A generalized principal components model in petrology

MICHAEL ED. HOHN & LA VERNE M. FRIBERG

LITHOS

Hohn, M. E. & Friberg, L. M. 1979: A generalized principal components model in petrology. *Lithos 12*, 317–324. Oslo. ISSN 0024-4937.

Principal components analysis of elemental distributions among minerals elucidate chemical, thermal, and equilibrium effects. Previous approaches in applying this multivariate statistical method to petrologic data are united into a single model, so-called three-mode principal components analysis. The major advantage of the model lies in the separate analysis of each of the three modes: element, mineral, and sample. The three sets of principal components are related through a 'core matrix'. The model accommodates the ordination of pair-wise combinations of the modes, such as samples and minerals in visualization of multidimensional tie-lines. Because of its generality, the model permits great flexibility in the study of petrologic data.

Michael Ed. Hohn, Organic Geochemistry Unit, School of Chemistry, University of Bristol, Bristol, England. Present address: West Virginia Geological and Economic Survey, P.O. Box 879, Morgantown, W. Va. 26505, U.S.A.
La Verne M. Friberg, Department of Geology, University of Akron, Akron, Ohio 44325, U.S.A.

Principal components analysis is useful in the study of the distribution of chemical species among mineral phases. Saxena (1969) begins with a matrix of oxide weight percents of coexisting minerals, derives a correlation matrix and calculates the eigenvectors in one approach to the *n*-dimensional tie-line problem. In such an application, the main interest lies in the examination of a plot of mineral-sample points on the principal axes and comparison of the tie-lines for position and orientation. The eigenvectors showing relationships among the chemical species possess secondary importance.

On the other hand, Lindh (1975) uses similar input data, but plots samples alone on principal axes, and examines relationships among mineral-chemical species combinations. His arrangement of the data differs from that of Saxena because of his purpose in evaluating element distributions among mineral phases. One can envisage a more complex study comprising more phases than the two coexisting pyroxenes that Lindh examines.

A formal relationship exists between these approaches in the form of three-mode principal components analysis (Tucker 1966). In the present context, the three modes are: chemical species, mineral phase, and sample. Exploitation of the three-mode character of petrologic data through this analytic model provides a number of vantage points for viewing patterns in a data set because the eigenvectors of each mode can be plotted separately, and studied with respective scores of combination modes on principal components. The approaches of Saxena (1969) and Lindh (1975) are two special cases of the three-mode formulation. Such a synthesis of principal components methods may be appropriate to an exhaustive study of a data set, in particular one in which the nature and significance of internal relations are poorly understood *a priori*.

An advantage of the model as adopted from Tucker (1966) is that a three-mode solution can be reduced to a common two-mode format for reasons of simplification or purposes of the analysis. This reduction can proceed to any of three formats, including those used by other workers.

Procedure

Because Saxena (1969) describes the rationale and calculations behind principal components analysis, only a few relevant details shall be repeated here, mainly for purposes of notation and analogy. Representing two-mode principal components analysis, i.e. the conventional model, as:

$${}_i\tilde{X}_k = {}_iD_mE_k \qquad (1)$$

the matrix ${}_iD_m$ contains the 'loadings' of i variables on m principal components; the matrix ${}_mE_k$ contains the 'scores' of k samples on principal components; and ${}_i\tilde{X}_k$ contains the original data entries as estimated from the m components. The original data matrix, ${}_iX_k$ might comprise a number of cations in a given mineral determined for i cations and k samples. From the product of this matrix by its transpose, ${}_iX_kX'_i$, eigenvectors normalized to a sum of squares equal to unity are computed, yielding ${}_iD_m$. The scores are calculated from:

$$ {}_mE_k = {}_mD'_iX_k \tag{2} $$

If the data include determinations of numbers of cations from each of several minerals from each sample, then there exists an ${}_iX_k$ for each of the j minerals. Concatenating these matrices yields ${}_iX_{(jk)}$. One can picture this matrix as a table having the cations arranged as rows, and the first k columns representing the measurements on all of the samples for one mineral, the next k columns representing those measurements for a second mineral, and so forth. The matrix multiplication of this matrix by its transpose gives a sum of squares and cross products amongst the cations: ${}_iX_{(jk)}X'_i$, from which a set of eigenvectors, ${}_iA_i$, can be calculated. The data matrix can be rearranged such that minerals correspond to rows: ${}_jX_{(ik)}$, or such that samples become rows: ${}_kX_{(ij)}$; again, a set of eigenvectors can be calculated from the respective sums of squares and cross products matrices, yielding ${}_jB_j$ and ${}_kC_k$. We shall follow the example of Tucker (1966) in calling (jk) or (ik) or (ij) a 'combination mode'.

A fundamental purpose of principal components analysis lies in the reduction of dimensionality, that is, the reduction of a table of many columns to one of a few, or the reduction of a multidimensional graph impossible to draw, to one or two axes – the principal components. The model described here allows the independent reduction of each mode according to some criterion such as the cumulative percentage of the total variance explained by the axes.

We can now state the three-mode principal components model in matrix form:

$$ {}_i\tilde{X}_{(jk)} = {}_iA_mG_{(pq)}\,({}_pB_j \otimes {}_qC_k) \tag{3} $$

where ${}_i\tilde{X}_{(jk)}$ is a principal components approximation of the input data. The eigenvector matrices, ${}_iA_m$, ${}_pB_j$ and ${}_qC_k$ have been reduced to m, p and q eigenvectors, respectively. Note that use of the combination mode (jk) permits us to represent the model in conventional, two-way matrices, but also necessitates us to use the Kronecker product $\otimes$, described by Pease (1965) and others. The 'core matrix', ${}_mG_{(pq)}$ relates the principal components of the three modes; as with eigenvector matrices, inspection of the salient values in ${}_mG_{(pq)}$ provides insight into the major patterns in the data. The model of (3) states that the original data is approximated by the sum of a number of rougher approximations. There are $m \cdot p \cdot q$ of these approximations, each calculated by considering the matrix multiplication of three eigenvectors, one from each mode. The $m \cdot p \cdot q$ entries of the core matrix weight each possible combination of eigenvectors such that those entries with large magnitudes indicate the combinations contributing most to the reconstructed data. The core matrix is calculated from the three eigenvector matrices and the input data:

$$ {}_mG_{(pq)} = {}_mA'_iX_{(jk)}\,({}_jB'_p \otimes {}_kC'_q) \tag{4} $$

Scores

One can rewrite (3) in a conventional two-mode form:

$$ {}_i\tilde{X}_{(jk)} = {}_iA_mL_{(jk)} \tag{5} $$

where ${}_iA_m$ represents loadings of the i variates on principal components, and ${}_mL_{(jk)}$ represents the scores of units in the combination mode (jk). Therefore:

$$ {}_mL_{(jk)} = {}_mG_{(pq)}\,({}_pB_j \otimes {}_qC_k) \tag{6} $$

The fundamental equation (3) can be rearranged three ways, conforming with the three possible arrangements of the data as described above. The matrices of loadings are the eigenvector matrices of (3), and each as a corresponding matrix of scores, calculated from, for instance:

$$ {}_mL_{(jk)} = {}_mA'_iX_{(jk)} \tag{7} $$

In the two examples below, the data comprising each mineral-cation combination were standardized to zero mean and unit variance, thus weighting the cations equally and offsetting the tendency for minerals of greatest contrasting composition to dominate the analysis. In computing the scores, the unstandardized data were used in order to restore the inter-mineral distances and therefore the tie-lines.

Example: Madras charnockites

The data of Howie (1955) were chosen because the internal relationships are reasonably simple,

Table 1. Eigenvalues and percent of total variance along each principal component calculated from Madras Charnockite data (Howie 1955).

	Principal component		
	1	2	3
Cation eigenvalues	15.53	5.36	3.29
Percent	51.80	17.90	11.00
Mineral eigenvalues	19.83	6.69	3.47
Percent	66.10	22.30	11.60
Sample eigenvalues	15.24	4.05	1.25
Percent	74.20	19.70	6.10

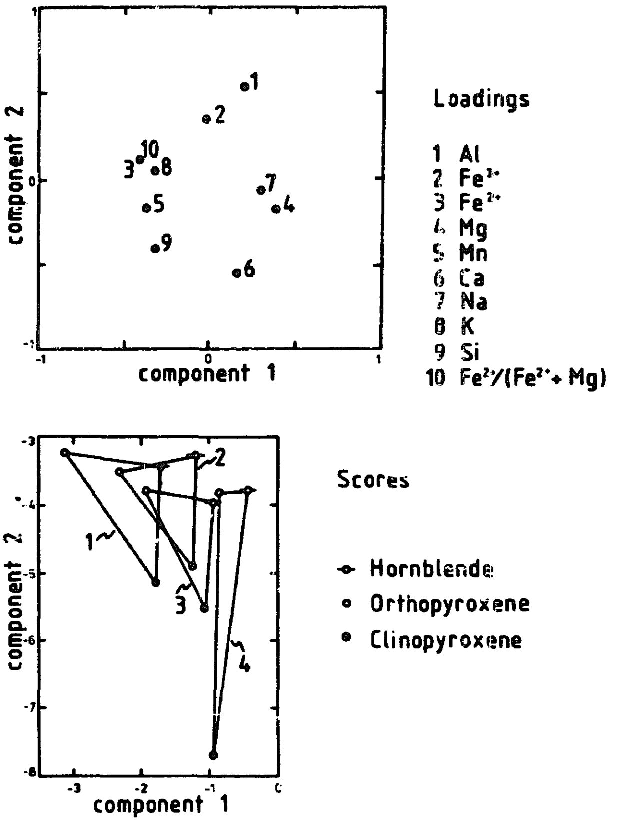

Fig. 1. Loadings and scores for principal components analysis of cation mode of Madras Charnockites (Howie 1955). Sample numbering corresponds to Howie's numbers in the order: 4642A, 2270, 2941, 3709.

and because they have been used previously in studying mineral compositions and equilibria (Kretz 1963) and in exemplifying multivariate graphical techniques (Saxena 1969). We used the number of cations and the ratio, X_{Fe} determined from the three minerals hornblende, orthopyroxene, and clinopyroxene, in each of four samples. As stated in the previous section, the four values in each of the thirty combinations of three minerals and ten cations were standardized by subtracting the mean of these four values, and dividing by the variance. Then, the three sums-of-squares-and-cross-products matrices were calculated, three respective sets of eigenvectors extracted, and the core matrix computed from (4). The scores were calculated from (7) using the unstandardized data. The computer program employed in the analysis limited the number of cations, oxides, or ratios to ten. Analysis of the cations mode shows that the first two eigenvectors account for 70 percent of the total variation in the data (Table 1). The ordination of the loadings (Fig. 1) provides information about the patterns among cations and the ratio, X_{Fe} across all minerals and samples. For instance, Fe^{2+} and X_{Fe} provide redundant information, attributable to the fact that Fe^{2+} and Mg^{2+} bear an almost perfectly inverse partitioning relationship with each other.

A plot of mineral-sample scores on principal axes is found to be useful in the viewing of n-dimensional tie-lines (Saxena 1969). As in the conventional graphing of tie-lines on ternary diagrams, features noted include the relative positions and angles among the lines. It should be borne in mind that a given projection might correctly represent some relationships and misrepresent others. In the present example, sample 4 has a different composition from the others, and, as an outlier, tends to dominate the positioning of the principal axes. We could have repeated the mathematical analysis with this sample excluded, allowing more subtle patterns among the apparently homogeneous subpopulation of samples 1, 2 and 3 to emerge.

Accounting for 88 percent of the total variance (Table 1), the first two eigenvectors of the minerals mode show that clinopyroxene and orthopyroxene are end members to the intermediate composition of hornblende (Fig. 2). This observation reflects a well-known property of these minerals, *viz.* a linear equation describes the composition of any one phase in terms of the other two.

Corresponding scores of cation-sample combinations are analogous to plots of numbers of cations and calculated cation ratios for pairs of minerals such as those found in Kretz (1963). In the case of principal components scores, each axis represents a linear combination of minerals, so that graphing these scores precludes the need for constructing all pair-wise combinations of the minerals yet permits one to detect straight-line

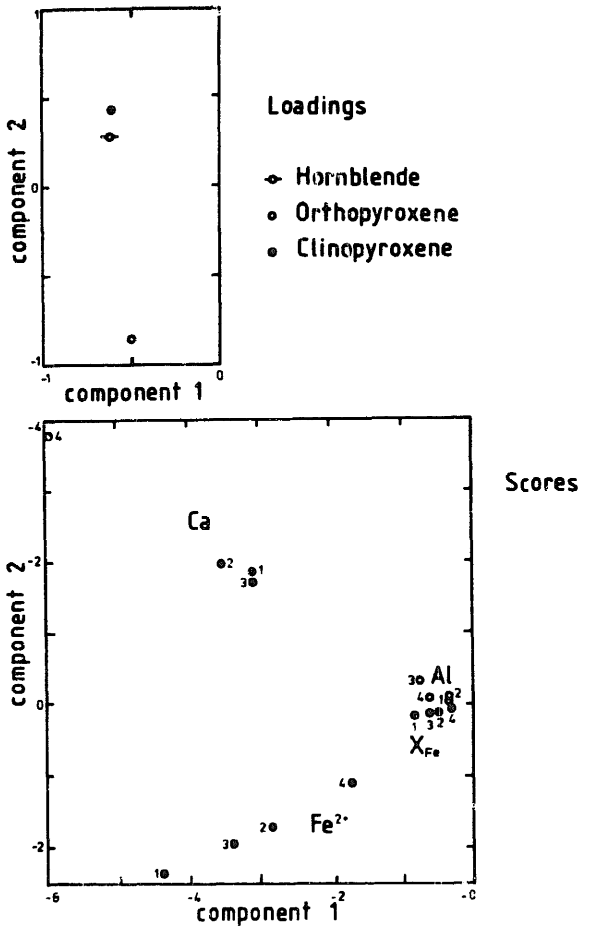

Fig. 2. Loadings and scores for principal components analysis of minerals in Madras Charnockites (Howie 1955).

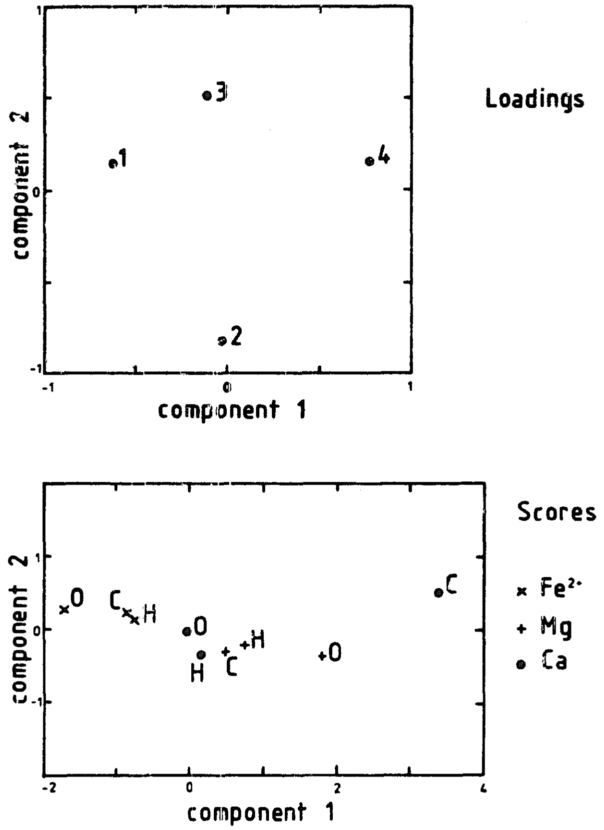

Fig. 3. Loadings and scores for principal components analysis of samples mode. Data are from Howie's study of Madras Charnockites (1955).

relationships. In the present example, only a small number of scores are shown in order to reduce the complexity of the presentation. Nevertheless, the near-linearity of the points for each cation is apparent. The orthogonality of Ca^{2+} with respect to Fe^{2+} is also obvious from the scores of Fig. 2, explaining the similar orthogonality of these cation loadings in Fig. 1.

The sample loadings and scores (Fig. 3) contrast samples 1 and 4 along an axis that accounts for a large proportion of the variance (Table 1). The calculations represented by Fig. 3 are analogous to Lindh's (1975) analysis of coexisting pyroxene pairs. A minor difference lies in the fact that Lindh scaled the sample eigenvectors to the respective eigenvalues, and has scaled the eigenvectors for the cation-mineral combination mode to unity. The relationships that Lindh found for approximately 100 pyroxene pairs compare well with those in the present study.

The core matrix (Table 2) relates the three principal components analyses, and can be used to inspect the interactions between the eigenvectors of different modes. The very large or small values indicate those combinations of eigenvectors that contribute a lot to the variation among the data; near-zero values indicate unimportant combinations. For instance, one value

Table 2. Core matrix in three-mode principal components model of Madras Charnockites. Components in each mode are labelled as in Figs. 1 to 3.

	Sample components	
	1	2
Cation component 1		
Mineral component 1	-3.846	-0.050
Mineral component 2	0.230	-0.467
Cation component 2		
Mineral component 1	0.097	-1.949
Mineral component 2	-0.620	-0.172

Table 3. Summary of the mineral assemblages and degree of retrograde metamorphism caused by the Tobacco Root Batholith in the Spuhler Peak samples.

Mineral	Sample number 1	2	3	4	5	6	7	8	9
Plagioclase	x	x	x	x	x	x	x	x	x
Orthoclase	–	–	–	x	–	–	x	–	–
Garnet	x	x	x	x	x	x	x	x	x
Biotite	x	x	x	x	x	x	x	x	x
Hornblende	x	x	x	x	x	x	x	x	x
Cummingtonite	–	x	–	–	x	–	x	–	x
Quartz	x	x	x	x	–	–	x	–	x
Calcite	–	–	–	–	–	x	–	–	–
Salite	–	–	–	–	–	x	–	–	–
Degree of retrogradation	S	S	H	S	M	M	M	M	M

S = slight; M = moderate; H = high; x = mineral present; – = mineral absent.

(–3.846) relates the first cation component, the first mineral component and the first sample component. Because all of the minerals weight equally and negatively on the first axis (Fig. 2), this core entry refers to a trend present throughout the minerals mode. In fact, it contrasts the high Mg^{2+} and Na^+ content of sample 4 against the high Fe^{2+} content of sample 1. A second important entry (–1.949) suggests a negative interaction between the second cation component, the first mineral component, and the second sample component. Again, this interaction cuts across all minerals, but perhaps not as consistently as the one above. The trend is one of Al-rich and Ca-poor minerals in sample 3 relative to 2.

Spuhler Peak Formation

A second example comprises a set of data from the Spuhler Peak Formation of the Central Tobacco Root Mountains in southwestern Montana. The samples were collected from a small area in order to keep T and P_t constant. The rocks were regionally metamorphosed to orthoclase-sillimanite grade (650 ± 50°C and 5–7 Kb) during the Precambrian, and later altered along fractures by the emplacement of the Tobacco Root Batholith (Laramide age; 70 m.y.) (Friberg 1976). All samples appear to have a basaltic or graywacke parentage and all contain the minerals: plagioclase, garnet, biotite, and hornblende (Table 3). The patterns within the Spuhler Peak data are probably more complex than the Madras data because of the more heterogeneous petrochemistry and a more complex petrologic history. The data comprise numbers of cations (see Fig. 4) in four minerals from each of nine samples.

For brevity, only the first two eigenvectors of the cation mode are presented (Fig. 4), although the eigenvalues in Table 4 suggest that more components give significant information. A third axis has salient loadings for Na^+ and Ti^{4+}, and a fourth axis has salient loadings for Mg^{2+} and Mn^{2+}.

In plotting the mineral-sample scores, the ordination of sample loadings on principal components is particularly useful in this example for locating outliers and defining homogeneous subpopulations. Indeed, samples 3, 6, and 7 constitute obvious outliers around a number of samples in the center of the plot (Fig. 5). Sample 3 has been highly altered by the metasomatic fluids along fractures and grain boundaries, resulting in a change in the chemistry of the minerals. Differences in bulk chemistry probably account for the anomalous positions of samples 6 and 7. Sample 6 shows the greatest chemical variance, with high Ca^{2+} and low Si^{4+} content reflected in the presence of calcite and salite, and the absence of quartz (Table 3). The presence of orthoclase in sample 7 suggests a high K^+ content. Note that the samples falling near the center of Fig. 5 include three samples of slight retrogradation, whereas the outliers have medium to high retrogradation. The plot sug-

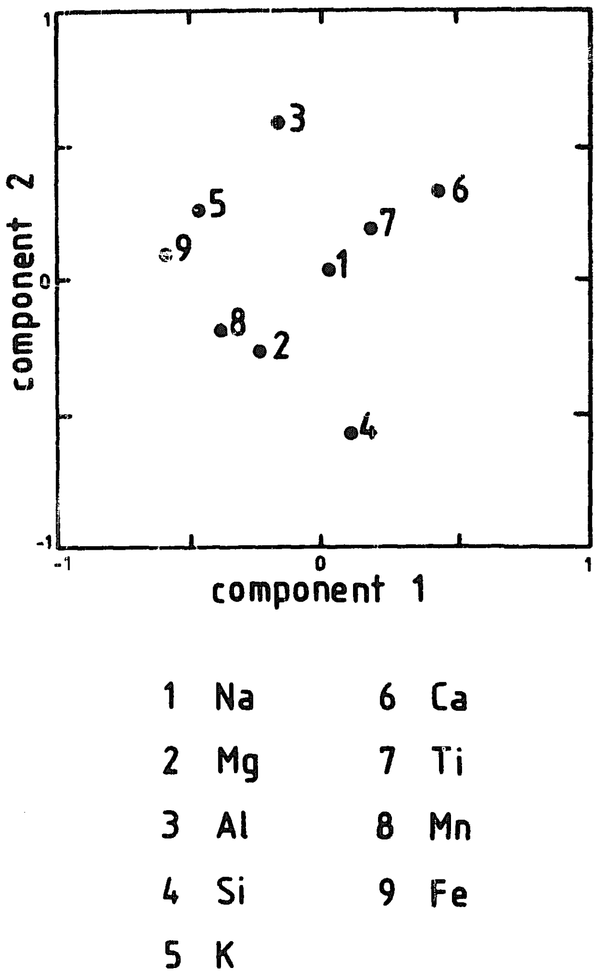

Fig. 4. Loadings of cations on principal components calculated from Spuhler Peak Formation data (Friberg 1976).

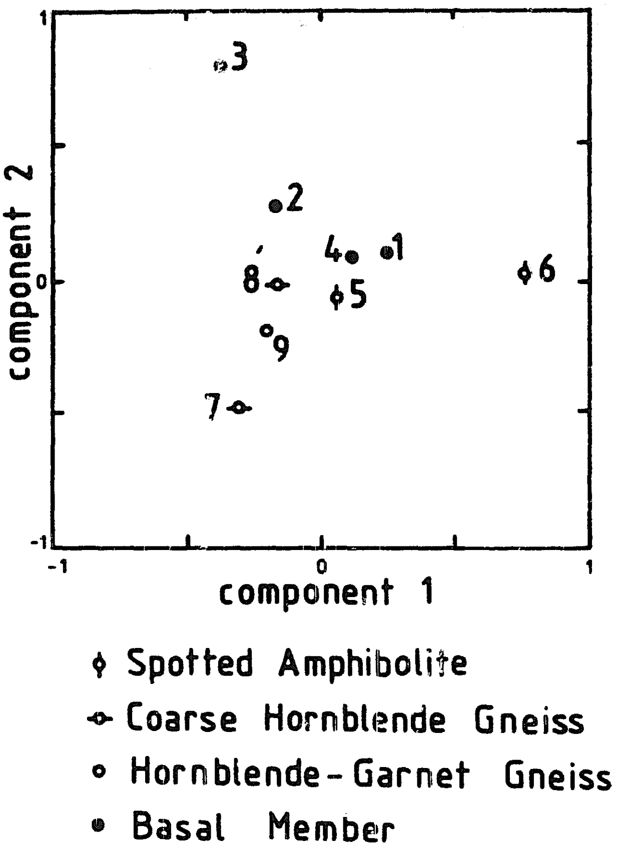

Fig. 5. Loadings of samples from Spuhler Peak Formation on principal components. The four rock types are described in Friberg (1976).

gests that samples 6 and 7 may reflect a higher level of retrogradation than thought when Table 3 was originally compiled.

The n-dimensional tie-lines show samples 4 and 5 to be very similar with some crossing of tie-lines (Fig. 6), and show a large similarity between these and sample 1 except for differences in plagioclase composition. These samples show a very different pattern from the outliers (Fig. 7). In the interest of a complete array for analysis, the data include a retrograde biotite in sample 6. Although this datum deviates from the other biotite points, the most distinctive feature of this sample lies in the unusual plagioclase composition.

Table 4. Eigenvalues and percent variance along principal components calculated from Spuhler Peak data.

	Principal components			
	1	2	3	4
Cation eigenvalues	8.33	8.11	6.55	4.81
Percent	23.10	22.50	18.20	13.40
Mineral eigenvalues	11.24	8.86	8.43	7.46
Percent	31.20	24.60	23.40	20.70
Sample eigenvalues	5.58	5.20	1.28	0.43
Percent	43.00	40.00	9.90	3.30

Eigenvalues from analysis of the minerals mode (Table 4) indicate that each of the four phases contribute to the overall pattern in some unique way. A two-component ordination of the eigenvector loadings of Table 5 does not adequately represent the observed data because no one phase can be described as a linear combination of the other phases. In addition, the scores for the cations do not show the degree of linearity evident in the previous example. When plotted, these scores detail the varying elemental compositions of each sample, *viz.* the low

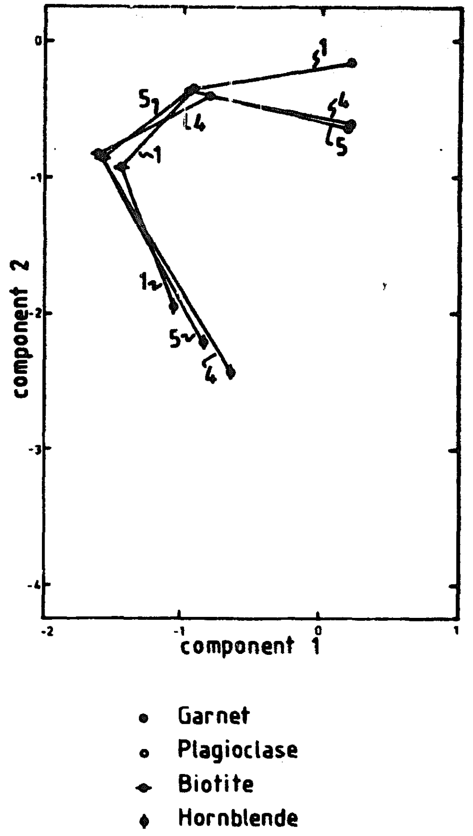

Fig. 6. Scores of mineral-sample combinations on principal components with a few of the tie-lines drawn in. Data are from Spuhler Peak Formation (Friberg 1976).

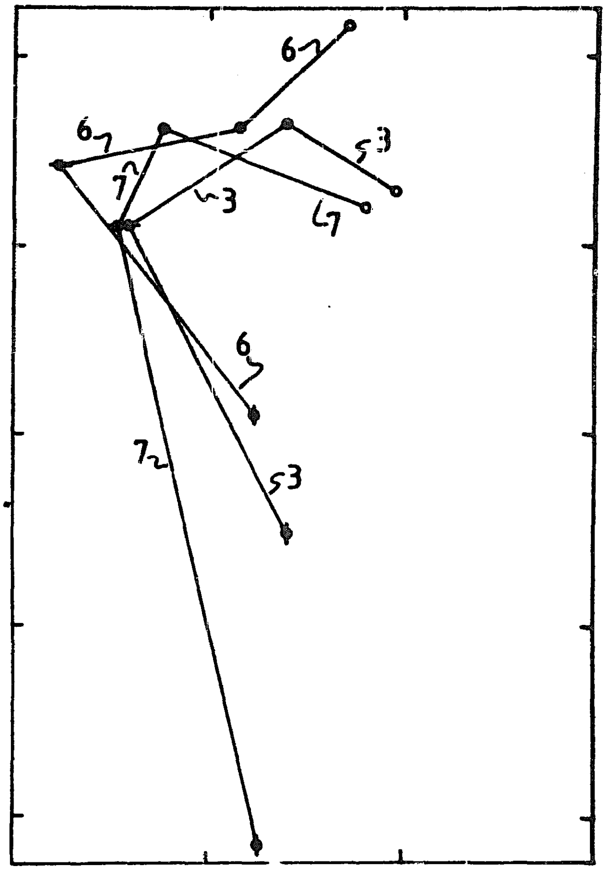

Fig. 7. Same as Fig. 6 but with scores of the atypical samples graphed for the four minerals.

sodium content of sample 7, the high silicon content of sample 6, and the nearly homogeneous compositions of the majority of samples.

The core matrix contains a number of values very different from zero, suggesting many interactions cutting across modes (Table 6). The negative interaction (–1.469) between the second cation, second mineral, first sample components expresses the Al-rich plagioclase and biotite in sample 6, consistent with observations made above on the mineral assemblage of this sample. Similarly, the entry of –1.860 concerns the Ca-poor and Fe-rich plagioclase relative to the other minerals in sample 3 which resulted from the retrograde metamorphic event.

Summary of conclusions

A problem with missing values will arise whenever samples do not contain the same minerals, or when a cation is absent throughout the sampling of a mineral. One can simply set the missing entries to zero, being careful not to attempt standardization of the relevant blocks in the array. If a cation is absent from a given mineral, then the sums of squares and cross products matrix among the cations will have unequal diagonal values; the result can be suppression of that cation on the principal components. The same holds true for the mineral concerned. Nevertheless, we have followed this procedure with data sets after we removed blocks of entries with only minor changes in the projection. Substituting theoretical or 'typical' values for missing entries (Saxena 1969) is probably not good practice. Procedures for estimating missing values from those present in the data array do not apply to systematic omissions.

In summary, the three-mode principal components model extends the multivariate approach

Table 5. Loadings of four minerals on principal components calculated from Spuhler Peak data.

	Principal component		
	1	2	3
Garnet	−0.621	0.125	−0.242
Plagioclase	0.395	−0.678	−0.561
Biotite	−0.354	−0.706	0.613
Hornblende	−0.577	−0.165	−0.500

Table 6. Core matrix in three-mode principal components model of Spuhler Peak data.

	Sample components	
	1	2
Cation component 1		
Mineral component 1	0.487	−1.860
Mineral component 2	0.412	1.068
Mineral component 3	−1.254	0.366
Cation component 2		
Mineral component 1	−0.286	−0.463
Mineral component 2	−1.469	−0.169
Mineral component 3	−1.166	−0.468

to analysis of petrologic data by allowing identical treatment of each mode. Previous application of principal components analysis has meant that one could obtain loadings of the units of one mode, but the other two modes were represented in combination as the scores. The present model comprises analysis of each mode in isolation, but a core matrix combines the separate sets of eigenvectors into a single, descriptive equation. Because of the difficulty in interpretation of the core matrix, calculation of the scores matrices expedites analysis of some data sets, and in fact yields displays equivalent to those used in previous studies. The model therefore incorporates the features of these specific applications, yet maintains a greater generality.

Acknowledgements. – The Computer Centre and the School of Chemistry of the University of Bristol provided logistical support. Prof. G. Eglinton also provided computing facilities. During this work, M.E.H. received support from the National Aeronautics and Space Administration (subcontracted under NGL 15-003-118), for which grateful thanks are due Prof. J. M. Hayes. Electron probe microanalyses were conducted at Indiana University by L.M.F. on a grant from the National Science Foundation to Cornelis Klein, Jr. (GA-37109).

References

Friberg, L. M. 1976: Petrology of a metamorphic sequence of upper-amphibolite facies in the Central Tobacco Root Mountains, southwestern Montana. Indiana University, U.S.A., unpublished Ph.D. thesis.

Hohn, M. E. 1979: Three-mode principal components analysis of geological data. (Submitted to *Mathematical Geology.*)

Howie, R. A. 1955: The geochemistry of the Charnockite Series of Madras, India. *Trans. Roy. Soc. Edinb. 62,* 725–768.

Kretz, R. 1963: Distribution of magnesium and iron between orthopyroxene and clinopyroxene in natural mineral assemblages. *J. Geol. 71,* 773–785.

Lindh, A. 1975: Coexisting pyroxenes – a multivariate statistical approach. *Lithos 8,* 151–161.

Pease, M. C. III 1965: *Methods of Matrix Algebra.* Academic Press, New York.

Saxena, S. K. 1969: A statistical approach to the study of phase equilibria in multicomponent systems. *Lithos 3,* 25–36.

Tucker, L. R. 1966: Some mathematical notes on three-mode factor analysis. *Psychometrika 31,* 279–311.

Accepted for publication April 1979
Printed October 1979