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

Band: 57 (1977)

Heft: 1

Artikel: Analysis of multivariate data : an application to some recent volcanics of the Central Andes

Autor: Frangipane-Gysel, Marisa

DOI: https://doi.org/10.5169/seals-44425

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 13.07.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

Schweiz. mineral. petrogr. Mitt. 57, 115–134, 1977

Analysis of Multivariate Data: An Application to some Recent Volcanics of the Central Andes

By *Marisa Frangipane-Gysel**)

Abstract

The application of Principal Component and Discriminant Analysis to recent volcanics from Central Andes illustrates a variation in chemical composition related to the distance from the Peru-Chile trench, mainly in terms of K_2O and TiO_2. The relationship between chemical composition (represented by several major oxide combinations, principal component and discriminant scores included) and depth to the seismic zone (or distance from the oceanic trench) is also considered and tested by means of Standardized Multilinear Regressions.

INTRODUCTION

The tectonic setting of the Central Andes was explained by several authors, taking into account the distribution of hypocenters beneath the continent, in terms of the Plate tectonics theory (James, 1971; Stauder, 1973; Prince and Kulm, 1975).

According to them, the oceanic Nazca plate is subducted under the S-American continental plate at the Peru-Chile trench. As the cordillera trend is strictly parallel to the oceanic trench, Prince and Kulm (1975) suggested a very close dependence of tectonic setting from subduction.

An attempt to relate a variation in chemical composition of recent volcanics from S-Peru to the distance from the oceanic trench and the depth to the seismic zone was made by Lefèvre (1973). As a result of his study, a classification of the analyses into two groups, A and B was made, mainly on the basis of diagrams such as $Na_2O + K_2O$ vs. SiO_2 and K_2O/Na_2O vs. SiO_2.

Group B, the more alkali-rich, is situated, according to Lefèvre (1973), eastward with respect to A and farther from the oceanic trench, the separating line being located at a distance of about 340 km from it.

---

*) Institut für Kristallographie und Petrographie, ETH-Zentrum, CH-8092 Zürich.

The purpose of the present study is to examine the situation in the Central Andes in a more general way, i.e., by using multivariate methods, taking into account also the author's own analyses from S-Peru and others of N-Chile, Bolivia and Argentina, available from the literature, in the attempt to give an answer to the following questions:

1. Are chemical composition (represented by the major oxides) and geographic position related in such a way to the distance from the oceanic trench, as to justify a division of the volcanics into two groups?
2. In the affirmative case, which oxides or combination of them become of relevance for such a classification?
3. Is there also a relationship between chemical composition and depth to the seismic zone?

Multivariate methods consider data in which each observational unit is characterized by several variables and evaluate changes in a large number of properties simultaneously. For example, determinations of chemical elements made on each rock specimen of a collection and arranged in tabular form, give a set of numbers referred to as a data matrix. Usually it is difficult to perceive the pattern of element associations by direct examination of such a matrix, simply because of the great number of elements and analyses involved. We need therefore, some type of quantitative method for summarizing and analysing the information hidden in a highly multivariate table. A number of statistical methods have as a common feature, the transformation of the data into fewer variables that are supposed to contain the essential information of the larger set of original observations. The overall complexity is therefore reduced in such a way as to take advantage of inherent interdependencies. This sort of data reduction becomes very useful if we want to classify patterns in terms of a number of classes defined on a priori grounds, as it is the case in the present study. The methods involved for the data reduction and the classification are the Principal Component and the Discriminant Analysis, and the results obtained from them are also taken into account in the further investigations of the relation between chemical variation and tectonic setting, by means of Standardized Multiple Regressions.

The many applications of these and other statistical methods to geological problems have been summarized by several authors, among which KOCH and LINK (1971), DAVIS (1973) and MC CAMMON (1975).

Of particular relevance for Principal Component Analysis are the works of LE MAITRE (1968), WAHLSTEDT and DAVIS (1968), TILL and COLLEY (1973), SAXENA and WALTER (1974) and BUTLER (1976), for Discriminant Analysis those of CHAYES (1964) and PEARCE (1976) and for Multiple Regression, those of NIELSON and STOIBER (1973) and HUTCHISON (1975).

COLLECTION OF DATA

Chemical analyses used in this study are of 124 recent lavas from S-Peru, N-Chile, Bolivia and Argentina, selected from those of LEFÈVRE (1973), FRANGIPANE-GYSEL (1976), PICHLER and ZEIL (1969), FRANCIS et al. (1974), FERNANDEZ et al. (1973) and HÖRMANN et al. (1973), on the basis of age and silica content.

The volcanics are of late cenozoic age and from stratovolcanoes in the Central Andes and should be representative of an active continental margin (LEFÈVRE, 1973). The SiO_2 values are of 52–63%, the generally accepted range of basaltic andesites and andesites (NIELSON and STOIBER, 1973). All analyses were recalculated to 100% after the subtraction of the volatiles. Nine variables, i.e. SiO_2, Al_2O_3, Fe_2O_3, FeO, MgO, CaO, Na_2O, K_2O and TiO_2 were considered for each rock.

In the last section of this study, the tectonic setting of each analysis is represented by the distance between the volcanic centre the analyses are from, and the Peru-Chile trench or the depth of the Benioff zone underneath. Three

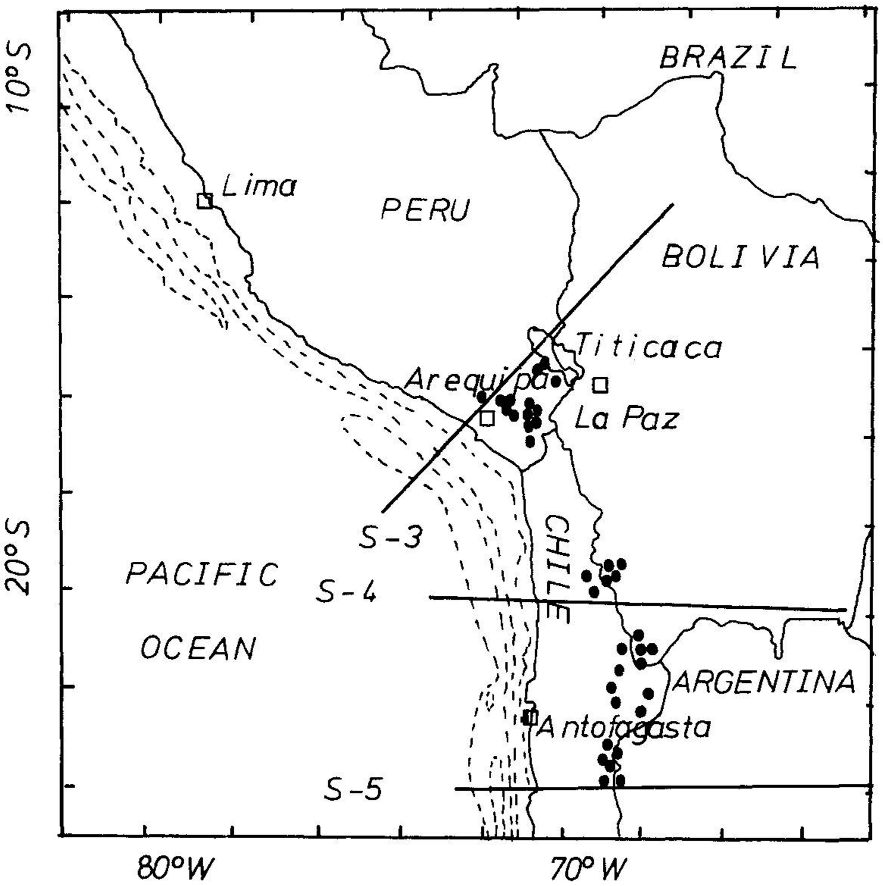

Fig. 1. The profiles are from JAMES (1971). The filled circles represent the approximative locations of the analyses.

seismic profiles from James (1971) are considered (Fig. 1) and the positions of the nearmost volcanic centers projected on them, the distance from the trench and the corresponding depth of the seismic zone are then obtained directly (Frangipane-Gysel, 1976). The distances range from 170 to 500 km and the depths from 110 to 240 km.

DATA REDUCTION AND CLASSIFICATION

As a theoretical approach is beyond the scope of the present study and the basic principles of matrix algebra and multivariate statistics were already presented at an elementary level by the authors cited in the Introduction, we will confine ourselves to a rather qualitative description of the techniques involved, giving more emphasis to their application.

Principal Component Analysis is used in the present study, as a "search procedure" in the hope to find underlying structure in the data, according to point 1.) and 2.) in the Introduction. The raw data are previously standardized and then transformed into new variables (principal components) which are independent and account for "the greatest amount of individual variable differences, not just of statistical variance" (Till and Colley, 1973). The transformation is therefore based upon a correlation matrix: standardized variables have equal weights and are independent of the magnitude of their measured units.

Till and Colley (1973) preferred to use the correlation matrix as the measure of similarity for chemical analyses in order to reduce the weight of SiO_2 (the use of a variance-covariance matrix yields a weighting according to the variance of each variable), Saxena and Walter (1974) consider that errors in the further calculations should be reduced if the correlation matrix is used.

The data are processed according to Davis (1973): from the $N \times N$ matrix of correlation (where N is the number of oxides in each analysis), N roots, called eigenvalues are extracted and, associate with each of them, there is an eigenvector or latent vector.

Elements in a $N \times N$ matrix can be regarded as defining points lying on an N-dimensional ellipsoid.

The N eigenvectors yield the principal axes of the hyperellipsoid and the sum of the diagonal elements of the matrix and the sum of the eigenvalues are identical and represent the total variance in the matrix, i.e. the trace of the similarity matrix. Thus, if the variance-covariance matrix is used, the trace will be equal to the sum of the variances of the variable, whereas if the correlation matrix is used, the trace is equal to the number of measured variables. If variables with small variances are eliminated, the effect will be more pronounced with the correlation matrix because the trace of the variance-covariance matrix will be reduced by only a small amount (BUTLER, 1976).

Eigenvalues appear in the order of their magnitude: if the first few account for an acceptable percent of the total, the remaining eigenvalues may be discarded. The percent contribution of the original variables (oxides in our case) to each principal component can be calculated and, if only a few of the original variables account for most of a principal component, it may be interpreted by considering the nature of this combination, ignoring the contribution of less significant oxides.

The problem has then been reduced in dimensionality, from one of N variables to one of $M<N$ principal components. By forming the vector product of each original analysis and the eigenvector terms, the samples can be plotted with reference ot the M new axes (principal latent vector variation diagram PLVVD, LE MAITRE, 1968).

If the inspection of such diagrams shows a grouping of analyses as a function of the corresponding oxide combination in the eigenvector term, in a petrologically meaningful way, a classification of the analyses may be made. This is achieved by means of the Discriminant Analysis. Once the classes are defined, some analyses are selected from each of them and the more significant oxides or oxide combinations, obtained from the Principal Component Analysis, considered. Let's examine the simple case in which there are two classes A and B. The problem is to find the linear combination of oxides which produces the maximum difference between A and B. If we find one or more such functions (discriminant functions) we can use them to allocate the remaining analyses or any new analysis to one of the two classes.

A simple linear discriminant function transforms an original set of measurements on a sample (in our case the most significant oxides selected by the Principal Component Analysis) into a single discriminant score. That score or transformed variable, represents the sample position along a line defined by the linear discriminant function. In this way a multivariate problem is reduced in dimensionality to a univariate one. Although the two classes may overlap, a line of division can be placed between them such that most observations from A are on one side and most observations from B on the other. Substitution of the mid point between the multivariate means of A and B into the discriminant function yields the discriminant index Ro, which is the point along the discriminant function line, exactly halfway between the center of A and B. If we calculate the discriminant function for the multivariate means of A and B, we obtain two values Ra and Rb, which define the centers of the two classes along the discriminant function. In fact, every observation can be entered into the equation and its position along the discriminant function located. The points of B that fall on the Ra side with respect to Ro and those of A on the Rb side are misclassified by the discriminant function.

A measure of the separation between the two classes is calculated by simply subtracting Ra from Rb, which is called Mahalanobis or generalized distance D^2.

APPLICATION TO THE CENTRAL ANDES VOLCANICS

The 124 analyses were arranged as the rows of a matrix, whose columns are the nine oxides mentioned previously, and then processed with a program in FORTRAN IV (Davis, 1973) to compute principal components. The results are illustrated in Table 1a and 2 and fig. 2.

Table 1a. *Correlation matrix*

	SiO_2	Al_2O_3	Fe_2O_3	FeO	MgO	CaO	Na_2O	K_2O	TiO_2
SiO_2	1.0000								
Al_2O_3	0.1740	1.0000							
Fe_2O_3	–0.1450	–0.0603	1.0000						
FeO	–0.6592	–0.1673	–0.6003	1.0000					
MgO	–0.8391	–0.4183	–0.1742	0.7814	1.0000				
CaO	–0.8535	0.0068	–0.0727	0.6844	0.7229	1.0000			
Na_2O	0.2761	0.3433	0.2488	–0.4910	–0.5578	–0.3055	1.0000		
K_2O	0.4139	–0.5075	0.2474	–0.4924	–0.3272	–0.6705	–0.0157	1.0000	
TiO_2	–0.6913	–0.5266	0.2962	0.3482	0.5113	0.4092	–0.1529	0.1168	1.0000

The elements of the correlation matrix (Tab. 1a) are the correlation coefficients r between oxides: the more their absolute values approach 1, the better the corresponding oxides are correlated. Table 1a shows conditions typical for a calc-alkaline suite: a negative correlation between CaO, MgO, FeO and SiO_2 and a positive one among MgO, FeO and CaO. As percentages (and closed data in general) are subject to the restriction that the sum of the variables measured on a given observation is a constant for all observations, the correlation coefficients were tested according to the Chayes-Kruskal test (Chayes, 1971) and the calculations performed with a program written by the author (Frangipane-Gysel, 1976).

This test involves determining the means and variances of the hypotetical open matrix X. The means and variances in X are such that closure to produce Y results in the means and variances of Y being equal to those in the observed data matrix. The correlations computed from the entries in X are zero. The closure of X to produce Y results in nonzero correlations in Y, entirely due to closure. These are used as null values against which to test the observed correlations of Table 1a. The variance of a variable is positive by definition, but negative open variances may occur in the analysis of closed chemical data.

A summary of the results is given in Table 1b. The hypothetical open variances are all positive (negative variances are considered significant if

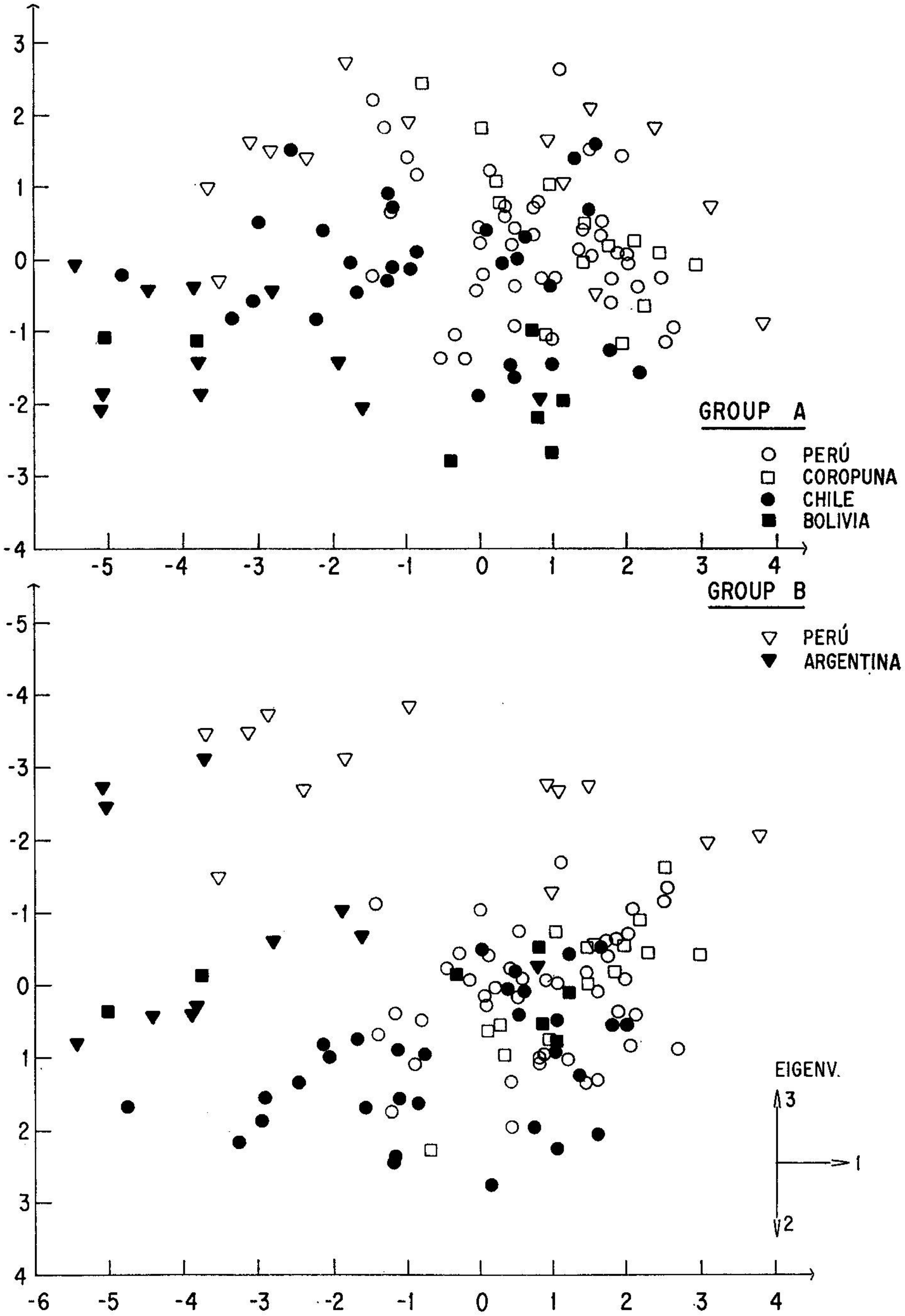

Fig. 2. PLVVD diagrams for the first three eigenvectors which represent 46.58, 22.68 and 16.55% of the trace (Table 2). The coordinates of the points are given by linear combinations of the oxides with the coefficients of the principal axes matrix (Table 2): SiO_2, MgO, FeO and CaO take on important loadings on the first eigenvector, Al_2O_3, TiO_2 and K_2O on the second and Fe_2O_3 on the third. The Coropuna analyses are from FRANGIPANE-GYSEL (1976). A and B correspond to the two groups of the Discriminant Analysis.

Table 1b

	SiO_2	Al_2O_3	Fe_2O_3	FeO	MnO	MgO	CaO	Na_2O	K_2O	TiO_2	P_2O_5
Mean	59.9	16.6	3.3	3.0	0.1	3.5	5.7	3.9	2.7	1.0	0.4
Variance	9	0.81	1.44	1.96	–	2.25	1.0	0.25	0.36	0.16	0.04
% Variance	52	4.7	8.3	11.3	–	13.0	5.8	1.50	2.1	0.9	0.23
St. dev.	3.0	0.9	1.2	1.4	–	1.5	1.0	0.5	0.6	0.4	0.2
Hypothetical open variance	41.97	–0.9	1.41	1.91	0.001	2.48	0.95	0.16	0.40	0.13	0.03
% open variance	84.9	–	2.9	3.9	–	5.02	1.92	0.32	0.8	0.26	0.06

Table 2

Eigenvector	1	2	3	4
Eigenvalue	4.1925	2.0414	1.4901	0.6186
Percent of total contribution per eigenvalue	46.58	22.68	16.55	6.87
Cumulative percent	46.58	69.27	85.82	92.7
		Principal axes matrix		
SiO_2	0.4397	0.0762	−0.3121	0.0020
TiO_2	−0.2870	−0.4340	0.2365	0.3346
Al_2O_3	0.1416	0.5584	0.2902	−0.2008
Fe_2O_3	0.1040	−0.3634	0.6052	−0.4443
FeO	−0.4277	0.1708	−0.2496	−0.2619
MgO	−0.4560	−0.0865	−0.0782	−0.0715
CaO	−0.4251	0.1620	0.2424	−0.1079
Na_2O	0.2601	0.1076	0.4568	0.7509
K_2O	0.2328	−0.5363	−0.2456	0.0345

less than −1.0, Butler, 1975). The correlations in Y are all very small and near to zero.

The apparent absence of closure effects may be the result of mixing two populations (Davis, pers. comm.).

The hypothesis of a significant correlation between variables, only due to the closure, was rejected. The four first eigenvetors describe 92.7% of the total variance of the matrix (Tab. 2) so that the remaining five may be neglected without a significant loss of information. The influence of the original variables is represented by the principal axes matrix (Tab. 2): SiO_2, MgO, FeO and CaO take on important loadings on the first eigenvector, Al_2O_3, TiO_2 and K_2O on the second and Fe_2O_3 and Na_2O on the third and fourth.

The first eigenvector defines the direction in N-space along which there is the maximum amount of "spread" in the distribution. The second eigenvector, that of maximum "spread" perpendicular to the first, the third that perpendicular to the first and second and so on. The diagrams (fig. 2) may be interpreted as variation diagrams, where the variables on the three axes are linear combinations of the above mentioned oxides. The position of an analysis along the first eigenvector is, therefore, the best possible indication of its position within the distribution or differentiation trend (Le Maitre, 1968, and Butler, 1976). The inspection of fig. 2 tells us that the differentiation trend is illustrated by a linear combination mainly of SiO_2, MgO, FeO and CaO and is independent from the geographic position of the analyses. More intersting is the fact that the analyses are roughly divided into two groups along the second eigenvector, which enables us the choice of Al_2O_3, TiO_2 and K_2O as the oxides significant for the Discriminant Analysis.

Of relevance for the last section of this study is the correspondence between the location of the analyses along the second eigenvector and their provenience (fig. 2).

The two classes in the Discriminant Analysis were chosen according to LEFÈVRE (1973), group A being separated from B along a line situated about 340 km eastward from the Peru-Chile trench.

To calculate the discriminant functions, 33 samples from S-Peru were selected for A and 15 analyses from S-Peru and 10 from Argentina for group B. Means and standard deviations are given in Table 3.

The most effective discriminant functions are listed in Table 4, Ra, Rb, Ro and the distances, in Table 5. The significance of the four functions were tested according to DAVIS (1973) and the null hypothesis that the multivariate means of A and B are equal, rejected at the 99% level.

The normality of the distribution of Al_2O_3, K_2O and TiO_2 in A and B (a

Table 3

	Group A (33 analyses)		*Group B* (25 analyses)	
	Mean	Standard deviation	Mean	Standard deviation
SiO_2	60.4	1.97	57.1	3.56
TiO_2	0.85	0.13	1.5	0.46
Al_2O_3	16.9	0.60	15.7	0.71
Fe_2O_3	3.6	0.72	3.8	1.7
FeO	2.6	0.66	3.5	1.9
MnO	0.1	0.02	0.12	0.03
MgO	2.9	0.67	4.8	1.99
CaO	5.5	0.78	6.0	1.26
Na_2O	4.0	0.28	3.6	0.52
K_2O	2.6	0.47	3.1	0.69
P_2O_5	0.32	0.05	0.59	0.25
	99.77		99.91	

Table 4. *Coefficients of the discriminant functions*

I $R = 1.3884\ Al_2O_3 - 1.2728\ K_2O - 5.6413\ TiO_2$
II $R = 1.9972\ Al_2O_3 - 4.8897\ TiO_2$
III $R = -1.9515\ K_2O - 7.0302\ TiO_2$
IV $R = -6.7449\ TiO_2$

Table 5. *Mahalanobis distance*

	R_a	R_b	R_o	$R_a - R_b$
I	15.44	9.20	12.32	6.24
II	29.72	23.89	26.81	5.83
III	−11.07	−16.81	−13.94	5.74
IV	− 5.74	−10.23	− 7.98	4.5

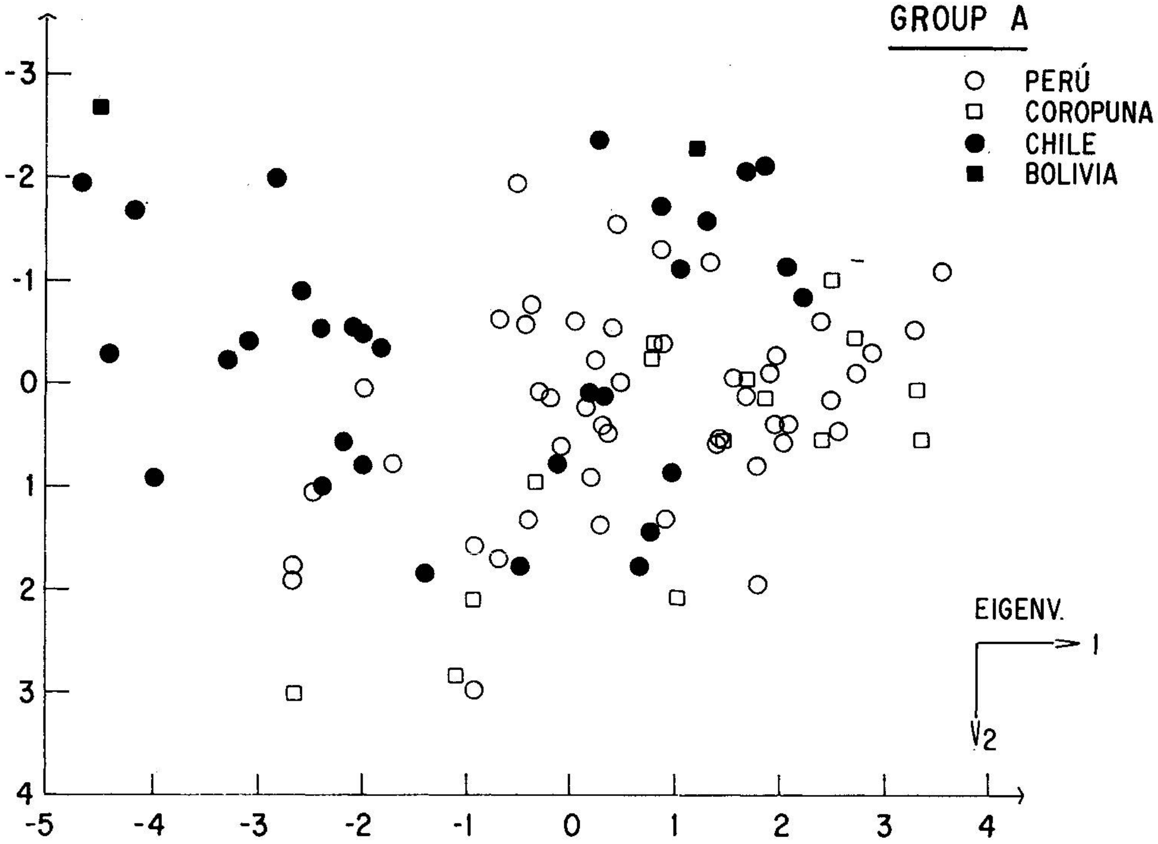

Fig. 3.

Fig. 3 and 4. Plot of the PCA scores (eigenvector 2 vs. eigenvector 1) for group A and B.

necessary requisite for the application of the Discriminant Analysis) was tested by means of the Chi-square and the Kolmogorov-Smirnov tests and the hypotesis accepted at the $\alpha = 0.05$ level (Frangipane-Gysel, 1976). The latter test, performed to compare the distributions of the oxides in the two groups, demonstrated that only Al_2O_3 belongs to the same population in A and B.

The contributions of the oxides to the discriminant functions are listed in Table 6.

When the four functions are applied to the remaining analyses, the most effective results discriminant III (K_2O and TiO_2). A reason for considering the two oxides alone is that Al_2O_3 is correlated with K_2O and TiO_2 ($r = -0.51$ and -0.53, Table 1), while the latter two are independent ($r = 0.12$).

Table 6. *Contribution (%) to the discriminant Efficiency (%)*

	Al_2O_3	K_2O	TiO_2	
I	28.65	11.03	60.32	89.4
II	44.08	–	55.92	92.4
III	–	18.37	81.63	95.5
IV	–	–	100.00	84.8

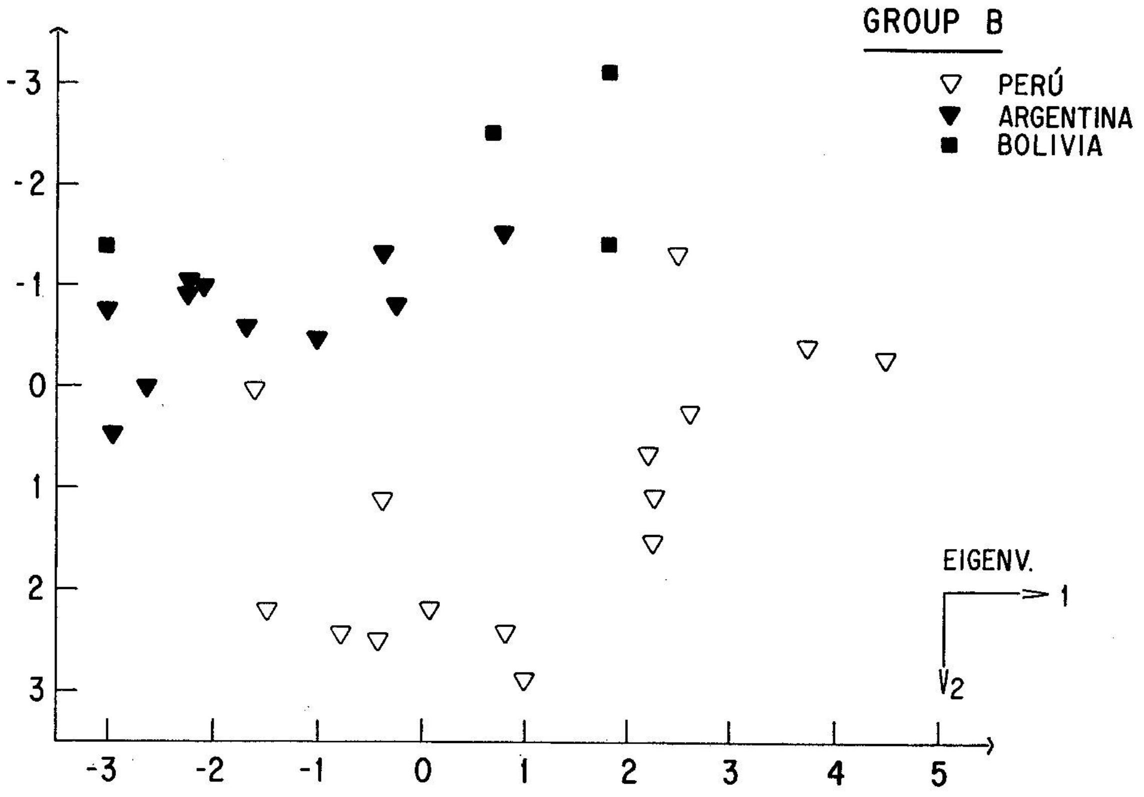

Fig. 4.

As DAVIS (1973) pointed out, if two or more variables in a discriminant function are not independent, their interaction may contribute to the distance D^2 to a greater extent than suggested by the values of Table 6. The results obtained insofar may be summarized as follows:

The number of variables is reduced from N = 9 oxides to M = 4 eigenvectors.

The calcalkaline characteristic of the rocks, already seen on diagrams such as AFM or K_2O vs. SiO_2 (FRANGIPANE-GYSEL, 1976) is supported by the Principal Component Analysis.

Eigenvector 2 isolates K_2O, TiO_2 and Al_2O_3 as the more significant to define a chemical variation among the analyses. A correspondence with areal distribution is confirmed by the Discriminant Analysis which considers also the increase of TiO_2 and K_2O with the distance from the Peru-Chile trench.

PCA analysis may be performed also on A and B separately (see fig. 3 and 4): the purpose is once more to illustrate eventual structures in the groups and to compare the results with the previous PCA:

Group A *(Table 6a):* K_2O shows a better correlation with SiO_2 ($r = 0.76$) and CaO ($r = -0.81$) if compared with *Table 1.*

Table 6a. *Correlation matrix for group A*

	SiO	Al_2O_3	Fe_2O_3	FeO	MgO	CaO	Na_2O	K_2O	TiO_2
SiO_2	1.0000								
Al_2O_3	−0.3523	1.0000							
Fe_2O_3	−0.1125	0.0227	1.0000						
FeO	−0.6573	0.1521	−0.6238	1.0000					
MgO	−0.7563	−0.1335	−0.1871	0.7295	1.0000				
CaO	−0.9217	0.3079	−0.0346	0.6883	0.7636	1.0000			
Na_2O	0.1087	0.1838	0.1162	−0.2582	−0.4463	−0.2601	1.0000		
K_2O	0.7641	−0.5035	0.1308	−0.6603	−0.6126	−0.8119	0.0371	1.0000	
TiO_2	−0.5421	−0.0266	0.0936	0.3380	0.3248	0.4709	0.0529	−0.3022	1.0000

Group B *(Table 6b)*: Na_2O and K_2O show slight better correlations with FeO as in *Table 1.*

Table 6b. *Correlation matrix for group B*

	SiO	Al_2O_3	Fe_2O_3	FeO	MgO	CaO	Na_2O	K_2O	TiO_2
SiO_2	1.0000								
Al_2O_3	0.4671	1.0000							
Fe_2O_3	−0.0871	−0.0041	1.0000						
FeO	−0.5767	−0.3225	−0.3223	1.0000					
MgO	−0.8532	−0.3975	−0.3446	0.7990	1.0000				
CaO	−0.8564	−0.4398	−0.1686	0.6871	0.7756	1.0000			
Na_2O	0.2237	0.2163	0.6435	−0.7267	−0.5510	−0.3279	1.0000		
K_2O	0.4885	−0.1379	0.3741	−0.6763	−0.5784	−0.6668	0.2977	1.0000	
TiO_2	−0.6885	−0.6509	0.4164	0.1683	0.3288	0.4553	0.1387	0.0063	1.0000

Table 6c illustrates the eigenvectors and the principal axes matrix for the two groups.

Table 6c

Group A				Group B			
Eigv.	Eigval.	% of total contribution	cumulative %	Eigv.	Eigval.	% of total contribution	cumulative %
1	4.3710	48.5672	48.5672	1	4.5910	51.0116	51.0116
2	1.5979	17.7549	66.3220	2	2.3107	25.6748	76.6863
3	1.2666	14.0736	80.3957	3	1.1100	12.3330	89.0193
4	0.9203	10.2261	90.6219	4	0.3947	4.3854	93.4043
5	0.4862	5.4026	96.0243				

Principal axes matrix

	Group A					Group B			
Eigenvector	1	2	3	4	5	1	2	3	4
SiO_2	0.4409	−0.2157	0.1515	0.0290	−0.1043	0.3922	−0.3005	−0.1863	0.2768
Al_2O	−0.1411	0.4953	0.5041	−0.3291	−0.4107	0.2276	−0.3356	0.5974	−0.3372
Fe_2O	0.1000	0.4976	−0.5919	−0.3207	0.1361	0.1974	0.5240	0.2525	−0.4418
FeO	−0.4052	−0.2695	0.2604	0.1933	−0.0262	−0.4242	−0.2235	−0.0923	0.1907
MgO	−0.4040	−0.2688	−0.2177	−0.0712	0.3729	−0.4329	0.0064	0.0367	−0.3619
CaO	−0.4515	0.0959	−0.0923	−0.1131	0.0293	−0.4142	0.1311	0.2114	0.1444
Na_2O	0.1300	0.4880	0.2495	0.5922	0.5376	0.2822	0.3421	0.3924	0.5660
K_2O	0.4097	−0.1973	−0.2085	0.1053	−0.2022	0.3148	0.1990	−0.5700	−0.2829
TiO_2	−0.2443	0.1714	−0.3783	0.6091	−0.5757	−0.1942	0.5465	−0.1049	0.1643

Table 7

	Dependent variable	Independent variables						Seismic profiles								
								S-3	70 analyses		S-4	24 analyses		S-5	30 analyses	
		K_2O	Na_2O	SiO_2	τ	eigen-vector 1	depth	r	$100 \times r^2$	signif. of r	r	$100 \times r^2$	signif. of r	r	$100 \times r^2$	signif. of r
1	K_2O						×	0.39	15	0.0005	0.45	20	0.025	0.47	22	0.005
2	K_2O			×			×	0.68	46	0.0005	0.88	78	0.0005	0.63	40	0.0005
3	K_2O		×	×			×	0.68	46	0.0005	0.89	79	0.0005	0.78	65	0.0005
4	K_2O			×	×		×	0.82	66	0.0005	0.95	90	0.0005	0.68	46	0.0005
5	TiO_2						×	0.50	25	0.0005	0.0027	0	0.0	0.51	27	0.0005
6	τ						×	0.42	17	0.0005	0.0	0		0.22	5	0.005
7	τ	×		×			×	0.81	66	0.0005	0.82	67	0.0005	0.88	77	0.0005
8	Na_2O+K_2O			×			×	0.57	32	0.0005	0.87	76	0.0005	0.63	39	0.0005
9	Na_2O+K_2O/Al_2O_3			×			×	0.52	27	0.0005	0.82	67	0.0005	0.61	37	0.0005
10	eigenvector 2					×	×	0.63	40	0.0005	0.43	19	0.05	0.62	39	0.0005
11	eigenvector 2						×	0.63	40	0.0005	0.28	7	0.05	0.58	34	0.0005
12	discriminant III						×	0.05	0.2	0.05	0.04	2	0.05	0.59	35	0.0005
							distance									
13	K_2O			×			×	0.69	47	0.0005	0.88	78	0.0005	0.74	55	0.0005
14	TiO_2						×	0.49	24	0.0005	0.0	0		0.59	35	0.0005
15	eigenvector 2					×	×	0.63	40	0.0005	0.40	16	0.05	0.71	50	0.0005
16	eigenvector 2						×	0.63	40	0.0005	0.21	5	0.05	0.69	48	0.0005

Group A: five eigenvectors explain 96.02 of the total variance in the similarity matrix and the contributions of Na_2O, K_2O and TiO_2 to eigenvector II are larger in *Table 2* and *6c*, respectively.

Group B: four eigenvectors explain 93.4 of the total variance and the contribution of K_2O to the second eigenvector is larger in *Table 2*. The PCA scores plots of fig. 3 shows a separation of the Peru and Argentina analyses in group B, along the second eigenvector (i.e. as a function, mainly of TiO_2 and Fe_2O_3). The Chayes-Kruskal test was possible for group A only, because the open variance of Al_2O_3 in group B was significantly negative (-1.84). The hypothesis of correlations due to closure was rejected for group A.

DEPENDENCE OF CHEMICAL VARIATION ON TECTONIC SETTING BY MEANS OF A STANDARDIZED MULTILINEAR REGRESSION

An attempt to test if the chemical variation, illustrated in the previous section, may also be related to the tectonic setting (see Collection of data) is made, according to 3.) in the Introduction, by means of several Standardized Multilinear Regressions (Davis, 1973). Some of the chemical parameters of Table 7 should illustrate a relationship between the characteristic and tectonic setting of a magma:

$(Na_2O+K_2O)/Al_2O_3$ combined with SiO_2 gives informations about the alkali contribution of the mantle to the parent magma (Myiashiro, 1974) $\tau = (Al_2O_3-Na_2O)/TiO_2$ indicates the difference between orogenic and cratonic andesites (Rittmann, 1973) and K_2O together with SiO_2 should be related to the depth of the seismic zone (Nielson and Stoiber, 1973).

The coefficient r is a measure of the linear correlation between the dependent variable and the independent ones. Only values above 0.48 are considered satisfactory according to Nielson and Stoiber (1973). The significance of r is the probability at which the hypothesis of null correlation may be rejected and the percentage of goodness of fit is represented by $r^2 \times 100$.

Some linear relationships are illustrated in Table 8. The best correlations are obtained when τ is considered together with K_2O.

Eigenvector 2 is correlated with depth and distance in profile S-3 and S-5. If the correlations of K_2O as a function of SiO_2 and depth or distance are compared with that of eigenvector 2 as a function of depth or distance, r is about the same in the two cases, for profile S-3 and S-5 but the eigenvector shows the advantage of a steeper regression line and a bidimensional representation.

It is often of interest to measure the correlation between a dependent variable and one particular independent variable when the effects of all other variables are removed.

Table 8. *Linear regression* $Y = a + \sum_{i=1}^{n} b_i x_i$

N represents the first column in Table 7

N	n	Variables		Coefficients S-3	S-4	S-5
2	2	$Y = K_2O$	$a =$	−6.07	−8.2	−3.8
		$x_1 = SiO_2$	$b_1 =$	0.13	0.17	0.08
		x_2 = depth	$b_2 =$	0.01	0.004	0.02
7	3	$Y = \tau$	$a =$	−62.5	−61.9	−31.9
		$x_1 = SiO_2$	$b_1 =$	1.44	1.51	0.85
		$x_2 = K_2O$	$b_2 =$	−4.48	−6.07	−1.14
		x_3 = depth	$b_3 =$	0.02	0.01	−0.03
10	2	Y = eigenvector 2	$a =$	2.71	1.1	4.1
		x_1 = eigenvector 1	$b_1 =$	0.02	0.16	0.14
		x_2 = depth	$b_2 =$	−0.03	−0.01	−0.04
11	1	Y = eigenvector 2	$a =$	2.75	2.09	4.37
		x_1 = depth	$b_1 =$	−0.02	−0.01	−0.04
13	2	$Y = K_2O$	$a =$	−6.48	−8.6	−5.55
		$x_1 = SiO_2$	$b_1 =$	0.13	0.17	0.11
		x_2 = distance	$b_2 =$	0.005	0.002	0.006
16	1	Y = eigenvector 2	$a =$	3.6	2.4	3.7
		x_1 = distance	$b_1 =$	−0.01	−0.004	−0.01

HATHERTON and DICKINSON (1969) used variation diagrams for individual volcanoes in order to obtain a K_2O value at specific SiO_2 values (55 and 60%) to be related with the depth of the Benioff zone.

The partial correlation analysis, preferred in this paper, is an excellent technique for understanding and clarifying relationships among variables.

Suppose that compositional variables a and b are positively correlated and b, in turn is positively correlated with c (for example, depth of the Benioff zone), the question is, wether the amount of a expectable at some particular b value is still correlated with depth. Another example of partial correlation is the uncovering of spurious relationships: the correlation between a and c results spurious if a varies along with some other variable b, which is indeed the true predictor of c. If the effects of b are controlled (b held constant), a no longer varies with c.

Partial correlation analysis may also be useful to locate relationships where none appear to exist. Variable a shows no relationship to c because a is negatively related to b, which is positively related to c. Hence a is positively related to c when one controls the effects of b.

The partial correlation coefficients for three variables (Table 10) are calculated from the correlation coefficients of Table 9, according to the following equation:

$$r_{ac \cdot b} = \frac{r_{ac} - r_{ab}\, r_{bc}}{\sqrt{(1 - r_{ab}^2)(1 - r_{bc}^2)}},$$

Table 9

$r_{a\cdot b}$				
a	b	S-3	S-4	S-5
K_2O	SiO_2	0.36	0.86	0.27
K_2O	distance	0.39	0.42	0.48
K_2O	depth	0.38	0.45	0.47
SiO_2	distance	−0.4	0.27	−0.3
SiO_2	depth	−0.4	0.28	−0.28
SiO_2	K_2O+Na_2O	0.44	0.8	0.37
τ	SiO_2	0.64	0.48	0.82
K_2O+Na_2O	depth	0.16	−0.11	0.38
τ	depth	−0.42	−0.003	−0.47
τ	K_2O	−0.2	0.07	−0.06
$(K_2O+Na_2O)/Al_2O_3$	SiO_2	0.26	0.71	0.08
$(K_2O+Na_2O)/Al_2O_3$	depth	0.24	−0.195	0.5
eigenvector 2	eigenvector 1	0.22	−0.36	0.39
eigenvector 2	distance	−0.63	−0.22	−0.69
eigenvector 2	depth	−0.63	−0.28	−0.58
eigenvector 1	depth	−0.3	0.1	−0.37
eigenvector 1	distance	−0.32	−0.1	−0.3
Na_2O	K_2O	−0.15	−0.12	−0.59
Na_2O	SiO_2	0.2	0.13	0.07
Na_2O	depth	−0.38	−0.74	−0.24

Table 10

			$r_{ac\cdot b}$			$r_{bc\cdot a}$		
a	b	c	S-3	S-4	S-5	S-3	S-4	S-5
K_2O	SiO_2	depth	0.61	0.43	0.59	−0.62	−0.23	−0.48
K_2O	SiO_2	distance	0.62	0.38	0.61	−0.62	−0.20	−0.51
Na_2O+K_2O	SiO_2	depth	0.41	−0.58	0.54	−0.53	0.68	−0.49
τ	K_2O	depth	−0.38	−0.04	−0.5	0.33	0.45	0.5
$(Na_2O+K_2O)/Al_2O_3$	SiO_2	depth	0.39	−0.58	0.55	−0.49	0.61	−0.37
eigenv. 2	eigenv. 1	distance	−0.61	−0.2	−0.64	−0.21	0.02	−0.15
eigenv. 2	eigenv. 1	depth	−0.61	−0.34	−0.53	−0.24	−0.22	−0.10
K_2O	Na_2O	depth	0.35	0.54	0.42	−0.35	−0.77	0.05
Na_2O	SiO_2	depth	−0.33	−0.82	−0.23	−0.36	0.57	−0.27
τ	SiO_2	depth	−0.23	−0.29	−0.44	−0.19	0.38	0.21

where $r_{ac.b}$ is the coefficient of partial correlation between a and c when b is kept constant and r_{ac}, r_{ab}, r_{bc} are the corresponding correlation coefficients between variables.

If four variables are considered (Table 11), the above equation takes the form:

$$r_{ad\cdot bc} = \frac{r_{ad\cdot c} - r_{ab\cdot c}\, r_{bd\cdot c}}{\sqrt{(1-r^2_{ab\cdot c})\,(1-r^2_{bd\cdot c})}},$$

where $r_{ad.bc}$ refers to the correlation between a and d, when b and c are held constant.

The partial correlation coefficients listed in Table 10 and 11 are compared with the correlation coefficients of Table 9 to evaluate the mutual influence of the variables involved. The effects of some compositional variables on the correlations between other compositional variables and depth of the Benioff zone may be summarized as follows:

$$a = K_2O, \quad b = SiO_2, \quad c = \text{depth}.$$

$r_{a.b}$ is always positive, if SiO_2 is held constant, the correlation improves in S-3 and S-5 but is not affected in S-4.

The situation in S-3 and S-5 may be explained taking into account the fact that a is positively correlated with c and b, while b is negatively correlated with c. Profile S-4 behaves in a different way: r_{ac} and $r_{ac.b}$ are almost equal because b is weakly (and positively) correlated with c.

$$a = Na_2O + K_2O \ (\text{or } [Na_2O + K_2O]/Al_2O_3), \quad b = SiO_2, \quad c = \text{depth}.$$

The situation in profile S-3 and S-5 is very similar to the above case. The negative correlation in S-4 improves greatly (from Table 9 to Table 10) if b is held constant, because of the strong positive value of r_{ab}.

$$a = \text{eigenvector } 2, \quad b = \text{eigenvector } 1, \quad c = \text{depth}.$$

Eigenvector 2 is negatively correlated with depth and the correlation coefficients are almost not affected at constant b.

$$a = \tau, \quad b = SiO_2, \quad c = \text{depth}.$$

In profile S-3 and S-5, a is strongly correlated with b, which is negatively correlated with c, if, therefore, b is held constant, $r_{ac.b}$ results smaller than r_{ac}. The situation is reversed in S-4, where a weak negative correlation results only if b is held constant (r_{ac} is almost zero).

The results of partial correlation for four variables are summarized in Table 11.

Table 11

					$r_{ad\cdot bc}$			$r_{bd\cdot ac}$	
a	b	c	d	S-3	S-4	S-5	S-3	S-4	S-5
K_2O	Na_2O	SiO_2	depth	0.58	0.10	0.59	−0.24	−0.78	0.23
K_2O	τ	SiO_2	depth	0.61	0.34	0.47	0.21	0.07	−0.20

The correlations between K_2O and depth improves or remains constant (for S-3 and S-5) with respect to Table 9 if SiO_2 and Na_2O or τ are held constant simultaneously, while the reverse effect may be seen for profile S-4.

The regressions of Table 7 are able to explain only 60–80% of the total variation of chemical characteristic as a function of the tectonic setting, there are certainly other factors involved, such as magma inhomogeneities at the origin and/or due to contamination during its rise, or errors in the location or accuracy of the analyses.

SUMMARY AND CONCLUSIONS

The present study illustrates the application of some statistical methods to the recent volcanics of the Central Andes.

TiO_2 and K_2O are isolated as the most effective oxides to define a chemical variation associated with the areal distribution of the analyses.

The increase in K_2O and TiO_2 eastward from the oceanic trench is related to the increasing depth to the underlying seismic zone. Should the parental magmas be more or less directly related with the depth of the subduction zone, then the increase in K_2O could be explained by the progressive breakdown of phlogopite from a source situated within a subducted portion of the upper mantle (BESWICK, 1976). Experimental data on titan-phlogopite (FORBES and FLOWER, 1974) show it to be stable to substantially higher temperatures than normal phlogopite and breakdown of this phase during melting at depth (more than 150 km) on subducted slabs is believed to be responsible for the concomitant increase in K and Ti observed in magmas erupted during orogenic volcanism (FORBES and FLOWER, 1974).

The advantage of multivariate data reduction becomes apparent in the present application and may be recomended to the treatment of many problems.

When applying the scheme proposed here, care should be taken to test the data for normality and, if necessary, for closure. More details about the data properties required to apply the multivariate methods of this study are to be found in DAVIS (1973), PEARCE (1976), LE MAITRE (1968) and CHAYES (1964).

As for the multilinear regressions, a partial correlation approach may be useful to investigate the mutual effects on the involved variables.

The rather cumbersome calculations are performed easily and quickly by the computer and the only problem is to select the input data and evaluate the results in a geologically meaningful way.

Acknowledgements

My thanks are due to Dr. F. Chayes and Prof. Dr. J. C. Davis for critically reading an earlier version of the paper.

The present study being a part of her PhD thesis, the autor wishes to thank Prof. Dr. M. Weibel and PD Dr. R. Schmid for many comments and useful discussions during her work.

References

BESWICK, A. E. (1976): K and Rb relations in basalts and other mantle derived materials. Is phlogopite the key? Geoch. and Cosm. Acta, 40, 1167–1184.

CHAYES, F. (1964): A petrografic distinction between Cenozoic volcanics in and around the open oceans. J. Geophys. Res., 69, 1573–1588.

CHAYES, F. (1971): Ratio correlation; a manual for students of petrology and geochemistry. Univ. Chicago Press.

DAVIS, J. C. (1973): Statistics and data analysis in Geology. Wiley and Sons.

FERNANDEZ, A., P. K. HOERMANN, S. KUSSMAUL, J. MEAVE, H. PICHLER and T. SUBIETA (1973): First petrologic data on young volcanic rocks of S.W. Bolivia. Tscherm. Mineral. Petrogr. Mitt., 19, 149–172.

FORBES, W. C. and M. F. J. FLOWER (1974): Phase relations of titan-phlogopite $K_2Mg_4TiAl_2Si_6O_{20}(OH)_4$: a refractory phase in the upper mantle? Earth Plan. Sci. Lett., 22, 60–66.

FRANCIS, P. W., M. J. ROOBOL, G. P. L. WALKER, P. R. COBBOLD and M. COWARD (1974): The S. Pedro and S. Pablo volcanoes of northern Chile and their hot avalanche deposits. Geol. Rdsch., 63, 357–388.

FRANGIPANE-GYSEL, M. (1976): Studio geochimico-petrografico del Nevado Coropuna (Perù Meridionale). Dissertation, ETH, Zürich.

HINCHEN, J. D. (1969): Practical statistics for chemical research. Methuen and Co and Science Paperbacks.

HOERMANN, P. K., H. PICHLER and W. ZELL (1973): New data on the young volcanism in the Puna of N.W. Argentina. Geol. Rdsch., 62, 397–418.

HUTCHISON, C. S. (1975): Correlation of indonesian active volcanic geochemistry with Benioff zone depth. Geol. en Mijnb., 54, 157–168.

JAMES, D. (1971): Plate tectonic Model for the evolution of the Central Andes. Geol. Soc. Amer. Bull., 82, 3325–3346.

KOCH, G. S. and R. F. LINK (1971): Statistical analysis of geological data. Wiley and Sons.

LEFÈVRE, C. (1973): Les caractères magmatiques du volcanisme plio-quaternaire des Andes dans le Sud du Pérou. Contr. Mineral. and Petrol., 41, 259–272.

LE MAITRE, R. W. (1968): Chemical variation within and between volcanic rock series – A statistical approach. J. Petrol., 9, 220–252.

MC CAMMON, R. B. (1975): Concepts in Geostatistics. Springer Verlag.

MIYASHIRO, A. (1975): Volcanic rock series and tectonic setting. Annu. Rev. of Earth and Planet. Sci., 3, 463 p.

NIELSON, D. R. and R. STOIBER (1973): Relationship of potassium content in andesitic lavas and depth to the seismic zone. J. Geophys. Res., 78, 6887–6892.

PARCE, J. A. (1976): Statistical analysis of major element patterns in basalts. J. Petrol., 17, 15–43.

PICHLER, H. and W. ZEIL (1969): Die Quartäre «Andesit»-Formation in der Hochkordillere Nord-Chiles. Geol. Rdsch., 58, 866–903.

PRINCE, R. A. and R. D. KULM (1975): Crustal rupture and the initiation of imbricate thrusting in the Peru-Chile trench. Geol. Soc. Amer. Bull., 86, 1639–1653.

RITTMANN, A. (1973): Stable mineral assemblages of igneous rocks. Springer Verlag.

SAXENA, S. K. and L. S. WALTER (1974): A statistical-chemical and thermodynamic approach to the study of lunar mineralogy. Geoch. and Cosm. Acta, 38, 79–95.

STAUDER, W. M. (1973): Mechanism and spatial distribution of Chilean earthquakes with relation to subduction of the oceanic plate. J. Geophys. Res., 78, 5033–5061.

TILL, R. and H. COLLEY (1973): Thoughts on use of Principal Component Analysis in Petrogenetic Problems. Math. Geol., 5, 341–350.

WAHLSTEDT, W. G. and J. C. DAVIS (1968): FORTRAN IV program for computation and display of principal components. Kansas Geol. Survey Computer Contr. 21, 27 p.

Other references

BUTLER, J. C. (1975): Occurrence of negative open variances in Ternary systems. Math. Geol., 7, 31–45.

BUTLER, J. C. (1976): Principal Component analysis using the hypothetical closed array. Math. Geol., 8, 25–36.

HATHERTON, W. and W. R. DICKINSON (1969): The relationship between andesitic volcanism and seismicity in Indonesia, the Lesser Antilles and other island arcs. Jour. Geophys. Res., 74, 5301–5310.

Manuscript received January 31, 1977.