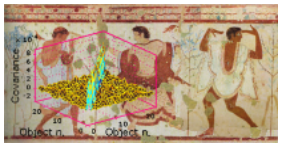

Feature Selection Applied to Study the Source of the Stone Used to Build a French Cathedral

C. Pizarro, C. Sáenz-González, S. Rodríguez-Tecedor, J. M. González-Sáiz

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

Abstract

The source of the stone used to build the "Nôtre-Dame de Senlis" cathedral in Senlis (Northern France) has been determined by using compositional data taken from the Limestone Sculpture Provenance Project database.

The application of stepwise linear discriminant analysis (SLDA) has enabled to select, with the classification algorithm, the most discriminant variables between five nearby limestone quarries around Senlis.

Once the quarries has been discriminated, SIMCA has been applied in order to obtain a class model for each quarry in an attempt to assign the samples from "Nôtre-Dame de Senlis" to one of the five categories considered.

Introduction

Knowing the source of the stone used to build a monument can help identify the different building campaigns and show patterns of exchange in stone [1]. Moreover, it could permit to discriminate the original construction from any restorations or additions.

Differences in the microscopic structure of stones from the same stone formation are often undetectable. Therefore, petrography is not capable of discriminating stone sources within the same sedimentary facie and new analytical methods are needed [2]. On the other hand, the composition of the stone is not constant in the whole stone formation and discrimination among different sources is possible by means of compositional fingerprinting.

The Limestone Sculpture Provenance Project [3] has built a database that includes compositional analysis of samples taken from sculptures in museum collections, quarries and monuments, carried out by Neutron Activation Analysis (NAA).

Since applications based on NAA coupled with multivariate analysis techniques seem to show great promise for provenance studies[1] and given the great improvement in the classification results obtained with the application of a variable selection algorithm[4], we decided to perform a more in-depth study to test the feature selection to compositional data from limestone.

However, taking into account the compositional variation of both the "object" under study and the quarries as well as the errors in the analytical procedure, results only can be shown as statistical probabilities [1].

Materials & Methods

The Limestone Sculpture Provenance Project database was used as source of the compositional data used in this study. The compositional data, obtained by NAA, comprises the concentrations of twenty five elements (Na, K, Rb, Cs, Ba, Sc, La, Ce, Eu, Lu, Hf, Th, Ta, Cr, Mn, Fe, Co, Sb, U, Zr, Ca, As, Sm, Yb, Sr), in the limestone samples. These concentrations, given as the concentrations of the oxides of those elements, have been considered as the sample compositional profile or "fingerprint".

For investigation, twenty nine quarry samples were collected from four quarries presumed to have furnished the stone used to built the Senlis cathedral because their proximity to Senlis. The

quarries are located in villages near Senlis, i.e., Mont l'Eveque, Ognon, Saint-Vaast-les-Mèllo and Courteuil. Eleven samples from the local quarry in Senlis were also included in the study.

Then, seven samples were taken from different parts of "Nôtre-Dame de Senlis".

Stepwise linear discriminant analysis (SLDA) has been applied as classification method for a simultaneous feature selection (based on Wilks' Lambda criterion) and classification. Additionally SIMCA has been used to define the limits of the "class space" for the five quarries under study. All calculations were carried out with V-Parvus software [5].

Results

Stepwise linear discriminant analysis has been performed on concentrations of elements data in order to highlight the most significant variables in order to discriminate between geographically separated quarries. The prediction ability of the SLDA model developed was evaluated by leave-one-out cross-validation, fig. 1.

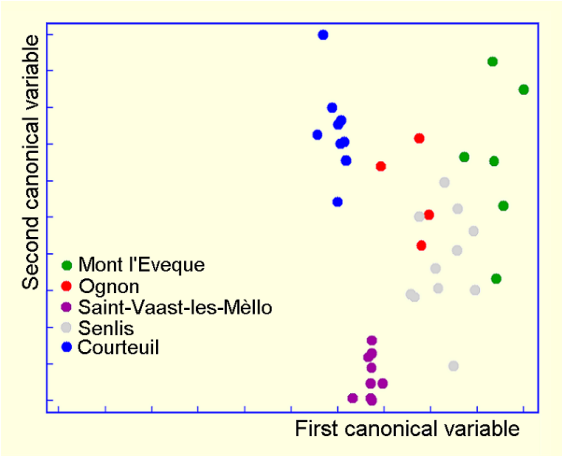

Fig. 1; Separation of the quarries on the first and second canonical variables obtained with LDA after variable selection.

The best results have been obtained with the selection of only ten oxides (Fe, Co, Sm, Sr, Ca, Ba. Eu, Hf, Mn, Cr) to compute a model capable of achieving a perfect classification of the quarries and a prediction ability of 95%.

After the classification step, SIMCA has been applied on the data matrix formed by the ten variables selected and the forty samples as class-modelling technique providing a sensibility of 97.5% and a specificity of 93.75%. Five of the seven samples taken from the cathedral were assigned to the same category corresponding to the quarry located in Mont l'Eveque.

Conclusions

The main strength and novelty of the classification approach reported has to do with the application of a variable selection method to extract a minimum number (maximum parsimony) of discriminant features from the data. This method improved the results obtained from the application of pattern recognition techniques to limestone compositional data.

The good classification results achieved for five nearby quarries around Senlis (Northern France) allowed for assigning the origin of samples taken from the "Nôtre-Dame de Senlis" cathedral to one of the five quarries studied, Mont l'Eveque.

The low number of predictors selected to the differentiation of these quarries can make it easier for petrographers and archaeologists to interpret the results.

References

1) L. L. Holmes, G. Harbottle, Old stones, new science, J. Radioanal. Nucl. Chem., 248(1) (2001) 75-79
2) J. Herrmann, N. Herz, R. Newman (eds.), ASMOSIA 5, Interdisciplinary Studies on Ancient Stone, Archetype Publications (2002), 103-109, ISBN: 1873132085
3) The Limestone Sculpture Provenance Project, www.limestonesculptureanalysis.com, accesed at May/2008
4) C. Pizarro, I. Esteban-Díez, C. Sáenz-González, J. M. González-Sáiz, Vinegar classification based on feature extraction and selection from headspace solid-phase microextraction/gas cromatography volatile analyses: A feasibility study, Anal. Chim. Acta, 608(1) (2008) 38-47
5) 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, 2007