Analytica Chimica Acta 646 (2009) 69–77

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Classification of archaeological sherds across the southeast United States based on variable selection from compositional fingerprints

C. Pizarro*, J.M. González-Sáiz, I. Esteban-Díez, S. Rodríguez-Tecedor, Nuria Pérez-del-Notario, C. Sáenz-González

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

ARTICLE INFO

Article history:
Received 13 March 2009
Received in revised form 6 May 2009
Accepted 14 May 2009
Available online 21 May 2009

Keywords:
Archaeological sherds
Compositional fingerprint
Feature selection
Stepwise linear discriminant analysis
Provenance study

ABSTRACT

The transfer of advances in chemometrics into archaeometric research opens a wide range of new application possibilities in this rapidly developing field. The present research represents a feasibility study aimed at showing how the huge potential that multivariate analysis and feature selection techniques have demonstrated for classification purposes can be extrapolated to archaeological provenance studies, thus pursuing an enhancement of the resulting classification performance. The classification problem studied here was related to the discrimination of pottery sherds from different sources across the southeast of the United States from their compositional fingerprints. The sample elemental concentrations were analyzed using the stepwise linear discriminant analysis (SLDA) method, thus simultaneously performing feature selection and classification. Several approaches, more or less restrictive according to the geographical scope and the number of considered classes, were explored, following a hierarchical classification approach. In contrast to previous studies on the same data set, the reliable and unequivocal classification strategy presented here did not merely focus on developing a large-scale classification into broad geographical areas, but finer classifications were also successively obtained until samples were assigned into individual regions. The great discrimination ability and effectiveness of the classification methodology proposed are promising and encourage its application to new samples of unknown provenance and the feasibility of using similar approaches in other archaeological studies. The high quality results obtained were even more remarkable considering the relatively small number of discriminant variables selected in each case by the stepwise procedure.

1. Introduction

Modern archaeology generates amazing amounts of data thanks to advances in instrumental analysis. Thus, in recent years, the characterization of ancient pottery to determine its provenance has involved the development, exploration, and testing of a large number of analytical techniques in order to compile and extract relevant compositional information. The main methodologies currently employed in archaeometry analysis [1–3] include inductively coupled plasma (ICP) coupled to mass spectrometry (ICP-MS) [4,5], atomic emission spectroscopy (ICP-AES) [6,7], optical emission spectrometry (ICP-OES) [8], Neutron Activation Analysis (NAA) [9–12], electron microprobe analysis (EMPA) [13], emission atomic spectroscopy (AES) [8], X-ray diffraction (XRD) [14,15], X-ray fluorescence (XRF) [16,17], Raman and micro-Raman spectroscopy [14,15], Fourier transform infrared spectroscopy (FT-IR) [18,19], laser-induced breakdown spectroscopy (LIBS) [20] and petrography [21,22]. In this context, Instrumental Neutron Activation Analysis (INAA) has been particularly consolidated as a powerful analytical tool widely employed in archaeometric studies for the chemical characterization of sherds [23]. The main reasons for the extended application of this method are its multi-element capability (over thirty elements can be determined simultaneously without chemical separation), its high sensitivity (high-accuracy determination of elemental concentrations, often in ppm or even ppb) and precision (measurement errors of <3–5% are standard), and the fact that minimal sample preparation is required (although pulverization and homogenization are needed, limited handling reduces sample contamination and accelerates analytical process). Nevertheless, the need for using a portion of the artifacts to carry out the analysis might become a disadvantage in the case of valuable archaeological samples [24].

However, it should be borne in mind that the passive collection of information does not generate useful knowledge on its own; hence, the application of efficient data processing methods for gathering and extracting significant information from raw data is essential in order to obtain the best knowledge of historical her-

* Corresponding author. Tel.: +34 941299626; fax: +34 941299621.
E-mail address: consuelo.pizarro@unirioja.es (C. Pizarro).

0003-2670/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2009.05.021

itage for society [25]. In this context, as far as we are aware, in most archaeometric studies described in recent literature only basic statistical tools have been commonly applied, which means that the full potential of archaeological research is far from fully exploited. In this respect, the notable success of chemometrics (a twin discipline of archaeometry by nature) in many application areas [26–32] should serve as an example and incentive for promoting the further development of this relatively young field of research. Thus, the integration of more powerful and versatile multivariate analysis tools in the analysis of large archaeological data sets could represent a highly beneficial symbiosis, and hold the key to future trends in provenance studies and open up new opportunities for development.

In this work, through the combination of INAA data and multivariate analysis, a classification problem addressed in previous studies was studied in an attempt to significantly improve the results obtained thus far: the classification of archaeological sherds from the southeast of the United States according to their geographical origin based on their compositional data. The data set used was downloaded from the Archaeometry Laboratory at the University of Missouri Research Reactor (MURR) database [23], freely available online through the MURR webpage for use in further research (as in this case for the present study). The Archaeometry Laboratory of MURR is specialized in trace-element analysis (i.e. chemical fingerprinting) of archaeological specimens for determining their provenance, thanks to its expertise in several powerful analytical techniques such as Instrumental Neutron Activation Analysis (INAA), X-ray fluorescence (XRF), and inductively coupled plasma-mass spectrometry (ICP-MS).

A previous study on the classification problem considered here [33] led to the assignment of sherds by principal component analysis of their compositional data into four distinct large-scale patterns that were associated with four geographical areas (Northern, Southern, Eastern and Western). However, as noted by the authors of that study, a number of questions remained unsolved in their work: a finer subdivision could not be achieved and 30% of the objects were unclassified. This result further justifies the aim pursued in the present study, i.e. to improve results obtained previously in order to provide a more reliable classification methodology in order to objectively discriminate between samples of different provenances.

The study of pottery distribution, trade and exchanges is a major issue in archaeology (and a vital source of information in culture change studies) from ancient times in human history, since information about artifact provenance (or geographical source) can be used, directly or indirectly, to draw inferences regarding many aspects of ancient cultures and their interactions (including technology, social structure, cultural influences, technology, and economics). In this context, the origin and extent of the pottery trade throughout the southeast of the United States still remain an open question. The feasibility of gaining further knowledge about the compositional fingerprint of groups of sherds belonging to this broad area and thus achieving an objective and reliable classification (focusing on providing more specific and refined subclassifications than those reported in previous studies on the same topic) highlights the importance and potential benefits of this study since it represents the most effective way to tackle the real core of any archaeometric study in a subsequent stage: a suitable interpretation of classification results in order to draw valuable archaeological conclusions (concerning the culture, trade and/or economical organization of ancient societies).

The classification methodology proposed was based on the combined use of INAA data for twenty-five elements of each archaeological sherd sample with the subsequent blind analysis of these chemical fingerprints using a stepwise linear discriminant analysis (SLDA) method [34,35] to distinguish between different geographical provenances. The application of stepwise LDA as a classification technique enabled us to simultaneously perform feature selection and classification by selecting the most significant elements for a reliable discrimination between sherd classes in order to use these later as input for LDA [27,36,37]. Different classification approaches were explored (differing in the number and properties of classification categories to be taken into account) in order to differentiate between not only the aforementioned broadest geographical areas but also more specific provenances and locations.

Although it should be borne in mind that this is only a feasibility study, the promising results obtained encourage similar approaches to be considered in the future in order to expand the field of application to a wider range of provenances and other archaeometric studies. However, it should be stressed that the classification methodology validated and presented in this paper should be considered at all times as an applicative tool to be used to draw more in-depth archaeological conclusions in subsequent studies. The present work does not intend to solve all the underlying issues related to archaeological sherds in the southeast of the United States. Actually, it would be impossible to cover in a single work a global study aimed at simultaneously accomplishing both the development of a reliable strategy for determining the geographic origin of samples and its application to draw archaeological conclusions on trade and exchange routes. A further complementary final stage devoted to the archaeological interpretation of the classification results should be carried out (preferably by the original research group that has available the whole archaeological information on the samples) in order to implement in practice the classification strategy here proposed, investigating the typochronology of the ceramics considered and checking these data against the geographic classification results obtained.

2. Materials and methods

2.1. Sherd samples

The Steponaitis data set [33] contained compositional information about one hundred and eighty-six sherds from twenty-one different regions across seven states in the southeast of the United States. The samples locations are listed in Table 1 and shown in Fig. 1.

Table 1
List of archaeological samples, showing their distribution in the different regions and states.

State	Region	Number of samples
Oklahoma	Spiro	9
	Little Rock	10
	Great Bend	10
	Big Lake	10
	Pecant Point	9
Louisiana	Natchitoches	3
Mississippi	Lower Yazoo	10
	Natchez	10
Tenessee	Nashville	11
	Lower Harpeth	10
	Servierville	10
	Tellico	10
Alabama	Wheeler Lake	9
	Gainesville	4
	Black Warrior	10
	Wetumpka	10
	Mobile Delta	9
	Mobile Bay	2
Georgia–Alabama border	Fort Gaines	10
	Eufaula	10
Georgia	Carters Lake	10