BJRS

BRAZILIAN JOURNAL
OF
RADIATION SCIENCES
09-01A (2021) 01-15

The effect of data standardization in cluster analysis

Nogueira[ab] A.L., Munita[b] C.S.

[a] *Instituto Federal de Sergipe (IFS), 49400-000, Lagarto, SE,*

andreln27@yahoo.com

[b] *Instituto de Pesquisas Energéticas e Nucleares (IPEN/CNEN-SP), 05508-000, São Paulo, SP,*

camunita@ipen.br

ABSTRACT

The application of multivariate techniques to experimental results requires a responsibility on behalf of the researcher to understand, evaluate and interpret their results, especially the ones that are more complex. The objective of this article is to evaluate the impact of three standardization techniques on the formation of clusters by means of the Kohonen neural network were studied. The standardization techniques studied were logarithm (log), generalized-log and improved minimum-maximum. The studies were performed using two different databases consisting of 298, named B1, and 146 samples, named B2. The B1 dataset is formed by samples that form two cluster very close. However, the B2 dataset form three diferent and separated cluster. The mass fractions of As, Ce, Cr, Cs, Eu, Fe, Hf, K, La, Lu, Na, Nd, Sc, Sm, Tb, Th, U, and Yb of each sample were determined by instrumental neutron activation analysis, INAA. Three validation indices : Jaccard, Fowlkes-Mallows and Rand were performed on the dataset. The results suggest that when the cluster are close, the improved minimum-maximum satandardization is better than the logarithm and generalized-log. However, when the cluster are separated, the logarithm and generalized-log are better than the improved minimum-maximum technique.

Keywords: **cluster analysis, INAA, neural network, standardization.**

ISSN: 2319-0612
Accepted: 2021-04-04

1. INTRODUCTION

The future advancement of physicochemical techniques means that the quantity of results generated will increase significantly. For results analysis, it is necessary to use more sophisticated methods, such as multivariate techniques. In general, multivariate statistical methods allow one to evaluate a set of samples in terms of the correlations between variables. These techniques consider that each sample can be represented as a point in multidimensional space, where each dimension of hyperspace corresponds to an axis determined by the physicochemical composition of the samples. One of the ways to verify the existence of similar behaviors between the samples in relation to the different variables is by carrying out a clustering analysis. A problem that arises during cluster analysis involves the decision to standardize the samples before calculating the distance measurements, while the existence of several standardization techniques complicates this decision further. This article aims to study the effect of three standardization techniques on cluster analysis, they are: log, log-generalized [1], and improved min-max [2]. After applying data standardization, they are submitted to a SOM (Self Organizing Map) neural network which aims to gather samples to create groups, so that there is internal homogeneity in the groups and external heterogeneity among them [3]. The SOM network is a self-organizing map of unsupervised training: the central idea of the SOM network is competitive learning, since when presenting the sample to the network, the neurons compete with each other and the winner has their weights adjusted to better answer to network stimuli. In addition, there is a process of cooperation between neurons and their neighbors, who also receive adjustments. The characteristics contained in the sample will stimulate a special region of the network associated with a particular group.

The study was performed using two databases, one of 298 samples and the other of 146 samples, in which the mass fractions of Na, K, La Yb, Lu, U, Sc, Cr, Fe, Cs, Eu, Tb, Hf, Tb and As, Ce, Cr, Eu, Fe, Hf, La, Na, Nd, Sc, Sm, Th, U repectively, were determined by instrumental neutron activation analysis, INAA. To evaluate the standardization techniques, three validation indices Jarccard [4], Rand [5] and Fowlkes-Mallows [6] were used.

2. MATERIALS AND METHODS

2.1. Neural network

Neural networks are made up of basic processing units called neurons. The figure below shows a neuron of an artificial neural network [7].

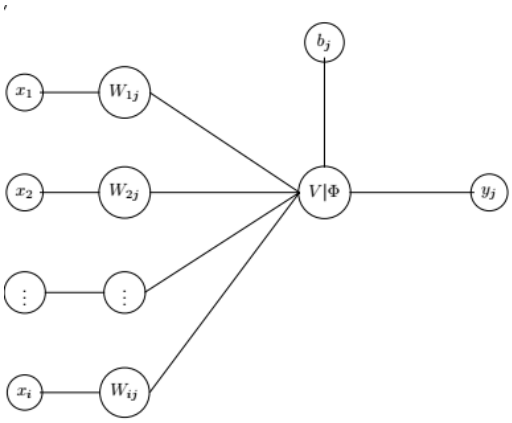

Figure 1: *Artificial neuron*

The artificial neuron, similar to the natural neuron, receives input signals and returns a single output signal, which may be the output of the network or the input signal to one or more neurons in the back layer.

The inputs of a neural network x_1, x_2, ..., x_i are multiplied by the corresponding synaptic weights W_{1j}, W_{2j}, ..., W_{ij} generating the following weighted sum:

$$V = \sum_{k=1}^{i} W_{kj} x_k \tag{1}$$

This function is called the activation function. The weighted sum is presented to a transfer function whose purpose is to avoid the progressive addition of output values [7].

The artificial neural model can also include an input bias (b_j) in order to increase the degree of freedom of the activation function [7].

An artificial neural network is a combination of neurons, their connections, and the algorithm used in training. The neural network has two stages of processing: learning and network application.

In learning, the adjustment of weights occurs in response to data presented to the network. In the application of the network, one has the way in which the network responds to the data without there being changes in the weights.

The SOM network is a self-organizing map model of unsupervised training [7, 8]. In this structure, the neurons are arranged in a normally two-dimensional grid, which can be square, rectangular, triangular, and so on. What characterizes the SOM network is the formation of a topological map of input data patterns in which the locations of the neurons indicate the characteristics of the input data.

The central idea of this model is competitive learning, because when presenting an input sample to the network, the neurons compete with each other and the winner has their weights adjusted to better respond to the stimulus presented to the network. In addition, there is a process of cooperation between neurons and their neighbors, who also receive adjustments. The characteristics contained in the input sample will stimulate some special region of the network and the sample is assigned to the corresponding group.

2.2. Standardization techniques

In many cluster analysis applications, raw data, or actual measurements, are not used directly unless a probabilistic model for pattern generation is available [9]. Thus, there is a need to prepare the data for cluster analysis through a transformation aimed at standardizing the data. In this work we studied three standardization techniques: log, min-max improved [2] and spectral log-generalized mean [1].

In the improved min-max standardization [2] a set of R_k is constructed for each column that is composed of values that occur more than once. The mean and standard deviation of R_k are summed to obtain R_{KA}. At the end, the improved min-max normalization is applied using the expression:

$$F(x) = \begin{cases} \dfrac{x_k - \min(x_k)}{2(R_{KA} - \min(x_k))}, x_k \leq R_{KA} \\ 0.5 + \dfrac{x_k - R_{KA}}{\max(x_k) - R_{KA}}, x_k > R_{KA} \end{cases} \quad (2)$$

where $R_{KA} = R_{Kavg}+R_{Kstd}$, $R_{Kavg} = mean(R_{KA})$, $R_{Kstd} = std(R_{KA})$ (standard deviation).

Log-generalized average spectral standardization is based on the q-logarithm function which is a generalization of logarithmic function [1], often called generalized logarithmic function, and is defined as:

$$\bar{x}^k = \exp_q\left(\frac{\log_q(x^k) - \frac{1}{N}\sum_{i=1}^{N-1}\log_q x^i}{1+q\frac{1}{N}\sum_{i=1}^{N-1}\log_q x^i}\right) \tag{3}$$

where

$$\log_q(x) = \begin{cases}\frac{x^q - 1}{q}, q \neq 0\\ \log(x), q = 0\end{cases} \tag{4}$$

$$\exp_q(x) = \begin{cases}(1+qx)^{\frac{1}{q}}, q \neq 0\\ \exp(x), q = 0\end{cases} \tag{5}$$

2.3. Validation indeces

The indices described below evaluate the quality of the clustering algorithm by comparing the results obtained from the SOM neural network with predefined information. The existence of two partitions is assumed, one obtained by the SOM neural network and the other with additional information about the base [10].

Let A and B be two partitions. The Jaccard index is a well known measure of similarity between groups described by the presence or absence of samples, used in cluster analysis. It counts the number of pairs of samples belonging to the same group in partitions A and B, and the number of pairs of samples that belong to the same group on at least one of the partitions. The Jaccard index [4] or coefficient of similarity is given by:

$$R = \frac{a}{a+b+c} \tag{6}$$

where a is the number of pairs of samples belonging to the same cluster, in A and B; b is the number of pairs of samples belonging to different groupings in A, but even group in B; c is the number of samples belonging to the same group in A, but different groups in B.

The Rand index is a statistical measure of the proportion of pairs of samples belonging to the same or different groups in both partitions and is defined by [5]:

$$R = \frac{a + d}{a + b + c + d} \tag{7}$$

where constants a, b, and c are the same as the previous index, and d is the number of samples belonging to different groups in A and B.

The Fowlkes-Mallows index [6] is a geometric mean of the proportion of pairs of samples belonging to the same group in both partitions. Let A and B be two partitions, with the same number of samples. Let m = [m_{ij}], i, j = 1, ..., k, where m_{ij} is the number of samples in common with the ith cluster of A and the jth cluster of B. The similarity measure proposed by [6]:

$$B_k = \frac{T_k}{\sqrt{P_k Q_k}} \tag{8}$$

where

$$P_k = \sum_{i=1}^{k} m_i^2 - n, Q_k = \sum_{j=1}^{k} m_j^2 - n\,, T_k = \sum_{i=1}^{k}\sum_{j=1}^{k} m_{ij}^2 - n, m_i = \sum_{j=1}^{k} m_{ij},$$

$$m_j = \sum_{i=1}^{k} m_{ij}, n = \sum_{i=1}^{k}\sum_{j=1}^{k} m_{ij}.$$

2.4. Datasets

In the tests, two databases were used, one consisting of 298 ceramic samples from the island of Marajó. The island of Marajó has 40.552 km^2 and is part of an archipelago that lies at the mouth of the river Amazon. The Amazon region has a humid winter, where pre-Columbian inhabitants had to deal with interspersed flooding and drought. About 1,500 years before the conquest of the continent came one of the most intriguing indigenous societies in America. Its culture was

characterized by the construction of enormous mounds of earth of up to 12 of height and 3 ha of area, by the production of elaborate ceramic vessels and other objects used in ceremonies [11]. Marajoaras ceramics have been studied intensively since the 19th century, especially in relation to the function of the vessel, production process and style.

During the second half of the twentieth century scholars made a study dedicated to society itself, focusing on social organization, subsistence pattern and material culture [12]. Radiocarbon dates place that the period of greatest growth and expansion of the Marajoara culture was between the 5th and 14th centuries.

Marajoara pottery belongs to the polychrome tradition, characterized by highly complex ceremonial items in shape and decoration. Decorative techniques involve painting, incision, exision and scraping. The ceramics are seasoned with crushed ash from the bark of a tree known as caraiapé. This material was used in the Amazon Basin at the end of the first millennium and is associated with other ceramics of the Polychrome tradition [13].

The other database with 146 samples, consists of three sites that are located superficially in the intermediate part of a hill with a water course in its inferior part [14]. The ceramics located in these sites are associated with food preparation, funeral urns and decorative use.

The Água Limpa site is located at the confluence of three small farms in the city of Monte Alto, in the north of the state of São Paulo 21º 15'40''S-48º 29'47"W. The site was divided into two excavation zones. In zone 1, an area of primary burials of young people and adults extended and semi-inflected was found. Ten other tombs were exhumed in addition to the exhumation of a secondary burial of an adult from an urn with lid [15]. There was one hearth on the spot, dated 1,524 ±50 year B.P. All other hearths were external [16]. In zone 2, the Village is formed by eight dark spots and several hearths, most of them inside the houses. Only a secondary burial of a child was found and exhumed.

The ceramics collected from this site are of two types: painted and plain. The painting is in red and white, the few painted and whole fragments that have been collected have no form. Grain selection is good with predominance of thin and medium grains.

The Prado site is located on the Engenho Novo farm, in the city of Perdizes, State of Minas Gerais, Brazil 19º14'25"LS-47º16'00"LW. It consists of seven dark spots (housing structure) and three hearths. The archaeological remains are collected are of two types: ceramic and lithic

(polished or not). The ceramic containers partially reconstructed in the field or in the laboratory, and the containers collected are smooth without plastic decoration, with predominance of medium and large granularity, with poor selection of grains. They Were dated 850±45 year B.P., and were produced for utility and funeral purposes [17-20].

The site of Resende is located in the farm Paiolão, in Piedade, in the Paranaíba Valley, 7km from the city of Centralina, State of Minas Gerais, Brazil 18º33'LS-49º13'LW. Archaeological studies indicate two occupations: the most recent is represented by occupation ceramics, and the fragments studied were dated 1,190 ± 60 B.P. It starts on the surface and goes up to 35/40 cm deep. Archaeological studies have shown that the population lived in oval huts forming villages and used fire for lighting, cooking and as a source of heat. The pottery produced was simple, utilitarian and funerary. The oldest occupation is pre-ceramic to 90/130 cm deep and was dated to 7,300 ± 80 B. P. They represent the first and oldest inhabitants of the state of Minas Gerais [15-22].

2.5. Sample Preparation and Description of the Method

The ceramic power samples were obtained by cleanning the outer surface and drilling, using a tungsten carbide rotary file attached to the end of a variable speed drill with a flexible shaft. After that, these materials were dried in an oven at 105 ^{o}C for 24 h, and stored in a desiccator.

Constituent Elements in Coal Fly Ash, NIST-SRM-1633b, were used as standards, and IAEA-Soil-7, Trace Elements in Soil, were used to check samples in every analysis. These materials were dried in oven at 105 ^{o}C for 4 h [23].

About 100 mg of different ceramic samples, NIST-SRM-1633b, and IAEA-Soil-7 were weighed in polyethylene bags and wrapped in aluminium foil. Groups of 8 samples, and one of each reference material were packed and irradiated inthe research reactor pool, IEA-R1, from the IPEN-CNEN/SP, at a thermal neutron flux of about 5×10^{12} cm$^{-2} \times$ s^{-1} for 8 h.

Two measurements series were carried out using Ge (hyperpure) detector, model GX 1925 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. Gamma ray spectra analysis and the concentrations were carried out using the Genie-2000 Neutron Activation Analysis Processing Procedure from Canberra. A detailed description of the method, the samples, the standard sample preparation, and the archaeological sites were published elsewhere [23-25].

The means and standard deviations of the B1 and B2 datasets are presented in Tables 1 and 2.

Table 1: Range, mean and standard deviation for ceramic samples from B1 dataset, in mgg^{-1}, unless otherwise indicated, n = 298 [26].

Element	Range	Mean ± sd
Na(%)	0.04 – 1,20	0.45±0.14
K(%)	0.01 – 4.20	1.96±0.70
La	4.01 – 73.30	54.68±7.70
Yb	0.50 – 5.00	3.83±0.47
Lu	0.25 – 0.77	0.59±0.06
U	2.60 – 6.90	4.06±0.65
Sc	12.30 – 88.13	18.74±4.30
Cr	57.80 – 933.34	94.50±49.96
Fe(%)	1.76 – 12.21	5.42±1.15
Cs	4.36 – 91.40	9.17±4.97
Eu	0.90 – 2.57	1.80±0.26
Tb	0.01 – 2.10	1.07±0.29
Hf	3.60 – 13.75	7.34±1.53
Th	14.10 – 23.10	18.17±1.40

Table 2: Means and standard deviations for ceramic samples from B2 dataset, in mg g^{-1}, n=146 [24, 25].

Elements	Água Limpa (n=81)	Prado (n=34)	Rezende (n=31)
As	2.23±1.01	1.57±0.38	1.86±0.49
Ce	122.68±20.93	115.11±9.92	85.21±34.71
Cr	160.73±30.48	13820±20.61	218.34±27.97
Eu	2.50±0.38	1.40±0.16	3.20±0.45
Fe	33461.73±7753.766	28535.29±5639.03	10821.61±2375.87
Hf	8.36±1.02	8.87±0.69	11.49±0.74
La	71.50±10.73	33.23±3.97	37.72±6.57
Na	1948.22±576.09	565.08±107.71	158.70±40.43
Nd	58.50±10.72	38.23±7.59	52.45±9.06
Sc	15.61±2.34	29.66±2.02	43.99±3.06

Sm	9.66±1.40	7.45±0.63	10.48±1.61
Th	12.78±1.91	17.47±0.96	6.40±0.77
U	1.37±0.29	4.24±0.87	1.37±0.23

3. **RESULTS AND DISCUSSION**

The tests were performed using two databases, one containing 298 samples, B1 (corresponds to Marajoara ceramics) , and the other with 146 samples, B2 (corresponds to the ceramics of the three sites). The elements determined for the datasets B1 were Na, K, La Yb, Lu, U, Sc, Cr, Fe, Cs, Eu, Tb, Hf, Tb. For the datasets B2, the elements determined were As, Ce, Cr, Eu, Fe, Hf, La, Na, Nd, Sc, Sm, Th, and U. The mass fraction of the each samples were obtained by INAA [23, 24, 26]. Tables 1 and 2 show the mass fractions for 298 and 146 samples, respectively.

Figures 2 and 3 shows the plots of the principal component analysis, PCA, PCA1 vs PCA2 of datasets B1 and B2, respectively. The PCA is a technique that transform linearly one set of *p* variables observed in a smaller set of *k* non-correlates variables, and that explains a substantial portion of the data covariance. In PCA, a transformation of the dataset, based on eigenvector methods, is performed to determine the direction and magnitude of maximum variance. Then PCA begins with the *p* correlated variables, and the procedure transforms these to an uncorrelated set of *p* new variables. So PCA provides a means for reducing the dimensionality of the dataset, with the minimum loss of information. The *p* transformed variables calculated from the original variables are denominated principal components. The PCs are ordered so that the first component explains the largest portion of the variability, the second component explains the second largest portion, and so on.

The plot of the two first principal components obtained by PCA for the samples B1 and B2 dataset is presented in Figs, 1 and 2, respectively. As can be observed, the B1 base graph presents a greater overlap of groups of samples, where the groups represent different chemical compositional groupings. However, as can be seen in Fig. 2, the plot shows three clusters with a greater separation between them. Each cluster is very well defined showing of the raw material used in the manufacture of the samples is different.

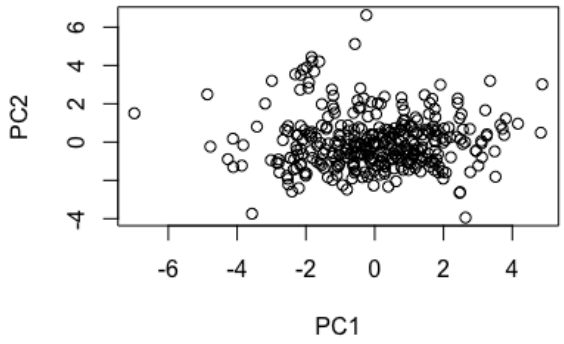

Figure 2: *Scatter plot of base B1 obtained after data projection using two principal components*

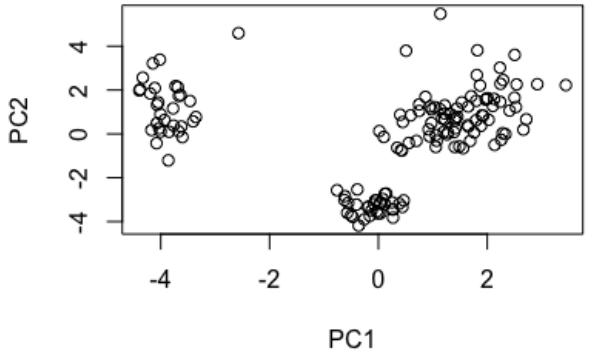

Figure 3: *Scatter plot of base B2 after data projection using two principal components*

The evaluation of the impact that standardization causes in clustering analysis using self-organizing maps was performed through the validation indexes of Jaccard [4], Rand [5] and Fowlkes-Mallows [6] the higher the index, the better result obtained by the SOM neural network [7].

The results obtained after cluster analysis of the transformed data corresponding to databases B1 and B2 are shown in Tables 3 and 4 respectively.

Table 3: Indices obtained after the application of the standardization techniques in the B1 database.

Transform/index	Jaccard	Fowlkes-Mallows	Rand
Logarithm	0.54	0.71	0.57
Log-generalized	0.33	0.50	0.62
Improved Min-max	0.64	0.79	0.77

Table 4: Indices obtained after the application of the standardization techniques in the B2 database.

Transform/index	Jaccard	Fowlkes-Mallows	Rand
Logarithm	1	1	1
Log-generalized	1	1	1
Improved Min-max	0.57	0.75	0.71

Table 3 shows that in the tests performed with B1, the standardization technique that presented better performance was the improved min-max. The values obtained from the Jaccard, Fowlkes-Mallows and Rand indices were, respectively, 0.64 , 0.79 and 0.77. However, The values obtained with log standardization were 0.54, 0.71, and 0.57 and finally the values corresponding to log-generalized standardization were 0.33, 0.50 and 0.62. The improved min-max was better that the other two (logarithm and log-generalized) because in the indexes are included the standard deviation of the concentrations and the two clusters have chemical composition very similar, as can be seen in Fig 2.

As can be seen in Fig 3, the PCA1 vs. PCA2 show that the samples of each cluster form a very tight chemically homogeneous group, shwing a high degree of chemical similaririty among them. From the archaeological point of view, the results showed that the clay from ceramics fragments from the three sites were originated from three distinct raw material

However, the improved min-max standardization presented the worst performance when is used the dataset B2, as can be seen in Table 4, since the values of the validation indexes of Jaccard, Fowlkes-Mallows and Rand were 0.57, 0.75 and 0.71. The Figure 3 show that the samples of each

cluster are well separated, form a very tight chemically homogeneous group, showing a high degree of chemical similarity among them. On the other hand, both log and log-generalized standardization presented all values of validation indices equal to 1.

4. CONCLUSION

Three standardization techniques (log, generalized-log and improved min-max) on the formation of clusters by means of the Kohonen neural network using two dataset, named B1 and B2, was made. The studies were performed using two different databases consisting of 298 and 146 samples. The results showed that the better performance was to improved minimum-maximum when the clusters does not show any significant differences in their composition. However, with the B2 dataset, the standardizations that presented the best performance were log and log-generalized because each cluster are separated and form homogeneous groups. Then, this study provided persuasive evidence that, when there is overlap between the groups, as in the case of dataset B1, the standardization technique with the best performance is improved min-max. On the other hand, if the groups do not present overlap, as in database B2, log and log-generalized techniques perform better.

REFERENCES

[1] PARDEDE, H. F.; SHINODA, K. Generalized-log spectral mean normalization for speech recognition, In: **TWELFTH ANNUAL CONFERENCE OF THE INTERNATIONAL SPEECH COMMUNICATION ASSOCIATION,** 2011, Florence/Italy, 2011, p.1645-1648.

[2] KABIR, W.; AHAMAD, M. O.; SWAMY M. N. S. A new anchored normalization technique for score-level fusion in multimodal biometrie systems, In: **2016 IEEE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS,** Montreal/Canada, 2016 p. 93-96.

[3] FÁVERO, L. P.; FÁVERO, P., **Análise de Dados: Técnicas multivariadas exploratórias com SPSS e STATA**, 1st ed., Brazil, Elsevier Brasil, 2017.

[4] JACCARD, P. Nouvelles recherches sur la distribution florale, **Bull. Soc. Vaud. Sci. Nat**., v. 44, p. 223-270, 1908.

[5] RAND,W. M. Objective criteria for the evaluation of clustering methods". **Journal of the American Statistical association**, v. 66, n. 336, p. 846-850, 1971.

[6] FOWLKES E. B.; MALLOWS, C. L. A method for comparing two hierarchical clusterings, **Journal of the American statistical association**, v. 78, n. 383, p. 553-569, 1983.

[7] HAYKIN, S. **Neural network: A comprehensive foundation**, 2sd ed., New Jersey, Prentice Hall PTR, 2004.

[8] VESANTO, J.; ALHONIEMI, E. Clustering of the self-organizing map. **IEEE Transactions on neural networks**, v. 11, n. 3, p. 586-600, 2000.

[9] JAIN, A. K.; DUBES, R. C. **Algorithms for clustering data** ", Englewood Clffs: Prentice hall, 1988.

[10] BRUN, M. Model-based evaluation of clustering validation measures, **Pattern recognition**, v. 40, n. 3, p. 807-824, 2007.

[11] KASZTOVSZKY, Z. ; BELGYA, T. ; KIS, Z. ; SZENTMIKLOSI, L. Nuclear Techniques for Cultural Heritage Research. **IAEA Radiation Technology Series**, n. 2, p. 121, 2011.

[12] SCHAAN, D. P. The nonagricultural chiefdoms of Marajó Island. In: **The handbook of South American archaeology**. Springer, New York, NY,. p. 339-357, 2008.

[13] SCHAAN, D. P. The Camutins chiefdom: rise and development of complex societies on Marajó Island, Brazilian Amazon. **Unpublished PhD dissertation, University of Pittsburgh**, 2004.

[14] 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**, v. 21, n. 4, p. 697-706, 2003.

[15] ALVES, M. A.; CHEUICHE MACHADO, L. M. Estruturas arqueológicas e padrões de sepultamento do sítio de Água Limpa, município de Monte Alto–São Paulo. **Anais da VII Reunião Científica da Sociedade de Arqueologia Brasileira**, n. 01, 1995, p. 295-310.

[16] ALVES, M. A.; CALLEFFO, M. E. V. Sítio de Água Limpa, Monte Alto, São Paulo-estruturas de combustão, restos alimentares e padrões de subsistência. **Revista do Museu de Arqueologia e Etnologia**, v. 6, p. 123-140, 1996.

[17] ALVES, M. A. Estudo do Sitio Prado, um sitio lito-cerâmico colinar. **Revista do Museu Paulista**, v. 29, p. 169-199, 1983.

[18] ALVES, M. A. Culturas ceramistas de São Paulo e Minas Gerais-estudo tecnotipológico. **Revista do Museu de Arqueologia e Etnologia da USP, São Paulo**, v. 1, n. 1, p. 71-96, 1991.

[19] ALVES, M. A. Estudo técnico em cerâmica pré-histórica do Brasil. **Revista do Museu de Arqueologia e Etnologia**, v. 4, p. 39-70, 1994.

[20] ALVES, M. A. O emprego da microscopia petrográfica, difratometria de raios X e microscopia eletrônica no estudo da cerâmica pré-colonial do Brasil. **Revista de Arqueologia**, v. 8, n. 2, p. 133-140, 1994.

[21] ALVES, M. A. As estruturas arqueológicas do Alto Paranaíba e Triângulo Mineiro-Minas Gerais. **Revista do Museu de Arqueologia e Etnologia**, n. 2, p. 27-47, 1992.

[22] ALVES, M. A. Metodologia e tecnicas de campo ea evidenciacao de Areas culturais. **Colecao Arqueologica. Edipucrs: Porto Alegre**, v. 1, p. 255-270, 1995.

[23] SANTOS, J. O. ; REIS, M. S. ; MUNITA, C. S.; SILVA, J. E. Box-Cox transformation on dataset from compositional studies of archaeological potteries. **Journal of Radioanalytical and Nuclear Chemistry**, v. 311, n. 2, p. 1427-1433, 2017.

[24] MUNITA, C. S. ; PAIVA, R. P.; ALVES, M. A. ; OLIVEIRA, P. M. S. ; MOMOSE, E. F. Provenance study of archaeological ceramic. **J. of Trace and Microprobe Techniques**, v. 21, n. 4, p. 697-706, 2003.

[25] MUNITA, C. S. ; PAIVA, R. P.; ALVES, M. A. ; OLIVEIRA, P. M. S. ; MOMOSE, E. F. Major and trace element characterization of prehistoric ceramic from Rezende archaeological site. **Journal of Radioanalytical and Nuclear Chemistry**, v. 248, n. 1, p. 93-96, 2001.

[26] MUNITA, C. S. ; TOYOTA, R.G. ; NEVES, E. G. ; DEMARTINI, C. C. ; SHAAN D.P. ; OLIVEIRA, P. M. S. Chemical characterization of Marajoara pottery. In : **Nuclear Techniques for Cultural Heritage Research**, IAEA Radiation Technology Series n. 2, Chapter 7, p. 133-145, 2011.