



Usefulness of Chemometrics to Classify Pottery Sherds Belonging to the Southeast of the United States

C. Pizarro, S. Rodríguez-Tecedor, Nuria Pérez-del-Notario, J. M. González-Sáiz

Department of Chemistry, University of La Rioja, C/Madre de Dios 51, 26006 Logroño, La Rioja, Spain

Abstract

The compositional information of an Instrumental Neutron Activation Analysis (INAA) of pottery sherds from the Southeast of United States belonging to the Steponaitis set in MURR database was combined with Chemometric techniques to create a geographical classification.

This study allows not only to confirming previous conclusions (Steponaitis et al.1996) for the classification in four geographical areas but also for subdividing the samples into finer units with good classification and prediction results validated by leave-one-out cross validation.

Introduction

Analytical techniques used in chemical composition analysis supply large sets of data and variables. Therefore, the application of multivariate treatments, variable selection, and pre-processing data is generally necessary in order to recognize compositionally homogeneous groups of samples that may be associated with geographical areas.

The present work aims improving the contribution by Steponaitis et al. (1996) [1] to the classification of pottery sherds from the Southeast of United States by subjecting the elemental composition data obtained by INAA to Chemometric analysis.

The authors of the previous work classified the sherds in four geographical regions (Northern, Southern, Eastern and Western) based on compositional information but some questions were unsolved in that work. We consider two of them, first the fact that the 30 per cent of the objects were ungrouped and second, that the subdivision of the regions into finer areas were not studied.

By applying different Chemometric tools such as LDA [2] combined with pretreatments and stepwiseLDA [2], we expect to achieve a successful classification in four large regions as well as their subdivision into finer units.

Material & Methods

The compositional information reported in the Steponaitis dataset available in the MURR database for about twentyfive elements, determined by Instrumental Neutron Activation Analysis (INAA) as previous described [1], for 186 archaeological ceramic sherds from 21 different regions across the Southeast of United States was analyzed using multivariate data analysis and scaling procedures.

Linear Discriminant Analysis (LDA) was applied after column pre-treatment as a classification method working on PCs when were necessary due to the large number of original variables, and using leave-one-out cross-validation for the validation of results. In many cases classification and prediction results were improved by applying StepwiseLDA as feature selection technique based on minimum Wilks Lambda to eliminate non informative variables. All calculations were carried out with V-Parvus software [3].

Results

Classification and prediction rates obtained after applying LDA with only 6 sherds not correctly classified are reported in Table 1.

In the first step, LDA was performed on the column centered data of the whole dataset, without any variable selection and using leave-one-out cross-validation to validate the results.

Secondly, each group was separately studied and subdivided by applying StepwiseLDA followed of the multivariate treatments used in the first step. Fig. 1 shows one of these separations.

Finally the categories that overlapped in the second step were studied again in order to separate them.

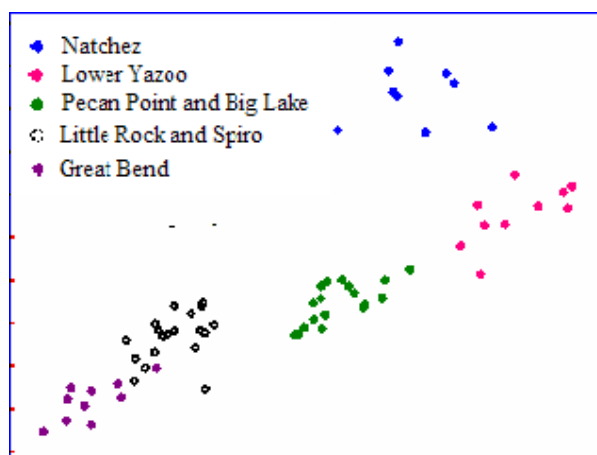


Fig 1. Projection on the two first canonical variables of LDA for Western region

Table 1: LDA results: percentage of correct classifications and predictions

	Subdivisions achieved	Classification ability (%)	Prediction ability (%)
Full set	Northern; Southern; Eastern; Western	94.70	86.56
North	Nashville; Tellico & Servieville; Harpeth	97.06	90.00
	Tellico & Servieville	99.75	90.00
South	Mobile Delta & Bay; Black Warrior & Gainesville; Wheeler Lake, Wetumpka	100.00	97.67
East	Carters Lake; Fort Gaines; Eufaula	96.43	72.41
West	Lower Yazoo, Natchez, Pecan Point & Big Lake; Great Bend, Little Rock & Spiro	100.00	87.88
	Pecan Point & Big Lake	94.74	89.47
	Little Rock & Spiro	95.06	88.89

Conclusions

The successful results obtained for the geographical discrimination of most of the sherds are an improvement on those obtained in a previous study, thus proving the potential of the multivariate data analysis applied. A large number of samples would be necessary to include Natchitoches region in the study and for the separation of the overlapped regions in Southern region.

References

- 1) V. P. Steponaitis, M. J. Blackman, H. Neff, Large-Scale patterns in the chemical composition of Mississippian pottery, *American Antiquity*, 6(3), (1996), 555-572
- 2) C. Pizarro, I. Esteban-Diéz, C. Saénz-Gonzalez, J.M. Gonzalez-Saiz, Vinegar classification based on feature extraction and selection from headspace solid-phase microextraction/gas chromatography volatile analyses: A feasibility study, *Anal. Chim. Acta*, 608(1), (2008), 38-47
- 3) M. Forina, S. Lanteri, C. Armanino, M. C. Cerrato-Oliveros, C. Casolino, V-Parvus. An extendable package of programs for explorative data analysis, classification and regression analysis, Dipartimento di Chimica e Tecnologie Farmaceutiche ed Alimentari, Università di Genova, Italy (2004), free available at <http://www.parvus.unige.it>