. Geological log analysis using computer methods, *AAPG Computer Applications in Geology,* No. 2.

Duda, R.O. & Hart, P.E., 1973. Pattern classification and scene analysis, *John Wiley, New York*.

Fisher, R.A., 1936. The use of multiple measurements on taxonomic problems, *Ann. Eugenics,* **7,** 179–188.

Fisher, R.A., 1938. The statistical utilization of multiple measurements, *Ann. Eugenics,* **8,** 376–386.

Gallinari, P., Thiria, S., Badran, F. & Fogelman-Soulie, F., 1991. On the relations between discriminant analysis and multilayer perceptrons, *Neural Networks,* **4,** 349–360.

Goncalves, C.A., 1995. Characterisation of formation heterogeneity, *PhD thesis,* University of Leicester.

Hartigan, J.A., 1975. *Clustering Algorithms,* John Wiley, New York.

Hartigan, J.A. & Wong, M.A, 1979. Algorithm AS 136: a K-means clustering algorithm, *Appl. Stat.,* **28,** 100–108.

Haykin, S.S., 1994. *Neural Networks: a Comprehensive Foundation,* Macmillan, New York.

Healy, M.J.R., 1986. *Matrices for Statistics,* Clarendon Press, Oxford.

Hiscott, R.N., Colella, A., Pezard, P., Lovell, M.A. & Malinverno, A., 1992. Sedimentology of deep-water volcaniclastics, Oligocene Izu-Bonin forearc basin, based on Formation Microscanner images, in *Proc. ODP. Sci. Results,* Vol. 126, pp. 75–96, eds Taylor, B., Fujioka, K. *et al.,* College Station, TX.

Jolliffe, I.T., 1972. Discarding variables in a principal components analysis, 1: Artificial Data, *Appl. Stat.,* **21,** 160–173.

Jolliffe, I.T., 1973. Discarding variables in a principal components analysis, 2: Real Data, *Appl. Stat.,* **22,** 21–31.

Kaiser, H.F., 1960. The application of electronic computers to factor analysis, *Educational and Psychological Measurements,* **20,** 141–151.

LeCun, Y., 1985. Une procedure d'apprentissage pour reseau a seuil assymetrique, *Cognitiva,* **85,** 599–604.

Lovell, M.A., Pezard, P.A. & Harvey, P.K., 1992. Chemical stratigraphy of boreholes in the Izu-Bonin arc from in situ nuclear measurements, in *Proc. ODP. Sci. Results,* Vol. 126, pp. 593–600, eds Taylor, B., Fujioka, K. *et al.,* College Station, TX. Mazullo, J., Meyer, A. & Kidd, R.B., 1987. A new sediment classification scheme for the Ocean Drilling Program, *ODP Technical Note,* **8.**

McCulloch, W.W. & Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity, *Bull. math. Biophysics,* **5,** 115–133.

Milligan, G.W., 1980. An examination of the effect of six types of error perturbation on fifteen clustering algorithms, *Psychometrika,* **45,** 325–342.

NeuralWare, 1993. *Using NeuralWorks: a tutorial for professional II/ plus and NeuralWorks explorer,* Neuralware Inc., Technical Publications Group, Pittsburgh.

Pezard, P.A., Lovell, M.A. & Hiscott, R.N., 1992a. Downhole electrical images in volcaniclastic sequences of the Izu-Bonin forearc basin, Western Pacific, in *Proc. ODP. Sci. Results,* Vol. 126, pp. 603–623, eds Taylor, B., Fujioka, K. *et al.,* College Station, TX.

Pezard, P.A., Hiscott, R.N., Lovell, M.A., Collela, A. & Malinverno, A., 1992b. Evolution of the Izu-Bonin intraoceanic forearc basin, western Pacific, from cores and FMS images, in *Geological Applications of Wireline Logs II,* eds Hurst, A., Griffiths, C.M. & Worthington, P.F., *Geol. Soc. Lond. Spec. Publ.,* **65,** 43–69.

Pratson, E.L., Reynolds, R., Lovell, M.A., Pezard, P.A. & Broglia, C., 1992. Geochemical well logs in the Izu-Bonin arc-trench system, sites 791, 792 and 793, in *Proc. ODP. Sci. Results,* Vol. 126, pp. 653–676, eds Taylor, B., Fujioka, K. *et al.,* College Station, TX.

Rao, C.R., 1952. *Statistical Methods in Biometric Research,* John Wiley, New York.

Rao, C.R. & Slater, P., 1949. Multivariate analysis applied to the difference between neurotic groups, *British J. Psycho. (Stat. Sect.),* **2,** 17–29.

Rogers, S.J., Fang, J.H., Karr, C.L. & Stanley, D.A, 1992. Determination of lithology from well logs using a neural network, *Am. Assoc. Petrol. Geol. Bull.,* **76,** 731–739.

Rumelhart, D.H., Hinton, G.E. & Williams, R.J., 1986. Learning representation by back-propagating errors, *Nature,* **323,** 533–536.

Silverman, B.W., 1986. *Density Estimation for Statistics and Data Analysis,* Chapman & Hall, London.

Sparks, D.N., 1973. Euclidean cluster analysis, *Appl. Stat.,* **22,** 126–130.

Taylor, B., Fujioka, A. *et al.,* 1990. *Proc. Ocean Drilling Program, Initial Results,* Vol. 126, College Station, Texas.

Wadge, G., Benaouda, D., Ferrier, G., Whitmarsh, R.B., Rothwell, R.G. & MacLeod, C., 1998. Lithological classification within ODP holes using neural networks trained from integrated core-log data, in *Core-Log Integration,* eds Harvey, P.K. & Lovell, M.A., *Geol. Soc. Lond. Spec. Publ.,* **136,** 129–140.

Wong, P.M., Jian, F.X. & Taggart, I.J. 1995. A critical comparison of neural networks and discriminant analysis in lithofacies, porosity and its permeability predictions, *J. petrol. Geol.,* **18,** 191–206.