

Dependence of chemistry on genesis in zeolites: multivariate analysis of variance and discriminant analysis

ALBERTO ALBERTI AND MARIA FRANCA BRIGATTI

*Istituto di Mineralogia e Petrologia dell'Università
Via S. Eufemia 19,
41100 Modena, Italy*

Abstract

Multivariate analysis of variance and discriminant analysis were used to study the relation between genesis and chemical composition of zeolites. Three hundred and three chemical analyses for 10 chemical elements (Na, K, Mg, Ca, Sr, Ba, Fe³⁺, Al, Si, H₂O) of the mineralogical families: heulandite, chabazite, erionite, phillipsite and analcime, 164 of hydrothermal and 139 of sedimentary genesis respectively were considered. Discriminant analysis showed a strong chemical differentiation between the two geneses for all five families; 95% of the samples are classified in the right genesis by the discriminant functions. Si and Al were important discriminating factors for all the families. The discriminant weight of the other elements varied strongly in the different families. Considering genesis only, 83% of the samples are correctly classified by the discriminant analysis. The difference between this value and 95% of the samples correctly classified when the families are separately examined is explained by the strong interaction between genesis and family shown by multivariate analysis of variance. Of the 10 elements considered only Na appeared unrelated to genesis.

Introduction

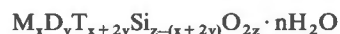
While it is a well known fact that the chemical composition of a mineral depends on thermodynamic conditions and genetic environment of growth, it is a problem to establish which chemical elements, and to what extent, are affected during growth, and how they interact. Such a problem can be studied by multivariate statistical analysis. The method was applied with success to some mineralogical problems that also involved zeolites. Hawkins (1974) used cluster analysis and discriminant analysis to make a chemical distinction between heulandites and clinoptilolites, and principal component analysis to study the relationships between chemical and structural variables. The latter analysis was also used by Alberti (1978) to verify a correlation found between chemical variables in hydrothermal and sedimentary zeolites. The same author (Alberti, 1979) employed factor analysis to evaluate whether each zeolitic species defines a characteristic field of chemical composition. Finally, other authors, (for instance, Passaglia, 1970; Galli and Loschi Ghittoni, 1972) employed multiple regression analysis to detect correlations between chemical, physical and crystallographic variables.

The aim of this work is to characterize differences in chemical composition which exist between hydrothermal and sedimentary zeolites. In agreement with Gottardi and Obradovic (1978) we consider a zeolite as "sedimentary" when it is "a significant constituent of sedimentary rocks (sensu lato), regardless of whether the zeolites were formed in a true sedimentary environment (diagenesis) or during low-grade metamorphism or as the results of low-

temperature hydrothermal activity, as long as the zeolites are homogeneously distributed throughout the rock", as "hydrothermal" when the zeolite is clustered in veinlets, dikes, geodes or fissures, without any evident reaction with the host rock. The study uses multivariate analysis of variance and discriminant analysis to establish the discriminating elements for five different families of zeolites. A similar problem was encountered by Alberti (1978) who obtained satisfactory results by the technique of factor analysis. In this paper a greater quantity of chemical data are available and more appropriate statistical tools are used.

Choice of the data

A general expression for the chemical composition of a zeolite may be written as:



where

M = Na, K

D = Mg, Ca, Sr, Ba and, rarely, subordinate amounts of Fe²⁺ and Mn

T = Al and Fe³⁺. The latter in lower but not negligible quantity.

Therefore, ignoring Fe²⁺ and Mn the principal chemical variables are ten. The range of values for the subscripts x and y and the multiplier n is as follows:

$$0.0 \leq x \leq 0.50z$$

$$0.0 \leq y \leq 0.25z$$

$$0.12z \leq x + 2y \leq 0.50z$$

$$0.30z \leq n \leq 1.20z$$

It has been observed that none of the zeolites covers the whole field of chemical variation.

Unfortunately, in published data, chemical analyses are sometimes incomplete and important elements such as Sr, Ba, Fe^{3+} , and Mg, have been neglected. The present study is restricted to chemical analyses from publications post-dating 1950, in which most of the principal chemical variables were determined, i.e., data for at least SiO_2 , Al_2O_3 , Na_2O , K_2O , and CaO are available. Where Fe is given as Fe_{tot} , this was recalculated as Fe^{3+} . Where H_2O was not determined, it has been assumed to be equal to the difference $100 - \Sigma$ analysis. Only analyses where the charge balance $E \leq 10\%$ with

$$E = \frac{(\text{Al} + \text{Fe}^{3+}) - (\text{Na} + \text{K} + 2[\text{Mg} + \text{Ca} + \text{Sr} + \text{Ba}])}{(\text{Na} + \text{K} + 2[\text{Mg} + \text{Ca} + \text{Sr} + \text{Ba}])} \cdot 100$$

have been selected. Less than 50% of the examined analyses satisfy the conditions posed.

Another constraint on the reliability of the analyses is due to the high capacity of zeolites for ionic exchange, this may produce transformations of the "original" chemical composition, i.e., composition during growth. The complication may be overcome if the composition of a phase, at the time of crystallization, and after ionic exchange in natural environments, is assumed as representative of the phase.

Even if the number of species grown either in hydrothermal or in sedimentary environments is high (see for instance Iijima and Utada (1972)), reliable chemical analyses of samples of sedimentary genesis, and in sufficient quantity for a statistical study, represent only the five species listed in Table 1.

Table 2 contains all zeolites which were used in the statistical analysis; the literature source is given for the chemical analyses from which the formulae were calculated¹.

Description of the method

Multivariate analysis of variance

The problem is a typical example of analysis of variance in a multidimensional measurement space, or Manova (Cooley and Lohnes, 1971). In this analysis, the dependent variable is a vector variable that should be normal in distribution and have the same dispersion for each group from which the samples are drawn. In practice, these assumptions need not be strongly adhered to.

The problem target of this research system is whether the groups are centered at different locations in the measurement space spanned by the dependent vector variable or not.

The linear model of Manova may be written as:

$$\mathbf{X}_{ki} = \mathbf{m} + (\mathbf{m}_k - \mathbf{m}) + (\mathbf{X}_{ki} - \mathbf{m}_k) \quad (1)$$

where \mathbf{X}_{ki} is the dependent vector variable for the i -th sample ($i = 1, \dots, N$) in the k -th group ($k = 1, \dots, g$) considered;

$$\mathbf{m} = 1/N \sum_{k=1}^g \sum_{i=1}^{N_k} \mathbf{X}_{ki} \text{ is the grand centroid, and} \quad (2)$$

$$\mathbf{m}_k = 1/N_k \sum_{i=1}^{N_k} \mathbf{X}_{ki} \text{ is the centroid for } k\text{-th group.} \quad (3)$$

Calling $\mathbf{x}_{ki} = \mathbf{X}_{ki} - \mathbf{m}$ the deviation of the i -th sample from the grand centroid, we may write

$$\mathbf{x}_{ki} = (\mathbf{m}_k - \mathbf{m}) + (\mathbf{X}_{ki} - \mathbf{m}_k) \quad (4)$$

The term $(\mathbf{m}_k - \mathbf{m})$ represents the difference in locations of the groups, (i.e., the "hypothesis effect") and the term $(\mathbf{X}_{ki} - \mathbf{m}_k)$ represents the deviation of the sample from its group centroid, (i.e., the "error effect") (Cooley and Lohnes, 1971).

Summing over all samples in all groups we have the partition theorem of Manova:

$$\sum_{k=1}^g \sum_{i=1}^{N_k} \mathbf{x}_{ki} \mathbf{x}'_{ki} = \sum_{k=1}^g \sum_{i=1}^{N_k} (\mathbf{m}_k - \mathbf{m})(\mathbf{m}_k - \mathbf{m})' + \sum_{k=1}^g \sum_{i=1}^{N_k} (\mathbf{X}_{ki} - \mathbf{m}_k)(\mathbf{X}_{ki} - \mathbf{m}_k)' \quad (5)$$

The term on the left of expression (5), usually called \mathbf{T} , for "Total", is the matrix of sum of squares and cross-products of deviations of all samples from the grand centroid:

$$\mathbf{T} = \sum_{k=1}^g \sum_{i=1}^{N_k} \mathbf{x}_{ki} \mathbf{x}'_{ki} \quad (6)$$

The first partition term on the right of expression (5) is the matrix of weighted squares and cross-products of deviations of the group centroids from the grand centroid. The second partition term on the right of expression (5) is the matrix of squares and cross product of deviations of samples from their group centroid, pooled over all groups.

The first partition term, identified by \mathbf{A} , for "Among-groups", is:

$$\mathbf{A} = \sum_{k=1}^g \sum_{i=1}^{N_k} (\mathbf{m}_k - \mathbf{m})(\mathbf{m}_k - \mathbf{m})' = \sum_{k=1}^g N_k (\mathbf{m}_k - \mathbf{m})(\mathbf{m}_k - \mathbf{m})' \quad (7)$$

The second partition term, identified by \mathbf{W} , for "Within-groups", is:

$$\mathbf{W} = \sum_{k=1}^g \sum_{i=1}^{N_k} (\mathbf{X}_{ki} - \mathbf{m}_k)(\mathbf{X}_{ki} - \mathbf{m}_k)' \quad (8)$$

In a simplified form, we can write

$$\mathbf{T} = \mathbf{A} + \mathbf{W} \quad (9)$$

Our purpose is to verify whether the "null hypothesis", that is, where the groups are not discriminated, can be rejected. The "Wilks" determinant ratio test statistic, named Wilks' Lambda, (Wilks, 1932) can be used as an estimator of the discrimination existing among groups. It is defined as:

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{T}|} \quad (10)$$

Decreasing values of this ratio accompany increases in the confidence with which the null hypothesis is rejected.

¹ To receive a copy of Table 2, order Document AM-84-272 from the Business Office, Mineralogical Society of America, 2000 Florida Avenue, N.W., Washington, D.C. 20009. Please remit \$1.00 in advance for the microfiche.

Discriminant analysis

In the case of rejection of the "null hypothesis" it is desirable to describe the measured differences of the groups with a simple model able to maximize these differences. Discriminant analysis faces this problem. Its objective is to weight and linearly combine the discriminating variables in some fashion so that the groups are forced to be as statistically distinct as possible.

The discriminant model is given by one or more discriminant functions which are linear combinations of the dependent discriminating variables. These functions are of the form

$$R_l = v_l X_{ki} \quad (11)$$

R_l is the score on discriminant function l , v_l is the vector of the weighting coefficients for function l and X_{ki} is the vector of the dependent variables for the i -th sample in the k -th group.

The coefficients of the weighting function are determined by solving the general eigenvalue problem

$$(W^{-1}A - \lambda)v = 0 \quad (12)$$

where v is the matrix of coefficients for the weighting function, λ is a diagonal matrix of eigenvalues, and W and A are, as previously defined, the "Within-groups" and "Among-groups" partition terms, respectively (Cooley and Lohnes, 1971). If p is the number of dependent variables, the rank of matrix W is of order p , whereas A is a $p \cdot (g - 1)$ matrix. Then the rank of $W^{-1}A$ will be of reduced rank $(g - 1)$ if $(g - 1) < p$, and $(g - 1)$ will be the number of discriminant functions. Consequently, the representative space of the samples is reduced to a $(g - 1)$ dimensional space and each discriminant function is a unique (orthogonal) dimension describing the location of each sample of each group (as measured by its centroid) relative to the others.

Every discriminant function has a different discriminating power: its relative importance is measured by the relative percentage of the eigenvalue associated with the function. However in some cases the number of functions can be less than $(g - 1)$, or p , without a considerable loss of discriminating power.

Once the discriminating functions have been derived, we can pursue the two objectives of this technique: discrimination and classification.

The success of the discrimination process is measured easily by many statistical tests which indicate how much the discriminating variables are effective when combined into the discriminating functions.

Classification permits: (1) identification of the likely group membership of a sample when only the discriminating variables are known; (2) testing of the adequacy of the derived discriminant functions; (3) measurement of the success in discrimination by observing the proportion of correct classification when the group membership is already known.

Our classification functions (Fischer's linear discriminant functions) are derived from the pooled within groups covariance matrix and from the centroids for the discriminant variables. Their form is:

$$C_k = c_{k0} + \sum_{j=1}^p c_{kj} Y_{jn} \quad (13)$$

where C_k is the classification score for the group k , Y_{jn} is the j -th row discriminant variable of the n -th sample to be classified, c_{kj} are the classification coefficients and c_{k0} is a constant. The Y_n vector variable will be one of the X_{ki} dependent vector variables if the sample has been used in the discrimination process, it will not be one of the X_{ki} if the sample is used only in the classification

process. For every sample, g classification scores C_k , one for each of the g groups considered, are obtained from the equations (13). The sample will be classified in the group with the highest score and the probability of group membership P_k (posterior probability) can be calculated from

$$P_k = \exp(C_k) / \sum_{k=1}^g \exp(C_k) \quad (14)$$

A detailed discussion on multivariate factor analysis can be found in Morrison (1978) and in Cooley and Lohnes (1971).

Computational details

SPSS, a Statistical Package for the Social Sciences (subprograms MANOVA and DISCRIMINANT) was used in this research (Nie et al., 1975). In the discriminant analysis, independent variables were selected on the basis of their discriminant power using an upward stepwise method. This method was preferred to the one that uses all p independent variables because generally a number of discriminating variables $p' < p$ achieves an equally satisfactory discrimination. The remaining $p - p'$ variables can be statistically significant but their importance in discrimination is debatable because all the variables are correlated. The discriminating criterion selected is the "Mahalanobis distance" which seeks to maximize the distance between the two closest groups (Mahalanobis, 1936).

From discriminant analysis we have $(g - 1)$ (or p' if $p' < (g - 1)$) discriminant functions. For each function the following information is given (SPSS-X Statistical Algorithms, 1983): (1) The eigenvalue λ_l ($l = 1, \dots, g - 1$). (2) The Wilks' Λ value. (3) The approximate F test for Λ . (4) The probability that the "null hypothesis" can be rejected, given the value of the Wilks' Λ . (5) The coefficients v_l of the unstandardized canonical discriminant functions. The sum of these coefficients multiplied by the respective raw discriminant variable values gives the discriminant score for this function. (6) The coefficients of the canonical discriminant functions. These coefficients give, when the sign is ignored, the relative contribution of their associated variables to the function.

Furthermore, the probability of the "null hypothesis" for each chemical element will be given. Lastly, the coefficient of the g classification functions c_{kj} (Fischer's linear discriminant functions) are given. For every sample, known its chemical formula on the basis of 200 oxygens and the c_{kj} coefficients, the equations (13) give the g classification scores and the equations (14) the probability of group membership for each of the g groups. The sample will be classified in the group with the highest score.

Results and discussion

Table 1 synthesizes the most significant data on the chemistry of the families studied.

An almost complete overlapping of chemical data of the two genetic groups is evident either when the five families are considered separately, or when they are treated as a whole. In addition, the differences among the mean values are not statistically significant. This implies that a comparison between the separate variables does not allow evaluation of differences between analysis of the two genetic groups. This fact was previously observed by Alberti (1978) who showed that the use of standard Si-divalent cations-monovalent cations, or Na-Ca-K ternary diagrams was not sufficient to discriminate between the two genetic groups.

The results of discriminant analysis are now examined

Table 1. Minimum, maximum, mean content and standard deviation for the ten elements analyzed and for the five families, on the basis of 200 oxygens

		No. of samples	Na	K	Mg	Ca	Sr	Ba	Fe	Al	Si	H ₂ O
Hydrothermal	38	x_{min}	0.0	0.6	0.0	1.9	0.0	0.0	0.0	17.9	73.6	47.6
		x_{max}	11.1	6.9	2.7	10.1	5.8	3.1	1.6	26.3	82.3	78.4
		\bar{x}	3.9	2.1	0.6	6.3	0.7	0.5	0.2	22.3	77.5	65.3
		σ_x	3.0	1.4	0.8	2.0	1.2	0.8	0.4	2.5	2.5	6.8
HEULANDITES												
Sedimentary	53	x_{min}	0.2	0.8	0.0	0.4	0.0	0.0	0.0	15.2	77.3	32.3
		x_{max}	12.3	12.3	3.8	5.3	0.6	0.4	2.1	21.9	84.7	85.9
		\bar{x}	4.6	6.2	1.2	2.2	0.0	0.0	0.3	17.9	81.9	53.8
		σ_x	3.1	3.1	0.9	1.4	0.1	0.1	0.5	1.2	1.6	11.7
CHABAZITES												
Hydrothermal	33	x_{min}	0.1	0.3	0.0	1.6	0.0	0.0	0.0	26.4	58.5	90.3
		x_{max}	25.9	17.1	2.5	15.5	5.1	1.2	1.4	41.1	73.4	127.9
		\bar{x}	4.7	4.3	0.3	10.2	1.1	0.2	0.2	32.2	67.5	107.0
		σ_x	6.7	3.6	0.6	3.3	1.6	0.3	0.4	4.1	4.1	8.5
ERIONITES												
Sedimentary	18	x_{min}	0.2	0.7	0.0	0.8	0.0	0.0	0.0	19.2	63.1	76.1
		x_{max}	24.2	11.5	5.6	10.7	7.3	0.7	5.7	36.4	80.4	112.1
		\bar{x}	5.0	5.8	1.5	6.6	0.6	0.1	1.0	27.4	71.7	92.4
		σ_x	6.7	3.8	1.3	3.0	1.7	0.2	1.3	5.3	5.0	10.3
ERIONITES												
Hydrothermal	10	x_{min}	0.4	4.3	0.0	3.3	0.0	0.0	0.0	23.6	73.8	44.2
		x_{max}	4.1	7.7	5.9	10.7	0.0	0.0	1.1	26.3	76.6	95.2
		\bar{x}	1.9	6.3	2.1	6.2	0.0	0.0	0.1	24.7	75.2	73.8
		σ_x	1.3	1.3	1.7	2.2	*	*	0.3	0.8	1.0	16.8
ERIONITES												
Sedimentary	10	x_{min}	1.6	5.5	0.3	0.0	0.0	0.0	0.1	20.0	73.4	44.3
		x_{max}	16.1	11.5	3.0	4.7	0.0	0.1	3.4	26.1	78.5	79.2
		\bar{x}	9.3	7.5	1.1	1.8	0.0	0.0	1.2	21.8	77.2	66.1
		σ_x	5.4	1.7	0.9	1.6	*	0.0	1.0	1.8	1.5	9.4
PHILLIPSITES												
Hydrothermal	39	x_{min}	0.3	3.9	0.0	1.6	0.0	0.0	0.0	23.8	57.6	46.3
		x_{max}	26.2	23.7	1.2	13.3	0.6	4.7	1.0	42.7	75.8	127.2
		\bar{x}	5.7	10.3	0.2	8.2	0.1	0.7	0.1	34.7	65.3	79.3
		σ_x	7.3	4.7	0.3	3.3	0.1	1.3	0.2	4.2	4.2	12.6
PHILLIPSITES												
Sedimentary	40	x_{min}	1.9	5.9	0.0	0.0	0.0	0.0	0.0	22.3	69.5	49.4
		x_{max}	17.9	17.1	1.6	7.9	0.1	1.0	1.4	30.0	77.0	80.2
		\bar{x}	9.3	11.6	0.6	2.4	0.0	0.1	0.2	27.1	72.6	68.0
		σ_x	4.3	2.4	0.5	2.2	0.0	0.2	0.4	2.1	1.8	7.7
ANALCITES												
Hydrothermal	44	x_{min}	20.8	0.0	0.0	0.0	0.0	0.0	0.0	29.8	59.3	32.0
		x_{max}	35.8	7.2	1.7	3.5	0.0	0.0	2.2	39.5	69.5	41.6
		\bar{x}	31.0	0.7	0.1	0.7	0.0	0.0	0.5	33.5	66.2	35.9
		σ_x	2.8	1.2	0.3	0.8	*	*	0.6	2.0	2.2	2.1
ANALCITES												
Sedimentary	18	x_{min}	9.8	0.0	0.0	0.0	0.0	0.0	0.0	25.7	68.3	33.4
		x_{max}	31.4	2.3	1.5	10.8	0.0	0.0	2.0	32.2	73.2	49.4
		\bar{x}	25.6	0.5	0.5	1.5	0.0	0.0	0.3	28.9	70.6	37.0
		σ_x	5.2	0.7	0.5	3.0	*	*	0.5	1.7	1.3	4.0
ALL FAMILIES												
Hydrothermal	164	x_{min}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	17.9	57.6	32.0
		x_{max}	35.8	23.7	5.9	15.5	5.8	4.7	2.2	42.7	82.3	127.9
		\bar{x}	11.6	4.4	0.4	6.0	0.4	0.3	0.3	30.4	69.4	69.6
		σ_x	12.9	4.7	0.8	4.3	1.0	0.8	0.4	5.9	5.9	26.4
ALL FAMILIES												
Sedimentary	139	x_{min}	0.2	0.0	0.0	0.0	0.0	0.0	0.0	15.2	63.1	32.3
		x_{max}	31.4	17.1	5.6	10.8	7.3	1.0	5.7	36.4	84.7	112.1
		\bar{x}	9.0	7.1	1.0	2.7	0.1	0.1	0.4	23.5	76.1	61.6
		σ_x	8.1	4.4	0.9	2.6	0.6	0.1	0.8	5.3	5.3	18.1
ENTIRE SAMPLE 303												
ENTIRE SAMPLE	303	x_{min}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	15.2	57.6	32.0
		x_{max}	35.8	23.7	5.9	15.5	7.3	4.7	5.7	42.7	84.7	127.9
		\bar{x}	10.4	5.6	0.7	4.5	0.3	0.2	0.3	27.2	72.5	65.9
		σ_x	11.0	4.7	0.9	4.0	0.9	0.6	0.6	6.6	6.6	23.3

Table 3. Probability of the "null hypothesis" of the chemical variables for the five families

Family	Na	K	Mg	Ca	Sr	Ba	Fe	Al	Si	H ₂ O
HEULANDITES	0.293	0.000	0.001	0.000	0.000	0.000	0.247	0.000	0.000	0.000
CHABAZITES	0.859	0.176	0.000	0.000	0.288	0.132	0.003	0.001	0.002	0.000
ERIONITES	0.000	0.071	0.140	0.000	*	0.331	0.007	0.000	0.002	0.223
PHILLIPSITES	0.009	0.137	0.000	0.000	0.014	0.003	0.233	0.000	0.000	0.000
ANALCIMES	0.000	0.634	0.001	0.105	*	*	0.314	0.000	0.000	0.182

The symbol * is used when the element was not determined or found equal to zero for all the samples in the family.

where the families are analyzed separately at first, and later all together. Tables 3 to 7 report the former in order: the significance of the "null hypothesis" for the ten elements; the parameters related to canonical discriminant functions; the unstandardized and canonical coefficients of the discriminant functions; the classification function coefficients; and the summary of the cases correctly and incorrectly classified. Figure 1 shows the distribution of the scores of the samples as given by the unstandardized discriminant function.

Heulandites

The discrimination between sedimentary and hydrothermal samples is very significant for this family as is shown by the values of the canonical correlation and Wilks' Λ reported in Table 4. Individually, Na and Fe only are not significant discriminating elements, as shown in Table 3. The canonical variables listed in Table 5, indicate that Si/Al ratio and Ca are better discriminating variables and this is in agreement with previous results (Alberti, 1978). The importance of CaO content as discriminant variable between heulandites and clinoptilolites was pointed out by Hawkins (1974). A direct comparison between Hawkins' and our classification functions is not possible both because he used wt% of oxydes and we chemical elements as dependent variables and mainly because there are differences in the classification of samples, as is stressed in the following. As can be seen in Table 7 and in Figure 1, four samples (4% of the total) are classified incorrectly. The

classification in hydrothermal and sedimentary heulandites differs strongly from the classification in heulandites and clinoptilolites. In fact, 8 clinoptilolites are almost certainly of hydrothermal genesis. Heulandites classified as type 2 by Alietti (1972) and Boles (1972) are sometimes of hydrothermal and sometimes of sedimentary genesis. Also heulandites of epimetamorphic genesis have been found, and at least one sample of certain sedimentary genesis, HEUS 75 in Table 2, (E. Passaglia, personal communication) has a heulandite type behavior.

Chabazites

Table 3 shows that for this family many ions (Na, K, Sr, Ba) are not significant discriminating elements. In practice the discrimination is due to the H₂O, Si, and Fe content. The significance of the Fe found here, has been emphasized elsewhere by Alberti (1978) but, in his work a limited number of sedimentary samples were used and the result was considered of uncertain significance. Sheppard and Gude (1970a) suggest that chabazites from tuffaceous sedimentary rocks can be distinguished from chabazites from igneous and metamorphic rocks on the bases of Si/(Al + Fe) ratio. This ratio should be greater than 3 for chabazites that occur in sedimentary rocks, less than 3, and generally near to 2, for the other chabazites. This hypothesis is in disagreement with the strong overlapping of the Si/(Al + Fe) ratios between the two geneses deducible from the data in Table 1. Tables 3, 5 and 6 suggest that this ratio is a noticeable variable, but they also show that the water content is by far the most important discriminating and classifying factor. This family has the least marked boundary between hydrothermal and sedimentary samples as is evident in Figure 1. This can also be seen from the values of the canonical correlation and of the Wilks' Λ in Table 4.

Erionites

This family shows a strong discrimination which is represented in Tables 4 and 7, and in Figure 1. Unfortunately there are few chemical analyses available. Na, Ca, Si/Al ratio, and Fe are better discriminating chemical variables, as can be seen in Table 3. The significance of Fe as a discriminating element, found for erionites and for chaba-

Table 4. Parameters of the canonical discriminant function for the five families

Family	Eigenvalue	Wilks' Λ	Approx. F	Probability
HEULANDITES	2.990	0.2506	35.4573	0.0000
CHABAZITES	1.466	0.4055	10.7533	0.0000
ERIONITES	10.213	0.0892	22.1273	0.0000
PHILLIPSITES	3.432	0.2256	34.8087	0.0000
ANALCIMES	1.625	0.3809	18.2051	0.0000

Table 5. Unstandardized discriminant function coefficients and canonical variables for the five families

Family		Na	K	Mg	Ca	Sr	Ba	Fe	Al	Si	H ₂ O
HEULANDITES	Unst. coefficients	*	0.11111	0.42580	-0.28577	*	-0.68391	*	-0.56474	-0.36849	-0.03397
	Canonical variables	*	0.4642	0.2026	-0.6929	*	-0.2336	*	-0.6793	0.6125	-0.3337
CHABAZITES	Unst. coefficients	*	-0.14486	*	*	-0.57740	1.63413	-0.60777	*	-0.18643	0.07808
	Canonical variables	*	-0.1622	*	*	0.1268	0.1805	-0.3703	*	-0.3764	0.6422
ERIONITES	Unst. coefficients	*	0.37870	-0.38561	-0.31182	*	74.98958	*	-1.25892	*	0.09003
	Canonical variables	*	0.1413	-0.1140	-0.3813	*	0.0738	*	-0.3564	*	-0.0931
PHILLIPSITES	Unst. coefficients	-0.06501	-0.24301	-1.13014	*	-3.09973	0.49124	*	*	-0.39299	-0.03444
	Canonical variables	-0.1658	-0.0925	-0.2698	*	0.1549	0.1866	*	*	-0.6291	0.2979
ANALCIMES	Unst. coefficients	-0.10560	-0.26427	0.92997	*	*	*	*	-0.38249	*	0.06911
	Canonical variables	-0.5471	-0.0484	0.3530	*	*	*	*	-0.8719	*	0.1368

The symbol * is used when the upward stepwise method does not include the element in the statistical analysis.

zites, is in agreement with the results of Sheppard and Gude (1969), who stated that "there is a strong suggestion that ferric ion can substitute for aluminum in zeolites of sedimentary deposits". For the other families, however, Fe content is not a discriminant element.

Phillipsites

Table 3 shows that, for this family, individually, K and Fe only are not significant discriminating elements. It can

be noted that potassium is the most abundant exchangeable cation (see Table 1). Tables 5 and 6 indicate that Si is the most important discriminating and classifying element, even if the overlap of Si/Al ratio between the two geneses, deducible from the data of Table 1, is much more evident than that shown by Sheppard and Gude (1970b); Mg and H₂O are also important. The discrimination between the two genetic groups is extremely significant as can be deduced by the data in Table 4 and Figure 1.

Table 6. Classification function coefficients (Fisher's linear discriminant functions) for the two geneses in the five families

Family	Genesis	Na	K	Mg	Ca	Sr	Ba	Fe	Al	Si	H ₂ O	Constant
HEULANDITES	Hydrothermal	*	-33.666	193.467	-51.870	*	-8.210	*	752.300	707.602	6.213	-35904.0
	Sedimentary	*	-33.281	194.944	-52.861	*	-10.582	*	750.342	706.324	6.095	-35754.0
CHABAZITES	Hydrothermal	*	9.108	*	*	12.026	-9.197	-4.256	*	12.319	2.971	-600.7
	Sedimentary	*	9.468	*	*	13.460	-13.256	-2.746	*	12.782	2.777	-617.0
ERIONITES	Hydrothermal	*	4.739	-1.976	-3.138	*	-2428.004	*	53.409	*	-1.450	-610.9
	Sedimentary	*	7.035	-4.314	-5.029	*	-1973.309	*	45.776	*	-0.904	-478.4
PHILLIPSITES	Hydrothermal	2.318	10.008	10.313	*	178.060	-20.027	*	*	17.276	3.661	-767.9
	Sedimentary	2.556	10.897	14.447	*	189.400	-21.824	*	*	18.713	3.787	-889.3
ANALCIMES	Hydrothermal	1.403	-2.113	12.844	*	*	*	*	8.703	*	4.559	-250.0
	Sedimentary	1.111	-2.843	15.413	*	*	*	*	7.646	*	4.750	-216.0

The symbol * is used when the upward stepwise method does not include the element in the statistical analysis.

Table 7. Summary of the cases correctly and incorrectly classified

Family	Predicted group membership		Cases correctly classified (%)
	1	2	
HEULANDITES	1	36	95.6
	2	2	
CHABAZITES	1	30	88.2
	2	3	
ERIONITES	1	10	100.0
	2	0	
PHILLIPSITES	1	36	96.2
	2	0	
ANALCIMES	1	42	96.8
	2	0	

1. Hydrothermal genesis. 2. Sedimentary genesis.

Analcimes

This family shows, among the five families considered, the narrowest field of chemical variation in exchangeable cations. Na is by far the most common extraframework cation, as was previously found by Wilkinson (1968) and is shown in Table 1. Coombs and Whetten (1967) and Iijima and Hay (1968) pointed out the relationship between Si/Al ratio and the mode of origin of analcime. Tables 3 and 5 show that, in addition to Si and Al, Na and Mg are discriminating elements. Canonical correlation and Wilks' Λ in Table 4, indicate that for this family the discrimination is not as high as for heulandites, erionites, and phillipsites although of the 62 samples considered only 2 are misclassified.

In conclusion, of the five families, only 15 samples (5% of the total) are misclassified and the significance of the discrimination is very high for all families.

Table 8 summarizes the results of the discriminant analysis when the samples are classified in two groups on the basis of genesis only, without taking into account the fact that they belong to different families.

Also in this case, the discrimination of the samples on the basis of their genesis is strongly significant. Only the Na content does not seem to depend on genesis. The significance of the dependence of the cation content on genesis is very strong for K, Mg, Ca, Si, Al and less evident for the other elements. It is impossible to establish whether this last result is a consequence of the fact that Sr, Ba, Fe and H₂O have not been determined in many samples, or whether the dependence of the elements on genesis is indeed lower. It can be noted that some of these elements, in particular K, are not discriminating elements for many of the families considered, as is shown in Table 3.

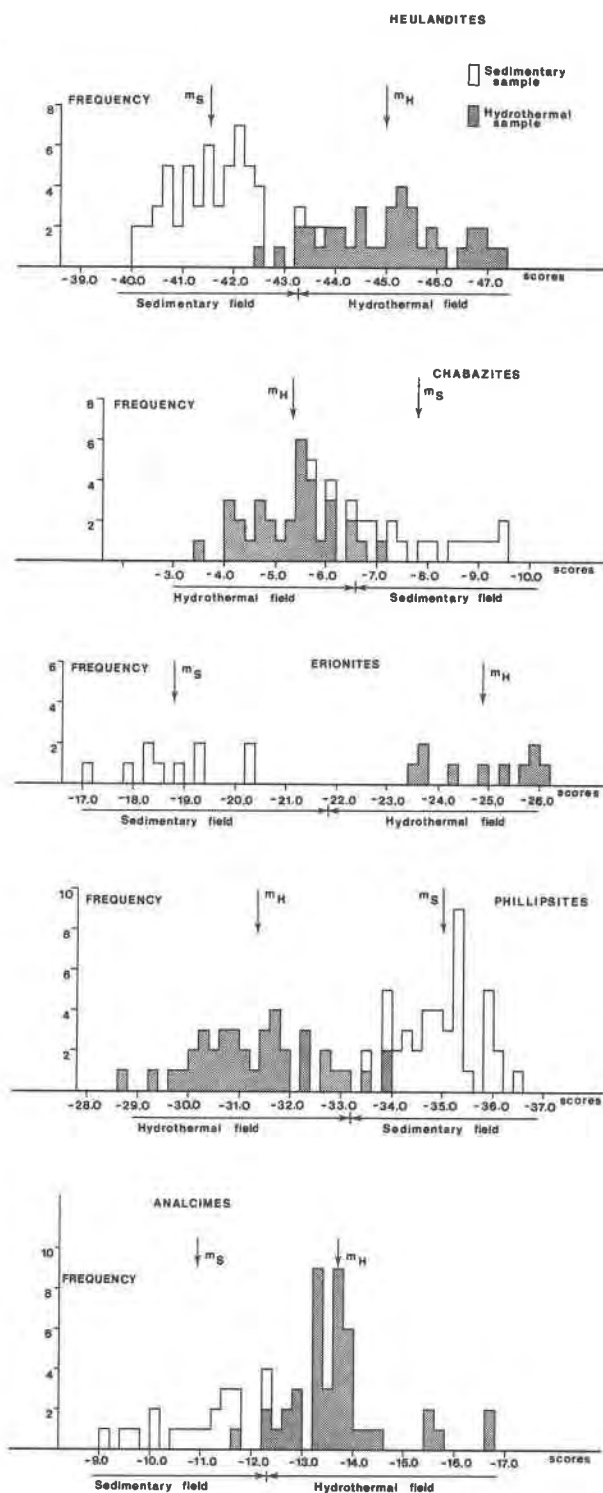


Fig. 1. Histogram of the sample scores, as given by the unstandardized discriminant function, and their classification in the hydrothermal or sedimentary group, on the basis of Fisher's classification functions, for the five families. m_H and m_S represent the centroids for the hydrothermal and sedimentary groups respectively.

Table 8. Probability of the chemical variables, unstandardized discriminant function coefficients, canonical variables, classification function coefficients, parameters of the canonical discriminant function, and summary of the cases correctly and incorrectly classified for the samples divided into two groups on the basis of genesis only.

	Na	K	Mg	Ca	Sr	Ba	Fe	Al	Si	H ₂ O	Constant	
Probability	0.041	0.000	0.000	0.000	0.002	0.000	0.011	0.000	0.000	0.003		
Unst. coefficients	0.20661	0.35018	0.52408	0.21114	0.52823	*	*	-0.33635	*	0.00933		
Canonical variables	-0.1136	0.2822	0.3207	-0.4387	-0.1709	*	*	-0.5903	*	-0.1685		
Classification coefficients	Hydrothermal	-0.5811	-0.9540	1.2785	-1.2996	-2.5540	*	*	1.7173	*	0.1352	-21.87
	Sedimentary	-0.1501	-0.2236	2.3716	-0.8593	-1.4523	*	*	1.0158	*	0.1546	-15.84
Eigenvalue: 1.087; Wilks' λ : 0.4791; Approx. F: 45.8212; Probability: 0.0000												
Genesis	Predicted group membership		Cases correctly classified (%)									
	1	2										
1	138	26	83.2									
2	25	114										

The symbol * is used when the upward stepwise method does not include the element in the statistical analysis.
1. Hydrothermal genesis. 2. Sedimentary genesis.

As expected, the values of the canonical variables indicate a direct correlation between Ca and Al whereas the discriminating power of Mg is unexpectedly high.

Figure 2 shows the distribution of the discriminant scores for the two genetic groups. Fifty-one samples (17% of the total) are incorrectly classified in comparison with the fifteen samples misclassified when the families are analyzed separately.

The justification for this discrepancy is obtained if we consider not only the contribution of genesis to the variance of the whole system, but also the contribution of the interaction between family and genesis. The value of Wilks'

Λ test of significance (0.4837) underlines the strong effect of family-genesis interaction. To this effect we can attribute the partial overlap between sedimentary and hydrothermal zeolites in Figure 2. All the elements considered are responsible for this effect, with the probable exception of Sr. The probability of the "null hypothesis" for Sr is in fact 0.047 whereas for the other 9 elements this value is always ≤ 0.002 .

This result is remarkable as the differences among the mean values for every chemical variable are always less than their standard deviation (see Table 1).

Conclusions

The results of analysis of variance and discriminant analysis indicate that for the five families considered a strong chemical boundary exists between hydrothermal and sedimentary zeolites, and, in fact, only 5% of the samples were incorrectly classified with regard to their genetic groups. Conversely, these results show the power of these statistical methods for problems of classification and discrimination.

The ability of the elements to discriminate between the two genetic groupings differs for the different families. For example, while the Si/Al ratio is generally an important discriminating factor, the weight of the other elements, in particular Ca, H₂O, and Mg, varies for the different families.

Discriminant analysis for the samples classified into two groups on the basis of genesis only, disregarding the different families, classifies 83% of the samples correctly. The analysis of variance justifies this high number of incorrectly classified samples compared with the number found when the families are separately considered, and shows the pres-

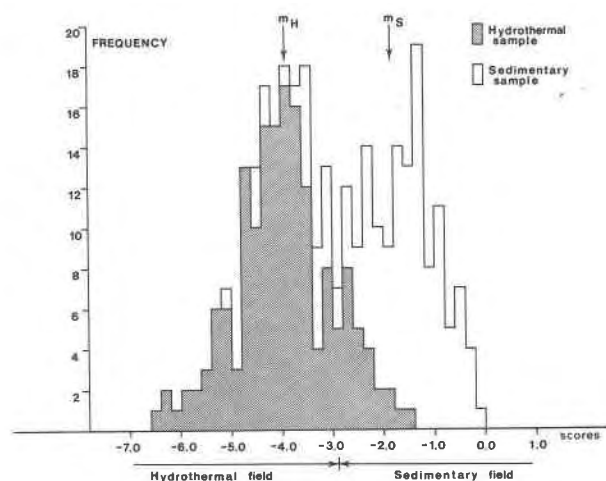


Fig. 2. Histogram of the sample scores grouped according to genesis only. The meaning of the symbols is as in Fig. 1.

ence of a strongly significant interaction between geneses and families. Only Sr does not seem to be affected.

Finally, the high probability of the discriminant functions provides a satisfactory criterion in identifying the genetic grouping of new unclassified samples.

The satisfactory results of this statistical analysis cannot be considered as a proof of the existence of a direct cause-effect relationship between genesis and chemistry. Certainly a lot of factors control zeolite composition (for example composition and temperature of the fluid) and undoubtedly the influence of the genetic environment of growth on some of these factors is remarkable also when hydrothermal zeolites are considered. Lastly, whereas the genesis of a zeolite is in most cases easily determinable, unfortunately no certain experimental information is known for many of the other factors controlling zeolite composition.

Acknowledgments

The authors are indebted to Prof. Renato Coppi, of Rome University and Prof. Edi Sommariva of Udine University for their helpful suggestions, to Prof. Andrea Fabbri of Bologna University and Prof. Pietro Faglioni of Modena University for their critical review of the manuscript. This research was supported by the "Consiglio Nazionale delle Ricerche" of Italy, and by "Centro di Calcolo" of Modena University.

References

- Alberti, A. (1978) Differences of chemistry between hydrothermal and sedimentary zeolites. (in Italian) *Rendiconti Società Italiana di Mineralogia e Petrologia*, 34, 471–484.
- Alberti, A. (1979) Chemistry of zeolites—Q-mode multivariate factor analysis. *Chemie der Erde*, 38, 64–82.
- Alietti, A. (1972) Polymorphism and crystal-chemistry of heulandites and clinoptilolites. *American Mineralogist*, 57, 1448–1462.
- Boles, J. R. (1972) Composition, optical properties, cell dimensions and thermal stability of some heulandite group zeolites. *American Mineralogist*, 57, 1463–1493.
- Cooley, W. W. and Lohnes, P. R. (1971) *Multivariate Statistical Methods*. J. Wiley & Sons, New York.
- Coombs, D. S. and Whetten, J. T. (1967) Composition of analcime from sedimentary and burial metamorphic rocks. *Geological Society of America Bulletin*, 78, 269–282.
- Galli, E. and Loschi Ghittoni, A. G. (1972) The crystal chemistry of phillipsites. *American Mineralogist*, 57, 1125–1145.
- Gottardi, G. and Obradovic, J. (1978) Sedimentary zeolites in Europa. *Fortschritte der Mineralogie*, 56, 316–366.
- Hawkins, D. B. (1974) Statistical analyses of the zeolites clinoptilolite and heulandite. *Contributions to Mineralogy and Petrology*, 45, 27–36.
- Iijima, A. and Hay, R. L. (1968) Analcime composition in tuffs of the Green River formation of Wyoming. *American Mineralogist*, 53, 184–2000.
- Iijima, A. and Utada, M. (1972) A critical review on the occurrence of zeolites in sedimentary rocks in Japan. *Japanese Journal of Geology and Geography*, 42, 61–84.
- Mahalanobis, P. C. (1936) On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India*, 12, 49–55.
- Morrison, D. F. (1978) *Multivariate Statistical Methods*. McGraw-Hill, Auckland.
- Nie, N. H., Hadlai Hull, C., Jenkins, J. G., Steinbrenner, K. and Bent, D. H. (1975) *SPSS Statistical package for the social sciences*. McGraw-Hill, New York.
- Passaglia, E. (1970) The crystal chemistry of chabazites. *American Mineralogist*, 55, 1278–1301.
- Sheppard, R. A. and Gude, A. J. (1969) Chemical composition and physical properties of the related zeolites offretite and erionite. *American Mineralogist*, 54, 875–886.
- Sheppard, R. A. and Gude, A. J. (1970a) Calcic siliceous chabazite from the John Day formation, Grant County, Oregon. U.S. Geological Survey Professional Paper, 700D, D176–D180.
- Sheppard, R. A. and Gude, A. J. (1970b) Chemical composition and physical properties of phillipsite from the Pacific and Indian Oceans. *American Mineralogist*, 55, 2053–2062.
- SPSS-X Statistical Algorithms (1983) SPSS Inc., Chicago.
- Wilkinson, J. F. G. (1968) Analcimes from some potassic igneous rocks and aspects of analcime-rich igneous assemblages. *Contributions to Mineralogy and Petrology*, 18, 252–269.
- Wilks, S. S. (1932) Certain generalizations in the analysis of variance. *Biometrika*, 24, 471–494.

*Manuscript received, February 17, 1984;
accepted for publication, February 12, 1985.*