2009 International Nuclear Atlantic Conference - INAC 2009
Rio de Janeiro, RJ, Brazil, September 27 to October 2, 2009
ASSOCIAÇÃO BRASILEIRA DE ENERGIA NUCLEAR - ABEN
***ISBN:* 978-85-99141-03-8**

NEURAL NETWORK APPLIED TO ELEMENTAL ARCHAEOLOGICAL MARAJORA CERAMIC COMPOSITIONS

Rosimeiri G. Toyota[1], Casimiro S. Munita[1], Clodis Boscarioli[2,3], Emilio D. M. Hernandez[3], Eduardo G. Neves[4] and Celia C. Demartini[4]

[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

[2]Universidade Estadual do Oeste do Paraná, Centro de Ciências Exatas e Tecnológicas, Colegiado de Informática.
Rua Universitária, 2069
85819-110 Cascavel, PR
boscarioli@gmail.com

[3]Universidade de São Paulo, Escola Politécnica.
Av. Professor Luciano Gualberto, 158, Travessa 3
05508-900 São Paulo, SP
boscarioli@gmail.com

[4] Museu de Arqueologia e Etnologia da Universidade de São Paulo (MAE-USP).
Av. Prof. Almeida Prado, 1466
05508-900 São Paulo, SP
eduardo@pq.cnpq.br

ABSTRACT

In the last decades several analytical techniques have been used in archaeological ceramics studies. However, instrumental neutron activation analysis, INAA, employing gamma-ray spectrometry seems to be the most suitable technique because it is a simple analytical method in its purely instrumental form. The purpose of this work was to determine the concentration of Ce, Co, Cr, Cs, Eu, Fe, Hf, K, La, Lu, Na, Nd, Rb, Sb, Sc, Sm, Ta, Tb, Th, U, Yb, and Zn in 160 original marajoara ceramic fragments by INAA. Marajoara ceramics culture was sophisticated and well developed. This culture reached its peak during the V and XIV centuries in Marajó Island located on the Amazon River delta area in Brazil. The purpose of the quantitative data was to identify compositionally homogeneous groups within the database. Having this in mind, the data set was first converted to base-10 logarithms to compensate for the differences in magnitude between major elements and trace elements, and also to yield a closer to normal distribution for several trace elements. After that, the data were analyzed using the Mahalanobis distance and using the lambda Wilks as critical value to identify the outliers. The similarities among the samples were studied by means of cluster analysis, principal components analysis and discriminant analysis. Additional confirmation of these groups was made by using elemental concentration bivariate plots. The results showed that there were two very well defined groups in the data set. In addition, the database was studied using artificial neural network with unsupervised learning strategy known as self-organizing maps to classify the marajoara ceramics. The experiments carried out showed that self-organizing maps artificial neural network is capable of discriminating ceramic fragments like multivariate statistical methods, and, again the results showed that the database was formed by two groups.

1. INTRODUCTION

Archaeometry uses several physical and chemical techniques to obtain as much information as possible about the materials being studied, such as ceramic remains. Techniques of chemical characterization of ceramic remains have been intensively used, due to the several hypotheses that can be clarified [1,2].

Because of their high resistance to environment conditions, ceramic remains are the most commonly found remnants in archeological excavations. Due to their abundance and durability, ceramics are extremely important indicators of socio-cultural and economic interactions of the peoples who used them. Hence, the characterization of ceramic remains is of great interest as it allows the clarification of aspects of the lifestile of ancient groups in order to assess their cultural, economic and technological status.

Techno-typological profiles such as decoration, color, form and function contribute to the classification of ceramics. However, these variables associated to results of the analytical techniques present extra information for the studies of the origin of the samples, as well as to the investigation of the production of the ceramic artifacts.

Here we report a study on ceramics finds from marajoara earth mounds (tesos). The marajoara culture, located in Marajó Island, northern Brazil, was one of the most complex societies in South America. They occupied almost the whole Marajó Island, between 400 and 1300 AD. Ceramics in this culture was a collective activity of certain localities. They produced vessels and objects for personal use or for the community. Since ceramic production is traditionally considered as a woman activity in Amazon, the participation of men was mainly in the transportation of clay, handling and firing of big pieces [3].

In this study, the 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 were defined in a group of 160 marajoara ceramic fragments from *Ethnology and Archaeology Museum* – São Paulo University, through instrumental neutron activation analysis (INAA). Presently, INAA is one of the most successful analysis techniques for composition analysis studies.[3] By this method it is possible to simultaneously determine more than 30 elements to trace and ultra-trace levels, with high levels of precision and accuracy. Additionally, being an instrumental technique, INAA poses a relative handiness for the preparation of samples, which results in a reduction in experimental errors and in the time spent in analysis. Origin studies may be performed by using Instrumental Neutron Activation Analysis (INAA), as the trace elements may present unique compositions and hence, be an indicator of the origin of the raw materials used in the ceramic production.

Due to the large amount of analyzed samples an to the number of determined variables (elements), the multivariate statistical analysis is an essencial technique for the result interpretation. The use of statistical methods in experimental data aims to classify and order objects related to each other as a function of their chemical composition.

Among the various statistic techniques, the discriminant analysis and the artificial neural network analysis are used in this study.

1.1 Discriminant Analysis

Discriminant analysis is a multivariate statistical technique used to discriminate populations and/or to classify objects in previously defined populations. Thus, the main objective of the technique is to find functions of the original variables (discriminant functions) that explain the differences among the populations and that permit to locate new objects in one ot the populations included in the analysis.

Discriminant analysis is a supervisioned techinique because in this kind of analysis it is necessary to know a priori the population to whom the objects belong to. To use discriminant analysis the number of populations must be well defined.

In archeometric studies, discriminant analysis has been used in studies of the origin of raw materials and in grouping ceramic remains according to their similarities [3,4].

1.2 Self-Organizing Maps (SOM)

SOM is based in a map of neurons, whose weighs are adapted to the similar input vectors present in a training set [5]. During the training, the SOM behaves as a flexible network, which folds inside clouds formed by the data vectors involved in the training. Due the neighbor relation, neighbor neurons are dragged in the same direction, indicating that the codebook vectors of the neighbor neurons become similar during the learning process. For each neuron is registered its proper value, since a proximity function to the input data.

The exploratory data analysis is an approach for the data analysis that applies a variety of visualization techniques and data clustering in the search for patterns in a database. The data set are reduces to codebook vectors of SOM, which it can be used by other techniques to visualization or to data clustering.

Besides the reduction of the data set for analysis, another advantage of SOM is that it is not necessary to recalculate the map for each new input item of data, therefore, if the statistics can be assumed as stationary, new data can directly be mapped to the codebook vector representative of the item of the nearest data to the old model.

To discovery information from clustering, a processing technique is necessary on the neurons. For such, there are several approaches being distinguished the visual inspection and the application of clustering techniques on the neurons.

One of the SOM objectives is to represent input patterns of high dimensionality with codebook vectors, so that they can be visualized, of a facilitated form, in a map of lower dimension, generally 2-D, once that the limited number of visual dimensions is one of the problems of the visualization of multidimensional data.

The SOM algorithm has been, for some years, used as base for the development of some algorithms for data clustering. This approach has two phases: first it uses SOM and after, applies some clustering techniques, to cluster the data [6,7,8,9].

In the literature there are few papers that use SOM in archaeometric studies [10].

2. EXPERIMENTAL

2.1 Equipment and sample preparation

The measurements of induced gama activity were executed using a Canberra Ge hiperpure detector, resolution of 1.90 keV at the 1332 keV of the ^{60}Co peak, a Canberra S-100 MCA plate with 8192 channels and associated electronics. The spectra of gama rays were obtained and analyzed through the Genie-2000 NAA Processing Procedure program, developed by Canberra.

The ceramics fragments were initially washed with Milli Q water, removing the external surface with a fine bristled brush. Then, the external surface of the fragments was eliminated with rotating tungsten carbide file, adapted to a multi-speed drill, in order to eliminate contamination. About 500 mg of sample powder were obtained by means drilling 3 to 5 holes in the internal part of the fragments, keeping the drill from perforating through their walls. This powder was then collected, dried in an oven at 105ºC for 24 hours and stored in a desiccators [11].

2.2 Analytical procedure

For the analysis, about 100 mg of each sample were weighed in polyethylene envelopes and sealed with a solder iron. The envelopes were wrapped in aluminum foil. A set of eight samples together with about 100 mg of reference material. Standard Reference Material - NIST-SRM 1633b Constituent Elements in Coal Fly Ash, was used as pattern. The samples were submitted to an 8 hour irradiation at the IEA Reactor - R1m at IPEN - CNEN / SP, at a thermal neutron flux of about $5 \times 10^{12} n\ cm^{-2} s^{-1}$ for 1 h.

Two measurements were made to determine, As, K, La, Lu, Na, Nd, Sb, Sm, U and Yb after seven days of decay time, and Ba, Ce, Co, Cr, Cs, Eu, Fe, Hf, Rb, Sc, Ta, Tb, Th and Zn after 25-30 days of decay time [12].

2.3 Analytical quality control

The precision, accuracy and the sensibility of the method were studied by means of the analysis of As, Ba, Ce, Co, Cr, Cs, Eu, Fe, Hf, K, La, Lu, In the, Nd, Rb, Sb, Sc, Sm, Ta, Tb, Th, U, Yb and Zn in the IAEA Soil 7 reference material.

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 [3]. 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 [13]. Zn was also eliminated because it undergoes an interference in the spectrum of gama rays at the Sc peak [14]. 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 (160) 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 [15]. 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, 13 outliers were found and not included in the experimental evaluation.

Next, a preliminary classification was accomplished through hierarchical cluster analysis by Ward and Euclidean distance methods in the dada of the 147 marajoara ceramics fragments.

The dendrogram showed the existence of two much defined groups. The study carried by means of the discriminante analysis also it presented the existence of two groups. Figure 1 shows the discriminant function 1 versus the discriminant function 2. It can be seen clearly that the elementary concentrations of the marajoara ceramics fragments are divided into two very well-defined groups.

With the purpose of confirming the existence of these two groups, the data were studied by means of the SOM to generate a topological ordinance of the data and then to find the existing groups in the data set. The main idea is to make use of the preservation property of topological neighborhood of the data in the SOM mapping as clustering criterion, using it as validation form to this process.

This methodology is divided in four steps [16].

1. To train the SOM map. In this stage a SOM map is trained.

2. To build a dendrogram. This stage corresponds to the calculation of the minimum proximity between the neurons of the map. The objective here is to find the neighboring neurons that group for the maximum similarity. Each neuron BMU is a leaf in the dendrogram, and the other levels are defined by the linking's (junctions) between the neurons, having formed an initial grouping.

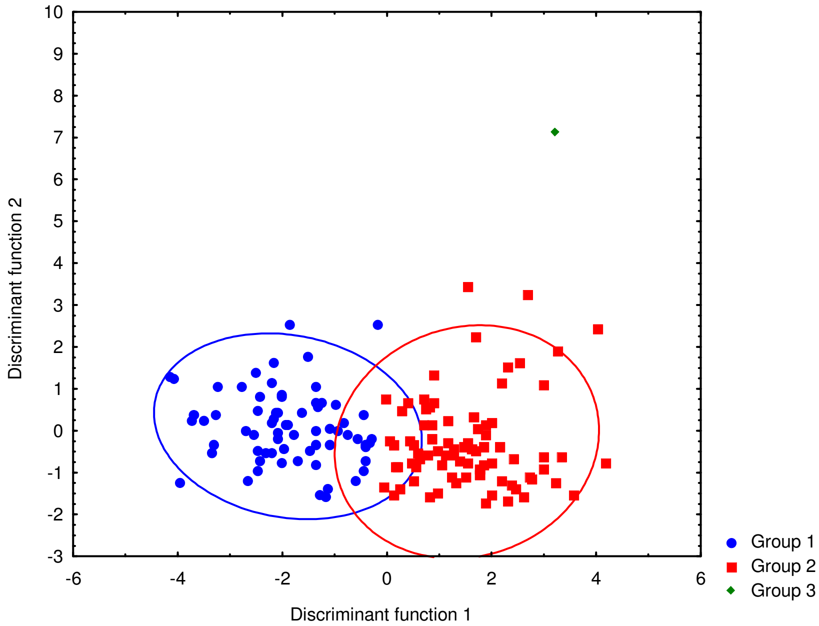

Figure 1. Discriminant function 1 versus discriminant function 2, the ellipses represent a confidence level of 95%.

3. Evaluate potential clusters. In each intersection in the dendrogram it's evaluated the potential groups, in intention to find the natural structure of existing grouping, in case that there is one, determining the number of the group that exist in the database in analysis.

4. To visualize the structure of groups in the map. From the found structure of grouping, a map of exit for visualization is colored, where each color represents a group.

Thus, the following parameterization for the network SOM was adopted for the 147 marajoara ceramic fragments: bi-dimensional map, linear initialization, batch training algorithm, hexagonal topology and Gaussian neighborhood kernel function. For the number of epochs (ne), we adopted the proposal of (VESANTO et al., 2000): ne = 10 q/n epochs for the initial phase and ne = 40 q/n to the convergence phase, where q is the number of map unit, and n is the database length. As learning rate, 0.5 was the value adopted in the initial phase and 0.05 in the convergence phase. The experiments were performed in SOM-Toolbox for MatLab.

The carried through analysis demonstrates the existence of two groups (Figure 1). Although this information is sufficiently interesting, it fits to the archaeologist to define if this fact if of a use of material of different region, or still, if the parts to have provided with same region but differed how much to the use, daily or ceremonial, for the fact to have difference in the used technology, as the all burnt range of the clay.

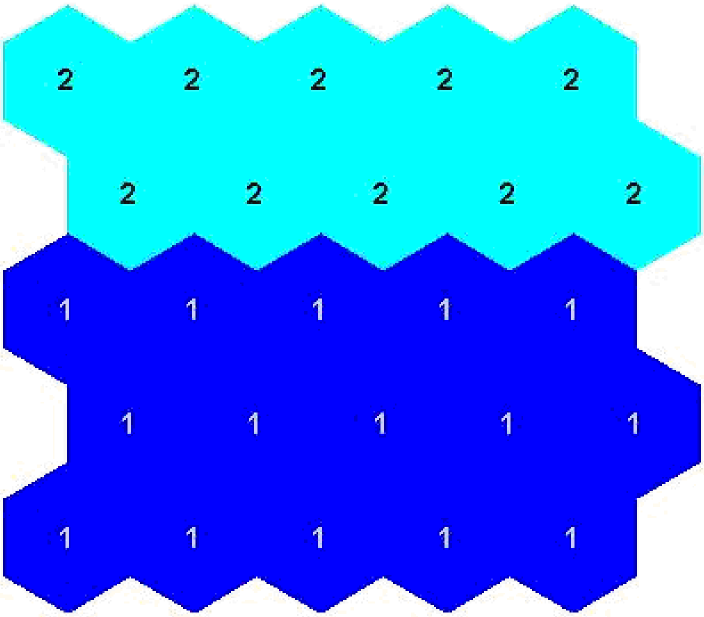

Figure 2. Cluster found for the database marajoara using the SOM analysis.

The Figure 2 shows that the methodology with SOM got a good representation of the groups and confirms the result found in the discriminant analysis, that points the existence of two groups, with some overlapping (Figure 1). Also analyzing the labels of the neurons and the vectors of data represented for each neuron of the map and, contrasting with the labels attributed in it discriminant analysis, evidenced it coincidence of 78,6% in the same ones. However, many of the data with different labels are in the border region enter the groups in the map of similarity for color. Although this information is sufficiently interesting, the archaeologist must define if this result happens because the raw material is provided or forward the difference use, daily or ceremonial, for the fact to have difference in the used technology, as the all burnt range of the clay. A time that does not have many information on the analyzed parts, exactly this information of uncertainty on the classification can be valid to the archaeologist in its work, a time that produced new knowledge that will assist in the cataloging of the parts of the MAE's collection.

4. CONCLUSIONS

In this work, it was verified that the precision of the analytical method for the studied elements was good. The discriminant analysis showed the existence of two groups, indicating that different raw-material were used in the production of the artifacts, and SOM confirmed this knowledge.

These studies through other analytical techniques will be made with the samples having in view complementing the studies here accomplished once SOM is widely used to generate an organization data in a learning process not supervised, with little knowledge about the data.

In archaeometry, the most often exploratory data analysis is performed using discriminant analysis. This work demonstrates that other tools like self-organizing maps can be improve the in-depth analysis of archaeometric data.

ACKNOWLEDGEMENTS

The authors wish to thank Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP, Process 2006/58555-4 and 2006/57343-3 and International Atomic Energy Agency, IAEA, Contract BRA 13048 for financial support.

REFERENCES

1. Gliozzo, E.; Vivacqua, P.; Memmi, I.T. "Integrating archaeology, archaeometry and geology: local production technology and imports at Paola (Cosenza, Southern Italy)", *Journal of Archaeological Science*, **35 (4)**, pp. 1074-1089 (2008).
2. Munita, C.S.; Paiva, R.P.; Alves, M.A.; Oliveira, P.M.S.; Momose, E.F. "Contribution of neutron activation analysis to archaeological studies", *Journal Trace Microprobe Techniques,* **18(3)**, pp. 381-387 (2000).
3. Toyota, R.G. *Caracterização química da cerâmica marajoara*. Dissertação (Mestrado) – Instituto de Pesquisas Energéticas e Nucleares, Universidade de São Paulo, São Paulo (2009).
4. Munita, C.S. "Contribuição da análise por ativação com nêutrons e estudos arqueométricos: estudo de caso", *Canindé,* **6,** pp.159-181 (2005).
5. Kohonen, T. *Self-Organizing Maps*. 3rd Edition. Springer-Verlag. Nova York, USA (2001).
6. Huntsberger, T. and Ajjimarangsee, P. "Parallel Self-organizing Feature Maps for Unsupervised Pattern Recognition". *Journal General Systems*. **16**, pp. 357-372 (1989).
7. Tasdemir, K.; Merényi, E. *Considering Topology in the Clustering of Self-Organizing Maps*. Proceedings of the 5th Workshop On Self-Organizing Maps (WSOM'2005), Paris, France, pp. 439-446 (2005).
8. Vesanto, J.; Alhoniemi, E. "Clustering of the Self-organizing Map". *IEEE Transactions on Neural Networks,* **11(3)**, pp. 586-600 (2000).
9. Wu, S.; Chow, T. W. S. *Self-Organizing-Map Based Clustering Using a Local Clustering Validity Index*. Neural Processing Letters, Kluwer Academic Publishers, Netherlands (2003).
10. Tanevska, V.; Kuzmanovski, I.; Grupče, O. "Provenance Determination of Vinica Terra Cotta Icons Using Self-Organising Maps", *Annali di Chimica by Società Chimica Italiana*, **97**, pp. 541-552 (2007).
11. Munita, C.S.; Paiva, R.P.; Alves, M.A.; Oliveira, P.M.S.; Momose, E.F. "Provenance study of archaeological ceramic", *Journal of Trace and Microprobe Techniques*, **21(4)**, pp. 697-706 (2003).
12. Munita, C.S.; Silva, M.A; Silva, S.A.; Oliveira, P.M.S. "Archaeometric study of clay deposits from the indigenous land of the Asurini do Xingu", *Instrumentation Science and Technology*, **33**, pp.169-181 (2005).
13. Attas, M.; Fossey, J.M.; Yaffe, L. "Corrections for drill-bit contamination in sampling ancient pottery for neutron activation analysis", *Archaeometry*, **26(1)**, pp. 104-107 (1984).
14. Perlman, I.; Asaro, F. *Science and Archaeology*. London: RH Brill (1971).

15. Wilks, S.S. “Multivariate statistical outliers”, *Sankhya*, **25**, pp. 407-426 (1963).
16. Boscarioli, C.; Del Moral Hernandez, E.; Silva, L. A. “Using self organizing maps and nearest-neighbor to data clustering”. *In: International Conference on Artificial Intelligence & International Conference on Machine Learning; Models, Technologies and Applications*, Las Vegas, USA. Proceedings: ICAI 2008, CSREA Press, pp. 463-467 (2008).