Zeitschrift: Schweizerische mineralogische und petrographische Mitteilungen = Bulletin suisse de minéralogie et pétrographie

Band: 77 (1997)

Heft: 3

Artikel: A new cluster analysis method for altered rock samples

Autor: Tóth, Tivadar M. / Engi, Martin

DOI: https://doi.org/10.5169/seals-58495

A new cluster analysis method for altered rock samples

by *Tivadar M. Tóth*[1,2] and *Martin Engi*[2]

Abstract

A variety of cluster analysis methods are commonly used by geochemists to answer genetic questions. Special problems may arise when one tries to group or classify metamorphic or highly altered samples. Various chemical elements may have been mobilized by a number of alteration processes, thus disturbing or even destroying the original pre-alteration group-structure.

This paper introduces a new hierarchical clustering procedure that is designed to be particularly useful in classifying altered rock samples. The basic objective of the method is to determine the pre-alteration structure of the groups. For this reason, during the analysis each pair of samples is compared element by element in order to decide their similarity. This procedure yields a specified, binary similarity matrix that is used to determine the groups of the samples. As with other hierarchical methods, the result of the new cluster analysis may be visualized with a dendrogram. Such a graph simplifies the evaluation of the group-structure. The homogeneity of each group as well as the heterogeneity between pairs of groups may be quantitatively measured by means of two newly introduced indices. These variables are helpful in interpreting geochemical data.

The procedure is demonstrated on a simple example of twelve historic lava samples from Mt. Etna.

Keywords: geochemical data analysis, cluster analysis, dichotomous function, chemical alteration, Mount Etna.

Introduction

Cluster analysis routines are frequently used to solve diverse classification problems in geochemistry. Because these methods can be applied to large sets of variables, they usually give more informative results than traditional discrimination diagrams. However, under special circumstances, e.g. when data from highly altered rocks are analyzed, specific problems of mathematical classification may arise. During various alteration processes, the concentrations of most chemical elements may change to different degrees (e.g. GRESENS, 1967; GRANT, 1986). For example, complex seawater-rock interactions take place in the early evolution of oceanic crust changing the chemical composition of the MOR-basalt due to Mg-, Ca-, and alkali-metasomatic reactions (SEYFRIED et al., 1988; ALT, 1995, and references therein). Wall-rock alteration processes also modify mineralogical and chemical composition of the host rock and yield ore deposits (BARNES, 1967; RIVERIN and HODGSON, 1980). Hydrothermal processes in shear zones often lead to pervasive changes in the concentration of many elements and composition of minerals (ALTENBERG, 1991). Many metamorphic processes are thought to run under open system conditions, i.e. gain and loss of different elements occurs (WOOD et al., 1976; CONDIE et al., 1977; SORENSEN and BARTON, 1987).

Due to these and other possible alteration processes, the composition of samples may change significantly, rendering their mathematical classification difficult. The aim of this paper is to discuss special problems that arise regarding the classification of metamorphic and altered rock samples and to introduce a new classification method that deals with these problems. This method has proven successful when applied to large sets of geochemical data (e.g. M. TÓTH, 1994). However, for illustrative purposes in this paper, only a small example is presented, consisting of geochemical data of historic lava samples from Mount Etna.

[1] Department of Mineralogy, Geochemistry und Petrology, Attila Jozsef University, Szeged, Hungary.
[2] Mineralogisch-petrographisches Institut, Universität Bern, Baltzerstrasse 1, CH-3012 Bern, Switzerland (corresponding author). E-mail: engi@mpi.unibe.ch.

Four samples from each of three eruptions (1865, 1911, 1974) were chosen for which trace element data exist (BARBIERI et al., 1993). Petrologically, all samples are hawaiites, but the rocks from 1974 are more mafic than others. Before the application of the new method, nine popular hierarchical cluster approaches were applied to classify the given samples. All combinations of complete linkage, centroid and Ward methods with Euclidean, cosine θ and Pearson correlation measures were used. None of these approaches could repeat the expected original structure; indeed these methods all more or less "mix" the samples.

Some comments on cluster analysis

While many textbooks introduce the mathematical background (e.g. ANDERBERG, 1973) of methods such as cluster analysis (CA), few of them refer to specific requirements and limits of their application in the Earth sciences. CA is a useful exploratory tool in the analysis of large, multivariate data sets, such as occur typically in geochemistry. Used most frequently in the geosciences are agglomerative CA methods; these recursively grow clusters of samples based on similarity between samples. Subsequent grouping of these clusters depends on the method chosen. For example, in the average linkage method, the similarity between clusters is determined using cluster centroids. This leads to a progressive averaging process, in which the multivariate characteristics of a cluster become less well defined as a cluster grows, accepting samples and other clusters of samples that "dilute" the original grouping.

In practice, the results are sometimes difficult to interpret; arbitrary boundaries between classes may be introduced. This need not be a flaw of the statistical method because, unlike in biology, samples in geochemistry typically do not constitute *a priori* classes. In many instances, it is useful to view rocks as transitional between two or more end-members, between which natural boundaries do not exist. In situations such as these, the classification approach may not be a suitable model. However, even where it is, conventional CA sometimes encounters difficulties, either because of the progressive averaging problem sketched out above, or because the data are multidimensional, yet the ultimate representation of the intersample relationships is reduced to a 2-D graph, a dendrogram (see below). A method is suggested in the present paper to circumroute some of the drawbacks of traditional CA, in that we use indicator transforms (GOOVAERTS, 1994) in the clustering problem.

However, we should stress that, as with all statistical analysis, it is the responsibility of the practitioner to ensure that one may expect CA to yield meaningful results when applied to the geochemical data in question. Where the significance of (geochemical) classes is not clear, the results of any classification approach may be difficult to interpret, no matter how good a CA method is used.

Statement of the problem

Let us assume that a suite of samples under investigation did at some earlier, pre-alteration stage show a geochemically recognizable group structure. One may then ask: how can this structure be recognized after the alteration?

Many studies on igneous and metamorphic geochemistry as well as soil chemistry deal with changes of element concentrations due to different alteration processes. It is widely recognized that some elements are mobile during many alteration processes (e.g. alkalies), while others may be referred to as immobile (e.g. rare earth elements). The main chemical and crystal structural factors controlling element mobility have also been known for decades (e.g. GOVETT, 1981). On the other hand, several studies (e.g. GAST, 1968; GIRARDI et al., 1986) show that, given certain geochemical circumstances, all elements may show mobile behaviour, and none of them may be judged completely immobile. Trace elements such as Zr, Y, Nb, Ti or P generally exhibit conservative behaviour during many alteration or metamorphic processes (PEARCE, 1975; GRAHAM, 1976; WOOD et al., 1976; COISH, 1977). On the other hand, under certain geochemical conditions (e.g. high activity of CO_2 or F^-), these elements may also be mobilized (MORRISON, 1978; PEARCE and NORRY, 1979, GIERÉ, 1986, 1990). These results suggest that the original group structure of the samples may be disturbed or destroyed by differential mobilization of chemical elements.

An obvious way to solve this problem would seem to be the elimination of the most mobile elements from the classification analysis. However, due to the different kinds of alteration processes, the elimination of almost all elements might theoretically be justified; few if any of them would remain suitable for use in classification. Worse, there is no generally accepted procedure to choose which elements remain suitable. Not only is the identification of relatively immobile elements a problem, but the extent of alteration may vary greatly from one sample to the next. Strongly altered samples preserve few or none of the geochemical characteristics of their original (i.e. pre-

metamorphic, pre-alteration) genetical group (cf. basalt-diabase alteration, spilitization, propylitization, rodingitization). Again, there is no general method to identify and exclude such samples. However, elimination of all suspicious samples and elements would lead to significant loss of information. Therefore, erasing variables from the data base cannot be the best procedure for dealing with the complex problem of mobility. A suitable cluster method should follow a different approach: The purpose must be to find the "essence" of the original group structure by determining the similarity of the sample pairs element by element. To determine the similarity between two samples, only those elements should be used which show immobile behaviour in both samples. A different set of elements may be necessary to classify different groups of samples. Samples with significantly changed composition must be eliminated during the analysis.

The algorithm

Let us take m (chemical) variables and n samples. Let us define the graphs $g_1, g_2, \ldots, g_m$ (one for each variable) as follows. Each graph has n vertices, representing the samples. In the case of the ith graph let two vertices be joined with a tie-line if the value of the ith variable measured in the two samples is close enough in value. (The expression "close enough" will be defined more precisely later.) Thus in the case of the graphs $g_1, g_2, \ldots, g_m$ those samples (vertices) would be joined which are similar to each other according to the 1st, 2nd,..., mth variables, respectively. The graph form is useful to visualize (Fig. 1) the first step of the algorithm; however, further steps are more easily described algebraically. Each graph can be represented by a symmetric, $n \cdot n$ type binary (each value is 0 or 1) matrix (A_k, k = 1,2,..., m). In the case of the kth variable, $A_k(i,j) = 1$ if there is a tie-line between

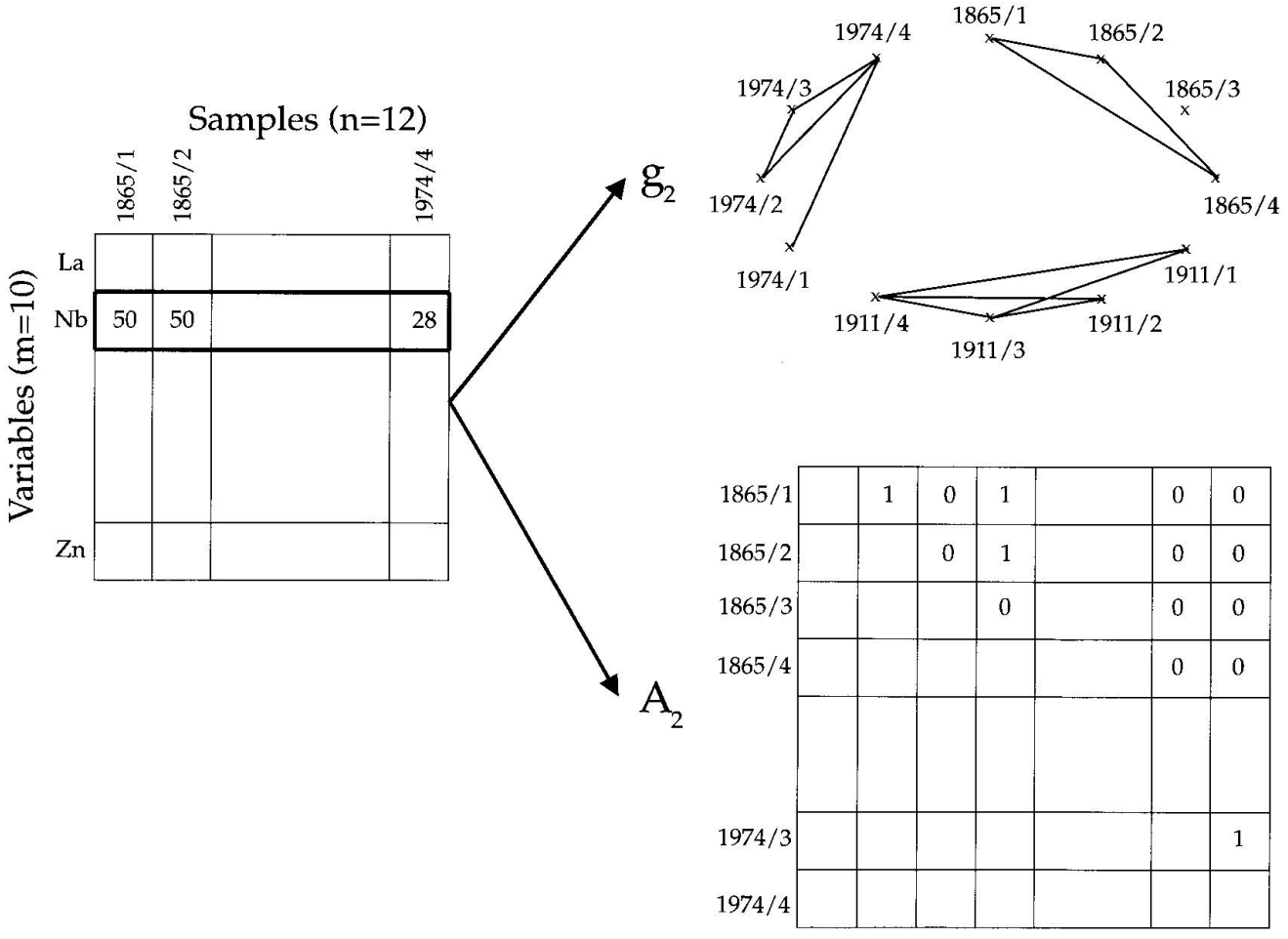

Fig. 1 Derivation of the g_i graphs and the A_i matrices from the original data matrix (in the case of 12 hawaiite lava samples of Mount Etna).

the *i*th and *j*th samples, otherwise $A_k(i,j) = 0$. Consequently, the first step of the algorithm yields a graph and as an equivalent form a binary matrix representing each row of the original data matrix (Fig. 1).

By summing up these matrices (or graphs), a new $n \cdot n$ matrix (A) is generated, where

$$A(i,j) = \Sigma A_k(i,j).$$

A(i,j) gives the number of variables which have similar values in samples *i* and *j*. The higher the value A(i,j), the more likely the *i*th and *j*th samples belong to the same genetic group. Hence let us define *m* different $n \cdot n$ matrices (B_Γ, Γ = 1,2,..., m) as follows:

$$B_\Gamma(i,j) = 1, \text{ if } A(i,j) \geq \Gamma, \text{ else } B_\Gamma(i,j) = 0$$

These matrices are binary and they determine G_1,..., G_m graphs. In these graphs two vertices are joined if the samples represented by them are similar to each other according to at least *1, 2,..., m* variables (Fig. 2). Let us call the Γ value in the above formula the *connection level.* If Γ is small, the G_Γ graph likely has many connections. However, as Γ increases, the graph must break up, forming smaller, independent groups. For any one of these groups, samples belonging to it are similar to each other in at least Γ variables, though the set of the variables may vary from one sample pair to another within this group. These samples are related to elements of other groups in fewer than Γ variables. The G_Γ graphs define clusters and yield a group-structure, one for each value of the connection level. Though the group-structure may be obtained in the presented way is not a hierarchical one by origin, by increasing (or decreasing) the Γ value step by step, the results may be visualized by common hierarchical forms (dendrograms).

To generate the g_i graphs, samples are connected that are "close enough" to each other. To define "close enough" precisely, a similarity measure is needed. However, unlike in commonly used agglomerative cluster analysis, only a uni-variate function is needed for the new method. An obvious definition for a uni-variate similarity

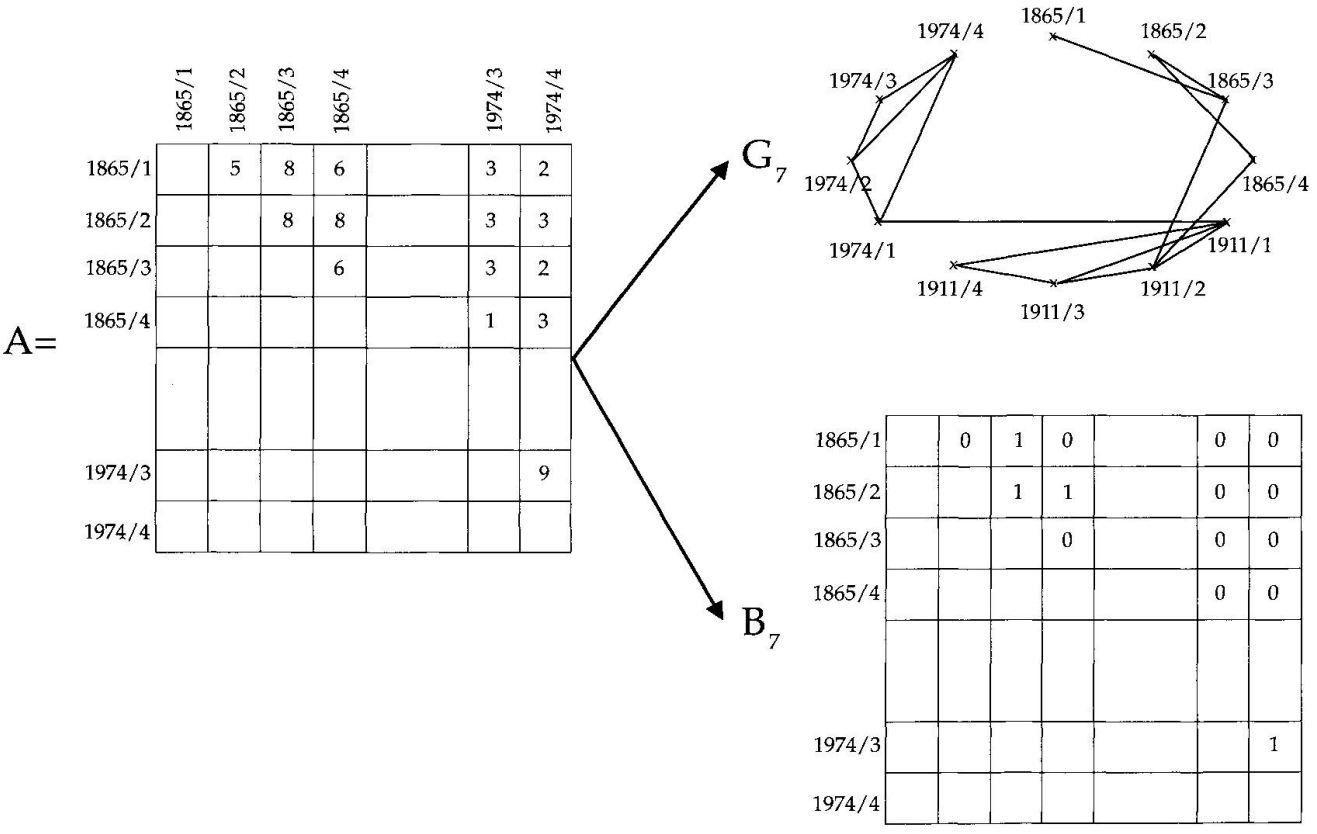

Fig. 2 Derivation of the B_i matrices as well as G_i graphs from the A matrix (in the case of 12 hawaiite lava samples of Mount Etna).

Tab. 1 Trace element content of the examined lava samples (BARBIERI et al., 1993).

Variables	1865/1	1865/2	1865/3	1865/4	1911/1	1911/2	1911/3	1911/4	1974/1	1974/2	1974/3	1974/4
La	67	82	82	78	66	74	67	69	73	65	61	67
Nb	50	50	57	51	36	41	40	37	31	26	24	28
Rb	47	44	41	40	37	35	31	26	43	47	48	45
Zr	229	249	229	264	220	249	233	211	203	215	180	194
Sr	1230	1160	1140	1180	1190	1210	1170	1140	1170	1150	1140	1140
Ni	30	22	24	24	31	27	31	33	50	54	51	47
Co	32	34	29	30	36	35	38	34	49	45	49	45
Cr	48	35	42	41	58	44	67	55	55	63	61	68
Cu	115	130	118	151	140	130	128	139	117	154	134	138
Zn	82	69	99	102	91	93	86	80	85	101	98	113

measure is the absolute value. Let us define the similarity of two samples according to a variable as the absolute value of their difference. Traditional cluster methods apply real numbers as similarity measure (e.g. Euclidean distance, cos θ). Here a dichotomous function is used, requiring that an absolute value function be transformed to an indicator function in order to create the g_i graphs. Dichotomy is the most essential feature of the method because it is used for eliminating disproportionately large differences between samples. Numerous possibilities may be suitable to define an indicator function. Here, we transform an absolute value function to a sequence of characteristic similarity functions (S_i) as follows:

$$S_i(x) = 1, \text{ if } x < i \cdot (\sigma/n), \text{ else } S_i(x) = 0$$

where σ is the standard deviation of the given variable and n is the number of samples. Let us call the multiplier of σ/n, (i), the *similarity level* (Π).

Evaluation of results

The ability to choose the similarity level renders the method flexible. Although the two determining factors of the method (connection level and similarity level) are independent from each other, they have a common influence on the grouping tendencies of the samples. If a relatively small similarity level is chosen to be characteristic, the groups start to form at a high Γ. This also means that relatively many variables are used to classify the samples. If, in the other extreme, a large similarity is accepted, the groups develop at a low Γ (small number of variables).

Given unaltered rocks, two samples belonging to the same genetic group are normally rather close to each other in the original variable space, so their connection is easy to recognize by using traditional cluster methods. In the same case, the value of A(i,j) will be relatively large (close to *m*), because the samples are similar to each other with regard to most of the variables. However, suppose that the effect of some alteration processes is to change some of these variables (elements). The two samples may get "far" from each other in the original variable space. Thus their correspondence may remain hidden when applying traditional classification methods. Given the same alterations, the value of A(i,j) will decrease because of this change in the variables, but this number will remain high relative to two samples that do not belong to the same genetic group because these latter two samples will be similar to each other in only a few variables.

In order to give the possibility of understanding the geological implication of each group and of the whole group-structure, the most characteristic parameters of the groups must be determined simultaneously while revealing the significant differences among them. In other words, the variables must be identified that play a significant role in determining the homogeneity of each group, and also those variables that are responsible for the heterogeneity between groups. In the present classification method, the homogeneity of a group means the number of samples that are similar to each other according to the given variable. An index of homogeneity can be defined:

$$I_{hom,i}(\text{group}) = 2 \cdot e_i / (k_1 \cdot (k_1 - 1)), \text{ where}$$

$I_{hom,i}$ is the *homogeneity index* of the *i*th variable within one group;

e_i is the number of the tie-lines (connections) according to the *i*th variable in the given group;

k_1 is the number of samples in the given

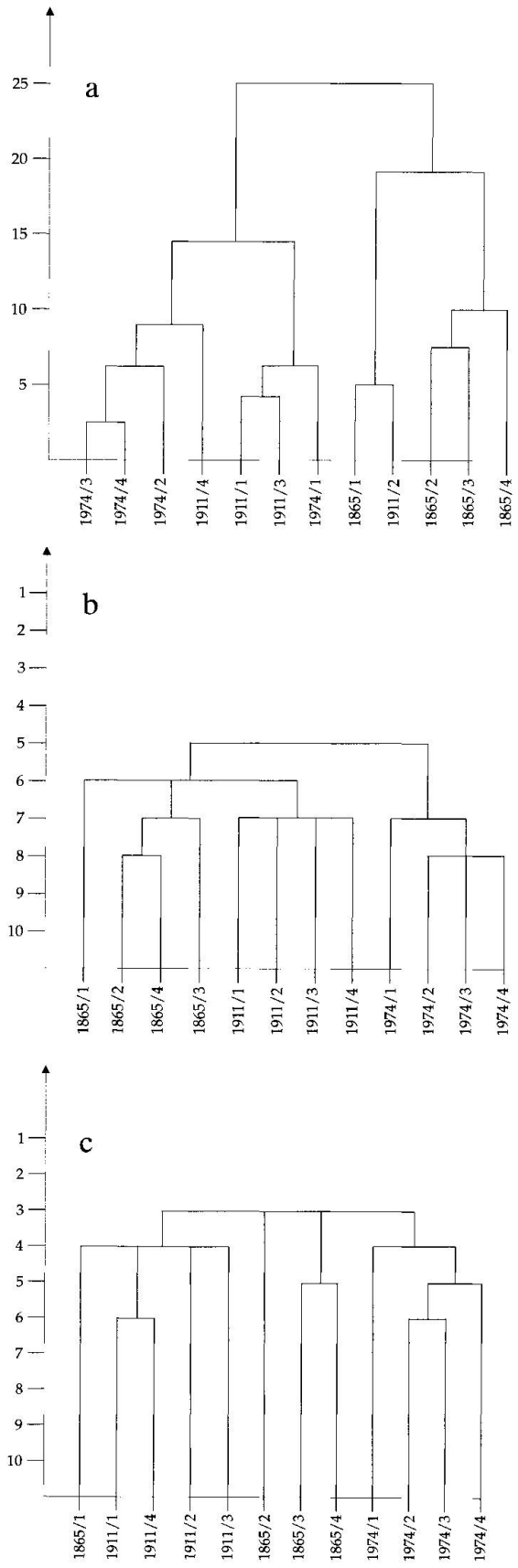

Fig. 3 Dendrograms for Mount Etna hawaiite lavas based on: a) complete linkage method, cos θ measure; b) the A matrix of figure 2 with Π = 9; c) the A matrix of figure 2 with Π = 5.

group (and thus $(k_1 \cdot (k_1 - 1)) / 2$ is the maximum number of possible tie-lines).

The homogeneity index alone is not sufficient to determine the role of the *i*th variable in the group. A second parameter is required to indicate the variables that cause the differences between two groups, i.e. the number of the tie-lines between the groups:

$$I_{het,i}(group_1, group_2) = 1 - f_i / (k_1 \cdot k_2), \text{ where}$$

$I_{het,i}$ is the *heterogeneity index* of the *i*th variable between two groups;

f_i is the number of the tie-lines (connections) according to the *i*th variable between two given groups;

k_1, k_2 are the number of samples in the two given groups (and thus $k_1 \cdot k_2$ is the maximum number of possible tie-lines).

These two indices ($I_{hom,i}$, $I_{het,i}$) provide a sensitive measure of the role of the *i*th variable in forming a given group. However, if (in this group) the homogeneity index of one of the variables is high (close to one) while the heterogeneity index is low (close to zero), the given variable has a similar value in both groups. Consequently, it may not indicate much about the important geochemical features of the groups. Another situation is when the homogeneity is low and the heterogeneity is high. In this case, the examined variable has a high variance in both groups, and neither may be characterized by it. These two examples may help to define the parameter which shows the differences between groups most sensitively:

$$I_{d,i}(group_1, group_2) = (I_{hom,i}(group_1) \cdot \cdot I_{hom,i}(group_2) \cdot I_{het,i}(group_1, group_2))^{1/3},$$

where $I_{d,i}$ is the *discriminant index* of the *i*th variable between $group_1$ and $group_2$.

The higher this value (close to unity), the more suitable the *i*th variable is for characterizing the given pair of groups.

A geochemical example

Trace element data on hawaiite lavas from three eruptions (1865, 1911, 1974) of Mount Etna were selected (Tab. 1) to test the new classification method. A detailed petrological and geochemical presentation as well as a volcanological interpretation of the samples examined were given by BARBIERI et al. (1993). As volcanoes from the Mediterranean region commonly do, Mount Etna shows significant chemical variation among different eruptions. The data set chosen here may be useful to demonstrate the new cluster analysis method introduced above and to compare its per-

Tab. 2 Homogeneity indices ($I_{hom,i}$) of trace elements in the three lava groups for $\Pi = \theta$.

Variables	Homogeneity indices		
	Group 1865	Group 1911	Group 1974
La	0.50	0.66	0.33
Nb	1.00	1.00	1.00
Rb	0.67	0.50	1.00
Zr	0.33	0.50	0.50
Sr	0.33	0.33	0.67
Ni	1.00	1.00	1.00
Co	1.00	1.00	1.00
Cr	0.83	0.17	0.83
Cu	0.17	0.50	0.17
Zn	0.17	0.67	0.17

Tab. 3 Heterogeneity indices ($I_{het,i}$) of examined elements between the groups for $\Pi = \theta$.

Variables	Heterogeneity indices		
	Groups 1865–1911	Groups 1865–1974	Groups 1911–1974
La	0.75	0.81	0.43
Nb	1.00	1.00	0.81
Rb	0.81	0.25	1.00
Zr	0.50	0.88	0.81
Sr	0.62	0.50	0.63
Ni	0.31	1.00	1.00
Co	0.38	1.00	1.00
Cr	0.75	0.94	0.43
Cu	0.81	0.69	0.56
Zn	0.75	0.69	0.56

formance to other methods. To be sure, these samples are not as highly altered as are many (meta)volcanics, especially following hydrothermal metamorphism. The set of hawaiite samples from Etna is sufficient, however, to compare some of the characteristics of the new grouping method to traditional methods.

Nine commonly used hierarchical cluster methods were chosen to classify the given samples. All combinations of complete linkage, centroid and Ward methods with Euclidean, cos θ and Pearson correlation measures were applied. None of these approaches turned out to repeat the expected original structure; indeed these methods all more or less mixed the samples. As an example, figure 3a shows that the four 1911 samples end up in three different groups, and only two of them are found to be similar (1911/1 and 1911/3).

When applying the new classification method, a high Π value is selected first. This ensures that, even in the case of a relatively large distance, two samples may be considered to be similar. Consequently, the G_Γ graph breaks up – subgroups start forming – only at high values of Γ, i.e. if a large number of variables is used for grouping. The set of variables may vary from one sample pair to another as mentioned earlier. For $\Pi = 9$, all samples are connected up to $\Gamma = 5$. When decreasing the Γ value, three groups form, one for each eruption, and the original sample structure is reproduced. The 1911 group remains stable even at $\Gamma = 7$ (Fig. 3b).

By decreasing the value of Π to 5 (Fig. 3c) the diagram suggests two groups, one for the 1974 samples and the other for those of 1911. One sample of the 1865 samples also belongs to this latter group. The rest of the 1865 samples cannot be grouped; they form independent groups at $\Gamma = 2$ (1865/3, 4) and $\Gamma = 3$ (1865/2).

The homogeneity indices for the groups ($\Pi = 9$) show that each group has rather uniform value of Nb, Ni and Co, while Cu is quite variable within each group (Tab. 2). Most variables behave differently within the three groups. The heterogeneity indices (Tab. 3) indicate that Nb might be the most suitable variable to distinguish the three groups. Other elements (e.g. Ni, Co) have high values between samples from 1865–1974 and 1911–1974 groups, but they play an insignificant role in distinguishing the 1865–1911 samples (Tab. 3).

There are excellent examples in the given set of variables to demonstrate the role of the discriminant index. Although $I_{hom,Ni}(1865) = 1$ and $I_{hom,Ni}(1911) = 1$, one finds that $I_{het,Ni}(1865,1911) = 0.31$; hence nickel has a similar value in both groups, and it cannot be used to distinguish the composition of the two eruptions. Similarly, though the value of $I_{het,Cu}(1865,1911)$ is rather high (0.86), $I_{hom,Cu}(1865) = 0.17$ and $I_{hom,Cu}(1911) = 0.33$ are low, making the variable (Cu) unsuitable for discrimination.

In the case of the examined lava rocks, Nb has high $I_{d,i}$ values for each group pair, showing that Nb is most important in distinguishing the groups. $I_{d,Ni}$ and $I_{d,Co}$ are also high for group pairs which contain the 1974 samples (Tab. 4).

Assessment of hierarchical methods

When classifying different cluster methods, hierarchical and nonhierarchical methods are distinguished (ANDERBERG, 1973). The procedure presented in this paper is not a hierarchical one, every

Tab. 4 Discriminant indices ($I_{d,i}$) of the variables.

Variables	Discriminant indices		
	Groups 1865–1911	Groups 1865–1974	Groups 1911–1974
La	0.62	0.51	0.45
Nb	1.00	1.00	0.94
Rb	0.64	0.55	0.79
Zr	0.43	0.52	0.59
Sr	0.41	0.48	0.51
Ni	0.67	1.00	1.00
Co	0.72	1.00	1.00
Cr	0.47	0.86	0.39
Cu	0.40	0.26	0.36
Zn	0.43	0.26	0.39

pair of (Γ, Π) determines a valid classification. However, a hierarchical group structure does in fact result from increasing (or decreasing) Γ step by step for each fixed Π. On increasing Γ, an analogy to divisive methods results; on decreasing Γ a similarity to agglomerative methods is evident. As with other hierarchical methods, the result may be visualized by dendrograms. Such a diagram may be suitable to determine the sample groups.

In the case study of the historical lava samples from Mount Etna, the method possesses the advantages of the hierarchical cluster analyses. Among these, agglomerative approaches build a tree from branches to root, while the divisive methods begin at the root and work towards the branches (ANDERBERG, 1973). The new method belongs to neither one of these groups. It has some of the advantages and avoids some of the difficulties of both families of cluster methods.

Agglomerative approaches usually have a recursive algorithm with the following two steps:

1) Calculation of the similarity matrix using one possible definition of a distance.

2) Connection of the most similar groups.

The distances usually applied are all defined for samples only, and not for groups. Because of their construction, all of these methods have the same problem: How does one define distances that are generally usable to calculate similarities between groups? There are many established approaches to answer this question. However, since the new algorithm uses the original similarity matrix during the entire classification procedure, it is not a recursive method; it does not require recalculation of similarities between groups after each step. This may be also a considerable computational advantage when handling large data sets.

Divisive methods operate by splitting the initial set of samples into two parts. To accomplish this, a dichotomized data base is commonly used instead of the original metric one. The basic concept is then to divide the sample group so as to minimize the appropriate measure of similarity between the newly formed subgroups. Many different approaches to do this were summarized and tested by GILL and TIPPER (1978). The usage of dichotomized (indicator) variables is familiar in geology and is popular in geo-mathematical and geostatistical practice (e.g. JOURNEL, 1983; GOOVAERTS, 1994). Divisive cluster methods yield respectable results and possess theoretically optimal characteristics (GILL and TIPPER, 1978).

Conclusions

The new cluster method, originally designed for classifying altered and geochemically changed rock samples, shows the following features. The algorithm utilizes a specified similarity matrix defined by binary variables. Dichotomy is used for eliminating disproportionately large differences. Similarity between two samples is measured by variables (element concentrations) and, in the second step, by the number of corresponding variables. As a consequence, samples are classified based only on variables that have not changed appreciably, and each variable is applied only to distinguish two samples at a time. This approach can handle samples with significant changes in some of their variables. Since grouping involves only Γ variables, a sample which is similar to the others according to fewer than Γ variables is excluded from the analysis. In fact, it will form an independent group. Because of this simultaneous elimination of some samples and variables, the result should be the framework of the pre-alteration group-structure. The choice of similarity level (Π) makes the method more flexible in generating groups. Due to the simultaneous application of the two determining parameters (Γ, Π), a hierarchical group-structure may also be obtained.

In addition, it may also be possible to determine variables which may be suitable for describing the groups geologically. Using the indices introduced above ($I_{hom,i}$, $I_{het,i}$, $I_{d,i}$), a detailed investigation of each group or pair of groups is possible and may yield geochemical insight.

In addition to the small case study of historic lava samples of Mount Etna, presented above primarily for didactic purposes, other geochemical datasets have been successfully analyzed using the CA method introduced above. While the detailed documentation of these applications is be-

yond the scope of the present paper, this method and its predecessors (M. TÓTH, 1992) have been found capable of handling fairly extensive datasets (M. TÓTH, 1994) as well, and the results do appear to be geologically sensible. Nevertheless, further tests are desirable, as are more applications of the new classification procedure in diverse areas, especially where effects of alteration processes render the application of traditional classification approaches difficult.

Acknowledgements

We are grateful to J. Geiger, Cliff Todd and L.Ó. Kovács, as well as two anonymous reviewers who provided constructive comments on earlier manuscripts of this work and lead us to clarify our presentation. George Skippen's efforts to improve the present paper are acknowledged as well.

References

ALT, J.C. (1995): Subseafloor processes in mid-ocean ridge. Hydrothermal systems. – In: Seafloor Hydrothermal Systems: Physical, Chemical, Biological and Geological Interactions. Geophys. Monogr. 91, 85–114.

ALTENBERG, U. (1991): Hochtemperierte Scherzonen in der Ivrea-Zone/N-Italien – Ein Beitrag zu deren Mikrogefüge, Metamorphose und Geochemie. Zbl. Geol. Paläont., H. 1, 3–20.

ANDERBERG, M.R. (1973): Cluster analysis for applications. Academic Press, New York, 359 pp.

BARBIERI, M., CRISTOFOLINI, R., DELITALA, M.C., FORNASERI, M., ROMANO, R., TADDEUCCI, A. and TOLOMEO, L. (1993): Geochemical and Sr-isotope data on historic lavas of Mount Etna. Jour. Volcan. Geotherm. Res., 56, 57–69.

BARNES, H.L. (1967): Geochemistry of Hydrothermal Ore deposites. Holt, Rinehart and Winston Inc.

COISH, R.A. (1977): Ocean floor metamorphism in the Betts Cove ophiolite, Newfoundland. Contrib. Mineral. Petrol., 60, 255–270.

CONDIE, K.C., VILJOEN, M.J. and KABLE, E.J.D. (1977): Effects of alteration on element distributions in Archean tholeiites from the Barberton greenstone belt, South Africa. Contrib. Mineral. Petrol., 64, 75–89.

GAST, P.W. (1968): Trace element fractionation and the origin of the tholeiitic and alkaline magma types. Geochim. Cosmochim. Acta, 32, 1057–1086.

GIERÉ, R. (1986): Zirconolite, annalite and hoegbomite in a marble skarn from the Bergell contact aureole: implications for mobility of Ti, Zr and REE. Contrib. Mineral. Petrol., 83, 459–470.

GIERÉ, R. (1990): Hydrothermal mobility of Ti, Zr and REE: Examples from the Bergell and Adamello contact aureole (Italy). Terra Nova, 2, 60–67.

GILL, D. and TIPPER, J.C. (1978): The adequacy of nonmetric data in geology: Test using a divisive-omnithetic clustering technique. Jour. Geol., 86, 241–259.

GIRARDI, V.A.V., RIVALENTI, G. and SINIGOI, S. (1986): The petrogenesis of the Niquelandia layered basic-ultrabasic complex. Jour. Petrol., 27, 715–744.

GOOVAERTS, P. (1994): Comparative performance of indicator algorithms for modeling conditional probability distribution functions. Math. Geol., 26/3, 389–411.

GOVETT, G.J.S., ed. (1981): Handbook of Exploration Geochemistry, I. Elsevier, Amsterdam.

GRAHAM, C.M. (1976): Petrochemistry and tectonic significance of Dalradian metabasaltic rocks of the SW Scottish Highlands. Jour. Geol. Soc. London, 132, 61–84.

GRANT, J.A. (1986): The isocon diagram – A simple solution to Gresens' equation for metasomatic alteration. Econ. Geol., 81, 1976–1982.

GRESENS, R.L. (1967): Composition-volume relationships of metasomatism. Chem. Geol., 2, 47–55.

JOURNEL, A.G. (1983): Nonparametric estimation of spatial distribution. Math. Geol., 15/3, 445–468.

M. TÓTH, T. (1992): Földtani objektumok csoportosítása gráfelmélet segítségével Szeghalmi amfibolitok példáján. (Classification of geological samples using graph theory, demonstrated on amphibolites from Szeghalom.) Földtani Közlöny, 122/2–4, 251–263 (in Hungarian with English summary).

M. TÓTH, T. (1994): Geochemical character of amphibolites from Tisza Unit on the basis of incompatible trace elements. Acta Mineral.-Petrograph., Szeged, XXXV, 27–38.

MORRISON, M.A. (1978): The use of "immobile" trace elements to distinguish the palaeotectonic affinities of metabasalts: application to the paleocene basalts of Mull and Skye, northwest Scotland. Earth Planet. Sci. Lett., 39, 407–416.

PEARCE, J.A. (1975): Basalt geochemistry used to investigate past tectonic environments on Cyprus. Tectonophys., 25, 41–67.

PEARCE, J.A. and NORRY, M.J. (1979): Petrogenetic implications of Ti, Zr, Y and Nb variations in volcanic rocks. Contrib. Mineral. Petrol., 69, 33–47.

RIVERIN, G. and HODGSON, C.J. (1980): Wall-rock alteration at the Millenbach Cu–Zn mine, Noranda, Quebec. Econ. Geol., 75, 424–444.

SEYFRIED, W.E., JR., BERNDT, M.E. and SEEWALD, J.S. (1988): Hydrothermal alteration processes at mid-ocean ridges: constrains from diabase alteration experiments, hot-spring fluids, and composition of the oceanic crust. Can. Mineral., 26, 787–804.

SORENSEN, S.S. and BARTON, M.D. (1987): Metasomatism and partial melting in a subduction complex: Catalina Schist, southern California. Geol., 15, 115–118.

WOOD, D.A., GIBSON, I.L. and THOMPSON, R.N. (1976): Elemental mobility during zeolite facies metamorphism of the tertiary basalts of Eastern Iceland. Contrib. Mineral. Petrol., 55, 241–254.

Manuscript received May 15, 1997; revised manuscript accepted September 17, 1997.