Cluster Analysis of Greek Pottery from Carthage

Herbert Pike, Archaeology Department,
University of Reading.

Introduction

The title is slightly misleading. In fact I was asked to analyse, using neutron activation, 39 sherds brought back by my colleague Dr. M. Fulford, from the British excavations on the Ilôt d'Amirante, Carthage. Of these 39, one was of red-figure and one of late Roman African red-slipped ware, the rest of black glaze (slip) ware. Such pottery is difficult to sort into fabric groups because of the fineness of the paste and surface treatment. All the material came from stratified contexts dating from about 400 BC to probably late 1st century BC. In addition two bases and a rim supplied by the Nancy Ure Museum at Reading and assigned as Attic Ware, were also analysed.

The samples were prepared at Reading but irradiation and counting of three weekly batches were carried out at the London University Reactor Centre at Sunninghill, and I am very grateful to the Director and staff for providing facilities and guidance. The numerical analysis at Reading made use of computer programs and advice on their use supplied by the Applied Statistics Department.

Duplicate samples were taken from 12 sherds and triplicate from 4 others; in addition 5 samples were counted twice, giving 67 sets of measurements for 19 gamma ray peaks for 12 elements. Three of the 19 were discarded as redundant and for four elements the two counts were added together, giving just one count for each element.

The 4000 channel LINK counting system was used, each channel being 1 keV wide. For 25 samples this printed out the counts for about 13 channels on each side of the peak, so that graphical estimates of background could be made - this was necessary because counting took place 5½ days after irradiation ended. For the remaining samples only integrals were printed out and percentage corrections were applied to these, derived from the graphs.

In discriminating between samples a given percentage change in count for any element was assumed to be equally important. Hence the counts were replaced by their logs.

Corrections were made for variations in irradiation time from week to week, and for variation in time at which counting commenced - these varied from element to element. Corrections for variations in sample mass and neutron flux could have been made but were vitiated by small changes in counting geometry due to movement of the counter from week to week. The best available estimate of the product of mass, flux and counter efficiency was taken to be the geometric mean of the twelve counts for each sample i.e. the mean of the 12 logs was subtracted from each.

The 12 elements counted were Sm, Ce, La, Pa, Yb, Cr, Co, Sc, Lu, As, Fe and Ti (counted as Sc 47).

Numerical Analysis

The first step was to look at the correlations between our twelve variables to see how many really were independent. For example one would expect the three lanthanides to be highly correlated. The results were astonishing; of the 36 correlation coefficients between 9 of the elements, 30 exceeded 0.8 and 17 exceeded 0.9 in magnitude. Only Ti was completely uncorrelated.

The reason for these results can be seen from Figure 1, which shows a plot on equal scales, of the first two principal components for our 67 points. The point numbers i.e. sherd numbers are arbitrary except that multiple samples from the same sherd are indicated by letters. The first principal component has separated the points into 3 distinct groups, the second principal component has merely spread the middle group out into a long chain.

The left hand group which we call Group B contains 37 points very tightly clustered together. I therefore hypothesised that these sherds are all made from a single clay, the scatter being purely random. If this be so then we can use some measure of the spread of Group B to test whether the small Group C on the right hand side is a single group and to divide the chain, Group A into groups of acceptable size.

Detailed study of Group B

First we make estimates of the errors due to various causes. From the 5 repeated counts on the same samples we can estimate the errors inherent in counting and correcting for background. From the duplicate and triplicate samples from the same sherd we can estimate the within sherd variability and from Group B we can estimate the overall variability for pots made from the same clay.

For the third purpose we must only have one set of numbers per sherd so duplicate measurements were averaged reducing Group B to 24 sherds, Group A to 14 and Group C to 3. This omits sherd 42 which will be discussed later. The results are given in Table 1, expressed as standard deviations of the differences between the logs for each element. The first column based on only 5 sets of results is erratic but on average the s.d.s. are about half those in the second column. The entries in the third column based in this case on differences from the mean multiplied by 1.414, are substantially greater for most elements especially Ti and As so we conclude that the bulk of the variance stems from pot to pot variations.

Next let us look at the differences between samples in Group B. Since all 12 variables are equally important we must standardise them i.e. from each variable subtract its mean value (averaged over the 24 samples) and then divide by its s.d. (also calculated over the 24 samples). Using our standardised variables as coordinates we can then calculate the Euclidean distance D between any two samples and look at the distribution of D values. For 12 normally distributed variables the fraction F of the values of D exceeding some value K should be given by

$$F = (1 + y + y^2/2 + y^3/6 + y^4/24 + y^5/120) \quad \exp(-y)$$

where $y=D^2/4$, and for 10 variables the same formula with the y^5 term omitted. Since we have 11 d. of f. i.e. effectively only 11 variables our values of D should lie between the two curves.

Figure 2 shows the two curves while the circles are calculated from the 276 values of D and the crosses from the 24 distances R from the centroid, multiplied by 1.414. The fit is disappointing, there being too many large values of D and R.

The reason for the bad fit is that our variables are partially correlated, even for Group B alone. In fact 11 of the 66 correlation coefficients exceed 0.5 in magnitude, the largest being -0.7.

To get some uncorrelated variables we used ASF2 to produce principal components from the correlation matrix and to calculate transformed variables as scores on the principal components. Unfortunately these new variables are not of equal importance, the first 4 accounting for 28, 22, 20 and 10 percent of the total variance, the remaining 7 for much less. Since the new variables had to be standardised the evidence would have been badly distorted by dividing the remaining variables by their small s.d.s. Hence only the first four of the new variables, accounting for 80% of the total variance, were standardised and used.

Since we now have only 4 d. of f. the formula reduces to

$$F = (1 + y) \exp(-y)$$

Figure 3 now shows excellent agreement between measurement and theory. The crosses are derived from the 24 values of R and the circles from the 276 values of D. We see that there is only a very small probability of any value of D exceeding 6.

We conclude that Group B is a single cluster of sherds all made from the same clay and only showing random variations from the mean.

Applications to the other groups

Since we feel sure that Group A consists of more than one group, it would be futile to find its principal components. The best we can do is to assume that each clay used for Group A was of different average composition from that used for Group B but of the same quality i.e. each giving the same spread of results. The most suitable variables we can use to analyse Group A are obtained by the same transformation as was used to find the four variables for Group B.

The variables for Group A were standardised, transformed using the same coefficients as for Group B and then the 4 new coordinates restandardised.

All 91 interpoint distances for Group A were then calculated and the points subdivided into groups, inside each of which no interpoint distance was much greater than 6. Sherds 1 and 8 could not be fitted either together or into any group but the remaining twelve sherds were fitted into two groups of 6. The interpoint distances are shown in Table 2 and a scatter diagram derived by multidimensional scaling in Figure 4.

In Table 2, just to the right of the data for Group A1 is included a column for sherd 23. 5 out of the six distances from the other sherds exceed 6, two exceed 8 and one exceeds 9. We therefore exclude this sherd from Group A1.

Even for just the four sherds 2, 30, 32 and 37 one value of D equals 7.9. We cannot therefore exclude sherd 23 from this group. Sherd 39 is so very close to sherd 37 that it seems very reasonable to include it in this group, increasing the largest interpoint distance to 8.5.

We can see what we are doing from Figure 4. Because Group A2 looks a more diffuse group than A1 it is entirely reasonable to include sherd 23 in it. On the other hand if we had to include sherd 8 in Group A1 there would no longer be any grounds for separating Groups A1 and A2.

We conclude that Group A1 is made from a good quality clay, comparable with that for Group B, especially when we remember that the use of variables which are not quite uncorrelated makes interpoint distances a little larger than theory suggests. On the other hand Group A2 is made from material of definitely more variable composition.

Carrying out the same procedure for the three sherds of Group C we find

$$D\,(12,\ 28) = 4.9$$
$$D\,(12,\ 29) = 2.8$$
$$D\,(28,\ 29) = 3.3$$

So clearly these three sherds are made from a single, good quality clay.

We are left with one fly which ought to be in the ointment, namely sherd 42 which ought to be inside Group B. To solve this problem let us look at the archaeological evidence.

The Archaeological Evidence

The following evidence was supplied by Dr. Fulford:-

(a) Typology

Group B contained the red-figure body sherd as well as several classified as typical Attic on the basis of fabric, slip and decoration, a number of other sherds in this group would probably not have been included on visual inspection because of wear, burning or slight coarseness of fabric.

Group A1 - origin unknown, the main difference in appearance from Group B appears to be in the presence of small amounts of mica.

Group A2 - coarser, grey, sandy fabric. In many cases the slip does not coat the entire vessel. There are a number of reasons to suppose this to be North African fabric.

Group C - almost certainly Campanian, having a distinctive fabric and slip.

Of the two outliers in Group A and fairly near A1, one is of late Roman African red slipped ware and almost certainly made in Tunisia, the other is of an uncertain black glazed ware.

Sherd 42 had been presumed to be Attic; it might be of earlier date and so from a different part of the clay body.

(b) Chronology

The early phases of the site contain almost exclusively Attic black glaze pottery. Groups A1 and A2 first appear in the late 4th or early 3rd century. Group C does not appear until about 200 BC.

Conclusion

From the neutron activation analysis there can be little doubt that the whole of Group A is of Tunisian origin and this analysis has provided very clear-cut groupings by country of origin. Also the Tunisian material appears to come from several clays, of which two have been fairly well defined.

Table 1

Standard deviations of differences between logs for each element.

Element	(a)	(b)	(c)
Sm	.018	.025	.021
Ce	.011	.018	.022
Ti	.009	.019	.086
Lu	.015	.034	.036
Pa	.023	.014	.025
Cr	.022	.013	.035
Yb	.031	.035	.042
As	.014	.051	.103
Sc	.008	.015	.024
Fe	.010	.011	.022
Co	.007	.024	.046
La	.006	.023	.029
No. of pairs	5	18	24

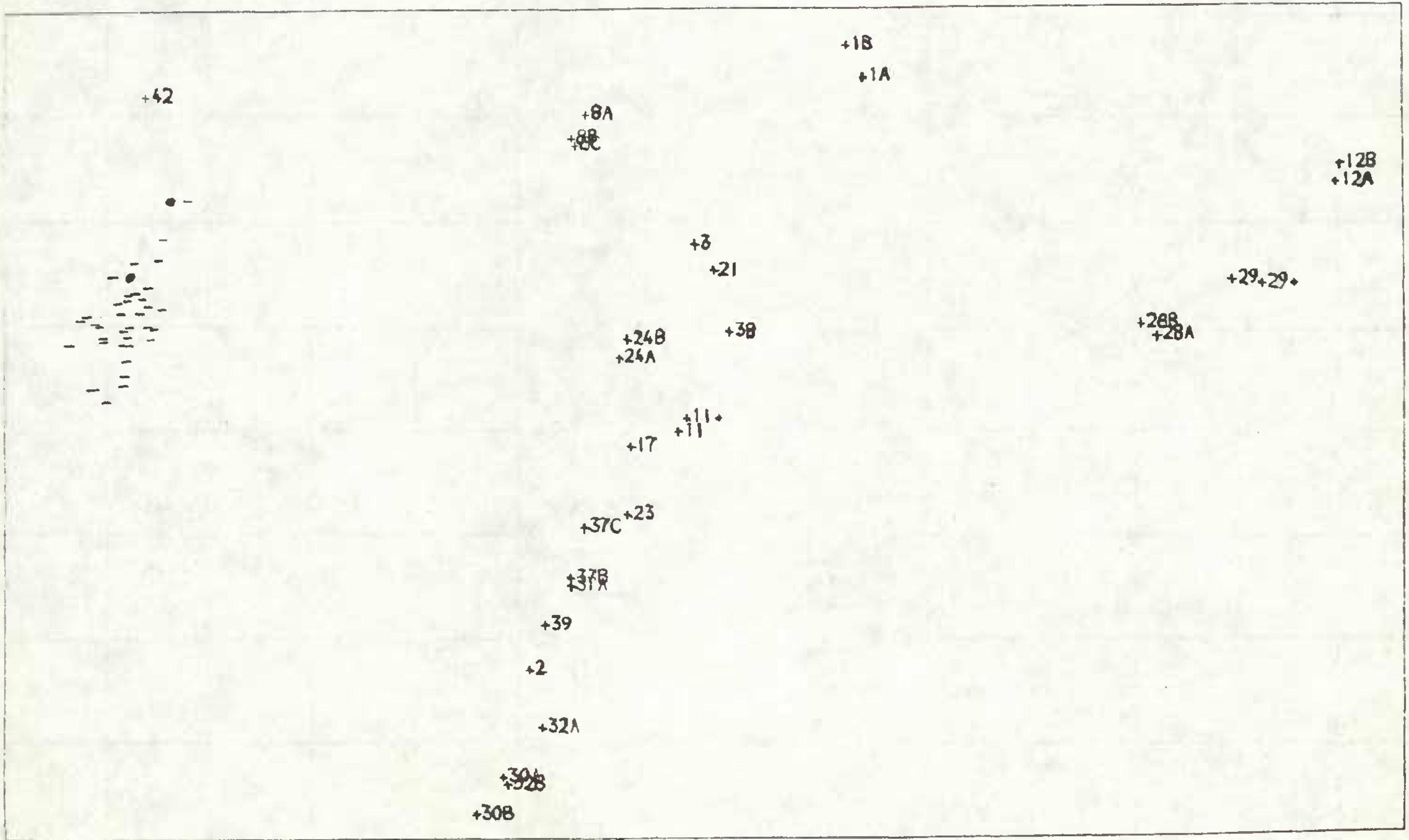
Table 2

Group A1

Sherd	11	17	21	24	38	23
3	4.4	3.3	3.4	5.0	5.8	8.1
11		3.5	5.6	5.7	4.9	5.7
17			5.7	5.9	6.6	6.9
21				3.8	6.5	9.4
24					4.9	7.0
38						8.5

Group A2

Sherd	30	32	37	23	39
2	4.4	3.8	5.4	5.9	5.9
30		4.5	5.3	7.1	5.4
32			7.9	4.8	8.4
37				7.6	1.3
23					8.5



+1B
+1A
+42
+8A
+3
+21
+12B
+12A
+29
+29
+28B
+28A
+24B
+24A
+3B
+11
+17
+23
+37C
+39
+2
+32A
+30B

FIGURE 1.

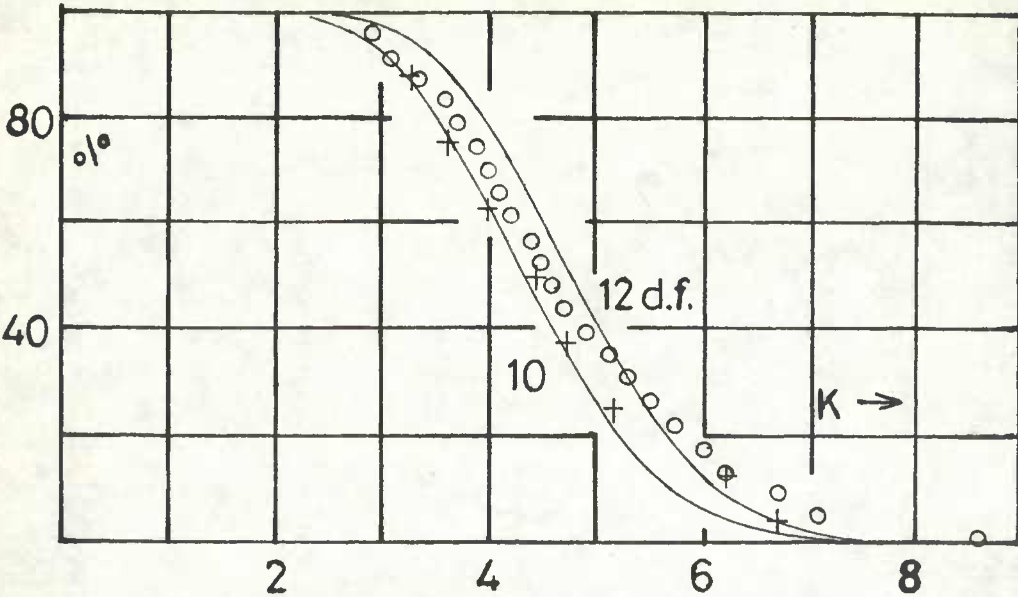
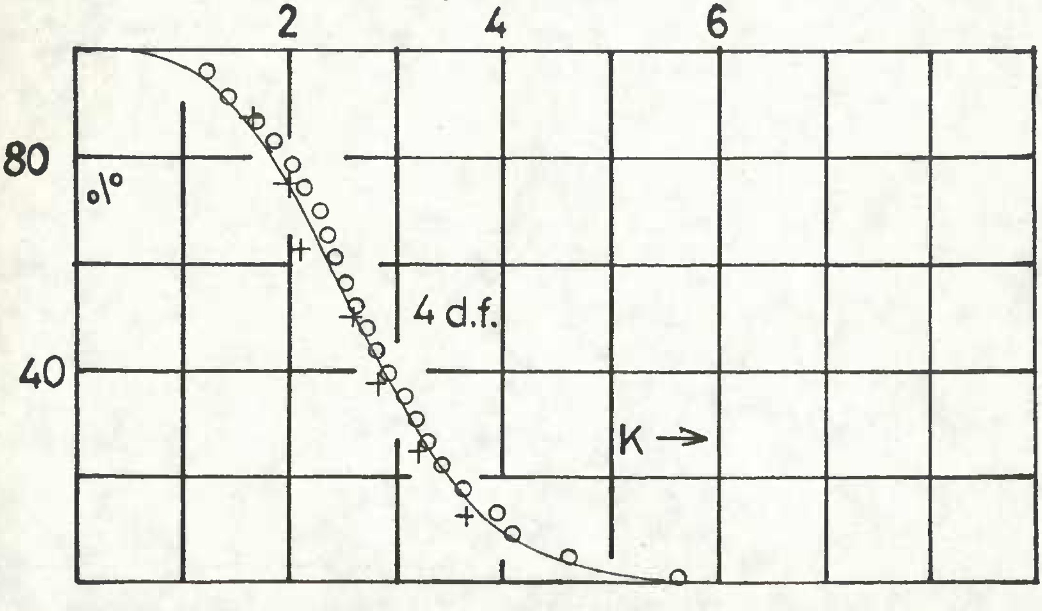


FIG. 2.

FIG. 3.



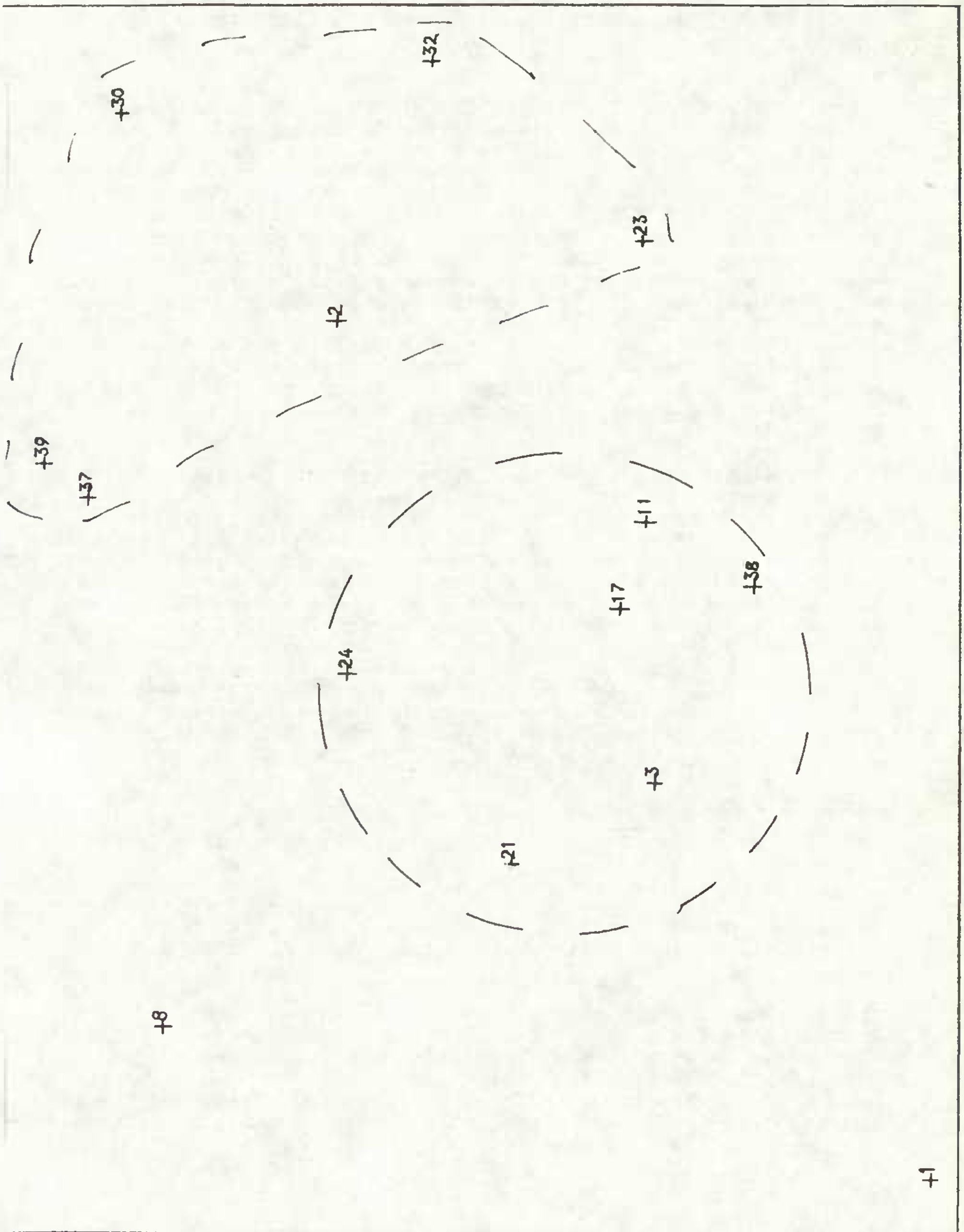


FIGURE 4