(1967); (d) T. Nagasawa, K. Kuroiwa, K. Narita, and Y. Isowa, *Bull. Chem. Soc. Jpn.*, **46** 1269 (1973); (e) S. Kim, J.I. Lee, and K.Y. Yi, *Bull. Chem. Soc. Jpn.*, in press (1985).
11. (a) K. Hoffmann, A. Johl, A.E. Furlenmeier, and H. Kappeler, *J. Am. Chem. Soc.*, **79**, 1636 (1957); (b) H. Schwarz, F.M. Bumpus, and I.H. Page, *J. Am. Chem. Soc.*, **79**, 5697 (1957); (c) W. Grassmann, E. Wünsch, and A. Riedel, *Chem. Ber.*, **91**, 449 (1958); (d) R.B. Merrifield, *J. Biol. Chem.*, **232**, 43 (1958); (e) F. Sakiyama, K. Okawa, T. Yamakawa, and S. Akabori, *Bull. Chem. Soc. Jpn.*, **31**, 926 (1958); (f) H. Voss, *Z. Naturforsch*, **206**, 116 (1965); (g) S. Sakakibara and M. Fujino, *Bull. Chem. Soc. Jpn.*, **39**, 947 (1966).

Characterization of Korean Clays and Pottery by Neutron Activation Analysis (I). Characterization of Korean Porcelainsherds

Chul Lee*, Oh Cheun Kwun, and Hyung Tae Kang

Department of Chemistry, Hanyang University, Seoul 133,

Received September 24, 1985

Data on the concentration of Na, K, Sc, Cr, Fe, Co, Cu, Ga, Rb, Cs, Ba, La, Ce, Sm, Eu, Tb, Lu, Hf, Ta, and Th obtained by neutron activation analysis have been used to characterize Korean porcelainsherds by multivariate analysis. The mathematical approach employed is principal component analysis (PCA). PCA was found to be helpful for dimensionality reduction and for obtaining information regarding (a) the number of independent causal variables required to account for the variability in the overall data set, (b) the extent to which a given variable contributes to a component and (c) the number of causal variables required to explain the total variability of each measured variable.

Introduction

The combination of the advances in multielement analysis and the development in the application of mathematical methods for the analysis of data has enabled valuable information to be extracted from analytical data, that would not have been readily accesible otherwise.[1-3] Archaeology is one of the major beneficiaries of such a combination approach,[4] the other notable areas of application being environmental science[5] and forensic science.[6] Pattern recognition (PR) techniques[1,2] have a significant role for the extraction of information from analytical data.

This paper reports on the work done in the development and application of principal component analaysis (PCA) for the grouping of Korean ancient porcelain-sherds, using elemental abundance data generated by present authors. The results obtained by PCA in the present work have been discussed in relation to those reported previously using other PR techniques such as single linkage and minimal spanning tree.[4]

Principal Component Analysis (PCA)

PCA is an approach akin to factor analysis and is known in pattern recognition literature as Karhunen-Loeve transformation. The set of data on N samples with M variables measured (N × M data) can be represented as a set of N points in M dimensional space. For convenience, the initial variables can be preprocessed and represented as deviation from their respective means so that all the variables are equally weighted. The new variables will be

$$Z_{ij} = (X_{ij} - \overline{X}_j) \quad (1)$$

for $i = 1, 2, 3, \cdots N$ and $j = 1, 2, 3, \cdots M$.

The product $Z' \cdot Z/N$ gives the variance-coveriance(VC) matrix S and is taken to represent the dispersion in the original data set. The objective of PCA is to generate abstract causal variables (factors) from matrix S using Eckart-Young theorem or by considering the projections of the N points in M dimensions onto some orthogonal axes such that the variances for the projected points are maximum in their respective directions.

The first principal component is a vector such that the perpendicular projections of the N points on to that vector have the largest possible variance. The second principal component is a vector such that it is orthogonal to the first and the projections of the N points onto it account for the largest fraction of remaining variance.

The new component variables Y_{ij} are given by the linear combination,

$$Y_{ij} = k1 \cdot Z1 + k2 \cdot Z2 + \cdots + km \cdot Zm \quad (2)$$

In matrix notion $Y = Z \cdot K$, where $Z(N,M)$ is the original data matrix, $K(M,P)$ is the matrix of coefficients and $Y(N,P)$ is the matrix of component scores. P is the rank of the VC matrix.

Derivation of the Coefficients in PCA.

N points are assumed to be projected onto a vector in M dimensional space such that the projections will have the maximum variance. For a vector to fit this condition in a M dimensional space, it is required that the sum of the squares of its direction cosines should be unity.

Let $K1$ be the $(M \times 1)$ vector of the direction cosines such that,

$$K1' \cdot K1 = 1 \quad (3)$$

The projection of a point Z_{ij} onto this vector is given by,

$$Y_{ij} = Z_{ij} \cdot K1 \quad (4)$$

For the projections of the N points, the variance of the projection is given by,

$$\frac{Y' \cdot Y}{N} = K1' \cdot Z' \cdot Z \cdot K1/N = K1'SK1 \quad (5)$$

To determine the linear combination with the requirement that $K1'K1 = 1$ and to maximize the variance of the projection given by eq.(5), we define,

$$F = K1'SK1 - L(K1'K1 - 1), \quad (6)$$

where L is the Lagrangian multiplier.

For maximising the variance, it is required that $\delta F/\delta K1 = 0$, *i.e.*,

$$\delta F/\delta K1 = 2SK1 - 2LK1 = 0. \text{ Thus,}$$
$$SK1 = LK1 \quad (7)$$

premultiplying eq.(7) by $K1'$, we have,

$$K1'SK1 = K1'LK1 = LK1'K1 = L \quad (8)$$

From eq.(5) and eq.(8), it is clear that L is the variance of the projections obtained using the set of coefficients $K1$.

Eq.(7) is the familiar eigen value problem involving the symmetric VC matrix S. In order to have nontrivial solutions for K it is required that the determinant $|S - LI|K = 0$. This implies that K is eigen vector of the matrix S and L is the corresponding eigen value. Since $|S - LI|$ is a Mth order polynomial, one can extract $P \leqslant M$ (P is the rank of matrix S) eigen vectors and corresponding eigen values.

It is obvious now, that to have the maximum variance for the projections of the N points, the eigen vector coefficients corresponding to the largest eigen value are to be used for obtaining the linear combinations. Procedures are available to extract the eigenvectors in a sequential manner such that $L1 > L2 > \cdots > LP$.[3] The different sets of eigen vector coefficients corresponding to the eigen values $L1$, $L2$ - - - - etc., are used to obtain the linear combinations generating the principal component 1, 2 and so on.

It is known that for a symmetric matrix like the one corresponding to S matrix the sum of eigen values is equal to the sum of diagonal elements. In the case of the VC matrix S, the corresponding diagonal elements are the variances of the individual variables and hence the sum of the eigen values is equal to the sum of the variances. Since the variances of the derived components are given by the eigenvalues the PCA transformation preserves the sum of the original variables.

Description and Analysis of Samples.

Eighty four samples of porcelainsherds from different sites in Korea were collected through museums. Archaeological data and their corresponding symbols are given in Table 1. In this table, the sites where the specimens were found are given together with the corresponding symbols used for the positions on the map shown in Figure 1. While the first column in the table gives the sample series number n for the present author's reference, the second column shows the symbols adopted to designate sampling sites in this work.

The third column gives group symbols defined previously[4] along with previous sample series number i of each group, those groupings had been carried out to find samples of known origin and known to be similar to each other. These samples were supposed to be used as a basis for latter identification and serves as batches of identified species. For this purpose, well selected samples had been grouped by single-linkage cluster analysis and minimal spanning tree analysis.[4] The number in the second part of the third column shows the number of samples in each previous group which has been included in this work because they are porcelainsherds. The

Table 1. Archaeological Data of Samples and Their Symbols

Series number n	Symbols with series number n	Symbols*	Number of samples	Sites		Item
1-9	▽(n)	▽i(n)	6	Kwangju	Kyonggi-do	White porcelain
10-15	▽(n)			Kwangju	Kyonggi-do	Punchong porcelain
16,17	▽(n)	▽i(n)	2	Yongin	Kyonggi-do	White porcelain
18-20	▼(n)	■i(n)	1	Yangju	Kyonggi-do	Celadon
21-23,27-29	◇(n)			Puan	Chollabuk-do	Celadon
24-26	◇(n)			Puan	Chollabuk-do	White porcelain
32-34	◈(n)			Kochang	Chollabuk-do	Punchong porcelain
30,31,35-39	◆(n)	■i(n)	1	Kangjin	Chollanam-do	Celadon
40-42	⊕(n)			Kwangju	Chollanam-do	Celadon
43	⊕(n)	□i(n)	1	Kwangju	Chollanam-do	Punchong porcelain
44-48	□(n)	□i(n)	4	Yongi	Chungchongnam-do	Punchong porcelain
49-52,54-61	■(n)	■i(n),□i(n)	4,1	Kongju	Chungchongnam-do	Punchong porcelain
53	■(n)			Kongju	Chungchongnam-do	White porcelain
62-70	○(n)	■i(n)	1	Total area	Kyongsangbuk-do	Punchong porcelain
71-73	○(n)			Yangsan	Kyongsangnam-do	Celadon
240-242	X(n)			Haenam	Chollanam-do	Celadon
243	X(n)			Shinan seabed	Chollanam-do	Celadon
244	X(n)			Shinan seabed	Chollanam-do	White porcelain
245-250	X(n)	X i(n)	5	Wando seabed	Chollanam-do	Celadon Bowl

*Symbols in previous work with sample series No. i of each group.

symbols in column 3 are adopted for the map of Figure 1 as well as for further explanation whenever the samples selected previously are concerned, otherwise the symbols in column 2 have been used.

The elemental analysis of porcelainsherds was carried out by thermal neutron activation analysis. The detailed analytical procedures had been described elsewhere.[4]

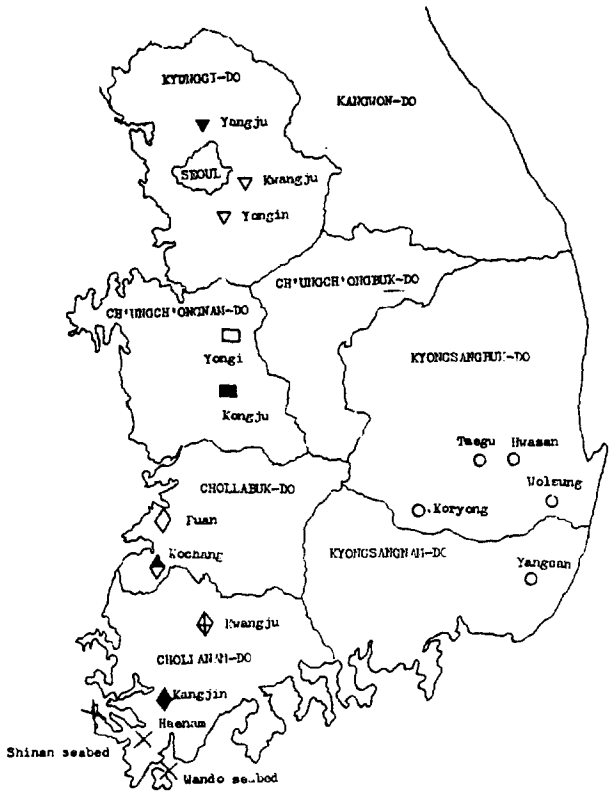

Figure 1. Map of Korea showing the sites at which samples were found. The sample symbols are defined in Table 1.

Results and discussion

Twenty elements (Na, K, Sc, Cr, Fe, Co, Cu, Ga, Rb, Cs, Ba, La, Ce, Sm, Eu, Tb, Lu, Hf, Ta, Th), which were analyzed by neutron activation analysis, have been used in the present PR for the grouping of 84 porcelainsherds collected from various areas in Korea. The *VC* matrix S(20×20) has been generated from original data set after transforming each data into the form of eq.(1). Twenty characteristic roots of S *i.e.*, 20 eigen values have been found. 20 characteristic vectors, *i.e.*, 20 eigen vectors corresponding to each eigen value has been calculated subsequently. Twenty principal components were finally obtained using similar equations to eq.(2) and (4). As the results, the main principal components, *i.e.*, first, second, third and forth components were found to be most attributed to major and minor constituent elements such as Na, K, Fe and Ba. The remaining components were found to be attributed by such ten trace elements as Sc, Cr, Co, Cu, Ga, Cs, La, Ce, Hf and Th. These elements were therefore selected for further PR analysis as follows.

Using the data of the selected ten elements, a *VC* matrix S(10×10) has been generated again. Ten eigen values thus obtained are given in Table 2. This table also shows the contribution (%) of each principal component to total variance S of the original variables. The eigen vectors corresponding to each eigen value, normalized with the sum of the squares being one, has been calculated and are given in Table 3. The communalities of the input variables in each of the derived components, *i.e.*, the fraction of total variance accounted for in each component are also given in Table 3 to show the importance of a given element in a component.

The principal components were derived using similar equations to eq.(2) and eq.(4).

The PCA plots of component 1 vs. component 2 and component 1 vs. component 3 are given in Figure 2 and Figure 3, respectively. The three different groups, covered by most samples, are clearly shown, *i.e.*, white porcelainsherds from Kyonggido symboled ▽(n) and celadon and punchong

Table 2. Eigen-values and Their Contribution (%)

Component	1	2	3	4	5	6	7	8	9	10
Eigen-value	8.4×10^4	1.0×10^4	7.3×10^3	1.1×10^3	2.6×10^2	2.3×10^2	1.2×10^2	7.2×10	3.7×10	8.0
Contribution (%)	81.22	9.87	7.10	1.11	0.00	0.00	0.00	0.00	0.00	0.00

Table 3. Eigenvector Coefficients and Communalities

Elements	Eigenvector coefficients			Communalities			Percentage of variance accounted for in		
	Comp 1	Comp 2	Comp 3	Comp 1	Comp 2	Comp 3	Comp 1	Comp 2	Comp 3
Cu	0.876	−0.440	0.180	64226.645	1970.050	236.858	96.67	2.97	0.36
Ga	0.002	−0.015	0.050	0.466	2.375	2.883	0.84	4.27	5.19
La	0.009	−0.078	−0.262	6.693	61.784	502.826	0.99	9.17	74.60
Ce	0.008	−0.195	−0.761	4.731	385.940	4238.322	0.09	7.70	84.56
Th	−0.001	−0.013	−0.056	0.082	1.831	20.260	0.06	1.37	15.21
Cr	0.056	−0.241	−0.494	261.676	1179.703	1789.222	6.62	29.82	45.23
Hf	0.002	−0.018	−0.049	0.214	3.390	17.412	0.66	10.52	54.03
Cs	−0.009	0.013	0.045	6.666	1.657	14.857	2.96	0.74	6.61
Sc	0.00	−0.027	−0.057	0.00	7.434	23.988	0.00	2.50	8.06
Co	0.479	0.838	−0.254	19188.463	7144.922	470.519	71.58	26.65	1.76

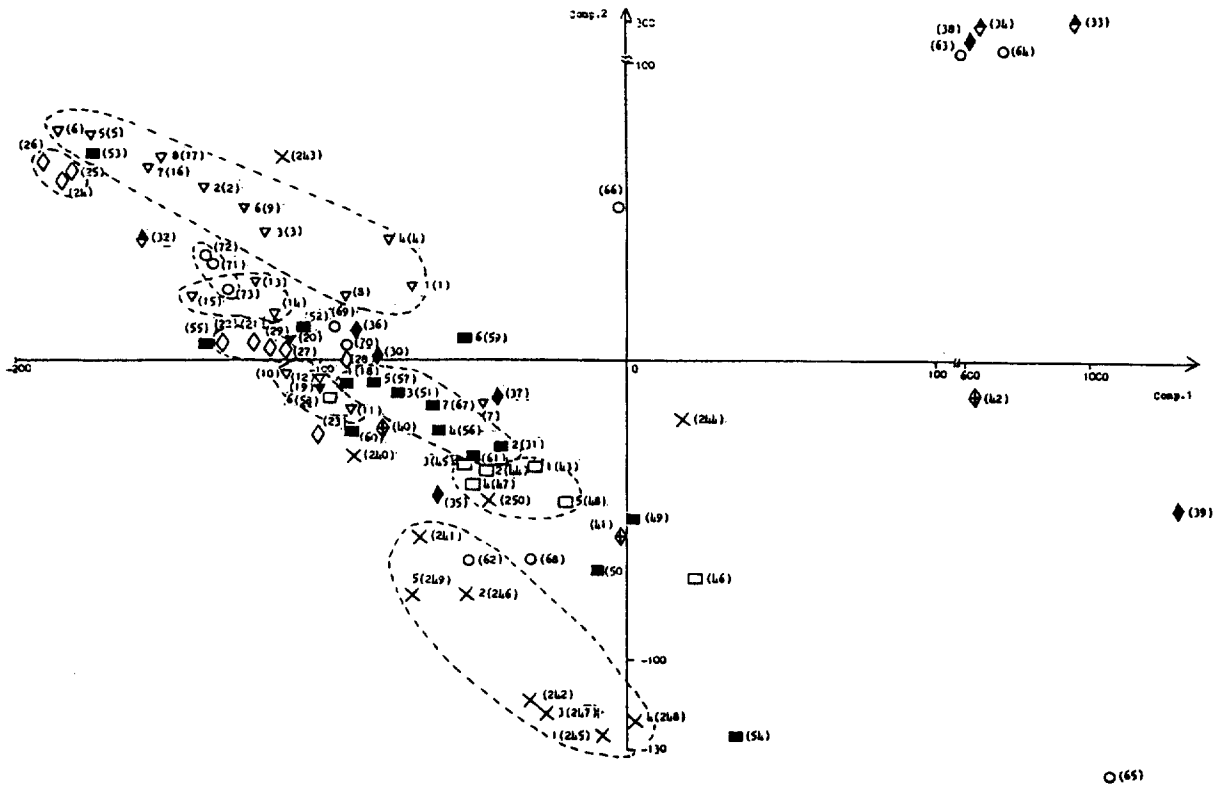

Figure 2. Principal components results.

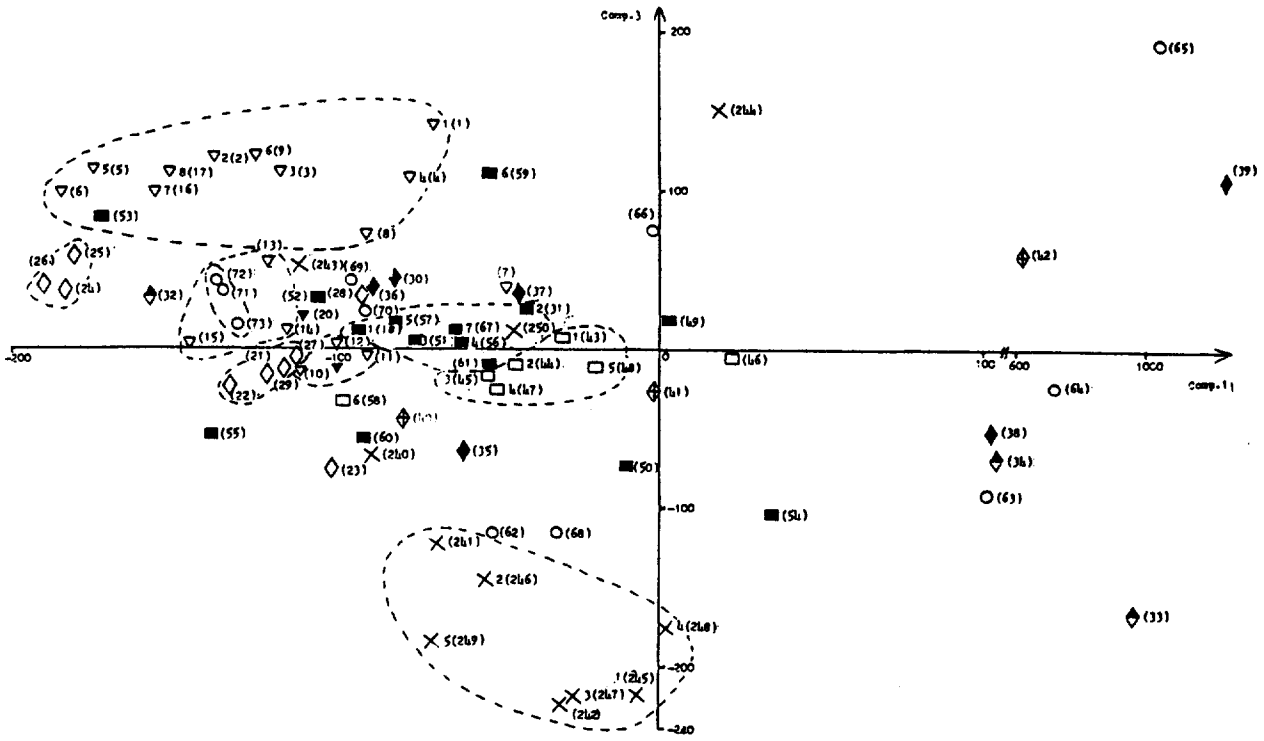

Figure 3. Principal components results.

porcelainsherds from the seashore areas of Chollanam-do symboled X(n) are clearly separated into respective groups, whereas the remaining samples are largely clustered into a group. This large group which is consisted of celadon and punchong porcelainsherds was found to be discriminated into several subgroups, *i.e.*, the groups symboled ◇(n), ■(n), □(n), etc.

It is also clear that the four groups, which were each symboled ∇_i(n), $\blacksquare_i$(n), $\square_i$(n) and X_i(n) and each defined as a batch of identified species as shown in column 3 of Table 1, are so tightly clustered that some other samples shown in column 2 of Table 1 can be assigned to any of these previous groups, based on the proximity to the groups.

Figure 2 shows that the samples excavated from Puan, covered by symbol ◇(21-29) are clustered into two subgroups, depending on difference between species, *i.e.*, white porcelainsherds covered by ◇(24-26) and celadonsherds covered by ◇(21-23) and ◇(27-29). The subgroup of white porcelainsherds is, however, relatively proximated to the group of other white porcelainsherds symboled ▽(n). These results show the proximity among white porcelainsherds even though their geological nature is different. Figure 3 shows the subgroup symboled by ◇(24-26) is a little further separated from the proximated group symboled ▽(n).

The celadon porcelainsherds recovered from a seabed near Wando, covered by symbol X(245-250), are found to form a group along with other celadon porcelainsherds excavated from Haenam. These results suggest that the samples recovered from the seabed are originated from Haenam. Two samples symboled X(243) and X(244) recovered from a Shinan seabed, *i.e.*, a celadonsherd and a white porcelainsherd, are found to be outlain as shown in Figure 2 and Figure 3. This result is in accord with the suggestion by some archaeological specialists who believe that the samples had been transferred from China along with a sunken ship. Punchong porcelainsherds symboled ▽(10-15) excavated from Kwangju are found to be separated into two subgroups as shown in Figure 2 and Figure 3. These two groups are found to have a little different geological origins.

While Figure 2 still shows two samples given by symbol ○(62, 68) and three samples given by ○ symbol (71-73) are proximated to the group covered by X (n) and to the group covered by ▽(n), respectively, Figure 3 shows the latter three samples are separated along with samples symboled ▽(13-15) into a independent subgroup.

Ninety percent of total variance had been used for the two dimensional plots in Figure 2 and Figure 3. While component 1 is most attributed by Cu and Co, component 2 is much attributed by Cr and Co as shown in Table 3. The same table also shows that component 3 is attributed by Ce, La, Cr and Hf.

Acknowledgement. The present studies were supported (in part) by the Basic Science Research Institute Program, Ministry of Education, 1985.

References

1. M.J. Karson, "Multivariate Statistical Methods", Iowa State University Press, 1982.
2. K. Varmuza, "Lecture Notes in Chemisty", "Pattern Recognition in Chemistry", New York, 1980.
3. D.N. Lawley and A.E. Maxwell, "Factor Analysis as a Statistical Method", Butterworths,, London, 1963.
4. C. Lee, O.C. Kwun, N.B. Kim and I.C. Lee, *Bulletin of Korean Chem. Soc.*, **6**, 241 (1985).
5. J. Arunachalam and S. Gangadharan, *Analytics Chimica Acta*, **157**, 245 (1984).
6. J. Arunachalam and S. Gangadharan, *J. Indian Acad. Forensic Sci.*, **20**, 54 (1981).