Q-Mode Factor Analysis of Geochemical and Petrologic Data Matrices with Constant Row-Sums

GEOLOGICAL SURVEY PROFESSIONAL PAPER 574–G

Q-Mode Factor Analysis of Geochemical and Petrologic Data Matrices with Constant Row-Sums

By A. T. MIESCH

STATISTICAL STUDIES IN FIELD GEOCHEMISTRY

GEOLOGICAL SURVEY PROFESSIONAL PAPER 574–G

An extension of the method of Q-mode factor (vector) analysis to increase its usefulness in geochemical and petrologic investigations

UNITED STATES GOVERNMENT PRINTING OFFICE, WASHINGTON : 1976

UNITED STATES DEPARTMENT OF THE INTERIOR

THOMAS S. KLEPPE, *Secretary*

GEOLOGICAL SURVEY

V. E. McKelvey, *Director*

Library of Congress Cataloging in Publication Data
Miesch, Alfred T.
Q-mode factor analysis of geochemical and petrologic data matrices with constant row-sums.
(Statistical studies in field geochemistry) (Geological Survey Professional Paper 574-G)
Bibliography: p.
Supt. of Docs. no.: I 19.16:574-G
1. Geochemistry—Statistical methods. 2. Petrology—Statistical methods. 3. Vector analysis.
I. Title. II. Series. III. Series: United States Geological Survey Professional Paper 574-G.
QE515.M54 551.9'01'82 75-619414

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402
Stock Number 024-001-02909-0

CONTENTS

ILLUSTRATIONS

TABLES

STATISTICAL STUDIES IN FIELD GEOCHEMISTRY

Q-MODE FACTOR ANALYSIS OF GEOCHEMICAL AND PETROLOGIC DATA MATRICES WITH CONSTANT ROW-SUMS

By A. T. MIESCH

ABSTRACT

Matrices of data representing all the major constituents in a suite of rock samples tend to have constant row-sums, a property that has caused considerable difficulty in attempts to interpret rock genesis from the correlations among matrix columns. *Q*-mode factor analysis may be used to interpret rock genesis from relations among matrix rows, and the constant row-sum is a definite asset. The constant can be used to compute scalers, which, in turn, can be used to adjust the factor loadings and scores to conform with the original data. Therefore, principal components and varimax models and a variety of oblique models may be derived that can be used to recompute estimates of the data in, for example, percent or parts per million. Any real or hypothetical composition may also be tested as a possible end-member in a petrologic mixing problem.

Other advantages of scaling the factor loadings and scores to conform with the original data, and thereby deriving composition loadings and scores, are that (1) the loadings sum to unity and may be interpreted as proportions of end-members; (2) the scores sum to the constant row-sum of the original data matrix and, if none are negative, may be interpreted as composition values in units of percent or parts per million; and (3) the signs of the loadings become fixed and are dependent only on the choice of end-members for the model and on the nature of the compositional variation in the rocks under examination.

When the number of factors in a *Q*-mode factor model is less than the number of variables in the data matrix, the model accounts for less than the total variation in the data and may be used to reproduce the data matrix only approximately. The degree of approximation with which the normalized form of the data may be reproduced is indicated precisely by the eigenvalues of the matrix of coefficients of proportional similarity and by the sample communalities. These measures, however, are only approximate indicators of the degree to which the derived model may be used to reproduce the original data. The proportion of the variance for each variable that the model acocunts for is given by the coefficient of determination between the original and reproduced data.

The coefficients of determination have been used to construct factor-variance diagrams for four suites of igneous-rock samples. The diagrams show the proportion of the total variance in each chemical constituent that can be explained by petrologic models containing any number of factors, or end-members, up to the number of constituents present in the samples. Petrologic models have been developed for all four suites by choosing reference vectors with composition scores that approach the compositions of end-members thought to have been involved in the petrologic system. The resultant models are, for the most part, in reasonable accord with geologic observations by other workers and serve to quantify various aspects of the processes by which the rocks may have originated.

The four suites of samples are from (1) a rhyolite-basalt complex in Yellowstone National Park, Wyo., (2) a granitoid intrusive in the southern part of the Snake Range, Nev., (3) lavas and pumices from the 1959 summit eruption of Kilauea, Hawaii, and (4) the layered series of the Skaergaard intrusion, Greenland.

GENERAL INTRODUCTION

Work of Chayes (1960) has shown that matrices of compositional data on rocks and other materials are difficult to examine for genetic implications because such matrices either have constant row-sums or are parts of larger matrices that have constant row-sums. A test designed to overcome the difficulty that the constant row-sum presents (Chayes and Kruskal, 1966) was found to be ineffective (Miesch, 1969). The situation is a serious one because petrologists work with this type of data almost constantly and the ultimate goal of their work is to understand rock genesis. Mathematical geologists have known for some time that the constant row-sum property of compositional data was no serious hindrance in the method of *Q*-mode factor analysis introduced by Imbrie (1963), but it is now apparent that in order to derive a *Q*-mode model in terms of the original data the constant row-sum is a definite asset.

Although the method being used is referred to as factor analysis throughout this report, this name may be unacceptable to many students and workers in multivariate statistics. The diagonal values in the similarity matrices have been unity in all applications of the method thus far, and the method, therefore, might best be referred to by the general terms "vector" or "components" analysis, but the term "factor" analysis is used because of its prevalence in geologic literature.

The purposes of this report are to describe the mathematical basis for an extended form of the *Q*-mode method originally developed by Klovan and Imbrie (1971) and to demonstrate its application to four problems in igneous petrology. The original method leads to a factor model that can be used to reproduce the observed data in a

normalized form only. Because of this, the models are difficult to interpret in terms of geologic conditions and processes. It will be shown here that if the row-sums in the original data are constant, these models can be modified to reproduce unbiased approximations of the original data in units of, for example, weight percent or parts per million. This modification further allows the geochemist or petrologist to use some other methods to develop a variety of models that are mathematically acceptable and to examine each of them for geologic plausibility.

(Note.—After this report was written, an abbreviated description of the extended *Q*-mode method (Miesch, 1975) and computer programs that perform most of the basic computations (Klovan and Miesch, 1975) were prepared for publication in Computers & Geosciences.)

Among the more significant of these other methods is one that enables the geochemist or petrologist to propose various compositions as those of end-members in a geochemical or petrologic system and then to test these propositions. If the compositional variation in a basaltic magma, for example, is thought to have resulted from the crystallization and redistribution of olivine, it is possible not only to test this hypothesis but also to determine the olivine composition most likely to have been involved. Further, for a given parent-magma composition, one can determine the amount of olivine that would have had to have been separated from or added to the magma to form each sample of basalt.

The compositional variation that will be observed in most sample suites will have resulted from processes that were more complex than the simple separation of one mineral from a magma. Other methods made possible by the presence of the constant row-sum enable one to determine the minimum number of end-members required to account for any given proportion of the variance in each compositional variable. This can be done before one knows what the end-members might be. The methods have clear advantages over eigenvalues, which are commonly used to determine the number of end-members required, and are suggested as alternative means for this purpose.

The principal advantage of the extended form of *Q*-mode factor analysis may be that the derived model is more easily interpreted; the composition scores (end-member compositions) are in the same units as the original data, and the composition loadings (mixing or unmixing proportions) sum to unity for each sample and have signs that are fixed rather than arbitrary. Thus, not only can the investigator examine the plausibility of proposed geologic processes more easily but also the results of his work can be presented in a form that will be more understandable and acceptable to nonmathematical colleagues.

In part I the mathematical bases of the methods are given where the number of factors in the model is equal to the number of columns (representing compositional variables) in the data matrix. The mathematical effects of reducing the number of factors in the model are described in part II, and applications of the method to four petrologic mixing problems are discussed in part III.

The methods described in this report are intended to supplement those described by Klovan and Imbrie (1971). Their computer program gives the method for derivation of the unscaled factor scores from which the composition scores, described here, are computed. Their program also provides an efficient method for deriving a *Q*-mode factor solution for large data matrices. Klovan and Imbrie's program was adapted to the U.S. Geological Survey's system of statistical programs (STATPAC) by George Van Trump.

Parts I and II of this report were read by J. J. Connor, L. J. Drew, and J. E. Klovan; part III was read by R. W. White. I am grateful to each of these reviewers for criticisms that led to improvement of the manuscript. In addition, I benefited from discussions of the petrologic aspects of the work with Fred Barker, D. E. Lee, G. J. Neuerburg, and R. E. Wilcox.

Part I. — Mathematical Basis Where the Number of Factors Equals the Number of Variables

INTRODUCTION

The type of data matrix being considered is one wherein the rows represent rock samples and the columns represent chemical or mineralogical variables. Where the data are expressed as percentages, each row sums to about 100, depending on whether all the major constituents present in the rocks are represented in the data and on the nature of the analytical errors. Large departures from 100 may be eliminated by recomputing the row to sum to 100, thereby spreading the effect of the analytical errors over all the constituents in proportion to their abundances and directing the *Q*-mode analysis at only the part of the rock that is composed of the constituents represented in the data. If all row-sums are less than 100, the departures may also be eliminated by subtracting the row-sum from 100 and treating the difference as a new compositional variable—either some major constituent not represented in

the data or a composite of constituents that may be referred to as "others".

As has been customary in most geologic applications of Q-mode factor analysis, the matrix of original data can be treated in a number of different forms. The first type of transformation consists of scaling columns (representing variables) of the matrix so that the magnitudes of the values are approximately the same from one column to another. This can be done by either of two options provided in the computer program of Klovan and Imbrie (1971). The first option is used to express each column as a proportion of the maximum value in the column so that no values exceed unity; the second option is used to express each value as a proportion of the total range for the column so that all values within each column are in the range from zero to unity. The purpose of this scaling is to give each compositional variable an approximately equal weight in the factor analysis. Without scaling, in treating conventional petrographic variables, for example, variables with large variance, such as SiO_2, may completely dominate the outcome, and variables with small variance, such as Na_2O, may exert little or no influence. In effect, the use of the scaling options in Klovan and Imbrie's program overcomes the problems caused by the fact that the means and variances in most geochemical and petrographic data are strongly related, and it recognizes the fact that minor compositional constituents can be just as diagnostic as major constituents in arriving at petrogenetic models.

Whether the original data are transformed by scaling or not, the data matrix is also treated in a row-normalized form. That is, each row of the matrix is adjusted so that the sum of the squares of the values within it is unity. This adjustment is done automatically when using Imbrie and Purdy's (1962) coefficient of proportional similarity (cosine theta); in effect, the matrix of cosine theta is computed as the major product of the row-normalized data matrix and its transpose. The appropriateness of cosine theta as a measure of similarity among rock compositions was shown by Imbrie (1963). Aside from providing an appropriate measure of compostional similarity, the use of cosine theta allows treatment of each row of the data matrix as a vector of unit length, thereby facilitating subsequent computation and interpretation.

The conventional forms for a hypothetical data matrix are given in table 1. Tables 1*A* and 1*AA* show, respectively, the original data with constant row-sums and the same data normalized so that the sum of squares for each row is unity. The normalization factors, t_i, which are the square roots of the sums of squares for the corresponding rows in the original data matrix, are also given. Table 1*B* shows the original data transformed so that the maximum value in each column is unity, and table 1*C* shows the original data transformed so that each column ranges from zero to unity. Tables 1*BB* and 1*CC* show the normalized forms of the transformed data.

The relationships among the various forms of the data are illustrated in figure 1, which shows that the row-normalized form of the data, derived from either the original or transformed data, is used to derive the factor model. Then, the factor model, developed by the conventional procedures, can be used to reproduce the row-normalized data. The extended method described here allows one to modify the model to reproduce the data in their original or scaled forms.

Q-MODE FACTOR ANALYSIS

The purpose of Q-mode factor analysis in geochemistry and petrology is to resolve a data matrix into a concise model wherein each analyzed sample of rock, soil, or other material is viewed as a mixture of a small number of theoretical or actual end-member samples. As has been shown by Imbrie (1963), the method is used to determine the number of end-members that are required to account for any given proportion of the total variability in the normalized data and then to estimate the proportions of each end-member present in each sample. Either the

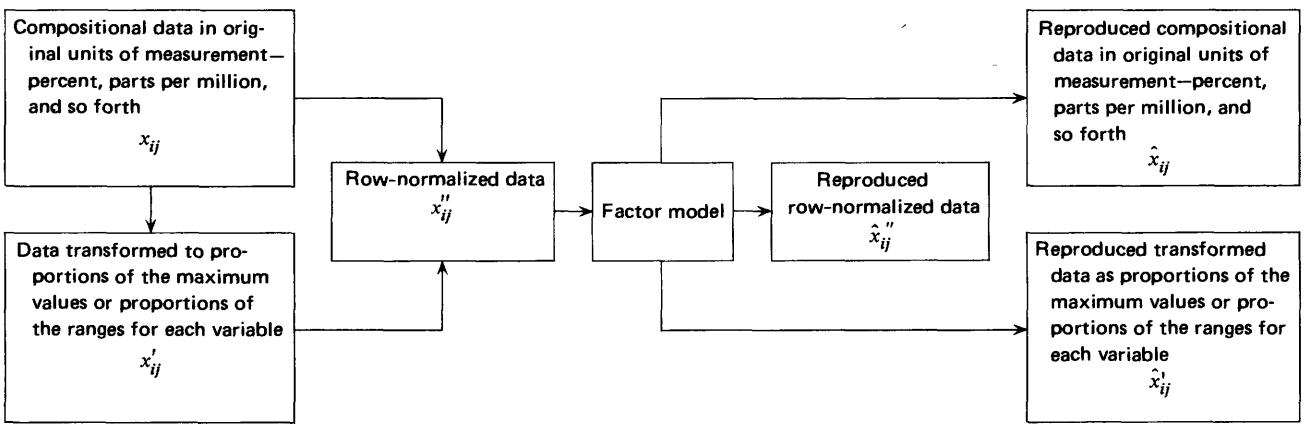

FIGURE 1.—Diagram showing the relationships between the factor model and the various forms of the original and reproduced data.

TABLE 1.—*Original hypothetical data, transformed data, and normalized data used to illustrate the methods of computation*

[t_i is the normalization factor for the ith row; $t_i = (\sum_j x_{ij}'^2)^{1/2}$; $x_{ij}'' = x_{ij}'/t_i$]

A. Original data, $x_{ij}' = x_{ij}$

				t_i	Row-sum
	90	7	3	90.32	100
	50	10	40	64.81	100
$x_{ij}' =$	80	15	5	81.55	100
	60	30	10	67.82	100
	85	10	5	85.73	100
	95	3	2	95.07	100
$xmax_j =$	95	30	40		
$xmin_j =$	50	3	2		

AA. Normalized original data

	0.996	0.078	0.033
	.772	.154	.617
$x_{ij}'' =$	.981	.184	.061
	.885	.442	.147
	.991	.117	.058
	.999	.032	.021

B. Data transformed to proportion of the maximum, $x_{ij}' = x_{ij}/xmax_j$

				t_i
	0.947	0.233	0.075	0.978
	.526	.333	1.000	1.178
	.842	.500	.125	.987
$x_{ij}' =$	.632	1.000	.250	1.209
	.895	.333	.125	.963
	1.000	.100	.050	1.006

BB. Normalized transformed data from *B*

	0.968	0.238	0.077
	.447	.283	.849
	.853	.506	.127
$x_{ij}'' =$	.522	.827	.207
	.929	.346	.130
	.994	.099	.050

C. Data transformed to proportion of the range, $x_{ij}' = (x_{ij} - xmin_j)/(xmax_j - xmin_j)$

				t_i
	0.889	0.148	0.026	0.902
	0	.259	1.000	1.033
	.667	.444	.079	.805
$x_{ij}' =$	.222	1.000	.211	1.046
	.778	.259	.079	.824
	1.000	0	0	1.000

CC. Normalized transformed data from *C*

	0.986	0.164	0.029
	0	.251	.968
	.828	.552	.098
$x_{ij}'' =$	.212	.956	.201
	.944	.315	.096
	1.000	0	0

end-members are theoretical, as in a principal components or varimax model, for example, or they are actual samples in the data.

Q-mode factor analysis, therefore, can be viewed as an attempt to reduce a data matrix into a smaller matrix that may facilitate interpretation in terms of petrologic mixing (for example, mixing of sediments, magmas, or magma and wallrock), or possibly in terms of petrologic mixing and chemical processes (such as solution or precipitation of material that tends to be of a specific composition). Alternatively, *Q*-mode factor analysis can be used simply as a means for classifying samples or for developing a concise description of the compositional variation in a sample suite. In using the method for classifying samples or for developing a concise description of variation, one is capitalizing on the fact that chemical and mineralogical variables are seldom, if ever, all unrelated to each other. The major chemical and mineralogical variations among samples can commonly be described more concisely in terms of a few actual or theoretical end-members than in terms of each one of the compositional variables.

The general *Q*-mode factor model is

$$\hat{x}_{ij}'' = \sum_k a_{ik}'' f_{kj}''. \qquad (1)$$

where $\hat{x}_{ij}''$ is an approximation of the jth element in the ith row of the normalized data matrix, a_{ik}'' is an initial factor loading for the ith sample on the kth factor, and f_{kj}'' is the kth unscaled factor score for the jth variable. The initial factor loadings, the a_{ik}'''s, are the conventional loadings for the principal components, varimax, or oblique models and are those that are provided by widely available computer programs. The unscaled factor scores, the f_{kj}'''s, are provided by the computer program of Klovan and Imbrie (1971) for the principal components model and varimax model.

The matrix of normalized data, x_{ij}'', is $N \times M$ in size ($1 \leqslant i \leqslant N$, $1 \leqslant j \leqslant M$). The matrix of initial loadings, a_{ik}'', is $N \times m$ ($1 \leqslant k \leqslant m$), and the matrix of unscaled factor scores, f_{kj}'', is $m \times M$. Where m, the number of factors in the model (the number of end-members), is set equal to M, the number of chemical or mineralogical variables, the loading

matrix is equal to the data matrix in size, and the model can be used to reproduce the normalized data matrix exactly. In most actual applications, however, m is less than M, and the reproduced data, $\hat{x}''_{ij}$, are only approximations of the original normalized data, x''_{ij}.

For purposes of illustration and verification of the extended form of the Q-mode method, the data given in tables 1*A*, 1*B*, and 1*C* will be used to generate a variety of factor models with m set equal to M. Verification of the models, therefore, will consist of using them to reproduce the original data exactly, except for round-off errors.

The first step in generating a factor model is to compute a matrix of sums of squares and cross products among the columns of the normalized original or transformed data matrices. Klovan and Imbrie (1971) have shown that (1) the eigenvalues of this matrix are equal to the nonzero eigenvalues of a matrix of coefficients of proportional similarity (cosine theta), and (2) the eigenvectors of the same matrix are the unscaled principal component factor scores. The coefficient of proportional similarity (cosine theta) is a measure of similarity between the compositions of two samples, i and p (Imbrie and Purdy, 1962):

$$\cos\theta_{ip} = \frac{\sum_j x'_{ij} x'_{pj}}{\left(\sum_j x_{ij}^2 \sum_j x'^2_{pj}\right)^{1/2}}. \qquad (2)$$

The unscaled varimax factor scores are computed as the product of the matrix of principal component scores and the varimax transformation matrix (Klovan and Imbrie, 1971).

Other details of the procedures for deriving the initial factor loadings are given by Imbrie (1963) and Klovan and Imbrie (1971). It will suffice here to point out that the only important difference among the principal components, varimax, and various oblique models is that the vectors representing the samples are described in terms of different sets of reference axes. (See Imbrie, 1963.) The initial factor loadings, a''_{ik}, and the unscaled scores, f''_{kj}, for the principal components and varimax models, and for the type of oblique model developed by Imbrie (1963), which uses extreme sample vectors as reference axes, are given in tables 2, 3, and 4. Those in table 2 were derived using the original data without transformation (table 1*A*). Those in table 3 were derived from the data transformed to proportions of the maximum value in each column (table 1*B*). Those in table 4 were derived from the data transformed to proportions of the total range for each column (table 1*C*). The products of the a''_{ik} and f''_{kj} matrices for each of the nine models represented are exactly equal to the corresponding normalized data matrices in tables 1*AA*, 1*BB*, and 1*CC*. The principal component and varimax loadings and scores in tables 2, 3, and 4 were derived using the computer program given by Klovan and Imbrie (1971). The loadings for the oblique models were derived using an addition to the program that follows the technique described by Imbrie (1963) and Manson and Imbrie (1964). Because m equals M in these examples, the scores for the oblique models are simply the normalized forms of the original and transformed data and are taken directly from tables 1*AA*, 1*BB*, and 1*CC*.

SCALED SCORES AND COMPOSITION LOADINGS

Models of the type represented in tables 2, 3, and 4 are difficult to interpret because the factor scores, the f''_{kj}'s, are normalized and, therefore, dimensionless. Also, the initial loadings, the a''_{ik}'s, do not sum to unity across k. It is preferable to express the scores in units conformable to those of the transformed data or those of the original data, commonly weight percent concentration. Klovan and Imbrie (1971) scale the factor scores by multiplying them by the square root of M, but this procedure is only for purposes of standardization. Alternative, and realistic, scaling is possible if the row-sums of the original data matrix are constant across all rows, so that

$$\sum_j x_{ij} = K, \qquad (3)$$

where $K = 100$ if the original data are percentages. Similarly, factor scores in units of the original data should sum to the same constant, inasmuch as the scores should represent the compositions of real or theoretical end-member samples:

$$\sum_j f_{kj} = K. \qquad (3a)$$

The transformed data may be regarded as having been derived from a general expression found useful for computer programming:

$$x'_{ij} = (x_{ij} - xmin_j)/(xmax_j - xmin_j), \qquad (4)$$

where $xmax_j$ and $xmin_j$ are the maximum and minimum values in the jth column of the original data matrix if the data were transformed to proportions of the total range. If transformed to proportions of the maximum value, all values of $xmin_j$ are set equal to zero, and, if not transformed, all values of $xmin_j$ are zero and all values of $xmax_j$ are unity. Similarly, the transformed scores can be defined as

$$f'_{kj} = (f_{kj} - xmin_j)/(xmax_j - xmin_j), \qquad (5)$$

and this can be rearranged to

$$f_{kj} = f'_{kj}\,(xmax_j - xmin_j) + xmin_j. \qquad (6)$$

TABLE 2.—*Initial loadings, a''_{ik}, unscaled scores, f''_{kj}, scale factors, s_k, and the estimated normalization factors, $\hat{t}_i$, for the three-factor principal components, varimax, and oblique factor models derived without transforming the data*

[The products of the matrices of a''_{ik} and f''_{kj} are equal to the normalized data in table 1*AA*]

Model	Initial loadings, a''_{ik}	Unscaled scores, f''_{kj}	Scale factors, s_k	Estimated normalization factors[1] $\hat{t}_i$
Principal components	a''_{ik} = 0.988 −0.138 −0.068; .871 .481 −.095; .996 −.087 .031; .960 .066 .273; .994 −.105 −.035; .981 −.159 −.110	f''_{kj} = 0.973 0.172 0.152; −.186 .199 .962; −.136 .965 −.226	s_k = 77.053; 102.499; 165.778	$\hat{t}_i$ = 90.4; 64.8; 81.5; 67.8; 85.7; 95.1
Varimax	a''_{ik} = 0.851 0.420 0.316; .416 .873 .255; .797 .443 .410; .608 .493 .623; .826 .442 .349; .871 .409 .274	f''_{kj} = 0.888 −0.280 −0.365; .388 .030 .921; .247 .960 −.136	s_k = 412.169; 74.647; 93.393	$\hat{t}_i$ = 90.3; 64.8; 81.6; 67.8; 85.7; 95.0
Oblique	a''_{ik} = 0.901 −0.004 0.112; 0 1.000 0; .661 −.012 .373; 0 0 1.000; .800 .020 .200; 1.000 0 0	[2]f''_{kj} = 0.999 0.032 0.021; .772 .154 .617; .885 .442 .147	s_k = 95.068; 64.807; 67.823	$\hat{t}_i$ = 90.4; 64.8; 81.5; 67.8; 85.7; 95.1

[1] $\hat{t}_i = 1/\sum_k (a''_{ik}/s_k)$.

[2] From table 1*AA*.

TABLE 3.—*Initial loadings, a''_{ik}, unscaled scores, f''_{kj}, scale factors, s_k, and the estimated normalization factors, $\hat{t}_i$, for the three-factor principal components, varimax, and oblique factor models derived after transforming each variable to proportions of the maximum value*

[The products of the a''_{ik} and f''_{kj} matrices are equal to the normalized transformed data in table 1*BB*]

Model	Initial loadings, a''_{ik}	Unscaled scores, f''_{kj}	Scale factors, s_k	Estimated normalization factors[1], $\hat{t}_i$
Principal components	a''_{ik} = 0.969 −0.234 −0.084; .706 .627 −.328; .990 −.037 .135; .851 .294 .435; .991 −.131 −.015; .927 −.324 −.189	f''_{kj} = 0.880 0.416 0.231; −.407 .408 .817; −.245 .813 −.528	s_k = 0.950; 16.001; −4.983	$\hat{t}_i$ = 0.978; 1.179; .987; 1.209; .963; 1.007
Varimax	a''_{ik} = 0.904 0.256 0.343; .261 .927 .268; .749 .300 .591; .371 .339 .864; .843 .307 .441; .951 .226 .211	f''_{kj} = 0.978 −0.128 −0.164; .173 .055 .983; .117 .990 −.076	s_k = 1.212; 1.742; 2.650	$\hat{t}_i$ = 0.978; 1.178; .987; 1.210; .963; 1.006
Oblique	a''_{ik} = 0.880 −0.006 0.185; 0 1.000 0; .578 −.019 .549; 0 0 1.000; .754 .032 .317; 1.000 0 0	[2]f''_{kj} = 0.994 0.099 0.050; .447 .283 .849; .552 .827 .207	s_k = 1.006; 1.178; 1.209	$\hat{t}_i$ = 0.978; 1.178; .988; 1.209; .963; 1.006

[1] $\hat{t}_i = 1/\sum_k (a''_{ik}/s_k)$.

[2] From table 1*BB*.

Substitution of the right side of equation 6 in equation 3a gives:

$$\sum_j \left(f'_{kj} \, (xmax_j - xmin_j) + xmin_j \right) = K. \qquad (7)$$

Although equations 5 and 6 define the relation between the values of f_{kj} and f'_{kj}, their absolute values have not been established. This is done by defining f'_{kj} as equal to $s_k f''_{kj}$, where s_k is a scale factor and f''_{kj} is an unscaled score as

TABLE 4.—*Initial loadings, a''_{ik}, unscaled scores, f''_{kj}, scale factors, s_k, and estimated normalization factors, $\hat{t}_i$, for the three-factor principal components, varimax, and oblique factor models derived after transforming each variable to proportions of the range*

[The products of the a''_{ik} and f''_{kj} matrices are equal to the normalized transformed data in table 1*CC*]

Model	Initial loadings, a''_{ik}	Unscaled scores, f''_{kj}	Scale factors, s_k	Estimated normalization factors [1], $\hat{t}_i$
Principal components	$a''_{ik} =$ 0.963 −0.249 0.101; .234 .830 .506; .985 .094 −.147; .603 .643 −.473; .995 −.095 .037; .906 −.375 .196	$f''_{kj} =$ 0.906 0.400 0.138; −.374 .608 .700; .196 −.686 .701	$s_k =$ 0.792; 1.720; 2.656	$\hat{t}_i =$ 0.902; 1.032; .805; 1.045; .823; 1.000
Varimax	$a''_{ik} =$ 0.984 0.012 0.178; .022 .979 .202; .822 .105 .559; .204 .244 .948; .942 .088 .324; .999 −.026 .016	$f''_{kj} =$ 1.000 −0.014 0.026; −.025 .051 .998; .016 .999 −.050	$s_k =$ 0.987; 1.179; 1.747	$\hat{t}_i =$ 0.902; 1.033; .805; 1.046; .823; 1.001
Oblique	$a''_{ik} =$ 0.949 −0.006 0.173; 0 1.000 0; .704 −.020 .583; 0 0 1.000; .876 .032 .321; 1.000 0 0	$^2f''_{kj} =$ 1.000 0 0; 0 .251 .968; .212 .956 .201	$s_k =$ 1.000; 1.033; 1.046	$\hat{t}_i =$ 0.902; 1.033; .805; 1.046; .824; 1.000

[1] $\hat{t}_i = 1/\sum_k (a''^2_{ik}/s_k)$.

[2] From table 1*CC*.

previously defined. Equation 7 may now be written as

$$\sum_j \left(s_k f''_{kj} (xmax_j - xmin_j)\right) + \sum_j xmin_j = K. \quad (8)$$

The scale factors for converting the unscaled scores to scores that are conformable to the transformed data may be derived by rearrangement of equation 8 to

$$s_k = \frac{K - \sum_j xmin_j}{\sum_j \left(f''_{kj} (xmax_j - xmin_j)\right)}, \quad (9)$$

where, again, $xmax_j$ and $xmin_j$ are the maximum and minimum values in the jth column of the original data matrix if the data were transformed to proportions of the total range. If the transformation was to proportions of the maximum value, then all values of $xmin_j$ used in equation 9 are set equal to zero, and if no transformation was made, all values of $xmin_j$ are set equal to zero, and all values of $xmax_j$ are set equal to unity.

The Q-mode factor analysis model in equation 1 can be rewritten as

$$\hat{x}''_{ij} = \sum_k \frac{a''_{ik}}{s_k} f''_{kj} s_k, \quad (10)$$

or

$$\hat{x}''_{ij} = \sum_k a'_{ik} f'_{kj}, \quad (11)$$

where $a'_{ik} = a''_{ik}/s_k$ and f'_{ik} is a scaled factor score. Equation 11, like equation 1, may be used to derive an estimate of the normalized form of the original or transformed data matrix or may be used to derive the normalized form of the data exactly, for models where $m = M$. However, the model may be changed to one that reproduces the transformed data (or the original data if the data were not transformed) if both sides are divided by the quantity $\sum_\kappa a'_{ik}$ giving

$$\hat{x}''_{ij} \Big/ \sum_k a'_{ik} = \sum_k (a'_{ik} \Big/ \sum_k a'_{ik}) f'_{kj}. \quad (12)$$

Although not proven here, it has been observed and abundantly verified in the present study that when $m = M$ the quantity $1/\Sigma a'_{ik}$ is, except for round-off errors, precisely equal to the normalization factor, t_i, which is the square root of the sum of squares of the ith row in the original or transformed data matrix, depending on whether or not a transformation of the data was made. This can be seen by comparing the values of $\hat{t}_i$ in tables 2, 3, and 4 with the corresponding values of t_i in table 1. Therefore, the left side of equation 12 may be written simply as $\hat{x}'_{ij}$, which is the approximation of an element in the transformed data matrix or in the original data matrix if no transformation was made. The quantity $a'_{ik}/\Sigma a'_{ik}$ in equation 12 may be written as a_{ik} : a matrix analogous to that of a_{ik} is referred to as the composition matrix by Imbrie and VanAndel (1964, p. 1141) where it is derived

TABLE 5.—*Composition loadings,* a_{ik}, *and composition scores,* $f_{kj} = f'_{kj}$, *for the three-factor principal components, varimax, and oblique factor models derived without transforming the data*

[The products of the a_{ik} and f_{kj} matrices are equal to the original data in table 1A]

Model	Composition loadings, a_{ik}			Composition scores, $f_{kj} = f'_{kj}$		
Principal components	1.159	−0.122	−0.037	74.99	13.29	11.72
	.732	.304	−.037	−19.04	20.41	98.63
	1.054	−.069	.015	−22.49	159.92	−37.43
	.854	.044	.112			
	1.106	−.088	−.018			
	1.211	−.148	−.063			
Varimax	0.186	0.508	0.306	365.93	−115.35	−150.58
	.065	.758	.177	29.00	2.25	68.75
	.158	.484	.358	23.04	89.62	−12.66
	.100	.448	.452			
	.172	.508	.320			
	.201	.521	:279			
Oblique	0.856	−0.005	0.150	95	3	2
	0	1.000	0	50	10	40
	0.567	−.016	.449	60	30	10
	0	0	1.000			
	.722	.026	.253			
	1.000	0	0			

[1] From table 1A.

for an oblique model. The quantity a_{ik} will be referred to here as a composition loading whether it pertains to an oblique model, a varimax model, or a principal components model. Equation 12 then becomes

$$\hat{x}'_{ij} = \sum_k a_{ik} f'_{kj}, \qquad (13)$$

where $\hat{x}'_{ij}$ refers to the transformed data if a transformation was made, or to the original data if there was no transformation. The m values of a_{ik} sum to unity across k and may be interpreted as proportions. Table 5 gives values of a_{ik} and f'_{kj} for the principal components, varimax, and oblique models derived without transformation of the data. The products of the corresponding a_{ik} and f'_{kj} matrices are equal to the original data matrix in table 1*A*. The matrices of a_{ik} and f'_{kj} in table 6 were derived after transforming the data to proportions of the maximum values, and their products are equal to the transformed data matrix in table 1*B*. The matrices of a_{ik} and f'_{kj} in table 7 were derived after transforming the data to proportions of the total ranges for each variable, and their products are equal to the transformed data matrix in table 1*C*.

COMPOSITION SCORES

If the data were transformed prior to the derivation of the a_{ik} and f'_{kj} matrices, the scaled scores must be converted to composition scores in order to reproduce the original data, or an approximation of the original data, from the factor model. Each side of equation 13 is multiplied by the quantity $(xmax_j - xmin_j)$ and increased by $xmin_j$ to obtain

$$\hat{x}'_{ij}(xmax_j - xmin_j) + xmin_j = \sum_k (a_{ik} f'_{kj} (xmax_j - xmin_j)) + xmin_j. \qquad (14)$$

By virtue of the fact that $a_{ik} = a'_{ik} / \sum_k a'_{ik}$, it follows that $\sum_k a_{ik} = 1$.

Therefore, equation 14 may be written as

$$\hat{x}'_{ij}(xmax_j - xmin_j) + xmin_j = \sum_k (a_{ik} f'_{kj} (xmax_j - xmin_j)) + \sum_k a_{ik} xmin_j \qquad (15)$$

and

$$\hat{x}'_{ij}(xmax_j - xmin_j) + xmin_j = \sum_k a_{ik} (f'_{kj} (xmax_j - xmin_j) + xmin_j). \qquad (16)$$

The effect of the values of $xmax_j$ and $xmin_j$ in equation 16 is to reverse the transformation that consisted of scaling each variable to a proportion of the total range for the variable. Or, if all values of $xmin_j$ are set equal to zero, the effect is to reverse the transformation that consisted of scaling each variable to a proportion of the maximum

TABLE 6.—*Composition loadings,* a_{ik}, *scaled scores,* f'_{kj}, *and composition scores,* f_{kj}, *for the three-factor principal components, varimax, and oblique factor models derived after transforming each variable to proportions of the maximum value*

[The products of the a_{ik} and f'_{kj} matrices are equal to the transformed data in table 1*B*. The products of the a_{ik} and f_{kj} matrices are equal to the original data in table 1*A*]

Model	Composition loadings, a_{ik}	Scaled scores, f'_{kj}	Composition scores, f_{kj}
Principal components	a_{ik} = 0.998 −0.014 0.016; .876 .046 .078; 1.029 −.002 −.027; 1.083 .022 −.106; 1.005 −.008 .003; .982 −.020 .038	f'_{kj} = 0.836 0.395 0.219; −6.515 6.533 13.074; 1.223 −4.050 2.633	f_{kj} = 79.39 11.84 8.77; −618.92 195.98 522.94; 116.18 −121.51 105.33
Varimax	a_{ik} = 0.730 0.144 0.127; .254 .627 .119; .610 .170 .220; .370 .235 .394; .670 .170 .160; .790 .131 .080	f'_{kj} = 1.185 −0.155 −0.199; .301 .096 1.714; .309 2.624 −.202	f_{kj} = 112.62 −4.65 −7.97; 28.57 2.89 68.54; 29.34 78.73 −8.07
Oblique	a_{ik} = 0.856 −0.005 0.150; 0 1.000 0; 0.567 −.016 .449; 0 0 1.000; 0.722 .026 .253; 1.000 0 0	[1]f'_{kj} = 1.000 0.100 0.050; .526 .333 1.000; .632 1.000 .250	[2]f_{kj} = 95 3 2; 50 10 40; 60 30 10

[1] From table 1*B*.
[2] From table 1*A*.

TABLE 7.—*Composition loadings,* a_{ik}, *scaled scores,* f'_{kj}, *and composition scores,* f_{kj}, *for the three-factor principal components, varimax, and oblique factor models derived after transforming each variable to proportions of the range*

[The products of the a_{ik} and f'_{kj} matrices are equal to the transformed data in table 1*C*. The products of the a_{ik} and f_{kj} matrices are equal to the original data in table 1*A*]

Model	Composition loadings, a_{ik}	Scaled scores, f'_{kj}	Composition scores, f_{kj}
Principal components	a_{ik} = 1.096 −0.131 0.034; .304 .499 .197; 1.001 .044 −.045; .795 .391 −.186; 1.034 −.045 .011; 1.144 −.217 .074	f'_{kj} = 0.718 0.317 0.109; −.644 1.045 1.204; .522 −1.822 1.861	f_{kj} = 82.31 11.55 6.14; 21.02 31.23 47.75; 73.47 −46.20 72.72
Varimax	a_{ik} = 0.899 0.009 0.092; .023 .858 .119; .671 .072 .258; .216 .216 .567; .786 .061 .153; 1.012 −.021 .009	f'_{kj} = 0.987 −0.014 0.026; −.030 .060 1.177; .027 1.745 −.088	f_{kj} = 94.40 2.62 2.98; 48.65 4.62 46.73; 51.24 50.11 −1.35
Oblique	a_{ik} = 0.856 −0.005 0.150; 0 1.000 0; 0.567 −.016 .449; 0 0 1.000; .722 .026 .253; 1.000 0 0	[1]f'_{kj} = 1.000 0 0; 0 .259 1.000; .222 1.000 .211	[2]f_{kj} = 95 3 2; 50 10 40; 60 30 10

[1] From table 1*C*.
[2] From table 1*A*.

value for the variable. In either case, because of equations 4 and 6, we can now write:

$$\hat{x}_{ij} = \sum_k a_{ik} f_{kj}, \qquad (17)$$

where $\hat{x}_{ij}$ refers to the original data in table 1*A*, the a_{ik}'s are composition loadings, and the f_{kj}'s are composition scores for the principal-components, varimax, or oblique factor models.

The composition scores are in units of the original data and sum to K across j (eq 3a). For the oblique model developed by Imbrie (1963), the composition scores, when $m = M$, are simply the compositions of the samples selected to serve as end-members in the factor solution. For the principle component and varimax models, the derived composition scores for the kth factor may be regarded as the composition of a theoretical sample that may serve as an end-member for the corresponding model. However, if any value of f_{kj} is negative or in excess of K, the derived composition of the kth factor is, of course, unrealistic and should be viewed as only an approximate indication of the nature of some theoretical end-member sample.

The composition scores, f_{kj}, derived from the data in table 1 are given in tables 5, 6, and 7 for the principal components, varimax, and oblique models. Those in table 5 were derived without transformation of the data. Those in table 6 were derived after the data were transformed to proportions of the maximum values, and those in table 7 were derived after the data were transformed to proportions of the total range for each variable. The matrices of composition loadings, a_{ik}, in tables 5, 6, and 7 may be multiplied by the corresponding matrices of composition scores, f_{kj}, to produce the matrix of original data in table 1*A*.

NEGATIVE LOADINGS AND NEGATIVE SCORES

In most factor analysis applications the signs of the factor loadings are regarded as arbitrary, and it is entirely valid to change all the signs in any column of the initial loading matrix. However, if factor scores are computed and the signs are changed in the kth column of the loading matrix, it is necessary to also change the signs in the kth row of the score matrix. When these sign changes are made for the composition loadings and scores, the scores will sum to minus K, rather than K, and the requirement in equation 3a is violated. This situation is not corrected by changing the signs in the initial loadings and unscaled scores from which the composition loadings and scores were derived. When the sign changes are made for the initial loadings and unscaled scores, the scale factor, s_k, also changes in sign, as may be seen from equation 9. Because of this, the sign change is cancelled out (eq 10) before the composition loadings and scores are derived. Therefore, unlike the conventional (initial) loadings, the composition loadings and scores have fixed signs that are determined by the choice of reference axes and the nature of the compositional variation in the rocks under examination.

In the use of Q-mode factor analysis to develop a genetic model, the composition scores for the selected reference vectors should be entirely nonnegative because these scores represent chemical or mineralogic compositions of end-members. However, if the factor solution is to be used as a device for summarizing geochemical or petrologic data or for purposes of sample classification, negative composition score values can be perfectly acceptable. I have found that the composition scores for the conventional varimax reference axes are commonly negative in part but that the varimax axes can still be useful for reference purposes and the set of scores for each axis is still indicative of the general compositional nature of the theoretical end-member. For example, a series of scores consisting of some value larger than the constant K for SiO_2, and of negative values for all other constituents indicates that a more realistic end-member might have the composition of a siliceous magma or a pure quartz sand.

In situations where negative composition scores are unacceptable, they can be avoided by the choice of a different set of reference axes. The method for finding the composition scores for any vector that may serve as a potential reference axis within the factor space of the Q-mode solution is demonstrated in the following section. After m reference axes with all nonnegative composition scores have been selected, there will be no difficulty in determining the composition loadings for each sample with respect to them as long as the selected axes are sufficiently independent. (That is, they are not colinear or approximately colinear in a two-factor space, coplanar in a three-factor space, and so forth.)

The composition loadings, a_{ik}, always sum to unity across k and their signs indicate the required addition or subtraction of the kth end-member to form the ith sample according to the model. Where either the addition of an end-member, as indicated by a positive loading, or the subtraction of an end-member, as indicated by a negative loading, is unreasonable, a new choice of reference axes may be desirable.

Composition scores for the first axis of a principal component model tend to be all nonnegative, as do the composition loadings on the first end-member. The scores and loadings for subsequent axes tend to be positive and negative in about equal number.

Although some of the composition scores for varimax models are commonly negative, the composition loadings are generally almost entirely positive. As the varimax axes are changed to oblique axes by moving them toward the vectors representing the actual samples, their composition scores become increasingly positive and the composition loadings become increasingly negative. The first of these properties can be useful in the modification of varimax models to form models with oblique reference axes that have all nonnegative composition scores, as described in part II.

Oblique models that employ the extreme sample vectors as reference axes, as used by Imbrie (1963), have all nonnegative composition scores if $m = M$, or if for some

other reason the model accounts for all the variability in the data. Where this is not true, the composition scores differ from the actual compositions of the reference samples and some may be negative. The composition loadings for this type of oblique model can be entirely nonnegative, but commonly some of them are slightly negative.

The signs of the composition scores and loadings for other types of oblique models depend entirely on the reference axes chosen to describe a particular set of sample data. These other types of oblique models may be developed by choosing alternative oblique reference axes. The choice of reference axes may be based either on examinations of the composition scores for selected vectors or on the determination of the vector representations for given compositions.

FINDING THE COMPOSITION SCORES FOR ANY VECTOR IN THE FACTOR SPACE

The oblique model developed by Imbrie (1963) is one wherein actual samples of extreme composition serve as end-members, or reference samples, for describing the compositional variations within the entire sample series. In many studies, however, there is no reason to expect that the actual end-members which contributed to the formation of a rock body are represented in the sample data. Thus, the extreme samples might be rejected as plausible end-members, and a search begun for alternative real or hypothetical end-members that have acceptable composition scores as well as acceptable composition loadings of all samples with respect to them. It is an easy matter to determine the composition scores for any vector that may potentially serve as a reference axis.

As an example, we may consider an arbitrary vector within the three-factor space of the varimax solution represented in tables 4 and 7. The initial loadings of the vector on the varimax axes are taken as

$$a''_{i1} = 0.03162,\ a''_{i2} = 0.03162,\ a''_{i3} = 0.9990. \quad (18)$$

These loadings (which define a vector close to but not coincident with the third varimax axis) and the unscaled varimax factor scores from table 4 are used to derive the normalized composition of the vector according to equation 1. Because the normalized composition is to be regarded as a set of unscaled scores, the notation is appropriately changed from $\hat{x}''_{ij}$ to f''_{kj}. The derived values are

$$f''_{k1} = 0.047,\ f''_{k2} = 0.999,\ f''_{k3} = -0.018. \quad (19)$$

Equation 9, then, gives $s_k = 1.586$, and the scaled factor scores for the vector are

$$f'_{k1} = 0.074,\ f'_{k2} = 1.584,\ f'_{k3} = -0.029, \quad (20)$$

and, from equation 6, the composition scores for the vector are

$$f_{k1} = 53.33,\ f_{k2} = 45.77,\ f_{k3} = 0.90. \quad (21)$$

Because none of the composition scores is negative, the vector may serve as a more acceptable reference axis than the third varimax axis which has a negative value for f_{33} (table 7).

Composition scores from some other arbitrary vectors within the same factor space are given in table 8. Note that the scores for colinear vectors having loadings that are opposite in sign are exactly the same. Note also that some vectors, such as vectors 3 and 4, may have composition score values that are negative or in excess of K and impossible to interpret in terms of rock composition.

TABLE 8.—*Computed composition scores, $\mathbf{f}_{ij}$, for some arbitrary vectors in the varimax solution of tables 4 and 7*

Vector, i	Loadings, a''_{ik}			Composition scores, f_{ij}		
	$k=1$	$k=2$	$k=3$	$j=1$	$j=2$	$j=3$
1	0.577	0.577	0.577	68.31	14.49	17.20
2	-0.577	-.577	-.577	68.31	14.49	17.20
3	.271	-.924	.271	12.06	-13.77	101.70
4	-.271	.924	-.271	12.06	-13.77	101.70
5	.924	.271	.271	81.38	8.55	10.08
6	.271	-.924	-.271	30.38	16.07	53.55
7	.382	.654	.654	62.85	16.97	20.18

FINDING THE VECTOR REPRESENTATION FOR A GIVEN COMPOSITION

Although the search for possible end-members in a compositional system may consist of the determination of composition scores for various vectors in the factor space, as described in the previous section, it is commonly more satisfactory to begin with the composition, z_j, of some material that may have been involved in the system and determine whether or not the composition can be satisfactorily represented as a vector in the m-dimensional factor space. Any composition that can be represented in the factor space can be arrived at by the mixing or unmixing of compositions equal to the composition scores of any other m vectors in the same space. Any composition that cannot be represented in the factor space could not have been an end-member in the m-component compositional system.

The procedure cannot be fully illustrated here because the example factor solutions have $m = M$, and any composition of the same M components can be completely represented in the m-dimensional space. However, the procedure begins with the transformation and normalization of the given composition in the same manner used to transform the original data before derivation of the

factor solution, using the same values of $xmax_j$ and $xmin_j$ (table 1). Representing the transformed and normalized composition by a row of z_j'', the relation given in equation 1 can be written as

$$z_j'' = \sum_k a_k'' f_{kj}''. \quad (22)$$

Equation 22 in matrix notation is

$$Z = AF. \quad (23)$$

Following Klovan and Imbrie (1971, p. 64), both sides of equation 23 are postmultiplied by the transpose of the unscaled score matrix to give

$$ZF' = AFF'. \quad (24)$$

However, the unscaled scores are in normalized form, and, if they are for orthogonal reference axes, such as the principal component or varimax axes, the product of F and F' is an identity matrix. Consequently, equation 24 becomes

$$A = ZF'. \quad (25)$$

This translates to

$$a_k'' = \sum_j z_j'' f_{jk}'', \quad (26)$$

which shows that the initial loadings on the principal components or varimax axes are derived simply by multiplying the array of normalized composition values by the transposed matrix of unscaled scores.

The sum of squares of the m values of a_k'' is the communality h^2 of the vector representing the given composition and indicates how well the composition fits into the factor space of the sample data and the degree to which the given composition in original units (generally percent or parts per million) may be reproduced from the factor model. To reproduce the original composition, the derived values of a_k'' are divided by the scale factors, s_k, for the principal components or varimax axes, and adjusted to sum to unity across k, giving a_k. The original composition, z_j, is then estimated by

$$\hat{z}_j = \sum_k a_k f_{kj}, \quad (27)$$

where the values of f_{kj} are the composition scores for the principal components or varimax axes. Wherever h^2 is unity, the quantities $(\hat{z}-z)_j$ will be zero for each variable, and, where h^2 is less than unity, the absolute values of $(\hat{z}-z)_j$ will be correspondingly larger.

The vector representations of some arbitrarily selected compositions in the varimax model of tables 4 and 7 are given in table 9. Because $m = M$ in this model, the value of h^2 is unity for each composition.

In real applications of the method, where m is less than M, the acceptability of the given composition z_j, as that of a possible end-member in a petrologic system will depend on its derived communality, h^2.

TABLE 9.—*Vector loadings, a_{ik}'', in the varimax solution of tables 4 and 7 for some given compositions, z_{ij}*

Phase, i	Composition, z_{ij}			Vector loadings, a_{ik}''		
	$j=1$	$j=2$	$j=3$	$k=1$	$k=2$	$k=3$
1	63	24	13	0.325	0.366	0.873
2	75	3	22	.744	.668	−.023
3	99	0	1	.995	−.054	−.084
4	0	82	18	−.361	.189	.913
5	12	5	83	−.344	.939	−.020
6	14	76	10	−.294	.130	.947
7	6	10	84	−.388	.920	.057

COMPUTATION OF INITIAL AND COMPOSITION LOADINGS WITH RESPECT TO OBLIQUE REFERENCE VECTORS

Determination of the initial factor loadings with respect to any oblique reference axes follows the technique described by Imbrie (1963). It is illustrated here using the varimax model represented in tables 4 and 7. For the purpose of illustration only, the oblique reference axes will be taken as vectors 5, 6, and 7 in table 8; the loading matrix of the new reference axis system is

$$A = a_{ik}'' = \begin{matrix} 0.924 & 0.271 & 0.271 \\ .271 & -.924 & -.271 \\ .382 & .654 & .654. \end{matrix} \quad (28)$$

The inverse of A is

$$A^{-1} = \begin{matrix} 1.306 & 0.0 & -0.541 \\ .859 & -1.531 & -.990 \\ -1.621 & 1.531 & 2.836. \end{matrix} \quad (28a)$$

The product of the matrix of initial varimax loadings in table 4 and the inverse matrix in 28a gives the matrix of initial loadings on the new reference axes

$$a_{ik}'' = \begin{matrix} 1.007 & 0.254 & -0.039 \\ .542 & -1.190 & -.408 \\ .258 & .695 & 1.037 \\ -1.061 & 1.078 & 2.337 \\ .781 & .361 & .322 \\ 1.256 & .064 & -.469. \end{matrix} \quad (29)$$

The matrix of unscaled factor scores, f_{kj}'', for the new reference axes is then determined as the product of the matrix of initial loadings, a_{ik}'', in 28 and the matrix of unscaled varimax scores, f_{kj}'', in table 4. The scale factors, s_k, for the new reference axes are then computed, using the new scores and the values of $xmax_j$ and $xmin_j$ from table 1 in equation 9. The scale factors are

$$s_1 = 0.7562, \quad s_2 = -1.5040, \quad s_3 = 0.7595. \quad (30)$$

The composition loadings for the new model are derived by dividing the columns of 29 by the corresponding values of

s_k in 30 and then adjusting each row to sum to unity. The composition loadings are

$$a_{ik} = \begin{matrix} 1.199 & -0.152 & -0.047 \\ .739 & .816 & -.556 \\ .274 & -.372 & 1.098 \\ -1.465 & -.749 & 3.214 \\ .849 & -.198 & .349 \\ 1.661 & -.043 & -.619. \end{matrix} \quad (31)$$

The original data in table 1*A* can be reproduced exactly, except for differences due to round off, by multiplying the composition loading matrix in 31 by the matrix of composition scores for vectors 5, 6, and 7 given in table 8.

SUMMARY

Q-mode factor analysis appears to be a method that is well suited for exploring the nature of the compositional variation in a suite of rock samples. It can be used in at least three ways: (1) to develop a geochemical or petrologic model, (2) to develop a scheme of classification, or (3) to summarize complex multivariate data. Applications of the second and third types can employ any set of reference axes, regardless of their composition scores. The development of tenable geochemical or petrologic models, however, requires that all composition scores be in the range from zero to K, and the acceptability of the model may depend on the consistency of the composition loadings, especially their signs, with other geologic observations. These conditions will usually require experimentation using various sets of reference vectors. Potential reference vectors can be examined by either finding the composition scores for selected vectors or finding whether or not certain compositions of interest can be satisfactorily represented as vectors in the factor space.

The ultimate test of the mathematical validity of any factor model is the closeness with which it can reproduce the original data. The only test of its geologic validity is whether or not it conforms to other geologic observations regarding the rock unit, or units, from which the series of rock samples was taken.

All the models developed here can be used to reproduce the normalized, transformed, or original data exactly because the number of factors they contain was made equal to the number of variables represented in the data matrix. In most real applications, m is less than M because the purpose of the analysis is to reduce the problem, and it cannot be expected that the data will be reproduced exactly. However, the closeness with which the data can be reproduced may or may not be satisfactory to the investigator, and the test is recommended.

Part II. Effects of Reducing the Number of Factors

INTRODUCTION

Part I has shown that the constant row-sum of matrices of compositional data on rocks and other materials can be used to advantage in *Q*-mode factor analysis. Where the matrix row-sums are constant, it is possible to scale the factor loadings and scores, which together comprise the factor model, so that they are conformable with the original data, which are commonly in units of weight percent or parts per million. Without such scaling, the loadings are difficult to interpret in terms of proportions of end-members, and the scores cannot easily be interpreted in terms of rock compositions. Moreover, the products of the matrices of loadings and scores are not approximations of the original data in percent or parts per million but of the data in their normalized form. If the loadings are scaled and adjusted to sum to unity for each sample, and, if the scores are scaled and detransformed (if a data transformation was used in their derivation), the products of the matrices of loadings and scores approach or equal the matrix of original data.

Examples of factor models were given in part I for which the number of factors was equal to the number of variables (columns) in the data matrix, and the products of the matrices of composition loadings and composition scores equaled the original data exactly, except for differences due to round off. The purpose here in part II is to show how the method and the mathematical basis are affected when the number of factors is reduced to something less than the number of variables. Thus, the computations and the models to be developed are more representative of most real factor analysis applications wherein the objective is to reduce, or simplify, the problem by developing a less complex model that explains some major part of the compositional variability in a suite of rock samples. The minor part of the variability left unexplained is ascribed to measurement errors and to the effects of geologic processes that have had only minor effects on variation in rock composition. The same hypothetical data used in part I (table 1) will be used here, but the models will be developed with two factors rather than three. The methods will be applied to real data in part III.

The initial factor loadings and unscaled scores for the principal components and varimax models developed here, as those in part I, were derived by the method of Klovan and Imbrie (1971).

ESTIMATES OF FACTOR LOADINGS

The geometric representation of a *Q*-mode factor model is a system of vectors with a common origin. Where the number of factors, m, is equal to the number of variables, M, the vectors are of unit length, and each represents a row of either the normalized original data matrix or the normalized matrix of transformed data where a transformation was used. The cosine of the angle between any two of the vectors is exactly equal to the coefficient of proportional similarity (cosine theta) for the two samples that the vectors represent. Because the minimum value of cosine theta for nonnegative data is zero, none of the sample vectors is separated by more than 90 degrees from the others. The positions of the vectors may be described in terms of any reference axis system, but the conventional systems are the principal components axes, the varimax axes, and the oblique axes as used by Imbrie (1963).

The principal components axes are positioned in the vector system so that the sum of squares of the projections of the sample vectors on one of the axes is as high as possible. In other words, one of the principal axes is placed near the center of the vector cluster, and all the sample vectors have positive projections on it. The second principal axis is positioned orthogonal to the first but is oriented so that the sample vectors have as large a projection on it as possible considering the orthogonality constraint. Subsequent axes are positioned in the M-dimensional space so that they are orthogonal to the preceding axes and so that the projections of the sample vectors on each of them are as large as possible. Consequently, the first column of the initial factor loading matrix for the principal-components model, which gives the loadings on the first axis, contains relatively large positive values, and subsequent columns contain both positive and negative values of decreasing absolute magnitude (tables 2, 3, and 4).

The sum of the squares of the kth column of the matrix of initial loadings on the principal components axes is equal to the kth eigenvalue of the cosine theta matrix (Harman, 1967, p. 167). Consequently, the eigenvalues indicate the number of dimensions occupied by the cluster of sample vectors within the M-dimensional space. For example, if the first two eigenvalues of the cosine theta matrix are large and subsequent eigenvalues are small, the sample vectors cluster about a plane. This has been taken to indicate that a major part of the compositional variation can be described in terms of two end-members and as justification for reducing the number of factors, m, to two so that m is less than M. Interpretation of the eigenvalues is discussed further in the following section.

In the derivation of the varimax model, where $m = M$, the principal components axes are rotated simultaneously, while maintaining their mutual orthogonality, so that the total sum of the squares of the projections of the sample vectors on all the axes is as large as possible. This is the varimax criterion. In the derivation of the oblique model, where $m = M$, using the method of Imbrie (1963), the reference axes are taken as the M sample vectors that occur at the extremes of the sample vector cluster and, if possible, enclose all the others.

If the eigenvalues of the cosine theta matrix indicate that the cluster of sample vectors occurs largely in m dimensions, where m is less than M, only the first m-principal-components axes are rotated in deriving the varimax model, and the sample vectors are then projected into m-dimensional space. Following this projection, the sample vectors tend to be of less than unit length. The squares of the vector lengths, computed as the sums of the squares of the initial loadings, are referred to as the sample communalities and indicate the degree to which each sample conforms with the final factor model.

These procedures are illustrated here using the same hypothetical data (table 1) used in part I, wherein all models were such that the number of factors equalled the number of variables. As before, the data matrix is treated in three ways: (1) the original data (table 1*A*), (2) the data transformed so that each variable is expressed as a proportion of the maximum value for that variable (table 1*B*), and (3) the data transformed so that each variable is expressed as a proportion of the total range for that variable (table 1*C*). The row-normalized forms of the data are given, respectively, in tables 1*AA*, 1*BB*, and 1*CC*. The sample vector cluster previously defined for the principal components model is projected into two-dimensional space merely by eliminating the third column of the initial factor loading matrix. (Compare tables 10, 11, and 12 with tables 2, 3, and 4 of part I.) The third row of each unscaled factor score matrix is also eliminated. Only two of the principal components axes are rotated in deriving the varimax model. The projections of the sample vectors on the two varimax axes are given in tables 10, 11, and 12. Computation of the initial loadings and unscaled scores for both the principal components and the varimax models follows the methods given by Klovan and Imbrie (1971).

According to the method of deriving the oblique factor model described by Imbrie (1963), the m end-member samples are first chosen as those that have the largest projections on each of the varimax reference axes. As shown in part I, the m by m matrix of initial varimax loadings, A, is then inverted to give A^{-1}, and the entire N by m matrix of initial loadings postmultiplied by A^{-1} gives the

TABLE 10.—*Initial loadings, a''_{ik}, unscaled scores, f''_{kj}, scale factors, s_k, estimated normalization factors, $\hat{t}_i$, and reproduced data matrix, $\hat{x}_{ij}$, for the two-factor principal components, varimax, and oblique factor models derived without transforming the data*

[The products of the matrices of a''_{ik} and f''_{kj} are approximations of the normalized data in table 1*AA*]

Model	Initial loadings, a''_{ik}	Unscaled scores, f''_{kj}	Scale factors, s_k	Estimated normalization factor[1], $\hat{t}_i$
Principal components[2]	a''_{ik} = 0.988 −0.138; .871 .481; .996 −.087; .960 .066; .994 −.105; .981 −.159	f''_{kj} = 0.973 0.172 0.152; −.186 .199 .962	s_k = 77.1; 102.5	$\hat{t}_i$ = 87.2; 62.5; 82.9; 76.4; 84.3; 89.5
Varimax[2]	a''_{ik} = 0.891 0.448; .443 .891; .869 .495; .752 .600; .877 .479; .898 .427	f''_{kj} = 0.906 0.029 −0.422; .400 .262 .878	s_k = 194.8; 64.9	$\hat{t}_i$ = 87.1; 62.5; 82.7; 76.3; 84.2; 89.4
Oblique	a''_{ik} = 0.975 0.036; 0 1.000; .908 .120; .663 .355; .933 .091; 1.000 0	[3]f''_{kj} = 0.98 0.14 0.00; .76 .25 .60	s_k = 89.3; 62.1	$\hat{t}_i$ = 87.0; 62.1; 82.6; 76.1; 83.9; 89.3

Reproduced data matrix (product of matrices of a''_{ik} and f''_{kj})

$\hat{x}''_{ij}$ = 0.99 0.14 0.02; .76 .25 .60; .99 .15 .07; .92 .18 .21; .99 .15 .05; .98 .14 .00

[1] $\hat{t}_i = 1/\sum_k (a''_{ik}/s_k)$.

[2] The communalities ($h_i^2 = \sum_k a''^2_{ik}$) are 1.00, 0.99, 1.00, 0.93, 1.00, and 0.99.

[3] From reproduced data matrix.

initial factor loading matrix for the oblique model. If any of these oblique loadings exceeds unity, an iterative procedure is begun to select different end-member samples for which all initial oblique loadings are unity or less.

A different iterative procedure has been used to derive the initial oblique loadings in tables 10, 11, and 12. The criterion in the iteration is that none of the oblique composition loadings, rather than the initial loadings, exceeds unity. The unscaled scores for the oblique models are taken directly from the reproduced data matrices (tables 10, 11, 12). The alternative procedure is discussed in a subsequent section dealing with the identification of samples of extreme composition.

INTERPRETATION OF EIGENVALUES AND COMMUNALITIES

The first three eigenvalues, λ, of the 6 by 6 matrices of cosine theta derived from the original and transformed data matrices (tables 1*A*, 1*B*, and 1*C*) and their cumulative proportions of *N*, here after abbreviated as CPN, are given in table 13. The values of CPN indicate the proportions of the total sums of squares in the normalized data matrices that will be explained by principal components, varimax, or oblique models having the corresponding numbers of factors. It was shown in part I that three-factor models account for all the variability in the data matrices. That is, the final factor models could be used to reproduce the original, transformed, and normalized data matrices exactly, except for differences due to round off. The values of CPN in table 13 show that two-factor models will explain 98 percent, 94 percent, or 91 percent of the variability (as a sum of squares) in the normalized data matrices, depending on whether or not a data transformation was used and on which type of transformation it was. For example, when the initial factor loadings for any of the three models represented in table 10 are multiplied by the corresponding matrices of unscaled scores, the product is an estimate of the normalized original data in table 1*AA*, referred to as the reproduced data matrix. The total sum of squares in the reproduced data matrix of table 10 is 98 percent of $N = 6$, the total

TABLE 11.—*Initial loadings,* a''_{ik}, *unscaled scores,* f''_{kj}, *scale factors,* s_k, *estimated normalization factors,* $\hat{t}_i$, *and reproduced data matrix,* $\hat{x}''_{ij}$, *for the two-factor principal components, varimax, and oblique factor models derived after transforming each variable to proportions of the maximum value*

[The products of the matrices of a''_{ik} and f''_{kj} are approximations of the normalized transformed data in table 1*BB*]

	Initial loadings, a''_{ik}	Unscaled scores, f''_{kj}	Scale factors, s_k	Estimated normalization factors[1], $\hat{t}_i$
Principal components[2]	a''_{ik} = 0.969 −0.234; .706 .627; .990 −.037; .851 .294; .991 −.131; .927 −.324	f''_{kj} = 0.880 0.416 0.231; −.407 .408 .817	s_k = 0.950; 16.001	$\hat{t}_i$ = 0.99; 1.28; .96; 1.09; .97; 1.05
Varimax[2]	a''_{ik} = 0.936 0.341; .241 .913; .845 .517; .546 .716; .898 .440; .951 .244	f''_{kj} = 0.958 0.120 −0.260; .148 .570 .808	s_k = 1.188; 1.575	$\hat{t}_i$ = 1.00; 1.28; .96; 1.09; .97; 1.05
Oblique	a''_{ik} = 0.954 0.119; 0 1.000; .799 .353; .403 .677; .882 .246; 1.000 0	[3]f''_{kj} = 0.95 0.25 −0.05; .37 .55 .68	s_k = 1.044; 1.268	$\hat{t}_i$ = 0.99; 1.27; .96; 1.09; .96; 1.04

Reproduced data matrix (products of matrices of a''_{ik} and f''_{kj})

$\hat{x}''_{ij}$ =

0.95	0.31	0.03
.37	.55	.68
.89	.40	.20
.63	.47	.44
.93	.36	.12
.95	.25	−.05

[1] $\hat{t}_i = 1/\sum_k (a''_{ik}/s_k)$.

[2] The communalities ($h_i^2 = \sum_k a''^2_{ik}$) are 0.99, 0.89, 0.98, 0.81, 1.00, and 0.96.

[3] From reproduced data matrix.

sum of squares in the normalized data of table 1*AA*. Similarly, the total sum of squares in the reproduced data matrix of table 11 is 94 percent of that in the normalized data of table 1*BB*, and the total sum of squares in the reproduced data matrix of table 12 is 91 percent of that in the normalized data of table 1*CC*. For those who check these assertions by computation, the minor differences that will be observed are due solely to round off. Thus, the eigenvalues of the cosine theta matrix indicate the overall degree to which principal components, varimax, and oblique *Q*-mode models will succeed in reproducing the normalized forms of the original data.

The communalities of the samples, as previously mentioned, are computed as the sums of squares of the initial loadings on the *m* principal components or varimax axes (tables 10, 11, 12). The sample communalities are also equal to the sums of squares of the corresponding rows in the data matrices reproduced from the initial loadings and unscaled scores (tables 10, 11, 12). Thus, the communalities, like the eigenvalues of the cosine theta matrices, indicate the degree to which the models will succeed in reproducing the original normalized data. In fact, the average sample communality is equal to the corresponding value of CPN.

It will be shown in a following section on goodness-of-fit measures that the eigenvalues and sample communalities can be misleading where they are used as indicators of the degree to which the factor model can be used to reproduce the original data.

SCALING THE FACTOR LOADINGS AND SCORES

The appropriate factors, s_k, for scaling the initial factor loadings and scores to conform with the original data were computed from equation 9 of part I and are given in tables 10, 11, and 12.The composition loadings, a'_{ik}, derived by dividing the initial loadings, a''_{ik}, by the corresponding scale factors to give a'_{ik} and adjusting the values of a'_{ik} so that they sum to unity across k, are given in tables 14, 15,

TABLE 12.—*Initial loadings,* a''_{ik}, *unscaled scores,* f''_{kj}, *scale factors,* s_k, *estimated normalization factors,* $\hat{t}_i$, *and reproduced data matrix,* $\hat{x}_{ij}$, *for the two-factor principal components, varimax, and oblique factor models derived after transforming each variable to proportions of the range*

[The products of the matrices of a''_{ik} and f''_{kj} are approximations of the normalized transformed data in table 1CC]

Model	Initial loadings, a''_{ik}	Unscaled scores, f''_{kj}	Scale factors, s_k	Estimated normalization factors,[1] $\hat{t}_i$
Principal components[2]	$a''_{ik} =$ 0.963 −0.249; .234 .830; .985 .094; .603 .643; .995 −.095; .906 −.375	$f''_{kj} =$ 0.906 0.400 0.138; −.374 .608 .700	$s_k =$ 0.792; 1.720	$\hat{t}_i =$ 0.93; 1.28; .77; .88; .83; 1.08
Varimax[2]	$a''_{ik} =$ 0.991 0.083; −.054 .861; .898 .414; .357 .806; .970 .239; .979 −.054	$f''_{kj} =$ 0.979 0.177 −0.101; −.054 .706 .706	$s_k =$ 1.000; 1.035	$\hat{t}_i =$ 0.93; 1.29; .75; .87; .83; 1.08
Oblique	$a''_{ik} =$ 1.021 0.161; 0 1.000; .947 .541; .418 .963; 1.010 .341; 1.000 0	$^3f''_{kj} =$ 0.96 0.13 −0.14; −.10 .60 .61	$s_k =$ 1.087; 1.290	$\hat{t}_i =$ 0.94; 1.29; .77; .88; .84; 1.09

Reproduced data matrix (product of matrices of a''_{ik} and f''_{kj})

$\hat{x}'_{ij} =$ 0.97 0.23 −0.04; −.10 .60 .61; .86 .45 .20; .31 .63 .53; .94 .34 .07; .96 .13 −.14

[1] $\hat{t}_i = 1/\sum_k (a''_{ik}/s_k)$.

[2] The communalities ($h_i^2 = \sum_k a''^2_{ik}$) are 0.99, 0.74, 0.98, 0.78, 1.00, and 0.96.

[3] From reproduced data matrix.

and 16. The scaled scores, f'_{kj}, derived simply by multiplying the unscaled scores, f''_{kj}, by the scale factors, are also given in tables 14, 15, and 16. The products of the matrices of a_{ik} and f'_{kj} are approximations of the original data where no data transformation was made,[1] or of the transformed data where transformations were used.[2] Where a transformation was not used (table 14), the scaled scores are equal to the composition scores. Where the data were transformed to proportions of the maximum values for each variable (table 15), the composition scores are derived by multiplying the scaled scores by the corresponding values of $xmax_j$ from table 1. Where the data were transformed to proportions of the ranges for each variable (table 16), the composition scores are derived by multiplying the scaled scores by the quantity $xmax_j - xmin_j$ and adding $xmin_j$.

[1] Compare tables 1A and 14.

[2] Compare tables 1B and 15, and 1C and 16.

TABLE 13.—*Eigenvalues of the matrices of coefficients of proportional similarity (cosine theta) derived from the data matrices in table 1, and CPN*[1]

Data not transformed table 1A		Data transformed to proportion of maximum values, table 1B		Data transformed to proportion of ranges, table 1C	
λ	CPN	λ	CPN	λ	CPN
5.5986	0.93	4.9840	0.83	4.1256	0.69
.2987	.98	.6577	.94	1.3225	.91
.1027	1.00	.3583	1.00	.5518	1.00

[1] Cumulative proportions of *N*.

GOODNESS-OF-FIT MEASURES FOR THE FACTOR MODEL

Because the number of factors, *m*, in the models represented in tables 10–12 and 14–16 is less than the number of variables, *M*, less then 100 percent of the variability in the data will be explained by the models, as

TABLE 14.—*Composition loadings,* a_{ik}, *composition scores,* $f_{kj} = f'_{kj}$, *and reproduced data matrix,* $\hat{x}_{ij}$, *for the two-factor principal components, varimax, and oblique factor models derived without transforming the data*

[The products of the matrices of a_{ik} and f_{kj} are approximations of the original data in table 1A]

Model	Composition loadings, a_{ik}	Composition scores, $f_{kj} = f'_{kj}$
Principal components	$a_{ik} =$ 1.117 −0.117; .707 .293; 1.070 −.070; .951 .049; 1.086 −.086; 1.139 −.139	$f_{kj} = f'_{kj} =$ 75.0 13.3 11.7; −19.1 20.4 98.6
Varimax ..	$a_{ik} =$ 0.399 0.601; .142 .858; .369 .631; .295 .705; .379 .621; .412 .588	$f_{kj} = f'_{kj} =$ 176.5 5.7 −82.2; 26.0 17.0 57.0
Oblique...	$a_{ik} =$ 0.950 0.050; 0 1.000; .840 .160; .565 .435; .877 .123; 1.000 0	$f_{kj} = f'_{kj} =$ 88 12 0; 47 15 37

Reproduced data matrix (product of matrices of a_{ik} and f_{kj})

				Row-sum
$\hat{x}_{ij} =$	86	12	2	100
	47	15	37	99
	82	13	6	101
	70	14	16	100
	83	13	4	100
	88	12	0	100

the eigenvalues in table 13 indicate. Consequently, the data matrices reproduced from the composition loadings and composition scores (tables 14, 15, 16) are not equal to the original data in table 1*A*. Appropriate comparisons are made as shown in table 17.

The values of d_j^* and d_j^{**} in table 17 are, respectively, the mean and standard deviation of the factor-model residuals, and the values of r_j and r_j^2 are, respectively, the correlations and coefficients of determination for the reproduced and original data. If the model explains all the data perfectly, all values of d_j^* and d_j^{**} are zero, and all values of r_j and r_j^2 are unity. In real applications of the method, the values of d_j^* are found to be small in comparison to the original data values, indicating that the models are unbiased. The coefficients of determination have been found to be particularly informative in that they show the proportions of the total variance in each variable explained by the model. That is

$$r_j^2 \approx \frac{s(x)_j^2 - (d_j^{**})^2}{s(x)_j^2}, \qquad (32)$$

where $s(x)_j^2$ is the variance in the jth column of the original data matrix.

The values of r_j^2 in table 17 show that none of the factor models explains much of the variance in variable 2. This shortcoming of the two-factor models is not evident from the eigenvalues given in table 13, and the indication is clear that the eigenvalues alone can be misleading measures of the mathematical adequacy of a factor model. It will be shown in the following section and in part III that diagrams—factor-variance diagrams—showing the proportion of the total variance in each variable, r_j^2, accounted for by factor models with up to M reference axes, are useful devices for selecting the number of end-members that the model should include.

The reason that the original data cannot be reproduced exactly by factor models that have $m < M$ is related to the equality $t_i = 1/\sum_k a'_{ik}$, where t_i is the quantity that was used to convert the ith row of the original or transformed data to the normalized form. Where less than the total variability in the data is explained by the model (that is, where the number of factors, m, is less than the rank of the matrix of $\cos\theta$), this equality does not hold true. However, estimates of the normalization factors will approach the correct values as the proportion of the variability explained by the model approaches one, just as the reproduced data matrices approach the matrix of original data. The estimates of the normalization factors, $\hat{t}_i$, are given in tables 10, 11, and 12 for comparison with the corresponding values of t_i in table 1.

AN EXAMPLE OF THE EFFECT OF THE CONSTANT ROW-SUM CONSTRAINT

Inspection of factor-variance diagrams in real applications shows that as the number of factors, m, in the model is increased, the model explains somewhat greater portions of the variance in all, or nearly all, of the compositional variables. Generally, however, as will be demonstrated in part III, the addition of a factor to the model will increase the variance accounted for in a few of the variables to a greater degree than it will increase the accountable variance in others. The particular variables that respond to the same factor depend, at least in part, on the correlations among the variables, and this raises the question of the effect of the constant row-sum constraint that was described by Chayes (1960). Chayes showed that if the row-sums of the data matrix are constant across all rows, each variable must show some negative correlation with at least one of the others, even if the variables have no genetic relationships whatsoever with each other. The question here is the degree to which the constant row-sum constraint affects the configuration of the factor-variance diagrams. The question cannot be fully answered at this time, but it is possible to show the nature of a diagram, derived from simulated data, wherein nearly all the intercolumn correlation is a result of the constraint.

TABLE 15.—*Composition loadings, a_{ik}, scaled scores, f'_{kj}, composition scores, f_{kj}, and reproduced data matrices for the two-factor principal components, varimax, and oblique factor models derived after transforming each variable to proportions of the maximum value*

[The products of the matrices of a_{ik} and f'_{kj} are approximations of the transformed data in table 1*B*. The products of the matrices of a_{ik} and f_{kj} are approximations of the original data in table 1*A*]

Model	Composition loadings, a_{ik}	Scaled scores f'_{kj}	Composition scores, f_{kj}
Principal components.	$a_{ik} =$ 1.015 −0.015; .950 .050; 1.002 −.002; .980 .020; 1.008 −.008; 1.021 −.021	$f'_{kj} =$ 0.835 0.395 0.219; −6.51 6.52 13.06	$f_{kj} =$ 79.3 11.9 8.8; −618 196 522
Varimax.....	$a_{ik} =$ 0.784 0.216; .259 .741; .684 .316; .503 .497; .730 .270; .838 .162	$f'_{kj} =$ 1.138 0.143 −0.309; .233 .898 1.273	$f_{kj} =$ 108.1 4.3 −12.4; 22.1 26.9 50.9
Oblique	$a_{ik} =$ 0.907 0.093; 0 1.000; .733 .267; .420 .580; .813 .187; 1.000 0	$f'_{kj} =$ 0.992 0.261 −0.052; .469 .697 .862	$f_{kj} =$ 94 8 −2; 45 21 34

Reproduced data matrices

(Product of matrices of a_{ik} and f'_{kj})	(Product of matrices of a_{ik} and f_{kj})
$\hat{x}'_{ij} =$ .94 0.30 0.03; .47 .70 .86; .85 .38 .19; .69 .52 .48; .89 .35 .12; .99 .26 −.05	$\hat{x}_{ij} =$ 90 9 1 (100); 44 21 34 (99); 81 12 8 (101); 65 16 19 (100); 85 10 5 (100); 94 8 −2 (100)

The exercise was carried out using Chayes and Kruskal's (1966) equation 7 to search for the column variances in an open matrix with no intercolumn correlation that would yield, on adjustment of all rows to sum to 100, a closed matrix with column means and variances equal to those computed from data given by Lee and Van Loenen (1971, table 5) for 81 samples of a granitoid intrusive in eastern Nevada. (The computed means and standard deviations are given in table 18 of part III.) Chayes and Kruskal's equation 7 yielded the following variances for the open matrix:

Column	*Representing*	*Variance*
1	SiO_2	122.417
2	Al_2O_3	−2.2181
3	Fe_2O_3	.2394
4	FeO	.4104
5	MgO	.3104
6	CaO	.9272
7	Na_2O	−.0658
8	K_2O	.4306
9	H_2O+	.0950. (33)

A 500 by 9 matrix was constructed from a population of normally distributed psuedorandom numbers with column means equal to the computed means for the granitoid intrusive and column variances equal to these open variances where they are nonnegative or equal to zero where they are negative. After setting all negative values that appeared in the 500 by 9 matrix to zero, each matrix row was adjusted to sum to 100 by dividing through by the row-sum and multiplying by 100. The matrix then became closed in the terminology of Chayes (1960).

The column means and standard deviations of the closed matrix were

Column	*Mean*	*Standard deviation*
1	71.29	3.59
2	15.20	1.78
3	1.15	.48
4	1.19	.68
5	.94	.53
6	2.37	1.00
7	3.75	.44
8	3.43	.78
9	.69	.31. (34)

The matrix of closed data was then used to derive varimax factor models containing two to eight factors (only the first eight eigenvalues of the matrix of cosine

TABLE 16.—*Composition loadings, a_{ik}, scaled scores, f'_{kj}, composition scores, f_{kj}, and reproduced data matrices for the two-factor principal components, varimax, and oblique factor models derived after transforming each variable to proportions of the range*

[The products of the matrices of a_{ik} and f'_{kj} are approximations of the transformed data in table 1C. The products of the matrices of a_{ik} and f_{kj} are approximations of the original data in table 1A]

Model	Composition loadings, a_{ik}	Scaled scores, f_{kj}	Composition scores, f_{kj}
Principal components.	a_{ik} = 1.135 −0.135; .379 .621; .958 .042; .670 .330; 1.046 −.046; 1.236 −.236	f'_{kj} = 0.718 0.317 0.109; −.643 1.045 1.203	f_{kj} = 82.3 11.6 6.1; 21.1 31.2 47.7
Varimax.....	a_{ik} = 0.925 0.075; −.069 1.069; .692 .308; .314 .686; .808 .192; 1.056 −.056	f'_{kj} = 0.979 0.177 −0.101; −.056 .731 .731	f_{kj} = 94.1 7.8 −1.8; 47.5 22.7 29.8
Oblique	a_{ik} = 0.883 0.117; 0 1.000; .675 .325; .340 .660; .779 .221; 1.000 0	f'_{kj} = 1.043 0.141 −0.152; −.129 .774 .787	f_{kj} = 97 7 −4; 44 24 32

Reproduced data matrices

(Product of matrices of a_{ik} and f'_{kj})	(Product of matrices of a_{ik} and f_{kj})
$\hat{x}'_{ij}$ = 0.90 0.22 −0.04; −.13 .77 .79; .66 .35 .15; .27 .56 .47; .78 .28 .06; 1.04 .14 −.15	$\hat{x}_{ij}$ = 91 9 0 (Row-sum 100); 44 24 32 (100); 80 12 8 (100); 62 18 20 (100); 85 11 4 (100); 97 7 −4 (100)

TABLE 17.—*Measurements of correspondence between the observed data x_{ij}, and the data reproduced from the two-factor models, $\hat{x}_{ij}$*

[d_j^* is the mean of the six values of $d_{ij} = (\hat{x}_{ij} - x_{ij})$; d_j^{**} is the standard deviation of the six values of d_{ij}. r_j is the correlation coefficient between $\hat{x}_{ij}$ and x_{ij} for the six pairs; r_j^2 is the coefficient of determination]

Transformation	d_j^* Variable (j)			d_j^{**} Variable (j)			r_j Variable (j)			r_j^2 Variable (j)		
	1	2	3	1	2	3	1	2	3	1	2	3
None (table 14)	-0.6	0.7	-0.1	5.5	8.4	2.9	0.94	0.28	0.98	0.88	0.08	0.96
To proportions of the maximum for each variable (table 15) ..	-.1	.1	.0	3.2	7.9	4.8	.98	.41	.93	.96	.17	.86
To proportions of the range for each variable (table 16) ..	-.2	.9	-.7	2.6	7.7	5.9	.99	.49	.90	.98	.24	.81

theta were greater than zero), and each of the models was used to reproduce an estimate of the original matrix. Although the models employed the varimax axes as reference vectors, use of the principal components axes or any oblique axes would have led to the same reproduced data and the same goodness-of-fit measures. The goodness-of-fit measures included the coefficients of determination, eight sets of r_j^2's, which were used to construct the factor-variance diagram of figure 2. In constructing figure 2, all values of r_j^2 for the single factor model were set equal to zero, because a model with only one factor would lead to a reproduced data matrix with no intracolumn variance.

The factor-variance diagram shows that, for the most part, the addition of a factor to the model results in an important increase in the accountable variance for only

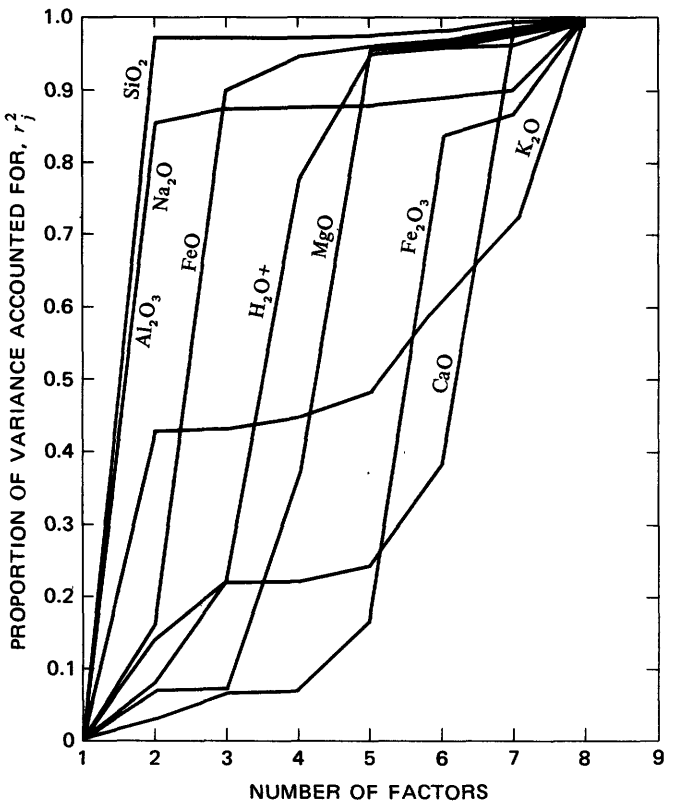

FIGURE 2.—Factor-variance diagram derived from a closed matrix of random numbers having column means and variances similar to those of the data matrix for the granitoid intrusive in the southern part of the Snake Range, Nev.

one variable. As will become apparent in part III, this is far different from most of the situations encountered in treating real geochemical and petrographic data and suggests that the effect of the constant row-sum on the nature of the factor-variance diagrams is exceedingly minor. However, even if this were not the case, the factor-variance diagrams would still provide a correct appraisal of the degree to which each factor model will account for the total variance in the original data.

COMMON AND UNIQUE FACTORS

In general, the coefficients of determination, r_j^2, increase for each variable with each increase in m, the number of factors in the model. This is in accord with the conventional concept of common factors. (See, for example, Harman, 1967, p. 15.) Common factors, representing processes or effects, are those that cause variability in more than one variable. However, the situation is not uncommon where the addition of one factor to the model will cause only small increases in r_j^2 for all but one variable but will cause a large increase in r_j^2 for that one variable. In conventional factor analysis terminology, this factor is said to be unique (Harman, 1967, p. 15). In some situations, more than one unique factor may appear.

It has been conventional in factor analysis applications to use common factors only, excluding those that are unique. If a major portion of the variation in a compositional variable ($j = u$) can be explained only with a unique factor, and the unique factor is not included in the model, the model will, of course, fail to account for some increment of this variable. The increment, which can be ascribed to a unique process or condition, is that amount, I_{iu}, which must be added to the reproduced value for the unique variable, $\hat{x}_{iu}$, to give the observed value for the sample, x_{iu}, as in

$$x_{iu} = \frac{K(\hat{x}_{iu} + I_{iu})}{K + I_{iu}}. \tag{35}$$

The increment that must have been added to or subtracted from the ith sample by some process attributed to the unique factor can be estimated by rearrangement of equation 35 and solving for I_{iu}. Similarly, if two of the compositional variables ($j = u$ and $j = v$) respond to unique factors not included in the model, the increments, I_{iu} and I_{iv}, can be estimated by rearrangement and simultaneous solution of equations 36a and 36b:

$$x_{iu} = \frac{K(\hat{x}_{iu} + I_{iu})}{K + I_{iu} + I_{iv}} \tag{36a}$$

$$x_{iv} = \frac{K(\hat{x}_{iv} + I_{iv})}{K + I_{iu} + I_{iv}}. \tag{36b}$$

COMPOSITIONS VERSUS COMPOSITION SCORES

Generally in this report, the composition scores have been referred to as properties of the reference axes and denoted by f_{kj}. However, the composition scores are essentially analogous to reproduced data designated by $\hat{x}_{ij}$ in equation 17 of part I. That is, each sample vector representing a composition, x_{ij} ($1 \leqslant j \leqslant M$), has a set of composition scores, $\hat{x}_{ij}$, which represents the original composition modified to correspond with the factor solution. Thus, the only difference is in notation. The symbol, f_{kj}, is used for the scores of the kth reference vector, and $\hat{x}_{ij}$ is used for the scores (reproduced data) of the ith sample vector.

When the number of factors in the model is less than the number of variables in the data matrix, most of the sample vectors are in different positions and are less than unit length after projection into the m-dimensional factor space. Consequently, the vectors no longer represent the exact compositions of the corresponding samples. That is, the composition scores, $\hat{x}_{ij}$, for the vectors are not exactly

equal to the sample compositions, x_{ij} . The same is true for the unscaled scores, $\hat{x}''_{ij}$, and the normalized data, x''_{ij} , and for the scaled scores, $\hat{x}'_{ij}$, and the transformed data, x'_{ij} . The degree of departure between the unscaled scores and the normalized data is inversely related to the sample communality.

The composition scores for the sample vectors or for reference axes that represent selected end-member compositions may be viewed as compositions that have been modified to conform with a compositional system. The degree of modification required will tend to be large when the vector communality is small and may be cause for declaring that a particular composition is inconsistent with the compositional series.

IDENTIFICATION OF THE SAMPLES OF EXTREME COMPOSITION

After it has been determined that the sample vectors occur mostly within a factor space of m dimensions and that the factor model, therefore, should contain m end-members, it may be helpful to identify the m samples of extreme composition. By samples of extreme composition, we mean the m samples having compositions that can be combined in various proportions, none exceeding unity, to give approximations of the compositions of all the other samples.

In Q-mode analyses of data matrices with variable row-sums, the investigator has no means of scaling the initial loadings and scores and computing the composition values; in a search for extreme samples, his only recourse is to find the sample vectors on which all the initial loadings for the other sample vectors are less than one—that is, the procedure suggested by Imbrie (1963). This will frequently serve to identify the same samples as being of extreme composition as would have been identified using the composition loadings. Examination of the initial varimax loadings in table 12, however, shows that this will not always be true. According to Imbrie's procedure, these loadings indicate that samples 1 and 2 are of extreme composition; the initial loadings of all samples with respect to the oblique vectors represented by them are all one or less:

$$a''_{ik} = \begin{matrix} 1.000 & 0 \\ 0 & 1.000 \\ .928 & .391 \\ .409 & .896 \\ .989 & .182 \\ .980 & -.157. \end{matrix} \qquad (37)$$

Although samples 1 and 2 are, indeed, of extreme normalized composition, the fact that they are not of extreme actual composition becomes evident when the composition loadings are derived. The scale factors to be used for deriving the composition loadings are computed from the unscaled scores for samples 1 and 2, taken from the reproduced matrix of normalized data at the bottom of table 12:

$$f''_{11} = 0.97 \quad f''_{12} = 0.23 \quad f''_{13} = -0.04$$

$$f''_{21} = -0.10 \quad f''_{22} = 0.60 \quad f''_{23} = 0.61. \qquad (38)$$

The scale factors, from equation 9, are 0.931 and 1.290. Division of the respective columns of matrix 37 through by these values and adjustment of each row to sum to unity give the composition loadings as follows:

$$a_{ik} = \begin{matrix} 1.000 & 0 \\ 0 & 1.000 \\ .767 & .233 \\ .387 & .613 \\ .883 & .117 \\ 1.131 & -.131. \end{matrix} \qquad (39)$$

These values indicate that sample 6 is beyond the range of compositions represented by samples 1 and 2, a fact that also seems apparent from the matrix of original data in table 1*A*. Alternatively, returning to the initial varimax loading matrix in table 12 and selecting samples 6 and 2, rather than 1 and 2, as the end-members for the oblique model, the initial oblique loadings are as given in table 12, and the oblique composition loadings, which are all in the range from zero to plus one, are as given in table 16. The revised scale factor, s_k, for the first end-member (sample 6) is found, using equation 9, to be 1.087 (table 12).

The products of the matrices of the composition loadings and composition scores for the oblique model are equal to the corresponding products for the principal components and varimax models to two significant figures (table 16). Thus, the oblique model is consistent with the other types.

The fundamental differences between the procedure suggested here and the method for deriving the oblique model given by Imbrie (1963) are that here (1) the reference vectors are taken as the vectors representing the samples of extreme composition rather than those of extreme normalized composition, and (2) the scale factors for adjusting the factor loadings are computed from the unscaled composition scores for the extreme samples rather than from their original or transformed compositions. The advantage of this second modification is that the reproduced data from the oblique model will be exactly the same as the data reproduced from the principal components or varimax models. In fact, derivation of the scale factors from the unscaled composition scores rather

than from the original or transformed compositions will lead to the same reproduced data regardless of any choice of reference axes within the m-dimensional factor space.

MODIFICATION OF A VARIMAX MODEL TO AN OBLIQUE MODEL

Commonly, the computed set of composition scores for one or more of the varimax factor axes contains one or more negative values. If this is viewed as undesirable for the purpose of the investigation and if the composition scores computed for the sample vectors tend to be all positive, it may be helpful to move the reference axes, one at a time, from the varimax positions by small increments toward the sample vectors. Computer programs can be easily prepared to do this, wherein, after each increment, the composition scores can be determined by the procedure give in part I, and the process terminated when all scores are nonnegative. The procedure leads to a model that approximates the conventional varimax solution, but the reference axes are oblique. The advantage of the oblique model is that the end-members for the model are more likely to be of realistic composition.

If the m-dimensional model does not account for large proportions of the variances for each variable, however, the sample vectors may have composition scores that are not all positive, and movement of the reference axes toward them will not have the desired effect. The only recourse then is to increase m.

In general, as the reference axes are moved from the varimax positions towards the sample vectors, the composition scores of the axes move towards the range of zero to K, and the loadings of the sample vectors on each of them become increasingly variable. As the axes are moved past sample vectors, the composition loadings move outside the range of zero to one.

TESTING THE PLAUSIBILITY OF ALTERNATIVE END-MEMBERS

As pointed out in part I, the unscaled varimax scores may be used to find the vector representation of any given composition in the varimax space and to determine whether or not the composition could be that of an end-member in the compositional system. For purposes of illustration, we may suppose that the varimax model derived after transforming the variables to proportions of their maximum values (table 15) is unsatisfactory because one of the composition scores, f_{13}, is negative and because the composition scores for the varimax axes, in general, are not sufficiently close to those believed, on the basis of other evidence, to have been involved in the genesis of the petrologic system under study. Then it may be desirable to test the following five observed, theoretical, or hypothetical phases as possible end-member compositions:

Phase					
1	z_{1j} =	60	30	10	
2	z_{2j} =	10	80	10	
3	z_{3j} =	5	90	5	
4	z_{4j} =	90	10	0	
5	z_{5j} =	25	25	50.	(40)

The transformed values in this case are obtained by dividing the value of z_{ij} by the corresponding values of $xmax_j$ from table 1A, giving

Phase					
1	z'_{1j} =	0.632	1.000	0.250	
2	z'_{2j} =	.105	2.667	.250	
3	z'_{3j} =	.053	3.000	.125	
4	z'_{4j} =	.947	.333	0	
5	z'_{5j} =	.263	.833	1.250.	(41)

The transformed compositions are normalized to give

Phase					
1	z''_{1j} =	0.523	0.827	0.207	
2	z''_{2j} =	.039	.995	.093	
3	z''_{3j} =	.018	.999	.042	
4	z''_{4j} =	.943	.332	0	
5	z''_{5j} =	.172	.546	.820.	(42)

Applying equation 25 of part I, the initial loadings on the varimax axes are computed for each phase. These are found to be

Phase	a''_1	a''_2	$h^2 = \Sigma(a''_k)^2$	
1	0.546	0.716	0.811	
2	.133	.648	.438	
3	.126	.605	.382	
4	.944	.329	.998	
5	.018	.999	.999.	(43)

The values of a''_k may be converted to composition loadings by dividing by the corresponding scale factors for the varimax model (table 11) and adjusting to sum to unity. The composition loadings can then be postmultiplied by the varimax composition scores, f_{kj}, from table 15 to give composition scores of the 5 phases. These are

Phase					
1	$\hat{x}_{1j}$ =	65.37	15.54	19.10	
2	$\hat{x}_{2j}$ =	40.50	22.10	37.40	
3	$\hat{x}_{3j}$ =	40.71	22.04	37.24	
4	$\hat{x}_{4j}$ =	90.23	8.99	0.81	
5	$\hat{x}_{5j}$ =	24.13	26.41	49.45.	(44)

Any of the phase composition scores could be the composition of an end-member in the petrologic system if

the composition loadings in the final model are subject to plausible geologic interpretation. However, the composition scores differ from the phase compositions tested to a degree indicated by h^2 in 43. Phases 1, 2, and 3 might be rejected because the revised compositions are appreciably different from the original compositions tested. (Compare 40 and 44.) Phases 4 and 5 may be accepted as possible end-members if the composition loadings of the other samples on the new reference axes are geologically plausible.

If the decision were made to describe the samples in terms of phases 4 and 5, one would form the following partition of the matrix in 43:

$$a''_{ik} = \begin{matrix} 0.944 & 0.329 \\ 0.018 & 0.999 \end{matrix} = A, \qquad (45)$$

and then proceed as described in equations 28–31 of part I.

SUMMARY

The recommended steps to be followed in the development of a geochemical or petrologic model using the extended form of *Q*-mode factor analysis can be summarized as follows. It is assumed that the rows of the data matrix have a constant sum.

1. Use the *Q*-mode factor analysis computer program (CABFAC) of Klovan and Imbrie (1971), rotating 2 to M (or 10 if $M > 10$) varimax axes. Using the unscaled varimax scores, output from CABFAC, compute all the varimax scale factors and varimax composition scores for each rotation.
2. Using the appropriate scale factors, adjust each matrix of initial varimax loadings to a matrix of varimax composition loadings.
3. Multiply each matrix of varimax composition loadings by the corresponding matrix of varimax composition scores to produce $M-1$ approximations of the original data, generally in units of percent or parts per million.
4. For each of the $M-1$ approximations, generate the measures of correspondence to the original data that are given in table 17.
5. Use the measures of r_j^2 to construct a factor-variance diagram and select the number of end-members, m, that the model should contain in order to account for as much of the variance in as many of the variables as possible. The principle of parsimony (Imbrie, 1963) should be applied.
6. The next step is to select the reference axes or end-member compositions, and here geologic criteria play an important part. For an m-dimensional model, it is necessary to form an m by m matrix of the initial loadings of the reference vectors on the varimax axes. One or more of the reference vectors may be those representing samples in the original data matrix if those samples are thought to represent end-members or extremes in the petrologic system. They also may be the varimax axes or any other vectors within the varimax space as long as the composition scores for the axes or vectors approach the compositions of materials thought possibly to have been end-members in the petrologic system. The reference vectors may also be vectors with composition scores approaching the compositions of materials that had been tested for fit in the factor space. In brief, if the reference vectors are not the varimax axes or vectors representing the original samples, they are selected either by examining the composition scores of selected vectors or by finding vectors with composition scores approaching the compositions of materials tested and found to be representable in the factor space. The techniques are given in parts I and II.
7. When an m by m matrix of initial loadings of the reference vectors on the varimax axes has been selected, the initial loadings on the reference axes (generally oblique axes) are derived using the method given here, which is originally from Imbrie (1963).
8. The unscaled scores of the new reference axes are computed by multiplying the matrix of initial loadings on the varimax axes by the matrix of unscaled varimax scores. The scale factors are then computed for the reference axes and their composition scores are derived.
9. The scale factors and the initial loadings on the new reference axes (derived in step 7) are then used to compute the composition loadings on the new reference axes. The composition loadings, especially their signs, are examined for geologic plausibility. If they are not acceptable, return to step 6 for the selection of alternative reference axes.
10. As a partial check for computational errors, postmultiply the matrix of composition loadings on the selected reference axes by the matrix of their composition scores. Errors are present if the departure of the product matrix from the corresponding matrix derived in step 3 is beyond that which can be attributed to round off.

The steps outlined here were followed in application of the *Q*-mode technique to some petrologic mixing problems, as described in part III.

Part III. Applications to some Petrologic-Mixing Problems

INTRODUCTION

The extended form of Q-mode factor analysis has been applied to four problems in igneous petrology to illustrate how it may be used and the kinds of information it may yield. Problems in igneous petrology were chosen because igneous systems generally exhibit compositional variations that are simpler than those in sedimentary and metamorphic systems and because relatively precise analyses of igneous suites are available in abundance. The simpler compositional variations in igneous rocks are attributed to the relative simplicity of the systems in which they formed, increasing the likelihood that a small number or realistic end-members can be found that will account for a large part of the compositional variation.

The four sample suites chosen for examination are from (1) a Pleistocene rhyolite-basalt complex on the Gardiner River, Yellowstone National Park, Wyo. (Fenner, 1938), (2) a Jurassic granitoid intrusive in the southern part of the Snake Range, Nev. (Lee and Van Loenen, 1971), (3) lavas and pumices of the 1959 summit eruption at Kilauea, Hawaii (Murata and Richter, 1966), and (4) the layered series of the Skaergaard intrusion, Greenland (Wager and Brown, 1968). Analyses of the samples are from the references cited, but only the determinations on constituents that are essential to the principal rock-forming minerals were used in this exercise; these constituents are SiO_2, Al_2O_3, Fe_2O_3, FeO, MgO, CaO, Na_2O, K_2O and H_2O+. The same nine constituents were used in all four studies so that the results would be more easily comparable from one to another. Subsequent studies should, perhaps, include such constituents as MnO, TiO_2, P_2O_5, and Cr_2O_3, especially if they might be diagnostic of processes that may have led to important amounts of variation in the rocks under study.

Means and standard deviations for the compositional data of all four sample suites are given in table 18.

Prior to the factor analyses, each data matrix was transformed so that each column (chemical variable) ranged from zero to one (that is, to proportions of the range).

The first step in the examination of each sample suite was the construction of a factor-variance diagram. The diagram, as shown in part II, is merely a plot of the values of r_j^2, as given in table 17, versus the number of factors in the model. The values of r_j^2 were derived from varimax models but would be identical if a principal components model or some oblique model had been used. Factor-variance diagrams show immediately and concisely the complexity of the petrologic system and indicate which compositional variables can be explained, and to what degree, by models with any given number of end-members.

TABLE 18.—*Means and standard deviations (in weight percent) for four sets of data examined by Q-mode factor analysis*

[All analyses were adjusted before computation of means and standard deviations so that the nine constituents summed to 100]

Constituents	Data sets: Rhyolite-basalt complex, Yellowstone National Park, Wyo. (15 samples) [1]	Granitoid intrusive in the southern Snake Range, Nev. (81 samples) [2]	Lavas and pumice from the 1959 summit eruption at Kilauea, Hawaii (22 samples) [3]	Layered series of Skaergaard intrusion, Greenland (19 samples) [4]
	Means			
SiO_2	62.62	71.72	49.89	45.34
Al_2O_3	14.60	14.97	12.12	12.52
Fe_2O_3	2.07	1.18	1.88	4.29
FeO	4.65	1.14	10.14	19.70
MgO	4.07	.94	13.09	5.65
CaO	5.98	2.32	10.34	9.00
Na_2O	3.11	3.69	2.00	2.69
K_2O	2.40	3.38	.49	.31
H_2O+	0.50	.67	.03	.50
	Standard deviations			
SiO_2	7.69	3.11	1.26	6.96
Al_2O_3	1.15	1.08	1.39	6.00
Fe_2O_3	.51	.50	.70	3.16
FeO	2.58	.64	.70	10.08
MgO	2.34	.56	4.12	6.15
CaO	3.00	.97	1.23	2.38
Na_2O	.31	.32	.25	1.25
K_2O	1.34	.73	.06	.15
H_2O+	.10	.31	.03	.38

[1] Data from Fenner (1938, table 6), excluding samples Y.P. 904, Y.P. 1024, and Y.P. 1025.
[2] Data from Lee and Van Loenen (1971, table 5), excluding samples 71, 72, 77, 81, 85, and 87.
[3] Data from Murata and Richter (1966, table 1) and Wright (1973, table 2). Same selection of samples as used by Wright.
[4] Data from Wager and Brown (1968, table 5).

RHYOLITE-BASALT COMPLEX ON THE GARDINER RIVER, YELLOWSTONE NATIONAL PARK, WYO.

Exposures of the rhyolite-basalt complex occur in an area about 76 m wide and 274 m long within the canyon of the Gardiner River in the northern part of Yellowstone National Park (Fenner, 1938). The complex forms part of the Obsidian Creek Member of the Plateau Rhyolite of Pleistocene age in the third volcanic cycle of Christiansen and Blank (1972, p. B13). The thickness of the complex is difficult to determine, but it has been measured as less than

30 m at one point. The complex consists of basalt impregnated with rhyolite and rhyolite containing abundant inclusions of basalt. The rhyolite penetrating the basalt forms veinlets and pseudodikes, and an intermingling of rhyolitic and basaltic constituents is evident from the groundmass and phenocrysts throughout most of the complex. Fenner showed that the compositions of most analyzed samples can be explained as mixtures of rhyolitic and basaltic end-members; the principal mathematical evidence is in the straight-line relationships as seen on silica-variation diagrams. According to the interpretation of Fenner, the complex formed by the extrusion of a rhyolitic lava that attacked an older basalt which constituted the floor and walls of the Gardiner River canyon. More than 68 percent of some basalt has been replaced by rhyolite, which entered the basalt as gaseous emanations from the rhyolite lava. In Fenner's interpretation, the gaseous emanations concept was used to overcome the difficulties of penetrating the basalt with rhyolitic liquid.

Fenner's interpretation of the origin of the complex was debated by Wilcox (1944), who explained the complex as the product of the mixing of rhyolitic and basaltic lavas that erupted simultaneously. Wilcox's principal evidence, rejected by Fenner (1944), consisted of observations of a chill phase in the rhyolite in contact with country rock but not in contact with basalt and the presence of quartz and orthoclase xenocrysts from the rhyolite in the basalt.

The factor analysis to be described here does not offer any new evidence with regard to the problem of the complex's origin, but it does emphasize and show some exceptions to the linearity of the compositional gradation between rhyolite and basalt. It also provides estimates of the relative proportions of basalt and rhyolite in each of the analyzed samples. The principal conclusion reached through the analysis is that the rhyolitic end-member, whether liquid rhyolite or gaseous emanations, had to have a composition near that of the least contaminated rhyolites now present in the complex. This tends to support the interpretation of Wilcox (1944). The exceptions to the linear compositional gradient from rhyolite to basalt involve Na_2O, H_2O+, and Fe_2O_3. The first two of these are late-stage volatile constituents that appear to have behaved independently of other constituents, and the Fe_2O_3 probably formed with oxidation after the mixing had occurred.

Fifteen of the 18 analyses given by Fenner (1938, p. 1466) were used to derive the factor-variance diagram. Analyses of samples 1025, 904, and 1024 were not used because the source of the samples has been a matter of debate (Wilcox, 1944, p. 1067) and Fenner referred to them as varied from the standard composition. The factor-variance diagram is given in figure 3 and shows the proportion of the total variance in each constituent that

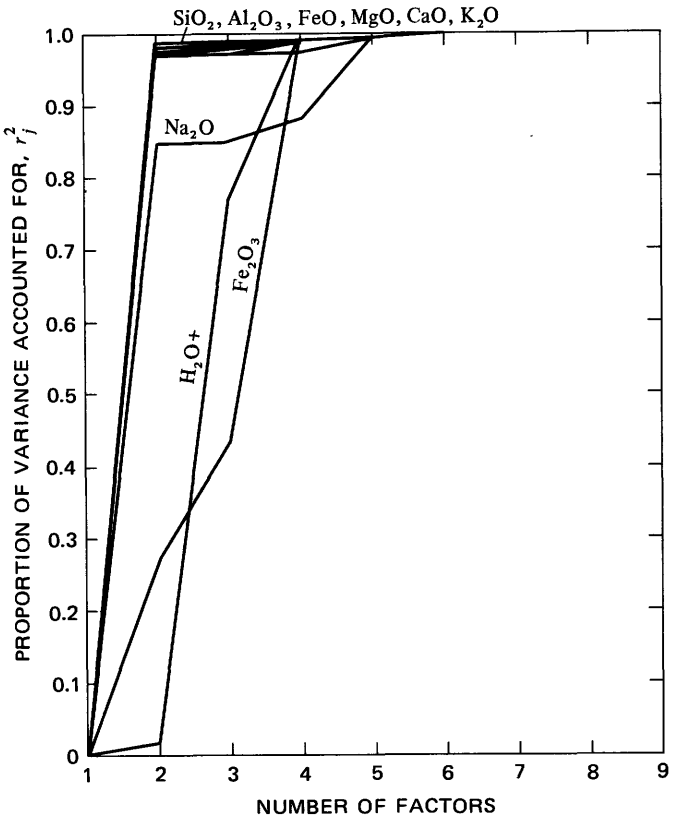

FIGURE 3.—Factor-variance diagram for the rhyolite-basalt complex on the Gardiner River, Yellowstone National Park, Wyo.

can be explained by models containing one to nine factors, or end-members. The curves for all constituents originate at zero variance for one factor, inasmuch as single-factor models would lead to reproduced data matrices with no intracolumn variation. All but two of the curves rise sharply, indicating that two-factor models can account for most of the chemical variation in the complex. That is, a linear combination of two end-members in varying proportions can account for nearly all of the variation in six of the major oxides and 85 percent of the variation in Na_2O. The variations in H_2O+ and Fe_2O_3, however, can be explained equally well only if the model contains at least four factors. These constituents appear to represent factors that are largely unique; that is, each of these constituents appears to have behaved independently of the others, and each may have been controlled by some process that affected it alone. The process that accounted for the 15 percent unexplained variance in Na_2O and most of the variance in H_2O+ is interpreted as migration of late-stage volatile constituents in the lavas. The process controlling Fe_2O_3 was probably later oxidation.

Because of the apparent importance of postvolcanic oxidation, the Fe_2O_3 in Fenner's analyses was recomputed and combined with FeO to give total iron as FeO, usually designated by 'FeO.' The complete set of analyses,

TABLE 19.—*Analyses of a rhyolite-basalt complex on the Gardiner River, Yellowstone National Park, Wyo., adjusted so that eight major oxides sum to 100 percent, and corresponding data reproduced from the factor model*

[Analyses from Fenner (1938)]

Sample No. Y.P.	SiO_2	Al_2O_3	'FeO'	MgO	CaO	Na_2O	K_2O	H_2O+
			A. **Adjusted analyses, x_{ij} (in percent)**					
969	51.64	16.25	10.41	7.44	10.53	2.77	0.52	0.44
918	54.33	16.06	9.49	6.70	8.98	2.87	1.04	.53
948	54.49	15.74	9.49	6.75	9.30	2.76	.98	.49
985	55.07	15.72	9.40	6.27	9.25	2.77	1.13	.40
512	55.33	15.74	9.40	6.34	8.94	2.61	1.13	.52
925	58.66	15.31	7.96	5.35	7.28	3.13	1.58	.72
514	59.81	14.97	7.76	5.09	7.02	2.94	1.97	.45
909A	62.24	14.82	6.79	4.27	6.09	3.27	2.02	.51
909B	64.94	14.11	5.78	3.45	5.15	3.36	2.66	.56
543	65.92	14.00	5.38	3.19	4.78	3.13	2.98	.61
467	67.30	13.94	4.99	2.55	4.22	3.22	3.26	.53
980	68.06	14.20	4.30	1.95	4.16	3.58	3.22	.53
505	72.23	13.13	3.26	1.02	2.22	3.37	4.16	.61
554	75.48	12.71	1.85	.37	1.10	3.58	4.59	.31
914	75.72	12.70	1.72	.40	.83	3.44	4.80	.37
			B. **Reprodnced dnta (in percent)**					
969	51.49	16.30	10.66	7.51	10.39	2.77	0.43	0.44
918	54.81	15.78	9.42	6.50	9.07	2.87	1.02	.53
948	53.85	15.96	9.81	6.82	9.48	2.76	.84	.49
985	53.94	15.96	9.80	6.81	9.47	2.77	.85	.40
512	53.63	16.01	9.92	6.91	9.60	2.61	.79	.52
925	59.65	14.99	7.56	4.99	7.08	3.13	1.89	.72
514	59.30	15.15	7.82	5.20	7.36	2.94	1.79	.45
909A	62.36	14.63	6.62	4.21	6.07	3.27	2.34	.51
909B	65.24	14.19	5.55	3.34	4.92	3.36	2.85	.56
543	65.51	14.19	5.50	3.30	4.87	3.13	2.88	.61
467	67.06	13.97	4.94	2.85	4.27	3.22	3.15	.53
980	68.15	13.73	4.45	2.45	3.75	3.58	3.37	.53
505	71.66	13.26	3.24	1.46	2.45	3.37	3.96	.61
554	76.72	12.57	1.45	.00	.54	3.58	4.82	.31
914	76.63	12.60	1.51	.05	.60	3.44	4.80	.37

adjusted so that the eight oxides sum to 100 for each sample, is given in table 19*A*. The recomputed factor-variance diagram is essentially the same as the one given in figure 3, except for the absence of the curve for Fe_2O_3.

The eigenvalues of the cosine theta matrix for the data in table 19*A* are given in table 20 and, like the factor-variance diagram, show that somewhere between two and four factors are required in the factor model. The eigenvalues, however, do not indicate the constituents whose behaviors depart from that of the others.

A plot of the sample vectors in the two-dimensional varimax space is given in figure 4 and shows that the 15 samples used in the analysis as well as the 3 samples (1025, 904, and 1024) that vary from the standard have communalities near 1. The largest departures from 1 among the 15 samples used in the analysis are for samples 925 and 554, both with communalities of 0.94. The largest departure among all 18 samples is for sample 1025 with a communality of 0.90. Although sample 1025 is one of the samples that varies from the standard, as pointed out by Fenner (1938, p. 1476), it does not vary much more than several of the samples accepted as taken from the rhyolite-basalt complex. It seems probable that some kind of genetic relation exists between the samples that vary from the standard and the rhyolite-basalt complex, even though they may differ in age and occurrence.

TABLE 20.—*First eight eigenvalues of the cosine theta matrix for the rhyolite-basalt complex on the Gardiner River and CPN* [1]

No.	Eigenvalue	CPN
1	11.353	0.7569
2	3.298	.9767
3	.258	.9939
4	.080	.9993
5	.007	.9997
6	.003	.9999
7	.001	1.0000
8	.001	1.0000
Total.....	15.001	

[1]Cumulative proportions of *N*.

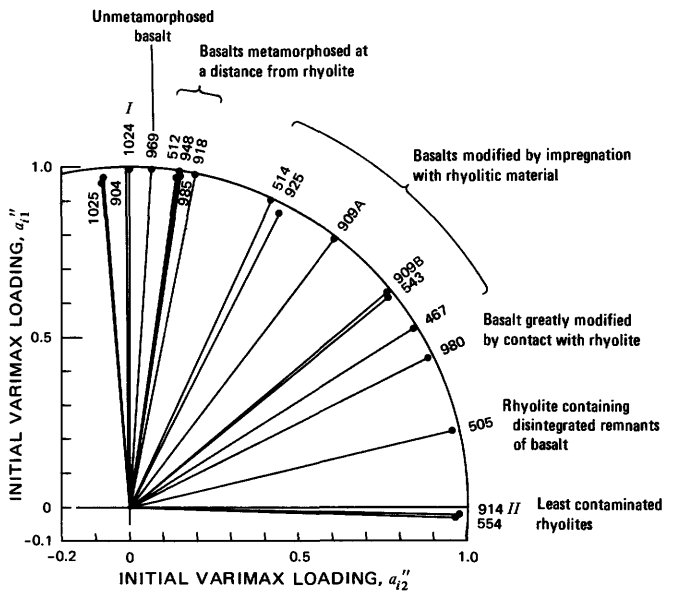

FIGURE 4.—*Q*-mode vector diagram for the rhyolite-basalt complex on the Gardiner River, Yellowstone National Park, Wyo. Numbers at ends of vectors are the sample numbers with prefix Y.P. in table 19. Vectors *I* and *II* are the varimax axes. Notes on vectors give interpretation of Fenner (1938). Wilcox (1944) interpreted each sample (except 904, 1024, and 1025) as resulting from the mixture of basaltic and rhyolitic lavas.

Composition scores for some of the vectors represented in figure 4 are given in table 21. It may be seen that vectors outside of the positive quadrant formed by the two varimax axes have at least one negative score value. This is true for two of the samples that vary from the standard (1025 and 904) as well as for the two most rhyolitic samples

TABLE 21.—*Composition scores, $\hat{x}_{ij}$, for some vectors in figure 4 (in percent)*

Vectors	SiO_2	Al_2O_3	'FeO'	MgO	CaO	Na_2O	K_2O	H_2O+
904	46.50	17.03	12.48	9.03	12.33	2.53	-0.44	0.54
1025	46.54	17.02	12.46	9.02	12.31	2.53	-.43	.54
I	49.41	16.60	11.42	8.14	11.19	2.63	.07	.54
1024	49.60	16.57	11.34	8.08	11.12	2.64	.10	.54
969	51.47	16.29	10.66	7.51	10.39	2.71	.43	.54
II	76.00	12.66	1.70	.02	.81	3.62	4.70	.49
554	76.69	12.56	1.45	-.19	.54	3.64	4.82	.49

from the complex (914 and 554). Because the negative score values are small, however, they can be set to zero, and the scores can then be recomputed to sum to 100 without seriously affecting the validity of subsequent arithmetic operations.

The linearity of the compositional gradation between rhyolite and basalt, as recognized by Fenner and verified by the factor-variance diagram of figure 3 and as represented in Fenner's silica variation diagrams and in the vector diagram of figure 4, makes it difficult to accept any other origin for the complex than the mixing of two end-members of nearly constant composition. One of these end-members can be taken as basalt, represented by sample 969 (fig. 4). This is the most basaltic sample that is unquestionably part of the complex. The other end-member could be the gaseous emanations referred to by Fenner, but it is not likely that such emanations would have a constant composition that could be represented by a vector with positive scores at the other end of the vector system, and there is little doubt that the other end-member could be so represented. Consequently, Wilcox's alternative explanation that the complex formed by the mixing of rhyolitic and basaltic lavas seems probable, and the second end-member is taken as rhyolitic lava similar in composition to the composition scores of the vector representing sample 554 (fig. 4). The composition scores for vector 969 and the adjusted scores for vector 554 are given as end-member compositions and designated by '969' and '554' in table 22.

TABLE 22.—*Compositions of the end-members, f_{kj}, for the factor model of the rhyolite-basalt complex on the Gardiner River (in percent)*

End-member	SiO_2	Al_2O_3	FeO	MgO	CaO	Na_2O	K_2O	H_2O+	Total
'969'	51.47	16.29	10.66	7.51	10.39	2.71	0.43	0.54	100.00
'554'	76.54	12.54	1.45	.00	.54	3.63	4.81	.49	100.00

The composition loadings for each of the 15 samples accepted as taken from the rhyolite-basalt complex are given in table 23, along with the increments (eq. 36a, 36b) of Na_2O and H_2O+ that must be added by independent processes to account for these constituents. For example, sample Y.P. 925 is interpreted to have formed by the mixing of 0.666 parts basaltic lava and 0.334 parts rhyolitic lava, with the addition of 0.12 percent Na_2O and 0.20 percent H_2O+. When the end-members of table 22 are mixed in these proportions and the increments of Na_2O and H_2O+ are added, the total mixture recomputed to sum to 100 is as given in table 19*B*. The entire matrix of recomputed data in table 19*B* is formed by postmultiplying the matrix of a_{ik} in table 23 by the matrix in table 22, adding the increments of Na_2O and H_2O+ in table 23, and adjusting each row of the product to sum to 100.

TABLE 23.—*Composition loadings, a_{ik}, and increments of Na_2O and H_2O+ for the factor model of the rhyolite-basalt complex on the Gardiner River*

Sample No. Y.P.	Composition loadings, a_{ik}		Increments (in percent)	
	End-member '969'	End-member '554'	Na_2O	H_2O+
969	1.000	0.000	0.06	-0.10
918	.866	.134	.04	.00
948	.907	.094	-.04	-.04
985	.905	.095	-.03	-.14
512	.918	.082	-.18	-.01
925	.666	.334	.12	.20
514	.691	.309	-.06	-.08
909A	.562	.438	.16	-.01
909B	.446	.554	.14	.05
543	.440	.560	-.09	.09
467	.379	.621	-.06	.02
980	327	.673	.26	.03
505	.194	.806	-.08	.11
554	.000	1.000	-.06	-.18
914	.006	.994	-.19	-.12

The correspondence of the matrix of reproduced data in table 19*B* to the matrix of original data in table 19*A* is measured by the goodness-of-fit statistics given in table 24. Although the correspondence is good, departures from perfect correspondence are attributed to (1) minor errors present in the analytical data, (2) some small amount of compositional variation in the actual basaltic and rhyolitic end-members, and (3) other processes that may have caused compositional variation in the complex, such as contamination of the lavas by other materials and later alterations brought about, perhaps, by weathering.

TABLE 24.—*Goodness-of-fit statistics for the factor model of the rhyolite-basalt complex on the Gardiner River*

[$d_{ij}=\hat{x}_{ij}-x_{ij}$; d_j^* = mean of d_{ij}; d_j^{**} = standard deviation of d_{ij}; r_j = correlation coefficient for $\hat{x}_{ij}$ and x_{ij}; r_j^2 = coefficient of determination]

Goodness-of-fit statistic	Variable (*j*)							
	SiO_2	Al_2O_3	FeO	MgO	CaO	Na_2O	K_2O	H_2O+
d_j^*	-0.08	-0.01	0.02	0.08	0.01	0.00	-0.02	0.00
d_j^{**}	.80	.23	.28	.33	.31	.00	.21	.00
r_j	1.00	.98	1.00	.99	1.00	1.00	.99	1.00
r_j^2	.99	.97	.99	.98	.99	1.00	.98	1.00

GRANITOID INTRUSIVE IN THE SOUTHERN SNAKE RANGE, NEV.

An exposure of granitoid intrusive rocks in the Snake Creek-Williams Canyon area of the southern Snake Range of eastern Nevada was interpreted by Lee and Van Loenen (1971) to have resulted from the crystallization of a magma after it intruded and assimilated one or more kinds of sedimentary country rock. The granitoid intrusive is exposed over an area of 34 km^2 and varies from granodiorite to quartz monzonite in composition. The predominant minerals are quartz, microcline, plagoiclase ($An_5 - An_{35}$), and biotite; hornblende and a number of other accessory minerals are also present. The intrusive is undeformed and domelike and probably has not been eroded to a depth of much more than 300 m. Its contacts with the overlying sedimentary rocks are roughly concordant with the bedding in the sediments at most localities, but exceptions that point clearly to stoping do occur. The eastern part of the intrusive contains abundant dark inclusions that Lee and Van Loenen referred to as xenoliths of Pioche Shale (Cambrian). They range in size from several centimetres in width to about 8,000 m^2 in area and are especially abundant where the intrusive comes into contact with the Pioche Shale. They tend to be ellipsoidal in shape and oriented with the long axis parallel to the intrusive margins. The preferred orientation was interpreted by Lee and Van Loenen to indicate that the magma flowed before solidification.

Sedimentary rocks in contact with the intrusive at the surface are the Osceola Argillite of Misch and Hazzard (1962), which is of Precambrian age and about 230 m thick, the Prospect Mountain Quartzite of Precambrian Z and Early Cambrian age, and the Pioche Shale and Pole Canyon Limestone of Cambrian age. The Prospect Mountain Quartzite is about 1,000 m thick, the Pioche Shale is 90 – 120 m thick, and the Pole Canyon Linestone is about 600 m thick. The Pioche Shale contains a limestone unit near its base, which ranges from 1.5 to 7.5 m in thickness and which is an informally named unit referred to locally as the Wheeler linestone. None of the sedimentary rocks shows evidence of appreciable metasomatism from the intrusive, whose principal effects on the sediments appear to have been some deformation and thermal metamorphism within about 1 m of the contacts. The Prospect Mountain Quartzite has the composition of a mature shelf sediment. The Pioche Shale is highly variable in composition both laterally and stratigraphically. The Pole Canyon Limestone is a rather pure limestone, whereas the Wheeler limestone, a bed in the Pioche Shale, tends to be more argillaceous and quite variable from one place of occurrence to another.

Chemical data on 81 samples of intrusive rocks and 6 samples of the dark inclusions are given in table 5 of Lee and Van Loenen (1971), along with normative mineral compositions and partial modal analyses for all samples. However, as for the other applications, only the data on the nine major oxides were used for the factor analysis. Lee and Van Loenen have shown that the concentrations of most of the chemical constituents vary with the CaO contents of the samples. The data on the nine major oxides, adjusted to sum to 100 for each sample, are given in table 25*A* for the five samples with the lowest and five samples with the highest CaO contents.

TABLE 25.—*Analyses of selected samples of the granitoid intrusive in the Snake Range, Nev. adjusted so that nine major oxides sum to 100 percent, and corresponding data reproduced from the factor model*

[Analyses from Lee and Van Loenen (1971, table 5)]

Sample No.	SiO_2	Al_2O_3	Fe_2O_3	FeO	MgO	CaO	Na_2O	K_2O	H_2O+
***A.* Adjusted analyses, x_{ij} (in percent)**									
1	76.33	13.48	0.33	0.23	0.17	0.47	3.72	4.42	0.84
2	75.91	13.15	.60	.24	.12	.52	5.18	4.08	.19
3	76.66	13.01	.42	.21	.21	.55	3.73	4.34	.86
4	75.44	14.05	.27	.44	.32	.61	4.11	4.21	.54
5	76.22	13.79	.38	.44	.10	.64	3.93	4.33	.16
80	66.20	16.65	2.03	2.03	2.23	4.16	3.76	2.23	.70
82	67.89	16.19	2.23	1.32	1.42	4.25	4.05	2.02	.64
83	64.12	17.21	2.52	2.32	2.21	4.33	3.42	2.72	1.16
84	62.84	18.17	2.54	2.54	2.13	4.37	3.76	2.34	1.32
86	63.84	17.49	1.93	3.05	2.13	4.57	3.86	2.34	.88
***B.* Reproduced data (in percent)**									
1	75.98	13.29	0.50	0.25	0.11	0.82	3.72	4.49	0.84
2	74.98	13.32	.55	.32	.17	93	5.18	4.36	.19
3	75.95	13.30	.50	.25	.11	.82	3.73	4.49	.86
4	75.27	13.52	.60	.37	.23	1.03	4.11	4.32	.54
5	76.26	13.40	.51	.27	.13	.85	3.93	4.49	.16
80	66.75	16.78	1.96	2.10	1.84	4.04	3.76	2.06	.70
82	67.35	16.46	1.84	1.94	1.70	3.77	4.05	2.25	.64
83	66.61	16.78	1.97	2.11	1.85	4.06	3.24	2.04	1.16
84	65.35	17.03	2.10	2.28	2.01	4.35	3.76	1.80	1.32
86	65.69	17.05	2.09	2.26	2.00	4.33	3.86	1.83	.88

The factor-variance diagram derived from the oxide data on all 81 intrusive samples is given in figure 5, and resembles, to some extent, that for the rhyolite-basalt complex on the Gardiner River. That is, a large part of the variance for most constituents can be accounted for by models with two factors, or end-members, and the variances in Na_2O and H_2O+ call for unique factors or for processes that controlled their variances alone. An important difference from the diagram for the rhyolite-basalt complex, however, is the generally lower proportions of the total variances accounted for by two-factor models. The reason for this may be partly due to the fact that the data of Lee and Van Loenen are based on an analytical method acknowledged to be less precise than conventional methods (Shapiro and Brannock, 1956, p. 19), but also the reason may be partly due to variation in the compositions of the end-members and to other geologic processes that caused minor compositional variations in the intrusive. Nevertheless, two-factor models that include unique processes to account for the variabilities in Na_2O

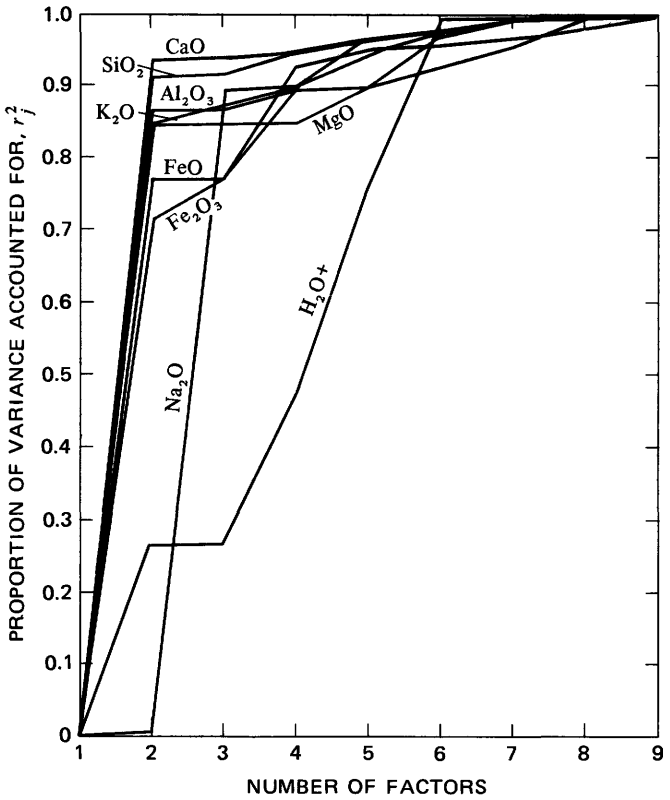

FIGURE 5.—Factor-variance diagram for the granitoid intrusive, southern Snake Range, Nev.

and H_2O+ can explain more than 70 percent of the variance in each constituent.

The eigenvalues of the cosine theta matrix are given in table 26.

TABLE 26.— *First nine eigenvalues of the cosine theta matrix for the granitoid intrusive in the southern Snake Range and CPN*[1]

No.	Eigenvalue	CPN
1	67.445	0.8327
2	10.230	.9590
3	1.013	.9715
4	.708	.9802
5	.572	.9873
6	.473	.9931
7	.318	.9970
8	.177	.9992
9	.064	1.0000
Total.....	81.000	

[1] Cumulative proportions of *N*.

If the interpretation of Lee and Van Loenen (1971) that the intrusive formed by solidification of a magma after assimilation of sedimentary country rocks is correct, the major questions raised pertain to which particular sedimentary units were assimilated and to what degree. The average compositions of the five major sedimentary units in contact with the intrusive at the surface and of the dark inclusions within the intrusive are given in table 27; the communalities of vectors representing these compositions in the two-factor varimax space are also listed. The same vectors are shown diagrammatically in figure 6. It is apparent that none of the sedimentary units is highly compatible with the compositional series formed by the intrusive samples in two-factor space; the composition scores of the vectors representing sedimentary units in figure 6 depart widely from the actual compositions of the units.

In contrast to the sedimentary units, the average composition of the dark inclusions within the intrusive has a large communality within the two-factor space (table 27) and is represented by a vector close to the margin of the range of vectors representing samples of the intrusive (fig. 6).

The consequences of increasing the number of factors in the model are examined in table 28, which gives the communalities of vectors representing the dark inclusions and of each of the sedimentary units in factor space of two to nine dimensions. As indicated in the table, the communalities of the limestone units and the Prospect Mountain Quartzite remain small until the model contains almost as many factors as there are chemical variables. When $m = M$, any combination of the nine variables can be perfectly represented in the factor space.

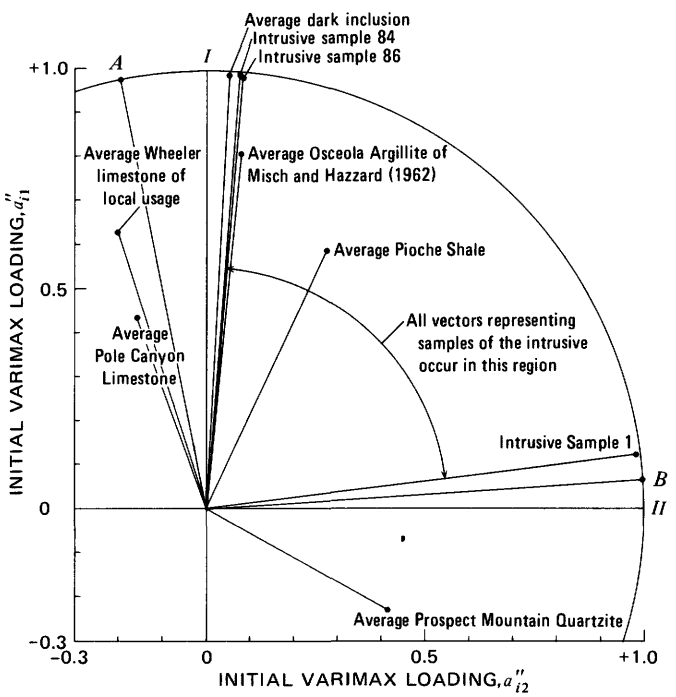

FIGURE 6.—*Q*-mode vector diagram for the granitoid intrusive. Vectors *I* and *II* are the varimax axes. See text for explanation of vectors *A* and *B*.

TABLE 27.—*Average compositions, in percent, of some sedimentary rocks and dark inclusions from the southern Snake Range and computed communalities,* h^2

Rock	Stratigraphic unit	SiO_2	Al_2O_3	Fe_2O_3	FeO	MgO	CaO	Na_2O	K_2O	H_2O+	Total	h^2
Limestone[1]	Pole Canyon Limestone.	4.5	1.1	0.33	0.25	2.0	91.0	0.12	0.37	0.07	99.74	0.218
Limestone[1]	Wheeler limestone (of local usage).	31.4	2.9	1.6	5.2	16.5	39.6	.3	.42	.19	98.11	.436
Argillite[1]	Osceola Argillite of Misch and Hazzard (1962).	62.9	16.2	4.3	1.9	2.2	4.7	1.5	3.0	2.2	98.90	.656
Quartzite[1]	Prospect Mountain Quartzite.	91.6	4.1	.79	.25	.23	.01	.13	2.1	.38	99.59	.227
Shale[1]	Pioche Shale	63.6	15.9	3.8	3.3	1.8	1.9	.58	5.3	2.7	98.88	.420
Dark inclusions[2]	Intrusive	62.9	16.5	2.4	3.0	2.3	4.2	4.0	2.1	1.0	98.40	.967

[1] Average composition is from Lee and Van Loenen (1971, table 9). Average compositions for the limestones were computed on a CO_2-free basis.
[2] Average of six analyses from Lee and Van Loenen (1971, table 5).

As shown in table 26, the average communality of the 81 vectors representing samples of the intrusive in the two-factor space is 0.9590. Table 28 shows that, in the two-factor space, only the vector representing the average dark inclusion has a communality this large. Of the sedimentary units, only the communality for the average Osceola Argillite is this great, but it requires a factor space of 5 dimensions. The communalities for the average Pioche Shale are second largest.

Although the evidence presented so far points strongly toward a model containing the dark inclusions as one of the end-members, a further comparison of the suitabilities of the dark inclusions, the Osceola Argillite, and the Pioche Shale was made by use of computer simulation. The assumption was made that the actual mafic end-member in the two-dimensional compositional system can be represented by a vector in the position of the vector representing the average dark inclusion in figure 6. The fact that the average Osceola Argillite and the average Pioche Shale are not so represented is attributed, for the present purpose, to errors in sampling—that is, the samples from which the averages were obtained are not perfectly representative of the sedimentary units. The composition loadings of all 81 intrusive samples were then determined with respect to the two reference vectors that represent the dark inclusions and intrusive sample 1 (fig. 6). The first matrix of simulated data, 81 by 9 in size, was then derived by mixing the composition of intrusive sample 1 with randomly selected analyses of individual dark inclusions, given by Lee and Van Loenen (1971, table 5), in the proportions indicated by the composition loadings. Six analyses of individual inclusions were available. Two other simulated data matrices were formed using six analyses of the Osceola Argillite and nine analyses of Pioche Shale (Lee and Van Loenen, 1971, table 1). The factor-variance diagrams derived from the three simulated data matrices are given in figure 7. There is no doubt that the diagram simulating the assimilation of dark inclusions bears the closest resemblance to the diagram derived from the 81 intrusive samples (fig. 5), and the results of the experiment are accepted as further evidence that the major part of the compositional variation in the intrusive can be correctly ascribed to the dark inclusions. Although Lee and Van Loenen (1971) interpreted the inclusions, apparently on the basis of field evidence, to be xenoliths of Pioche Shale, it is obvious from figure 6 that they are much closer to the

TABLE 28.—*Values of* h^2 *for the average compositions of dark inclusions and some sedimentary rocks from the southern Snake Range in models containing two to nine factors*

[Line in body of table separates values of h^2 above and below the value of 0.9590 (table 26)]

Rock	Stratigraphic unit	Number of factors in model, *m*							
		2	3	4	5	6	7	8	9
Dark inclusions	Intrusive	0.967	0.968	0.969	0.977	0.985	0.987	0.992	1.000
Argillite	Osceola Argillite of Misch and Hazzard (1962)	.656	.895	.903	.964	.989	.989	.996	1.000
Shale	Pioche Shale	.420	.792	.816	.926	.939	.941	.942	1.000
Quartzite	Prospect Mountain Quartzite	.227	.379	.422	.453	.459	.890	.943	1.000
Limestone	Wheeler limestone (of local usage)	.436	.469	.471	.471	.586	.669	.852	1.000
Limestone	Pole Canyon Limestone	.218	.237	.237	.292	.292	.293	.686	1.000

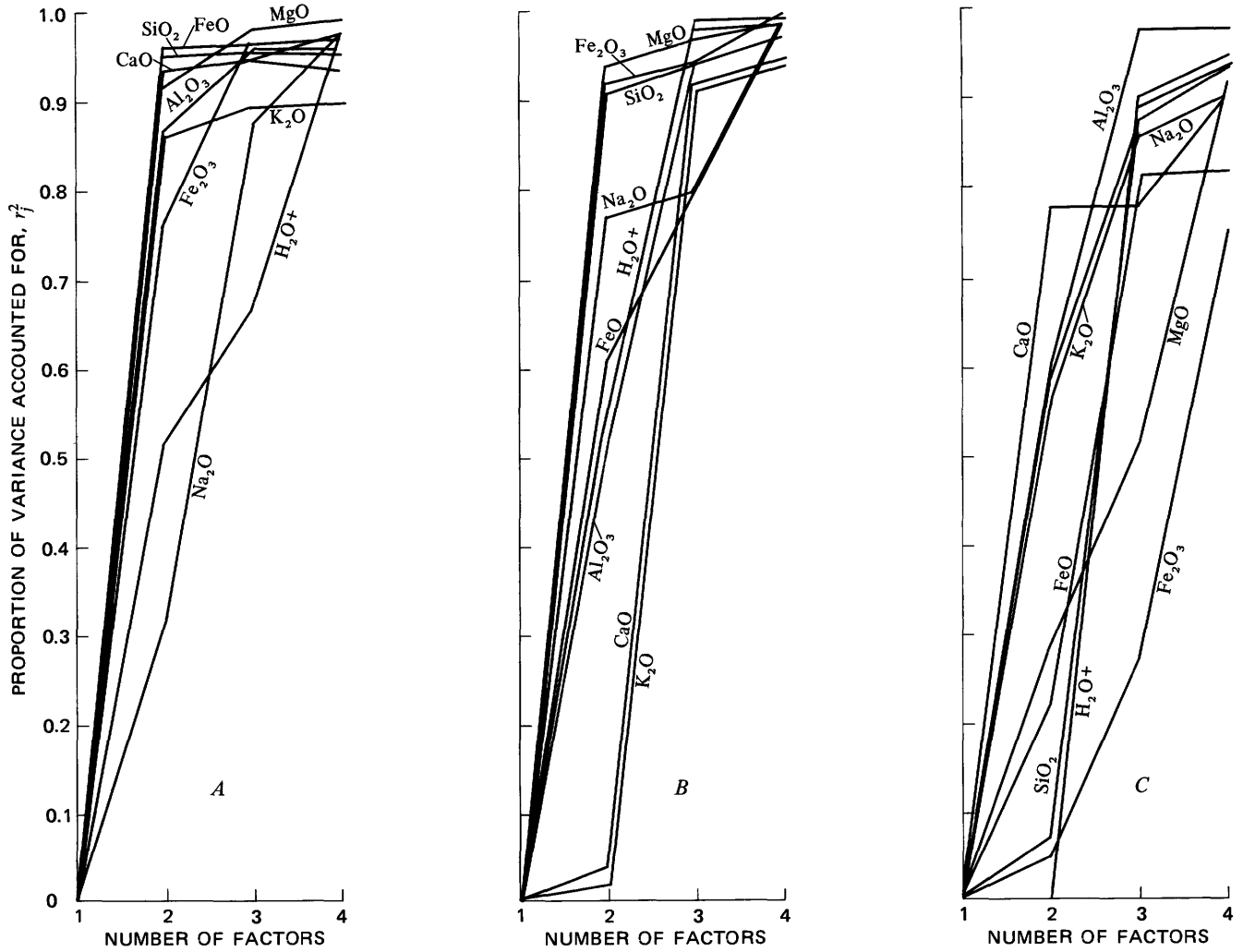

FIGURE 7.—Factor-variance diagrams derived from data obtained by mathematical mixing of sample 1 of the granitoid intrusive and randomly selected samples of (*A*) dark inclusions from the intrusive, (*B*) Osceola Argillite, and (*C*) Pioche Shale.

Osceola Argillite in bulk chemical composition. The source of the inclusions, therefore, appears to be unknown; judging from Lee and Van Loenen's (1971) petrographic descriptions, even an igneous origin is possible.

If the inclusions are igneous, they may represent segregations that were separated from the magma that formed the intrusive. The composition loadings (mixing proportions) for at least some of the samples, in this situation, would be negative with respect to the end-member representing the inclusions. All that can be said of the composition of the magma is that it would be represented by a vector somewhere within or close to the plane of sample vectors represented in figure 6. This vector, and the magma composition, could be fixed more closely only if some assumption were made as to which, if any, samples formed by incorporation of inclusions rather than separation of inclusions from the magma. However, we shall proceed with the assumption that all mixing proportions are positive, as is necessarily the situation if Lee and Van Loenen (1971) are correct in their interpretation that the inclusions are of sedimentary origin.

The composition scores for the vectors represented in figure 6 are given in table 29. It may be seen that vectors lying outside the positive region formed by vectors *A* and *B* have composition scores that are partly negative and that all the vectors representing samples of the intrusive occur in the positive quadrant formed by the varimax axes. If it is assumed that the compositional variation in the intrusive has been due mostly to the mixing of two end-members in positive proportions rather than to the separation of some constituent, one of the end-members must have been of a composition that can be represented by a vector between that of intrusive sample 1 and vector *B* (fig. 6). The other end-member must have been of a composition that could be represented by a vector somewhere between vector *A* and the vector representing sample 84. The composition

TABLE 29.—*Composition scores, $\hat{x}_{ij}$, of some vectors in figure 6 (in percent)*

Vector	SiO_2	Al_2O_3	Fe_2O_3	FeO	MgO	CaO	Na_2O	K_2O	H_2O+
Average Pole Canyon Limestone.	51.19	22.02	4.24	4.99	4.56	9.07	3.94	−1.81	1.80
Average Wheeler Limestone.	53.60	21.20	3.88	4.53	4.13	8.28	3.91	−1.20	1.67
A	58.39	19.55	3.17	3.63	3.29	6.71	3.85	.00	1.41
I	64.36	17.50	2.29	2.51	2.24	4.76	3.77	1.50	1.08
Average dark inclusion..	65.17	17.22	2.17	2.36	2.09	4.50	3.76	1.70	1.03
Intrusive sample 84 .	65.57	17.09	2.11	2.29	2.02	4.36	3.75	1.80	1.01
Intrusive sample 86 .	65.69	17.05	2.09	2.26	2.00	4.33	3.75	1.83	1.00
Average Osceola Argillite ..	65.90	16.98	2.06	2.22	1.97	4.26	3.75	1.88	.99
Average Pioche Shale.....	69.39	15.78	1.54	1.57	1.35	3.11	3.70	2.76	.80
Intrusive sample 7 ..	76.37	13.38	.51	.26	.12	.83	3.61	4.51	.41
Intrusive sample 1 ..	76.42	13.37	.50	.25	.11	.81	3.60	4.52	.41
B	77.05	13.15	.41	.13	.00	.61	3.60	4.68	.38
II	77.56	12.97	.33	.04	−.09	.44	3.59	4.81	.35
Average Prospect Mountain Quartzite	86.67	9.85	−1.02	−1.67	−1.70	−2.54	3.47	7.10	−.15

loadings for the 81 intrusive samples will not vary greatly with any choice of possible end-members within these limits. However, inasmuch as the dark inclusions and intrusive sample 1 are known to exist, the vectors representing these materials were used as reference vectors. That is, the intrusive is interpreted as having formed by assimilation of material similar in composition to the average dark inclusion into a magma similar in composition to intrusive sample 1. The composition scores of the reference vectors are brought together in table 30, where they are interpreted as end-member compositions.

TABLE 30.—*Compositions of the end-members for the Snake Range factor model (in percent)*

End-member	SiO_2	Al_2O_3	Fe_2O_3	FeO	MgO
Magma.....	76.42	13.37	0.50	0.25	0.11
Inclusions.....	65.17	17.22	2.17	2.36	2.09

End-member	CaO	Na_2O	K_2O	H_2O+	Total
Magma	0.81	3.60	4.52	0.41	99.99
Inclusions.....	4.50	3.76	1.70	1.03	100.00

The composition loadings for the 81 intrusive samples with respect to these end-members are given in table 31, along with the increments (eq 36a, 36b) of Na_2O and H_2O+ that must have been added to each sample by independent geochemical processes in order to account for the variabilities in these constituents. Postmultiplication of the matrix of loadings in table 31 by the matrix of scores in table 30, addition of the increments of Na_2O and H_2O+

TABLE 31.—*Composition loadings and increments of Na_2O and H_2O+ for the Snake Range factor model*

Sample	a_{ik}		Increments (in percent)		Sample	a_{ik}		Increments (in percent)	
	Magma	Inclusions	Na_2O	H_2O+		Magma	Inclusions	Na_2O	H_2O+
1	1.000	0.000	0.13	0.44	42	0.593	0.407	-0.58	0.17
2	.966	.034	1.64	−.24	43	.497	.503	−.26	.22
3	.999	.001	.14	.45	44	.593	.407	.57	−.27
4	.940	.060	−.07	−.09	45	.540	.460	.04	−.15
5	.991	.009	.32	−.25	46	.672	.328	−.24	−.38
6	.944	.056	.12	.44	47	.596	.404	−.48	.15
7	.996	.004	.11	−.05	48	.532	.468	.06	.15
8	.998	.002	−.09	.04	49	.564	.436	−.15	.20
9	.928	.072	.10	.03	50	.516	.484	.07	−.07
10	.952	.048	−.01	−.05	51	.583	.417	−.16	−.22
11	.970	.030	−.12	−.04	52	.602	.398	−.15	.01
12	.909	.091	−.34	−.33	53	.634	.367	−.36	.02
13	.881	.119	1.25	.18	54	.589	.411	−.05	−.15
14	.892	.108	−.41	−.30	55	.654	.346	−.24	−.31
15	.764	.237	−.03	.17	56	.545	.455	−.03	.19
16	.806	.194	−.23	−.01	57	.602	.399	.37	.13
17	.793	.207	−.23	−.11	58	.371	.629	−.06	−.43
18	.793	.207	−.35	−.28	59	.604	.396	−.31	−.11
19	.786	.215	−.09	.06	60	.423	.577	.25	.18
20	.739	.261	.18	.01	61	.256	.744	.11	−.24
21	.763	.237	−.22	.00	62	.520	.480	−.37	.04
22	.686	.314	−.05	.19	63	.443	.557	.05	.15
23	.794	.206	−.22	−.08	64	.364	.636	.02	−.37
24	.647	.353	−.03	.22	65	.368	.632	.64	−.20
25	.813	.187	−.02	−.28	66	.413	.587	.15	.15
26	.780	.221	−.23	−.20	67	.505	.495	.16	.26
27	.651	.349	.16	−.23	68	.378	.623	.15	−.25
28	.658	.342	−.14	.04	69	.254	.746	.46	1.43
29	.670	.330	−.45	.07	70	.327	.673	−.05	−.15
30	.659	.341	−.04	−.08	73	.251	.750	−.10	−.22
31	.630	.370	−.01	.30	74	.237	.763	.01	−.13
32	.635	.365	.07	.18	75	.243	.757	−.11	.00
33	.613	.387	−.16	−.06	76	.170	.830	.12	−.30
34	.568	.432	−.06	−.01	78	.129	.871	.33	.47
35	.580	.420	−.03	.19	79	.125	.875	.12	−.17
36	.637	.363	−.35	.26	80	.127	.873	.01	−.25
37	.558	.442	.38	.28	82	.197	.803	.32	−.27
38	.596	.404	.60	−.15	83	.121	.879	−.32	.20
39	.573	.427	−.02	.44	84	.036	.964	.01	.32
40	.448	.552	−.07	.12	86	.046	.954	.11	−.12
41	.563	.437	−.05	−.15					

to the product matrix and, then, adjustment of each row of the product matrix to sum to 100 yield the matrix of reproduced data partially given in table 25*B*.

The correspondence between the original data and the reproduced data, given partially in table 25, is measured by the goodness-of-fit statistics given in table 32. As before, the lack of perfect correspondence is attributed to analytical errors, variation in the compositions of the actual end-members, and other geologic and chemical processes that caused minor compositional variations in the intrusive. It seems likely that these processes included

TABLE 32.—*Goodness-of-fit statistics for the Snake Range factor model*

[$d_{ij} = \hat{x}_{ij} - x_{ij}$; d_j^* = mean of d_{ij}; d_j^{**} = standard deviation of d_{ij}; r_j = correlation coefficient for $\hat{x}_{ij}$ and x_{ij}; r_j^2 = coefficient of determination]

Goodness-of-fit statistic	Variable (*j*)								
	SiO_2	Al_2O_3	Fe_2O_3	FeO	MgO	CaO	Na_2O	K_2O	H_2O+
d_j^*.....	0.06	−0.02	0.00	−0.02	−0.01	0.01	0.00	−0.01	0.00
d_j^{**}.....	.86	.40	.26	.31	.22	.23	.00	.28	.00
r_j	.96	.93	.85	.88	.92	.97	1.00	.93	1.00
r_j^2.....	.93	.87	.72	.78	.85	.94	1.00	.86	1.00

the stoping and assimilation of sedimentary country rock as reported by Lee and Van Loenen (1971). The major variation in the intrusive, however, seems certain to be related to the dark inclusions of unknown source and origin.

LAVAS AND PUMICES OF THE 1959 SUMMIT ERUPTION AT KILAUEA, HAWAII

On November 14, 1959, Kilauea Volcano on the Island of Hawaii began a series of eruptions from Kilauea Iki near its summit, that consisted of 17 phases and lasted until the 20th of December. About 3 weeks later another series of eruptions began on the flank of the volcano and continued almost without interruption for a period of more than a month. The volcanic activity and the detailed chemistry, mineralogy, and petrology of the volcanic products have been described by Murata and Richter (1966), Richter and Moore (1966), Murata (1966), Richter and Murata (1966), and Richter, Eaton, Murata, Ault, and Krivoy (1970). The volcanic products consisted principally of lavas and pumices of tholeiitic basalt. The less siliceous samples of the lavas contain phenocrysts of olivine (Fa_{13}) only, and the more siliceous samples contain clots of clinopyroxene and plagioclase, along with a more ferrous olivine (Fa_{18}) (Richter and Murata, 1966, p. D3—D4). The groundmass portions of the lavas are either glassy or consist of fine-grained plagioclase, clinopyroxene, and opaque minerals.

Murata and Richter (1966, p. A1) summarized their interpretation of the origin of the summit lavas as follows:

> The hottest and most mafic lavas were produced in the 1959 summit phase of the eruption. The compositional variation (7 to 19 percent MgO) among most summit lavas is ascribable solely to fractional crystallization of olivine (Fa_{13}), the first silicate mineral to separate from the primitive magma. A direct relationship between the olivine content of summit lavas and the rate of lava discharge suggests that strong currents of magma erode beds of previously sedimented olivine crystals lying on the bottom of the magma chamber. The coolest, olivine-poor summit lavas contained some small phenocrysts of clinopyroxene and plagioclase as well as olivine. The compositions of these lavas indicated a slight overall accumulation of clinopyroxene, the second mineral to start separating from the cooling magma.

A mathematical study of the compositional variations in the lavas and pumices from the summit eruption was made by Wright (1973), who derived a model that accounts for nearly all of the variation in 11 oxide constituents by the mixture of two magmas and addition of chromite and olivine of varying composition ($Fa_{9.4}$ to $Fa_{13.9}$, or $Fa_{13.1}$ to $Fa_{18.9}$ in weight percents as given by Wright). Wright's model, therefore, contained five end-members: (1) magma A, (2) magma B, (3) an impure chromite, (4) a magnesium-rich olivine (Fa_{10}), and (5) a less magnesium-rich olivine (Fa_{30}). One possible difficulty with the model is that it calls for the addition of olivine and chromite to the parent magmas for all samples, rather than the separation of these minerals by fractional crystallization and settling as might be expected to occur in the upper parts of a magma chamber. As noted in the preceding paragraph, Murata and Richter (1966) regard the separation of olivine as the dominant process that caused compositional variation among the lavas. Another possible difficulty with Wright's model is that, although it includes a chromite end-member, it does not include the more prominent minerals, clinopyroxene and plagioclase.

The factor-variance diagram for the lavas and pumices of the summit eruption (fig. 8) was derived from the analytical data on the same samples as those used by Wright (1973, table 1) which, except for four samples, are from Murata and Richter (1966, table 1). Because of the apparent chemical distinctiveness of samples S – 1 and S – 2, as noted by Wright (1973, p. 850), the data on these samples were not used in the derivation of the factor-variance diagram, but they were later tested for their conformity with the compositional series formed by the other 22 samples. Wright (1973, table 5) also withheld the data on these samples from the data to be explained by the model and used them to represent the compositions of his two end-member magmas. The analyses used in the derivation of figure 8 are given in table 33*A*.

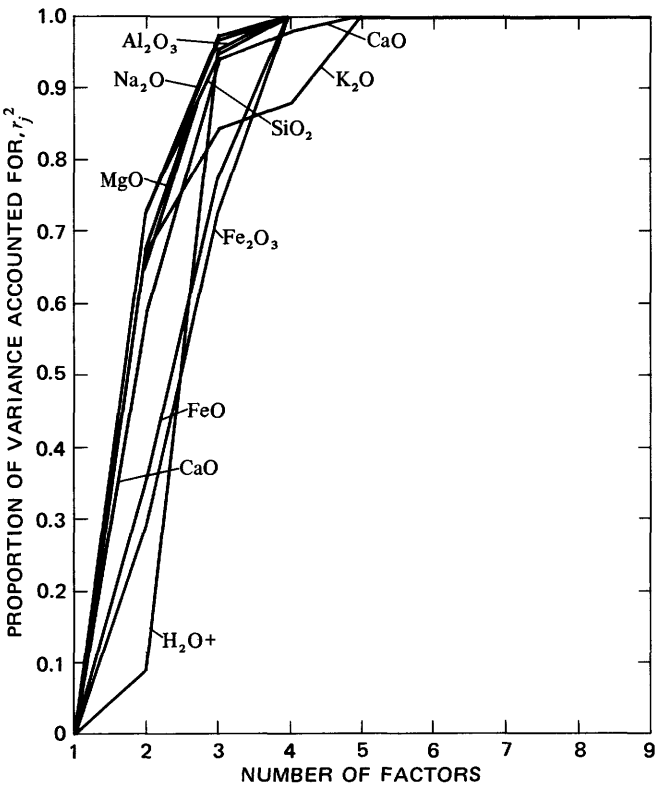

FIGURE 8.—Factor-variance diagram for the lavas and pumices of the 1959 summit eruption at Kilauea, Hawaii.

TABLE 33.—*Analyses of lavas and pumices from the 1959 summit eruption at Kilauea, adjusted so that nine major oxides sum to 100 percent, and corresponding data reproduced from the factor model*

[Analyses from Murata and Richter (1966) and Wright (1973)]

Sample	SiO_2	Al_2O_3	Fe_2O_3	FeO	MgO	CaO	Na_2O	K_2O	H_2O+
A. **Adjusted analyses, x_{ij} (in percent)**									
S–3	51.16	13.22	2.78	9.09	9.12	11.85	2.18	0.52	0.08
S–3A	51.18	13.18	1.88	9.96	9.18	11.87	2.21	.52	.02
S–4	50.42	12.30	1.51	10.37	11.84	10.91	2.11	.48	.05
S–5	47.77	9.74	1.61	10.54	19.98	8.45	1.56	.36	.01
S–7	49.58	11.89	1.40	10.60	14.06	9.95	1.99	.46	.07
S–8	50.77	12.84	1.34	10.51	10.80	10.93	2.16	.56	.09
S–9	49.78	11.95	1.40	10.58	13.72	10.05	2.00	.48	.04
Iki–51	51.08	12.97	1.66	10.26	9.62	11.70	2.18	.51	.03
S–10	50.03	12.46	1.44	10.76	11.96	10.72	2.09	.49	.04
S–11	49.43	11.73	1.47	10.57	14.51	9.87	1.95	.46	.01
S–12	51.29	13.79	2.38	9.55	8.51	11.55	2.30	.57	.06
S–13	51.04	13.65	1.81	10.04	9.21	11.41	2.27	.56	.00
S–14	51.07	13.74	4.39	7.77	8.83	11.39	2.26	.55	.00
S–14A	51.00	13.64	2.74	9.39	8.98	11.38	2.27	.54	.05
S–15	47.95	10.06	1.40	10.78	19.32	8.51	1.61	.37	.00
S–19	48.71	10.77	1.74	10.37	17.07	9.18	1.74	.42	.00
S–20	49.00	11.27	1.44	10.62	15.81	9.48	1.87	.51	.00
S–21	47.88	9.94	1.39	10.70	19.72	8.36	1.59	.42	.00
S–22	48.27	10.12	1.27	10.81	18.75	8.68	1.65	.41	.03
S–24	48.17	10.31	1.85	10.29	18.71	8.64	1.65	.39	.00
S–25	51.08	13.57	2.42	9.58	9.14	11.34	2.25	.55	.07
S–25A	51.00	13.58	2.00	9.96	9.24	11.34	2.22	.56	.09
B. **Reproduced data (in percent)**									
S–3	51.76	14.08	2.08	9.84	7.20	12.10	2.37	0.52	0.06
S–3A	50.84	13.20	2.38	9.64	9.95	11.27	2.19	.52	.02
S–4	50.26	12.46	1.72	10.28	12.02	10.67	2.07	.48	.04
S–5	47.73	9.74	1.33	10.82	20.20	8.26	1.57	.36	–.01
S–7	49.74	11.80	1.23	10.76	13.86	10.13	1.96	.46	.06
S–8	50.69	12.85	1.43	10.50	10.68	11.06	2.16	.56	.08
S–9	49.68	11.84	1.61	10.42	13.85	10.12	1.96	.48	.03
Iki–51	50.68	12.98	2.15	9.86	10.55	11.10	2.16	.51	.03
S–10	50.02	12.21	1.74	10.27	12.76	10.45	2.02	.49	.03
S–11	49.35	11.54	1.83	10.24	14.85	9.83	1.89	.46	.01
S–12	51.42	13.78	2.24	9.71	8.11	11.81	2.31	.57	.04
S–13	50.67	13.08	2.54	9.49	10.34	11.15	2.17	.56	.00
S–14	51.82	14.48	3.63	8.43	6.44	12.30	2.40	.55	–.04
S–14A	51.31	13.69	2.36	9.62	8.45	11.72	2.29	.54	.03
S–15	47.91	9.94	1.36	10.79	19.61	8.43	1.60	.37	–.01
S–19	48.64	10.79	1.72	10.40	17.13	9.16	1.75	.42	–.01
S–20	49.05	11.26	1.85	10.24	15.69	9.57	1.84	.51	.00
S–21	47.95	10.00	1.40	10.74	19.40	8.49	1.61	.42	–.01
S–22	48.24	10.24	1.11	10.98	18.60	8.73	1.67	.41	.02
S–24	48.17	10.26	1.57	10.58	18.69	8.70	1.66	.39	–.01
S–25	51.35	13.66	2.06	9.88	8.44	11.72	2.29	.55	.05
S–25A	51.28	13.52	1.72	10.19	8.75	11.63	2.27	.56	.08

The factor-variance diagram in figure 8 shows that the compositional variation in the lavas and pumices of the 1959 summit eruption at Kilauea can be almost completely accounted for by the mixing, or unmixing, of five factors, or end-members. This is in accord with the fact that Wright's model, containing five end-members, accounts for the observed data almost perfectly. However, the diagram shows further that the final two factors are required mainly to account for about 15 percent of the variance in K_2O, 23 percent of the variance in FeO, and 29 percent of the variance in Fe_2O_3. The FeO and Fe_2O_3 contents of the lavas and pumices have undoubtedly been modified by oxidation after they reached the atmosphere, and no attempt will be made to account for this process in the derived model. This leaves K_2O as the only other variable that a three-factor model would not explain almost entirely. Therefore, a three-factor (end-member) model will be derived and increments of K_2O will be added to, or subtracted from, each sample, as required (eq 35), to account for the additional 15 percent of the variation in this constituent.

The eigenvalues of the cosine theta matrix for the data in table 33*A* are given in table 34 and show that the three-factor models will account for 98.56 percent of the total variability, as a sum of squares, in the normalized data matrix. They also indicate that the average sample communality in the three-factor space is 0.9856.

TABLE 34.—*First nine eigenvalues of the cosine theta matrix for the lavas and pumices of the 1959 summit eruption at Kilauea and CPN*[1]

No.	Eigenvalue	CPN
1	16.868	0.7667
2	4.347	.9643
3	.469	.9856
4	.209	.9951
5	.086	.9990
6	.016	.9997
7	.003	.9999
8	.002	1.0000
9	.001	1.0000
Total	.....22.001	

[1]Cumulative proportions of *N*.

The compositions of samples S – 1 and S – 2 (table 35*A*), used as end-member compositions by Wright (1973), were tested by the method illustrated in parts I and II and were found to have communalities in the three-factor space of 0.9710 and 0.9460, respectively. These values indicate that sample S – 1 conforms with the compositional series of the summit lavas almost as well as the average sample in the series; sample S – 2 does not conform as well. The composition scores for the vectors representing these samples in the three-factor space are given in table 35*B* and, except for iron and MgO, are close to the actual sample compositions.

TABLE 35.—*Compositions of two samples of spatter from the 1959 summit eruption at Kilauea, adjusted so that nine major oxides sum to 100 percent, and composition scores of vectors representing these compositions in the three-factor Kilauea model*

[Analyses from Murata and Richter (1966)]

Sample No.	SiO_2	Al_2O_3	Fe_2O_3	FeO	MgO	CaO	Na_2O	K_2O	H_2O+
A. **Adjusted compositions, x_{ij}, in percent**									
S–1	51.57	13.11	2.64	9.15	8.34	12.32	2.20	0.57	0.09
S–2	51.70	14.15	1.44	10.33	7.47	11.93	2.38	.62	.00
B. **Composition scores, $\hat{x}_{ij}$, in percent**									
S–1	51.93	14.30	2.06	9.82	6.52	12.28	2.42	.58	.09
S–2	50.94	13.37	2.56	9.43	9.51	11.39	2.22	.54	.02

A vector diagram for the three-factor solution is given in figure 9*A*. The diagram was constructed after restoring all represented vectors to unit length. Interpretation of the

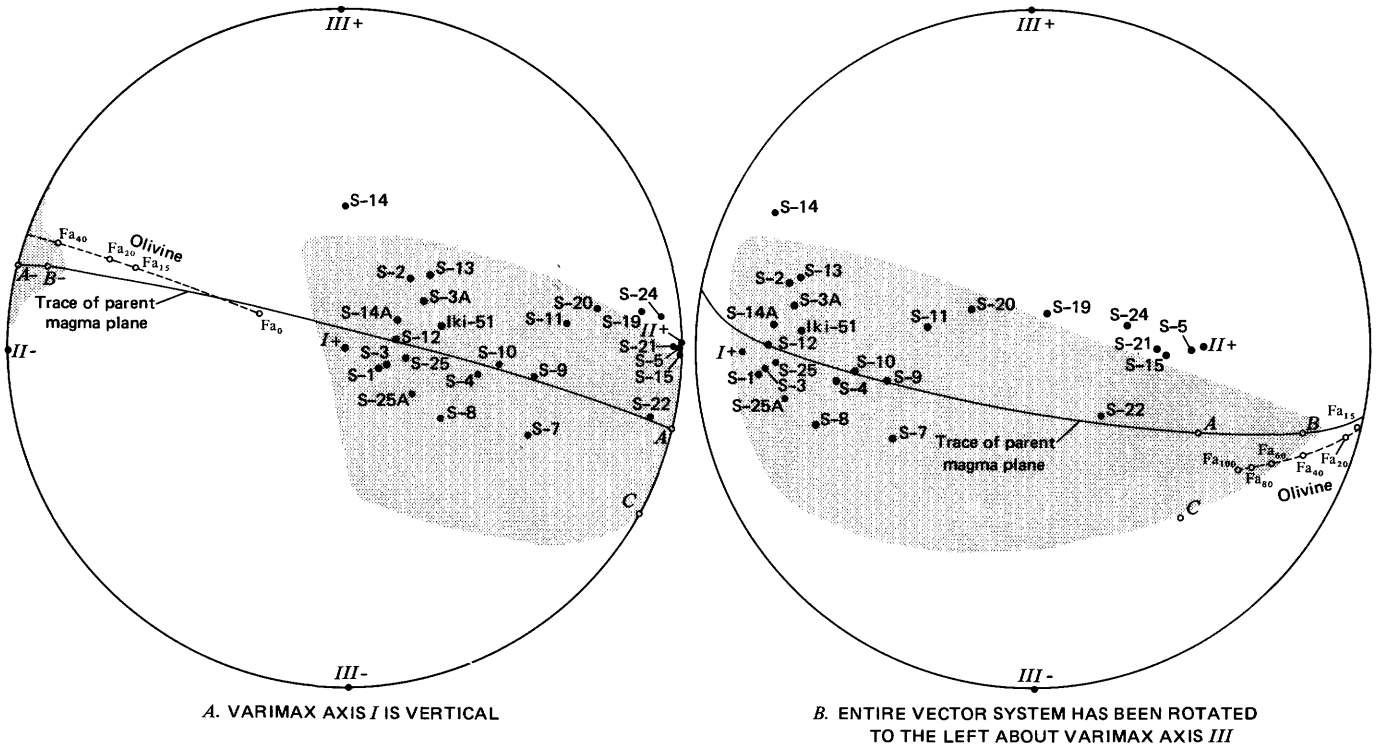

FIGURE 9.—*Q*-mode vector diagrams for lavas and pumices of the 1959 summit eruption at Kilauea, Hawaii. Solid dots with Roman numerals represent the three varimax axes. Solid dots with sample numbers represent sample vectors. Open circles represent vectors discussed in text. Shaded areas indicate the portions of the vector space where vectors have all nonnegative composition scores. All points have been projected from an upper hemisphere in a direction perpendicular to the plane of the diagram.

diagram is similar to interpretation of the stereograms used in structural geology. Points in the diagram represent the intersections of the vectors with the upper surface of a sphere formed by the varimax axes (*I, II,* and *III* in fig. 9). The shaded areas identify the parts of the sphere wherein all contained vectors have composition scores that are entirely nonnegative. All vectors outside of these areas have one or more composition score values that are negative. Thus, any composition represented by a vector far outside of these areas must be regarded as incompatible with the compositional series formed by the 22 lava and pumice samples and not possibly that of one of the three end-members for the factor model.

The small shaded area near the negative end of varimax axis *II* in figure 9*A* reflects a continuation of the larger shaded area at the positive end of varimax axis *II* on the lower half of the sphere. As pointed out in part I, colinear vectors have identical composition scores. In order to show the area of nonnegative composition scores in a contiguous form, the entire vector system was mathematically rotated 60 degrees to the left about varimax axis *III*. The rotated system is shown in figure 9*B*.

The vector configuration in figure 9*B* shows the compositional series of the lavas and pumices as it varies in three dimensions. The major dimension roughly parallels the plane of varimax axes *I* and *II*, and the lesser dimension is across this plane. The general nature of the compositional variation in each direction can be seen by inspection of the composition scores for the three varimax axes given in table 36.

The more siliceous samples are represented by vectors near varimax axis *I* and the more mafic samples by vectors near axis *II*. The composition of sample S - 5 (table 33), for example, is nearly identical to the composition scores for axis *II* (table 36). Variation across the plane of varimax axes *I* and *II* is less pronounced than variation along it and appears to be caused by relatively minor variation in certain constituents, especially Fe_2O_3, FeO, and H_2O+. (See varimax axis *III* in table 26.)

Even though the varimax axes serve as an adequate reference system for describing the sample vector configuration, there is no reason to believe that their composition scores even approximate the compositions of any materials actually involved in the magmatic differentiation. Alternative reference vectors having more probable composition scores must be sought.

In view of the observation of Murata and Richter, (p. G34), regarding the importance of fractional crystallization of olivine (Fa_{13}) in determining the compositional variation in the 1959 summit lavas, olivine

TABLE 36.—*Initial loadings, a''_{ik}, and composition scores, $\hat{x}_{ij}$, for selected vectors in the varimax solution for Kilauea, and compositions of some volcanic materials*

[Leaders (...) indicate no data]

Vector	Initial loadings, a''_{ik}			Composition scores, $\hat{x}_{ij}$, *in percent*								
				SiO_2	Al_2O_3	Fe_2O_3	FeO	MgO	CaO	Na_2O	K_2O	H_2O+
				***A.* Vectors in the varimax solution**								
I	1.0	0.0	0.0	52.80	15.30	2.46	9.39	3.61	13.14	2.60	0.63	0.09
II	.0	1.0	.0	47.66	9.68	1.38	10.81	20.39	8.20	1.55	.38	−.01
III	.0	.0	1.0	49.26	12.59	7.21	5.32	13.43	10.18	1.90	.48	−.36
A	.0	.9682	−.2500	47.52	9.42	.82	11.30	21.01	8.02	1.52	.37	.02
B	−.4146	.8750	−.2500	45.94	7.65	.25	11.94	26.25	6.48	1.20	.29	.01
C	.0	.8660	−.5000	47.31	9.03	.03	12.03	21.94	7.75	1.47	.36	.07

Material	Nature	Compositions, in percent								
		SiO_2	Al_2O_3	Fe_2O_3	FeO	MgO	CaO	Na_2O	K_2O	H_2O+
	***B.* Composition of volcanic materials**									
a	Average oceanite of Macdonald (1968, p. 502)	46.4	8.5	2.5	9.8	20.8	7.4	1.6	0.3	....
b	Crystal-rich fraction separated from lava. Contaminated with glass (Murata and Richter, 1966, table 6, column 1)	45.76	[1]7.59	....	12.20	24.50	6.59	1.22	.28	....

[1] Includes Cr_2O_3.

must be considered as one possible end-member in the three-factor model. The complete series from fayalite to forsterite was tested at increments of one molar percent for compatibility with two- to five-factor solutions; the resultant communalities are shown in figure 10 and indicate that Fa_{15} is the most compatible with the compositional system formed by the lava and pumice samples as represented in a three-factor solution. The computed communality, however, is only 0.9834, and four of the nine composition score values for the representative vector are negative. The positions of vectors representing olivines that range from Fa_0 to Fa_{100} are shown in figure 9; all the magnesium-rich varieties lie outside of the areas of nonnegative composition scores. This and the fact that the iron-rich varieties all have low communalities (fig. 10) are interpreted as indications that the addition or separation of any pure olivine alone cannot help to account for compositional variations in the lavas and pumices of the 1959 summit eruption. Rather, magnesium-rich olivine may have separated from the parent magma along with some other phases. If the amounts of olivine and of other phases that were separated were highly correlated, their combined composition might represent, in effect, that of a single end-member.

The composition scores for vector *B* in figure 9 were computed because this vector lies near one extreme of the elongate area of nonnegative composition scores and because it occurs relatively close to the vector representing the most compatible olivine composition (Fa_{15}). The composition scores for vector *B* are given in table 36, and the normative mineral composition is given in table 37. The

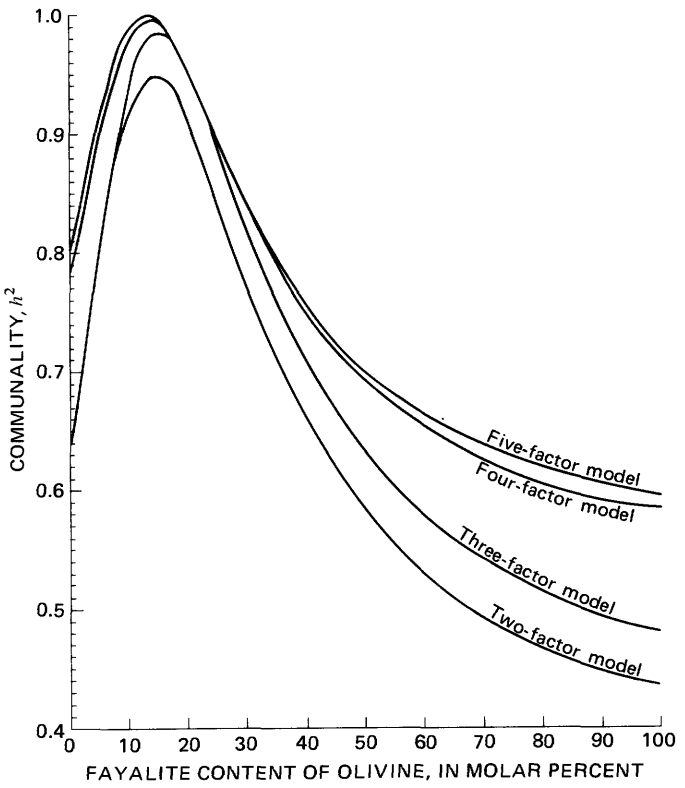

FIGURE 10.—Communalities of olivines in two- to five-factor varimax solutions for the lavas and pumices of the 1959 summit eruption at Kilauea, Hawaii.

composition scores are closely similar to the composition of crystalline material separated from one of the flank lavas by Murata and Richter (1966, p. A20) which

TABLE 37.—*Chemical compositions, f_{kj}, and norms of the end-members for the Kilauea factor model (in percent)*

End-member	SiO_2	Al_2O_3	Fe_2O_3	FeO	MgO	CaO	Na_2O	K_2O	H_2O+	Total
A. Compositions, f_{kj}										
A	47.52	9.42	0.82	11.30	21.01	8.02	1.52	0.37	0.02	100.00
B	45.94	7.65	.25	11.94	26.25	6.48	1.20	.29	.01	100.01
C	47.31	9.03	.03	12.03	21.94	7.75	1.47	.36	.07	99.99

End-member	Or	Ab	An	Di			Hy		Ol		Mt	Total
				Wo	En	Fs	En	Fs	Fo	Fa		
B. Norms												
A	2.19	12.86	17.79	9.20	6.15	2.36	5.75	2.21	28.33	11.98	1.19	100.01
B	1.71	10.15	14.63	7.31	5.05	1.68	3.59	1.19	39.76	14.56	.36	99.99
C	2.13	12.44	16.98	8.97	5.93	2.39	4.39	1.77	31.06	13.83	.04	99.93

contained plagioclase, clinopyroxene, olivine ($Fa_{14.8}$), and contaminations of glass. The composition of the crystalline material is given in table 36*B*. The composition of the normative olivine derived from the composition scores for vector *B* is Fa_{20} (table 37), about the same as the most magnesium-rich normative olivines computed for the summit lavas. (See Murata and Richter, 1966, p. A5.) Vector *B* will be accepted as representing one of the three end-members for the factor model. The end-member is interpreted to have been a mixture that tended to consist of about one-half olivine plus pyroxene and plagioclase in about equal amounts.

The composition scores were computed and were examined for a series of vectors located along the lowermost margin of the shaded area of figure 9, in search of a vector that might represent the glomeroporphyritic clots of clinopyroxene and plagioclase, which are associated with a relatively iron-rich olivine (Fa_{18}), as described by Richter and Murata (1966, p. D4). Such clots were also described by White (1966, p. 270, 308). It was found that vector *C*, as represented in figure 9, has composition scores (table 36) that yield a normative olivine (table 37) of Fa_{24}. The normative composition also shows plagioclase, pyroxene, and about 2 percent orthoclase. This vector was taken as representing the second of the three end-members for the factor model. The principal reason for selecting it was that the normative olivine computed from the composition scores is a few molar percent richer in fayalite than the normative olivine of vector *B*; thus, the first two end-members, although hypothetical and of unknown actual mineral composition, have normative compositions that are at least roughly in accord with the principal mineral assemblages described by Richter and Murata (1966). These are (1) a less fayalitic olivine (Fa_{13}) that is not in equilibrium with the enclosing glass (vector *B*), and (2) a slightly more-fayalitic olivine (Fa_{18}) associated with clots of plagioclase and clinopyroxene (vector *C*). The only major discrepancy apparent is that the first of these end-members (represented by vector *B*) contains normative plagioclase and pyroxene as well as olivine. It is possible, however, that any plagioclase and pyroxene that may have actually separated from the magma along with the olivine either could be present in the fine-grained groundmasses of the lavas or may have settled into a hotter magma and been partially or wholly resorbed.

The third and final end-member for the factor model will represent the parent magma, and its selection will determine the composition loadings for each of the 24 lava and pumice samples—that is, the proportions of each end-member that must be combined with or extracted from the others to approximate the composition of each sample. A positive loading will indicate the net addition of the end-member to the parent magma, even though it may also have been precipitated and separated from the parent to some degree. A negative loading will indicate a net loss of the end-member from the parent magma, even though it may also have been added to some lesser degree. This kind of interpretation is necessary in order to reconcile the model with the concept of a magma chamber that is both precipitating mineral phases at least throughout its upper part, losing these phases by gravitational settling, and receiving precipitated phases from above.

According to the petrographic descriptions of the lava and pumice samples that were given by Richter and Murata (1966, p. D3 - D4), the more fayalitic olivine (Fa_{18}) and clots of plagioclase and pyroxene were found only in samples S - 1, S - 3, and S - 25; they were not found in samples S - 9 or S - 20. Therefore, a parent magma must be selected such that the composition loadings on vector *C* (representing the more fayalitic olivine, plagioclase, and pyroxene) will be positive for samples S - 1, S - 3, and S - 25 and zero or negative for samples S - 9 and S - 20. This requires that the composition of the parent magma be such that it can be represented by a vector that lies in the parent magma plane indicated in figure 9. The plane was drawn through vector *B* so that it separates the vectors representing samples S - 25 and S - 9.

TABLE 38.—*Composition loadings,* a_{ik}, *and increments of* K_2O *for the Kilauea factor model*

Sample No.	a_{ik}			Increment of K_2O (in percent)	Sample No.	a_{ik}			Increment of K_2O (in percent)
	A	*B*	*C*			*A*	*B*	*C*	
S—1	3.386	−2.850	0.464	−0.02	S—12	3.481	−2.461	−0.019	0.00
S—2	3.806	−2.064	−.741	.08	S—13	3.708	−1.890	−.818	.03
S—3	3.344	−2.701	.357	−.07	S—14	5.246	−2.465	−1.781	−.04
S—3A	3.532	−2.020	−.512	−.02	S—14A	3.603	−2.353	−.250	−.02
S—4	2.622	−1.738	.116	−.02	S—15	1.726	−.170	−.556	−.02
S—5	1.660	−.045	−.616	−.02	S—19	2.316	−.617	−.699	.00
S—7	1.929	−1.461	.532	−.02	S—20	2.553	−.898	−.655	.07
S—8	2.355	−2.107	.753	.03	S—21	1.787	−.201	−.587	.03
S—9	2.380	−1.363	−.017	.00	S—22	1.499	−.452	−.046	.00
Iki—51	3.228	−1.951	−.276	−.02	S—24	2.032	−.315	−.717	−.01
S—10	2.606	−1.569	−.037	.00	S—25	3.250	−2.432	.182	−.02
S—11	2.585	−1.087	−.499	.00	S—25A	2.822	−2.452	.630	.00

Selection of the vector in the parent magma plane that may represent the composition of the parent magma is difficult, but I assumed that the lavas and pumices were erupted from the upper parts of the magma chamber and that each sample formed from parent magma that had undergone at least some net loss in the less fayalitic olivine. Thus, the parent magma vector must be to the right (toward the positive end of varimax axis *II*) of a plane through vector *C* and the vector representing sample S - 5. Vector *A* (fig. 9) was selected for this reason, but its composition scores (table 36) are, for the most part, strikingly similar to the average composition of oceanites as given by Macdonald (1968, p. 502). The only notable exceptions are the score values for Fe_2O_3 and FeO (compare vector *A* and material "a" in table 36), but the total iron values are in almost perfect accord. Macdonald (1968, p. 511) judged oceanite to be the most likely single parent for the Hawaiian lavas.

The composition loadings for all samples with respect to reference vectors *A*, *B*, and *C* in figure 9 are given in table 38; also included are the increments of K_2O that must be added to each sample to account for the additional 15 percent of the variability in this constituent. Sample S - 1, for example, is viewed as having originated from 3.386 parts original magma (vector *A*) from which 2.850 parts of the less fayalitic olivine as well as plagioclase and pyroxene (vector *B*) have been subtracted and 0.464 parts of the more fayalitic olivine plus plagioclase and pyroxene (vector *C*) have been added. The sum of these three loadings is unity; the source of the 3.386 parts original magma, presumably, was a lower part of the magma chamber where magma was displaced upwards by settling crystalline phases from above. The magnitudes of the loadings on the vector representing the parent magma (vector *A* in table 38) indicate that the model requires a magma chamber at least several times the volume of the erupted materials.

Combination of the three end-members represented in table 37 in the proportions indicated in table 38, addition of the corresponding increment of K_2O, and minor adjustment of each analysis to sum to 100 yield the reproduced data given in table 33*B*. Correspondence of the reproduced data to the original data given in table 33*A* is measured by the goodness-of-fit statistics given in table 39. The values of r_j^2 show that the model explains more than 90 percent of the variance in all constituents except Fe_2O_3 and FeO.

TABLE 39.—*Goodness-of-fit statistics for the Kilauea factor model*

[$d_{ij} = \hat{x}_{ij} - x_{ij}$; d_j^* = mean of d_{ij}; d_j^{**} = standard deviation of d_{ij}; r_j = correlation coefficient for $\hat{x}_{ij}$ and x_{ij}; r_j^2 = coefficient of determination]

Goodness-of-fit statistic	Variable (*j*)								
	SiO_2	Al_2O_3	Fe_2O_3	FeO	MgO	CaO	Na_2O	K_2O	H_2O+
d_j^*	0.04	0.03	−0.01	0.02	−0.11	0.04	0.00	0.00	0.01
d_j^{**}	.28	.29	.39	.35	.82	.33	.06	.00	.01
r_j	.98	.98	.84	.88	.98	.97	.97	1.000	.96
r_j^2	.96	.96	.71	.77	.97	.94	.95	1.000	.93

LAYERED SERIES OF THE SKAERGAARD INTRUSION, GREENLAND

The Skaergaard intrusion in eastern Greenland has been the object of some classical studies in igneous petrology by L. R. Wager and his colleagues, and it is generally acknowledged as having attained its compositional variation almost entirely by fractional crystallization of a basaltic magma and by redistribution of crystals through settling and convection. The geology, petrology, and chemistry of the intrusion have been described in detail by Wager and Brown (1968), who also tabulated selected analytical data. The intrusion is exposed over an area of 50 km^2 and is interpreted to have the form of an inverted cone. Most of it consists of a layered series that formed by gravity stratification of plagioclase, pyroxene, and olivine. The layered series has been subdivided into zones and subzones on the basis of cumulus minerals, those that precipitated and collected at the bottom of the magma. The three zones of the exposed part of the layered series are a lower zone (LZ) containing cumulus olivine, a middle zone (MZ) in which cumulus magnesian olivine is absent,

and an upper zone (UZ) containing cumulus olivine that is richer in iron than that in the lower zone (Wager and Brown, 1968, p. 30). Wager and Brown estimated that 70 percent of the layered series is hidden beneath the part that is exposed (1968, p. 204).

The factor-variance diagram derived from 19 analyses of specimens from the layered series (table 40) is given in figure 11. It shows that a petrologic model with five end-members could account for more than 85 percent of the variance in each compositional variable; each addition of an end-member beyond five will result in only small increases in the accountable variance for each constituent. However, an extensive search for five end-members failed to reveal a set that could be accepted as geologically plausible. Similarly, a search was made for sets of six to eight end-members. Again, no models that seemed geologically acceptable could be found. The most common reason for the rejection of models containing five to eight factors was the appearance of negative composition loadings on the reference vector representing the composition of the parent magma, which would indicate that some samples formed by subtraction of the parent magma from crystallizing minerals. The composition of the parent magma was taken as that of specimen 4507 of Wager and Brown (1968, p. 150 – 151 and table 4) with the value for H_2O+ set to 0.25 percent (Wager and Brown, 1968, p. 192), and the composition was adjusted so that the nine constituents summed to 100 percent.

TABLE 40.—*Analyses, x_{ij}, of samples from the layered series of the Skaergaard intrusion, adjusted so that nine major oxides sum to 100 percent*

[Analyses from Wager and Brown (1968, table 5 and fig. 15)]

Zone of layered series and sample Nos.	x_{ij}								
	SiO_2	Al_2O_3	Fe_2O_3	FeO	MgO	CaO	Na_2O	K_2O	H_2O+
UZc:									
5166	50.83	9.61	3.80	21.57	0.09	9.29	3.19	0.63	0.99
1881	50.00	8.89	4.21	23.71	1.25	7.69	2.74	.35	1.17
4139	46.55	9.65	5.95	24.48	.44	9.32	2.50	.50	.59
4142	46.02	8.22	4.22	27.77	.26	10.46	2.24	.49	.31
1974	45.63	8.04	4.15	26.12	.39	10.92	2.88	.33	1.55
UZb:									
4145	46.68	12.39	2.17	24.01	1.81	9.22	3.12	.37	.23
4272	45.58	11.20	4.75	19.30	4.15	10.98	3.59	.32	.12
UZa:									
5181	47.62	15.29	2.97	16.48	4.35	8.73	3.86	.28	.42
1907	46.07	14.35	3.86	17.13	5.70	8.77	3.44	.34	.35
5322	29.39	1.71	8.68	44.60	10.88	4.16	.21	.06	.32
5321	54.65	22.44	2.47	5.38	.51	8.26	5.69	.50	.09
MZ:									
3661	46.82	15.47	3.50	15.24	6.51	9.41	2.54	.29	.23
3662	49.43	18.50	2.59	9.75	5.39	10.44	3.55	.14	.21
LZa:									
2308	24.43	5.63	15.60	37.60	9.23	6.49	.61	.07	.33
2307	47.33	13.75	5.16	13.70	6.59	10.67	2.34	.23	.22
LZb:									
4077	46.84	16.99	1.54	10.55	9.71	11.41	2.48	.20	.29
LZc:									
5109	50.98	25.71	1.12	4.10	1.91	11.82	3.56	.36	.44
5108	40.44	3.36	2.73	23.39	26.35	2.34	.49	.08	.83
4087	46.18	16.67	2.12	9.43	11.83	10.62	2.09	.27	.78

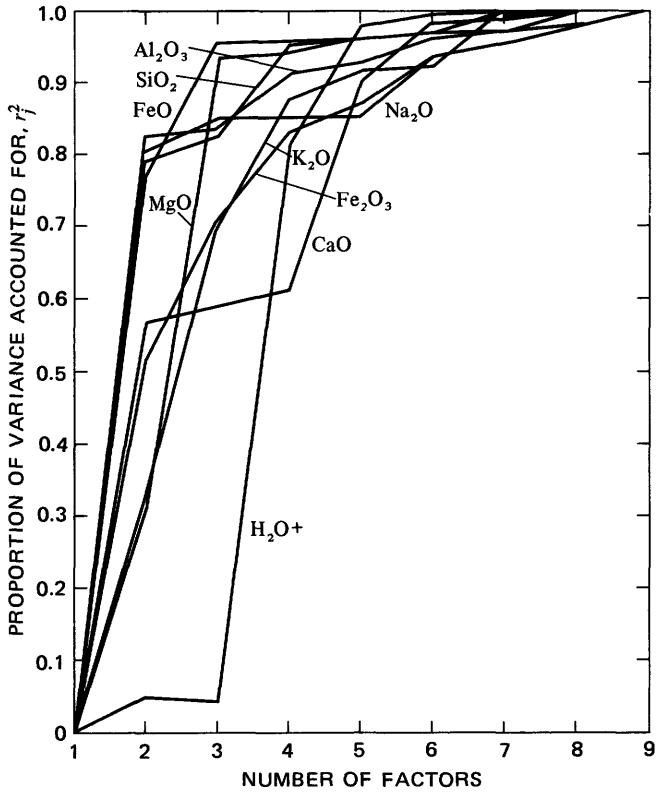

FIGURE 11.—Factor-variance diagram for the layered series of the Skaergaard intrusion.

The only geologically acceptable factor model that could be derived contained nine end-members, equal to the number of compositional variables used in the analysis. The disadvantage of a nine-factor model is that none of the variance in each constituent can be attributed to either analytical error, to compositional variation within each end-member, or to other geologic and petrologic processes not included in the model. Although the resultant model accounts for the observed data exactly, the three other general sources of variation are almost certain to have exerted at least a small amount of influence on the variation in the data. Consequently, the exactness with which the observed data are reproduced is, to some degree, a false indication of the correctness of the model. Nevertheless, the model can be expected to at least approximate a general process by which the Skaergaard intrusion may have formed.

The eigenvalues of the cosine theta matrix for the data in table 40 are given in table 41; the corresponding values of CPN when plotted against the number of factors display a smooth curve, having no prominent breaks that might indicate the proper number of end-members to include in the model.

The concept of the general process by which the layered series of the Skaergaard intrusion formed is that of a parent magma which was modified in composition during

TABLE 41.—*First nine eigenvalues of the cosine theta matrix for the layered series of the Skaergaard intrusion and* CPN [1]

No.	Eigenvalue	CPN
1	15.240	0.8021
2	1.935	.9040
3	.688	.9402
4	.613	.9724
5	.241	.9851
6	.131	.9920
7	.091	.9968
8	.038	.9988
9	.022	1.0000
Total.....	18.999	

[1]Cumulative proportions of *N*.

the process of solidification from the bottom of the magma chamber upward. Modification resulted from the crystallization of mineral phases and redistribution of these phases throughout the chamber by gravity settling and convection (Wager and Brown, 1968, p. 101). The problem of deriving a factor model, therefore, is one of finding mineral phases that (1) are conformable with the compositional series formed by the samples from the layered series and (2) can be represented by reference vectors on which the sample vectors have composition loadings that are subject to plausible interpretation. Above all, as stated previously, the loadings on the vector representing the composition of the parent magma must be all nonnegative. As pointed out in part I, however, when the number of factors in the model equals the number of compositional variables ($m = M$), any combination of the variables fits the model perfectly (all communalities are exactly one), and testing of mineral phases for conformity with the samples is impossible. Because of this, the mineral phases were tested within eight-factor space, even though the phases found to be satisfactory were then used in the nine-factor model. The validity of this procedure seemed apparent from the examination of the feldspar system.

The feldspar system of albite-anorthite-orthoclase was examined throughout at equal increments of 2 molar percent. That is, 1,326 compositions within this system were tested using a computer program based on the procedures given in parts I and II. Of these, only 54 showed communalities equal to or greater than the average sample communality (0.9988) for the eight-factor solution (table 41). Nearly all of these 54 feldspar compositions included 2 molar percent orthoclase (fig. 12). This finding is in excellent accord with the data and the observations of Wager and Mitchell (1951, p. 146 - 147), who found, by analysis of three plagioclase specimens, 2 molar percent orthoclase in two specimens and 3 molar percent orthoclase in the third. They observed further that during fractionation of the magma the change in the K_2O content of the plagioclase was small. Two of the end-members for

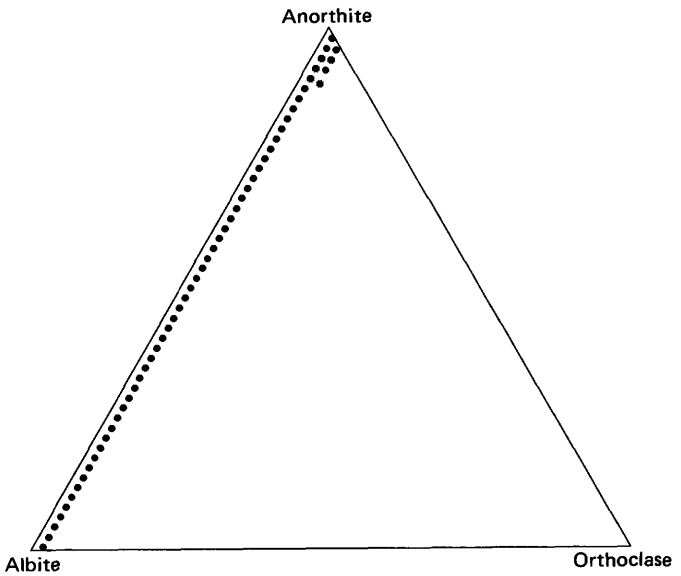

FIGURE 12.—Molar compositions in the system albite-anorthite-orthoclase that have communalities in excess of 0.9988 in the eight-factor varimax solution for the layered series of the Skaergaard intrusion.

the nine-factor model are taken as $Ab_{98}Or_{02}$ and $An_{98}Or_{02}$.

Cumulus olivines from the layered series vary in composition from Fa_{33} to almost pure fayalite (Wager and Brown, 1968, p. 28). Testing of the complete range of olivine compositions throughout the series, at increments of one molar percent, showed that all compositions have high communalities in the eight-factor space (fig. 13). Therefore, two additional end-members for the nine-factor model were taken as pure forsterite and pure fayalite.

The most abundant pyroxenes in the layered series are

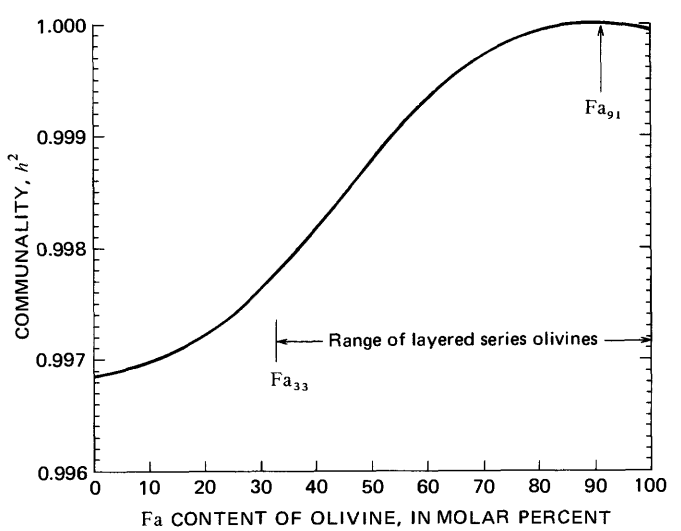

FIGURE 13.—Communalities of olivines in the eight-factor varimax solution for the layered series of the Skaergaard intrusion.

augites that vary in composition between diopside ($CaMg(SiO_3)_2$) and hedenbergite ($CaFe(SiO_3)_2$). An Mg-Ca-Fe diagram given by Wager and Brown (1968, p. 39), however, shows that these ideal compositions are not those of the actual end-members in the augite series. The compositional series formed by five augite specimens for which analyses are available (Wager and Brown, 1968, p. 42) was examined by means of a special *Q*-mode analysis using the same procedures described throughout this report. The two-factor varimax solution is represented by the vector diagram given in figure 14 and shows that the augites can be well represented as a two-component system—that is, all vectors representing pyroxene specimens are of near unit length. The theoretical extreme end-members in the system are represented by the two extreme vectors with all nonnegative composition scores—vectors *A* and *B* in figure 14. Because vector *A* is close to vectors 13 and 15 (representing augite specimens 13 and 15 of Wager and Brown), no particular advantage is gained by using the theoretical composition that vector *A* represents as an end-member composition in the factor model. Wager and Brown (1968, p. 39) regarded specimen 15 as the extreme in the series and suggested that the composition of specimen 13 had been affected by processes of exsolution. Consequently, the compositions of the end-members for the augite series, used in the nine-factor model for the layered series, are taken as the composition of augite specimen 15 adjusted to sum to 100 and as the composition scores for vector *B* in figure 14. The end-members are referred to, respectively, as Fe-augite and Mg-augite, and their compositions are given in table 42.

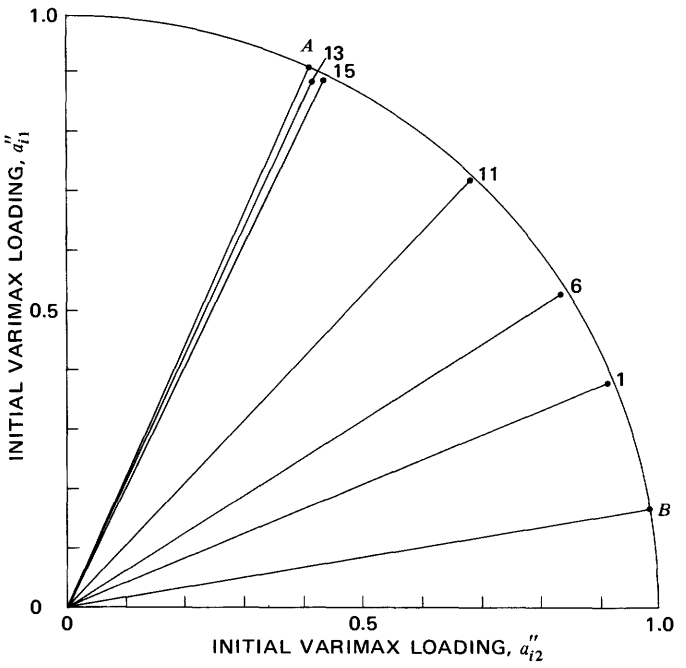

FIGURE 14.—*Q*-mode vector diagram for a series of augite specimens from the layered series of the Skaergaard intrusion. Vectors *A* and *B* represent, respectively, the theoretical Fe- and Mg-augite extremes. Numbers on other vectors refer to specimen numbers of Wager and Brown (1968, p. 42).

Calcium-poor pyroxenes are far less abundant than augites in the layered series (Wager and Brown, 1968, table 5) but are present there as well as in the border phases of the intrusion (Wager and Brown, 1968, p. 42). Analyses of two Ca-poor pyroxenes, a bronzite from the border phase and an inverted pigeonite from the middle zone of the layered series, are given by Wager and Brown (1968, p. 42), and the average, adjusted to sum to 100, is taken as the composition of another end-member for the nine-factor model (table 42).

TABLE 42.—*Compositions, f_{kj}, of end-members for the Skaergaard factor model*

End-member	f_{kj} (in percent)								
	SiO_2	Al_2O_3	Fe_2O_3	FeO	MgO	CaO	Na_2O	K_2O	H_2O+
Parent magma[1]	49.10	17.58	1.35	8.62	8.80	11.62	2.42	0.26	0.25
Forsterite[2]	42.69	0	0	0	57.31	0	0	0	0
Fayalite[2]	29.48	0	0	70.52	0	0	0	0	0
Ca-poor pyroxene[3]	52.70	2.10	1.00	18.10	22.70	3.20	.20	0	0
Mg-augite[4]	51.68	3.33	1.52	.48	22.16	20.21	.59	.03	0
Fe-augite[4]	47.23	.94	0.60	31.83	.14	18.96	.26	.03	0
Iron ore[5]	0	0	31.74	68.26	0	0	0	0	0
Anorthite[6]	43.62	36.28	0	0	0	19.76	0	.34	0
Albite[6]	68.65	19.42	0	0	0	0	11.57	.36	0

[1] Sample 4507 of Wagner and Brown (1968, table 7) with H_2O+ set to 0.25 percent, adjusted sum to 100.
[2] Theoretical.
[3] Average of two analyses from Wager and Brown (1968, table 1), adjusted to sum to 100.
[4] The composition of the Mg-augite is the theoretical extreme in the series of five analyses given by Wager and Brown (1968, table 1). The composition of the Fe-augite is taken as the analysis of augite specimen 15 from Wager and Brown (1968, table 1), adjusted to sum to 100.
[5] Theoretical mixture of iron ores. (See text.)
[6] Anorthite is theoretical $An_{98}Or_{02}$. Albite is theoretical $Ab_{98}Or_{02}$.

The reference vectors chosen thus far for the nine-factor model represent the parent magma, two extremes in the plagioclase series, two extremes in the olivine series, two extremes in the augite system, and a Ca-poor pyroxene. It seems that the remaining reference vector should represent the composition of the iron-ore minerals, principally ilmenite and magnetite, which constitute the only other cumulus mineral group of quantitative importance throughout the layered series. Although ilmenite is thought to have formed a cumulus mineral phase slightly before magnetite in the differentiation process, ilmenite and titaniferous magnetite tend to be found together (Wager and Brown, 1968, p. 49). It is appropriate, therefore, to represent these minerals by a single reference axis in the factor model. The composition of magnetite was taken as 69 percent Fe_2O_3 and 31 percent FeO; because the data for TiO_2 were not used in this exercise, the composition of ilmenite in terms of the constituents that were used is 100 percent FeO. Neither of these compositions has a high communality in the eight-factor space for the layered series rocks, but, as shown in figure 15, a mixture of 46 percent magnetite and 54 percent FeO has a communality of

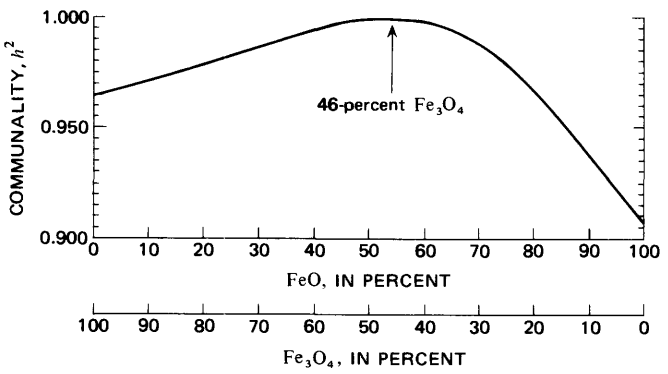

FIGURE 15.—Communalities of theoretical mixtures of ferrous iron oxide and magnetite in the eight-factor varimax solution for the layered series of the Skaergaard intrusion.

almost precisely 1 ($h^2 > 0.9999$). Using the ideal compositions of magnetite and ilmenite, this mixture is that present in a cluster consisting of 29 percent magnetite and 71 percent ilmenite, having a chemical composition of about 32 percent Fe_2O_3 and 68 percent FeO. This chemical composition is accepted as that of the final end-member for the Skaergaard model (table 42).

Composition loadings for the 19 samples from the layered series with respect to the nine selected reference vectors are given in table 43. The loadings on the vector representing the parent magma range from about 0.4 to 6.2, indicating that the model calls for some samples to have formed with a net addition of the constituents of the cumulus minerals to the parent magma ($a_{i1} < 1.0$) and others to have formed with a net loss of these constituents ($a_{i1} > 1.0$). The greatest losses of the constituents of the cumulus minerals, according to the model, were from the uppermost parts of the upper zone (UZc) and from the lowermost parts of the lower zone (LZa). As the cumulus minerals precipitated and either settled or were swept away by convection currents, they were replaced by influxes of magma from elsewhere in the magma chamber. Some of this additional magma, according to the model, must have been derived from outside of the volume occupied by the exposed layered series, because the exposed layered series as a whole requires about 1.454 parts parent magma (table 43*C*). This value is the weighted average of the composition loadings on the parent material reference vector, using the volume weights for the zones and subzones of the layered series given by Chayes (1970, p. 4). The most likely source of the additional parent material is the part of the magma chamber now occupied by the hidden part of the layered series.

The composition loadings of the 19 layered series samples on the vectors representing cumulus minerals (table 43) should not be interpreted as gains and losses due to precipitation of the end-members in the following two-component compositional series: forsterite-fayalite,

TABLE 43.—*Composition loadings, a_{ik}, for the Skaergaard factor model, average loadings for each zone of the layered series, and weighted averages for all zones*

Zone of layered series[1] and sample Nos.	Forsterite		Ca-poor pyroxene		Fe-augite	Anorthite		
	Parent magma		Fayalite	Mg-augite		Iron ore		Albite
				A. Composition loadings, a_{ik}				
UZc:								
5166	3.975	3.171	−3.922	−1.029	−8.776	8.484	0.244	−0.864	−0.283
1881	4.686	−1.132	.577	.198	.931	−1.747	−.085	−1.674	−.754
4139	2.344	3.026	−3.630	−.775	−7.995	7.999	.345	−.280	−.035
4142	1.257	4.020	−4.442	−1.048	−9.876	10.276	.392	.201	.219
1974	6.204	−2.814	2.459	.017	4.851	−5.840	−.256	−2.457	−1.164
UZb:								
4145	.933	1.992	−2.035	−.626	−4.830	5.043	.185	.122	.218
4272	.469	1.741	−1.794	−.640	−3.873	4.332	.254	.190	.323
UZa:								
5181	1.669	−.385	.404	−.100	.637	−.803	.010	−.405	−.028
1907	1.400	.898	−.946	−.392	−2.234	2.212	.140	−.153	.075
5322	1.280	−.480	.845	.084	1.146	−1.280	.185	−.500	−.281
5321	.378	2.507	−2.787	−.624	−6.009	6.144	.253	.548	.590
MZ:								
3661	.907	.683	−.771	−.096	−1.746	1.753	.125	.063	.081
3662	.825	−1.871	1.957	.485	4.286	−4.396	−.092	−.201	.007
LZc:								
2308	1.322	−.931	1.017	.119	2.193	−2.325	.370	−.480	−.284
2307	.886	.172	−.400	.132	−.639	.714	.138	−.034	.031
LZb:								
4077	1.171	−.878	1.043	.088	2.168	−2.238	−.066	−.196	−.092
LZa:								
5109	1.772	−.653	.520	.041	1.036	−1.402	−.065	−.166	−.084
5108	3.301	−1.946	2.341	.466	4.467	−5.283	−.184	−1.398	−.764
4087	3.123	−1.457	1.459	.092	2.990	−3.567	−.145	−.950	−.545
				B. Zone averages				
UZc....	3.693	1.255	−1.792	−0.527	−4.173	3.834	0.128	−1.015	−0.404
UZb....	.701	1.867	−1.915	−.633	−4.352	4.687	.220	.156	.270
UZa....	1.182	.635	−.621	−.258	−1.615	1.568	.147	−.127	.089
MZ	.866	−.594	.593	.195	1.270	−1.322	.017	−.069	.044
LZ.....	1.929	−.949	−.997	.156	2.036	−2.350	.008	−.537	−.290
				C. Weighted averages for all zones [2]				
.....	1.454	−0.266	0.248	0.014	0.425	−0.552	0.051	−0.288	−0.085

[1] From Wager and Brown (1968, p. 15 and table 5).
[2] Weights used for UZc, UZb, UZa, MZ, and LZ were, respectively, 6.0, 11.4, 6.2, 37.1, and 39.3 (Chayes, 1970, p 4).

Mg-augite – Fe-augite, and anorthite-albite. Rather, the loadings indicate the net effects of the precipitation of species of varying composition within each of these series. For example, the first sample (No. 5166) in table 43 formed from the parent magma partly by the precipitation of olivine of varying composition and partly by the concentration of the constituents of olivine in the residual magma; and so the net effect was the gain of 3.171 parts forsterite and the loss of 3.922 parts fayalite. This amounts to a net loss of 0.751 parts olivine, even though the constituents of olivine were both gained and lost during the course of differentiation. A detailed interpretation of the chemical and mineralogical changes in the parent magma that produced sample 5166 is given in table 44. It shows that the precipitation of olivine led to a net loss of FeO and net gains in MgO and SiO_2, even though there are no means of specifying the compositions of individual olivine crystals that were added to and lost from the parent magma. Similarly, the precipitation of compositions in the

TABLE 44.—*Interpretation of chemical and mineralogical modifications of the parent magma required to produce sample 5166 from the Skaergaard intrusion*

Minerals and mineral groups	SiO_2	Al_2O_3	Fe_2O_3	FeO	MgO	CaO	Na_2O	K_2O	H_2O+	Total
***A.* 3.9752 parts of parent magma**										
	1.9519	0.6988	0.0537	0.3427	0.3498	0.4619	0.0962	0.0103	0.0099	3.9752
***B.* Gains and losses by precipitation of minerals and mineral groups**										
Olivine	0.1975	0	0	−2.7659	1.8174	0	0	0	0	−0.7510
Ca-poor pyroxene	−.5420	−.0216	−.0103	−.1862	−.2335	−.0329	−.0021	0	0	−1.0286
Augite .	−.5277	−.2125	−.0828	2.6586	−1.9330	−.1652	−.0295	−.0001	0	−.2922
Iron ore	0	0	.0775	.1666	0	0	0	0	0	.2441
Plagioclase ...	−.5716	−.3686	0	0	0	−.1708	−.0328	−.0040	0	−1.1478
Total	−1.4438	−.6027	−.0156	−.1269	−.3491	−.3689	−.0644	−.0041	0	−2.9755
***C.* Net results as percentages[1]**										
	50.81	9.61	3.81	21.58	0.07	9.30	3.18	0.62	0.99	99.97

[1] Compare with analyses of sample 5166 given in table 40. Subtract total in *B* from corresponding value in *A* and multiply by 100.

augite series caused a net loss of all chemical constituents except FeO. The net gain in FeO (table 44) may have resulted from the introduction of Fe-rich augite or its dissolved constituents into that part of the magma chamber at a late stage in the differentiation process, whereas augites that formed earlier and settled out were Mg-rich. The composition loadings for anorthite and albite show that the parent magma underwent a net loss of plagioclase in the formation of sample 5166, even though it is possible that some plagioclase, of undetermined composition and in either crystalline or dissolved form, was introduced into that part of the magma chamber sometime during the differentiation process. In fact, Wager and Brown (1968, table 5) reported the presence of cumulus plagioclase in this sample. The net result of the introduction and loss of crystalline or dissolved plagioclase from that part of the magma, however, is interpreted, on the basis of table 44, to have been a loss in all the constituents of plagioclase.

Interpretation of the positive composition loadings of table 43 as proportions of mineral phases in dissolved as well as crystalline form is important and necessary. For example, the composition loadings for sample 5166 on the iron-ore end-member is +0.244, indicating that more than 24 percent iron ore (comprised of 32 percent Fe_2O_3 and 68 percent FeO, table 42) has been added to this sample by the differentiation process. However, Wager and Brown (1968, table 5) reported only 6 percent iron ore in sample 5166. If the iron ore had been added entirely in crystalline form, it is unlikely that it could have been resorbed into the magma, and it should be present in the rock. If added at least partly in dissolved form, however, it could provide the iron needed for the formation and precipitation of olivine and pyroxene (tables 43, 44). Thus, some of the positive composition loadings on mineral end-members are interpreted to reflect residual concentration of the constituents of these minerals in the magma. Residual concentration of iron in the Skaergaard magma has been firmly established (Wager and Brown, 1968, p. 242).

Several features of the model presented in tables 42 and 43 support its creditability, and others must be rationalized. Notable among the latter are that (1) the model does not account for the effects of other cumulus minerals present in minor amounts, such as apatite, and (2) it leaves none of the variance in each constituent to be attributed to the effects of analytical error, to minor petrologic processes, or to compositional variation within each of the end-member mineral phases. Regarding variation in the end-member mineral phases, the compositions of the Ca-poor pyroxenes in the Skaergaard intrusion, for example, are not perfectly represented by the end-member composition given in table 42, and the olivine compositions are not perfectly explained as mixtures of ideal forsterite and fayalite. At best, the model can be accepted as only an approximation of the effects of magmatic differentiation on each sample, in spite of the fact that it accounts for the observed data exactly.

On the other hand, the eight end-member compositions representing the average compositions of Ca-poor pyroxene, the iron-ore minerals, and the end-members in the olivine, augite, and plagioclase series are at least good approximations of the compositions certain to have been involved in the differentiation of the layered series. The least certain end-member composition is that of the parent magma, which was taken as interpreted by Wager and Brown (1968). The model, therefore, may be viewed as a

test of their interpretation, and the test shows no evidence that they were wrong.

If the model is accepted as approximately correct, the net gains and losses of the constituents of each mineral in the parent magma can be shown for each sample. In figure 16 the gains and losses are plotted against the respective structural elevations of each sample in the layered series. The plots show that the constituents of olivine and Ca-poor pyroxene were largely lost from the parent magma in the upper parts of the magma chamber and that they were both lost from and added to the parent magma in the lower parts. Addition is interpreted to have been by crystal settling from upper parts of the chamber and by the transfer of crystalline material and magma with convection. Augite and plagioclase were mostly lost from the parent magma in the lower parts of the magma chamber, but they were both lost from and added to the upper parts. Inspection of the relative loadings on the Mg-augite and Fe-augite end-members (table 43) indicates that the added augites tended to be rich in iron. Similarly, the added plagioclases tended to be richer in sodium than in calcium. The addition of augite and plagiolcase, therefore, is interpreted to have been by concentration in residual magma and transfer of the magma either by convection or by upward displacement due to settling crystals. Iron-ore minerals tended to be lost from the lower parts of the lower zone (LZa and LZb), which accords with Wager and Brown's (1968, p. 71) observation that magnetite and ilmenite are not present there as cumulus minerals. Above these zones, however, the constituents of these minerals were concentrated in the residual magma or were added as cumulus minerals—perhaps both.

The concentration of iron-ore minerals in the residual magma in the upper parts of the magma chamber is significant with respect to the volume of the hidden part of the layered series. The requirement that the exposed part of the series formed from 1.454 parts parent magma implies that the hidden part forms at least 31 percent of the series as a whole. However if the model is correct with respect to iron-ore minerals, the part of the magma chamber now occupied by the exposed layered series received an addition of about 0.051 parts (table 43*C*), or 5 percent, iron-ore minerals, either as crystalline magnetite and ilmenite or as dissolved phases. Judging from the communalities of the total range of mixtures (fig. 15), the iron-ore minerals would contain 46 percent magnetite, indicating the

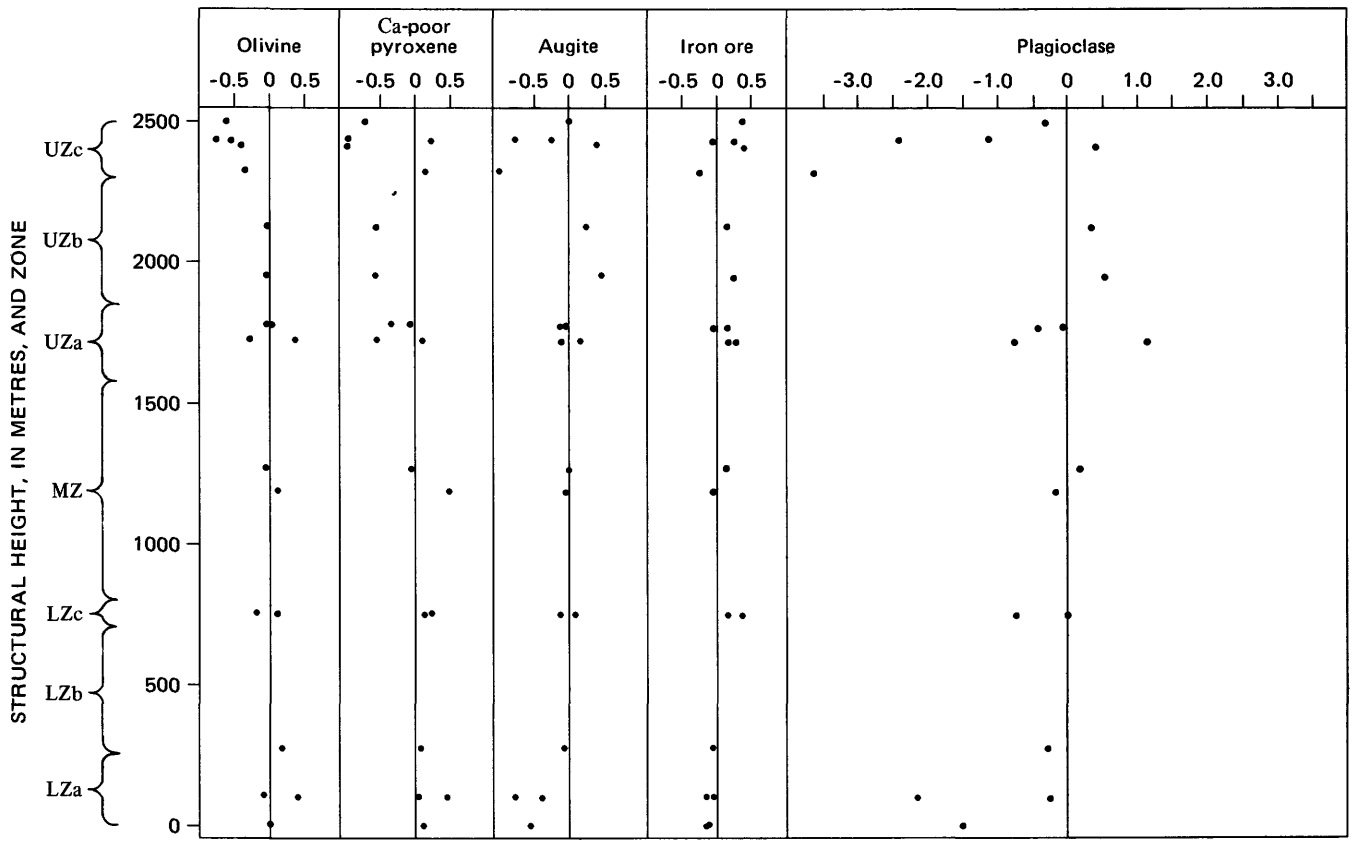

FIGURE 16.—Interpretation of the net gains and losses of minerals in the parent magma plotted against structural height of sample in the layered series of the Skaergaard intrusion. The horizontal scale, representing gains and losses, is in units of proportions. The amounts of parent magma that contributed to the formation of each sample are given in table 43.

addition of 2.3 percent magnetite to the exposed layered series, presumably from the part of the layered series that is hidden. If the original magma of the hidden layered series had the composition assumed for the parent magma (table 42), it contained a maximum of 2 percent magnetite and would have had to more than equal the exposed layered series in volume in order to have supplied the constituents of magnetite in the amounts required. If only one-half of the dissolved magnetite in the hidden layered series were transferred to the exposed part by residual concentration or by any other mechanism, the hidden layered series would necessarily have composed 70 percent of the total intrusion in order to supply the required amount. The estimate of 70 percent, however, is given only to illustrate how such estimates will vary with the assumed amounts of iron depletion in the hidden zone; the only absolute requirement of the model is that the hidden layered series forms at least one-half of the layered series as a whole. As pointed out before, Wager and Brown gave an estimate of 70 percent; Chayes (1970) suggested that 81 percent is more correct.

SUMMARY

The four series of igneous rock analyses to which the extended method of *Q*-mode factor analysis has been applied vary a great deal in complexity. The series of analyses from the layered series of the Skaergaard intrusion is complex, not because the analyses for each constituent exhibit large variation among the samples (table 18) but because the compositional variation within the intrusion was caused by a complex interaction of many processes—specifically, the precipitation of minerals of changing composition from a melt that also changed in composition as the processes continued. In contrast, although similar processes must have occurred to some extent in the formation of the rhyolite-basalt complex on the Gardiner River, the major part of the compositional variation in the complex was caused by only two processes—the eruption of a rhyolite lava and the incorporation of basalt or basaltic lava. The factor-variance diagrams given in figures 3 and 11 clearly and immediately reflect this difference in origin. Moreover, the diagrams for the granitoid intrusive in eastern Nevada (fig. 5) and for the lavas and pumices of the 1959 summit eruption at Kilauea (fig. 8) reflect origins of intermediate complexity that are in at least fair accord with those arrived at after exhaustive field and laboratory study by other workers. After the samples and the data have been selected, the factor-variance diagrams are derived by a straightforward procedure with complete objectivity, and if the samples are representative and the analyses reasonably correct, they can be regarded as fundamental descriptors of the rock bodies or series from which the samples came.

The *Q*-mode factor analysis procedures that follow derivation of the factor-variance diagram may or may not contain some element of subjectivity. If the purpose of the factor analysis is only to condense geochemical or petrologic data into fewer variables without significant loss of information, then the choice of end-members is unimportant, and the varimax axes will serve as well as any other possible reference vectors. All that is required is to present the composition loadings for each sample with respect to each varimax axis in tables or on maps—or to use these loadings to classify the samples into compositional groups. As a classification technique, the *Q*-mode method has an important advantage over common cluster methods and many petrographic diagrams that have been derived in that the classification scheme is not forced into a two-dimensional framework. Where the varimax axes are used as reference vectors for the compositional system, subjectivity enters the procedure only if and when an interpretation is made of what the varimax axes may represent. If no such interpretation is made, the varimax axes serve only as useful and conventional reference axes. Of course, the choice of the number of varimax reference axes could be partly subjective if, as can be expected to happen on some occasions, the factor-variance diagram does not clearly indicate how many there should be.

In the derivation of petrologic or geochemical models, it will usually be necessary to reject the principal components and varimax axes as representative of the compositions of the actual end-members in the compositional system. If the proper number of end-members in the system is clear from the factor-variance diagram, the search for more realistic end-members can be made either by the examination of the composition scores of selected vectors or by finding the vector representations of compositions of interest in the problem. If the proper number of end-members is not clear, the same procedures may have to be followed using factor spaces of several different dimensions. It will always be possible to derive a number of models that are mathematically satisfactory. The ultimate test of the model is its geologic plausibility, which must be inferred from the compositions of the selected end-members and the composition loadings of the samples with respect to them. The best reason for developing a factor model may be to subject a model developed by other means to rigorous quantitative examination.

REFERENCES CITED

Chayes, Felix, 1960, On correlation between variables of constant sum: Jour. Geophys. Research, v. 65, no. 12, p. 4185–4193

———1970, On estimating the magnitude of the hidden zone and the compositions of the residual liquids of the Skaergaard layered series: Jour. Petrology, v. 11, no. 1, p. 1–14.

Chayes, Felix, and Kruskal, William, 1966, An approximate statistical test for correlations between proportions: Jour. Geology, v. 74, no. 5, pt. 2, p. 692–702.

Christiansen, R. L., and Blank, H. R., Jr., 1972, Volcanic stratigraphy of the Quaternary rhyolite plateau in Yellowstone National Park: U.S. Geol. Survey Prof. Paper 729-B, 18 p.

Fenner, C. N., 1938, Contact relations between rhyolite and basalt on Gardiner River, Yellowstone Park: Geol. Soc. America Bull., V. 49, no. 9, p. 1441-1484.

_____1944, Rhyolite-basalt complex on Gardiner River, Yellowstone Park, Wyoming—A discussion: Geol. Soc. America Bull., v. 55, no. 9, p. 1081-1096.

Harman, H. H., 1967, Modern factor analysis [2d ed. rev.]: Chicago, University of Chicago Press, 474 p.

Imbrie, John, 1963, Factor and vector analysis programs for analyzing geologic data: Office Naval Research, Geography Branch, Tech. Rept. 6 [ONR Task No. 389-135], 83 p.

Imbrie, John, and Purdy, E. G., 1962, Classification of modern Bahamian carbonate sediments, *in* Classification of carbonate rocks—A symposium: Am Assoc. Petroleum Geologists Mem. 1, p. 253-272.

Imbrie, John, and VanAndel, T. H., 1964, Vector analysis of heavy-mineral data: Geol. Soc. America Bull., v. 75, no. 11, p. 1131-1155.

Klovan, J. E., and Imbrie, John, 1971, An algorithm and FORTRAN-IV program for large-scale Q-mode factor analysis and calculation of factor scores: Jour. Internat. Assoc. for Mathematical Geology, v. 3, no. 1, p. 61-77.

Klovan, J. E., and Miesch, A. T., 1975, Extended CABFAC and QMODEL computer programs for Q-mode factor analysis of compositional data: Computers & Geosciences, v. 1, no. 3 [1976].

Lee, D. E., and Van Loenen, R. E., 1971, Hybrid granitoid rocks of the southern Snake Range, Nevada: U.S. Geol. Survey Prof. Paper 668, 48 p.

Macdonald, G. A., 1968, Composition and origin of Hawaiian lavas, *in* Coats, R. R., Hay, R. L., and Anderson, C. A., eds., Studies in volcanology: Geol. Soc. America Mem. 116, p. 477-522.

Manson, Vincent, and Imbrie, John, 1964, FORTRAN program for factor and vector analysis of geologic data using an IBM 7090 or 7094/1401 computer system: Kansas Geol. Survey Spec. Distrib. Pub. 13, 46 p.

Miesch, A. T., 1969, The constant sum problem in geochemistry, *in* Merriam, D. F., ed., Computer applications in the earth sciences—An international symposium: New York, Plenum Press, p. 161-176.

Miesch, A. T., 1975, Q-mode factor analysis of compositional data: Computers & Geosciences, v. 1, no. 3 [1976].

Misch, Peter, and Hazzard, J. C., 1962, Stratigraphy and metamorphism of Late Precambrian rocks in central northeastern Nevada and adjacent Utah: Am. Assoc. Petroleum Geologists Bull., v. 46, no. 3, p. 289-343.

Murata, K. J., 1966, An acid fumarolic gas from Kilauea Iki, Hawaii: U.S. Geol. Survey Prof. Paper 537-C, 6 p.

Murata, K. J., and Richter, D. H., 1966, Chemistry of the lavas of the 1959-60 eruption of Kilauea Volcano, Hawaii: U.S. Geol. Survey Prof. Paper 537-A, 26 p.

Richter, D. H., Eaton, J. P., Murata, K. J., Ault, W. U., and Krivoy, H. L., 1970, Chronological narrative of the 1959-60 eruption of Kilauea Volcano, Hawaii: U.S. Geol. Survey Prof. Paper 537-E, 73 p.

Richter, D. H., and Moore, J. G., 1966, Petrology of the Kilavea Iki lava lake, Hawaii: U.S. Geol. Survey Prof. Paper 537-B, 26 p.

Richter, D. H., and Murata, K. J., 1966, Petrography of the lavas of the 1959-60 eruption of Kilauea Volcano, Hawaii: U.S. Geol. Survey Prof. Paper 537-D, 12 p.

Shapiro, Leonard, and Brannock, W. W., 1956, Rapid analysis of silicate rocks: U.S. Geol. Survey Bull. 1036-C, p. 19-56.

Wager, L. R., and Brown, G. M., 1968, Layered igneous rocks: Edinburgh and London, Oliver & Boyd, Ltd., 588 p.

Wager, L. R., and Mitchell, R. L., 1951, The distribution of trace elements during strong fractionation of basic magma—A further study of the Skaergaard intrusion, East Greenland; Geochim. et Cosmochim Acta, v. 1, p. 129-208.

White, R. W., 1966, Ultramafic inclusions in basaltic rocks from Hawaii: Contr. Mineralogy and Petrology, v. 12, no. 3, p. 245-314.

Wilcox, R. E., 1944, Rhyolite-basalt complex on Gardiner River, Yellowstone Park, Wyoming: Geol. Soc. America Bull., v. 55, no. 9, p. 1047-1080.

Wright, T. L., 1973, Magma mixing as illustrated by the 1959 eruption, Kilauea Volcano, Hawaii: Geol. Soc. America Bull., v. 84, no. 3, p. 849-858.

☆ U.S. GOVERNMENT PRINTING OFFICE 1976—777-034/17