J Radioanal Nucl Chem (2013) 298:1257–1272
DOI 10.1007/s10967-013-2603-y

k0-INAA of Venezuelan ceramics and complete statistical analysis to establish their provenance

F. Pino · A. M. Sajo-Castelli · H. Barros · P. Vermaercke · L. Sneyers · L. Sajo Bohus · Ma. M. Mackowiak de Antczak · A. Antczak

Received: 8 February 2013 / Published online: 16 July 2013
© Akadémiai Kiadó, Budapest, Hungary 2013

Abstract A group of 46 archaeological figurines samples (AD 1300 and 1500) from Venezuelan mainland and northern island were analyzed by k0-instrumental neutron activation analysis (k0-INAA) to obtain their elemental content and give a step ahead to establish the provenance of the island figurines. In total 37 elemental concentrations were measured with uncertainties between 3 and 20 %. To make the study of provenance, a complete statistical analysis was achieved; Fisher linear discriminant, principal component analysis, hierarchical clustering and the Hotelling T^2 test were used to this end. Furthermore, not only the 46 samples analyzed in this work by k0-INAA were used, but also 40 samples analyzed by PGNAA and reported by Sajo-Bohus et al. (JRNC 265(2):247–256, 2005) were included in the statistical analysis. It was done in order to increase the size of the data set, and then to obtain from the statistical techniques more reliable results. It was found that a very good differentiation exits between the figurines from the island and from the mainland supporting the idea that the raw materials of the figurines come from different places.

Keywords Neutron activation analysis · Provenance study · Statistical techniques · Pre-hispanic figurines

F. Pino (✉) · A. M. Sajo-Castelli · H. Barros · L. Sajo Bohus · Ma. M. Mackowiak de Antczak · A. Antczak
Universidad Simon Bolívar, Apdo 89000, Caracas 1080, Venezuela
e-mail: felixpino@gmail.com

P. Vermaercke · L. Sneyers
SCK CEN Belgian Nuclear Research Centre, Mol, Belgium

Introduction

The presence of a high number of anthropomorphic figures in its practically original undisturbed context at the small oceanic island Dos Mosquises (Los Roques Archipelago, Venezuela) is an exceptional archeological phenomena, which point out to deep interpretations. The present work is one of the steps forward to provide material evidence of use for archeologists in its transcendental work of reconstruct the meaning of this cultural manifestation in relation with the Valenciod sphere of interaction along the central Venezuelan coast and the surrounding islands.

A set of figurines samples of Valencia Lake Basin (V) and Los Roques Archipelago (R) were analyzed as part of a large study to establish the extension of the Valenciod influence during pre-Columbian centuries. Amerindian population navigated seasonally between the Venezuelan north central regions to the Caribbean islands and it is assumed that they participated in the diffusion of the Valencia style pottery, from the Valencia Lake Basin to Los Roques islands around AD 1300 and 1500. The archeological hypothesis is that Los Roques figurines have the same origin than Valencia Lake figurines. This possibility is based on stylistic and morphologic analysis and because their geographic proximity [1]. However, the statistical tests were carried out trying to make as less assumptions as possible. Results reported here together with those already available contribute to establish colonization or demographic expansion of the V-settlers [1–5].

From almost 800 pre-hispanic pottery figurines found, since now more than 100 have been analyzed and it is believed that more samples should be scrutinized to establish on firm ground the origin of R-island figurines and their relation with Valencia Lake Basin culture. In this study 46 artifacts of archaeological interest are included

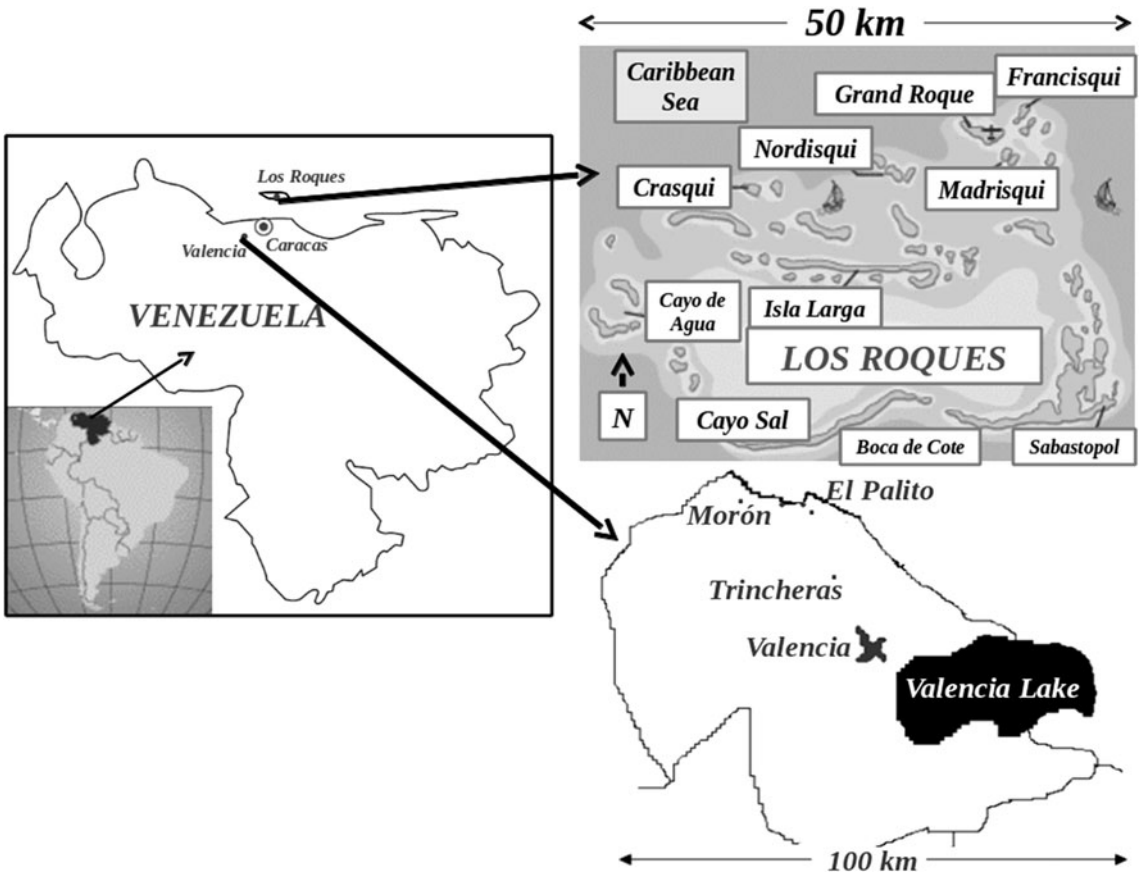

Fig. 1 Maps of the two regions of interest: Valencia Lake Basin and Los Roques Archipelago

with the purpose to shed further light on their provenance as a complementary data. Figurines excavated in the Valencia lake region and Los Roques Islands have been analyzed for different techniques as described in detail in [1–4]. In Fig. 1 the archaeological sites are shown.

k0-Instrumental neutron activation analysis (k0-INAA)

Brief description of the technique

Neutron activation analysis is an analytical technique based on the neutron capture of a stable nucleus that compounds a material when it is irradiated with neutrons, usually in irradiation channels of nuclear reactors. The neutron capture of a stable nucleus may lead the formation of a radioactive isotope of the same element, it can be use to quantify the corresponding elemental concentration by implementing gamma spectrometry. The classical NAA is based on a comparison of activities of the unknown samples and a known standard co-irradiated under similar conditions, but the problem is that it is needed one standard per elemental concentration to be determined [5]. The k0-INAA method, developed in the mid 70s by Prof. F. De Corte (INW, Gent) and Dr. A. Simonits (KFKI, Budapest) is more flexible to implement [6, 7]. The idea behind this method is based on the accurate knowledge of the neutron flux parameters—the thermal to epithermal flux ratio (f) and the shape parameter of the epithermal flux distribution (α), the characteristics of the gamma spectrometry detector (detection efficiency, detailed information about of the geometry and composition of the detectors, etc.) and the fundamental physical constants, known as k0 factors, proportional to molar mass, natural abundance, gamma ray intensities and cross sections [7]. These requirements and the sample characteristics, such as mass, geometry and matrix composition, are sufficient to determine elemental concentrations in the sample, without the use of (mono- or multi) elemental standards. The mains advantages of the k0-INAA technique are sensitivity, freedom of contamination after irradiation, selectivity, and with only few milligrams of sample it is possible to make a high quality and precise analysis. Besides, a complete "panoramic" analysis can be performed between 20 and 25 days.

Sample preparation

A set of 46 representative samples of pre-Hispanic figurines excavated in the Valencia Lake Basin (18) region and the Los Roques Archipelago (28) were analyzed, in Fig. 2 is shown a typical figurine of "Valencioide" style.

The preparation of the samples was achieved very carefully scraping with an agate stone the external and/or the internal body of the figurines. To avoid large damages to the invaluable archaeological material, in some cases only one kind of sample (inner or outer) was taken, of course, tests to probe that there are not systematic differences between both kind of samples were done. After that, the scraped material was pulverized using an agate mortar. All the set of samples were oven dried for 48 h at 100 °C. The samples were weighed directly into small polyethylene vials, capped and heat sealed. The weight of the samples varied between 10 and 200 mg. Besides, synthetic multi-elemental standard (SMELS) I, II, III [8] and the IAEA Soil 5 reference material were included in the set of samples in order to validate the obtained results. Finally, a set of gold comparators (Al with 0.1 % Au foil) were included uniformly distributed in the package of samples to be irradiated.

Irradiation

All the samples, the multi-elemental standards and the gold comparators were irradiated in the BR1 Reactor at the SCK-CEN, Mol-Belgium, which is a 4 MW graphite moderated and air-cooled Research Nuclear Reactor, working with natural uranium and it has two very good characterized irradiation channels [9]. The channel "S84" was used for the short irradiations, it is equipped with a pneumatic system in the top of the reactor which let load/unload samples into/out very quickly. The time of the short irradiations was 5 min. In the short irradiation the samples were irradiated one by one. On the other hand, the channel "Y4" was used for the long irradiations (7 h), it must be loaded with a package of samples before starting up the reactor and may only be unloaded after shutting off the reactor and after cooling the samples.

Fig. 2 Typical figurine of "Valencioide" style

Counting

The measurements of the γ-ray spectra were performed on four conventional gamma ray spectrometers equipped with p-type coaxial HPGe-detectors (Canberra) with relative efficiency 40 %, and an electronic unit to compensate for losses pulses (Loss Free Counting, LFC) which is very useful when it is desirable to measure samples with high activities (it occurs very often in the measurements to determine elements with short half lifes).

The measurements of the γ-ray spectra after short irradiations were performed on the top of the reactor BR1 with two spectrometers. Three spectra of each sample were taken; the first one was taken immediately after the irradiation during 5 min (dominated mainly for Al-28). The second one was taken after around 5–10 min of cooling to determine elements like Cl, Mg, Ti, V during 5 min. And the last spectrum was taken after a few hours of cooling with special interest to determine Dy ($t_{1/2} = 2$ h) during 10 min. The reference material used to validate the results in short irradiations was the SMELS I, and the procedure to measure it was the same than the sample's one.

On the other hand, the procedure to measure the samples and references materials SMELS II, SMELS III and IAEA Soil-5 after of the long irradiations was in the following way: approximately three spectra were taken of each sample in two gamma spectrometers located at the laboratory; the first one was taken after about 15 h of cooling during 30 min to measure elements like Na, K, Mn, etc. The second one was taken after 2 or 3 days of cooling to determine Sm, As, La, Fe, etc. and reduce the Compton Effect in the spectra produced mainly by Na-24. And the last measure was made, after about 2 weeks of cooling, during 24 h in order to determine elements with very long half-lives.

Finally, after 2 weeks of the irradiations, were measured the gold comparators in all the different positions which were used to measure the samples. These measurements were performed very quickly due to the activity of the comparators, which were high enough to get a good statistic in the determination of the peak area of the 411.8 keV peak of Au-198.

It should be mentioned that the measurement times have been determined by choosing times which were long enough to provide gamma peaks (most of them) with areas of minimum 10.000 counts (~ 1 % standard deviation) as a general rule (if the background is not too high).

PGAA versus INAA

Due to previous reported results of samples analyzed by prompt gamma activation analysis (PGAA) [2] are used in this work to make a complete statistical analysis, it is worth to make a brief description of PGAA. It is also a analytical technique based in the capture of a low energy neutron by a stable nucleus. Because of the binding energy of the added neutron, after capture, the new nucleus is in an excited state, about 10^{-14} s later characteristics gamma rays are emitted as a consequence of the deexcitation of that state. The emission rate of these gamma rays from a sample in a neutron beam can be used to measure the amount of that nucleus in the sample, which is proportional to the corresponding elemental composition.

Although the greatest number of publications using this technique is concerned with light elements, and that is the reason why it is considered as a complementary analytical technique of INAA, today there are around 82 elements which can be currently determined by PGAA.

PGAA shares many characteristics with INAA: Independence of matrix and chemical state of the sample, good sensitivity for many elements, highly penetrating probe and response radiations. Also, the equipment needed for PGAA is roughly the same as for INAA, but differs on: characteristics of the neutron source, a means for introducing samples and the way to shield the HPGe detector. Commonly, in both cases the used neutron flux comes from a nuclear reactor, but in INAA the sample is inserted in an irradiation channel, while in PGAA is needed to transport cold neutrons tens of meters from the reactor to reduce the background because the gamma spectrometry is made at the same time than the irradiation. Usually the amount of sample in INAA must be very low, and in case of ceramics, it must be in form of powder. In PGAA the samples may in theory be nearly any size and shape, this is a great advantage for archaeological samples, but in practice there are often limitations about the sample size, for example, if the analysis of the entire sample is desired the sample must be smaller than the neutron beam spot. And finally, the counting of the sample in INAA can be made as a conventional gamma spectrometry with an HPGe since immediately after irradiation until few weeks later, while in PGAA the counting is made at the same time of the irradiation and for that reason the characteristics of the shielding of the detector is different than the conventional gamma spectrometry, it must include gamma and neutron shielding.

INAA and PGAA are affected equally by many sources of error: count rates must be corrected for dead time and pulse pileup, and often for spectral interference by other elements, corrections for gamma ray attenuation and neutron self-shielding may also be necessary (neutron self-shielding is mostly significant in PGAA) and the sample′s homogeneity is always a factor to take into account in both techniques. Other source of error unique in PGAA is the gamma ray background, neutron capture by detectors, shieldings and the atmosphere may give rise to gamma rays. Also, if the sample to be analyzed is hydrogenous it is often necessary to calibrate the gamma ray background. In INAA one has to pay attention in the activation of the vials (container of the sample), in our work it was found traces of Cr in the vials, so these elemental concentrations were corrected. Also, corrections for uranium composition in the sample are made, because the presence of fission fragments could over estimate the concentration of other elements like La and Cs.

Results and discussion

Elemental concentrations

The used method provides a "panoramic" study by k0-instrumental neutron activation analysis. Collected spectra were deconvoluted and fitted (see Fig. 3) using the commercial software "Hyperlab 2005 for Windows" from KFKI (Budapest, HU) [10], from here were obtained a peak table file (PTF) which contains the areas, and uncertainties, of all the gamma peaks present in the spectra calibrated in energy.

After that, the last step to obtain the elemental concentrations was performed, it was used the package "KAYZERO/SOLCOI software for windows" from DSM (Geleen, NL) [11]. First, SOLCOI software was used to compute the effective solid angles and the coincidence correction factors to each sample, comparator and validation standard. The information required to make this calculation were all the geometrical and the compositional parameters of the detector crystals, the thickness and the composition of the absorbers between the sample and the crystal and the distant between the sample and the crystal. Besides, referring to the sample, it should be introduced approximately the major material composition (in this case it is an alumina-silicate [12]) and the geometrical parameters of the sample and of the vial (polyethylene).

With the geometry well defined to each sample, comparator and validation standard the next step consisted in used "KAYZERO software for windows" to compute the

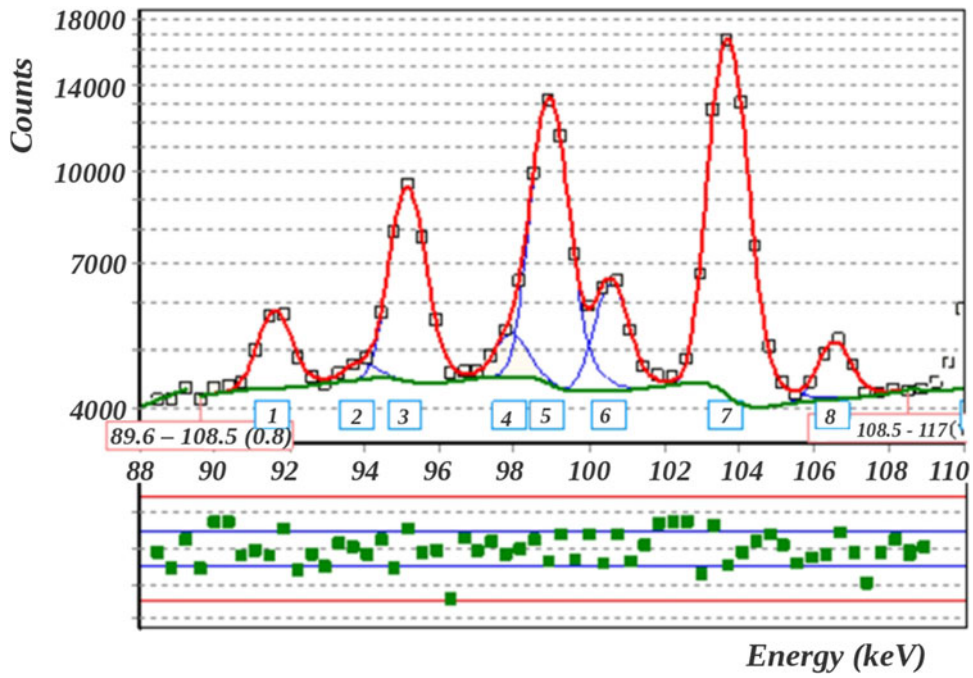

Fig. 3 Deconvoluted and fitted region of the typical spectrum with Hyperlab software

"Comparator Factor" [7] which is a factor that contains all the information about the gold comparator (specific activity, efficiency, etc.), to do this the monitor data (weight of gold into the monitor and isotope comparator, Au-198), the irradiation data (t_i, channel of irradiation depending if it was a short, S84, or a long irradiation, Y4) and the measurement data (td, tm, detector and position when it was measured, geometry and the PTF).

Finally, the elemental concentrations of the validation standards and the unknown samples were obtained by introducing into "KAYZERO" the sample data, the irradiation data (here the Comparator Factor was introduced), the measurement data of each sample and standard. Besides, corrections for reactions which produce interferences were done [13]. In total, approximately, 37 elements were analyzed in 46 samples. Also it is shown in the Fig. 4 that the obtained results are validated.

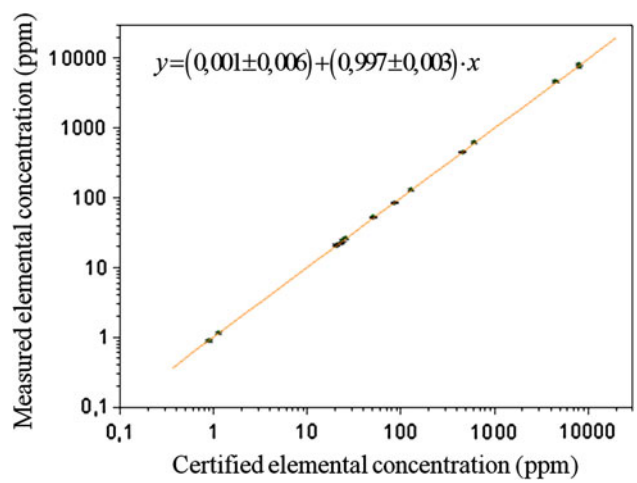

Fig. 4 Comparison between the certified values of the multi-elemental standard "SMELS III" and the elemental concentrations obtained by k0-INAA

Statistical analysis

This section discusses the statistical analysis of the elemental concentrations of the figurines excavated at two distinct locations: Los Roques Archipelago Islands and the Valencia Lake Basin. The fundamental aim of this analysis is to determine whether the data has enough evidence to support species differentiation. Four methods are presented. The methods discussed are the Fisher linear discriminant, principal component analysis, hierarchical clustering and a multivariate statistical hypothesis test for two populations means known as the Hotelling T^2 test. Some assumptions were made. First, the experimental data to be analyzed is of the required quality. Second, there exist contextual reasons to believe that there is a maximum of two species. All computations were done using the open source R statistical software [14]. This sections is structured has follows. A small description of the dataset is presented, followed by three differentiation techniques: Fisher linear discriminant (FLD), principal components analysis (PCA) and hierarchical clustering (HC). The discriminability of the datasets is also verified by the Hotelling T^2 statistical multivariate test. The section concludes presenting some overall comments.

Datasets and pre-processing

The data comprises three sets of the elemental composition of the analyzed figurines and their respective location of discovery. The characteristics of the datasets are presented in Table 1. Missing values were replaced by the detection limit of the equipment. Due to weathering some elements such as Br and Cl are not taken into account. The choice of missing value threshold is a necessary and sensible matter. A compromise between imputed and measured data is hard to define. For our case study, elements with less than 15 % of measured data were discarded.

There is a variety of standardizing techniques available in the literature, e.g. [15–17], each one has its own advantages. Here we used four commonly standardizing techniques used in statistical studies: $\hat{x} = \frac{x-\mu}{\sigma}$ (a); $\hat{x} = \frac{x-\hat{\mu}}{\hat{\mu}}$ (b); $\hat{x} = \log(x)$ (c) and $\hat{x} = \frac{x}{\hat{\mu}}$ (d). Where $\hat{x}$ and $\hat{\mu}$ are the estimated population mean and standard deviation. Subsequent studies make use of all four standardizations and study their effect on group discriminability. It is worth noting that all projection methods are invariant to translation, so standardizations (b) and (d) should give the same results.

After standardizing the sets, no outliers or atypical values were detected. Next step was the identification and handling of elements that have strong linear dependence (correlation greater than 70 %). For the F set, we found that Tb, Yb, La, Dy, Cs, Sc, Ti, V, K, Rb and Al have strong dependence. On the other hand, for set B, only Al was found to have strong relationship with Sm.

After the pre-processing, each of the F, B, C sets was reduced to (See Table 2). Table 2 shows the resulting sets after the pre-processing. Some care has to be taken when deciding if 15 or 7 are too few variables for the task of successfully classifying the two species. Figure 5 shows box-plots for the three data sets F, B and C. It can be observed that some elements can be easily associated to a species (R or V). Some hints on separability are evident (for example, see elements Mn and Cr) encouraging analysis to see whether the species "R" and "V" can be systematically discriminated successfully or not.

Fisher linear discriminant (FLD)

The Fisher linear discriminant is a very powerful classification technique introduced by Fisher in 1936 [18]. It is very closely related to linear discriminant analysis (LDA), PCA and factor analysis. The fundamental idea behind this method is to project n-dimensional data onto the "best" one-dimension component that discriminates two or more species. Details on the derivation can be found in [19, pp. 117–120]. Projecting the data onto the line described by the "best" one-dimension component does not classify the species, it is necessary to estimate a threshold. In the case where it can be assumed that the conditional densities are multivariate normals with equal covariance matrices, the optimal decision boundary involves only the projection direction and the prior probabilities [19]. Fortunately, there are other approaches that do not require any assumptions on the conditional densities in order to estimate best threshold constants. For our case study, we used a simple binary classification. This method moves on the projected data from one extreme to the other while it accounts the number of incorrect classified individuals and has at least one global minimum. As a result of the classification there can be more than one optimal threshold. In such cases, the average is taken to be the final threshold. The main

Table 1 Characteristics of the datasets

Set name	Sample size	Elements of set
F	46: 28 R + 18 V	37: Al, As, Au, Ba, Br, Ca, Ce, Cl, Co, Cr, Cs, Dy, Eu, Fe, Ga, Hf, K, La, Mg, Mn, Na, Nd, Rb, Sb, Sc, Sm, Sr, Ta, Tb, Th, Ti, U, V, W, Yb, Zn, Zr.
B	40: 21 R + 19 V	12: Al, Ca, Cr, Fe, K, Mg, Mn, Na, Sc, Sm, Ti, V.
C = F ∪ B	Common elements from F and B 86: 49 R + 37 V	7: Cr, Fe, K, Mn, Na, Sc, Sm.

Table 2 Characteristics of the datasets after pre-processing

Set name	Sample size	Elements of set
F	46: 28 R + 18 V	15/24: As, Ba, Ce, Co, Cr, Eu, Fe, Ga, Hf, Mn, Na, Sm, Ta, Th, Zn.
B	40: 21 R + 19 V	7/12: Ca, Fe, K, Mg, Na, Sm, Ti.
C = F ∪ B	Common elements from F and B 86: 49 R + 37 V	6/7: Cr, Fe, K, Mn, Na, Sm.

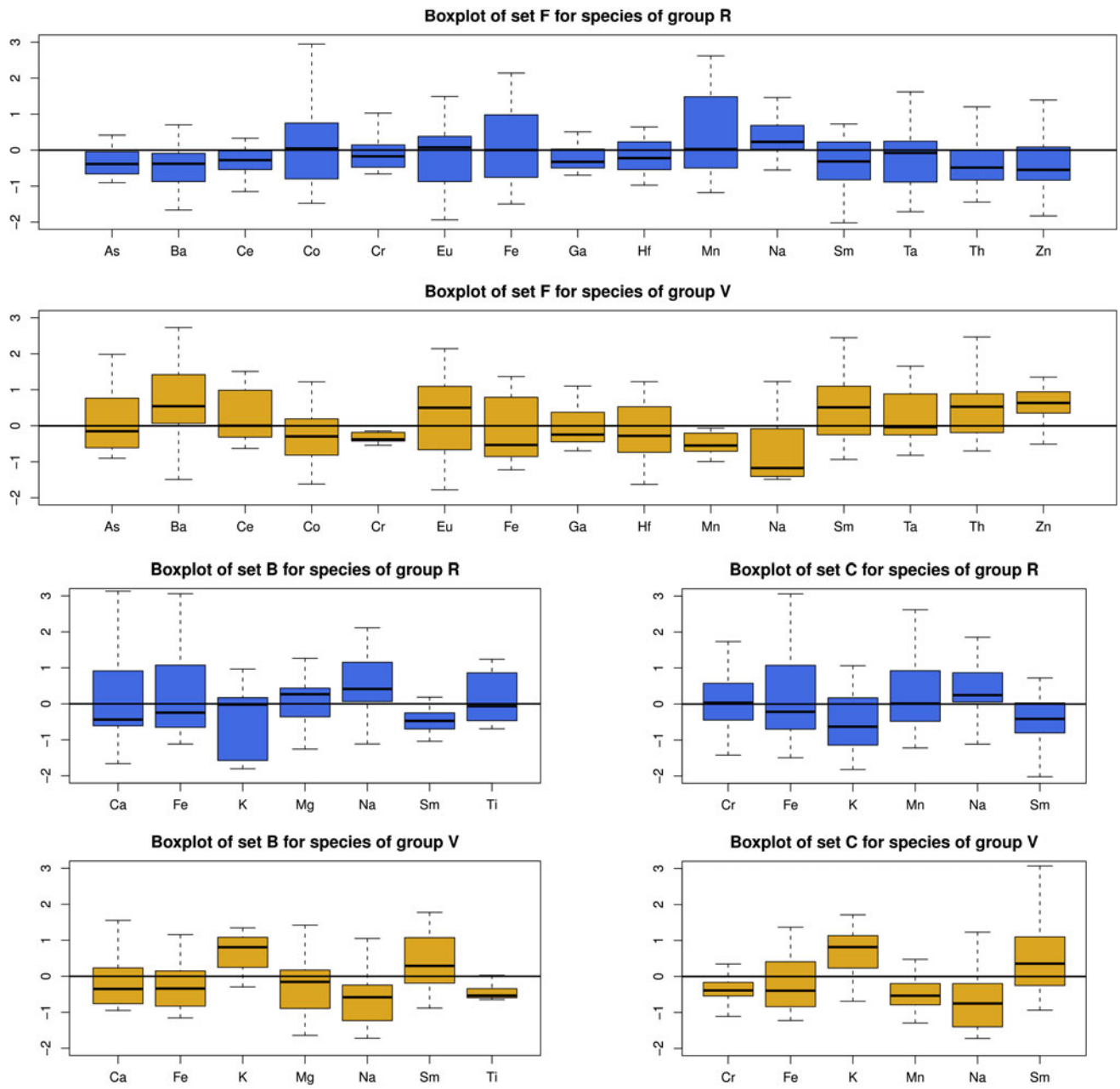

Fig. 5 Counter clockwise from *top*: Boxplots for the F, B and C data sets. Some hints on separability can be observed

difference between FLD and LDA is that it approaches the classification problem by assuming that the conditional probability density functions are normally distributed. This subject is thoroughly studied in [19, chapter 5] and [18, chapter 12]. The R statistical software has readily available LDA support through the MASS package. Figure 6 shows the distinctness of the two species for the three data sets; here from the B data set can clearly be seen that there exist more than two optimal thresholds. The histograms show the estimation of the underlying conditional densities and below them a smoothed version. Some overlapping is observable. The thresholds are shown with black vertical lines. Prior conditional density functions are also shown. To understand the effect of standardization, Fig. 7 shows the conditional densities estimated of the species R and V for set C. From these figures it can be observed that, in general terms all standardizations have the same behavior. For some cases density traits are more readily visible than others (logarithmic standardization). To see how they affect classification efficiency, Table 3 shows the results of applying FLD, PCA and HC to all the available data. Data set F is separable while the sets B and C present some

overlap, at least in the projected component. Results from Table 3 suggest separability of the two species (R and V). Given the nature of mapping many-to-one dimension, these results seem very attractive. As expected, the (b) and (d) standardizations perform identically. Logarithmic transformation on the other hand can have somewhat distinct output. To evaluate the classifier robustness, cross-validation was used.

Robustness analysis

Cross-validation is a very common technique to evaluate the quality of a classifier and it has many variants. Cross-validation is most useful when multiple trials are performed, having the effect of reducing variability, among other benefits. The leave-one-out cross-validation method (LOOCV) consists of using one individual for validation and the rest is used as the training set. This is repeated once for each individual of the data set. When computationally possible, redundant trials (choosing randomly the validation individual) can be performed providing a more realistic bound of the robustness. In our case study, LOOCV with 10.000 trials for each set was performed. Table 3 reports the computational results.

The results report that eight out of ten individuals are correctly identified. Considering the upper bounds reported in Table 3 and the small data sets, these results seem to indicate that the FLD could be a robust classifier.

The results using FLD not only showed that simple mapping many-to-one dimensions can be very powerful but it is also a computationally cheap. A key aspect on FLD (or LDA) is that it seeks directions that are efficient for discrimination while PCA seeks those that are efficient for representation [19]. In the next section we discuss how PCA was used to classify the R and V species.

Principal components analysis (PCA)

Principal components analysis [20] in its simplest form can be interpreted as a projection method that seeks directions of maximal variability. A formal introduction to this method can be found in [17, 21, 22]. This method is commonly used in exploratory data analysis and predictive models as well as a tool for data dimensionality reduction.

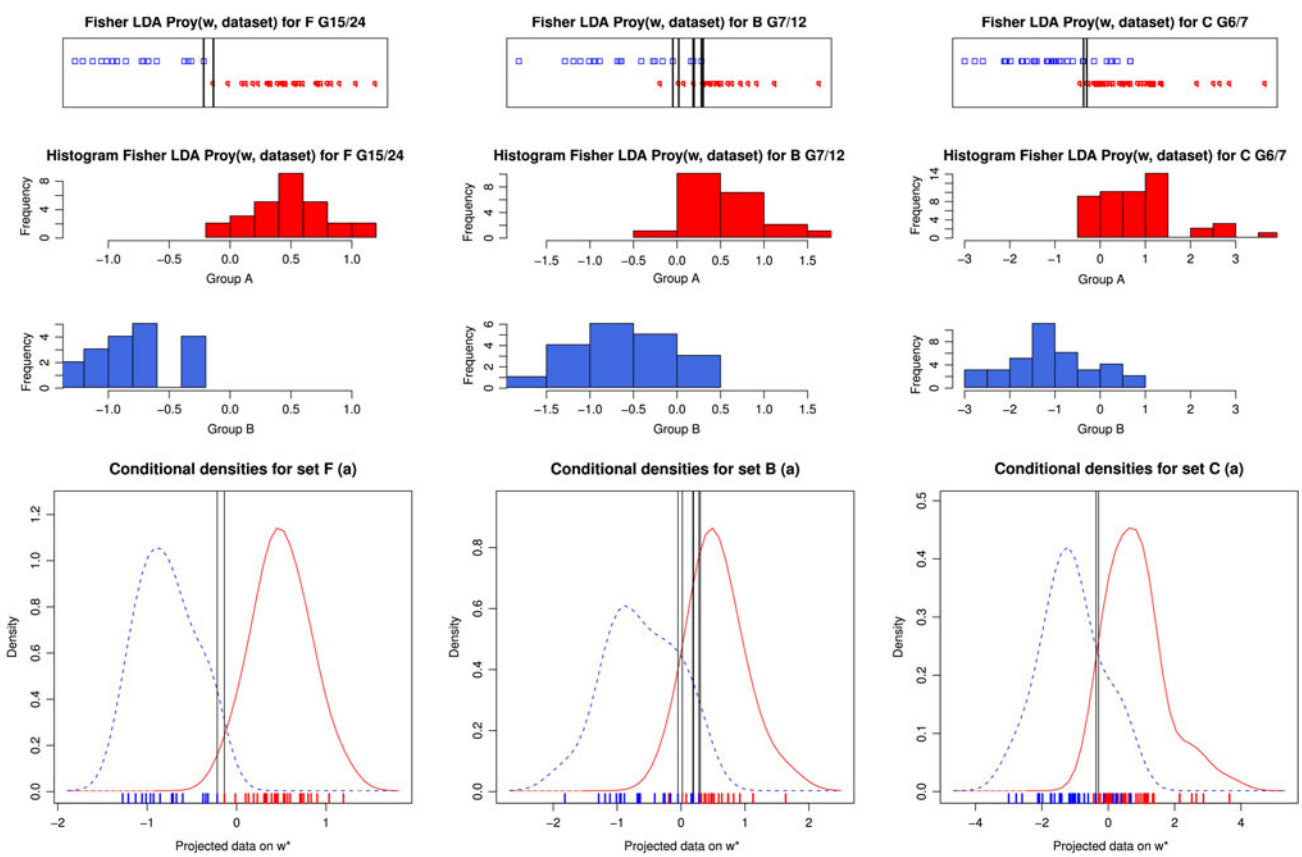

Fig. 6 From *left* to *right*, classification of the two species using Fisher linear discriminant for the F, B and C sets. The *red color* or *lower circles* represent the R species while the *blue color* or *top circles* represent the V species. The *top* section of each sub-figure represent the projected data on w^*. (Color figure online)

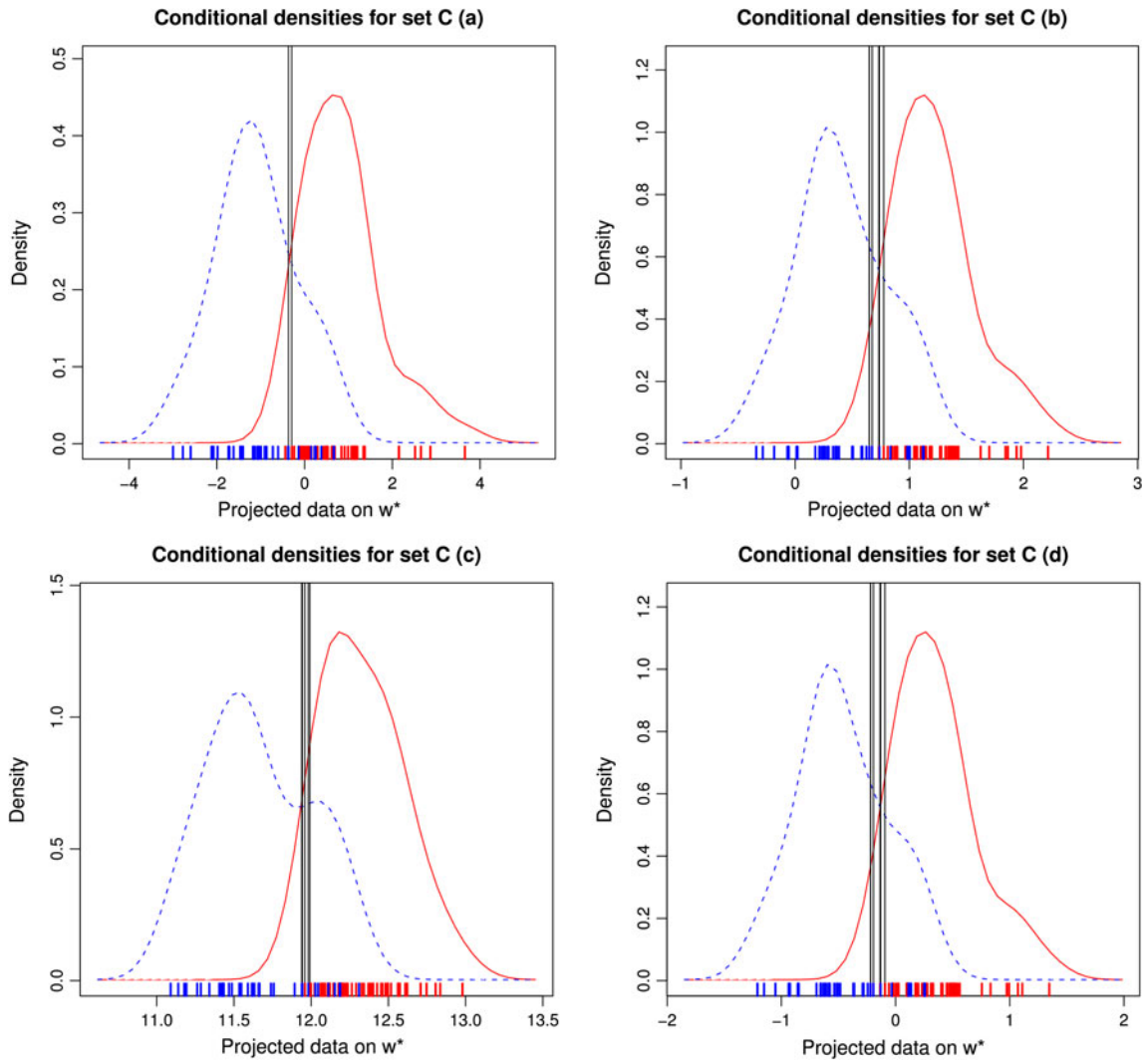

Fig. 7 Conditional densities estimate of species R and V for set C using the proposed standardizations. From *top left* to *bottom right*: normalizing (**a**), standardizing by mean (**b**) and (**d**) and logarithmic (**c**)

For our needs of discrimination we use PCA as an essentially geometric method for species classification. The idea is to project the data onto the two or three most significant components and geometrically identify any kind of patterns that differentiate the two species.

Given the sets F, B and C, Table 4 shows the proportion of variance each component explains with the logarithmic standardization. For our data sets, this transformation provided higher weights for the most significant components.

Using only the first two components, depending on the set, from 60 to 80 % of the variability is taken into account. Figure 8 shows scatter plots of these two components as well as the optimal linear discriminant bound constructed applying the same binary classification method used in FLD. Table 3 reports the classification results.

The results reported for PCA are similar to those for FLD, accentuating the differentiability of the two species. Since PCA is invariant to translation, standardizations (b) and (d) are equivalent. All standardizations seem to perform similarly. Figure 8 shows that despite the slight overlapping, differentiation can be achieved. Given the nature of PCA to isolate hidden dependencies and capture the behavior of each species, it allows a straightforward classification. In our study, PCA has similar discrimination efficiency to FLD. R statistical software [14] has readily available PCA and a wide variety of companion tools [17]. In the following section we present an effort to discriminate the two species using hierarchical clustering techniques.

Hierarchical clustering

Hierarchical clustering belongs to cluster analysis which is an important step of data mining. It is a common technique to hierarchically group, tag, organize and classify individuals. A

Table 3 Results of using all the data available in the construction of the FLD, PCA and CA classifiers

Fisher linear discriminant

Data set	Standardization technique	Classification efficiency (%)	Data set	Standardization technique	Classification efficiency (%)	Data set	Standardization technique	Classification efficiency (%)
F	(a)	100	B	(a)	90	C	(a)	90
	(b)	100		(b)	90		(b)	90
	(c)	100		(c)	93		(c)	88
	(d)	100		(d)	90		(d)	90

Leave-one-out cross-validation using 10.000 trials

Data set	Standardization technique	Classification efficiency LOOCV 10 K (%)	Data set	Standardization technique	Classification efficiency LOOCV 10 K (%)	Data set	Standardization technique	Classification efficiency LOOCV 10 K (%)
F	(a)	89	B	(a)	80	C	(a)	83
	(b)	90		(b)	80		(b)	85
	(c)	88		(c)	83		(c)	83
	(d)	90		(d)	80		(d)	86

Principal components analysis

Data set	Standardization technique	Classification efficiency PCA (%)	Data set	Standardization technique	Classification efficiency PCA (%)	Data set	Standardization technique	Classification efficiency PCA (%)
F	(a)	83	B	(a)	90	C	(a)	86
	(b)	85		(b)	88		(b)	88
	(c)	87		(c)	85		(c)	88
	(d)	(b)		(d)	(b)		(d)	(b)

Cluster analysis

Data set	Standardization technique (cluster quantity)	Classification efficiency (%)	Data set	Standardization technique (cluster quantity)	Classification efficiency (%)	Data set	Standardization technique (cluster quantity)	Classification efficiency (%)
F	(a) [7]	72	B	(a) [10]	73	C	(a) [14]	77
	(b) [7]	74		(b) [10]	70		(b) [14]	81
	(c) [7]	80		(c) [10]	88		(c) [14]	80
	(d) [7]	(b)		(d) [10]	(b)		(d) [14]	(b)

Table 4 Proportion of variance of each component

Components		1	2	3	4	5	6	7	8
Proportion of variance	F	0.40	0.61	0.72	0.79	0.84	0.88	0.90	0.93
	B	0.53	0.78	0.87	0.94	0.97	0.99	1.00	
	C	0.54	0.74	0.85	0.95	0.98	1.00		

sound introduction to unsupervised learning and clustering are available in [23, 24] and [19, chapter 10]. Venables and Ripley in [17] do not assume that clustering methods are the best way to discover interesting groupings. Alternatively, they suggest more visual methods. Hierarchical clustering requires determine two parameters: distance metric and agglomeration algorithm. In the area of application, many works [16, 25, 26] suggest the use of the Mahalanobis (or some variant) metric while others (e.g. [15]) the use of Euclidean distance or its square. There is no de facto metric that always leads to interesting groupings. A problem associated to clustering is that different metrics and agglomeration algorithms generally produce quite different groupings. For our case study, a range of metrics were tested, contending favorably in all cases the Minkowski metric that is a generalization of the Euclidean and Manhattan metrics. Given that this metric is a generalization of the Euclidean, it should emphasize the effect of those variables that have the major contribution to variability. During the experimentation, values of $k = 2$, 3 and 4, gave good classification results. Mahalanobis metric was also tested giving poor results.

For the agglomeration algorithm, literature of interest [15] suggest using a centroid agglomerative method. Given

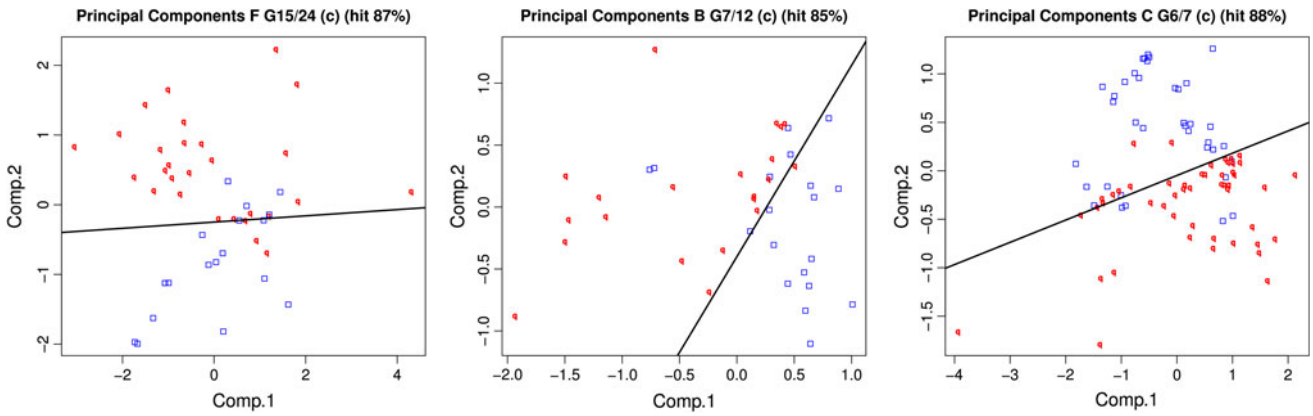

Fig. 8 Scatter plots of the first two most significant PCA components and an optimal linear threshold. *Red circles* indicate the individual of the R species while the *blue squares* represent the V species. (Color figure online)

the nature of the R and V species of having a sensible overlapping, an agglomeration algorithm that formed compact groups was necessary. The single link (nearest neighbor) algorithm would be the preferred choice. Unfortunately, the groupings constructed by this procedure have a tendency to be oversensitive to extremes. The average agglomerative algorithm ameliorates this behavior [19] and it was chosen in this case. A list of efficient agglomerative algorithms for hierarchical clustering can be found in [27].

The hierarchy of clusters is obtained by cutting the tree at different heights. This has the additional problem of estimating the right height to prune the tree. The chosen heights were experimentally established but there exist a variety of methods providing optimal partitions: K-means [28], K-medoids [29], etc. Clusters with mixed species can be tagged in a variety of manners. An attractive choice is to tag the mixed group by population census, assigning the same species tag as of the majority of its individuals.

The classification results obtained are reported in Table 3; Figs. 9, 10 and 11. In Table 3 the classification efficiency refers to the amount of well classified individuals, the mapping is performed by population census. For comparison purposes a K-means grouping was also performed and it is shown in the same figures (bottom right scatter plots). For the K-means clustering method, the Forgy algorithm [30] was used with Euclidean metric. The centers were carefully hand chosen.

Pruning the tree and tagging each individual with the group's number or color reveals the structure of each species in a much more useful way (bottom left scatter plots). These plots not only show the hierarchical groups but also the specie to which each individual belongs to. This permits visually to identify mixed groups and single individual clusters.

The choice of the number of clusters to build is far from easy. Fortunately, contextual archaeological information indicates that the possible number of species is one or two. So the aim was not to have a large quantity of clusters. On the other hand, in the overlapping space where classification is difficult, small homogeneous groups (groups having all individuals of the same species) would be desired. A convenient number that allows small homogeneous groups and avoids having many clusters of single individuals is not easy indeed. For the data set F, considered as well-behaved, a number of 7 clusters attained good results. Under that configuration (see Fig. 9), there were two main groups, one mixed, two subgroups and only two clusters of single individuals.

For sets B and C correct grouping was a harder task and required incrementing the cluster quantities to 10 and 14 respectively. Under these configurations, Grouping of set B resulted in (see Fig. 10) four main groups, one mixed, one subgroup and four single individual clusters. Finally for set C, the groupings were (see Fig. 11) nine groups, three mixed clusters and only two clusters of single individuals. Results from Table 3 indicate that differentiation is plausible even though classification efficiencies are somewhat lower that those reported for FLD and PCA. It is also evident that standardization is of importance. For all three data sets, transformation by logarithm gave the best species differentiation. As expected, standardizations (b) and (d) are equivalent. The results of two species classification using this technique are comparable with those obtained using PCA and FLD, adding to the evidence in favor that the two species indeed could be different. All clustering tools used are available in the R statistical software.

All the previous methods do not make assumptions about the multivariate probability distribution of data. We conclude the discrimination attempts of the two species by presenting a multivariate test of hypothesis.

Hotelling T^2 test

The Hotelling T^2 test [31] is a powerful multivariate statistic used to test if two populations have the same mean or

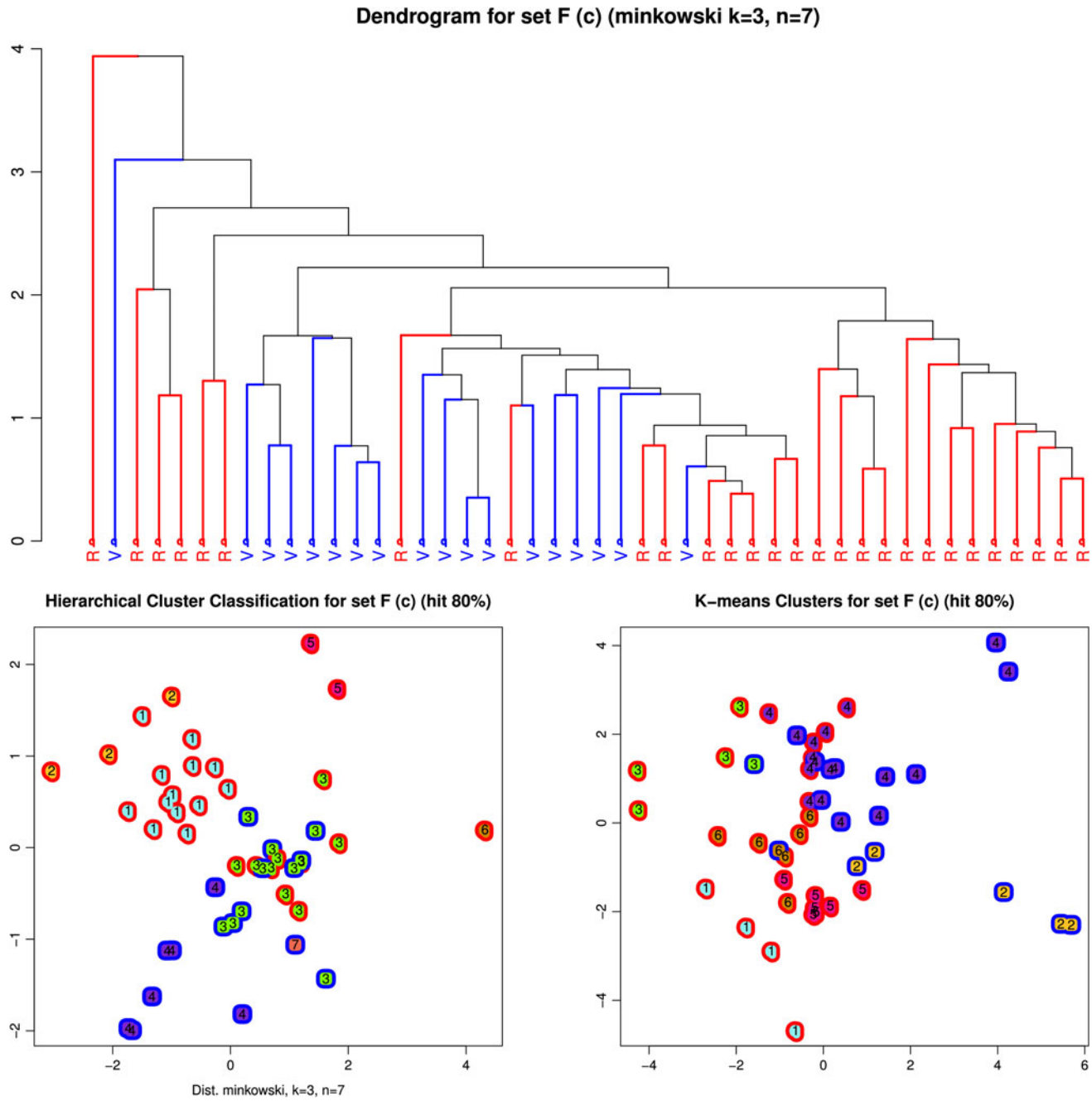

Fig. 9 Cluster results for the set F. *Bottom left scatter plot* shows the identified groups. *Outer color ring* represents the species (*red* for R and *blue* for V). *Inner color* represents the groups to which each individual is assigned to. *Bottom right scatter plot* show the result of grouping using the K-means algorithms. (Color figure online)

centroid. The Hotelling T^2 test is the multivariate generalization of Student's t test and quantifies the differences between the centroids of two groups using the unbiased pooled covariance matrix estimate. It is coupled with a hypothesis test [32, chapter 8] where the null hypothesis is that centroids do not differ between the two groups.

The main underlying assumption of this statistic is that both populations have multivariate Normal distributions. Generally, to prove a population has multivariate Normal distribution is seldom easy.

Before the above hypothesis test can be performed, each species distribution must be tested for Normality. This is performed via the multivariate version of the Shapiro–Wilk test [33] proposed by Villasenor-Alva and Gonzalez-Estrada in [34].

In order to check if both species have multivariate Normal distribution, the multivariate version of Shapiro–

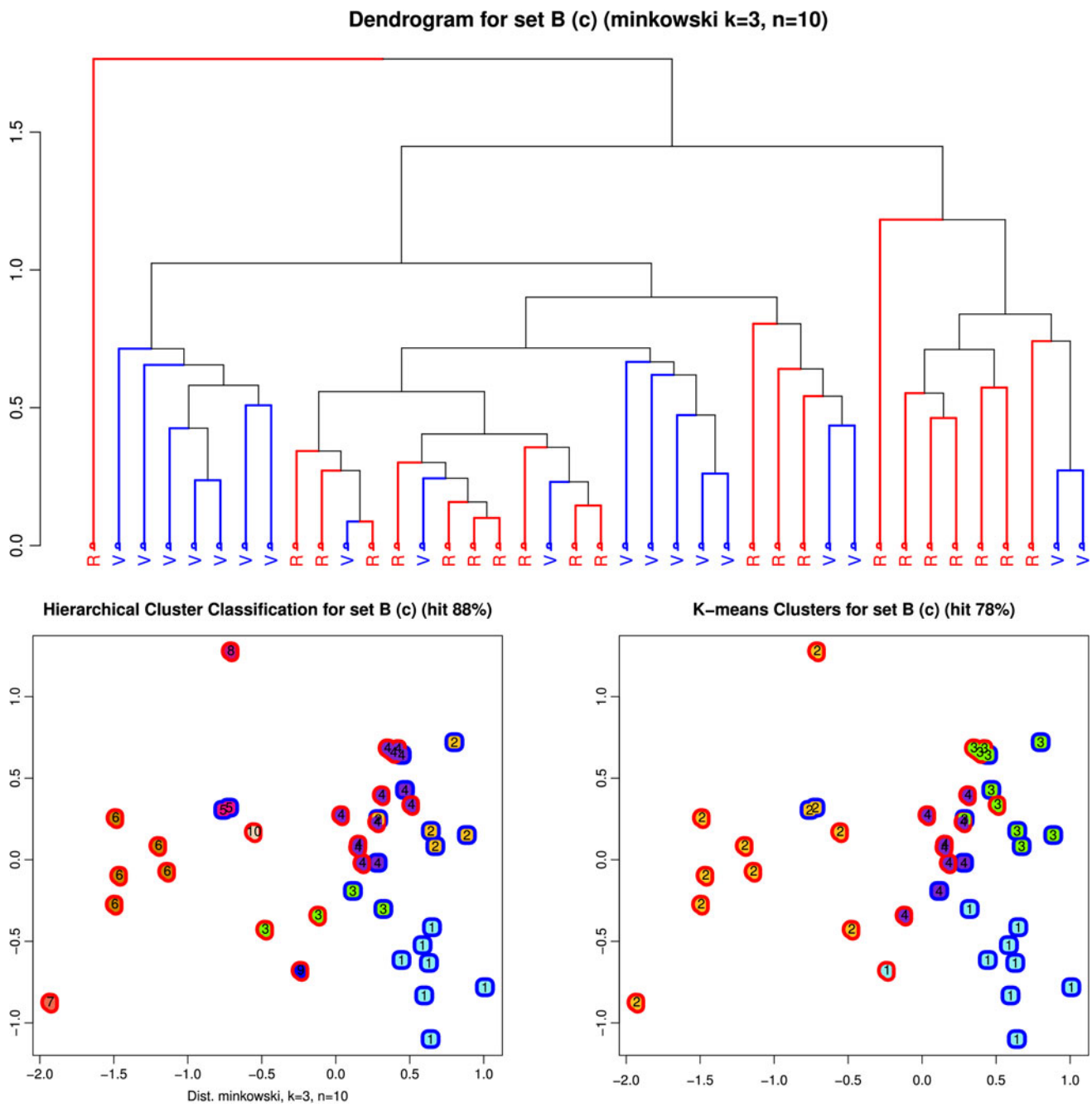

Fig. 10 Cluster results for the set B. Same legend than Fig. 9

Wilk test is applied not directly to the two groups R and V defined previously but to a set of principal components that describe at least 90 % of the variability. This technique is not uncommon and helps further remove hidden linear dependencies.

The sets were standardized before applying PCA. A significance level of 1 % was chosen. Test results are reported in Table 5 and are for each species in each set. Values for the normalized data are rather low but not unexpected. The multivariate data behavior is very sensible to the standardization method used. Standardization (a) is the only method that didn't reject the null hypothesis. For all other transformations, it was rejected at least once. Given the normalized data sets, it is now possible to test for equality of the centroids.

The results obtained use the same significance level and are reported in Table 6. The data sets show enough evidence to reject the hypothesis that the two species have the same centroids. In other words, there is not enough evidence to sustain the fact that the sets contain only one

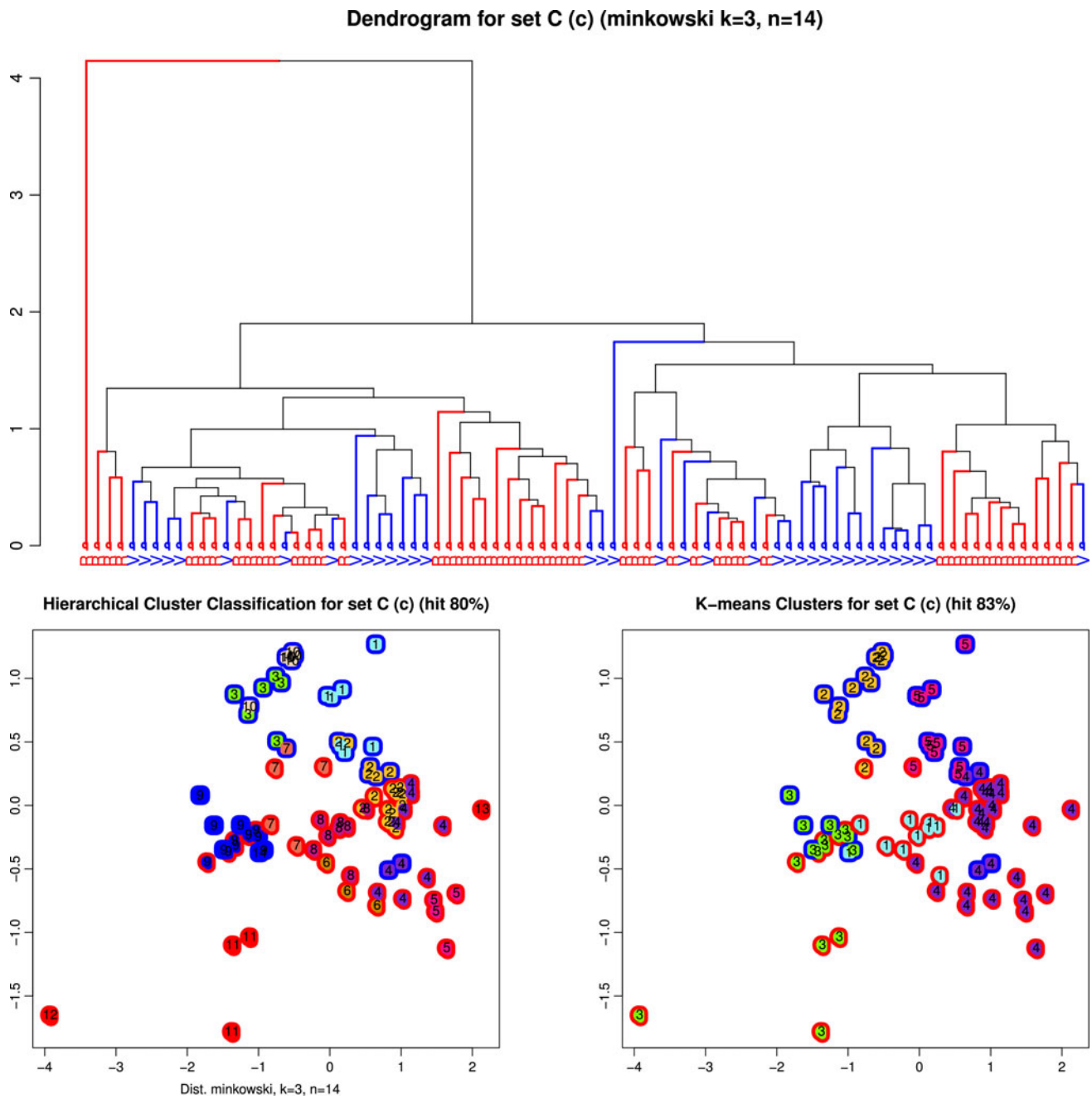

Fig. 11 Cluster results for the set C. Same legend than Fig. 9

Table 5 Results of multivariate normality test for the various standardizations

Data set	Specie and components count	*p* value	Data set	Specie and components count	*p* value	Data set	Specie and components count	*p* value
F (a)	R, 8	0.03	B (a)	R, 4	0.01	C (a)	R, 5	0.02
	V, 8	0.09		V, 4	0.92		V, 5	0.03
F (b)	R, 8	*	B (b)	R, 4	0.01	C (b)	R, 5	*
	V, 8	0.91		V, 4	0.23		V, 5	0.03
F (c)	R, 8	0.56	B (c)	R, 4	0.02	C (c)	R, 5	*
	V, 8	0.92		V, 4	0.43		V, 5	0.12

* The corresponding value is negligible (approx. 0)

Table 6 Result of applying Hotelling T^2 test to the normalized sets

Set	Test statistic T^2 (n, d)	H_0 rejected?	p value
F	11.79 (8, 37)	Yes	4e-08
B	9.21 (4, 35)	Yes	3e-05
C	17.53 (5, 80)	Yes	1e-11

$\alpha = 1$ %. All three sets reject the null hypothesis showing enough evidence that the two species can have different centroids

species, providing evidence supporting species differentiation. This statistical result corroborates the findings of the previous methods. The R statistical software's 'Hotelling' and 'mvShapiroTest' libraries were used to perform the test describe herein.

Conclusions

The k0-instrumental neutron activation analysis is a nuclear technique very suitable for the elemental compositional analysis of archaeological samples. A very low amount of sample ($\sim$100 mg) and no big pre-processing was required to make a panoramic study (37 elemental concentrations were determined with uncertainties between 3 and 20 %) of each examined figurine. The use of the Hyperlab y Kayzero/Solcoi softwares was determining and makes all the spectra analysis very comfortable and easy.

Over the literature is found that there are only few criteria to choose the elements used to perform the statistical analysis in provenance analysis. In this work we present a procedure based in physical and statistical arguments to get the minimum number of elements that should be used to obtain reliable results. We found that there are some elements which are linearly correlated, in statistical terms they contain the same information and to include all of them in the analysis introduce "noise" in the results.

Regarding the statistical analysis we used: the Fisher linear discriminant method which is very suitable for species differentiation. It is a simple and powerful discrimination technique. The threshold estimation problem can be solved even without assuming anything on the conditional densities; principal components approach is a well known method and easy to use for species differentiation. The main virtue of this method is dimensionality reduction; hierarchical clustering techniques are widely known to be very sensitive to metric and agglomeration algorithms of which exist a great variety, we found favorable results using the Minkowski metric and the average agglomeration algorithm. We found that in this method experimentation is fundamental, dendrogram visual interpretation might be cumbersome and choosing the right number of clusters is difficult; the Hotelling T^2 test presented relies on data belonging to multivariate Normal distributions, it is paramount test the groups for normality. From all statistical techniques we found results very similar which support the species differentiation. Also these results present the possibility to used only one or two of these methods to perform similar studies. All this analysis shows that the raw material used to make the figurines found in Los Roques Archipelago is not the same used to make the continental figurines despite the morphological similarity of the figurines. However, increase the number of analyzed specimens to get more reliable results is recommended. Also, it is planned to make XRD analysis to introduce a new variable in the statistical analysis and also to establish the manufacturing techniques.

It would be of interest to try other unsupervised as well as supervised learning methods, specially LDA, Kn-nearest neighbor estimation, support vector machines and neural network techniques.

Acknowledgments We wish to thank the Department of International Relationships of the Simón Bolívar University for the financial support.

References

1. Mackowiak de Anctzak MM, Anctzak A (2006) "Los Ídolos de las Islas Prometidas Arqueología Pre-Hispánica del Archipiélago de los Roques", Editorial Equinoccio, 1st edition, Venezuela
2. Sajo-Bohus L, Mackowiak de Antczak MM, Greaves E, Antczak A, Bermudez J, Kasztovszky Zs, Poirier T, Simonits A (2005) JRNC 265(2):247–256
3. Mackowiak de Anctzak, Kasztovszky Zs, Révay Zs, Anctzak A, Molnár GL, Sajó-Bohus L, Greaves ED (2002) Proceedings of the international radiation education symposium, pp 71–78
4. Barros H, Sajo-Bohus L, Kasztovszky Zs (2006) 11th International conference on nuclear reaction mechanisms, Villa Monastero, Varenna, 12–16 June 2006
5. Travesi A (1975) "Análisis por Activación Neutrónica: Teoría, práctica y aplicaciones". Ediciones J.E.N, Madrid, 1era Edición
6. De Corte F (2003) k0-NAA past the turn of the century: problems, concepts, insights, prospects. Czechoslov J Phys 53(Suppl. A):A161
7. De Corte F (1987) "The k0-standardization method: a move to the optimization of neutron activation analysis". Ph.D. thesis, University of Gent, Belgium
8. Vermaercke P, Robouch P, Eguskiza M, De Corte F, Kennedy G, Smodis B, Jacimovic R, Yonezawa C, Matsue H, Lin X, Blaauw M, Kucera J (2006) Nucl Instrum Methods Phys Res A 564:675–682
9. De Wispelaerea A, De Cortea F, Bossusb DAW, Swagtenb JJMG, Vermaercke P (2006) Nucl Instrum Methods Phys Res A 564:636–640
10. Simonits A, Östör J, Kálvin S, Fazekas B (2003) J Radioanal Nucl Chem 257:589–595
11. Jacimovic R, Smodiš B, Bucar T, Stegnar P (2003) J Radioanal Nucl Chem 257(3):659–663
12. Read HH (1970) Rutley's elements of mineralogy, 26th edn. Thomas Murby and CO, London
13. Xilei L, Van Renterghem D, De Corte F, Cornelis R (1989) J Radioanal Nucl Chem 133:153–165
14. R Core Team (2012) R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna,

Austria, 2012, ISBN 3-900051-07-0, http://www.R-project.org. Accessed Sep 2012
15. Buxeda i Garrigós J (1999) Alteration and contamination of archaeological ceramics: the perturbation problem. J Archeol Sci 26:295–313
16. Vitali V, Franklin UM (1986) New approaches to the characterization and classification of ceramics on the basis of their elemental composition. J Archaeol Sci 13:161–170
17. Venables WN, Ripley BD (2002) Modern applied statistics with S. Springer, New York
18. Fisher RA (1936) The use of multiple measurements in taxonomic problems. Ann Eugenics 7(2):179–188. doi:10.1111/j.1469-1809.1936.tb02137.x
19. Duda RO, Hart PE, Stork DH (2000) Pattern classification, 2nd edn. Wiley Interscience, New York. ISBN 0-471-05669-3
20. Pearson K (1901) On lines and planes of closest fit to systems of points in space. Philos Mag 2:559–572. http://pbil.univ-lyon1.fr/R/pearson1901.pdf
21. Jolliffe IT (2002) Principal components analysis (PCA) series: Springer series in statistics, 2nd edn. Springer, New York
22. Bourlard Herve, Kamp Yves (1988) Auto-association by multilayer perceptrons and singular value decomposition. Biol Cybern 59:291–294
23. Gordon AD (1999) Classification, 2nd edn. Chapman & Hall/CRC, London
24. Kaufman Leonard, Rousseeuw PeterJ (1990) Finding groups in data: an introduction to cluster analysis. Wiley, New York
25. Glascock MD, Neff H (2003) Neutron activation analysis and provenance research in archaeology. Meas Sci Technol 14:1516–1526
26. Hein A, Mommsen H (1999) Element concentration distributions and most discriminating elements for provenancing by neutron activation analyses of ceramics from Bronze Age sites in Greece. J Archaeol Sci 26:1053–1058
27. Day WHE, Edelsbrunner H (1993) Efficient algorithms for agglomerative hierarchical clustering methods. J Classif 1(1):51–74
28. Lloyd SP (1982) Least squares quantization in PCM. IEEE Trans Inf Theory 28(2):129–137
29. Vinod H (1969) Integer programming and the theory of grouping. J Am Stat Assoc 64:506–517
30. Forgy E (1965) Cluster analysis of multivariate data: efficiency vs. interpretability of classifications. Biometrics 21:768
31. Hotelling H (1931) The generalization of student's ratio. Ann Math Stat 2(3):360–378. doi:10.1214/aoms/1177732979
32. Casella G, Berger RL (1990) Statistical inference. Duxbury Press, Pacific Grove. ISBN 0-534-11958-1
33. Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). Biometrika 52(3–4):591–611
34. Villasenor-Alva JA, Gonzalez-Estrada E (2009) A generalization of Shapiro–Wilk's test for multivariate normality. Commun Stat: Theory Methods 38(11):1870–1883