FINAL REPORT

NSG-7198

Submitted by:

John C. Butler
Department of Geology
University of Houston
Houston, Texas 77004

(NASA-CR-153063) CLOSURE AND RATIO CORRELATION ANALYSIS OF LUNAR CHEMICAL AND GRAIN SIZE DATA Final Report (Houston Univ.) 50 p HC A03/MF A01 CSCL 03B
N77-25038
Unclas
G3/91 24474

CLOSURE AND RATIO CORRELATION ANALYSIS OF LUNAR CHEMICAL AND GRAIN SIZE DATA

TABLE OF CONTENTS

EFFECTS OF CLOSURE ON MAJOR AND TRACE ELEMENT VARIATIONS IN APOLLO 11, 12 AND 15 MARE BASALTS AND REGOLITH SAMPLES

ABSTRACT

Major element and major element plus trace element analyses have been selected from the lunar data base for Apollo 11, 12 and 15 basalt and regolith samples. Summary statistics for each of the six data sets have been compiled and the effects of closure on the Pearson product moment correlation coefficient have been investigated using the Chayes and Kruskal approximation procedure. In general there are two types of closure effects evident in these data sets: (1) negative correlations of intermediate size which are solely the result of closure and (2) correlations of small absolute value which depart significantly from their expected closure correlations which are of intermediate size. The Apollo 15 regolith samples do not conform to the Chayes and Kruskal model because SiO_2 is behaving as a constant and this set of analyses can not be meaningfully compared with the other five data sets. Particular attention was directed towards the possibility of inducing a strong positive correlation between variables present in trace amounts as a result of closure. It is shown that a positive closure correlation will arise only when the product of the coefficients of variation is very small (less than 0.01 for most data sets) and, in general, trace elements in the lunar data sets exhibit relatively large coefficients of variation. Preliminary experiments suggest that closure strongly effects both the Spearman and Kendall rank correlation coefficients and, at the present time, the closure effect can not be accounted for. These rank correlations may be better suited to assessing the strength of association between geochemical variables than is the Pearson product moment correlation coefficient and simulation experiments are being continued.

EFFECTS OF CLOSURE ON MAJOR AND TRACE ELEMENT VARIATIONS IN APOLLO 11, 12 AND 15 MARE BASALTS AND REGOLITH SAMPLES

Introduction

It is a well established fact that percentages formed from uncorrelated parent variables will be correlated as a result of percentage formation (summarized in Chayes, 1971). As an extreme examples, one can form percentages from a set of binary data (% X + % Y = 100%), and regardless of the correlation between X and Y the correlation between X and Y will be -1.0. If the parent variables are homogeneous in mean and variance the correlation induced by percentage formation is given by:

$p_{ij} = -1.0/(M-1.0)$ Equation (1) (Chayes, 1971)

where M is the number of variables in the closed data set (set of percentages). As M increases the magnitude of the induced correlation decreases ($p_{ij} = -0.5$ for $M = 3$ and $p_{ij} = -0.1$ for $M = 11$). However, a data array of geochemical interest with variables homogeneous in mean and variance is extremely unlikely so that one can not argue that the closure effect can be reduced simply by increasing the number of variables in the closed data array.

Chayes and Kruskal (1966) developed an approximation method for computing the correlation that will arise solely as the result of percentage formation from uncorrelated parent variables. This closure correlation (p_{ij}) can be used as the null value for testing an observed correlation for significant departures from randomness. Indiscriminant use of zero as a null value or a visual interpretation of the strength of a variation diagram may result in believing that a correlation is significant when in fact closure alone is responsible. Similarly, correlations of relatively small magnitude, which require accepting the null hypothesis when zero is used as the null, may be significant if the closure correlation is relatively large.

Whole sample major and trace element analyses of Apollo 11, 12 and 15 basalts and regolith samples have been compiled and closure effects assessed. Particular attention has been directed towards examining the possibilities of induced positive correlation between variables present in trace amounts.

The Chayes and Kruskal Test

Approximation of the expected closure correlations for the observed data array U requires estimation of the means and variances of the variables in the X array - the hypothetical closed array - which is characterized by possessing zero covariances and hence zero correlations. When X is closed to form Y - the hypothetical closed array - the means and variances of the variables in Y are equal to those in the observed data array U. As the correlation in X are zero, the correlations in Y are solely the result of closure.

The means of the variables in X are the observed mean proportions of the variables in U (Chayes, 1971). Hypothetical open variances can be computed from the observed mean proportions and variances of the parent variables in U. The sum of the open variances (σ_t^2) is given by:

$$\sigma_t^2 = \frac{\mathop{}\limits_{i=1}^{M} (s_i^2/(1.0-2.0\bar{y}))}{\mathop{}\limits_{i=1}^{M} ((\bar{y}_i(1.0-\bar{y}_i))/(1.0-2.0\bar{y}_i))}$$

Equation (2) (Chayes, 1971, Eq. 9.1)

where $\bar{y}_i$ and s_i^2 are the observed mean proportion and variance of the ith variable respectively. Each open variance (o_i^2) can be computed from:

$$o_i^2 = \frac{\bar{y}_i^2}{(1.0-2.0y_i)} \quad (C_i^2 - o_t^2)$$

Equation (3) (Chayes, 1971, Eq. 9.2)

where C_i is the observed coefficient of variation of the ith variable.

If none of the variables has a mean proportion greater than 0.50 an open variance will be negative if and only if C_i^2 is less than σ_t^2 (Equation 3).

Assuming that all of the open variances are positive one can compute the correlation induced by closure (p_{ij}) as follows:

$$P_{ij} = (\bar{y}_i \bar{y}_j \sigma_t^2 - \bar{y}_j \sigma_i^2 - \bar{y}_i \sigma_j^2)/s_i s_j$$ Equation (4) (Chayes, 1971, Eq. 6.1)

where $\bar{y}_n$, s_n and σ_n^2 are the observed mean proportion, the observed standard deviation and the open variance of the nth variable respectively.

In general, the greater the variances of the variables the greater the induced negative cor-elation and, for variables with small variances, the induced correlation may be positive. Chayes and Kruskal (1966) noted that the conditions for positive null correlation would most likely be satisfied by variables with small means and variances and concluded that null correlations between variables present in trace amounts would often (if not always) be positive.

Thus, the investigator is faced with a definite problem in assessing the strength of a geochemical variation diagram and/or in testing an observed correlation for significant departures from randomness. The apriori use of zero as a null value must be avoided for percentage data and, in general, each correlation must be considered to possess its own null value. Also, one must have a complete chemical analysis to evaluate σ_t^2; that is, partial chemical analyses can not be tested using the Chayes and Kruskal model.

Effect of Closure on Correlation Between Major Elements

The lunar data base (updated version of March, 1976, supplied by J. Warner of JSC) served as the primary data base. Subsets of Apollo 11, 12 and 15 basalts and regoliths have been extracted at the present time. Each subset has been carefully examined, duplicate analysis or entries removed and anomalous entries verified by checking the original reference. These corrected subsets are serving as the data bases for continuing studies of within and between mission natural groupings (Butler, 1976), theoretical frequency distributions of major and trace elements and the effects of closure.

Whole sample (coded ALL in the lunar data base) major element and major plus trace element analyses were selected for each mission using the following criteria:

(1) SiO_2, Al_2O_3, TiO_2, MnO, MgO, CaO, Na_2O, K_2O and Cr_2O_3 were analyzed for by the same investigator(s) and published in the same reference;

(2) the sum of the major elements was between 98.5 and 101.5

(3) trace element analyses were published in the same original reference as were the major element analyses.

Originally it was desired to include P_2O_5 and S but there is an apparent paucity of information concerning these two variables (especially in the Apollo 15 basalt and regolith samples). Whenever possible S and P_2O_5 analyses or averages were included to provide more complete data arrays.

In general there are approximately 2 to 3 times as many major element analyses as there are major plus trace element analyses for a given mission. In fact, there were less than 10 samples of major plus trace element analyses for the Apollo 11 and 12 regolith samples so these were excluded from further consideration.

Several previous investigators have published averages of basalts and regoliths (for example, Taylor, 1975 and Rose et al., 1976). Frequently such averages have been computed using limited portions of the total available information and I am not aware of published averages making use of all available information. This is especially critical for least squares mixing model studies. However, one must question the meaning of any set of computed averages. For example, there are some samples for which many analyses are available (15555 for example) and averages created without weighting may prove difficult to interpret. If one computed an average composition for each sample and averaged the results to give an average composition for the mission the weighting would be equal for each sample and not enough is known about the sampling efficiency at each site to warrant an equal weighting assumption. Difficulties arise, for example, when averages of basalt and regolith samples are compared. If one has confidence in the averages one may be compelled to postulate the presence of an "exotic" component or process to account for any observed differences. Taylor (1975) concluded that the "true abundance" of aluminous basalts has been underestimated at the Apollo 11 and 12 sites with the result that a source for the higher Al_2O_3 of the regolith must be identified.

Summary statistics for the major element analyses which satisfied the previously cited criteria are given in Table 1 (mean, coefficient of variation (C) and % of the total variance (% s^2)). More detailed sets of summary statistics for each analyzed sample may be obtained by writing the author. Use of such summaries, however, must be undertaken with considerable discretion.

Chayes (1965) advocated computing the sum of the variances as a measure of the total variability of the data array. From Table 1 it is evident that the regolith samples have significantly lower total variances than the basalts for a given mission. Also, the variability of a given variable (as estimated from the coefficient of variation) appears to be greater for the basalts than for the regoliths for a given mission. In part this might be the result of fewer samples in the regolith suite but most likely reflects the "gardening" processes accompanying regolith formation.

Open variances were computed for each of the 6 data sets summarized in Table 1. The open variance for SiO_2 in the Apollo 15 regolith suite is negative so that individual correlations can not be tested for significant departures from randomness. The coefficient of variation of SiO_2 is 0.013 for the Apollo 15 regolith samples and C^2SiO_2 is less than the sum of the open variances (0.00456) so that the open variance for SiO_2 is negative because SiO_2 is behaving as a quasi constant. All of the remaining 5 suites possess positive open variances. Rather than list all 66 correlations for each of the 5 data sets that conform to the Chayes and Kruskal model it was decided to list only those for which the result of testing against the closure correlation (Equation 4) significantly differed from testing against a null of zero (Table 2).

It is evident from Table 2 that two types of misinformation (Butler, 1976) are present in the 5 sets of lunar chemical analyses:

(1) correlations which would be significant if tested against a null of zero but fail to significantly depart from their expected closure correlation;

(2) correlations of small absolute value which would require acceptance of the null hypothesis if tested against zero but significantly depart from their intermediate in size closure correlation.

It is apparent from Table 2 that large negative closure correlations occur between variables with large variance. Of particular interest are the relatively large positive closure correlations which will be considered in the following section.

Many geochemists have been using multivariate statistical methods which require that the matrix of correlation coefficients be used as a measure of similarity for an R mode analysis. As demonstrated previously (Butler, 1976) such methods can produce biased results unless the effects of closure are somehow taken into account. In Tables 3, 4, and 5 correlations are given for Apollo 11, 12 and 15 respectively. Those that are significant at the 99% confidence level when tested against their appropriate closure correlations are indicated by either a + or - sign. Correlations for basalts are given in the upper half and correlations for regoliths are given in the lower half of the matricies in Tables 3, 4, and 5. There is a great deal of information contained in Tables 3, 4 and 5 and more detailed comparisions are currently being prepared. However, it is apparent that there are major correlation coefficient pattern differences between basalt and regolith samples from the Apollo 11 and 12 sites. It is likely (as many previous investigators have noted) that the regolith at a given site is not derived by simple mechanical breakup of basalt(s) with an average composition of that given in Table 1.

As noted previously, the Apollo 15 regolith samples do not conform to the Chayes and Kruskal model. Suggested stratigies for the elimination of negative open variances (Chayes, 1971 and Butler, 1975) have involved using normative quantities of the alkalis ($KAlO_2$, orthoclase or albite). Such transformations, however, were designed for the elimination of negative open variances for Al_2O_3 and Na_2O in data sets in which SiO_2 was greater than 50%. Rational transformations for removing the negative open variance of SiO_2 in the Apollo 15 regolith suite and in the three postulated suites of Apollo 15 basalts (Butler, 1976) are currently being investigated.

The Effect of Closure on Correlations Between Trace Elements

Trace elements possess very small variances as compared with the variances of the major elements and one would not expect that the closure correlations between trace elements would be negative and of intermediate size. Positive correlation is possible if and only if the following (a rearrangement of Equation 4) holds:

$$o_t^2 \quad (o_i^2/\bar{y}_j + o_j^2/\bar{y}_i \qquad \text{Equation (5).}$$

As a complete chemical analysis is required to compute o_t^2, trace element data sets were compiled from the sets of major element analyses summarized in Table 1. Relatively few major plus trace element analyses have been published so an attempt was made to maximize the number of samples in each of the four selected arrays (Apollo 11, 12 and 15 Basalts and Apollo 15 Regolith) by reducing the number of trace elements included. These trace element analyses are summarized in Table 6. Again, the open variance for SiO_2 was negative in the Apollo 15 regolith suite so that only three suites conform to the Chayes and Kruskal model. It is apparent from Table 6 that all of the expected closure correlations are positive and there is, in a practical sense, little difference between using one of these p_{ij}'s or zero as the null value. From Equation (4) one can argue that the maximum positive value of p_{ij} can be approximated by:

$$p_{ij} = o_t^2/C_iC_j \qquad \text{Equation (6)}$$

There appears to be a positive correlation between observed variance and hypothetical open variance for those arrays that conform to the Chayes and Kruskal model and for most trace elements o_t^2 will be much greater than the individual open variances so that the expression for maximum p_{ij} (Equation 4) can be used to examine the conditions that must be satisfied to induce positive correlation.

Results for a simple analysis of the requirements for positive correlation (using the o_t^2 values for the Apollo 11 and 12 basalts) are given in Table 7. It is apparent from data presented in Table 7 that large positive closure correlations are to be expected only between variables with a small product of their coefficients of variation. As o_t^2 increases the product of the coefficients of variation required to produce a given maximum closure correlation increases. Thus, unless the trace elements have very small coefficients of variation one need not expect to find intermediate in size positive closure correlations and the Chayes and Kruskal (1966) admonition concerning positive closure effects should be modified accordingly. As evidenced by comparing the data in Tables 2 and 1, relatively large positive null correlations are encountered in the basalt and regolith suites but in all such situations they involve major elements with very small coefficients of variation. From Table 6 it is apparent that the majority of the trace elements included in the three data arrays have coefficients of variation that are quite large. Shaw (1961) argued that when the coefficient of variation exceeds 0.20 to 0.25 a lognormal distribution model was better than a normal distribution model for predictive and descriptive purposes

A great deal of the chemical information in the lunar data base consists of trace element analyses for which there are no accompanying major element analyses. In a strict sense such analyses do not conform to the Chayes and Kruskal model as o_t^2 can not be computed. In a practical sense, however, it seems appropriate to use the values of the coefficients of variation as a guide if one knows that all of the open variances for the major elements are positive for the suite from which the trace elements were drawn. If the product of the coefficients of variation is less than 0.01 (an admittedly arbitrary selection) a positive null correlation of intermediate level should be expected. If the product exceeds 0.01 and the individual correlations are less than 0.25 then it appears appropriate to use zero as a null value. When the individual coefficients of variation exceed 0.25 the required assumption of bivariate normality is probably not satisfied and alternative approaches are required to assess the degree of association between the two variables.

Alternative Forms of Correlation

In the preceeding sections the Pearson product moment correlation coefficient has been used as the measure of the linear association between a pair of variables. Tests of independence of this correlation coefficient require the assumption of bivariate normality which, as noted above, may not be met by geochemical variables. Demirmen (1976) recently reviewed the application of Spearman's and Kendall's rank correlation coefficients to situations in which the degree of association between a pair of variables is being examined. Both of these correlation coefficients are computed from ranks rather than the actual values of the variables and tests of independence are nonparametric. As described by Demirmen (1976) the Spearman correlation is obtained by treating the ranks as though they were the actual scores and is a measure of the monotonicity between the two sets of scores. High values of Spearman's correlation indicate that the association is monotonic (an increase in the rank of one variable is accompanied by an increase in the rank of the other or vice versa) although the relationship need not be linear. Kendall's correlation is a measure of the degree to which the rankings of the two variables agree.

Preliminary experiments have been undertaken to examine the effect of closure of uncorrelated parent variables on the values of these two alternative approaches to correlation. A set of 100 samples with 10 variables each was drawn from uncorrelated parent variables of specified mean and variance. The computed Kendall and Spearman rank correlations all lead to acceptance of the null hypothesis; that is, the true values of the correlations are zero and hence the ranks are independent. When this data set was closed by transforming each row into percentage form and the Kendall and Spearman correlations computed for the closed set the null hypotheses are rejected for the association between variables with large variance. Thus, the preliminary work confirms intuition in that closure seriously modifies the rank correlations as well as the Pearson product moment correlation coefficient. There is apparently no procedure available for taking the effects of closure on the rank correlations into account. This is obviously an area that should be investigated as the nonparametric aspects of hypothesis testing are certainly appealing.

Summary and Conclusions

Many, but certainly not all, negative correlations of intermediate size are solely the result of closure. Similarly, it is possible that positive correlations of intermediate size arise as the result of the closure process - especially those between variables with very small coefficients of variation. Unless one makes use of the Chayes and Kruskal approximation method for assessing the effects of closure, however, recognition of the extent of the effects of closure and rational hypothesis testing are impossible. The automatic selection of zero as a null value is definitely to be avoided. Unfortunately, the practice of an intuitive assessment of association of variation diagrams is widespread and almost always involves a picture of randomness for which the Pearson product moment correlation coefficient is zero. Thus, the intuitive assessment amounts to a formal hypothesis test for independence using zero as the null but lacking in any quantitative input.

From Table 2 and the accompanying discussion it is apparent that the effects of closure in the suites of Apollo 11, 12 and 15 basalt and regolith data suites is not negligible. Unfortunately, the set of Apollo 15 regolith samples does not conform to the Chayes and Kruskal model with the result that correlation coefficient based comparisons with other data sets is impossible. It appears that there is a very small probability that closure will induce intermediate in size positive correlations between variables present in trace amounts in these data sets. Analyses of the trace element portion of a sample technically do not conform to the Chayes and Kruskal model but it has been suggested that a careful examination of the values of the coefficients of variation may allow the degree of association to be evaluated. Continued studies of the effects of closure on the Spearman and Kendall rank correlation are planned as the nonparametric nature of the tests for independence is appealing for geochemical problems.

The average compositions presented in Tables 1 and 6 may prove useful to other investigators but a certain amount of prudence must be used as is always the case when dealing with averages.

Acknowledgements

The work reported on in this note was sponsored by a grant from NASA - NSG-7198.

REFERENCES CITED

Butler, J. C., 1975, The occurrence of negative open variances in ternary systems; Math. Geol., v. 7, no. 1, p. 31-45

Butler, J. C., 1976, Chemical classification of Apollo 15 mare basalts: Proc. Lunar Sci. Conf., 7th, in press.

Chayes, F., 1965, Variance-covariance relations in some published Harker diagrams of volcanic suites: Jour. Petr., v. 5, part 2, p. 219-237.

Chayes, F., 1971, Ratio Correlation: University of Chicago Press, Chicago, Ill.

Chayes, F. and Kruskal, W., 1966, An approximate statistical test for correlations between proportions: Jour. Geol., v. 74, no. 5, pt. 2, p. 692-702.

Demirmen, F., 1976, Rank: a Fortran IV program for computation of rank correlations: Computers and Geosciences, v. 1, p. 221-229.

Rose, H. J., Baedecker, P. A., Berman, S., Christian, R. P., Dwornik, E. J., Finkelman, R. B., and Schnepfem, M. M., 1975, Chemical composition of rocks and soils returned by the Apollo 15, 16 and 17 missions: Proc. Lunar Sci. Conf. 6th, v. 2, p. 1363-73.

Shaw, D. M., 1961, Element distribution laws in geochemistry: Geochimica et Cosmochimica Acta, v. 23, p. 116-134.

Taylor, S. R., 1975, Lunar Science: A Post-Apollo View: Pergamon Press, New York.

TABLE 1.

SUMMARY STATISTICS FOR SUITES OF APOLLO 11, 12 AND 15 BASALTS(B) AND REGOLITH(R)

Suite	#	SiO_2	Al_2O_3	TiO_2	FeO	MnO	MgO	CaO	Na_2O	K_2O	P_2O_5	S	Cr_2O_3	$\sum s^2$
Apollo 11 B	28													
$\bar{X}$		40.61	9.43	11.13	18.93	0.27	7.07	11.26	0.47	0.20	0.13	0.19	0.31	
C		0.026	0.145	0.104	0.040	0.382	0.115	0.095	0.107	0.590	0.384	0.222	0.249	
% s^2		16.25	27.65	20.14	8.49	0.16	9.92	16.99	0.04	0.20	0.04	0.09	0.03	
														6.719
Apollo 11 R	18													
$\bar{X}$		41.90	13.73	7.52	15.67	0.21	7.84	11.94	0.43	0.14	0.18	0.11	0.31	
C		0.019	0.034	0.069	0.019	0.083	0.041	0.036	0.079	0.112	0.521	0.225	0.367	
% s^2		41.68	14.55	17.63	5.82	0.02	6.71	11.95	0.08	0.02	0.64	0.04	0.87	
														1.518
Apollo 12 B	56													
$\bar{X}$		44.63	9.68	3.22	20.23	0.27	11.04	9.90	0.32	0.08	0.08	0.05	0.51	
C		0.054	0.203	0.217	0.078	0.112	0.349	0.145	0.483	1.31	0.771	0.751	0.314	
% s^2		19.75	13.07	1.65	8.39	0.00_2	49.94	6.97	0.08	0.04	0.01	0.00_4	0.09	
														29.720
Apollo 12 R	31													
$\bar{X}$		45.69	13.84	2.84	15.43	0.21	9.97	10.44	0.51	0.29	0.36	0.08	0.35	
C		0.035	0.092	0.135	0.089	0.106	0.112	0.045	0.259	0.346	0.307	0.391	0.200	
% s^2		32.61	21.30	1.90	24.42	0.00_7	16.28	2.89	0.23	0.14	0.15	0.13	0.06	
														7.670
Apollo 15 B	75													
$\bar{X}$		45.83	9.12	2.24	21.45	0.27	10.27	9.75	0.32	0.06	0.09	0.06	0.52	
C		0.047	0.198	0.205	0.128	0.130	0.221	0.093	0.380	1.51	1.07	0.302	0.263	
% s^2		21.69	15.01	0.96	34.74	0.00_6	23.56	3.79	0.[illegible]	0.04	0.04	0.00_2	0.08	
														21.79
Apollo 15 R	47													
$\bar{X}$		46.64	15.01	1.53	14.16	0.19	10.69	10.62	0.39	0.17	0.16	0.07	0.37	
C		0.013	0.158	0.189	0.188	0.173	0.049	0.087	0.161	0.275	0.279	0.256	0.267	
% s^2		2.39	39.39	0.58	49.63	0.00_7	1.89	5.97	0.03	0.01	0.01	0.00_2	0.07	
														14.33

TABLE 2.

COMPARISON OF OBSERVED AND CLOSURE CORRELATIONS*

Suite	Variables	Observed Correlations	Closure Correlations
Apollo 11 Basalts	SiO_2-Al_2O_3	-.0129	-.4583b
	SiO_2-TiO_2	-.3991a	-.3057
	SiO_2-FeO	-.3532	+.3529b
	SiO_2-Na_2O	-.3851	+.2231
	Al_2O_3-FeO	-.5418a	-.3431
	TiO_2-FeO	+.2828	-.2626b
	FeO-MgO	+.3014	-.1457b
Apollo 11 Regolith	SiO_2-TiO_2	-.4435a	-.3836
	SiO_2-FeO	-.4977a	-.3662
	SiO_2-MgO	-.5801a	-.3043
Apollo 12 Basalts	SiO_2-Al_2O_3	+.2453	-.1753b
	SiO_2-FeO	-.7147a	+.3790b
	SiO_2-MnO	-.2326	+.4404b
	SiO_2-MgO	-.6297a	-.7470
	MnO-MgO	+.1487	-.3589b
Apollo 12 Regolith	SiO_2-Al_2O_3	+.0535	-.4127b
	SiO_2-FeO	-.5494a	-.4593
	Al_2O_3-MgO	-.5375a	-.3005
Apollo 15 Basalts	SiO_2-TiO_2	-.2867	+.1340b
	SiO_2-FeO	-.6542a	-.5130
	SiO_2-MnO	-.3269	+.3390b
	SiO_2-CaO	+.5472a	+.3284
	FeO-MgO	+.2198	-.3212b
	FeO-Cr_2O_3	+.3018	-.1019b
	Al_2O_3-CaO	+.3223	-.1444b
Apollo 15 Regolith	Open variance for SiO_2 is negative		

* a indicates an observed correlation that is significant (99%) tested against 0
b indicates an observed correlation that is significant (99%) tested against the closure correlation

TABLE 3.

CORRELATION COEFFICIENT MATRIX FOR APOLLO 11 SAMPLES*

	SiO_2	Al_2O_3	TiO_2	FeO	MnO	MgO	CaO	Na_2O	K_2O	P_2O_5	Cr_2O_3	S
SiO_2	1.00	+		-							-	
Al_2O_3		1.00	-				+		-	-	-	-
TiO_2			1.00	+		+	-		+	+	+	+
FeO				1.0		+						
MnO			+		1.0							
MgO				+		1.0	-			+	+	+
CaO							1.0		-	-	-	-
Na_2O								1.00	+	+		
K_2O	+			-					1.0	+	+	+
P_2O_5				-						1.0	+	+
Cr_2O_3			+								1.0	
S												1.0

TABLE 4.

CORRELATION COEFFICIENT MATRIX FOR APOLLO 12 SAMPLES

	SiO_2	Al_2O_3	TiO_2	FeO	MnO	MgO	CaO	Na_2O	K_2O	P_2O_5	Cr_2O_3	S
SiO_2	1.00	+		-	-		+				-	-
Al_2O_3	+	1.00	+	-		-	+	+	+	+	-	
TiO_2			1.00			-	+				-	
FeO		-		1.0		+	-	-	-	-	+	
MnO	-		+	+	1.0	+						
MgO	-			+	+	1.0	-				+	
CaO		+		-	-		1.0	+			-	
Na_2O			-	-	-	-		1.0			-	
K_2O	+	+	-	-				+	1.0			
P_2O_5			-	-				+	+	1.0		
Cr_2O_3		-		+		+	+				1.0	
S												1.0

TABLE 5.

CORRELATION COEFFICIENT MATRIX FOR APOLLO 15 SAMPLES

	SiO_2	Al_2O_3	TiO_2	FeO	MnO	MgO	CaO	Na_2O	K_2O	P_2O_5	Cr_2O_3	S
SiO_2	1.00	+	-		-	-				+	-	
Al_2O_3		1.00		-	-		+	+	+	+		
TiO_2			1.00	+								
FeO				1.0	+	+		-	-	-	+	
MnO					1.0				-	-		
MgO						1.0	-				+	
CaO							1.0				-	
Na_2O								1.0	+	+		
K_2O									1.0	+		
P_2O_5										1.0		
Cr_2O_3											1.0	
S												1.0

* + indicates positive correlation significant at the 99% level against closure
- indicates negative correlation significant at the 99% level against closure

TABLE 6.

SUMMARY STATISTICS FOR TRACE ELEMENTS(PPM) IN APOLLO 11, 12, AND 15 BASALTS

Suite	#	Rb	Sr	Zr	Y	Ni	Co	V	Ba	La	s^2	σ_t^2	$\bar{p}_{ij}$	Max p_{ij}	Min p_{ij}
Apollo 11	12														
	$\bar{X}$	3.63	167.0	43.63	151.2						5.19	5.57	.0057	.0112	.0021
	C	.668	.166	.366	.294										
Apollo 12	24														
	$\bar{X}$	1.29	105.5	130.6	49.10	36.94	44.09	145.0	80.29	9.70	22.30	21.54	.0150	.0378	.0070
	C	.530	.298	.258	.425	.699	.346	.219	.426	.715					
Apollo 15	31														
	$\bar{X}$	2.12	100.9	89.80	38.70	57.98	54.34	167.2	64.02		8.68	10.34	.0054	.0152	.0013
	C	.789	.237	.411	1.01	.543	.319	.282	.669						

TABLE 7.

REQUIREMENTS FOR POSITIVE CLOSURE CORRELATIONS

$$p_{ij} \text{ (maximum)} = \hat{\sigma}_t^2/C_iC_j$$

$\sigma_t^2 = 21.54\text{x}10^{-4}$*

p_{ij}(max)	C_iC_j	$C_i = C_j$
.80	$2.69\text{x}10^{-3}$	.052
.60	$3.59\text{x}10^{-3}$	.066
.40	$5.39\text{x}10^{-3}$	.073
.20	$1.08\text{x}10^{-2}$	.104

$\sigma_t^2 = 5.57\text{x}10^{-4}$*

p_{ij}(max)	C_iC_j	$C_i = C_j$
.80	$6.96\text{x}10^{-4}$	.026
.60	$9.27\text{x}10^{-4}$	.035
.40	$1.39\text{x}10^{-3}$	.037
.20	$2.78\text{x}10^{-3}$	.053

* Open variances scaled to agree with using the mean proportion instead of the obser mean.

APPENDIX I.

PRELIMINARY R MODE ANALYSIS OF APOLLO 15 BASALTS TAKING THE EFFECTS OF CLOSURE AND RATIO FORMATION INTO ACCOUNT

Paper presented to the Conference on : Origins of Mare Basalts and their Implications for Lunar Evolution. November 17-19, 1975

PRELIMINARY R MODE ANALYSIS OF APOLLO 15 BASALTS TAKING THE EFFECTS OF CLOSURE AND RATIO FORMATION INTO ACCOUNT:

J.C. Butler, Department of Geology, University of Houston, Houston, Texas 77004

Faced with the interpretation of a collection of chemical analyses, one can look for relationships between the variables (R mode) or between the samples (Q mode). The choice of a suitable measure of similarity for either Q or R mode has been discussed by many authors and in this study, the Pearson product moment correlation coefficient is taken as the measure of similarity for an R mode analysis of selected Apollo 15 basalt chemical analyses as it provides a dimensionless index of the joint behavior of each pair of variables with equal weighting of all variables. Each pair of variables can be examined using a graphical (scatter diagram) or statistical approach. The goal of the examination is to determine which pairs of variables exhibit a significant relationship. To make this determination a statement of what constitutes an insignificant relationship is necessary. In constructing a model to account for the petrogenesis of a selected group of samples, one needs to be able to focus attention on those relations that significantly depart from randomness. If relationships that fail of significance are included or if significant relationships are omitted, the model will not accurately portray the available information. Intuition and introductory statistics suggest that zero is the appropriate measure of randomness and if an observed correlation differs significantly from zero, the relationship is judged to be significant. The purpose of this study was to examine the chemical analyses of a suite of Apollo 15 lunar basalts using three data forms in which zero may not be the appropriate null value: (1) closed data, (2) normalized data, (3) ratios with common terms.

Data sets in which the sum of the variables measured for each sample is constant are termed closed and chemical data, by their very nature, are closed (1). Deviations between 100.00% and the sum of the measured variables result from a combination of errors of commission and omission. It can be shown that there are no restrictions on the row sums of the variance-covariance matrix generated from open data (1). Open data being defined as data in which the constant item row sum restriction does not hold. For closed data, however, each row of the variance-covariance matrix sums to zero; that is, $\mathrm{Var}_A + \mathrm{COV}_{AB} + \mathrm{COV}_{AC} \ldots + \mathrm{COV}_{AM} = 0.00$ (1)

Thus, for closed data, the sum of the covariances between A and all other variables is equal to the negative of variance of A which is positive by definition. There must be at least one negative covariance in each row of the variance-covariance matrix for closed data; there is no such restriction on the variance-covariance matrix for open data. As the correlation coefficient is simply the covariance divided by the product of the standard deviations, a negative covariance manifests itself as a negative correlation coefficient. In a data set in which there was zero correlation between the variables in the open form, closure (by forming percentages for example) would induce some degree of correlation and, in general, the larger the variances of the larger the induced correlation. Thus, the observed correlation between variables in a closed data set contains a contribution due to closure as well as a measure of the joint behavior of the variables. Chayes (1) argued that the contribution due to closure should be used as the null value instead of zero. This strategy often leads to a conflict with an intuitive assessment of randomness as large negative induced correlations are quite common; especially between the variables that are used in the frequently invoked Harker diagrams. Several stratigies have been suggested which may appear to reduce if not eliminate the closure effect: (1) normalization to a constant composition and (2) formation of ratios.

Suess and Urey (2) appear to have introduced the strategy of normalizing chemical compositions to the composition of the average carbonaceous chondrite by dividing the amount of each variable by the amount of that variable in average carbonaceous chondrite (CI). Thus, comparision with average CI is facilitated. Examination of a normalized data matrix reveals that it is not closed in the sense that the row sums are not constant. However, because each variable in each sample is divided by the same value (the amount of that variable in CI), the correlation between normalized variables (such as, SiO_2 /SiO_2 CI versus CaO/CaOCI) is identical to that between the raw or non-normalized data (SiO_2 versus CaO).

Creation of ratios by dividing the value of one variable by the value of another variable in the same row does indeed remove the effects of closure as the row sums are no longer constant. Depending upon the manner in which the ratios are formed, however, another form of induced correlation may arise (summarized in 1). For example, if the ratios have a common denominator (CaO/SiO_2 versus Al_2O_3 /SiO_2) there will be a correlation induced between the ratios even though there may be zero correlation between the parents of the ratios. It is possible to compute and simulate the correlation between ratios with common terms as functions of the coefficients of variation of the parents. This expected correlation can be used as the null value for determining whether the observed correlation differs significantly from that expected for randomness.

Chemical variations of Apollo 15 lunar basalts have been the subjects of many previous investigations; including, the recognition of olivine and quartz nromative subgroups. The apollo 15 basalts appear to be separable (on basis of major element variations) from other lunar basalts (3) and this study is a precursor to a more detailed statistical analysis of major and minor element chemical variations of all of the lunar basalts. A total of 28 major element analyses of Apollo 15 basalts were chosen as a trial set of data (3 and 4). This set of analyses was closed by recalculating SiO_2, Al_2O_3, TiO_2, FeO, MgO, CaO, Na_2O, K_2O and $P_2O_5$to 100%. The summary statistics, variance-covariance matrix and the correlation coefficient matrix were computed and, to compare with previous work, the correlations (r_{ij}) were tested against zero (Table 1.). To use the Chayes and Kruskal test for significant departures from randomness one computes the means and variances of X - the hypothetical open array. This array is characterized by having zero covariances and hence, zero correlations. When X is closed to form Y - the hypothetical closed array - the means and variances of the variables in Y are equal to those in the array being tested. As there are zero correlations in X the correlations in Y (p_{ij}) arise solely as a result of closing the data; these are the null values for the Chayes and Kruskal test. The open variances of the Apollo 15 lunar basalt suite are all positive; a prerequisite for application of the test (1). Values of p_{ij} must be considered as approximate as the equation for their calculation is derived using a first order delta approximation. A computer program in Fortran IV was written that enables simulation of the X and Y arrays. The observed correlations (r_{ij}) were also tested against the simulated (q_{ij}) and approximated (p_{ij}) closure correlations and the results are given in Table 1. Of particular interest is that fact that correlations between SiO_2 - MgO and Al_2O_3 - MgO would be judged significant (99%) when tested against either p_{ij} or q_{ij}. In fact, the expected closure correlation between SiO_2- MgO would be significant (99%) if tested against zero. Yet, this is the correlation induced solely by the process of closing the data. Perhaps of greatest interest, however, are the correlations between SiO_2 - Al_2O_3 and FeO-MgO which fail of significance when tested against zero yet, as the result of relatively large negative closure correlations, are significant when tested against both p_{ij} and q_{ij}. Thus, a petrogenetic model for the Apollo 15 lunar basalts should include the correlations between SiO_2 - Al_2O_3, and FeO - MgO. That these correlations are significant can only be detected if the effects of closure are considered. Similarly, such a model need not include the correlation between SiO_2 - MgO and Al_2O_3- MgO.

The 28 Apollo 15 basalts were normalized by dividing each variable by the amount of that variable in average CI (recalculated to 100% on a volatile-free basis; Mason). These basalts contain greater amounts of SiO_2 , Al_2O_3 TiO_2, MnO, and CaO than the average CI and lesser amounts of FeO, MgO, Na_2O, K_2O, and P_2O_5. Correlations that are significant against their appropriate null values are those given in Table 1.

Characterization and comparison of lunar sample chemistry by using ratios has been quite common. To illustrate the problems inherent in the interpretation of ratio correlations, the ratios FeO/Al_2O_3, MgO/Al_2O_3 and CaO/Al_2O_3 were computed. If the parents of ratios with a common denominator were uncorrelated and homogeneous in coefficient of variation a correlation of + 0.5 would be induced between ratios. This is the correlation referred to as spurious by Pearson (5). In general, as the coefficient of variation of the denominator increases relative to those of the numerators of the ratios, the induced correlation increases. Ratios in which the numerators are in common or in which the numerator of one ratio is the denominator of the other ratio are also subject to a spurious correlation; that is, a correlation that results solely from the formation of the ratios. Observed (r_{ij}), approximated (p_{ij}), and simulated (q_{ij}) correlations between the three ratios with Al_2O_3 as a common denominator are given in Table 2. The parents of the ratios (see Table 1) are far from being uncorrelated yet large samples of ratios formed from uncorrelated variables with means and variances equal of those of the observed Al_2O_3. FeO, CaO and MgO might exhibit correlations as strong as those exhibited by the ratios FeO/Al_2O_3- CaO/Al_2O_3 and CaO/Al_2O_3 - MgO/Al_2O_3. There is a definite need for more experimentation with ratio formation before it will be possible to unambigously assess the significance of ratios formed with common parts. Testing ratios with common parts requires that the parent variables be measured individually so that their means and standard deviations are known. Situations in which the ratio itself is measured instead of the individual variables are not testable. Thus, if one measures the ratios of Al/Si and Ca/Si, it will not be possible to isolate the component of the observed correlation that is the result of working with ratios with a common denominator (the null portion) from the true correlation between the ratios.

Use of multivariate statistical models such as cluster analysis, principal components analysis and various factor analysis models are being used with increasing frequency in the analysis of chemical data. All of these models require a measure of similarity as a starting point. If the similarity matrix is the correlation coefficient matrix or the variance-coveriance matrix then some method of taking the effects of closure normalization or ratio formation must be derived. If these effects are ignored, the results will be difficult to interpret as the investigator will not be able to separate out the bias that definitely is introduced when one deals with ratios or percentage data forms.

REFERENCES

1. Chayes, F.C., (1971) Ratio Correlation, University of Chicago Press.
2. Suess, H.E. and Urey, H.C., (1956) Rev. Mod. Phys. v 28, p. 53-74.
3. Thodes, J.M. and Hubbard, N.J., (1973) Proc. Lunar Sci. Conf. 4th, 1127
4. Cuttitta, F., Rose, J.H., Annell, C.S., Carron, C.P., Christian, D.T., Ligon, D.T., Dworni, E.J., Wright, T.L., and Greenland, L.P., (1973) Proc. Lunar Sci. Conf. 4th, 1081
5. Pearson, K., (1896) Proc. Roy. Soc. (London) 489.

R MODE - APOLLO 15 BASALTS

BUTLER, J.C.

TABLE 1. Observed (r_{ij}), Approximated (p_{ij}) and Simulated (q_{ij}) Correlations for the Apollo 15 Lunar Basalt Data

Variables	r_{ij}[1]	p_{ij}[2]	q_{ij}[3]
SiO_2-Al_2O_3	.1351	$\underline{-.2721}$	$\underline{-.2951}$
SiO_2-TiO_2	$\underline{-.7156}$	$\underline{-.0839}$	$\underline{-.0733}$
SiO_2-FeO	$\underline{-.7699}$	$\underline{-.4409}$	$\underline{-.4432}$
SiO_2-MgO	$\underline{-.5662}$	$\underline{-.4712}$	$\underline{-.4710}$
SiO_2-CaO	$\underline{.5281}$	$\underline{-.1150}$	$\underline{-.1217}$
Al_2O_3-FeO	$\underline{-.6104}$	$\underline{-.1634}$	$\underline{-.1272}$
Al_2O_3-MgO	$\underline{-.4517}$	$\underline{-.1634}$	$\underline{-.1836}$
Al_2O_3-CaO	$\underline{.6586}$	$\underline{-.0552}$	$\underline{-.0350}$
TiO_2-FeO	$\underline{.8841}$	$\underline{-.0539}$	$\underline{-.0688}$
TiO_2-MnO	$\underline{.4768}$	$\underline{.0498}$	$\underline{.0160}$
FeO-MnO	$\underline{.4935}$	$\underline{.0049}$	$\underline{.0066}$
FeO-MgO	.3608	$\underline{-.2727}$	$\underline{-.2670}$
FeO-CaO	$\underline{-.5336}$	$\underline{-.0822}$	$\underline{-.0942}$
MnO-Na_2O	$\underline{-.4242}$	$\underline{.1014}$	$\underline{.1202}$
MgO-CaO	$\underline{-.9219}$	$\underline{-.1980}$	$\underline{-.1954}$
Na_2O-K_2O	$\underline{.6833}$	$\underline{.0260}$	$\underline{.0383}$
K_2O-P_2O_5	$\underline{.5512}$	$\underline{.0099}$	$\underline{.0624}$

[1] an underlined r_{ij} indicates significance against zero at 99%

[2] an underlined p_{ij} indicates significance against p_{ij} at 99%.

[3] an underlined q_{ij} indicates significance against q_{ij} at 99%

TABLE 2. Observed (r_{ij}), Approximated (p_{ij}) and Simulated (q_{ij}) Correlations for Selected Ratios of the Apollo 15 Lunar Basalt Data

Variables	r_{ij}[1]	p_{ij}[2]	q_{ij}[3]
FeO/Al_2O_3 : CaO/Al_2O_3	$\underline{.8488}$	.7443	.7487
FeO/Al_2O_3 : MgO/Al_2O_3	$\underline{.8260}$	$\underline{.5001}$	$\underline{.5164}$
CaO/Al_2O_3 : MgO/Al_2O_3	$\underline{.5296}$	$\underline{.5002}$	$\underline{.5163}$

[1] an underlined r_{ij} indicates significance against zero at 99%

[2] an underlined p_{ij} indicates significance against p_{ij} at 99%

[3] an underlined q_{ij} indicates significance against q_{ij} at 99%

APPENDIX II.

VARIATIONS IN CHEMICAL COMPOSITIONS OF APOLLO 15 MARE BASALTS

Abstract of Paper Submitted to the Seventh Lunar Science Conference
March 15-19, 1976, Lunar Science Institute, Houston, Texas

VARIATIONS IN CHEMICAL COMPOSITIONS OF APOLLO 15 MARE BASALTS;

J.C. Butler, Department of Geology, University of Houston, Houston, Texas, 77004.

Sufficient major element analyses of lunar mare basalts have been accomplished so that within and between mission comparisons can be made. Hodges and Papike (1) have presented evidence that mare basalt texture is related to major element chemistry and that textural variations within each group of mare basalts is a function of cooling rate. Rhodes et al. (2) effectively argue that careful investigation of major element variations may allow recognition of samples that retain information about their source area.

Many investigators of lunar and terrestrial basalts resort to binary or ternary scatter diagrams (a form of R mode analysis) as the means of identifying groups of related samples and recognizing differentiation trends within each group. TiO_2, for example, has been found to effectively discriminate against samples classified as to mission and plots of TiO_2 versus SiO_2 or MgO have been used to classify mare basalts (2).

It has been demonstrated, however, that failure to take the effects of closure into account may lead to accepting correlations as significant when in fact they are probably due entirely to closure (3). In a set of 49 chemical analyses of Apollo 15 mare basalts (Table 1) SiO_2 and MgO account for some 58% of the total variance of the data set but their observed correlation of -0.50 can be accounted for by closure effects (3). In addition reliance on scatter diagrams may result in a classification scheme in which the variables selected account for only a small amount of the total information content of the matrix. In the set of 49 Apollo 15 mare basalts (Table 1) used in this study, TiO_2 and MgO (used by Rhodes et al. (2) to group lunar basalts) account for only some 30% of the total variance. Sole reliance on R mode techniques for classification purposes may fail to detect relations between the samples. In fact, many methods used to relate samples on the basis of their chemistry are in fact discrimination techniques rather than true classification techniques.

There are a variety of techniques (Q mode), however, that are designed to classify samples as a function of values of the measured variables. Q mode cluster analysis and non linear mapping techniques were used to classify members of the set of 49 analyses of Apollo 15 mare basalts. Interpretation of the results of such analyses is fraught with difficulty as a result of distortions in the summary plots and pre-existing bias of the interpreter. However, the results of application of these classification techniques are interesting. As would be predicted the quartz normative Apollo 15 (pigeonite) basalts and the olivine normative Apollo 15 basalts are easily separated. In addition to these two groups it appears that there are two groups of Apollo 15 olivine normative basalts. Group 1 olivine normative basalts are characterized by low SiO_2 (44.86%), intermediate TiO_2 (2.21%), high FeO (22.53%), high MgO (11.54%) and an MgO/CaO ratio greater than 1.0. Group 2 olivine normative basalts are characterized by intermediate SiO_2 (45.43%), high TiO_2 (2.54%), high FeO (22.53%), low MgO (9.62%) and an MgO/CaO ratio of less than 1.0. Values given in parentheses are group averages.

VARIATIONS IN CHEMICAL COMPOSITIONS

J.C. Butler

Normative compositions also emphasize the differences between groups 1 and 2. Using norms published by Rhodes and Hubbard (4) and Chappell and Green (5) the group 2 olivine normative basalts have an average of 10.2% normative olivine whereas the group 1 basalts have an average of 18% normative olivine (10 and 11 samples respectively). The largest normative olivine content for a group 2 basalt is 14.7% and the smallest normative olivine content for a group 2 basalt is 16.6%; thus, the two postulated groups show no overlap in normative olivine content. Trace element data have not been used in this study but published values show some differences between groups 1 and 2. For example, the average Y content for type 1 is 28 ppm as compared with an average of 36 ppm (based on only 5 and 4 analyses respectively (5)). The same three-fold subdivsion (quartz normative and two olivine normative) can be recognized using non linear mapping and the recognized groups are identified by sample number in Table 1.

Discriminant function analysis can be used to assess the assignment efficiency of the postualted groupings (although this involves admittedly somewhat circular reasoning). SiO_2 and TiO_2 reclaim the postulated groupings with an efficiency of 94%. Nava (6) stated that the olivine-quartz normative dichotomy disappeared unless one used the normative mineralogy or the SiO_2 content as variables. Q mode analysis was repeated after eliminating SiO_2 as a variable with the result that the same three-fold subdivision was evident with MgO and FeO yielding an assignment efficiency of some 90%. It is interesting that the observed correlation of 0.24 between MgO and FeO is not significant (99%) if tested against a null of zero but is significant (99%) if tested against its expected closure correlation of -0.32 (3). Previous investigators may have overlooked the utility of the FeO-MgO plot to portray differences between samples on the basis that the observed correlation appeared to be insignificant.

Previous investigators have noted that the olivine and quartz normative basalts can not be genetically related by a near surface crystal fractionation mechanism involving olivine, pyroxene or spinel (4,5). Variations within quartz normative Apollo 15 mare basalts, however, can be related to near surface olivine fractionation and cooling rate variation (1). Rhodes and Hubbard argued (4) that chemical variation within the olivine basalts was the result of olivine fractionation whereas Chappell and Green (5) believed that trace and minor element variations were indicative of derivation from more than one cooling unit. Along this line of reasoning, Papike and Hodges (7) reported that there are slower and faster cooling units or groups of Apollo 15 quartz and olivine normative basalts. It is possible that the two sub-groups of the olivine normative basalts recognized in this study (Table 1) could reflect cooling rate variations although an in-depth interpretation of the significance of the postulated groupings is uncertain at the present time. In part, this reflects a lack of field control for the lunar basalts in the sense that there remains some uncertainity as to how many cooling units were sampled at the Apollo 15 site (4).

There is no doubt that two groups of olivine normative basalts can be recognized from the Apollo 15 site using variations in major element chemistry as the basis for discrimination. All splits of the same parent sample are placed in the same sub-group and the classification reflects the views of previous investigators as to the relations between some of the smaller samples. For example, Chappell and Green (5) considered that samples 15622 and 15636 came from the same parent and that 15658, 15668 and 15674 were splits from the same parent. The former are assigned to group 1 and the latter to group 2 (Table 1). The postulated two-fold subdivision of the olivine normative basalts may reflect: (1) a hiatus due to inadequate sampling; (2) the presence of two cooling units of olivine normative basalt at the Apollo 15 site; or (3) sampling of different parts of a single olivine normative basalt cooling unit.

TABLE 1

CLASSIFICATION OF THE APOLLO 15 MARE BASALTS

Group 1 - Olivine Normative - low SiO_2

15016(3)
15379(1)
15555(6)
15622(1)
15636(1)
15643(1)
15659(1)
15672(1)

Group 2 - Olivine Normative - intermediate SiO_2

15119(1)
15256(2)
15545(2)
15556(1)
15557(2)
15607(1)
15658(1)
15668(2)
15674(1)
15676(1)

Group 3 - Quartz Normative - high SiO_2

15058(1)
15065(4)
15076(3)
15117(1)
15475(2)
15486(1)
15495(1)
15499(2)
15595(1)
15597(2)
15118(1)

REFERENCES

(1) Hodges, F.N. and Papike, J.J., 1975, Origins of Mare Basalts, L.S.I., 72.
(2) Rhodes, J.M., Hodges, F.N., and Papike, J.J., 1975, Origins of Mare Basalts L.S.I., 135-39.
(3) Butler, J.C., 1975, Origins of Mare Basalts, L.S.I., 15-19.
(4) Rhodes, J.M. and Hubbard, N.J., 1973, Proc. Lunar Sci. Conf. 4th, 1127-48.
(5) Chappell, B.W. and Green, D.H., 1973, Earth Planet. Sci. Lett.,19,237-46.
(6) Nava, D.F., 1974, Proc. Lunar Sci. Conf. 5th, 1087-96.
(7) Papike, J.J. and Hodges, F.N., 1975, Origins of Mare Basalts, L.S.I.,120.

APPENDIX III.

VARIATIONS IN CHEMICAL COMPOSITION OF APOLLO 15 MARE BASALTS

Paper Accepted for Publication in the Proceedings of the Seventh Lunar Science Conference. (reprint) mapping analysis in particular) to these data in an attempt to discover natural groupings based on all of the major oxide data rather than relying on one or more of the many binary scatter diagrams advocated by previous investigators. It appears that three chemically distinct groups of Apollo 15 mare basalts exist: (1) olivine normative basalts with low SiO_2, intermediate TiO_2, and an MgO/CaO ratio greater than 1.0; (2) olivine normative basalts with intermediate SiO_2, high TiO_2, high MgO, and an MgO/CaO ratio less than 1.00; (3) quartz normative basalts with high SiO_2, low TiO_2 and FeO. Stepwise discriminant function analysis reclaims these three sub-groups with an assignment efficiency of greater than 95%.

APOLLO 15 MARE BASALTS

Sample making up the set of Apollo 15 mare basalts are identified as to the source reference, chemical, and textural properties in Table 1. Sample numbers given in Table 1 will be used to refer to the specific samples in the following illustrations. There are a considerable number of published partial analysis of the Apollo 15 mare basalts but only samples for which SiO_2, Al_2O_3, FeO, MnO, MgO, CaO, Na_2O, K_2O, TiO_2, and P_2O_5 were analyzed were included. The set of 49 analyses identified in Table 1 was selected to provide a training set of data; seven chemical analyses of Apollo 15 mare basalts (LSPET, 1972) were intentionally reserved as a set of unknowns in the event that a classification of the 49 training analyses proved feasible.

Differences between the sum of the major oxides and 100% are the results of errors of commission and/or omission. Each analysis was recalculated to 100%. Closure to 100% changes slightly (0.01) the correlation coefficients but causes the sum of each row of the variance-covariance matrix to be zero. Summary statistics for the set of 49 analyses are given in Table 2. SiO_2, CaO, and FeO have very small coefficients of variation (.035, .058, .066, respectively) and hence are the least variable of the ten oxides. Of the oxides present in amounts greater than 1.0%, MgO and TiO_2 are the most variable (coefficients of variation of 0.18 and 0.17, respectively).

The analysis of correlations between percentages of the same whole is difficult. There is no guarantee that zero is a suitable null value or that the same null value can be used for all possible pairs of correlations. It can be readily demonstrated (Chayes, 1971) that closure of uncorrelated parent variables will induce correlation as the sole result of percentage formation. Thus, the correlation between percentages contains a component that is a measure of the linear relationship between the variables. The problem is to ascertain the magnitude of the effect due to closure.

Chayes and Kruskal (1966) developed a model that allows prediction of the closure effect. A hypothetical open array X (with zero covariances and correlations) is closed to form a hypothetical closed array Y, the variables of which have the same means and variances as those in the observed array U. Correlations in Y arise solely as the result of closure and thus may be used as null values.

Table 1. Make-up of the Apollo 15 mare basalt analyses.

Sample reference number[1]	Sample number	Chemical type	Petrographic type[2]	Source reference
1	15556,5	olivine normative	highly vesicular basalt	Rhodes and Hubbard, 1973
2	15256,15	olivine normative	recrystallized basalt breccia	Rhodes and Hubbard, 1973
3	15256,22	olivine normative	recrystallized basalt breccia	Rhodes and Hubbard, 1973
4	15668,2	olivine normative	highly vesicular basalt	Rhodes and Hubbard, 1973
5	15545,13	olivine normative	porphyritic olivine basalt	Rhodes and Hubbard, 1973
6	15016,4	olivine normative	highly vesicular basalt	Rhodes and Hubbard, 1973
7	15555,8	olivine normative	porphyritic olivine basalt	Rhodes and Hubbard, 1973
8	15475,35	quartz normative	porphyritic clinopyroxene basalt	Rhodes and Hubbard, 1973
9	15499,2	quartz normative	porphyritic clinopyroxene basalt	Rhodes and Hubbard, 1973
10	15058,5	quartz normative	porphyritic clinopyroxene basalt	Rhodes and Hubbard, 1973
11	15076,2	quartz normative	porphyritic clinopyroxene basalt	Rhodes and Hubbard, 1973
12	15076,21	quartz normative	porphyritic clinopyroxene basalt	Rhodes and Hubbard, 1973
13	15118,4	quartz normative	porphyritic clinopyroxene basalt	Rhodes and Hubbard, 1973
14	15076,24	quartz normative	porphyritic clinopyroxene basalt	Cuttitta *et al.*, 1973
15	15065,31	quartz normative	porphyritic clinopyroxene basalt	Cuttitta *et al.*, 1973
16	15486,10A	quartz normative	?	Cuttitta *et al.*, 1973
17	15495,23	quartz normative	porphyritic clinopyroxene basalt	Cuttitta *et al.*, 1973
18	15065,8	quartz normative	porphyritic clinopyroxene basalt	Cuttitta *et al.*, 1973
19	15117,8	quartz normative	porphyritic clinopyroxene basalt	Cuttitta *et al.*, 1973
20	15557,28	olivine normative	porphyritic olivine basalt	Cuttitta *et al.*, 1973
21	15607,3	olivine normative	porphyritic olivine basalt	Cuttitta *et al.*, 1973
22	15659,4	olivine normative	porphyritic olivine basalt	Cuttitta *et al.*, 1973
23	15555,27	olivine normative	porphyritic olivine basalt	Cuttitta *et al.*, 1973
24	15643,5	olivine normative	porphyritic olivine basalt	Cuttitta *et al.*, 1973
25	15672,4	olivine normative	highly vesicular basalt	Cuttitta *et al.*, 1973
26	15379,2	olivine normative	porphyritic olivine basalt	Cuttitta *et al.*, 1973
27	15016,37	olivine normative	highly vesicular basalt	Cuttitta *et al.*, 1973
28	15676,6	olivine normative	porphyritic olivine basalt	Cuttitta *et al.*, 1973
29	15475	quartz normative	porphyritic clinopyroxene basalt	Chappell and Green, 1973
30	15499	quartz normative	porph. clino. basalt vitrophyre	Chappell and Green, 1973
31	15595	quartz normative	porph. clino. basalt vitrophyre	Chappell and Green, 1973
32	15597	quartz normative	porph. clino. basalt vitrophyre	Chappell and Green, 1973
33	15016	olivine normative	highly vesicular basalt	Chappell and Green, 1973
34	15119	olivine normative	porphyritic olivine basalt	Chappell and Green, 1973
35	15545	olivine normative	porphyritic olivine basalt	Chappell and Green, 1973
36	15555	olivine normative	porphyritic olivine basalt	Chappell and Green, 1973
37	15555	olivine normative	porphyritic olivine basalt	Chappell and Green, 1973
38	15622	olivine normative	porphyritic olivine basalt	Chappell and Green, 1973
39	15636	olivine normative	porphyritic olivine basalt	Chappell and Green, 1973
40	15658	olivine normative	porphyritic olivine basalt	Chappell and Green, 1973
41	15668	olivine normative	highly vesicular basalt	Chappell and Green, 1973
42	15674	olivine normative	porphyritic olivine basalt	Chappell and Green, 1973
43	15545,35	olivine normative	porphyritic olivine basalt	Maxwell *et al.*, 1972
44	15555,153	olivine normative	porphyritic olivine basalt	Maxwell *et al.*, 1972
45	15557,40	olivine normative	porphyritic olivine basalt	Maxwell *et al.*, 1972
46	15065,6	quartz normative	porphyritic clinopyroxene basalt	Nava, 1974
47	15065,42	quartz normative	porphyritic clinopyroxene basalt	Nava, 1974
48	15555,18	olivine normative	porphyritic olivine basalt	Nava, 1974
49	15597,19	quartz normative	porph. clino. basalt vitrophyre	Nava, 1974

[1] Used in the illustrations in the paper to refer to specific samples.

[2] Taken primarily from LSPET (1972) and Rhodes and Hubbard (1973).

Table 2. Summary statistics for the 49 Apollo 15 mare basalts.

	SiO_2	Al_2O_3	TiO_2	FeO	MnO	MgO	CaO	Na_2O	K_2O	P_2O_5
Mean	46.41	9.03	2.19	21.77	0.29	9.81	10.05	0.30	0.05	0.08
Variance	2.674	1.625	0.133	2.044	0.000_5	3.126	0.337	0.005	0.000_3	0.000_6
Standard deviation	1.635	1.275	0.364	1.430	0.023	1.768	0.580	0.069	0.016	0.024
Coefficient of variation	0.035	0.141	0.166	0.066	0.078	0.181	0.058	0.228	0.323	0.295
Maximum	49.27	10.80	2.86	24.75	0.34	12.31	11.74	0.66	0.11	0.15
Minimum	44.12	5.36	1.45	18.65	0.22	6.54	8.91	0.21	0.03	0.03
Range	2.83	5.44	1.41	6.10	0.12	5.77	2.83	0.45	0.08	0.12
% Variance	26.88%	16.35%	1.33%	20.54%	0.005%	31.43%	3.39%	0.05%	0.003%	0.005%

In order to obtain the expected closure correlations the means and variances of the X array must be assigned as described by Chayes (1971). Expected closure correlations can be approximated (p_{ij}, Chayes, 1971, Eq. 4.10) or simulated (q_{ij}) using Monte Carlo techniques. If negative open variances exist in the X array individual correlations can not be tested. This amounts to a crude rejection of the null hypothesis for the entire data set (Chayes, 1971) in that one or more of the correlations significantly departs from randomness but it is not possible to specify which one(s).

Observed correlations (r_{ij}) were tested against zero, the approximated closure correlations (p_{ij}) and the simulated closure correlations (q_{ij}). Those that were found to be significant when tested against one or more of these null values are given in Table 3 along with their appropriate null values.

That a number of correlations are significant when tested against all three null values is not particularly surprising. Of importance are correlations such as those between SiO_2–Al_2O_3, FeO–MgO, and FeO–MnO (Table 3) that have relatively small positive values and are not significant when tested against zero. Yet, these correlations are significant when tested against their appropriate null values which are of intermediate size and negative. All three of these correlations would most likely be judged to be insignificant applying an intuitive appraisal to the scatter diagrams and would not be used in constructing a petrogenetic model. The FeO–MgO correlation, however, is extremely important in considering apparent sub-groups of the Apollo 15 mare basalts.

Also of interest are the correlations between SiO_2–MgO and Al_2O_3–TiO_2 which are of intermediate size, negative, and significant when tested against zero. However, all fail of significance when tested against their expected closure correlations which are fairly large and negative. Such correlations should not be used in the development of a petrogenetic model except, perhaps, to consider why they were not significant.

Rose *et al.* (1974) made use of what they referred to as correlation family diagrams in which elements that were positively correlated were linked together. Signs of the significant correlations (against zero and against the expected closure correlations) are given in Table 3, and a schematic representation of the correlation families for those significant against the expected closure correlations is given in Fig. 1a. None of the oxides are in common with the two groups and, in general, there is a significant negative correlation between oxides in the two groups. What could be called the mafic group is made up of two sub-groups. Sub-group 1 links TiO_2–FeO–MnO and could be considered as an opaque or oxide group. Sub-group 2 links FeO–MnO–MgO and could be considered as an olivine/pyroxene association. In the felsic group, SiO_2–Al_2O_3–CaO represents an anorthite association whereas P_2O_5–Na_2O–K_2O suggests a KREEP component. This relationship (Fig. 1a) suggests that it would be possible to reduce the number of measured variables from the ten oxides to four (opaque, olivine/pyroxene, anorthite, and KREEP) and still retain a great deal of the information content of the data matrix. This would allow characterization of each sample in terms of the four new variables it contains. Procedures designed to accomplish this (such as

Table 3. Observed (r_{ij}), approximated (p_{ij}) and simulated (q_{ij}) correlations.

Variable 1	Variable 2	r_{ij}^1	p_{ij}^2	q_{ij}^2
SiO_2	Al_2O_3	0.265	*−0.306*	*−0.324*
SiO_2	TiO_2	*−0.653*	*−0.010*	*−0.002*
SiO_2	FeO	*−0.773*	*−0.291*	*−0.298*
SiO_2	MnO	*−0.365*	*0.193*	*0.184*
SiO_2	MgO	*−0.530*	−0.516	−0.510
SiO_2	CaO	*0.492*	*0.133*	*0.132*
SiO_2	Na_2O	*0.314*	0.143	0.142
SiO_2	K_2O	*0.408*	*0.042*	*0.042*
Al_2O_3	TiO_2	*−0.319*	−0.065	−0.039
Al_2O_3	FeO	*−0.464*	−0.231	−0.200
Al_2O_3	MnO	*−0.582*	*−0.085*	*−0.073*
Al_2O_3	MgO	*−0.669*	*−0.187*	*−0.206*
Al_2O_3	CaO	*0.370*	*−0.147*	*−0.132*
Al_2O_3	Na_2O	*0.630*	*−0.033*	*−0.022*
TiO_2	FeO	*0.869*	*−0.062*	*−0.079*
TiO_2	MnO	*0.501*	*0.041*	*0.005*
FeO	MnO	*0.495*	*−0.017*	*−0.016*
FeO	MgO	0.240	*−0.329*	*−0.323*
FeO	CaO	*−0.540*	*−0.091*	*−0.109*
MnO	MgO	0.251	*−0.191*	*−0.190*
MnO	Na_2O	*−0.462*	*0.055*	*0.067*
MnO	K_2O	*−0.299*	*0.045*	*0.032*
MnO	P_2O_5	*−0.297*	*0.046*	*0.066*
MgO	CaO	*−0.576*	*−0.283*	*−0.282*
MgO	Na_2O	*−0.621*	*−0.069*	*−0.080*
MgO	K_2O	*−0.508*	*−0.047*	*−0.035*
MgO	P_2O_5	*−0.328*	−0.052	−0.073
Na_2O	K_2O	*0.641*	*0.010*	*0.001*
K_2O	P_2O_5	*0.649*	*0.018*	*0.043*
Na_2O	P_2O_5	*0.346*	*0.001*	*0.003*

[1] if r_{ij} is italic it is significant at the 99% level against a null of zero.

[2] if p_{ij} or q_{ij} is italic the observed correlation (r_{ij}) is significant against the appropriate null value at the 99% level.

R-mode principal components, cluster analysis, and factor analysis) are complicated by the closure effect. If, for example, one used zero as the null value against which to test the observed correlations, the correlation families would appear as in Fig. 1b, in which there is only the mafic opaque link. The olivine/pyroxene link disappears unless one uses the more appropriate correlation as the null value. The anorthite link is broken in Fig. 1b and, in addition, SiO_2 links with both Na_2O and K_2O. If one were to use the standard computer routines for *R*-mode analysis in which the matrix of correlation coefficients is used as the measure of similarity, the new variables would be similar to those in Fig. 1b and the effect of closure would have been ignored. It has been suggested (Butler, 1976) that the matrix of

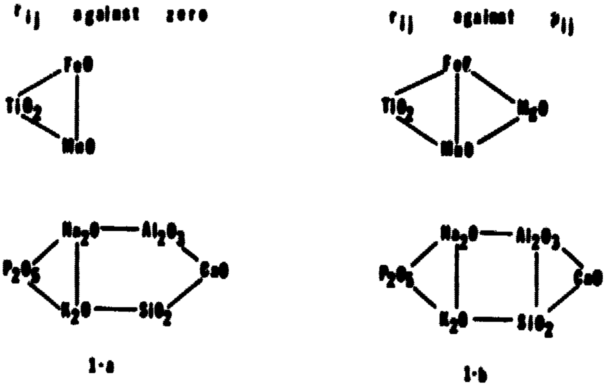

Fig. 1. Correlation family diagrams for the set of Apollo 15 mare basalts.

expected closure correlations could be subtracted from the matrix of observed correlations and the difference matrix used as the measure of similarity. Unless this or a similar strategy is adopted, the effects of closure will continue to influence the interpretation of the chemical data matrices. The danger is obvious. Without considering closure effects, one may both receive misinformation and lose valuable information. Preliminary results suggest that the proper *R*-mode will be useful in characterizing the Apollo 15 mare basalts and comparing them with mare basalts from other missions.

CLASSIFICATION OF THE APOLLO 15 MARE BASALTS

It is obvious that there are at least two chemically recognizable groups of Apollo 15 mare basalts—the olivine and quartz normative sub-groups recognized by many previous investigators. The SiO_2 content, as might be expected, provides an efficient discriminant. In fact, Nava (1974) stated that unless SiO_2 or the normative amount of quartz were used that it would not be possible to recognize any groupings within the Apollo 15 mare basalts unless one considered coarse-grained basalts as a third group. Nava's statement is backed up with reference to an MgO–TiO_2 scatter diagram which does indeed reveal the difficulty in recognizing discrete groups. However, from Table 2, the pair of variables TiO_2 and MgO account for only 32.8% of the total variance of the 49 mare basalts—using Chayes's (1964) suggestion that the sum of the variances provides a measure of the information content of the data matrix. It would appear advisable to base a classification scheme on the total data matrix if one is interested in ascertaining whether natural groupings exist within the data array.

Scanning the raw data suggested that further subdivisions of the Apollo 15 mare basalts might be possible. For example, more than 90% of the olivine normative basalts contain either greater than 10% MgO and less CaO or greater than 10% CaO and less MgO.

There are several models that are designed to search out natural groupings of samples based on variations of the measured variables. A number of these models were tried with the hypothesis that a successful model would minimally reveal the olivine-quartz normative dichotomy.

Q-mode cluster analysis is perhaps the simplest (conceptually) that could enable recognition of sample groups. However, results are difficult to interpret because of distortions that result in similar samples being placed in different groups or clusters. Results are depicted using a dendrogram which is a tree-like plot in which pairs of similar samples are linked together and treated as single samples. Grouping continues until all samples have been linked. A dendrogram for the 49 Apollo 15 mare basalts is given in Fig. 2 using the simple distance function (Davis, 1971) as the measure of between sample similarity. Any pre-existing bias of the investigator will certainly influence the interpretation of such a dendrogram. It is conceivable that a splitter would recognize the presence of 49 separate samples whereas a lumper would argue for the presence of one group composed of 49 entries. A similarity of 1.20 was selected as the cut-off for group recognition because at this level the olivine-quartz normative dichotomy was preserved. Samples that linked below 1.20 were considered to form a discrete group of sub-set of the entire sampled population. When the cosine theta measure of similarity was used or when scaled or transformed (0.0–1.0) data were used with the simple distance function the olivine-quartz normative dichotomy was not retained satisfactorily.

Two major subsets of the sampled population are recognized in Fig. 2. Group A contains the olivine normative basalts and group B the quartz normative basalts. Within the olivine normative cluster (Group A) two sub-groups can be recognized: (1) A-1 (henceforth referred to as group (1)) and (2) A-2 (henceforth referred to as group (2)). The quartz normative basalts will be referred to as group (3).

As is frequently the case in *Q*-mode cluster analysis there are samples that stand out as being sufficiently different from the major clusters. Sample number 23 (15555,57) clusters with the combined groups 1 and 2 which suggests that,

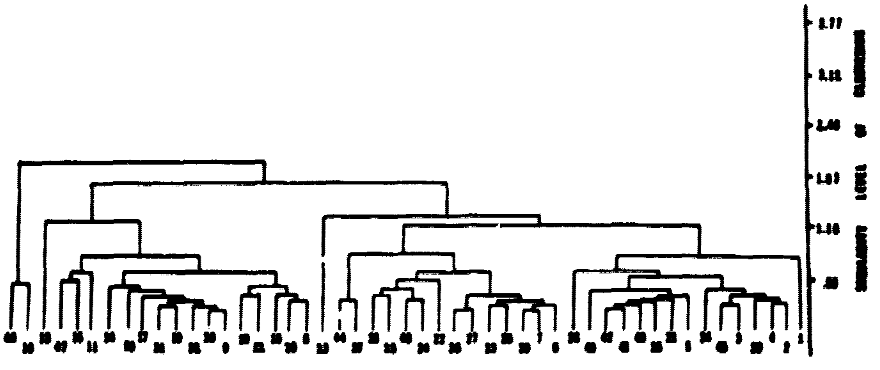

Fig. 2. *Q*-mode dendrogram (using raw distance) for the set of Apollo 15 mare basalts.

although it is properly classified with the other olivine normative basalts, it does not fall into either of the postulated sub-groups. Two splits of 15065 (sample numbers 18 and 46 in Table 1) are separated from the major clusters. These samples are characterized by abnormally low Al_2O_3 (5.36 and 6.05%, respectively) and high FeO (23.75 and 23.77%, respectively). Cuttitta *et al.* (1973) suggested that gabbro 15065 probably contained two distinct chemical and petrographic domains: (1) one with a relatively high mafic to plagioclase ratio (samples 18 and 46 in Table 1) and (2) one with a mixture of mafic and felsic components (samples 31 and 42 in Table 1). Chappell and Green (1973) considered that samples 15622 and 15636 came from the same parent rock and that 15658, 15668, and 15674 also were splits from the same parent sample. The former are assigned as group 1 and the latter as group 2.

In order to assess the effects of possible distortion in Fig. 2, a model referred to as non-linear mapping analysis (NLMA) was applied to the Apollo 15 mare basalt data. Howrath (1973) conducted a rather extensive series of tests of NLMA as applied to geologic data and a complete description of the method is given by Howrath (1973) and Salmon (1972). In general terms, the data are considered to be plotted in *M*-dimensional space (where *M* is the number of measured variables and the reference axes are orthogonal). NLMA attempts to locate a plane that can be passed through the cluster of points in *M*-space upon which the projection of sample points reveals any natural groupings of points within the cluster with a minimum of distortion. The pair of vectors corresponding to the variables with greatest variance are used to define the starting projection plane and adjustments are made in an iterative fashion until the relationship between the data points projected onto the plane and the points in *M*-space is maximized. In other words, the starting plane is tilted with respect to the selected vectors until the best possible two-dimensional representation of the *M*-dimensional collection of points is attained. Experience with NLMA (Howrath, 1973 and Butler, unpublished data) reveals that the groupings recognizable using NLMA are quite similar to those revealed by *Q*-mode form of analysis. Output from NLMA consists of a plot of the sample points on the final projection plane. It often occurs that the clusters of points would require greater than two dimensions to reveal inherent arrangements within the data and an estimate of the probable dimensionality is given as part of the output. Interpretation of the results of NLMA, however, is as difficult as with the *Q*-mode dendrogram as the investigator must make a decision as to how much space must occur between adjacent points before they are considered as belonging to different groups.

The training set of 49 Apollo 15 mare basalts was treated using NLMA and the plot of sample projection points is given in Fig. 3. The coordinate directions with greatest variance are those corresponding to the SiO_2 and FeO reference axes and the probable dimensionality of the plot is four. Thus, some distortion is present in the distribution of points in Fig. 3. As with the interpretation of the groupings in Fig. 2 there is a considerable degree of subjectivity associated with interpretation of the distribution of points in Fig. 3. The quartz normative basalts are well separated from the olivine normative group(s). Although it is possible to recognize

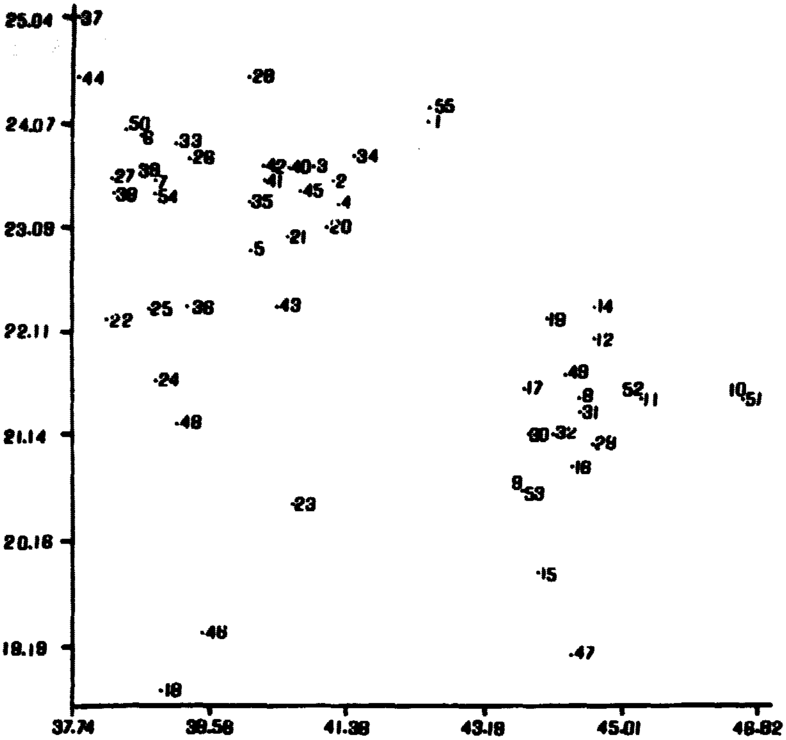

Fig. 3. Non-linear mapping projection for the full set of Apollo 15 mare basalts.

groups 1 and 2 of the olivine normative basalts one could reasonably argue that further subdivision is also possible. For example, the group 1 olivine basalts (located in the upper left hand corner of Fig. 3) could be subdivided using a boundary with a Y (the vertical axis) of value of 22.5. As might be expected, samples 18 and 46 appear again to form a distinct group and samples 13 and 23 are isolated from the major concentrations of points.

In order to assess the hypothesis (Nava, 1974) that the olivine-quartz normative dichotomy disappeared unless one used the normative mineralogy or the SiO_2 content of the samples, the set of 49 Apollo 15 mare basalts was analyzed using NLMA without SiO_2 as a variable. Results of NLMA using the nine variables are given in Fig. 4. The probable dimensionality of the cluster of points is two and the coordinate directions with maximum variance are those corresponding to FeO and MgO. The probable dimensionality of two suggests that essentially the same information could be contained on a binary scatter diagram using FeO and MgO as coordinate axes. To facilitate further discussion, the postulated groups are outlined in Fig. 4.

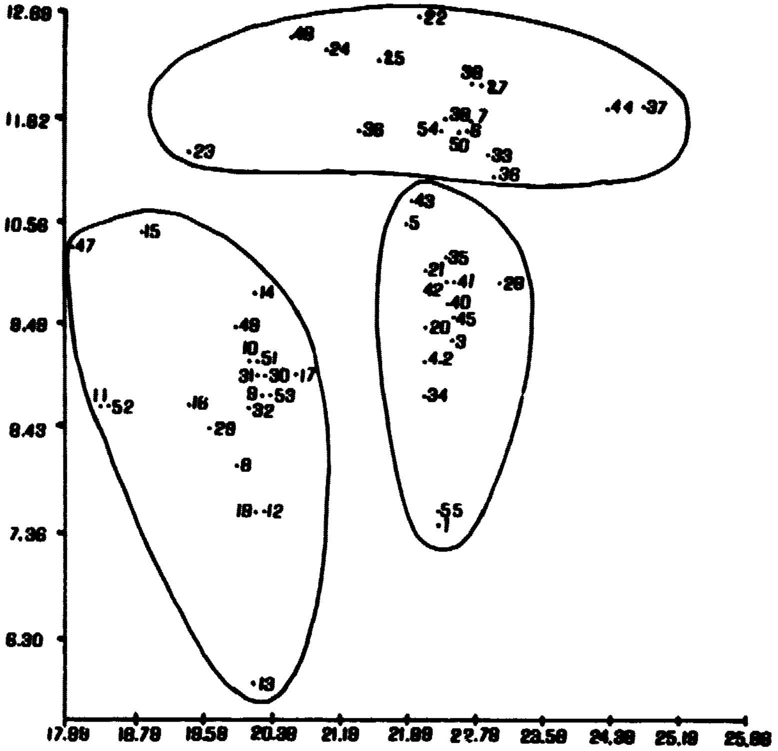

Fig. 4. Non-linear mapping projection for the set of Apollo 15 mare basalts in which SiO_2 was eliminated. Samples within a circled area belong to the groups defined in the text.

DISCRIMINATION OF THE APOLLO 15 MARE BASALTS

If one accepts the existence of three major groups of Apollo 15 mare basalts the strength of the classification can be assessed using discriminant function analysis (DFA). In general, most applications of DFA to chemical data (Chayes, 1964; Chayes and Velde, 1965; Butler and Scotford, 1973) have involved situations in which the classification being examined was based on other than chemical data (such as modal mineralogy, age, structure, stratigraphic position, etc.). In the present investigation, chemistry is being used to assess a classification based on chemical data. Although admittedly circular, it was decided that this was worthwhile as a result of the uncertainty involved in interpreting the relationships in Figs. 2–4. If a classification based on the somewhat arbitrary classification techniques employed can be reclaimed with reasonable efficiency then one can begin to question whether such groups have petrologic significance.

Samples belonging to the three groups are identified in Table 4. Samples 18 and 46 were assigned to the quartz normative group and sample 23 was assigned to group 2 along with all other splits of 15555. Biomedical computer program BMDO

Table 4. Classification of the Apollo 15 mare basalts.

Sample number NASA	Sample number (This report)
Group 1—olivine normative—low SiO_2	
15016	6, 27, 33
15379	26
15555	7, 23, 36, 37, 44, 48
15622	38
15636	39
15643	24
15659	22
15672	25
Group 2—olivine normative—intermediate SiO_2	
15119	34
15256	2, 3
15545	35, 5, 43
15556	1
15557	20, 45
15607	21
15658	40
15668	41, 4
15674	42
15676	28
Group 3—quartz normative—high SiO_2	
15058	10
15065	15, 18, 46, 47
15076	11, 12, 14
15117	19
15475	8, 29
15486	16
15495	17
15499	30, 9
15595	31
15597	32, 49
15118	13

M (stepwise DFA) was employed (Dixon, 1970 describes the routine and the method in considerable detail). Discriminant functions were computed for the three groups of Apollo 15 mare basalts in a stepwise fashion. The variable which had the largest influence on discriminating between the three groups was added to the discriminant functions at step 1 and additional variables were added to the functions using the same criterion. This continued until either all variables were added to the functions or until the addition of another variable did not change the discrimination. The stepwise procedure is highly appealing as the sequence of importance of the variables in the classification can be determined in addition to assessing differences between the groups. In the following discussion of assignment, efficiency is defined as the percentage of the total number of samples correctly assigned.

SiO_2 is the single most important discriminant. All 17 of the group 3, 13 of the group 2, and 12 of the group 1 samples were correctly assigned for an assignment efficiency of 89%. TiO_2 was added at step number 2 with an increase in efficiency to 94%. In other words, the pair SiO_2–TiO_2 provides a better discrimination of the 49 samples into their three groups than any other pair of variables. The difference between this result and the choice of SiO_2 and FeO as the coordinates of the NLMA projection plane results from the fact that DFA makes use of the within groups correlation coefficient matrix which effectively equally weights all variables. In NLMA, on the other hand, the variables are weighted according to their variance. Al_2O_3 is the third variable added which gives an assignment efficiency of 98%. Addition of the remaining seven variables does not change this result. Coefficients of the three discriminant functions are given in Table 5. The discriminant functions are evaluated by multiplying the measured value of a variable for a particular sample by its appropriate coefficient and the function that yields the largest algebraic solution identifies the group to which the sample is assigned. Sample 27 (classified as a group 2 basalt, Table 4) is consistently assigned to the group 1 mare basalts as a result of its low (44.43%) SiO_2 content.

DFA was repeated omitting SiO_2 to assess Nava's (1974) implication that SiO_2 was not an essential part of a classification attempt for some coarse-grained mare basalts. Three variables lead to an assignment efficiency of 98%: (1) FeO (76%); (2) MgO (94%); and (3) Al_2O_3 (98%). Coefficients of the three discriminant functions are given in Table 5b.

Thus, the three groups of Apollo 15 mare basalts can be satisfactorily assigned using no more than three of the ten major oxides and recognition of the three groups does not require using SiO_2 as a variable. Seven analyses of Apollo 15 mare basalts (LSPET, 1972) were assigned to one of the three groups using the discriminant functions in Tables 5a and 5b. Sample numbers, chemical nature,

Table 5a. Coefficients of discriminant function equations using ten oxides.

Variable	Group 1	Group 2	Group 3
SiO_2	537.409	526.326	559.876
TiO_2	90.228	87.091	91.983
Al_2O_3	784.926	761.137	796.706
Constant	−13607.566	−13023.107	−14703.089

Table 5b. Coefficients of discriminant function equations using nine oxides (no SiO_2).

Variable	Group 1	Group 2	Group 3
FeO	343.539	351.576	326.821
MgO	282.244	289.326	265.493
Al_2O_3	205.147	213.315	193.520
Constant	−5700.992	−6017.839	−5082.127

Table 6. Classification of the Apollo 15 mare basalts (LSPET, 1972).

Sample number	Chemical type	Petrographic type	Classification
15016	olivine normative	highly vesicular basalt	Group 1—low SiO_2
15058	quartz normative	porph. clino. basalt	Group 3—high SiO_2
15076	quartz normative	porph. clino. basalt	Group 3—high SiO_2
15256	olivine normative	recry. basalt breccia	Group 2—intermediate SiO_2
15499	quartz normative	porph. clino. basalt	Group 3—high SiO_2
15555	olivine normative	porph. olivine basalt	Group 1—low SiO_2
15556	olivine normative	highly vesicular basalt	Group 2—intermediate SiO_2

petrographic type, and assignments are given in Table 6. Splits of the same sample are assigned to the same group and both sets of discriminant functions lead to the same assignment.

PROPERTIES OF THE THREE GROUPS OF APOLLO 15 MARE BASALTS

Textural and modal mineralogical data for the 49 member training set of analyses and the set of seven "unknowns" do not indicate a strong relationship between petrographic type (LSPET, 1972) and the three chemically defined groups. Of the five olivine basalt vitrophyres, three are classified as group 1 basalts, and two as group 2. There is a better agreement with Nava's postulated third group of coarse-grained olivine basalts and group 1 basalts although Nava (1974) was working with a fewer number of samples.

Summary statistics for the three groups of basalts are given in Table 7. Group 1 basalts have the lowest SiO_2, intermediate TiO_2, MgO greater than 10%, and an average MgO/CaO ratio of 1.22. FeO accounts for the greatest percentage of the variance and, for variables present in amounts greater than 1%, TiO_2 is the most variable. The hypothetical open variance of SiO_2 is negative with the result that individual correlations can not be tested for significant departures from randomness. Group 2 basalts have intermediate SiO_2, high TiO_2, and an average MgO/CaO ratio of 0.94. MgO accounts for the greatest percentage of the total variance and is the most variable of the major oxides. Again, the hypothetical open variance of SiO_2 is negative. Group 3 basalts are the quartz normative basalts with high SiO_2, low TiO_2, and an MgO/CaO ratio of 0.85. The hypothetical open variance of SiO_2 is again negative.

The normative compositions of the basalts also emphasize the differences between groups 1, 2, and 3. Using published norms by Rhodes and Hubbard (1973) and Chappell and Green (1973), the group 2 basalts contain an average of 10.2% normative olivine (ten analyses) whereas the group 1 basalts contain an average of 18%. The largest normative olivine content for a group 2 basalt is 14.7% (15545, Chappell and Green, 1973) and the smallest olivine content for a group 1 basalt is 16.6% (15555, Chappell and Green, 1973). Thus, there is no overlap between groups 1 and 2 on the basis of normative olivine. Similarly, normative ilmenite is

Table 7. Summary statistics for groups 1, 2, and 3 Apollo 15 mare basalts.

	SiO_2	Al_2O_3	TiO_2	FeO	MnO	MgO	CaO	Na_2O	K_2O	P_2O_5
Group 1										
Mean	44.86	8.64	2.21	22.59	0.29	11.54	9.43	0.28	0.05	0.08
Coefficient of variation	0.009	0.081	0.122	0.051	0.085	0.043	0.027	0.127	0.233	0.291
Maximum	44.54	10.39	2.65	24.75	0.34	12.31	10.03	0.35	0.06	0.12
Minimum	44.12	7.50	1.61	20.31	0.25	10.82	8.91	0.21	0.03	0.05
% Variance	6.65%	20.96%	3.08%	55.82%	0.03%	10.58%	2.86%	0.05%	0.005%	0.03%
Group 2										
Mean	45.43	8.93	2.54	22.53	0.30	9.62	10.23	0.45	0.04	0.08
Coefficient of variation	0.009	0.031	0.042	0.011	0.042	0.069	0.028	0.189	0.060	0.213
Maximum	46.11	9.55	2.79	23.21	0.32	10.48	10.97	0.45	0.06	0.11
Minimum	44.44	8.37	2.35	22.18	0.28	7.83	9.70	0.21	0.03	0.05
% Variance	18.04%	9.41%	1.36%	7.45%	0.02%	53.50%	9.81%	0.36%	0.001%	0.03%
Group 3										
Mean	48.93	9.48	1.82	20.09	0.28	8.94	10.51	0.32	0.06	0.11
Coefficient of variation	0.008	0.063	0.108	0.031	0.069	0.105	0.047	0.130	0.272	0.986
Maximum	49.27	10.89	2.09	20.81	0.30	10.62	11.74	0.40	0.08	0.50
Minimum	47.53	8.13	1.45	18.65	0.23	6.54	9.67	0.24	0.03	0.03
% Variance	7.86%	17.28%	1.86%	18.47%	0.02%	43.32%	11.57%	0.08%	0.01%	0.52%

high for the type 2 and low for the type 1. Trace element data have not been used in this study, but published values show some differences between groups 1 and 2. For example, the average Y content for the type 1 is 28 ppm as compared with an average of 36 ppm for type 2 (based on 5 and 4 analyses respectively (Chappell and Green, 1973).

In all three groups, SiO_2 has the greatest mean and the smallest coefficient of variation. For closed variables with means less than 50%, the open variance will be negative if the square of the coefficient of variation is less than the sum of the open variances (Chayes, 1971). In effect, assessment of the strengths of linear relationships in the three groups is made impossible because SiO_2 is behaving as a constant within each of the three groups. In published analyses of the effects of closure on sets of chemical analyses (Chayes, 1971; Butler, 1975), the occurrence of negative open variances for SiO_2 has not been observed. However, published sets of chemical analysis usually reflect the desire of the investigator to illustrate some differentiation scheme and as a result an attempt is made to include a wide range of analyses. This effect is well illustrated in Table 2 giving the summary statistics for the combined set of analyses. The range of all variables is increased and there are no negative open variances.

Various transformations have been designed to eliminate negative open variances (Chayes, 1971). However, these transformations have involved recalculation of Na_2O to normative albite and/or recalculation of K_2O to normative orthoclase with proper adjustments made to SiO_2 and Al_2O_3. Because the total alkali content of the Apollo 15 mare basalts is very small the previously used transformations do not eliminate the negative open variances for SiO_2 in the individual groups. Until a suitable transformation is found, R-mode analyses of the Apollo 15 mare basalts must be confined to the complete set of data.

PETROGENESIS OF THE APOLLO 15 MARE BASALTS

A considerable amount of information and speculation concerning the petrogenesis of lunar basalts in general and Apollo 15 mare basalts in particular has already been published and only a brief synopsis will be presented here. The group 3 quartz normative Apollo 15 mare basalts were collected from a wide range of geographic positions, samples were collected from Elbow and Dune Craters, the Apennine Front (Station 2), the edge of the Hadley Rille (Station 9) and at the ALSEP site. At Dune Crater, the quartz normative vitrophyres appear to underlie the porphyritic quartz normative basalts (Rhodes and Hubbard, 1973). Primarily on the basis of regolith compositions, Rhodes and Hubbard (1973) suggested that the olivine normative basalts were younger than the quartz normative basalts and that the quartz normative basalts were present at the surface as the result of excavation by impact.

Most investigators have stated that the quartz and olivine normative types of Apollo 15 mare basalts can not be genetically related through a near-surface crystal fractional mechanism involving olivine, pyroxene, or spinel (Rhodes and Hubbard, 1973; Chappell and Green, 1973). Within the quartz normative group the observed chemical variation can be related to varying amounts of fractionation of olivine at the surface or near the surface.

Rhodes and Hubbard (1973) noted that although the majority of the olivine normative basalts were collected from Station 9A at the edge of Hadley Rille, there were samples collected that extended the range of this group to the extremities of the Apollo 15 site. The olivine normative basalts may have formed by a higher degree of partial melting than was involved in the genesis of the Apollo 12 basalts (Chappell and Green, 1973). Chemical variations within the olivine normative basalts are related to near-surface olivine fractionation by Rhodes and Hubbard (1973) although Chappell and Green (1973) state that trace- and minor-element variations within the group are indicative of derivation from more than one unit.

As a result of previous investigations of the genesis of the Apollo 15 mare basalts, it is necessary to question the validity of the two groups of olivine normative basalts postulated in this paper. There can be no doubt that the two groups can be very efficiently separated on the basis of modal and normative mineralogy, major-element chemistry, and perhaps minor-element chemistry as well. Does this reflect a hiatus to inadequate sampling or do these two groups represent two different lunar units? Analysis of the trace-element distributions may provide a means of differentiating the two groups. Occurrence of negative open variances of SiO_2 in the three groups of Apollo 15 mare basalts unfortunately precludes comparisons based on correlations between major and minor elements.

SUMMARY AND CONCLUSIONS

(1) If one uses the expected closure correlations as null values a set of significant correlations is obtained that differs from the set obtained using zero as the null value.

(2) It may be possible to describe the Apollo 15 mare basalts in terms of an opaque, an olivine/pyroxene, an anorthite, and a KREEP component if significant correlations are identified using the expected correlations as null values.

(3) Using Q-mode cluster analysis and non-linear mapping it is possible to recognize three groups of Apollo 15 mare basalts. Groups 1 and 2 belong to the previously recognized olivine normative basalt cluster and group 3 to the quartz normative cluster.

(4) DFA reclaims the three-fold subdivision with a high assignment efficiency and the discriminant functions can be used to assign other Apollo 15 mare basalts to one of the three postulated groups. Only three variables are required to yield the most efficient discriminant functions.

(5) Group 1 Apollo 15 basalts have low SiO_2, intermediate TiO_2, high FeO, high MgO, and an MgO/CaO ratio greater than 1.0. Group 2 Apollo 15 mare basalts have intermediate SiO_2, high TiO_2, high FeO, low MgO, and an MgO/CaO ratio of less than 1.0.

(6) The apparent subdivision of the olivine normative basalts into groups 1 and

2 could reflect inadequate sampling at the Apollo 15 site. If, however, the groups are real, in the sense that the chemical hiatus does indeed exist, then it is possible that the two groups represent different lunar units (flows?) or variations within the same unit with the group 1 basalts being enriched in normative olivine and group 2 enriched in the non-olivine mafic silicate phases. It is also possible that the two groups are not genetically related but owe their genesis to partial melting at different levels in the lunar interior.

(7) The occurrence of negative open variances for SiO_2 in all three groups precludes testing individual correlations for significant differences or for creating correlation family diagrams that might aid in elucidating differences between the two groups. Rational transformations designed to eliminate the negative open variance of SiO_2 are currently being investigated.

Acknowledgments—The author would like to thank D. Nava and J. Lindsey for their competent and critical reviews of the manuscript, J. Butler for preparing the illustrations, and M. Howe for typing various versions. The work reported in the paper was supported by NASA grant NSG-7198.

REFERENCES

Butler J. C. (1974) Analysis of correlations between percentages. *J. Geol. Educ.* **1**, 56–61.

Butler J. C. (1975) The occurrence of negative open variances in ternary systems. *J. Math. Geol.* **7**, 31–45.

Butler J. C. (1976) Principal components analysis using the hypothetical closed array. *J. Math. Geol.* **8**, 47–70.

Chayes F. (1960) On correlation between variables of constant sum. *J. Geophys. Res.* **65**, 4184–93.

Chayes F. (1962) Numerical correlation and petrographic correlation. *J. Geol.* **70**, 440–452.

Chayes F. (1971) *Ratio Correlation*, p. 98. Univ. Chicago Press, Chicago, Illinois.

Chayes F. and Kruskal W. (1966) An approximate statistical test for correlations between proportions. *J. Geol.* **74**, 692–702.

Chappell B. W. and Green D. H. (1972) Chemistry, geochronology and petrogenesis of lunar sample 1555. *Science* **175**, 415–416.

Chappell B. W. and Green D. H. (1973) Chemical compositions and petrogenetic relationships in Apollo 15 mare basalts. *Earth Planet. Sci. Lett.* **18**, 237–246.

Cuttitta R., Rose H. J., Annell C. S., Carron M. K., Christian R. P., Ligon D. T., Dwornik E. J., Wright T. L., and Greenland L. P. (1973) Chemistry of twenty-one igneous rocks and soils returned by the Apollo 15 mission. *Proc. Lunar Sci. Conf. 4th*, p. 1081–1096.

Dixon W. J., ed. (1970) *BMD Biomedical Computer Programs*. Univ. California Press, Berkeley, California. 600 pp.

Helmke P. A., Blanchard D. P., Haskin L. A., Telander K. L., Weiss K. C., and Jacobs J. W. (1973) Major and trace elements in crystalline rocks from Apollo 15 (abstract). In *Lunar Science IV*, p. 349–351. The Lunar Science Institute, Houston.

Howrath R. J. (1973) Preliminary assessment of a non-linear mapping algorithm in a geological content. *J. Math. Geol.* **5**, 38–58.

LSPET (1972) The Apollo 15 lunar samples; a preliminary description. *Science* **175**, 363–375.

Mason E., Jarosewich E., and Melson W. G. (1972) Mineralogy, petrology, and chemical composition of lunar samples 15085, 15256, 15271, 15475, 15476, 15535, 15555, 15556. *Proc. Lunar Sci. Conf. 3rd*, p. 758–796.

Nava D. F. (1974) Chemical compositions of some soils and rock types from the Apollo 15, 16, and 17 lunar sites. *Proc. Lunar Sci. Conf. 5th*, p. 1087–1096.

Rhodes J. M. (1972) Major element chemistry of Apollo 15 mare basalts. In *The Apollo 15 Lunar Samples* (J. Chamberlain and C. Watkins, eds.), p. 250–252. The Lunar Science Institute, Houston.

Rhodes J. M. and Hubbard N. J. (1973) Chemistry, classification, and petrogenesis of Apollo 15 mare basalts. *Proc. Lunar Sci. Conf. 4th*, p. 1127–1148.

Rose H. J. Cuttitta F., Berman S., Carron M. K., Christian R. P., Dwornik E. J., Greenland L. P., and Ligon D. T., Jr. (1974) Compositional data for twenty-two Apollo 16 samples. *Proc. Lunar Sci. Conf. 5th*, p. 1149–1158.

APPENDIX IV.

Al/Si VARIATIONS IN APOLLO 11, 12, AND 15 MARE BASALTS AND REGOLITH SAMPLES

Paper presented to the Conference on Comparisons of Mercury and the Moon, Houston, Texas, 15-17 November, 1976.

Accepted for publication in the Conference Proceedings.

Al/Si VARIATION IN APOLLO 11, 12, AND 15 MARE BASALTS AND REGOLITH SAMPLES:

John C. Butler, Geology, University of Houston, Houston, TX 77004

The direct measurement of Al/Si ratios in the uppermost portion of the lunar regolith by the Apollo 15 and 16 missions makes one optomistic concerning possibilities for future unmanned geochemical surveys of the moon and other planetary bodies. These data have proven extremly useful in examining large scale lateral chemical heterogenieties and in mapping local chemical variations. Previous investigators have noted that the between site chemical variations reflect compositional differences in the parent material which is primarly the local bedrock. The ability of variations in measured Al/Si to mark the highlands-mare material clearly has been demonstrated and a nearly complete trace and major element analysis has been obtained for the lunar highlands regolith by making certain assumptions and having access to ground-truth observations. Without access to ground-truth chemical analyses (which will be the situation for initial geochemical surveys of other planetary bodies) the present system will not yield absolute values of Al and Si. Chemical analyses of Apollo 11, 12 and 15 basalt and regolith samples have been extracted from the lunar data base in an attempt to discover how much information in contained in the Al, Si, and Al/Si variation (Table 1.)

TABLE 1.

Sample	Mean Si	Mean Al	C Si	C Al	rAl:Si	rAl:Al/Si
Apollo 11 Basalts	18.98	5.00	.026	.145	-.044	0.980
Apollo 11 Fines	19.61	7.21	.026	.036	-.024	0.782
Apollo 12 Basalts	20.87	5.12	.054	.204	.248	0.938
Apollo 12 Fines	21.61	7.30	.083	.083	.432	0.961
Apollo 15 Basalts	21.50	4.84	.047	.187	.373	0.954
Apollo 15 Fines	21.83	7.70	.019	.211	.379	0.964
Apollo 12 and 15	21.44	5.93	.044	.294	.350	0.977

At a given site the regolith samples tend to have higher Al, essentially the same Si and higher Al/Si ratios than the mare basalts from the same site. The correlation between a ratio and its numerator is a simple function of the coefficients of variation of the parent variables (Al and Si) if the Al:Si correlation is zero. As Si has a relatively small coefficient of variation the correlations between Al and Al/Si are quite large except for the Apollo 11 fines. Plots of Al/Si versus Al exhibit a strong positive linear trend. There is a considerable difference between the trends for the Apollo 11 basalt and regolith samples. However, there is no statistical difference between the lines of organic correlation for the Apollo 12 and Apollo 15 basalt and regolith samples. In fact, the Apollo 12 and 15 basalt and regolith samples can be combined into a single data set (Table 1) for which the equation of organic correlation is $Al=22.89(Al/Si) - 0.397$. These preliminary analyses suggest that it may be possible to estimate the Al content of the lunar surface materials from the measured Al/Si ratios with relatively small associated errors.