



## **Application of Exploratory Data Analysis: HCA and PCA in the Classification of Tubular Wells of the North of Brazil**

**F. S. Panero<sup>1</sup>, H. E. B. Silva<sup>2</sup>**

<sup>1</sup>Student; Department of Physics, University Federal de Roraima, Boa Vista, Brasil

<sup>2</sup>Department of Chemistry; University Federal de Roraima; Av Ene Garcez 2413 - Aeroporto CEP 69304000, Boa Vista-RR, Brasil

### **Abstract**

Roraima is located in the extreme north of Brazil, in the Western Amazon Region. The groundwater has an important role in the public supply of potable water on the Roraima State. The objective of this work is the application of PCA (Principal Component Analysis) and HCA (Hierarchical Cluster Analysis) for the classification tubular wells of the Roraima State, as regards composition physical-chemistry of water. In the physical-chemistry terms, the groundwater samples analyzed not indicate not any restriction to the human consumption and the two chemometrics methods of exploratory data analysis PCA and HCA are efficient for discrimination of groundwater bicarbonated (tubular wells located in the south of the Roraima State) and not bicarbonated.

### **Introduction**

All the existing forms of life in the planet depend on the water. Each human being needs to consume some liters of potable water daily to remain living. However, the potable water is a prize. Of the whole volume of potable water available in the planet, the subterranean water constitutes more than 98%. Roraima is located in the extreme north of Brazil, in the Western Amazon Region [1]. The groundwater has an important role in the public supply of potable water on the Roraima State, 87% of the cities in the State makes use partial or totally of the groundwater [2].

The objective of this work is the application of two methods of exploratory data analysis, PCA (Principal Component Analysis) and HCA (Hierarchical Cluster Analysis) for the classification tubular wells of the Roraima State, as regards composition Physical-Chemistry of water.

### **Materials & Methods**

The physical-chemistry analysis of groundwater was made in the full season, in September and October of 2000. Through the following parameters: pH, electric conductivity (Cond), Total Hardness (Dur Tot), Total Alkaline (Alc Tot), Bicarbonate ( $\text{HCO}_3$ ), Calcium (Ca), Magnesium (Mg), Sodium (Na), Potassium (K), Chloride (Cl), Nitrate ( $\text{NO}_3$ ) and Total Iron (Fe T) in 39 points of sampling (tubular wells) [3]. In the chemometrics analysis through PCA and HCA, was used the whole data set, resulting in one matrix 39 x 12. The preprocessing used were the autoscale. After the preprocessing were calculated the scores and loadings of the data, using the software EIN\*SIGHT [4]. To obtain the dendrogram was used the Euclidean Distance and the Incremental connection (cluster) method.

### **Results**

Through the chemometrics analysis for Principal Component Analysis – PCA, only analysing PC1 x PC2, it is observed formation of two great groups and the discrimination between the groundwater bicarbonated (tubular wells located in the south of the Roraima State) and not

bicarbonated, as it can be observed in Figure 1. The Figure 2 it is observed the factors responsible for the classification. That two principal components (PC1 x PC2) describe 74,93% of the total variation of data. It classification was confirmed by the Hierarchical Cluster analysis - HCA. That provide obvious separation between the groundwater bicarbonated and not bicarbonated (Fig. 3).

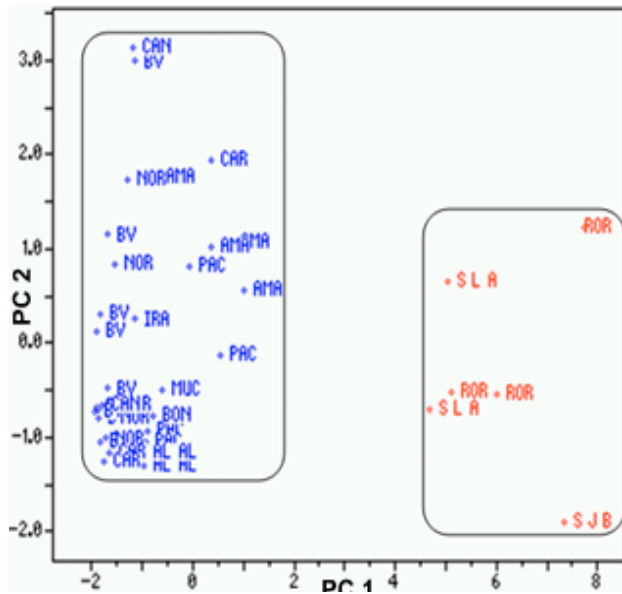


Fig.1; scores, PC2 vs PC1, autoscaling, tot. var. 74.93%

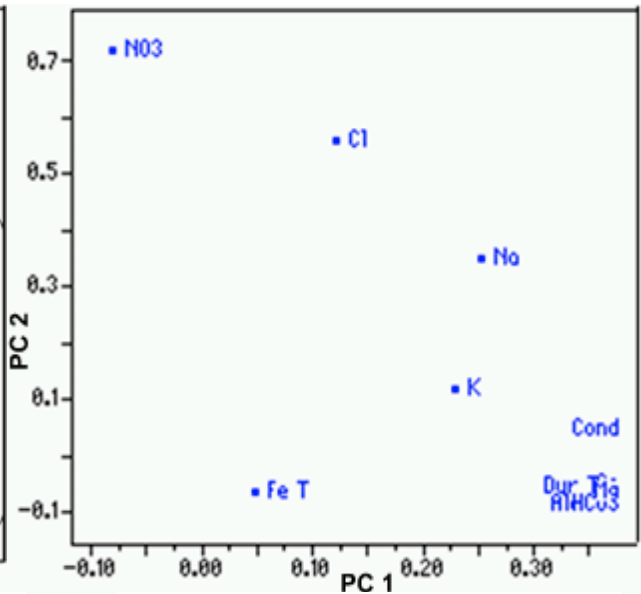


Fig.2; loadings, PC2 vs PC1, autoscaling

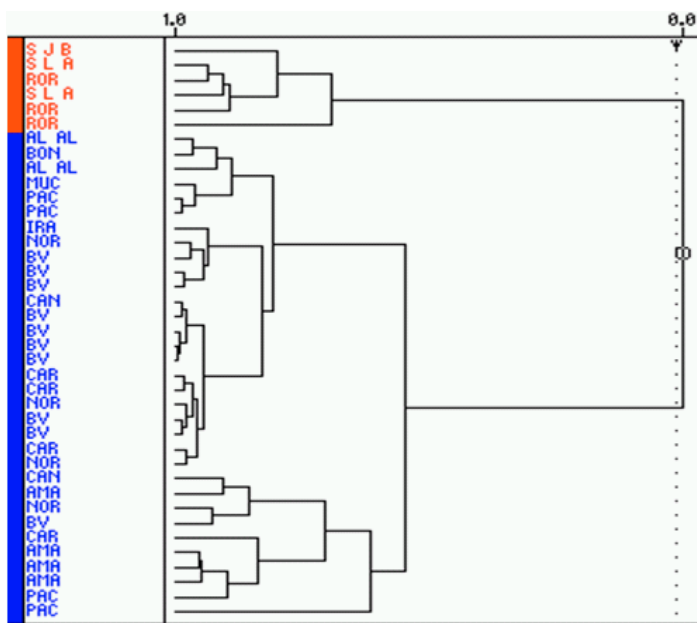


Fig.3; dendrogram of samples (HCA), autoscaling

## Conclusions

In the physical-chemistry terms, the groundwater samples analyzed not indicate not any restriction to the human consumption. The PCA (Principal Component Analysis) was efficient for classification of tubular wells of the Roraima State, in groundwater bicarbonated (tubular wells located in the south of the Roraima State) and groundwater not bicarbonated. Results also obtained with the HCA (Hierarchical Cluster Analysis).

## References

- 1) Roraima: O Brasil do Hemisfério Norte, AMBTEC, 1994.
- 2) CAER: Companhia de Água e esgoto de Roraima, 2000.
- 3) Zoneamento Ecológico-Econômico da Região Central do Estado de Roraima – ZEER-RR , CPRM, Hidrogeologia, Tomo III, 2002.
- 4) EIN\*SIGHT 3.0, Infometrix, Seattle, 1991.