

Lithology discrimination from physical rock properties

Miguel Bosch*, Maria Zamora‡, and Widya Utama‡

ABSTRACT

The estimation of lithology from multiple geophysical survey methods needs to be addressed to develop advanced tomographic methods. An initial requirement for lithology discrimination is that lithology should be discriminable from the media properties physically related to the geophysical observations. To test this condition for different combinations of the most common crustal rocks, we performed several lithology discrimination exercises on rock samples under laboratory conditions. The physical properties included mass density, compressional velocity, shear velocity, electric conductivity, thermal conductivity, and magnetic susceptibility. A categorical description of the sample lithology was followed; hence, the inference consisted of predicting the sample rock category (lithotype) membership. The

joint information provided by the physical properties of the rocks allowed us to discriminate the sample lithotype correctly, with an overall success rate of 100% in the most favorable situation and over 85% in the least favorable situation. We obtained successful classification results for a variety of common lithotypes (granite, gabbro, limestone, tuff, marble, basalt, and gneiss) using three common classification methods: clustering analysis, Gaussian classification, and discriminant analysis. Although discrimination was positive with each of these multivariate classification techniques, discriminant analysis showed some advantages for the classification and graphic analysis of the data. These results support our postulate that lithology can be estimated reliably if multiple geophysical observations are considered jointly.

INTRODUCTION

Physical rock properties are macroscopic parameters caused by the mineralogical composition, texture, and genesis of the rocks being measured. The influence of lithology on the physical behavior of rocks is well recognized, and a large amount of work has been devoted to its study. However, the inverse problem of estimating rock lithology from (several) rock physical properties has been less frequently addressed, despite its practical importance. One may postulate that, if we have sufficient information about the macroscopic physical behavior of the rock, we should be able to infer the lithology. Moreover, this task would be simplified if we had prior knowledge of the possible types of rocks in a particular region. Hence, the problem involves discriminating among the various rock types.

A background to this study is the work developed in the domain of well-log interpretation, where lithologic discrimination and classification methods have been extensively developed. Several multivariate statistical methods are routinely applied when analyzing well-log data. Most common among

these approaches are clustering analysis (Serra and Abbott, 1980) and Gaussian classification (Delfiner et al., 1984). Less frequently, discriminant analysis has been applied to well-log data (Sakurai and Melvin, 1988; Eberle, 1992). The logs used as parameters to estimate the lithology are, among others, density, neutron, sonic, resistivity, gamma ray, self-potential, photoelectric cross-section, and element (thorium, potassium, uranium) concentrations. These logs provide very detailed information about the logged media. In this work, we will study the possibility of discriminating lithology from a different kind of parameter: rock physical properties that can be estimated away from the well from geophysical tomographic methods, such as mass density, magnetic susceptibility, compressional velocity, shear velocity, and electrical conductivity.

The work is motivated by the search for advanced tomographic methods to infer the lithologic structure of earth media from observations provided by multiple geophysical techniques (Bosch, 1999). Conventionally, tomography consists of inverting a particular geophysical data set to estimate the physically related property. For example, we can obtain

Manuscript received by the Editor November 30, 1998; revised manuscript received November 20, 2000.

*Universidad Central de Venezuela, Facultad de Ingeniería, Departamento de Física Aplicada, Los Chaguaramos, Caracas, Venezuela. E-mail: mbosch@reaccun.ve; bosch@esc.cam.ac.uk.

‡Institut de Physique du Globe de Paris, Laboratoire des Géomatériaux, 4 place Jussieu, F-75252 Paris Cedex 05, France. E-mail: zamora@ipgp.jussieu.fr; utama@ipgp.jussieu.fr.

© 2002 Society of Exploration Geophysicists. All rights reserved.

compressional velocity from seismic arrival times (Bosch, 1997), conductivity from electromagnetic data (Newman and Alumbaugh, 1997), mass density from gravity data (Camacho et al., 1997), and magnetic susceptibility from magnetic data (Li and Oldenburg, 1996). But none of these estimated images is sufficient for lithology interpretation. From the scale of the globe to the scale of the reservoir or the ore body, interpretation is based on multidisciplinary evidence. Interpretation teams must match different images provided by independent geophysical techniques into a lithologic representation, consistent with the petrophysics and the geology of the region. In this interpretation process, either qualitative or quantitative, the problem of inferring lithology from multiple estimated physical rock parameters is obviously posed.

We focus on a required condition for lithology inference. Suppose we have precise knowledge of several physical properties of a rock sample and prior knowledge that the sample belongs to one of various lithologic groups. Can we successfully predict the group of the sample? To address this problem experimentally, we applied three conventional classification methods to a data set characterizing various common types of rocks. The data set contained laboratory measurements of several physical rock properties. This study is needed for further estimation of the lithology from multiple geophysical data and is required to support the development of the appropriate inversion methodologies.

Theory and methodology for lithologic tomography from multiple geophysical data sets has been considered in the work by Bosch (1999), and a field case has been presented by Bosch (1998) and by Bosch et al. (2001). Also, several inversion methods have been described for estimating the lithology from seismic data alone (e.g., Lortzer and Berkhout, 1992; Fichtl et al., 1997; Torres-Verdin et al., 1999). In both cases, the statistical description of the rock physical properties is incorporated into the inversion process. For lithology estimation purposes, the collection and analysis of petrophysical data is relevant and should be an integral component of geophysical exploration planning and geophysical inversion. With this in mind, the classification techniques used in the present work have the added value of being useful tools for (1) the multivariate statistical description of the dependence between the rock properties and the lithologic groups and for (2) the choice of the set of properties and geophysical data appropriate for lithology discrimination in a particular exploration scenario.

The following sections present a brief description of the three classification methods considered, as well as a description of the petrophysical data. They are followed by the presentation of the results, a discussion of these results, and our conclusions.

DATA AND METHODS

Three common methods for lithology classification of well logs were applied to our data sets: clustering analysis, Gaussian classification, and discriminant analysis. They all are based on multivariate statistics, and they all aim to define or identify groups of samples from some set of observations. The observations we used are the property values measured in the laboratory for each rock sample, after suitable transformation and standardization. Considering N_p properties in the database, we represent the property observations by a vector in an N_p -dimensional space, $\mathbf{z} = \{z_1, z_2, \dots, z_{N_p}\}$. The components

are the transformed standardized values of each property. This N_p -dimensional space is called the property space.

For each data set, the samples in our problem belong to various rock types (lithotypes) that we describe by a categorical variable G_i , with $i = 1, 2, \dots, N_g$. Here, the lithologic space is the discrete space of the N_g possible values of the lithotype.

Rock samples and properties

Two different data sets were considered for the discrimination tests. Our first data set, A, was integrated with samples representing four lithotypes: gabbro, limestone, marble, and sedimentary tuff. For this set of samples, we had laboratory measurements of five physical properties: mass density, thermal conductivity, compressional velocity, shear velocity, and electrical conductivity. To be as close as possible to real conditions, the rock samples were saturated with water. For the electrical conductivity measurements, the water conductivity was 5.27 S/m. This value is centered within the estimated range of electrical conductivity for crustal fluids (Nesbitt, 1993).

Lithotypes in data set A belong to the three major rock families—igneous, metamorphic, and sedimentary—and they are very different types of rocks, which represents a favorable case for discrimination. To test the discrimination ability of the techniques in a less favorable situation, a second data set, B, included more closely related types of rocks. It was integrated by samples representing four lithotypes: granite, gneiss, gabbro, and basalt. In addition to the five properties considered for data set A, we also had measurements of the magnetic susceptibility for data set B.

The rock samples came from various regions. The igneous and metamorphic rocks were taken from several drillholes in the Superior Province, Greenville Province, and Appalachian Mountains in North America (Mareschal et al., 1989; Pinet et al., 1991; Guillou et al., 1994; Guillou et al., 1995). The volcanic rocks (tuffs) originated from Champs Phelegreens in Italy (Zamora et al., 1994b; Yven, 1996). The sedimentary rocks came from different French quarries (Zamora et al., 1994a; Cattin, 1993). Within the same lithotype, the samples come from different sites or well depths, so they reasonably represent variations of composition and texture.

For the results shown in this work, the logarithm of the property value is used as the variable characterizing the sample. This transformation proved to be convenient to better represent (normalize) the property variability in our data sets. Additionally, there are theoretical arguments for the use of the logarithmic transformation. These properties are defined over the real positive semiaxis; hence, they are not normally distributed. After the logarithmic transformation, the data were standardized to have the same variance for each of the properties. These operations are generalized throughout the work. For simplicity we call properties the transformed and standardized values of the laboratory measurements.

Clustering analysis

Given a set of samples, clustering analysis is a method for constructing groups based on the proximity of the samples in the property space. The construction of these groups is based only on the properties and does not take into account information about sample membership to a particular lithologic

group. This differs from the other two methods, which use a training population to construct the statistical model for the classification.

Starting with a notion of distance in the property space (commonly Euclidean), there are several ways to form clusters of cases. Here we used hierarchical agglomeration, a common approach in which clusters are formed by grouping samples into bigger and bigger clusters until all cases are members of a single cluster. For the starting point, all cases are considered separate clusters. To decide which clusters should be combined at each step, various criteria of cluster proximity can be used. These include the average distance between samples within the resulting cluster (average linkage within groups), the smallest distance between samples across the clusters (nearest neighbor), and the distance between the centroids (centroid method). The two most proximate clusters, among all possible combinations, are linked.

The result of the clustering analysis can be visualized by representing the linkages and the combination distance with a display called a dendrogram. A review of clustering techniques can be found in the work of Gnanadesikan et al. (1989). Different applications to well-log data can be found in the work of Serra and Abbot (1980) and Gill et al. (1993).

Gaussian classification

Whereas clustering analysis constructs groups of similar data, other classification methods predict the membership of a sample from a well-known set of possible groups.

The Gaussian classification method is based on constructing a statistical model that relates property space with lithologic space. Within each lithologic group, the property vector \mathbf{z} is assumed to have an n -variate Gaussian distribution with conditional probability density function (PDF), denoted as $f(\mathbf{z} | G_i)$. The centroid and covariance matrix of the Gaussian density are inferred from a training population of samples that should be available for each lithotype.

To predict the group membership of a sample, the conditional $P(G_i | \mathbf{z})$ is needed. It is calculated using the inference rule,

$$P(G_i | \mathbf{z}) = \frac{f(\mathbf{z} | G_i)P(G_i)}{\sum_j P(\mathbf{z} | G_j)P(G_j)}, \quad (1)$$

Where $P(G_i)$ denotes the prior probabilities for each group. With the Gaussian model for the conditional PDF and using equal prior probabilities for each group,

$$P(G_i | \mathbf{z}) \propto \exp\left[-\frac{1}{2}(\mathbf{z} - \bar{\mathbf{z}}_i)' \mathbf{C}_i^{-1}(\mathbf{z} - \bar{\mathbf{z}}_i)\right], \quad (2)$$

with $\bar{\mathbf{z}}_i$ and \mathbf{C}_i^{-1} being the centroid and the covariance matrix inferred for the i th lithotype. The predicted lithotype is taken to be the one that maximizes $P(G_i | \mathbf{z})$.

Application of this method to classifying lithotypes from well-log data can be found in the work of Delfiner et al. (1984) and King (1990).

To test the performance of the classification, the membership of the samples in the training data set is predicted and compared with the true membership. When the same data set is used for training (calculating centroids and covariance matrix) and for testing (predicting) the classification, the procedure yields to the commonly called apparent error rate. To test

the actual performance of the classification over new samples, we use the jackknife sorting method. It consists of (1) omitting a test sample from the training data set used to calculate the statistical classification model, (2) classifying the test sample, and (3) repeating the procedure for all samples. The method provides a good estimate of the error rate (Gnanadesikan et al., 1989). Obviously, as the size of the population increases, both error estimates get closer.

Discriminant analysis

Discriminant analysis looks for functions of the properties that optimally (in a least-squares sense) separate the groups. These discriminant functions are linear functions of the properties and are calculated from a training data set of samples with known membership.

To introduce the discriminant analysis problem, consider a linear function of the data z_{pn} ,

$$d(\mathbf{z}_n) = \sum_{p=1}^{N_p} \alpha_p z_{pn}, \quad (3)$$

where α_p are the coefficients defining the linear function and the indices p and n identify the property and the case, respectively. The purpose of the method is to select a combination of linear coefficients that maximize the dispersion of group centroids yet minimize the dispersion of cases within groups. More precisely, discriminant analysis looks for linear functions that maximize the ratio between the former and the latter dispersions.

Figure 1 illustrates the problem with two discriminating variables, z_1 and z_2 , and two lithotype groups. It is clear in this

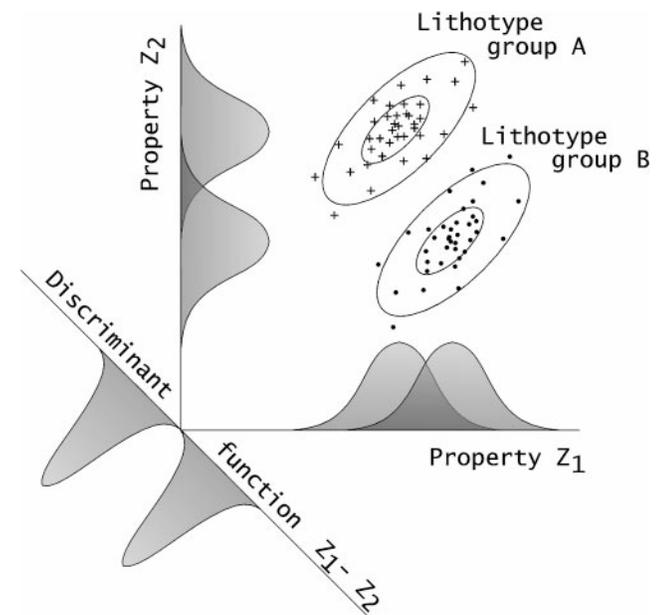


FIG. 1. Distribution of samples from two lithotype groups in a space of two physical properties z_1 and z_2 . The marginal probability densities for each group are represented near to the axes, showing significant overlap between the groups for both z_1 and z_2 . The group separation is improved for the discriminant function $z_1 - z_2$. Among all possible linear combinations of the original variables, this particular combination optimizes the group separation.

simple case that the function $d = z_1 - z_2$ optimizes the group separation. When more than two groups and more than two variables are involved, as in the present work, more than one discriminant function is obtained by solving the optimization problem. These functions are linearly independent.

The discriminant function coefficients are the eigenvectors of a matrix depending on the data covariances. Namely, they are the eigenvectors of $\mathbf{W}^{-1}\mathbf{B}$, with \mathbf{W} being the within-groups covariance matrix and \mathbf{B} being the between-groups covariance matrix. The discriminant functions are ranked according to the magnitude of the associated eigenvalue, which is related to the discriminating ability of the function. The number of discriminant functions depends on the number of groups N_g and the number of properties N_p ; there are no more than $N_f = \min(N_g - 1, N_p)$ discriminant functions. For a review on the mathematics of discriminant analysis, see, for instance, the work of Gnanadesikan et al. (1989). Applications of discriminant analysis to well-log data interpretation can be found in the work of Sakurai and Melvin (1988), Eberle (1992), and Wong et al. (1995).

The discriminant functions are used as an auxiliary coordinate system to represent data—the discriminant space. Common applications of discriminant analysis are graphic representation of multivariate data and classification. Classification of samples is performed in the discriminant space, following the same procedure as already explained for Gaussian classification. The only difference between the two methods consists of whether the classification procedure is performed in the property space or in the discriminant space. Discriminant analysis facilitates the visual representation of multivariate data. Commonly, the two or three first discriminant functions are used for graphical description and data analysis. This is useful because, in the property space, visual analysis involves the examination of a large number of crossplots.

RESULTS FOR DATA SET A

Figure 2 shows the property crossplot matrix for data set A. Several properties are significantly correlated in this data set: compressional velocity, shear velocity, and mass density. The information provided by these three correlated properties seems useful to distinguish between sedimentary lithotypes (limestone and tuff) and more competent rocks (gabbro and marble). Additional information is provided by the thermal conductivity, which seems to be useful for separating marbles from gabbros and, with less emphasis, tuffs from limestones. The electric conductivity is anticorrelated with the rest of the properties, and the plots show an intermediate behavior between the two types of information already mentioned.

Clustering analysis was performed with this data set using a Euclidean distance in the property space and the average distance between groups as the agglomeration criterion. A dendrogram of the links within groups and their hierarchy is shown in Figure 3. This figure depicts a pure branch of gabbros, a marble branch with one gabbro impurity, a predominantly limestone branch with two tuffs, and a mixed branch with five tuffs linked with a smaller branch of four limestones. This mixed cluster can be identified at the bottom of some of the crossplots of Figure 2 (for the density, compressional, and shear velocities).

As said before, the clustering procedure does not use information about sample membership. In Gaussian classification

and discriminant analysis, this information is used to construct, from the training population, a statistical model for classification. Table 1 shows the classification matrix obtained after applying the Gaussian classification procedure to data set A. The classification has been tested over all samples of the training population. An accuracy of 95% is obtained in the classification. The classification matrix using the jackknife sorting method is shown in Table 2, presenting an 87% overall success rate. The jackknife success rate is higher for the gabbro and limestone groups, which have the largest number of samples. When the number of samples is large, the jackknife rates approach the apparent rates obtained by using the training samples, as in Table 1. For the lithotypes with a small number of samples (marble and tuff), the covariance estimation is very sensitive to eliminating a single sample from the training data set, producing an artificially pessimistic jackknife error rate.

Discriminant analysis was performed on data set A, producing three discriminant functions. Table 3 shows the coefficients of the discriminant functions as well as the correlations between the properties and the discriminant function values. The hierarchy of the functions corresponds to their significance in the discrimination. Functions 1, 2, and 3 represented 81%, 17%, and 2% of the between-group variance, respectively. The crossplot of the samples for the two first discriminant functions is shown in Figure 4. Good separation of the four lithotypes can be seen in the discriminant space.

The prediction of the sample membership based on the three discriminant functions is shown in Table 4. A high success rate of 97% is obtained. To test the actual classification over samples outside the training population, the jackknife sorting method was performed. This implies recalculating the discriminant functions and the statistical model to test each sample omitted

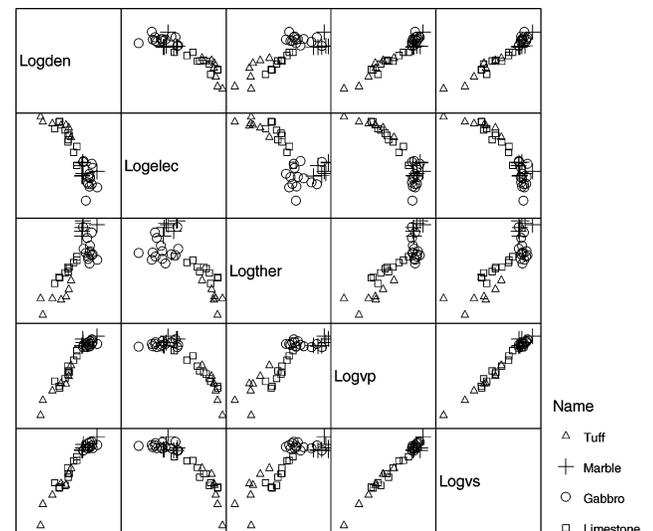


FIG. 2. Property crossplot matrix for data set A. The properties in the data set are logarithm of the density (Logden), logarithm of the electric conductivity (Logelec), logarithm of the thermal conductivity (Logther), logarithm of the compressional velocity (Logvp), and logarithm of the shear velocity (Logvs). The lithotypes represented in the data set A are limestone (11 samples), gabbro (14 samples), marble (7 samples), and tuff (7 samples). Each cell shows the plot for the two properties annotated at the diagonal. Plots that are symmetric by the diagonal present the same information with permuted axes.

from the training population. The classification matrix for the jackknife method is shown in Table 5; the method produced an 87% success rate. These classification results were very similar to those obtained with the Gaussian classification method.

RESULTS FOR DATA SET B

The property crossplot matrix for data set B is shown in Figure 5. Gabbro samples represented in data set B are the same as in data set A. For Figure 5, gabbro samples seem to be more dispersed. This is an effect of zooming in the property

Table 1. Results of the Gaussian classification for data set A.

Actual group	Predicted group membership			
	Tuff	Limestone	Marble	Gabbro
Tuff	7			
Limestone	1	10		
Marble			7	
Gabbro			1	13

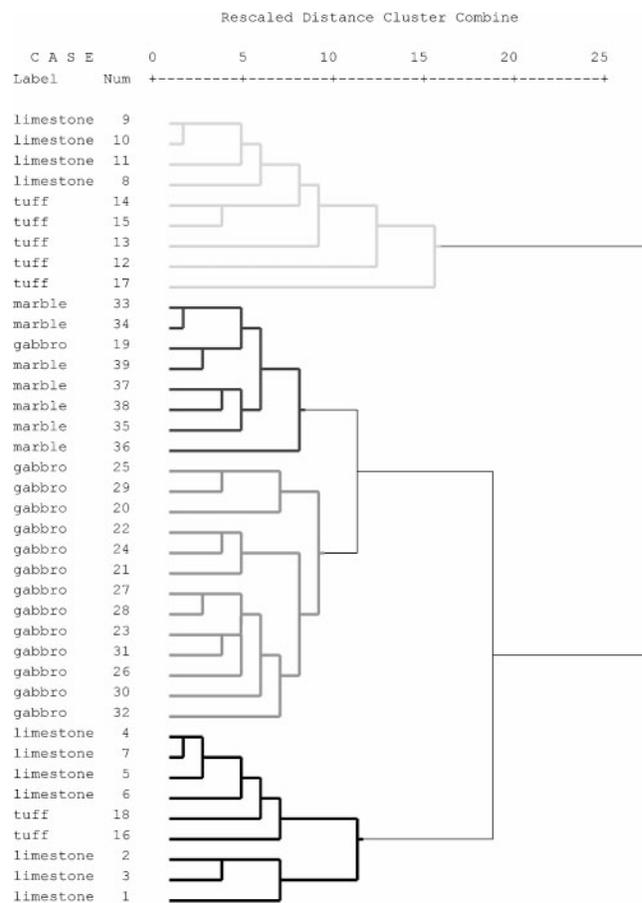


FIG. 3. Clustering dendrogram for data set A. Cases are labeled by the true lithotype of the sample and a sample number. Solid lines indicate the links between the cases created by the agglomeration procedure. The horizontal scale represents the distance in the property space between the combined groups. At the beginning of the procedure the cases belong, one by one, to single-element groups. The group agglomeration progresses from left (linking the closer groups) to right (linking the more separate groups). Gray levels in lines highlight branches showing a predominant lithotype.

space. Data set B is concentrated over a smaller region of the property space, and samples across lithotypes are closer than in data set A. Overall correlation between the compressional velocity, the shear velocity, and the density is also seen, but it is not as strong as for data set A. Each property seems to provide some independent information. Gabbro samples are in the upper range of the compressional velocity, shear velocity, mass density, and magnetic susceptibility. Gneiss samples are in the lower range of the compressional velocity, shear velocity, thermal conductivity, and mass density. These two groups are the most distant. Granite samples show lower density and lower magnetic susceptibility than basalt samples. In several crossplots granite samples are in between gneiss and basalt, and basalt samples are in between granites and gabbros.

Results of the clustering analysis are shown in Figure 6. The procedure was performed using a Euclidean distance and the average distance within groups criterion, as in Figure 3.

Table 2. Results of the Gaussian classification for data set A with jackknife sorting.

Actual group	Predicted group membership			
	Tuff	Limestone	Marble	Gabbro
Tuff	6	1		
Limestone	1	10		
Marble			5	2
Gabbro			1	13

Table 3. Discriminant functions for data set A, and function correlation with the discriminating variables.

Property name	Coefficients			Correlation		
	Func 1	Func 2	Func 3	Func 1	Func 2	Func 3
Logther	0.68	0.75	0.13	0.85	0.46	0.17
Logelec	-0.53	0.61	0.56	-0.70	0.55	0.04
Logvs	-0.64	0.80	3.08	0.60	-0.24	0.55
Logvp	0.46	0.01	-2.67	0.67	-0.19	0.34
Logden	0.19	-1.15	0.34	0.60	-0.44	0.51

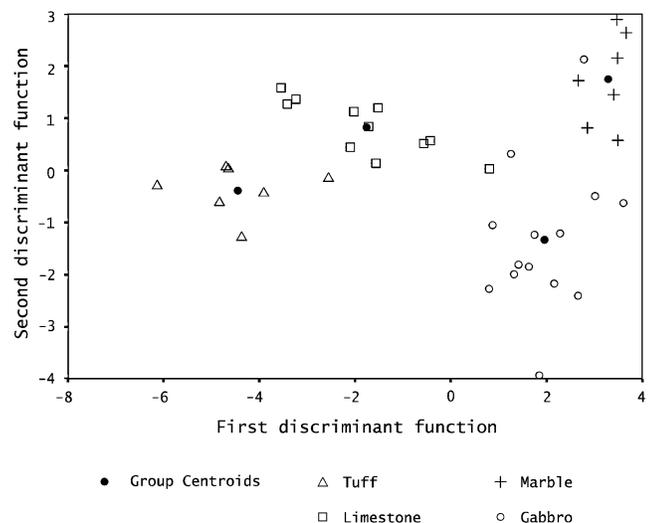


FIG. 4. Plot of cases in data set A in the space formed by the first and the second discriminant functions. The cases are labeled by the true lithotype.

Branches of the dendrogram were integrated mostly by samples of the same lithotype. Only one basalt case and one gneiss case were linked outside its lithotype group. The agglomeration distance between granite and basalt branches is shorter than the distance between the gneiss group and the gabbro group.

The predicted lithotypes produced by the Gaussian classification procedure are shown in Table 6 for samples in the training population. The overall classification success is 100%; no errors in the classification were present. The jackknife classification results are shown in Table 7 with an overall success rate of 85%. Most groups in this data set were represented by

Table 4. Results of the discriminant analysis classification for data set A.

Actual group	Predicted group membership			
	Tuff	Limestone	Marble	Gabbro
Tuff	7			
Limestone		11		
Marble			7	
Gabbro			1	13

Table 5. Results for the discriminant analysis classification of data set A with jackknife sorting.

Actual group	Predicted group membership			
	Tuff	Limestone	Marble	Gabbro
Tuff	6	1		
Limestone		11		
marble			5	2
Gabbro		1	1	12

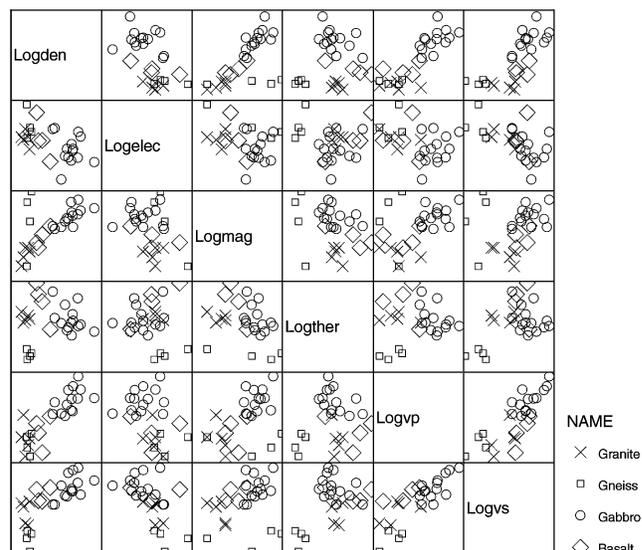


FIG. 5. Property crossplot matrix for data set B. The properties in the data set are logarithm of the density (Logden), logarithm of the electric conductivity (Logelec), logarithm of the magnetic susceptibility (Logmag), logarithm of the thermal conductivity (Logther), logarithm of the compressional velocity (Logvp), and logarithm of the shear velocity (Logvs). The lithotypes represented in the data set A are gabbro (14 samples), granite (5 samples), gneiss (4 samples), and basalt (4 samples). Each cell shows the plot for the two properties annotated at the diagonal. Plots that are symmetric by the diagonal present the same information with permuted axes.

a small number of samples; consequently, the jackknife success rate is quite conservative. Nevertheless, 85% is still a good rate when compared with the 25% rate that would be expected with a blind classification of the cases in the four groups.

The discriminant analysis performed with data set B provided three discriminant functions (see Table 8). The first function was the strongest correlated with the mass density, the shear velocity, and the compressional velocity; the second was the strongest correlated with the thermal conductivity; and the third was the strongest correlated with the magnetic susceptibility and the electric conductivity. The first function represented 78% of the between-group variance, the second represented the other 19%, and the third represented 2%. The crossplot of samples for the first two discriminant functions is shown in Figure 7. Note a large separation between the gneiss and gabbro. The closest samples across groups belong to basalt and granite. The granite lithotype is between the basalt and gneiss lithotypes. The basalt lithotype is closer to gabbro and granite lithotypes than to the gneiss lithotype. These relations are consistent with the description provided by the crossplots and the clustering analysis.

The classification of samples based on the three discriminant functions provided by the analysis is shown in Table 9.

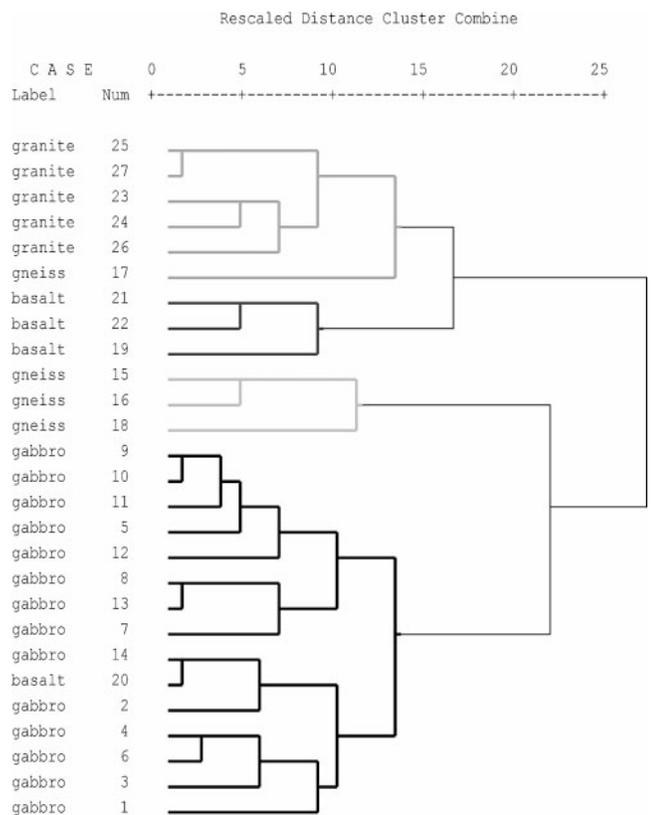


FIG. 6. Clustering dendrogram for data set B. Cases are labeled by the true lithotype of the sample and a sample number. Solid lines indicate the links between the cases created by the agglomeration procedure. The horizontal scale represents the distance in the property space between the combined groups. At the beginning of the procedure, the cases belong, one by one, to single-element groups. The group agglomeration progresses from left (linking the closer groups) to right (linking the more separate groups). Gray levels in lines highlight branches showing a predominant lithotype.

The success rate is 100%. To test lithotype prediction on samples outside the training population, the jackknife method was applied. The classification matrix for the jackknife method is shown in Table 10. Only one basalt, one gneiss, and one gabbro were misclassified—the same samples already misplaced in the clustering analysis. The overall jackknife success rate was 88% for data set B—a good result. Moreover, the jackknife

rate is a conservative estimate because various rock groups are represented by few cases.

DISCUSSION

We have used three different multivariate classification techniques to show that the sample lithotypes, represented in data sets A and B, could be successfully discriminated from the physical properties considered in this work. Data set A was dispersed over a large region of the property space, and data set B was dispersed over a smaller one (more similar lithotypes). In both cases the sample lithology was discriminated successfully. The lithotypes considered were common types of rocks. The results presented here support the idea that samples significantly different in a lithologic sense are also significantly different in the sense of the physical properties considered here.

Because of the correlations between properties, the property relevance in the multivariate discrimination is relative to the other properties included in a data set. For the same reason, it is unreliable to interpret the discriminant function coefficients as a quantitative measure of a property's importance in the discrimination. Selecting convenient property combinations is a better posed problem than evaluating the relevance of individual properties. Hence, the different property subsets could be ranked easily by their success rate in the classification.

As an example with data set A, the two less correlated variables—electrical and thermal conductivity—were used as a smaller data set for the discriminant analysis. The variables also have important correlations with the first and second discriminant functions. These two properties were taken as a smaller set of variables for performing the discriminant analysis, and they provided a 92% success rate in the posterior classification of samples—acceptable if compared with the 97% success rate obtained with all the properties. Data set B also used a smaller set of properties that contained enough information about the sample's lithology. Discriminant analysis based on mass density, shear velocity, and thermal conductivity data produced a 100% success rate. The same result was obtained with the full-variable data set.

An additional application of the type of analysis presented here would be to identify the most convenient property

Table 6. Results of the Gaussian classification for data set B.

Actual group	Predicted group membership			
	Gabbro	Basalt	Granite	Gneiss
Gabbro	14			
Basalt		4		
Granite			5	
Gneiss				4

Table 7. Results of the Gaussian classification for data set B with jackknife sorting.

Actual group	Predicted group membership			
	Gabbro	Basalt	Granite	Gneiss
Gabbro	14			
Basalt	2	1	1	
Granite	1		4	
Gneiss				4

Table 8. Discriminant functions for data set B and function correlation with the discriminating variables.

Property name	Coefficients			Correlation		
	Func 1	Func 2	Func 3	Func 1	Func 2	Func 3
Logden	0.87	0.38	0.14	0.67	0.49	0.18
Logvs	-0.07	-1.02	0.82	0.49	-0.24	0.02
Logvp	0.51	0.67	-1.07	0.42	0.26	-0.27
Logther	0.89	-0.39	-0.01	0.22	-0.58	0.16
Logmag	-0.08	0.33	0.76	0.15	0.50	0.55
Logelec	-0.39	0.05	0.96	-0.17	0.02	0.26

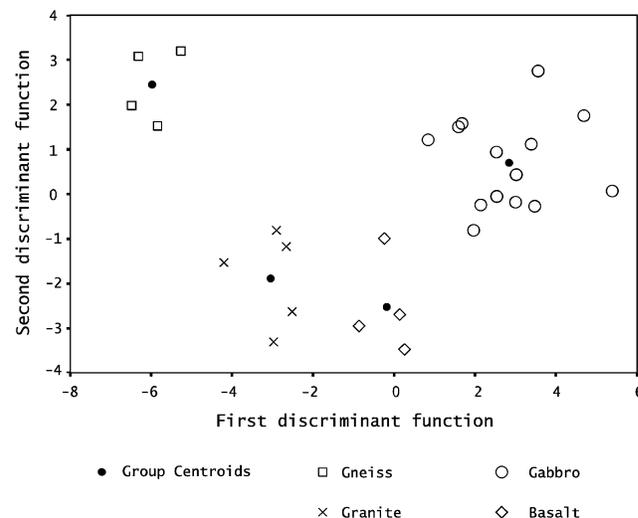


FIG. 7. Plot of samples in data set B in the space formed by the first and second discriminant functions. The cases are labeled by the true lithotype.

Table 9. Results of the discriminant analysis classification for data set B.

Actual group	Predicted group membership			
	Gabbro	Basalt	Granite	Gneiss
Gabbro	14			
Basalt		4		
Granite			5	
Gneiss				4

Table 10. Results for the discriminant analysis classification of data set B with jackknife sorting.

Actual group	Predicted group membership			
	Gabbro	Basalt	Granite	Gneiss
Gabbro	13	1		
Basalt	1	3		
Granite			5	
Gneiss			1	3

combinations for lithology discrimination, according to a particular set of lithotypes and samples from a study region. Identifying these properties would in turn imply the definition of the appropriate geophysical surveys for their estimation.

The properties in our data sets corresponded to water-saturated samples to better reproduce common in-situ conditions. For the electric conductivity measurement, the conductivity of the water was 5.27 S/m. This value is approximately in the middle of the range of estimated conductivity for common crustal fluids (Nesbitt, 1993). For practical reasons, our laboratory measurements did not reproduce all in-situ rock conditions: temperature, pressure, type of fluid, saturation, fluid conductivity. Nevertheless, we think that variations of these conditions introduce systematic changes in physical rock properties. This can be seen, for instance, in the work of Christensen and Mooney (1995) studying the variation on density, compressional velocity, and shear velocity with pressure and temperature for a variety of common crustal rocks. Our analysis showed that the variability of rock physical properties is often smaller within our lithotypes than across our lithotypes. If the variation in physical properties with rock conditions is mostly systematic, the statistical characteristics of the inference problem should not change drastically.

Models describing the lithologic structure for a region are likely to represent a larger scale of spatial resolution than the centimeter scale considered in our rock samples. As well, properties estimated from geophysical surveys represent averages over larger volumes of rock. For seismic surveys this spatial range is related with wavelength, which is typically around 50 m or more. Another interesting problem consists of extrapolating these discrimination results if larger rock volumes are considered.

The effect of changing the volume support for the property measurements has been studied in geostatistical work (e.g., Isaaks and Srivastava, 1989). Physical properties such as mass density or compressional velocity are additive. Increasing the volume of the rock block results in averaging the property inside a larger volume. Hence, the block property distributions have smaller variances than the property distributions of rock samples, whereas the mean values are the same. This effect, of course, favors the discrimination of lithologies because the physical properties characterizing the lithotype tend to be less scattered. Other properties, such as electrical conductivity, are not additive and follow more complicated regularizations. Nevertheless, in both cases, physical properties for larger block volumes of the same lithotype always result in less-variable property values than for small laboratory samples. Hence, the change for larger-scale estimation of the properties helps discriminate lithotypes.

Among the different classification techniques used, discriminant analysis can combine the influence of the properties to get an optimum discrimination. This operation cannot be obtained by a straightforward Gaussian classification from the property space. Discriminant analysis also provides a convenient framework for graphical analysis of the data by providing a discriminant space of fewer dimensions than the property space (usually no more than two or three discriminant functions are significant).

In this study, we represent lithology by categorical variables: lithologic groups of rocks. Lithology and rock structure in some cases can be described by continuous variables, such as silica

weight content or porosity. These cases are not considered in this work, but they are equally relevant. We believe lithologic continuous parameters can be inferred, as we have shown for lithotypes, from multiple physical rock properties. Different approaches to regression would provide, in this situation, the appropriate statistical methods for the inference, instead of the classification methods used here for categorical variables. Work in this line has been done in the domain of well-log interpretation (Mendelson and Toksöz, 1985; Sakurai and Melvin, 1988; Zimmermann et al., 1992).

We have considered physical measurements made on the rock samples under laboratory conditions. The more general problem of inferring the spatial lithologic structure of a region from geophysical survey data (Lortzer and Berkhout, 1992; Bosch, 1998) is of major interest in geophysical exploration. Geophysical techniques provide information about the physical properties inside a volume, but these estimates are affected by larger uncertainties than the laboratory measurements considered in our work. Nevertheless, our work represents a preliminary step toward the development of lithologic tomographic techniques. The possibility of estimating lithology from physical properties is required for estimating lithology from the geophysical data associated with these properties (e.g., gravity data for mass density, magnetic data for magnetic susceptibility, etc.). Roughly speaking, one could say that the uncertainties of estimating lithology from geophysical data would combine uncertainties from two levels: (1) uncertainties of estimating lithology from physical properties and (2) uncertainties of estimating the physical properties from the geophysical data.

Several methods have already been applied for discrimination of subsurface lithostratigraphy, combining attributes from seismic reflection sections and well-log data in the domain of oil reservoir description (Angeleri and Carpi, 1982; Sinval and Khattri, 1983; Doyen, 1988; Dumay and Fournier, 1988; Fournier and Derain, 1995). In these works, statistical relations are built between seismic traces attributes and reservoir lithofacies from a calibration population consisting of wells and their adjacent traces. The information is used to condition the interpolation of lithofacies between wells. Work by Fichtl et al. (1997) shows an interesting direction in this domain by establishing statistical relations between the lithofacies and the physical media properties inferred from the seismic data.

CONCLUSION

Several physical properties measured in rock samples were used to discriminate the sample lithology following three different multivariate classification methods: clustering analysis, Gaussian classification, and discriminant analysis. Several common lithotypes were considered. The physical properties used were mass density, compressional velocity, shear velocity, electrical conductivity, thermal conductivity, and magnetic susceptibility.

The results of the lithologic discrimination, based on the sample's physical properties, were positive for all three techniques considered. The techniques proved useful for analyzing multivariate data and presented consistent results among them. These results let us emphasize that the lithology can be estimated reliably if one has enough information about the physical behavior of the rocks. In particular, physical rock properties

that can be estimated away from the well by geophysical tomography are useful for lithology estimation if they are considered jointly.

The influence of each property in the discrimination depends on the particular rock types under consideration. In a particular area under exploration, a smaller set of relevant properties could be identified and used for the classification. In turn, this helps define the geophysical data to acquire in the area.

A few methodological advantages of discriminant analysis, compared with the other techniques, have been considered in this work. This may encourage its more extensive use in other domains, such as well-log interpretation.

ACKNOWLEDGMENTS

The authors want to specially acknowledge J. P. Pozzi (Ecole Normale Supérieure de Paris) for providing the measurements of the magnetic susceptibility for our data sets. C. Jaupart (Institut de Physique du Globe de Paris) and G. Bienfait (Institut de Physique du Globe de Paris) provided part of the rock samples. Discriminant analysis and clustering analysis were performed using the licensed software package SPSS. This research was partially funded by the TEKVOLC program (European Union), the GeoFrance-3D program (France), the CDCH of the Universidad Central de Venezuela, and the ITS of Surabaya-Indonesia. The authors acknowledge Andreas Ganps, an anonymous referee, and Associate Editor Keith Hirsche for their constructive reviews.

REFERENCES

- Angeleri, G. P., and Carpi, R., 1982, Porosity prediction from seismic data: *Geophys. Prosp.*, **30**, 580–607.
- Bosch, M., 1997, P-wave velocity tomography of the Venezuelan region from local arrival times: *J. Geophys. Res.*, **102**, 5455–5472.
- 1998, Lithologic tomography: Method and application to the Côte d'Armor region (France): GOCAD-ENSG Conference, Proceedings, **2**, 1–4.
- 1999, Lithologic tomography: From plural geophysical data to lithology estimation: *J. Geophys. Res.*, **104**, 749–766.
- Bosch, M., Guillen, A., and Ledru, P., 2001, Lithologic tomography: An application to geophysical data from the Cadomian belt of northern Brittany, France: *Tectonophysics*, **331**, 197–227.
- Camacho, A. G., Montesinos, F. G., and Vieira, R., 1997, A three-dimensional gravity inversion applied to Sao Miguel Island (Azores): *J. Geophys. Res.*, **102**, 7717–7730.
- Cattin, R., 1993, Relation entre les propriétés élastiques, thermiques et de transport des roches poreuses: Rapport du DEA en Géophysique, Institut de Physique du Globe de Paris.
- Christensen, N. I., and Mooney, W. E., 1995, Seismic velocity structure and composition of the continental crust: A global view: *J. Geophys. Res.*, **100**, 9761–9788.
- Delfiner, P., Serra, O., and Peyret, O., 1984, Automatic determination of lithology from well-logs: 59th Ann. Tech. Conf. and Exhib., Soc. Petrol. Eng., paper SPE 13290.
- Doyen, P. M., 1988, Porosity from seismic data: A geostatistical approach: *Geophysics*, **53**, 1263–1275.
- Dumay, J., and Fournier, F., 1988, Multivariate statistical analyses applied to seismic facies recognition: *Geophysics*, **53**, 1151–1159.
- Eberle, D., 1992, Interpretation of multi-method geophysical borehole logging data from KTB Oberpfalz VB using multivariate statistical analyses: *Scientific Drilling*, **3**, 16–26.
- Fichtl, P., Fournier, F., and Royer, J.-J., 1997, Cosimulation of lithofacies and associated properties using well and seismic data: 72nd Ann. Tech. Conf. and Exhib., Soc. Petrol. Eng., Proceedings, **1**, 381–393.
- Fournier, F., and Derain, J. F., 1995, A statistical methodology for deriving reservoir properties from seismic data: *Geophysics*, **60**, 1437–1450.
- Gnadesikan, R., Blashfield, R. K., Breiman, L., Dunn, O. J., Friedman, J. H., Fu, K., Hartigan, J. A., Kittenring, J. R., Lachenbruch, P. A., Olshen, R. A., and Rohlf, J., 1989, Discriminant analysis and clustering: *Statistical Science*, **4**, 34–69.
- Gill, D., Shomrony, A., and Fligelman, H., 1993, Numerical zonation of log suites and log facies recognition by multivariate clustering: *AAPG Bull.*, **77**, 1781–1791.
- Guillou, L., Mareschal, J. C., Jaupart, C., Cariépy, C., Bienfait, G., and Lapointe, R., 1994, Heat flow, gravity and structure of Atibi Belt, Superior Province, Canada: *Earth and Planet. Sci. Lett.*, **122**, 103–123.
- 1995, Heat flow variation in Grenville Province, Canada: *Earth and Planet. Sci. Lett.*, **136**, 447–460.
- Isaaks, E. D., and Srivastava, R. M., 1989, *Applied geostatistics*: Oxford Univ. Press.
- King, D. E., 1990, Incorporating geological data in well-log interpretation, in Hurst, A., Lovell, M. A., and Morton, A. C., Eds., *Geological applications of wireline logs*: *Geol. Soc. Special Publ.* **48**, 45–55.
- Li, Y., and Oldenburg, D. W., 1996, 3-D inversion of magnetic data: *Geophysics*, **61**, 394–408.
- Lortzer, G. J. M., and Berkhout, A. J., 1992, An integrated approach to lithologic inversion, part I—Theory: *Geophysics*, **57**, 233–244.
- Mareschal, J. C., Pinet, C., Gariépy, C., Jaupart, C., Bienfait, G., Dalla Coletta, G., Jolivet, J., and Lapointe, R., 1989, New heat flow density and radiogenic heat production data in the Canadian Shield and the Quebec Appalachians: *Can. J. Earth Sci.*, **26**, 845–852.
- Mendelson, J. D., and Toksöz, M. N., 1985, Source rock characterization using multivariate analysis of log data: 26th Ann. Logging Symp., Soc. Prof. Well Log Anal., Proceedings, 1–20.
- Nesbitt, B. E., 1993, Electrical resistivities of crustal fluids: *J. Geophys. Res.*, **98**, 4301–4310.
- Newman, G. A., and Alumbaugh, D. L., 1997, Three-dimensional massively parallel electromagnetic inversion—II: Analysis of a cross-well electromagnetic experiment: *Geophys. J. Internat.*, **128**, 355–363.
- Pinet, C., Jaupart, C., Mareschal, J.-C., Gariépy, C., Bienfait, G., and Lapointe, R., 1991, Heat flow and structure of the lithosphere in the eastern Canadian Shield: *J. Geophys. Res.*, **96**, 19941–19963.
- Sakurai, S., and Melvin, J., 1988, Facies discrimination and permeability estimation from well logs for the Endicott field: 29th Ann. Symp., Soc. Prof. Well Log Anal., Proceedings, F1–F18.
- Serra, O., and Abbot, H. T., 1980, The contribution of logging data to sedimentology and stratigraphy: 55th Ann. Tech. Conf. and Exhib., Soc. Petrol. Eng., paper SPE 9270.
- Sinvhal, A., and Khattri, K. N., 1983, Application of seismic reflection data to discriminate subsurface lithostratigraphy: *Geophysics*, **48**, 1498–1513.
- Torres-Verdin, C., Victoria, M., Merletti, G., and Pendrel, J., 1999, Trace-based and geostatistical inversion of 3-D seismic data for thin-sand delineation: An application in San Jorge basin, Argentina: *The Leading Edge*, **18**, 1070–1077.
- Wong, P. M., Jian, F. X., and Taggart, I. J., 1995, A critical comparison of neural networks and discriminant analysis in lithofacies, porosity and permeability predictions: *J. Petrol. Geol.*, **18**, 191–206.
- Yven, B., 1996, Propriétés physiques des roches volcaniques des Champs Phlégréens: Rapport du DEA en Géophysique, Institut de Physique du Globe de Paris.
- Zamora, M., Cattin, R., and Bienfait, G., 1994a, Empirical relationship between thermal conductivity and seismic velocities: Fall meeting, Am. Geophys. Union, T22D-7.
- Zamora, M., Sartoris, G., Chelini, W., 1994b, Laboratory measurements of ultrasonic wave velocities in rocks from the Campi Flegrei volcanic system and their relation to other field data: *J. Geophys. Res.*, **99**, 13553–13561.
- Zimmermann, G., Burkhardt, H., and Melchert, M., 1992, Estimation of porosity in crystalline rock by a multivariate statistical approach: *Scientific Drilling*, **3**, 27–35.