Archaeometry 36, 1 (1994), 115–130. Printed in Great Britain

RQ-MODE PRINCIPAL COMPONENTS ANALYSIS OF CERAMIC COMPOSITIONAL DATA*

H. NEFF

Research Reactor Center, University of Missouri, Columbia, MO 65211, U.S.A.

RQ-mode principal components analysis (PCA) is a means for calculating variable and object loadings on the same axes, so that elements can be displayed along with data points on a single diagram. The biplots resulting from RQ-mode PCA preserve both Euclidean relations among the objects and variance-covariance structure. When used with data on the chemical composition of archaeological pottery, such biplots facilitate recognizing compositional subgroups and determining the chemical basis of group separation. RQ-mode PCA is illustrated in this paper with neutron activation data on Mesoamerican Plumbate pottery.

KEYWORDS: PRINCIPAL COMPONENTS, BIPLOTS, R-MODE, Q-MODE, RQ-MODE, COMPOSITIONAL ANALYSIS, CERAMICS, PROVENANCE, SOURCE IDENTIFICATION

INTRODUCTION

Principal components analysis (PCA) is often used in the search for source-related subgroups in ceramic compositional data (e.g., Baxter 1992a and 1993b; Bishop and Neff 1989; Glascock 1992; Leese *et al.* 1989; Pollard 1986). While not precisely a group-formation technique, PCA facilitates recognition of compositional groups by identifying the orientation of axes along which the data set as a whole is most elongated. Dimensions of maximum elongation often coincide with axes along which separation between subgroups is most easily visible. Information about the analyses, such as provenance, type, or initial group membership derived from cluster analysis, may be attached to symbols on the principal component plot to assist the search for structure.

Beyond recognizing subgroups, another goal of PCA is to understand the chemical basis of group separation. In R-mode analysis, the loadings of the original variables (elemental concentrations) on each principal component may be inspected to obtain the desired information, whereas in Q-mode analysis the factor scores for the original variables may be inspected. A more direct approach is to perform a simultaneous RQ-mode analysis, which permits display of the data points and the variables (elements) on the same set of diagrams. An RQ-mode technique known as correspondence analysis (Teil 1975) has been employed previously in archaeological compositional investigations (e.g., Peisach *et al.* 1982; Underhill and Peisach 1985). Correspondence analysis was originally developed for nominal or ordinal data, and its application to continuous data, whether as an R-mode or RQ-mode technique, requires scaling the rows of the data matrix to induce closure (Davis 1986). This is an unappealing procedure for the mixture of trace and minor element log-concentrations that usually make up data matrices in ceramic compositional studies. Biplotting techniques

* Received 26 February 1993, accepted 23 July 1993.

of PCA pose no such conceptual difficulties for applications to ceramic compositional data. However, as Baxter (1992a) has pointed out, biplotting has seen little use either in compositional investigations or in archaeology generally.

This paper provides some additional discussion of the use of RQ-mode PCA in compositional investigations. In particular, use of a variant of Baxter's (1992a) type 3 biplot (see below) is advocated as a useful means of simultaneously assessing inter-object and inter-variable relations. The technique of simultaneous RQ-mode PCA discussed here was introduced about a decade ago in geology by Zhou *et al.* (1983; also see Davis 1986; Walden *et al.* 1992).

SIMULTANEOUS RQ-MODE PRINCIPAL COMPONENTS ANALYSIS

PCA involves extraction of eigenvalues and eigenvectors (also called characteristic roots and characteristic vectors) from a minor product matrix, $\boldsymbol{X'X}$, where $\boldsymbol{X}$ is the matrix of data consisting of n observations (rows) and m variables (columns) (Davis 1986). The eigenvectors, which are orthogonal, become a new set of reference axes on which the data can be displayed. That is, if $\boldsymbol{U}$ is a square matrix with the eigenvectors of $\boldsymbol{X'X}$ in successive columns, then $\boldsymbol{S}$, the scores of the observations on successive principal components, are found by $\boldsymbol{S}=\boldsymbol{XU}$. In other words, the elements of each eigenvector are coefficients of a linear equation that define a transformation of measured elemental concentrations into a score on that eigenvector. In principal components factor analysis, the eigenvectors are scaled by the singular value matrix, $\boldsymbol{E}$, to yield $\boldsymbol{A}^{\boldsymbol{R}}=\boldsymbol{UE}$, the factor loading matrix, before calculating $\boldsymbol{S}^*=\boldsymbol{XA}^{\boldsymbol{R}}$ ($\boldsymbol{E}$ has square roots of the eigenvalues of $\boldsymbol{X'X}$ on the diagonal and zeros elsewhere). PCA based on the minor product matrix is a R-mode analysis.

The Q-mode counterpart of PCA begins with calculation of a major product matrix, $\boldsymbol{XX'}$. Whereas the R-mode analyses just discussed are thought of as focusing on interrelationships among variables, Q-mode analyses are thought of as focusing on interrelationships among objects. Thus, the 'loadings' in Q-mode analysis represent the proportion of variance of a variable that is accounted for by a particular data point. The Eckart-Young Theorem establishes a close relation between R-mode and Q-mode analysis (Davis 1986, 519–24): Q-mode loadings are proportional to R-mode scores and vice versa, and the non-zero eigenvalues of the major product and minor product matrices are identical. Davis (1986) provides a clear and detailed discussion of R-mode and Q-mode techniques.

The connections between R-mode and Q-mode techniques that are spelled out by the Eckart-Young Theorem are embodied in the singular value decomposition of a data matrix, $\boldsymbol{X}$ (Davis 1986). The singular value decomposition (see Gower 1984; Jolliffe 1986) can be written as $\boldsymbol{X}=\boldsymbol{VEU'}$, where $\boldsymbol{E}$ and $\boldsymbol{U}$ are the singular value and eigenvector matrices, respectively, as described above; and $\boldsymbol{V}$ is a $n \times p$ matrix (in which p is the rank of $\boldsymbol{X}$) that must, given $\boldsymbol{E}$ and $\boldsymbol{U}$, contain the standardized principal component scores (Jolliffe 1986, 38). The singular values (on the diagonal of $\boldsymbol{E}$) are the standard deviations of scores on each principal component, so it follows that $\boldsymbol{VE}$ contains R-mode principal component scores from which standardization has been removed, that is, $\boldsymbol{VE}=\boldsymbol{S}=\boldsymbol{XU}$. As Baxter (1992a) points out following Gabriel (1971), Gower (1984), and Jolliffe (1986), the singular value decomposition gives rise to three alternative types of biplots, that is, simultaneous (RQ-mode) representations of the variables and observations:

(1) object coordinates from $\boldsymbol{VE}$, variable coordinates from $\boldsymbol{U}$;
(2) object coordinates from $\boldsymbol{V}$, variable coordinates from $\boldsymbol{UE}$;
(3) object coordinates from $\boldsymbol{VE}$, variable coordinates from $\boldsymbol{UE}$.

Type 1 and type 2 plots both specify combinations of variable and object coordinates whose inner product reproduces the original data matrix exactly. Geometrically, a particular object's score on one of the original variables is represented by the projection of its biplot coordinates on to the vector connecting the origin with the coordinates for the variable of interest (Jolliffe 1986, 79–80). However, both type 1 and 2 biplots have disadvantages: the principal component coefficients (i.e., the eigenvectors of $\boldsymbol{X'X}$) used to represent variables on a type 1 plot do not represent the variance-covariance structure of the data as faithfully as does $\boldsymbol{UE}$, the scaled loading matrix used to represent variables in type 2 plots (Baxter 1992a; Baxter and Heyworth 1991; Corsten and Gabriel 1976; Gabriel 1971; Gower 1984; Jolliffe 1986); and $\boldsymbol{VV'}$, the distances between objects on a type 2 biplot, do not preserve Euclidean relationships (Baxter 1992a; Gower 1984; Jolliffe 1986; Zhou *et al.* 1984). These observations lead Gower (1984) to prefer type 3 plots, which preserve both Euclidean relationships and variance-covariance structure. Furthermore, despite loss of the exact relationship between the biplot and the original data that is preserved by types 1 and 2 (Jolliffe 1986; also see Baxter 1992a), this relationship remains straightforward, since weighting the inner product of the object and variable coordinate matrices by the singular value matrix reproduces the original data, that is, $\boldsymbol{X}=[\boldsymbol{VE}]\boldsymbol{E}^{-1}[\boldsymbol{UE}]'$.

The simultaneous RQ-mode PCA proposed by Zhou *et al.* (1983) for geological applications is an example of Baxter's (1992) type 3 biplot. Without any initial scaling of the data, non-zero eigenvalues and eigenvectors of $\boldsymbol{X'X}$ and $\boldsymbol{XX'}$ are the same, and R-mode and Q-mode factor loadings are defined for the same factor loading space (Davis 1986, 595); the R-mode loadings are the coordinates of the variables, $\boldsymbol{UE}=\boldsymbol{A}^R$, and the Q-mode loadings are the coordinates of the data points, $\boldsymbol{VE}=\boldsymbol{A}^Q$ (these are also the non-standardized R-mode principal component scores, $\boldsymbol{S}$, as pointed out above). An important drawback of using raw data in a simultaneous RQ-mode analysis is the extreme sensitivity to differences to magnitudes of the variables (Davis 1986, 595).

In practice, raw data are almost always scaled prior to analysis so that eigenvalues and eigenvectors are extracted from a matrix of meaningful measures of similarities among variables or objects. In R-mode analysis, centring the data by columns produces a variance-covariance matrix as the minor product matrix, $\boldsymbol{X'X}$. If the data matrix is first not only centred but standardized, the minor product matrix is a correlation matrix; this is the most common starting point for a PCA. In Q-mode analysis the data matrix is scaled so that the major product matrix, $\boldsymbol{XX'}$, contains meaningful measures of similarity between observations; two widely used measures are Euclidean distance in principal coordinates analysis (Gower 1966; Joreskog *et al.* 1976) and the cosine-theta measure of proportional similarity in Q-mode factor analysis (Joreskog *et al.* 1976). Simultaneous RQ-mode PCA could be based on any of the foregoing scaling procedures, and in fact biplots based on standardized compositional data (so $\boldsymbol{X'X}$ is a correlation matrix) have been presented by Baxter (1992a and 1993b). However, for RQ-mode analysis to be congruent with R-mode and Q-mode analyses and in order to ensure a meaningful dual solution, one seeks a scaling procedure that yields meaningful measures of *both* inter-variable similarity in $\boldsymbol{X'X}$ and inter-object similarity in $\boldsymbol{XX'}$ (Zhou *et al.* 1983, 586).

The approach to simultaneous RQ-mode PCA introduced by Zhou *et al.* (1983; Davis 1986) involves a scaling of X to produce W, as follows,

$$w_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{n}} \qquad (1)$$

where w_{ij} are the transformed data and x_{ij} are the original data, $\bar{x}_j$ being the means of each variable over all objects. Using this transformation, WW', the major product matrix, is a principal coordinates matrix, which incorporates Euclidean distances, and $W'W$, the minor product matrix, is a variance-covariance matrix (Zhou *et al.* 1983, 588). Including the standard deviation of the jth variable in the denominator yields a correlation matrix as the minor product matrix (as would a simple standardization of the original data) and a matrix incorporating Euclidean distances between standardized data for the major product matrix. R-mode factor loadings are computed from the minor product matrix, and Q-mode loadings are given by the product of the scaled data matrix and R-mode eigenvectors, that is, $A^Q = WU$ (Zhou *et al.* 1983). Note that the Q-mode loadings are simply non-standardized R-mode scores for the scaled data. From an R-mode perspective, this is a PCA (see Davis 1986), while from a Q-mode perspective, it is a principal coordinates analysis (see Gower 1966; Joreskog *et al.* 1976).

Note that nothing about RQ-mode PCA precludes scaling the rows of the data matrix prior to scaling the columns by equation (1). One possible row transformation is the centred log-ratio suggested by Aitchison (1986) for 'fully-compositional' data (summing to 100%), which has been used in PCA of archaeometric compositional data by Baxter (1989, 1991, 1992b, and 1993a; but see Tangri and Wright 1993). Aitchison's transformation is

$$y_{ij} = ln\left(\frac{x_{ij}}{g\ (x_i)}\right) \qquad (2)$$

where $g(x_i)$ is the geometric mean of the constituents in the ith analysis. Even in the absence of 'full' compositional characterization, Aitchison's transformation may be useful because it cancels out proportional dilution that may be introduced by tempering with inert material (Leese *et al.* 1989; see also Baxter 1993b). Aitchison (1986) suggested using the variance-covariance matrix of the transformed data, a practice followed by Leese *et al.* (1989); in RQ-mode PCA, this would entail using the unstandardized version of equation (1) to effect a column scaling after transforming the rows using Aitchison's centred log-ratio. On the other hand, Baxter (1992b) has shown that PCA of the correlation matrix of centred log-ratio data may lead to more satisfactory results with some compositional data sets; in RQ-mode PCA, this would entail using the standardized version of equation (1) to scale the columns of the centred log-ratio data.

RQ-mode PCA by the above means, like other type 3 biplots, retains the best features of type 1 and type 2 biplots (Gower 1984). The R-mode and Q-mode loadings can be plotted on the same factor axes, giving a simultaneous representation of variables and data points in the same space (Zhou *et al.* 1983, 589). Within this space, clusters of similar observations and contributions of the variables to the object scores on the depicted components may be identified by inspection. As in standard PCA, Euclidean relations among objects are preserved, and clustering of objects in the component space indicates possible compositional group affiliation. At the same time, information on variance is preserved in the length of the vector from the origin to a variable point, and information on inter-variable correlations is preserved in the cosine of the angle formed between two such vectors (Baxter and Heyworth 1991, 1992a, and 1993b; Corsten and Gabriel 1976; Gower 1984; Jolliffe 1986). In ceramic compositional investigations, the simultaneous plotting of objects and variables makes it possible to determine rapidly the contribution of each element or set of intercorrelated elements to group separation (Baxter 1992a). Variables contributing toward elongation of groups in the compositional space may also be identified. Variables that plot close to the origin are relatively unimportant in determining object locations on the components depicted (Baxter 1990; Tangri and Wright 1993). These features of RQ-mode PCA are illustrated in the next section.

An important practical consideration in the use of any biplot is that whether or not variables and objects can be represented in the same space depends upon the extent to which their coordinates take on a similar range of values. For example, Baxter (1992a) utilizes separate plots to depict variables and observations in his illustrations of type 2 biplots, apparently because the variables (for which [scaled] principal component loadings are plotted) fall into a small cluster lying close to the origin, while the objects (for which standardized PCA scores are plotted) are more dispersed. As the number of objects in the analysis increases, the opposite problem becomes an increasing concern because the total object loading on each component is divided among a larger number of relatively smaller pieces. A RQ-mode plot of Walden *et al.* (1992, fig. 4) illustrates this problem: variables are distant from the origin compared to the objects, and separate variable and object plots are employed for full illustration of the results. Fortunately, the sizes of most compositional data sets seldom make this a serious problem in sourcing studies; for example, the 241-specimen data set used for illustration in the following section is comparatively large but does not require separate variable and object plots. A general means to circumvent this problem is to explore any principal components space through plots of different subsets of the variables and objects (or even just ellipses representing probability cut-off levels for the object subgroups) and to zoom in on portions of the plots; an interactive computing environment that permits such operations is clearly an advantage (see Appendix).

RQ-MODE PCA IN CERAMIC COMPOSITIONAL ANALYSIS

PCA techniques are best regarded as exploratory devices whose usefulness for ceramic compositional analysis is measured by the extent to which they promote recognition and interpretation of subgrouping tendencies in the compositional data (e.g., Baxter 1993a and 1993b). Do RQ-mode PCA or other kinds of biplots meet this standard of utility?

In a recent review, Baxter (1992a) found only two examples of the 'neglected technique' of biplotting in chemical characterization studies of archaeological artefacts (Berthoud *et al.* 1979, cited in Baxter 1992; Poirier and Barrandon 1983). Although neither of the two pre-1992 examples involved ceramics, Baxter (1992a and 1993b) has now added several examples of the use of type 2 biplots with ceramic compositional data. Casting the net somewhat more widely so as to include correspondence analysis as a biplotting technique adds a few more publications (e.g., Peisach *et al.* 1982; Underhill and Peisach 1985). Clearly, however, biplotting remains underused in ceramic compositional analysis. This is unfortunate because ceramic compositional investigations typically address both subgrouping tendencies of the objects and variance-covariance structure (Bishop and Neff 1989; Glascock 1992; Pollard 1986), and biplots, particularly of type 3 (Gower 1984), would seem ideally suited for such an application. Some of the potential utility of biplots in ceramic compositional investigations is illustrated in the following example, in which the RQ-mode PCA (type 3 biplot) of Zhou *et al.* (1983) is applied to ceramic compositional data from southern Mesoamerica. A data set with known, well-defined subgroups is chosen deliberately so that the ability of RQ-mode PCA to reveal group affiliations and group shape (rather than just individual sample-to-sample affinities) can be examined.

A study of Mesoamerican Plumbate pottery by neutron activation analysis (Neff 1984; Neff and Bishop 1988) revealed two distinct compositional groups that correspond to petrographic groups recognized previously (Shepard 1948). To explore the potential of the RQ-mode technique discussed above, the two compositional reference groups, San Juan Plumbate and Tohil Plumbate, were amalgamated into a single data set (n = 241), on which RQ-mode PCA was carried out.[1] A total of four RQ-mode analyses were undertaken: (1) base-10 log concentrations using the unstandardized version of equation (1); (2) base-10 log concentrations using the standardized version of equation (1); (3) Aitchison's log-ratio transformation on the rows together with the unstandardized version of equation (1); and (4) Aitchison's log-ratio transformation on the rows together with the standardized version of equation (1).

Baxter (1993b) has presented a series of examples based on type 2 biplots similar to the type 3 biplots presented here. As in Baxter's examples, the first two components extracted

Table 1 *Eigenvalues and percentage variance explained, first ten components*

	Unstandardized (variance-covariance matrix)			*Standardized (correlation matrix)*		
	Eigenvalue	*% variance*	*Cumulative % variance*	*Eigenvalue*	*% variance*	*Cumulative % variance*
Raw data (log base-10 ppm)						
1	0.0901	43.99	43.99	6.0190	31.68	31.68
2	0.0274	13.38	57.37	3.2500	17.10	48.78
3	0.0231	11.27	68.64	2.0400	10.74	59.52
4	0.0133	6.51	75.15	1.4400	7.58	67.10
5	0.0112	5.49	80.64	1.0540	5.55	72.65
6	0.0094	4.60	85.24	0.9945	5.23	77.88
7	0.0066	3.22	88.46	0.7251	3.82	81.70
8	0.0055	2.70	91.16	0.6609	3.48	85.18
9	0.0041	2.00	93.16	0.5242	2.76	87.94
10	0.0032	1.56	94.72	0.4344	2.29	90.22
Centred log-ratio data (Aitchison's row transformation)						
1	0.0884	48.59	48.59	6.7171	35.35	35.35
2	0.0231	12.68	61.27	2.4445	12.87	48.22
3	0.0155	8.54	69.81	2.1000	11.05	59.27
4	0.0116	6.36	76.17	1.3707	7.21	66.49
5	0.0094	5.18	81.35	1.2225	6.43	72.92
6	0.0074	4.05	85.40	0.9904	5.21	78.13
7	0.0066	3.61	89.01	0.7848	4.13	82.26
8	0.0055	3.03	92.04	0.6332	3.33	85.60
9	0.0032	1.76	93.80	0.5191	2.73	88.33
10	0.0025	1.37	95.17	0.4199	2.21	90.54

using these four approaches provide slightly different pictures of structure in the data. Because the membership of the two groups is known beforehand in this case, a rough evaluation of the performance of the four approaches is indicated by the number of misassignments implied by plots of the first two components.

Table 1 shows the eigenvalues obtained from the four RQ-mode PCAs, and Figures 1 to 4 show variables and objects plotted on the first two components derived from each analysis. In all four analyses, the two known groups appear as distinct centres of mass in the principal component space, with the axis of group separation being parallel to component 1 in all but the second analysis (standardized log concentration data). Although the groups are distinct on all four plots, the second analysis also appears to yield the most clear-cut separation. Reinforcing the latter observation, posterior classification using Mahalanobis distances from the two centroids on components 1 and 2 (Table 2) indicate no misclassifications for the scores derived from the second PCA (standardized log concentrations). Thus, the best two-dimensional picture of these data is achieved with standardized log concentration data; the lack of improvement with Aitchison's row transformation suggests that dilution is not a major contributor to variability in the compositional profiles of the Plumbate groups.

The four biplots provide differing but complementary perspectives on the chemical basis of group separation in the Plumbate data. The analyses based on unstandardized data (Figs 1 and 3) suggest that antimony contributes most to the divergence of San Juan from Tohil on component 1, while analyses based on standardized data (Figs 2 and 4) indicate more nearly equivalent contributions from antimony, chromium, thorium, and cesium. The PCAs based on standardized data also suggest a more important role for higher hafnium and iron concentrations in Tohil. These observations suggest that variation in antimony may be greater than the other elements but that it may not contribute any more significantly to group separation than other elements with which it is correlated. This suspicion is confirmed by the bivariate antimony-chromium plot (Fig. 5), which shows that, although both elements efficiently separate the two groups, the mean antimony concentration in San Juan is about 200% higher than Tohil, while there is only about 50% difference in the chromium means. Similarly, hafnium and iron discriminate the groups reasonably well, although differences in the means of the two groups are only around 25% or less.

Table 2 *Classification success based on Mahalanobis distances calculated on principal components 1 and 2*

		Unstandardized (variance-covariance matrix)				*Standardized (correlation matrix)*	
		Raw data (log base-10 ppm)					
		Into: San Juan	Tohil			Into: San Juan	Tohil
From	San Juan	153	2	From	San Juan	155	0
	Tohil	2	84		Tohil	0	86
		Centred log-ratio data (Aitchison's row transformation)					
		Into: San Juan	Tohil			Into: San Juan	Tohil
From	San Juan	152	3	From	San Juan	154	1
	Tohil	2	84		Tohil	1	85

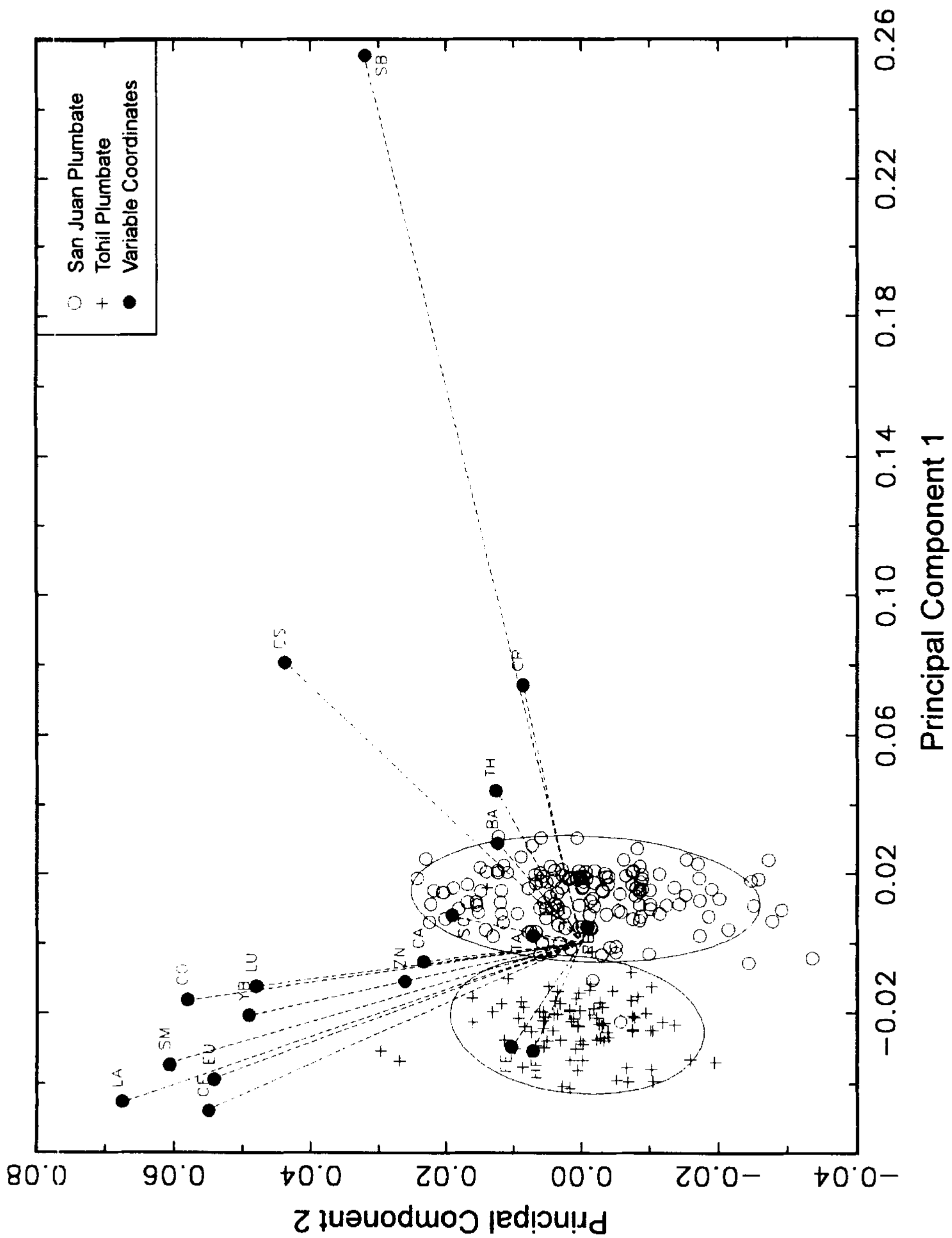

Figure 1 *Plot of variable and object loadings on the first two axes obtained from RQ-mode PCA of the Plumbate data using the unstandardized version of equation (1) without row scaling. Ellipses indicate 90% confidence level for membership in the San Juan and Tohil Plumbate subgroups.*

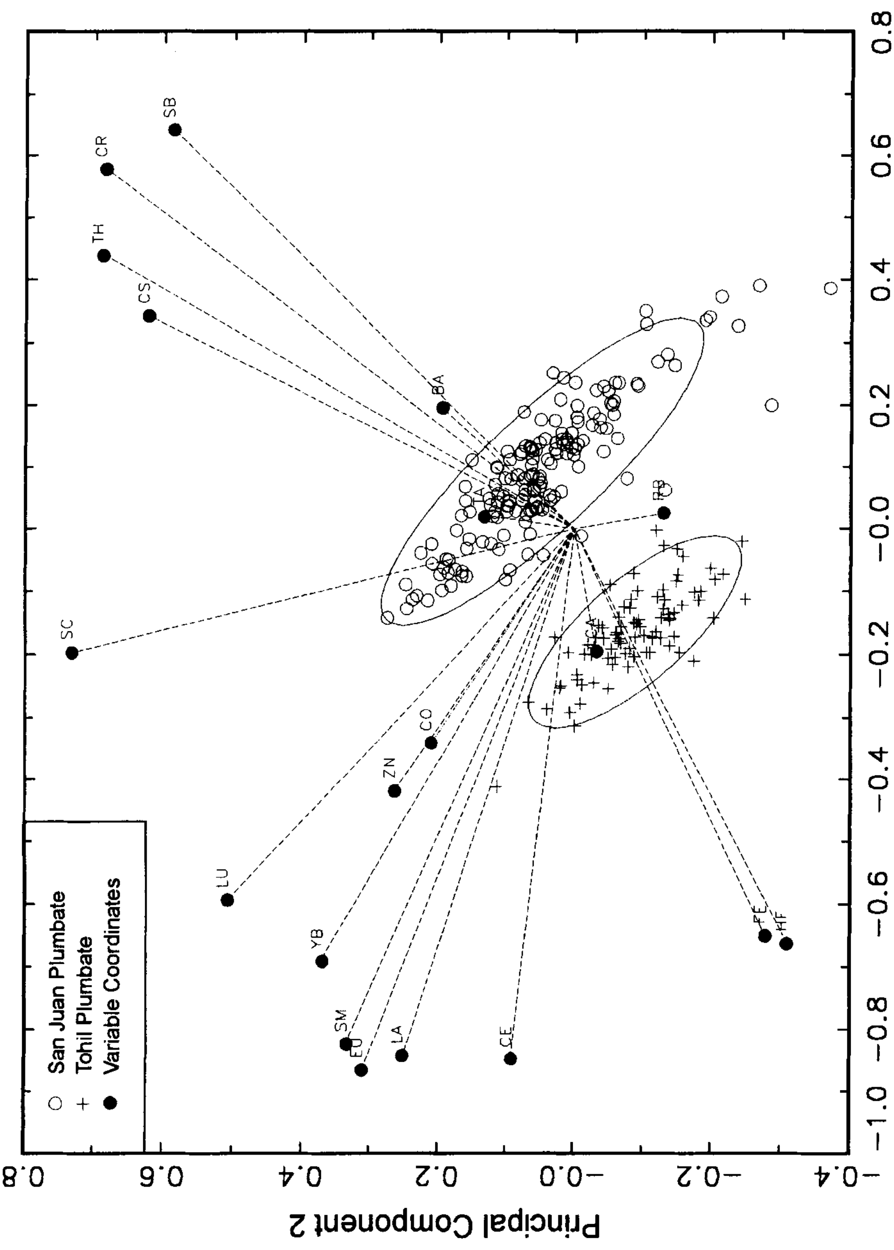

Figure 2 *Plot of variable and object loadings on the first two axes obtained from RQ-mode PCA of the Plumbate data using the standardized version of equation (1) without row scaling. Ellipses indicate 90% confidence level for membership in the San Juan and Tohil Plumbate subgroups.*

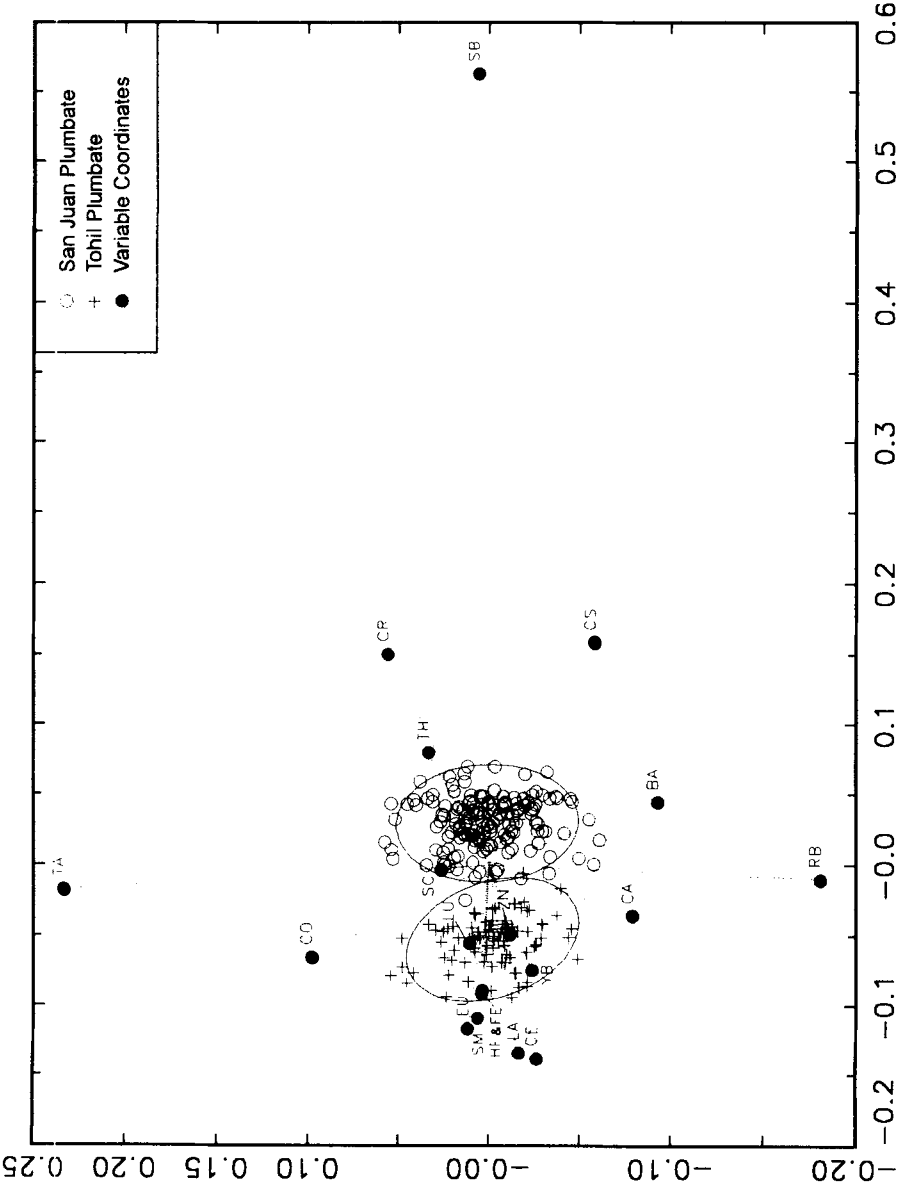

Figure 3 *Plot of variable and object loadings on the first two axes obtained from RQ-mode PCA of the Plumbate data using Aitchison's (1986) centred log-ratio transformation of the rows and the unstandardized version of equation (1). Ellipses indicate 90% confidence level for membership in the San Juan and Tohil Plumbate subgroups.*

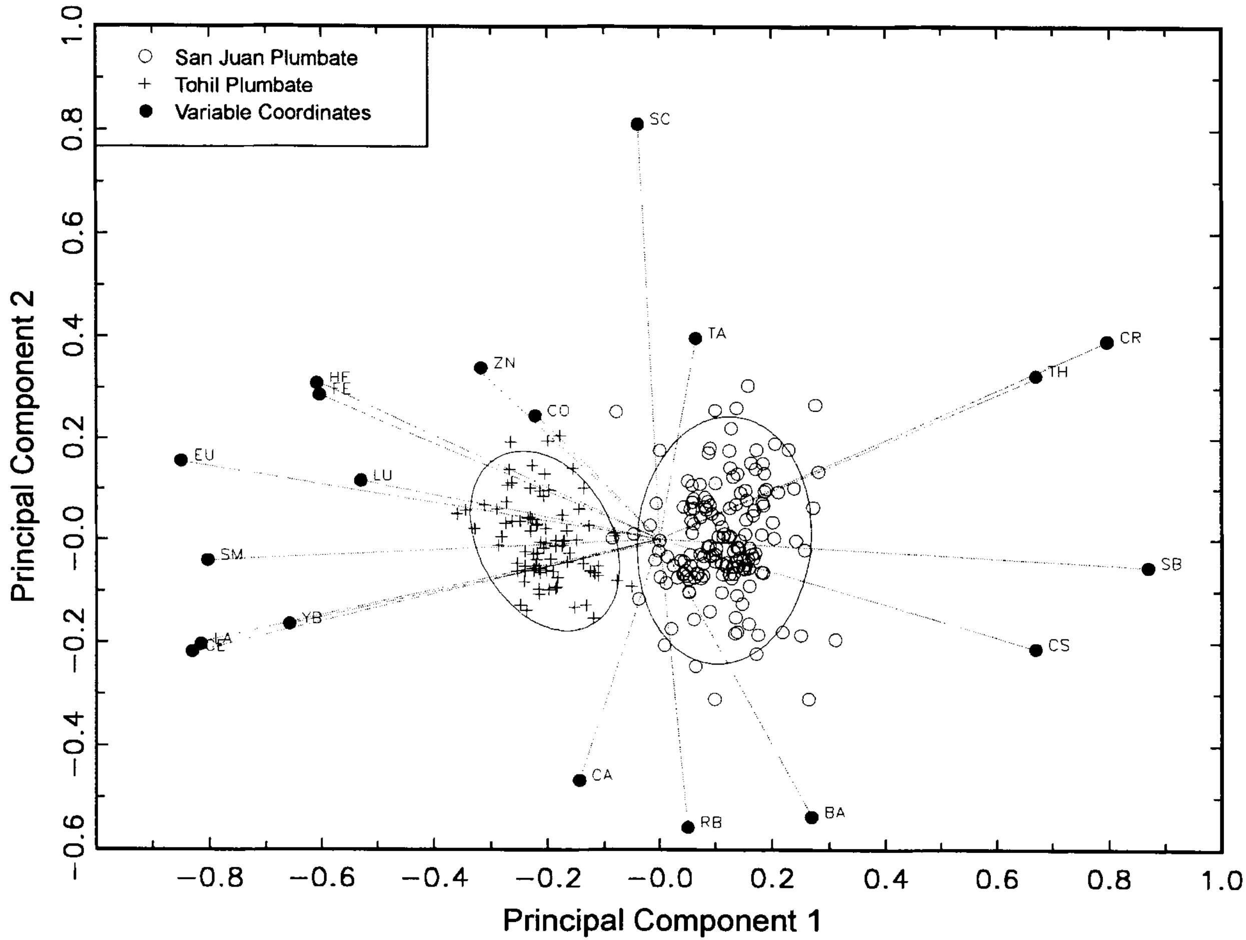

Figure 4 *Plot of variable and object loadings on the first two axes obtained from RQ-mode PCA of the Plumbate data using Aitchison's (1986) centred log-ratio transformation of the rows and the standardized version of equation (1). Ellipses indicate 90% confidence level for membership in the San Juan and Tohil Plumbate subgroups.*

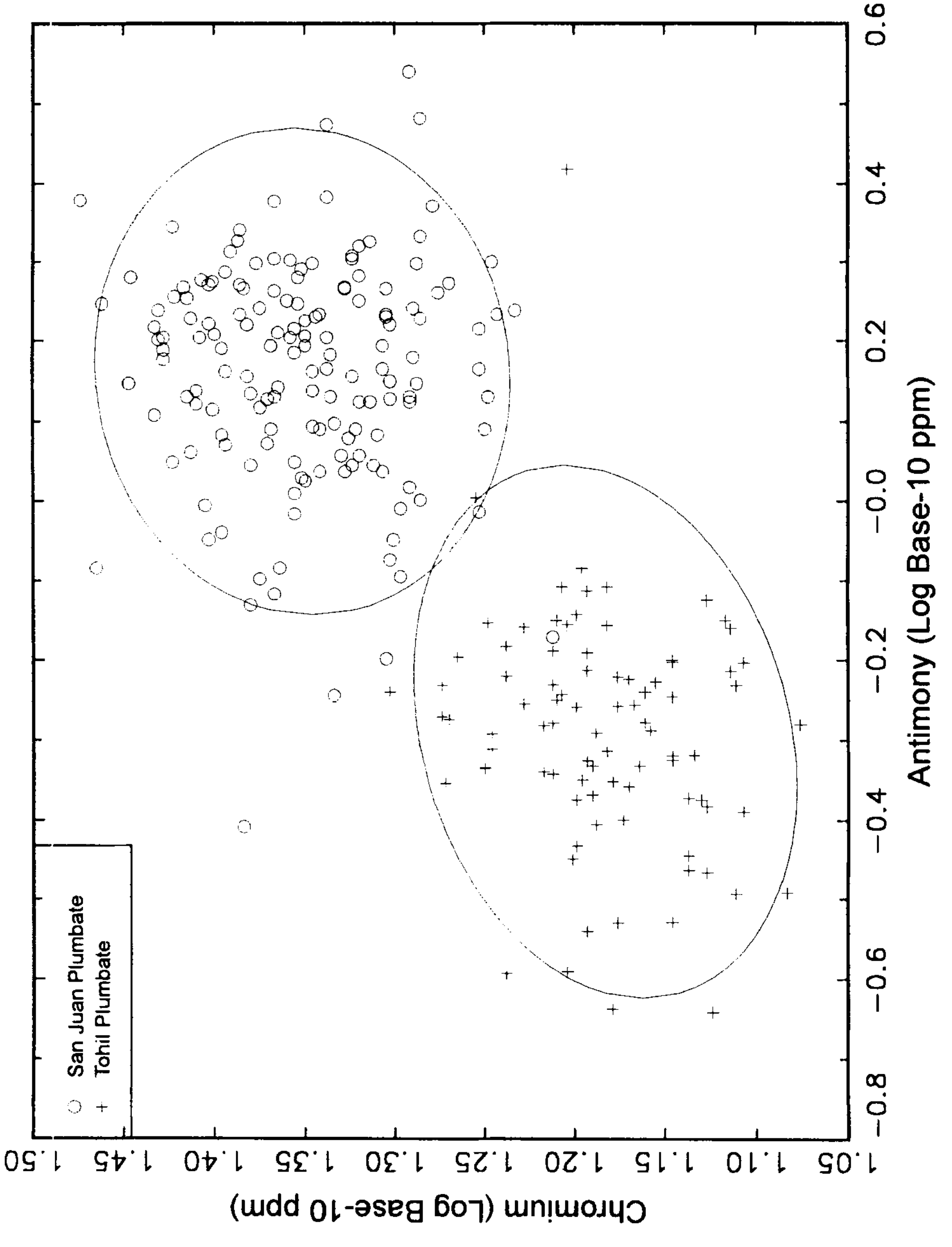

Figure 5 *Bivariate scatterplot of antimony and chromium log-concentrations in Plumbate data. Ellipses indicate 90% confidence level for membership in the two Plumbate subgroups.*

As this example illustrates, inspection of biplots is an efficient means to identify which of the numerous possible bivariate scatterplots of raw data will best represent the subgroup structure in a given data set. Thus, it is immediately obvious from inspection of any one of the biplots (Figs 1–4) that a bivariate plot of antimony and chromium will clearly show the separation between the Plumbate groups (Fig. 5).

Other sorts of contributions to overall chemical variation can also be identified by inspecting biplots. For the analyses based on log concentrations (Figs 1 and 2), the rare earths appear to contribute to elongation of the groups along parallel axes and not to be very important in creating separation between the groups. The Tohil group shows less elongation, suggesting a tighter range of values on these elements. The analyses based on centred log-ratio data (Figs 3 and 4) tell a slightly different story, the rare earths appearing to make a more important contribution to group separation in a direction similar to hafnium and iron. Bivariate lanthanum-samarium plots confirm these inferences, showing a tighter range of concentrations for the Tohil group and a tendency for the groups to differentiate into high (Tohil) and low (San Juan) rare earth groups following transformation. In other words, once proportional effects (i.e., dilution) are removed by expressing all concentrations as ratios to the geometric mean across all elements within each sample, Tohil appears to be enriched in rare earths relative to San Juan, and the rare earths acquire a very strong negative correlation with the suite of elements that tend to be enriched in San Juan (antimony, chromium, thorium, and cesium).

A number of elements can be identified as contributing very little to intergroup separation. For instance, rubidium and tantalum show negligible correlation with the main axis of group separation on all four biplots. In the case of centred log-ratio data (Figures 3 and 4), these elements make a strong contribution to variation in a direction perpendicular to the axis of group separation. In accord with these observations, a bivariate plot shows that the two groups overlap almost completely on rubidium and tantalum.

The foregoing observations and others that could be advanced on the basis of the biplots shown in Figures 1–4 provide a means for addressing the natural and cultural bases of separation between the Plumbate groups. For example, elongation of the two groups along the same rare earth element correlation axis in the PCAs based on raw log concentrations (Figs 1 and 2) is consistent with the inference that a common ceramic resource base was used in production of the two technologically-related wares. This is consistent with the inference that both varieties originated on the Pacific coast of southern Mesoamerica, near the current Mexico-Guatemalan border (Neff 1984; Neff and Bishop 1988). At the same time, the fact that the Tohil group is generally less spread out along this axis suggests greater specificity in exploitation of this resource base (Figs 1 and 2). Further, the suite of elements that are enriched in San Juan (cesium, chromium, thorium, and antimony) tend to be enriched in primary clays developing on rhyolitic volcanic tuffs elsewhere in coastal Guatemala (Neff *et al.* 1992); minute volcanic glass sherds incorporated into San Juan Plumbate paste (Shepard 1948) are consistent with such a derivation. Concentrations of these elements are depleted in more highly weathered clays of the region, while some other elements, including hafnium and iron, are sometimes enriched (as in Tohil Plumbate). Enhancement of rare earth differences following transformation to centred log-ratios may also be explicable in terms of greater weathering of clays originally derived from volcanic tuffs.

The foregoing example illustrates how the joint display of variables and objects made possible by simultaneous RQ-mode PCA facilitates recognition of the chemical basis of subgroup differentiation in ceramic compositional data. That information, in turn, bears on the nature of natural and cultural processes that produced the observed subgroup structure. Although the specifics of the Plumbate case are largely beyond the scope of this paper, it is worth noting that the RQ-mode PCA has contributed additional detail not indicated by more conventional analyses (cf. Neff 1984; Neff and Bishop 1988).

CONCLUSION

This paper illustrates the use of one of the many tools available for identifying and interpreting source-related subgroups in ceramic compositional data. The column scaling procedure recommended by Zhou *et al.* (1983) simultaneously yields major and minor product matrices containing conceptually reasonable measures of inter-variable and inter-observation similarity. The R-mode and Q-mode component loadings occur in the same factor space, so variables and observations can be represented on the same plots. The resulting type 3 biplots simultaneously provide the best two-dimensional representation of both Euclidean relations among objects and variance-covariance structure among variables. Biplots of types 1 and 2 provide alternative representations of variables and objects, but type 1 biplots are weak with respect to variables, while type 2 biplots are weak with respect to objects (Gower 1984, 737).

The Plumbate example illustrates how different versions of RQ-mode PCA (standardized versus unstandardized, raw data versus row transformed data) provide different perspectives on the inherent structure in the data. This example reinforces Baxter's (1991; 1993b; Baxter and Heyworth 1989) advice that any application of PCA to compositional data benefits from the use of multiple approaches. Although not illustrated here, RQ-mode PCAs based on different subsets of variables and different subsets of observations are other means for generating mutually reinforcing or complementary perspectives on data structure. Inspection of components beyond the first two should not be neglected either. Finally, as mentioned previously, display of different subsets of object and variable coordinates obtained from a single analysis may be required for effective visualization of data structure on a given set of axes. Effective application of multiple analyses and multiple displays of a single analysis ultimately provide a foundation for assessing how natural and cultural processes created structure in the data.

NOTE

[1] The author will furnish copies of the raw data on request. Please send a MS-DOS-formatted diskette and indicate the most useful kind of file format (ASCII, dBase, Gauss, etc.).

ACKNOWLEDGEMENTS

I thank Ronald L. Bishop for bringing to my attention the possibility of carrying out simultaneous RQ-mode analysis and directing me to some of the relevant literature. Thanks also to Michael D. Glascock for reading and commenting on an earlier draft of this paper. I am indebted to Michael Baxter, who provided copies of recently published and in-press articles. Finally, Morven Leese and two anonymous reviewers for *Archaeometry* provided extremely valuable critique based on review of an earlier version of this paper. Remaining problems in the paper are, of course, my own responsibility. I apologize to the reviewers for any misuse I may have made of their suggestions. Funding for the Archaeometry Laboratory at the Missouri University Research Reactor, where this research was carried out, comes from the U.S. National Science Foundation (DBS-9342768).

REFERENCES

Aitchison, J. A., 1986, *The statistical analysis of compositional data*, Chapman and Hall, London.

Baxter, M. J., 1989, Multivariate analysis of data on glass compositions: a methodological note, *Archaeometry*, **31** (1), 45–53.

Baxter, M. J., 1991, Principal component and correspondence analysis of glass compositions: an empirical study, *Archaeometry*, **33** (1), 29–41.

Baxter, M. J., 1992a, Archaeological uses of the biplot—a neglected technique?, in *Computer applications and quantitative methods in archaeology, 1991* (eds G. Lock and J. Moffett), 141–8, Brit. Archaeol. Rep. Internat. ser., **S577**, Tempus Reparatum, Oxford.

Baxter, M. J., 1992b, Statistical analysis of chemical compositional data and the comparison of analyses, *Archaeometry*, **34** (2), 267–77.

Baxter, M. J., 1993a, Comment on D. Tangri and R. V. S. Wright, 'Multivariate analysis of compositional data . . .', *Archaeometry*, **35** (1) (1993), *Archaeometry*, **35** (1), 112–15.

Baxter, M. J., 1993b, Principal component analysis in archaeometry, *Archaeologia e calcolatori* (in press).

Baxter, M. J., and Heyworth, M. P., 1989, Principal components analysis of compositional data in archaeology, in *Computer applications and quantitative methods in archaeology, 1989* (eds S. Rahtz and J. Richards), 227–40, Brit. Archaeol. Rep. Internat. ser., **S548**, Oxford.

Baxter, M. J., and Heyworth, M. P., 1991, Comparing correlation matrices: with applications in the study of artefacts and their chemical compositions, *Archaeometry '90* (eds E. Pernicka and G. A. Wagner), 355–64, Birkhäuser Verlag, Berlin.

Berthoud, T., Besenval, R., Cesbron, F., Cleuziou, S., Pechoux, M., Francuix, J., and Lisak-Hours, J., 1979, The early Iranian metallurgy: analytical studies of copper ores from Iran, *Archaeo-Physika*, **10**, 68–74.

Bishop, R. L., and Neff, H., 1989, Multivariate analysis of compositional data in archaeology, in *Archaeological Chemistry IV* (ed. R. O. Allen), 576–86, Am. Chem. Soc. Advances in Chemistry ser., **220**, Washington, D.C.

Corsten, L. C. A., and Gabriel, K. R., 1976, Graphical exploration in comparing variance matrices, *Biometrics*, **32**, 851–63.

Davis, J. C., 1986, *Statistics and data analysis in geology*, John Wiley and Sons, New York.

Glascock, M. D., 1992, Characterization of archaeological ceramics at MURR by neutron activation analysis and multivariate statistics, in *Chemical characterization of ceramic pastes in archaeology* (ed. H. Neff), 11–26, Monographs in World Archaeology, **7**, Prehistory Press, Madison, WI.

Gabriel, K. R., 1971, The biplot-graphic display of matrices with application to principal component analysis, *Biometrika*, **58**, 453–67.

Gower, J. C., 1966, Some distance properties of latent root and vector methods used in multivariate analysis, *Biometrika*, **53**, 453–67.

Gower, J. C., 1984, Multivariate analysis: ordination, multidimensional scaling and allied topics, in *Handbook of applicable mathematics VIB-statistics* (ed. E. Lloyd), 727–81, John Wiley and Sons, Chichester.

Jolliffe, I. T., 1986, *Principal component analysis*, Springer-Verlag, New York.

Joreskog, K. G., Klovan, J. E., and Reyment, R. A., 1976, *Geological factor analysis, methods in geomathematics, 1*, Elsevier Scientific Publishing Co., New York.

Leese, M., Hughes, M. J., and Stopford, J., 1989, The chemical composition of tiles from Bordesley: a case study in data treatment, in *Computer applications and quantitative methods in archaeology, 1989* (eds S. Rahtz and J. Richards), 241–9, Brit. Archaeol. Rep. Internat. ser., **S548**, Oxford.

Neff, H., 1984, *Developmental history of the Plumbate pottery industry in the eastern Soconusco region, A.D. 600–A.D. 1250*, unpublished Ph.D. dissertation, University of California, Santa Barbara.

Neff, H., and Bishop, R. L., 1988, Plumbate origins and development, *Am. Antiquity*, **53**, 505–22.

Neff, H., Bove, F. J., Lou, B., and Piechowski, M. F., 1992, Ceramic raw materials survey in Pacific coastal Guatemala, in *Chemical characterization of ceramic pastes in archaeology* (ed. H. Neff), 59–84, Monographs in World Archaeol., **7**, Prehistory Press, Madison, WI.

Peisach, M., Jacobson, L., Boulle, G. J., Gihwala, D., and Underhill, L. G., 1982, Multivariate analysis of trace elements determined in archaeological materials and its use for characterization, *J. Radioanalytical Chem.*, **69**, 349–64.

Poirier, J., and Barrandon, J. N., 1983, Non destructive analysis of Roman and Byzantine gold coins by proton saturation, *The proceedings of the 22nd symposium on archaeology* (eds A. Aspinall and S. E. Warren), 235–44, University of Bradford, Bradford.

Pollard, A. M., 1986, Multivariate methods of data analysis, in Jones, R. E., *Greek and Cypriot pottery: a review of scientific studies*, 56–83, Fitch Lab. Occas. Pap., **1**, Brit. Sch. Athens, Athens.

Shepard, A. O., 1948, *Plumbate: a mesoamerican tradeware*, Carnegie Institution of Washington Publications, **573**, Washington D.C.

Tangri, D., and Wright, R. V. S., 1993, Multivariate analysis of compositional data: applied comparisons favour standard principal components analysis over Aitchison's loglinear contrast method, *Archaeometry*, **35** (1), 103–15.

Teil, H., 1975, Correspondence factor analysis: an outline of its method, *Math. Geol.*, **7**, 3–30.

Underhill, L. G., and Peisach, M., 1985, Correspondence analysis and its application to multielemental trace analysis, *J. Trace & Microprobe Techniques*, **3**, 41–65.

Walder, J., Smith, J. P., and Dackombe, R. V., 1992, The use of simultaneous R- and Q-mode factor analysis as a tool for assisting interpretation of mineral magnetic data, *Math. Geol.*, **24**, 227–47.

Zhou, D., Chang, T., and Davis, J. C., 1983, Dual extraction of R-mode and Q-mode factor solutions, *Math. Geol.*, **15**, 581–606.

APPENDIX: COMPUTATION

The analyses presented in this paper, including basic data scaling operations, were undertaken using programs written by the author in Gauss, a matrix programming language and statistical package for IBM-compatible personal computers. The PCA program calculates principal components using Gauss eigenvalue/eigenvector extraction functions; an alternative approach would be to calculate principal components using a singular value decomposition function, which is also available in Gauss. Scatterplotting routines based on Gauss graphics functions permit display of various subsets of observations and variables, display of confidence ellipses calculated for the subgroups, and interactively changing the variable plotted on the y-axis in order to facilitate inspection of multiple principal components.

Data transformations and principal components analysis can also be accomplished with other computer statistical packages. SYSTAT, available for both PCs and Macintoshes, permits data transformations in its Data module, and provides standardized PC scores along with variable loadings through its Factor module. This output makes up a type 2 biplot; multiplying the component scores by the square root of the corresponding eigenvalue (singular value) removes the standardization, thus providing the scores plotted in biplots of types 1 or 3; for type 1 biplots, the component loadings must be divided by the square of the corresponding eigenvalue to yield variable coordinates. SAS Proc Princomp calculates either standardized or non-standardized scores along with eigenvectors and eigenvalues, thus also providing the information necessary for creating a type 3 biplot. Both SYSTAT and SAS also have graphics capabilities. Baxter (1992a) utilizes MINITAB along with STATGRAPHICS to generate biplots.