COMPOSITIONAL GROUPINGS OF SOME ANCIENT NATIVE POTTERY MADE IN MARAJÓ ISLAND

Rosimeiri G. Toyota[1], Casimiro S. Munita[1], Mário O. Menezes[1], Eduardo G. Neves[2], and Célia C. Demartini[2]

[1] Instituto de Pesquisas Energéticas e Nucleares (IPEN / CNEN - SP)
Av. Professor Lineu Prestes, 2242
05508-000 São Paulo, SP
rosimeiritoy@yahoo.com.br
camunita@ipen.br
mario@ipen.br

[2] Museu de Arqueologia e Etnologia da Universidade de São Paulo (MAE-USP)
Av. Prof. Almeida Prado, 1466 Cidade Universitária
05508-900 São Paulo, SP
egneves@usp.br
crismartine@yahoo.com

ABSTRACT

In this work 161 Marajoara ceramics fragments were studied by means of the Instrumental Neutron Activation Analysis (INAA) to determine As, Ba, Ce, Co, Cr, Cs, Eu, Fe, Hf, K, La, Lu, Na, Nd, Rb, Sb, Sc, Sm, Ta, Tb, Th, U, Yb, and Zn concentrations with the purpose to classify and to order the ceramics which are related to one another in their chemical compositions. The analytical method is right for this type of study because it is a non-destructive technique, however, the sample procedure is destructive, besides being highly sensitivity, accuracy and precision to determine chemical elements in trace and ultra trace levels. To study the data set multivariate statistical analysis techniques were used. Initially the concentrations were normalized using log base 10. Following the outliers were studied by means of Mahalanobis distance. Finally the data were interpreted by means of components analysis and discriminant analysis.

1. INTRODUCTION

The understanding and characterization of archeological objects, as a means to rebuild a ancient native culture has an important significance in archaeological studies. In recent years, research about ceramic artifacts has perfected itself due to contribution from sciences such as physics and chemistry. [1]

Archaeometry uses several physical and chemical techniques in order to obtain as much information as possible about the materials being studied. Especially, the techniques for the chemical characterization of ceramic remains have been intensively used given the several issues that can be addressed through their use.[2] Various analytical techniques such as AAS, ICP, PIXE, XRF, PIGE and INAA, are used for elementary analysis. Among several techniques, INAA, using gamma ray spectrometry, seems to be one of the most appropriate analytical technique, while not demanding dissolution of samples and allowing for simultaneous multielementary chemical determination, with high levels of sensitivity, accuracy and precision. In multi-elementary chemical analysis a great number of samples are analyzed, in which several elements with similar chemical properties are determined. In this

kind of study it is indispensable to apply multivariate statistical techniques. In this work the concentrations of As, Ba, Ce, Co, Cr, Cs, Eu, Fe, Hf, K, La, Lu, Na, Nd, Rb, Sb, Sc, Sm, Ta, Tb, Th, U, Yb and Zn were determined in a set of 161 Marajoara ceramic fragments.

1.1. Statistical Treatment

The application of the statistical method in experimental multivariate results seeks the classification and the ordination of related objects regarding their chemical composition. The data set were studied by means of cluster analysis [3], principal components analysis [4], discriminate analysis [5] and Procrustes analysis [6].

Procrustes analysis consists of a criterion with the following algorithm: an $\mathbf{X}$ (data matrix with an $\mathbf{n}$ x $\mathbf{p}$ dimension matrix, where $\mathbf{n}$ = sample and $\mathbf{p}$ = variable) is initially considered; assuming that the essential data dimension to be used in a given comparison is $\mathbf{k}$. (Attention to insure that enough data variability was explained as to the choice of $\mathbf{k}$); next, $\mathbf{Y}$ being the matrix ($\mathbf{n}$ x $\mathbf{k}$) of the scores of principal components (it produces the best $\mathbf{k}$-dimensional proximity of the original $\mathbf{X}$ data configuration); to select $\mathbf{q}$ of the original $\mathbf{p}$ variables. Being $\mathbf{q} < \mathbf{p}$ (for the selection to take place), $\mathbf{q} \geq \mathbf{k}$ (so that there is the possibility of recovering the real structure with the selected variables); the $\mathbf{n}$ x $\mathbf{q}$ dimension $\tilde{X}$ is considered as selected variable matrix, and finally, $\tilde{Z}$, being ($\mathbf{n}$ x $\mathbf{k}$) dimension matrix of the principal components scores of the reduced data (it produces the best $\mathbf{k}$-dimensional proximity of the defined $\mathbf{q}$-dimensional configuration in the data subset).

If the real data dimension is really $\mathbf{k}$, then $\mathbf{Y}$ can be seen as the real configuration, and $\tilde{Z}$ is the corresponding approximate configuration based only in $\mathbf{q}$ variables.

To measure the discrepancy among these configurations the Procrustes analysis was used, measuring the adjustment among the two configurations by applying the residual (M^2) sum of squares, which measures the loss of information on the data structure when only the selected $\mathbf{q}$ variables are used instead of $\mathbf{p}$ among the two configurations.[6]

Being the configurations: $\mathbf{Y}$ dimension matrix $\mathbf{n}$ x $\mathbf{k}$ and $\tilde{Z}$ dimension matrix ($\mathbf{n}$ x $\mathbf{k}$), then

$$M^2 = \mathrm{tr}\{YY' + \tilde{Z}\tilde{Z}' - 2\tilde{Z}Q'Y'\} \tag{1}$$

which can be rewritten as

$$M^2 = \mathrm{tr}[YY'] + \mathrm{tr}[\tilde{Z}\tilde{Z}'] - 2\mathrm{tr}[\Lambda] \tag{2}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, ... , \lambda_k)$

$$Q = VU' \tag{3}$$

where $\mathbf{Q}$ is the orthogonal matrix with $\mathbf{k}$ x $\mathbf{k}$ dimension;

$\mathbf{U}$, Λ and $\mathbf{V}$ are obtained from the decomposition by singular value of the dimension matrix $\mathbf{k}$ x $\mathbf{k}$, that is

$$_k\tilde{Z}'Y_k = U\Lambda V' \qquad (4)$$

A stop rule to determine an appropriate value for **i**, given **k**, have been discussed by Krzanowski [7], who showed that the number of **k** variables which should be kept in this process is related with the equation

$$\left(1 + e^2\right)\sigma^2\chi^2_{(a)} \qquad (5)$$

with [**nk** - 1/2 **k** (**k**+1)] degrees of exemption and $e = \sqrt{(p\text{ - }i\text{ - }k)/(p\text{ - }k)}$, where **i** is the number of eliminated variables. According to with Krzanowski, if the removed variable is important to explain the data structure, the sum of residues will be higher than the critic values (*cv*). Thus, Krzanowski establishes a stop rule for Procrustes analysis: the process of variable removal continues until $M^2 > cv$. In the expression the critical value σ^2 is unknown. However, it can be substituted by its estimator.[7]

The Procrustes analysis constitutes an objective tool for variables elimination without altering the data structure. The reduction of data dimensionality in the ceramic compositional analysis is desirable, since it facilitates the interpretation of the results. Another advantage constituted by the dimensionality reduction regards the subject of the variables that should be determined for samples to be analyzed later. This procedure seeks to save human and materials resources, while reducing the time spent in analysis.

2. EXPERIMENTAL

2.1. Sample Preparation and Standard

Ceramic powder samples were obtained by cleaning the outer surface and drilling to a depth of 1-2 cm using a tungsten carbide rotary file attached to the end of a flexible shaft, variable speed drill. Depending on the thickness, 3 or 5 holes were drilled as deep into the core of the fragment as possible without drilling through the walls. Finally, the powered samples were dried in an oven at 105°C for 24 h and stored in desiccators.

Constituent Elements in Coal Fly Ash - NIST-SRM-1633b, were used as standard in all analysis. The standard reference material Brinck Clay - NIST-SRM-679 was used to check the analytical quality of the results The standard and the samples were dried in an oven at 105°C, the standards for 4 h and samples for 24 h and stored in a desiccator until weighing.

2.2. Irradiation and Radioactivity Measurements

About 100 mg of ceramics samples, and NIST-SRM-1633b were weighed in polyethylene bags and involved in aluminum foil. Groups of 8 ceramic samples and two reference materials were packed in aluminum foil and irradiated in the swimming pool research reactor, IEA-R1m (IPEN/CNEN – SP) at a thermal neutron flux of about 5 x 10^{12} n·cm^{-2}·s^{-1} for 8h.

Two measurement series were carried out using Ge (hyperpure) detector, model GX 2020 from Canberra, resolution of 1.90 keV at the 1332.49 keV gamma peak of ^{60}Co, with S-100 MCA of Canberra with 8192 channels. As, K, La, Lu, Na, Nd, Sm, U, and Yb were measured

after 7 days decay and Ba, Ce, Co, Cr, Cs, Eu, Fe, Hf, Rb, Sb, Sc, Ta, Tb, Th, and Zn after 25-30 days. Gamma ray spectra analysis was carried out using the software Genie 2000 NAA Procedure from Canberra.

3. RESULTS AND DISCUSSION

With the purpose of studying the quality control of the analytical method, the concentrations of As, Ba, Ce, Co, Cr, Cs, Eu, Fe, Hf, K, La, Lu, Na, Nd, Rb, Sb, Sc, Sm, Ta, Tb, Th, U, Yb and Zn were determined in 18 samples of the IAEA Soil 7 reference material. From these data some statistical parameters were calculated, such as the average, standard deviation, level of average reliability of the, homogeneity of results, and precision and accuracy. The results showed that most of the elements presented a $\leq 10\%$ precision. This precision is considered appropriate by several authors as for the choice of the chemical elements for chemical characterization studies of archeological objects using multivariate statistical methods.[2]

Elements such as Co and Ta were eliminated, although they presented a precision lower than 10%, because there is evidence of contamination by the tungsten drill during the preparation of the sample.[8] ^{65}Zn was also eliminated because it undergoes an interference in the spectrum of gama rays at the ^{46}Sc peak.[9] Although As, Nd, Ba, Sb and Rb present a good precision, previous studies showed that they are not reliable elements to be inserted in the data base because they show a significant dispersion in ceramics concentrations. Therefore, the elements used were K, La, Yb, Lu, U, Sc, Cr, Fe, Cs, Ce, Eu, Tb, Hf, and Th.

Initially, the data of elementary concentrations of the samples of Marajoara ceramics fragments (161) were transformed in $\log_{10}$ to compensate for the difference in magnitude among elements given in percentages and at trace level. The transformation of the concentrations into $\log_{10}$ before applying multivariate statistical methods is a usual procedure in archaeometric studies. One of the reasons for this is that the normal distribution of the elements in the soil is logarithmic. Another reason for the logarithmic transformation it tends to stabilize the variance of the variables, which would have an approximately equal weight in a multivariate statistical analysis.

The study of the outliers values was made by means of the *Mahalanobis* distance, while using the lambda *Wilks* criterion as critical value.[10] For the sample that showed Mahalanobis distance values higher than the critical value were eliminated. After the elimination of the outliers, the *Mahalanobis* distance was calculated again in the new group of data. This process was repeated until all of the samples showed Mahalanobis distance values lower than the critical value. In all, 12 outiliers were found.

Next, a preliminary classification was accomplished through cluster analysis by Ward and Euclidian distance methods. The dendrogram showed the existence of two very defined groups. The study carried by means of the discriminant analysis also it presented the existence of two groups. Figure 1 shows the discriminant function 1 versus the discriminant function 2 with a confidence level of 95%. It can be seen clearly that the elementary concentrations of the Marajoara ceramics fragments are divided into two very well-defined groups.
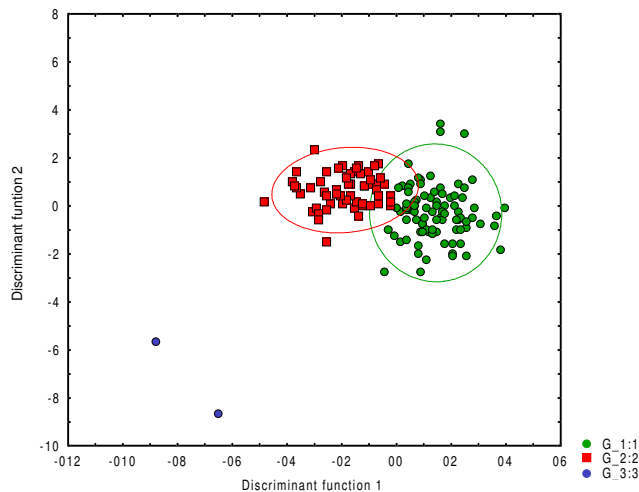
Figure 1. Discriminant function 1 versus discriminant function 2 using all variables. The ellipses represent a confidence level of 95%.

The data were submitted to the principal component analysis and again it showed the existence of two groups. Figure 2 shows the scores of principal component 1 versus principal component 2.
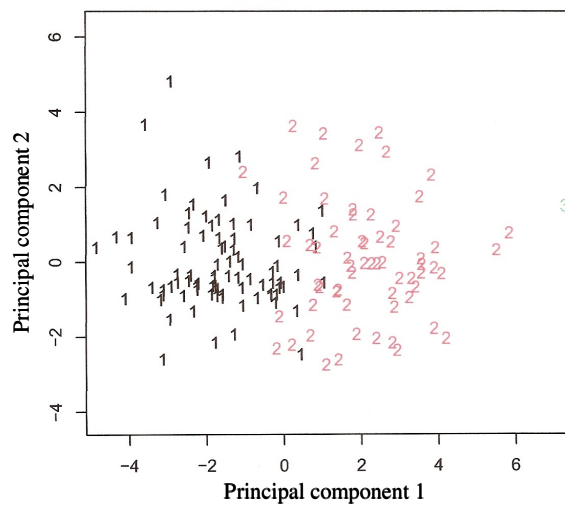

Figure 2. Principal component 1 versus principal component 2 using all variable.

With the purpose of verifying the possibility of the reduction of data dimensionality in compositional analysis or, in other words, to eliminate variables without altering data structure, the data were analyzed through the Procrustes analysis, in charge of selecting the interesting variables. Table 1 shows the variables selection procedure including the elimination sequence, the bidimensionality distance of the configuration of the main components using all the data (M^2) and the critical value (cv) of 5% (stop rule cv). In Table 1 La is the first element to be eliminated, since the M^2 value was 10.5, which measures the proximity of the bidimensionality of the configuration of the principal components. This

leads to the fact that for La the 577.6 value is the *cv* with 5% of significance level. Since M^2 is lower than that value, the elimination of La does not significantly affect the configuration of the principal components. Once the variables are eliminated by the configuration distance of the principal components, M^2 increases and the critical value decreases, which, in turn, depends on the number of variables, which also decreases.

Table 1. Selection of variables with elimination sequence

	La	**U**	**Lu**	**Cr**	**Na**	**Th**	**Yb**	**Hf**	**Ce, Fe, Eu, K e Sc**
M^2	10.5	29.4	73.5	99.0	144.3	189.0	241.0	***385.1***	
cv	557.6	531.1	478.0	451.4	424.8	398.3	371.7	***318.6***	

When the variable is eliminated, the associated configuration is altered to the point when it becomes unacceptable, which happens when M^2 is larger than the *vc*. That point is reached when the M^2 value is higher than the *vc*. This procedure suggests that Na, La, Yb, Lu, U, Cr, Cs, Tb and Th are eliminated and that Hf, Ce, Fe, Eu, K and Sc should be considered.

To verify if the behavior of the selected variables represents the complete data structure, a new analysis of principal components was made using the data matrix with 149 samples and the 6 selected elements (Hf, Ce, Fe, Eu, K and Sc). Figure 3 shows the principal component 1 scores versus principal component 2 of the selected variables.
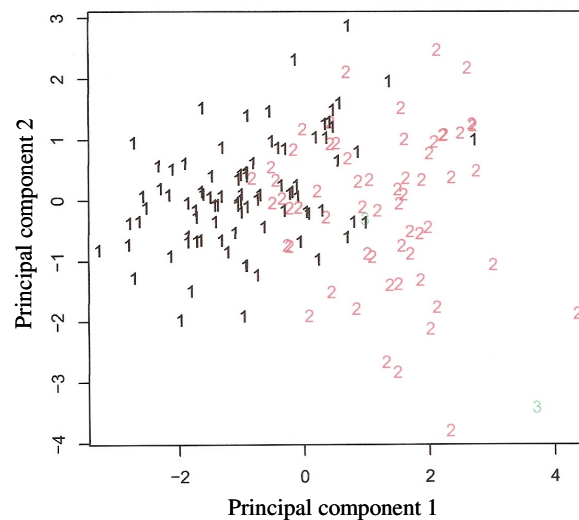

Figure 3. Principal component 1 versus principal component 2 using the select variables (Hf, Ce, Fe, Eu, K and Sc).

The formation of two compositional groups can be clearly observed, confirming that the result of the principal component analysis using the six elements reproduces results that are similar to those from principal component analysis using all the variables. It shows that the reduction of variables is viable, permitting a savings of human and materials resources,

besides allowing for a reduction of the time spent in analysis. This reduction technique of data dimension can be useful with authenticity tests of ceramic artifacts, because a small number of variables can be used for this purpose.

3. CONCLUSIONS

In this work, it was verified that the precision of the analytical method for the studied elements was good. The analyses by discriminant, principal components and Procrustes showed the existence of two groups, indicating that different raw materials were used in the manufacturing of the artifacts. The Procrustes analysis showed that the reduction of variables is viable, permitting savings of human and materials resources, while reducing the time spent in analysis. Furthermore, complementary studies of data dimension reduction could be useful to authenticity tests of ceramic artifacts.

ACKNOWLEDGMENTS

REFERENCES

1. C.S. Munita, M.A Silva, S.A.Silva, P.M.S. Oliveira. "Archaeometric study of clay deposits from the indigenous land of the Asurini do Xingu", *Instrum. Sci. .Technol.*, **33**, pp. 169-181 (2005).
2. C.S. Munita, R.P. Paiva, M.A. Alves, P.M.S. Oliveira, E.F. Mamose. "Contribution of neutron activation analysis to archaeological studies", *J. Trace Microprobe Tech,* **18(3)**, pp. 381-387 (2000).
3. I.T. Jolliffe, B. Jones, J.T. Morgan. "Identifying influencial observations in hierarchical cluster analysis", *App. Statist.*, **22(1)**, (1995).
4. U.M. Vinagre Filho, R.M. Latini, A.V.B. Bellido, A. Buarque, A.M. Borges. "Ancient ceramic analysis by activation in association with multivariate methods*". Brazilian Journal of Physics*, **35(3)**, pp. 779-781 (2005).
5. M.J. Baxter, C.M. Jackson. "Variable selection in artificial compositional studies", *Archaeometry*, pp. 253-268 (2001).
6. R. Sibson. "Studies in the robustness of multidimensional scaling", *J. Roy. Stat. Soc. B*, **40**, pp. 234-238 (1978).
7. W.J. Krzanowski. "A stopping rule for structure – preserving variable selection", *Statistics And Computing*, **6(1),** pp. 51-56 (1996).
8. M. Attas, J.M. Fossey, L. Yaffe. "Corrections for drill-bit contamination in sampling ancient pottery for neutron activation analysis", *Archaeometry*, **26(1)**, pp. 104-107 (1984).
9. I. Perlman, F. Asaro. "Pottery analysis by neutron activation", in *Science and Archaeology,* RH Brill, London (1971).
10. K.I. Penny. "Appropriate critical values when testing for a single multivariate outlier by using the Mahalanobis distance", *Appl. Statist.* ; **35:** 153-162 (1987)**.**