Neural Networks (SOM) Applied to INAA Data of Chemical Elements in Archaeological Ceramics from Central Amazon

R. Hazenfratz[1], C. S. Munita[1] and E. G. Neves[2]

(1) *Nuclear and Energy Research Institute, Research Reactor Center, Av. Lineu Prestes 2242 - Cidade Universitária, Zipcode: 05508-000 - São Paulo, SP - Brasil*

(2) *University of São Paulo – Museum of Archaeology and Ethnology, Av. Prof. Almeida Prado 1466, Cidade Universitária, Zipcode: 05508-900 - São Paulo, SP - Brasil*

ABSTRACT

In this study, the Self-Organizing Maps were applied to concentration data of chemical elements measured in archaeological ceramics from Central Amazon using instrumental neutron activation analysis (INAA). The main objective was testing the chemical patterns previously identified using cluster and principal component analysis, forming groups of ceramics according the multivariate chemical composition. It was verified by statistical tests that the chemical elemental data was not normally distributed and did not present homogeneity of covariance matrices for different groups, as requested by principal component analysis and other multivariate techniques. The maps obtained were consistent with the patterns identified by cluster and principal component analysis, forming two chemical groups of pottery shards for each archaeological site tested. Finally, it was verified the potential of SOMs for testing if failures in underlying hypotheses of traditional multivariate techniques might be critically influencing the results and subsequent archaeological interpretation of archaeometric data.

INTRODUCTION

Artificial neural networks, AAN, represent an alternative to traditional multivariate techniques, such as principal component and discriminant analysis, which rely on hypotheses regarding the normal distribution of the data and homoscedasticity. They also may be a powerful tool for multivariate modeling of systems that do not present linear correlation between variables, as well as to visualize high-dimensional data in bi or trivariate structures. One special kind of neural network of interest in archaeometric studies

is the Self-Organizing Map (SOM). SOMs can be distinguished from other neural networks for preserving the topological features of the original multivariate space.

Two main applications of ANN may be identified: function approximation by retropropagation algorithms, and grouping or classification of input vectors (Merdun, 2011).

The main applications of artificial intelligence to archaeometric studies regards the comparison of geochemical data classification by traditional algorithms, such as hierarchic cluster analysis (CA) and principal component analysis (PCA), and neural networks based methods. One main advantage is that the latter do not depend on any specific data distribution (Bell & Croson, 1998). A recurrent method for such a purpose is the application of self-organizing maps (SOMs) (Toyota et al., 2009), developed by Teuvo Kohonen. Those maps are based in AAN in a frequently non-supervised leaning strategy, and do not need a previous set of test samples of an already known structure (Lopes-Molinero *et al*., 2000). A graphical representation of a SOM is presented in Fig. 1. The method is partially based in the way sensorial information are processed in separate parts of the human brain cortex.
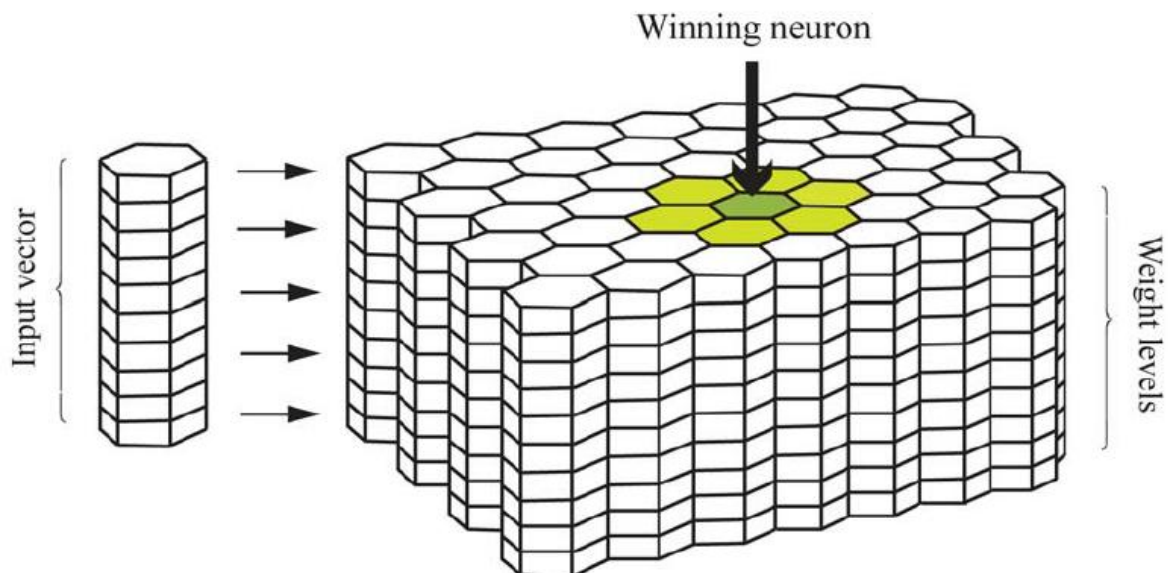
Figure 1 – Graphical representation of a self-organizing map (SOM) showing the neurons represented by vertical columns. The structure surface corresponds to the map which is visualized after the application of the method. Source: Tanevska et al. (2007).

As in the case of principal components and discriminant functions, those maps are a useful tool for graphical visualization of high dimensional data in bi- or tridimensional structures. The maps as formed by nodes, or neurons, and every node has a weigh vector with the same dimension as the data, or inputs. The neuronal geometry may be hexagonal or rectangular, which influences the number of neighbors. SOMs can be distinguished from other ANNs by the preservation of the topological properties of the original space, by using neighborhood functions.

In the SOM method, an input vector is allocated in the neuron unit that has the most similar weigh vector, or the nearest, based in an adequate metric to measure the distances between vectors. The general idea is that the weigh vectors representative of each neuron are spatially correlated, such as near representative vectors in the grade are more similar between them when compared to distant vectors.

The SOM training consists in a competitive unsupervised learning, which assign input vectors to the most similar neuron, represented by its weigh vector. In this way, this method aids in identifying grouping patterns in the data set. As a new input vector is presented to the net, its Euclidean distance to every weigh vector associated to a neuron is calculated. The "winning" neuron, called best matching unit (BMU), and neighboring vectors have their weigh adjusted due to the input vector. The magnitude of the adjustment decreases over time and with the distance to the BMU, according the following updating equation (adapted from Lopez-Molinero et al., 2000)

$$W_v(t+1) = W_v(t) + \theta(v,t,d)\alpha(t)[D(i) - W_v(t)] \tag{1}$$

where t is the iteration or epoch of the algorithm, $D(i)$ is the input vector (i is the sample size of the data set for training, varying from 1 to n), $\theta(v,t,d)$ is the neighborhood function (frequently Gaussian or Triangular) depending on the distance d to the BMU, and $\alpha(t)$ is the monotonically decreasing learning coefficient.

The learning process initiate with large neighborhoods to the point where its definition turns more restrict, and the weigh vector of each neuron converge to local estimates in the map. An epoch of training is complete when all input vectors are presented once to the net for correction of neuronal weigh vectors. At the end of the procedure, the neurons, or output nodes, may be associated to the groups present in the multidimensional space of input by graphical visualization, with color scales representing the Euclidean distances between the neurons, stored in a U matrix.

The batch version of SOM algorithm, where all the input vectors are presented simultaneously to the net in a unique update is more recent. In the literature, there are still few works dealing with the application of self-organizing maps to archaeometric data (Lopez-Molinero et al., 2000; Toyota, 2009).

Archaeological background

This work comprises the analysis of pottery shards data from two large and important archaeological sites in Central Amazon, namely Lago Grande and Osvaldo. The main archaeological objective is the verification of potential cultural and commercial exchange networks in the region. Once confirmed, it would have impact on the traditional theories about the occupation of pre-colonial Amazon, based on environmental determinism, which conclude that large sedentary settlements and hierarchical and complex human organizations could not emerge in the tropical rain forest environment (Lima, 2008; Hazenfratz et al., 2016).

Previous archaeometric analysis of data from Lago Grande and Osvaldo led to the determination of two chemical groups of pottery for each archaeological site. They presented pairwise superposition, which were further explored seeking for archaeological correlations regarding territorial integration, commercial exchange and exogamic marriage between the inhabitants of the region (Hazenfratz-Marks et al., 2012). The statistical significance of the chemical groups was teste indirectly by comparison with variation coefficients found in archaeometric studies of pottery around the world (Harbottle, 1982; Bishop, 2003), by discriminat analysis and by multivariate analysis of variance.

Table 1 present the average concentrations of nine chemical elements used in previous studies for each chemical group defined for the two archaeological sites. Fig. 2 presents a refined projection of the multivariate data in the two first principal components.

RESULTS AND DISCUSSION

In this work, a study with self-organizing maps was developed for comparison with multivariate patterns identified by CA and PCA of geochemical data of pottery shards from two large archaeological sites in central Amazon. The results may be found elsewhere (Hazenfratz-Marks et al., 2012; Hazenfratz et al., 2016). The total data set comprises the concentrations of 9 chemical elements (Sc, Cr, Fe, La, Ce, Eu, Yb, Lu, Th) in 141 samples.

Table 1 – Mean concentrations of nine chemical elements in the chemical groups of pottery defined for Lago Grande and Osvaldo archaeological sites. Concentrations are in $\mu g.g^{-1}$.

Element	Lago Grande		Osvaldo	
	A (n=61)	B (n=28)	A (n=37)	B (n=15)
Sc	14.8 ± 1.9	18.6 ± 2.1	13.8 ± 2.0	18.7 ± 2.3
Cr	64.3 ± 8.3	79.1 ± 10.2	59.0 ± 7.4	75.3 ± 9.6
Fe	36543 ± 7991	44197 ± 6505	32432 ± 5723	39165 ± 3728
La	38 ± 5	50 ± 5	38 ± 4	47 ± 7
Ce	75 ± 11	104 ± 17	71 ± 12	107 ± 24
Eu	1.1 ± 0.2	1.7 ± 0.2	1.1 ± 0.2	1.5 ± 0.2
Yb	2.7 ± 0.4	3.5 ± 0.4	2.9 ± 0.4	3.3 ± 0.4
Lu	0.45 ± 0.07	0.56 ± 0.07	0.47 ± 0.07	0.55 ± 0.06
Th	13.7 ± 1.5	16.6 ± 2.0	13.8 ± 2.2	17.2 ± 2.4

The parameter *n* represents the number of samples in each group.
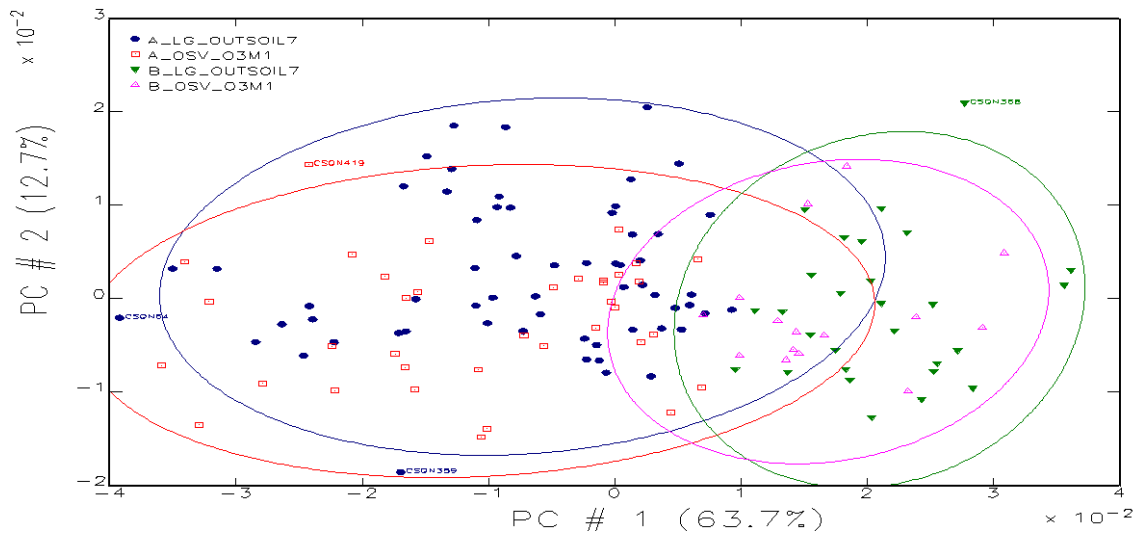
Figure 2 – PCA scores of elemental concentration data from Lago Grande and Osvaldo archaeological pottery combined after classification refinement. Ellipses represent the 95% confidence region.

The computer packages Kohonen and CP-ANN (implemented in Matlab 7.6.0 R2008a with the function *newsom*) were employed to calculate the SOMs (Ballabio *et al*., 2009; Ballabio & Vasighi, 2012). It was adopted the non-supervised algorithm so that the previous classifications of samples by CA and PCA might not interfere in the learning procedure. After the convergence of maps comparisons between the classifications of samples were made. The parameters selected for the algorithm were (Lopez-Molinero et al., 2000; Tanevska et al., 2007; Toyota, 2009; Ballabio & Vasighi, 2012):

- Model: non-supervised Kohonen map
- Net geometry: hexagonal
- Number of neurons: 7x7
- Number of epochs: 500
- Neighborhood function: Gaussian
- Training algorithm: batch
- Initialization of neuronal weighs: linear (from data eigenvectors)
- Data transformation: base-10 logarithm
- Initial learning rate: 0.5
- Final learning rate: 0.01

In Fig. 3 and 4 the SOMs for Lago Grande and Osvaldo are presented, respectively. Fig. 3 shows that the SOM converged to a configuration where the samples from different chemical groups were allocated in different regions of the map. The white and hachured neurons represent approximately the different group domains. The group g1 has a higher number of neurons associated to it, and it may be correlate of a higher geochemical variation in the samples. In fact, by comparison of Fig. 2 and 3, regarding the refined PCA results, it is possible to observe that the 95% confidence ellipse associated to the group A (called g1 here) is larger than the ellipse associated to group B (called g2 here), which may be interpreted as another indication of the same higher geochemical variability. The same interpretation may be claimed for Fig. 3, regarding the SOM for Osvaldo archaeological site.

The higher chemical variability of group g1 in Lago Grande was interpreted as possible superposition of different clay recipes, and it was hypothesized that it may represent pottery shards from other archaeological sites in the region that have might participated in larger exchange networks than previously thought initially, and not only between Lago Grande and Osvaldo (Hazenfratz-Marks, 2014).
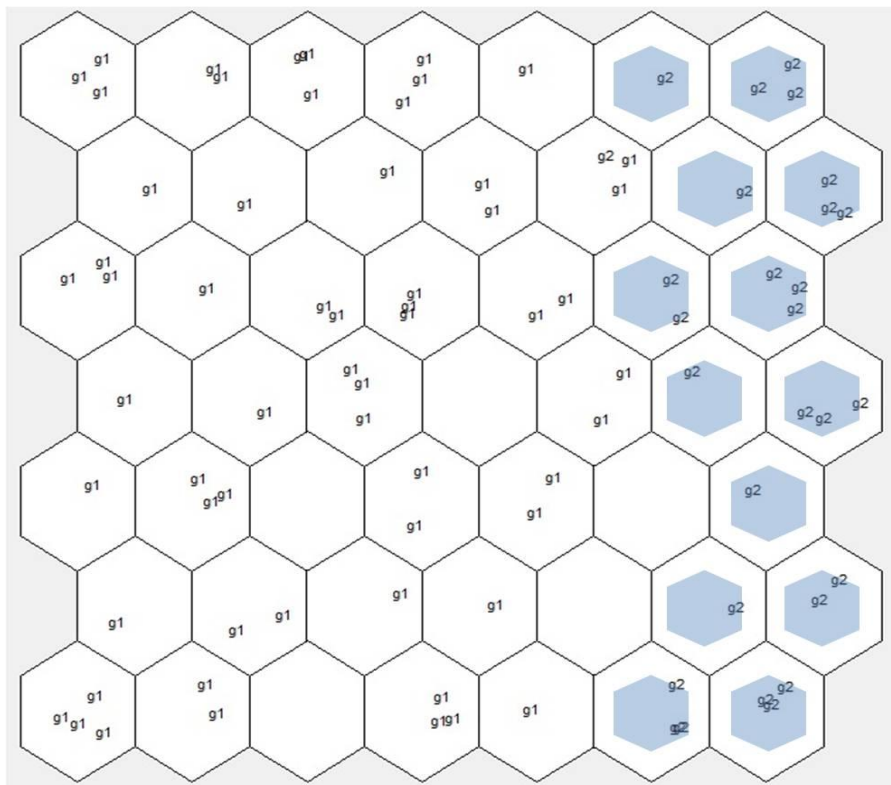
Figure 3 – Kohonen Map for the chemical groups of Lago Grande archaeological site. The symbols g1 and g2 represent different groups identified by CA and PCA analysis.

Fig. 4 refers to the application of SOM algorithm for the samples of pottery shards from Lago Grande and Osvaldo in a combined fashion. It was observed that the map converged to a configuration where the pottery shards of group g1 of Lago Grande and group g3 of Osvaldo presented superposition and were allocated in a common region in the map (white hexagons). This region is distinct from the region where the pottery shards from groups g2 of Lago Grande and g4 from Osvaldo were allocated, presenting the same superposition observed for groups g1 and g3 (blue hexagons). The three neurons highlighted in orange represent a transition region, with different mixing of samples. However, they represent the smallest portion of the map.

The results of the samples from Lago Grande and Osvaldo combined are in agreement with the combined CA and PCA analysis which identified at least two chemical groups of pottery shards for each archaeological site, with pairwise superposition, as it can be observed in Fig. 2.
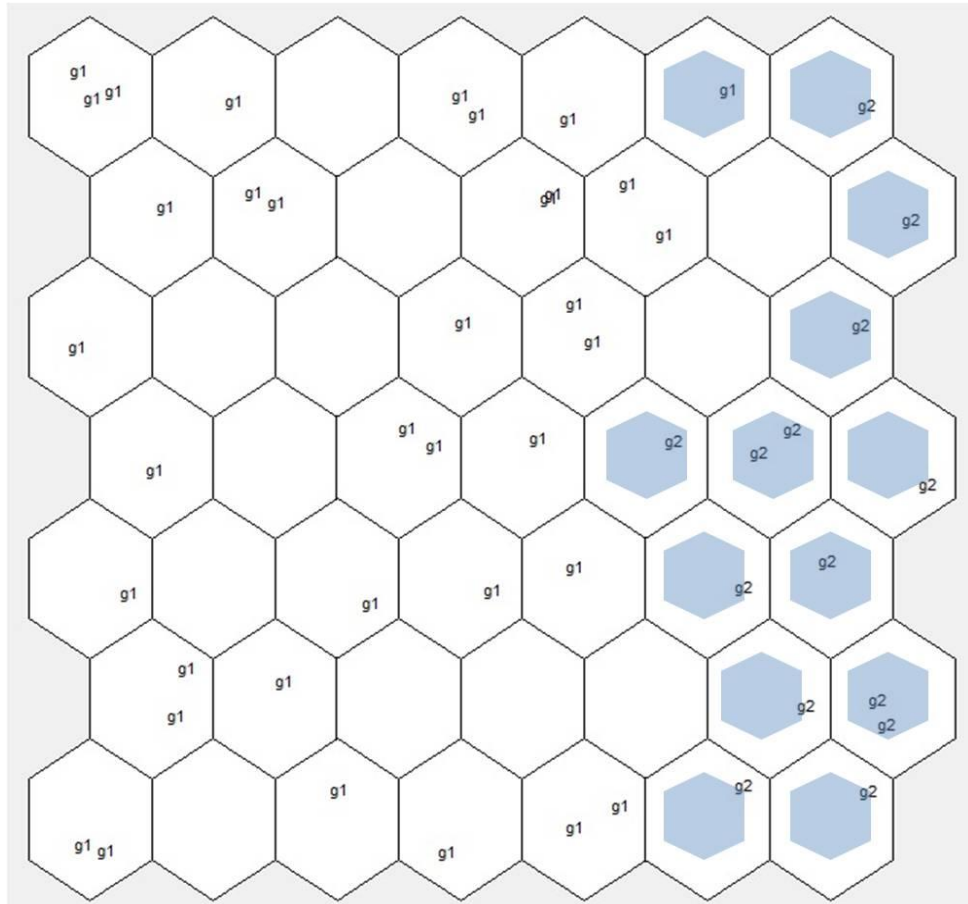
Figure 4 – Kohonen Map for the chemical groups of Osvaldo archaeological site. The symbols g1 and g2 represent different groups identified by CA and PCA analysis.

CONCLUSION

The Kohonen maps applied to the geochemical data of pottery shards from Lago Grande and Osvaldo, two large and important sites in central Amazon, agreed with the results of previous multivariate statistical methods using cluster and principal component analysis. It indicates that the archaeological community can rely on the chemical groups of ceramic artifacts defined for those archaeological sites so far, and can use them for archaeological interpretation and basis for planning further research regarding the pre-colonial Amazonian occupation.

It was also verified the potential of the self-organizing maps for archaeometric data analysis and comparison with multivariate statistical methods which rely upon hypotheses regarding the normal distribution of data and the homogeneity of the covariance matrices. Such a comparison could indicate if the deviation of underlying hypotheses of more

8

traditional multivariate analysis would affect the results in pattern recognition, used frequently to analyse archaeometric data sets with the objective of identifying chemical groups of artifacts and provenance studies.

ACKNOWLEDGEMENTS

REFERENCES

BALLABIO, D.; CONSONNI, V.; TODESCHINI, R. The Kohonen and CP-ANN toolbox: a collection of MATLAB modules for Self Organizing Maps and Counterpropagation Artificial Neural Networks. *Chemometrics and Intelligent Laboratory Systems*, 98, pp. 115-122, 2009.

BALLABIO, D.; VASIGHI, M. A MATLAB toolbox for Self Organizing Maps and supervised neural network learning strategies. *Chemometrics and Intelligent Laboratory Systems*, 118, pp. 24-32, 2012.

BELL, S.; CROSON, C. Artificial Neural Networks as a Tool for Archaeological Data Analysis, *Archaeometry*, 40(1), pp. 139-151, 1998.

BISHOP, R. L. Instrumental Neutron Activation Analysis of Archaeological Ceramics: Progress and Challenges. In: *IAEA-TRS416 - Nuclear Analytical Techniques in Archaeological Investigations*, Vienna, cap. 2, pp. 35-44, 2003.

HARBOTTLE, G. Provenance studies using neutron activation analysis: the role of standardization. In: Olin, J. S., and Franklin, A. D. (eds.). *Archaeological Ceramics*, pp. 67–77, Washington, D.C., Smithsonian Institution Press, 1982.

HAZENFRATZ-MARKS, R. MONGELÓ, G.; MUNITA, C.S.; NEVES, E.G. Comparison of INAA elemental composition data between Lago Grande and Osvaldo archaeological sites in the central Amazon: a first perspective. *Journal of Radioanalytical and Nuclear Chemistry*.291, pp. 43-48, 2012.

HAZENFRATZ-MARKS. *Nêutrons, Radiação e Arqueologia: Estudo de Caso Multianalítico de Cerâmicas da Tradição Borda Incisa na Amazônia Central*. Tese de Doutorado em Tecnologia Nuclear – Aplicações, Instituto de Pesquisas Energéticas e Nucleares, Universidade de São Paulo, 2014.

HAZENFRATZ, R.; MUNITA, C. S.; GLASCOCK, M. D.; NEVES, E. G. Study of Exchange Networks between two Amazon Archaeological sites by INAA. *Journal of Radioanalytical and Nuclear Chemistry*. 2016, accepted.

LIMA, H. *História das Caretas: A Tradição Borda Incisa na Amazônia Central*. Tese de Doutorado em Arqueologia, Museu de Arqueologia e Etnologia, Universidade de São Paulo, 2008.

LOPEZ-MOLINERO, A.; CASTRO, A.; PINO, J.; PEREZ-ARANTEGUI, J.; CASTILLO, J. R. Classification of ancient Roman glazed ceramics using the neural network of Self-Organizing Maps. *Fresenius´ Journal of Analytical Chemistry*, 367, pp. 586-589, 2000.

MERDUN, H. Self-organizing map artificial neural network application in multidimensional soil data analysis. *Neural Computing and Applications*, 20, pp. 1295-1303, 2011. Doi: 10.1007/s00521-010-0425-1.

TANEVSKA, V.; KUZMANOVSKI, I.; GRUPČE, O. Provenance determination of Vinica terra cotta icons using self-organizing maps. *Annali di Chimica*, 97, pp. 541-552, 2007.

TOYOTA, R. G. *Caracterização Química da Cerâmica Marajoara*. Dissertação de Mestrado em Tecnologia Nuclear – Aplicações, Universidade de São Paulo, 2009.