2
Experiments with Detrended Correspondence Analysis

Kris Lockyear
Institute of Archaeology, University College London

2.1 Introduction

Correspondence Analysis (CA) is now a widespread and accepted technique within archaeology (see, for example, the many papers in CAA87 — Ruggles & Rahtz 1988). A common, and oft-looked for, result in CA is a horseshoe curve (*e.g.,* Lockyear 1996a, figs. 1–2). If the aim of the analysis is the seriation of the objects/variables included, whether the principal gradient be time, or perhaps social status, geographical location or some other variable, then the results can be judged successful. If, however, the principal gradient creating the sequence is already known, then the horseshoe effect can mask other variation which may be of interest. Simply examining lower order axes does not solve this problem as the quadratic curve becomes a cubic curve and so on (Hill & Gauch 1980, p. 48).

This problem was encountered in my analysis of coin hoards of the Roman Republic (Lockyear 1996b). A CA which included 241 hoards was dominated by the time gradient, which was already known (Figs. 2.1–2.2). The dataset was sufficiently large that it was possible to divide it into 22 subsets on the basis of the closing date[1] of the hoard, and to analyse each set individually. The principal problem with this strategy is that it prevented sources of variation which cross-cut the closing dates of the hoards to be identified. In order to look at the dataset as a whole, three further statistical techniques were tried.

Two of these analyses using *Dmax-based Cluster Analysis* and *Principal Coordinates Analysis* have already been presented to CAA and published (Lockyear 1996a; see also Lockyear 1995). The remaining analysis using *Detrended Correspondence Analysis* (DCA), forms the subject of this paper.[2]

This paper will not consider the theoretical background to the technique (for which see Hill & Gauch 1980) but will present an 'experiment' into the use of the method in the analysis of hoard data which will hopefully illustrate the uses and difficulties of using this method for other types of archaeological data.

The aim of the method is to, in effect, straighten out the curve and to present the deviations from that curve. In the process, it is expected that the 'bunching' effect often seen at the ends of a horseshoe curve will also be eliminated. Two methods have been used to achieve this. The original method detrended 'by segments' (Hill & Gauch 1980). Fig. 2.3 shows, in a simplified fashion, how the method works. The detrending process takes place as part of the calculation of the ordination axes, not as a post-analysis transformation of the scores. The second method uses polynomial curves in the detrending process, rather than segments.

The technique has, however, been met with some caution by writers on the subject. Greenacre states that:

> In the process [of detrending], however, control over the geometry is lost and it is possible that... the detrending might introduce further artifacts into the results. (Greenacre 1984, p. 232).

Baxter is also unenthusiastic about the method stating:

> Since the horseshoe effect is natural to CA in the presence of seriation structure, and since in many archaeological uses an unambiguous ordering is all that is wanted, there may be little need to worry about it. (Baxter 1994, p. 120).

The only archaeological application of DCA of which I am aware is that by Shennan in which he examines assemblages of amber artefacts from Bronze Age Britain (Beck & Shennan 1991, pp. 85–98). Shennan states that:

> It was decided to carry out a detrended correspondence analysis to avoid the problems of the so-called horseshoe effect which introduces a correlation between the first and second principal axes (Beck & Shennan 1991, p. 85).

Unfortunately for our purposes the normal CA of the data was not presented. Shennan also used a further method, Detrended *Canonical* Correspondence Analysis which compares the ordination axes with a further set of axes derived from 'environmental' data, in this case associated finds, but found that his data were too sparse and the results were not significant (Beck &

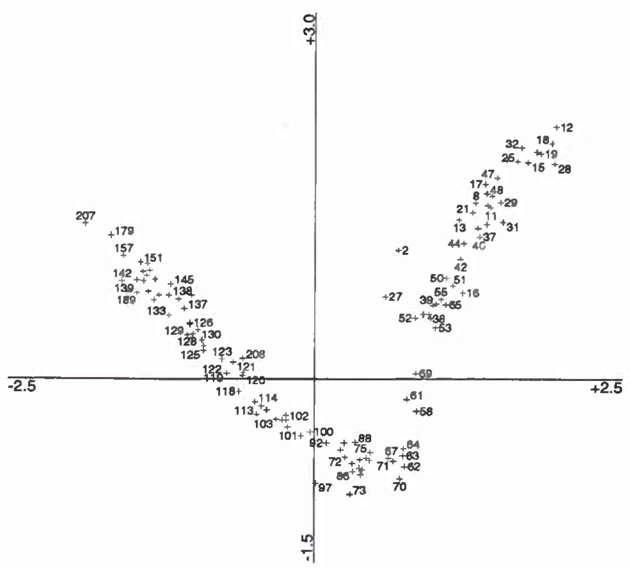

Figure 2.1: Species map from CA of 241 Roman Republican coin hoards. Data points are years of issue BC.

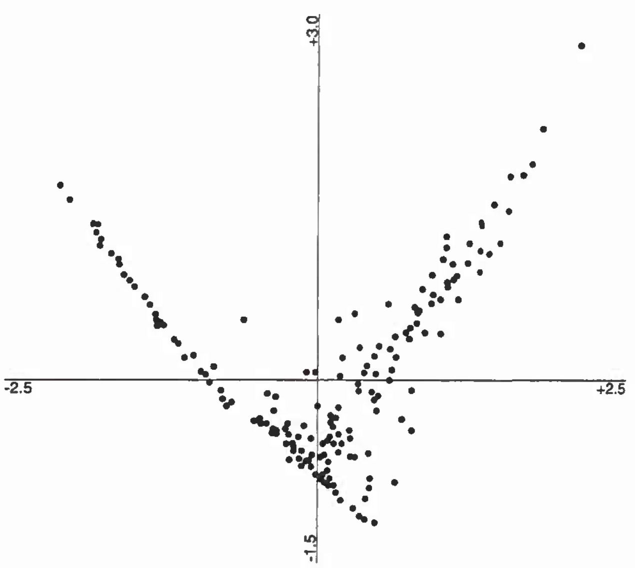

Figure 2.2: Sample map from CA of 241 Roman Republican coin hoards. Data points are coin hoards.

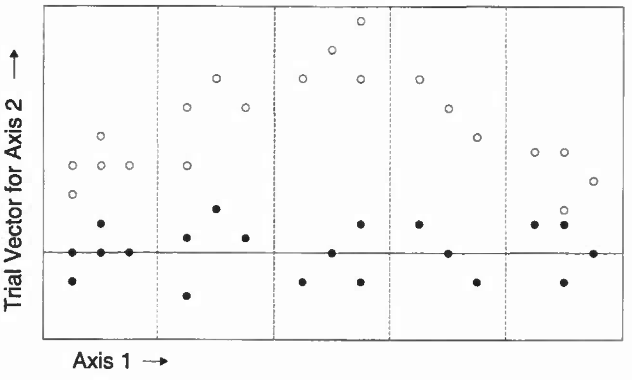

Figure 2.3: Detrending by segments. The gradient along axis 1 is divided into a number of segments, and then within each segment the values on axis 2 are adjusted (after Hill & Gauch 1980, Fig. 3).

Shennan 1991, p. 91). The final method he employed derived a correlation matrix between the ordination axes and the associated finds.[3]

2.2 Examining a test data set

It was decided to test the method on a relatively small data set. Twenty-four hoards were chosen consisting of 11,161 *denarii*. These hoards had been included in the cluster analysis previously (Lockyear 1996a) and came from clusters *b*, *f* and *g*.[4] All the hoards either came from Italy or Romania (see Table 2.1). To recap, the principal features of these clusters were:

Cluster *b*. This was the largest cluster from the analysis and consisted principally of Italian hoards closing from 82–71 BC, *or* Romanian hoards which could close anywhere from 77 to 32 BC. Six Italian and six Romanian hoards where chosen from this group.

Cluster *f*. This cluster consisted almost entirely of Italian hoards the majority of which closed in the 40s BC. Only two Romanian hoards were included in this cluster, one closing in 42 BC and one in 29 BC. Four Italian and the two Romanian hoards were selected.

Cluster *g*. This cluster of hoards included 11 Romanian and 7 Italian hoards. The Italian hoards mainly date to the late 50s–early 40s BC whereas the Romanian hoards date to the mid- to late 40s BC. Two Italian and four Romanian hoards were included in the test data set.

2.2.1 Analysis one — 'ordinary' CA

CA was performed on the test data set using the package CANOCO. Asymmetric maps were produced (Figs. 2.4–2.5)[5]. The first two axes accounted for 48.2% of the variation in the data set.

The results of this analysis illustrate a number of classic features encountered in the analysis of coin

cluster	code	hoard	country	closing date	total
b	CAR	Carovilli	Italy	82	40
b	COS	Cosa	Italy	74	1999
b	CST	Castelnovo	Italy	71	391
b	OSS	Ossero	Italy	72	465
b	PL2	Palestrina	Italy	74	357
b	VPT	Villa Potenza	Italy	71	411
b	CUC	Cuceu	Romania	48	484
b	FA2	Fărcaşele II	Romania	42	113
b	FND	Frauendorf	Romania	56	563
b	GUR	Gura Padinii	Romania	32	232
b	SDS	Sălaşul de Sus	Romania	54	103
b	SFI	Sfinţeşti	Romania	71	91
f	BOR	Borzano	Italy	42	582
f	CR1	Carbonara	Italy	48	383
f	CR2	Carbonara	Italy	36	2371
f	SPN	Spoiano	Italy	46	264
f	BPT	Bran Poartă	Romania	42	59
f	SEI	Şeica Mică	Romania	29	346
g	CAS	Casaleone	Italy	51	712
g	GRA	Grazzanise	Italy	54	256
g	ILI	Ilieni	Romania	46	108
g	PRS	Poroschia	Romania	39	541
g	TI2	Tîrnava	Romania	46	148
g	VIS	Vişina	Romania	41	139

Table 2.1: Details of the hoards used in the correspondence analyses. The total number of coins cited are those coins which have been identified to a reasonable degree of accuracy.

hoards (which in themselves may illuminate the results of other less well-dated archaeological assemblages). The sample map (Fig. 2.5) has cluster *b* hoards, those which have an 'Italian 70s BC' profile, plotted close together in the bottom left-hand quadrant of the map in a tight group. Although cluster *g* hoards form a separate group on the map, the four Romanian hoards are plotted close to the cluster *b* hoards while the two Italian hoards are plotted at the top of the second axis. Cluster *f* hoards are plotted towards the right of the map, relatively close together on the first axis, but spread out along the second. The two Romanian hoards (BPT and SEI) are 'pulled away' towards cluster *b*.

The variable map (Fig. 2.4) shows a horseshoe curve with the variables, in this case years of issue BC, plotted in an approximate sequence. We can interpret the axes in the following manner: the first axis represents time with the earliest issues to the left, and the latest issues to the right; the second axis represents relative abundance of middle period coins (roughly those from the 60s and 50s BC) at the top, with relative (or absolute) lack of those issues to the bottom. This is a classic seriated sequence. If we take the six Italian hoards on the right side of the map, they are in order with Carbonara (CR2) closing in 36 BC, followed by Borzano (BOR, 42 BC) through to Casaleone (CAS, 51 BC), with Grazzanise (GRA, 54 BC) lying on the return of the curve.

The Romanian hoards in clusters *f* and *g*, despite having overall profiles like their Italian counterparts as defined by the cluster analysis, are at a detailed level still 'drawn towards' cluster *b* with its 70s BC profile, *i.e.*, still have more of this early coinage than the Italian hoards. This reinforces the results of the cluster analysis previously reported (Lockyear 1995, 1996a).

The spacing of the hoards on the map is also of interest. The within-cluster variation of these groups (using the Dmax values as an approximate indication) is more-or-less the same, but the dispersion of the hoards on the maps is anything but. This can be explained relatively simply. Table 2.2 represents ten hypothetical hoards and 12 coin types. All hoards have types A–C, only 7 have E–F and 3 hoards have J–L. Although the hoards α–γ vary amongst themselves just as much as hoards θ–κ, the fact that all the hoards have types A–C means that the variation appears less significant on the CA maps than the variation in types J–L. In other words, hoards α–γ and types A–C will be plotted close together on the maps, whereas hoards θ–κ and types J–L will be spread out, despite intra-group variation being more-or-less equal.

Another good example of this can be drawn from the analysis of two chronologically overlapping data sets, the first containing hoards closing 147–118 BC, the second containing hoards closing 118–108 BC (Lockyear 1996b, sections 8.3.2–8.3.3, pp. 166–173). In the analysis of the first data set the three hoards closing in 118 BC were very widely spaced on the resulting map, but the same three hoards were plotted very close together on the map from the second analysis (Lockyear 1996b, Fig. 8.15 *cf.* Fig. 8.18b).

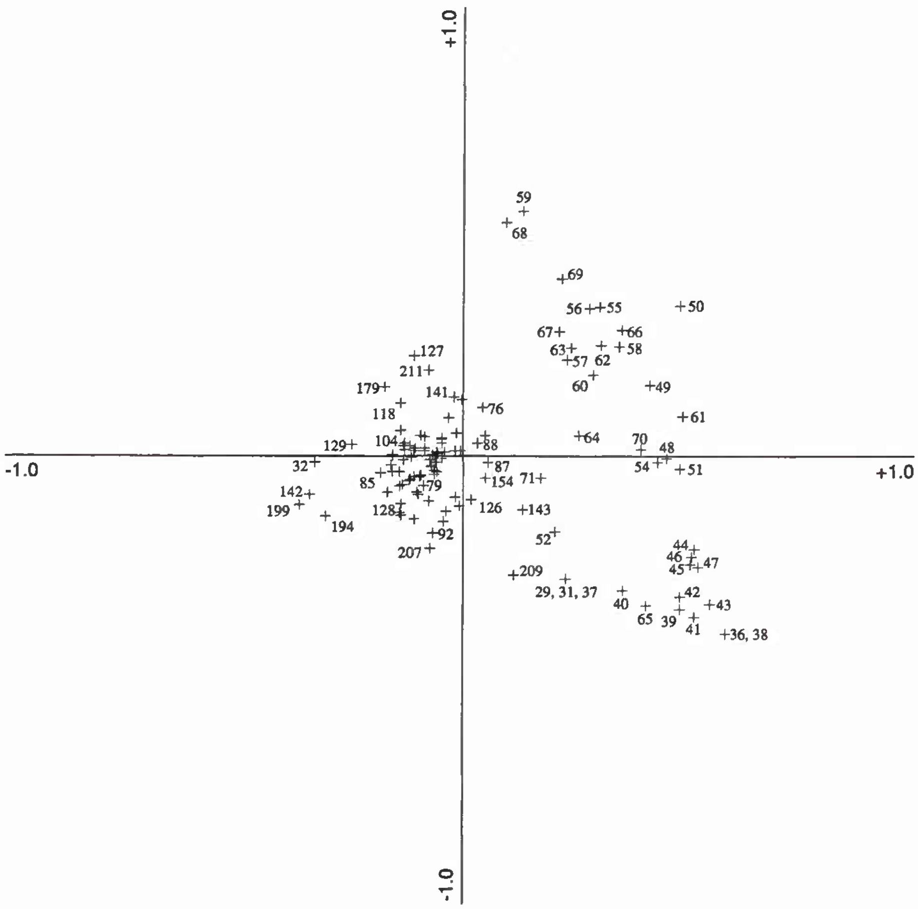

Figure 2.4: Species map derived from ordinary CA of 24 hoards as listed in Table 2.1. Data points are years of issue. First (horizontal) and second axes of inertia.

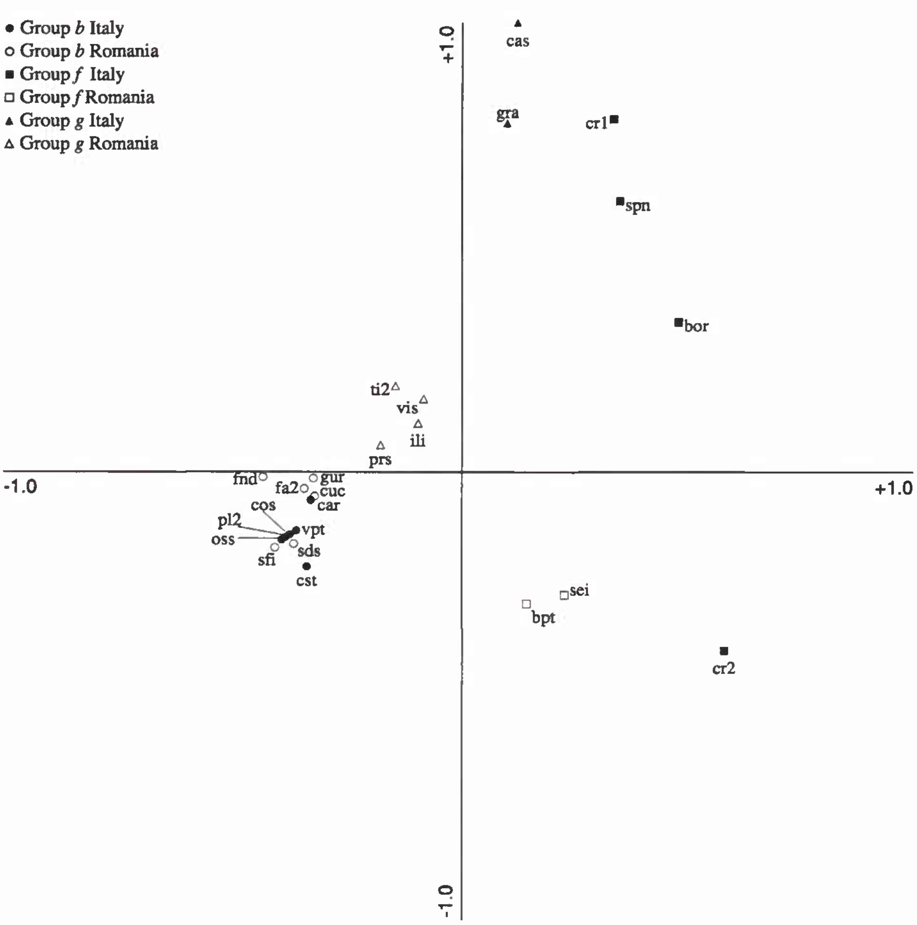

Figure 2.5: Sample map from ordinary CA of 24 hoards as listed in Table 2.1. Data points are coin hoards. First (horizontal) and second axes of inertia.

Table 2.2: Table showing ten hypothetical hoards (α–κ) with twelve hypothetical coin types (A–L). ○ represents a low occurrence of that coin type in the hoard; ● represents a high occurrence. See text for details.

	A	B	C	D	E	F	G	H	I	J	K	L
α	●	●	●									
β	●	○	●									
γ	●	○	●	●								
δ	○	○	●	●	●	●	●					
ϵ	○	○	●	●	●	○	●					
ζ	○	○	●	●	○	●	○					
η	○	○	○	○	●	○	●	●	●			
θ	○	○	○	○	●	○	●	●	○	●	●	○
ι	○	○	○	○	●	○	●	●	○	●	●	●
κ	○	○	○	○	●	○	●	●	○	○	○	○

2.2.2 Analysis two — detrended CA

CANOCO presents the user with a wide variety of options. For this experiment it was decided to use detrending by third order polynomials,[6] and to produce asymmetric maps as before, which are shown in Figures 2.6–2.7. The first two axes accounted for 43.7% of the variation in the data set.

As can be seen, the technique has removed the horseshoe pattern from the distribution of hoards on the map (Fig. 2.7). Obviously, the hoards are in the same order on the first axis as previous analysis but the second axis is somewhat different, and affects the overall map. Within group *f* the Italian hoards (SPN, CR1, BOR & CR2) now cluster tightly whereas the two Romanian hoards (SEI & BPT) form an isolated pair at the top of the plot. The twelve hoards from group *b* on the left of map are now split into Romanian hoards with negative scores on the second axis, and Italian ones with positive scores. Cuceu (CUC) is, however, nearer to the Italian group. The final group, *g*, falls between the other two groups on the first axis but is spread along the second with, notably, the two Italian hoards in this group separated from the Romanian hoards and the latter group clustering near to the Italian hoards of group *b*.

The detrended analysis has achieved its two aims of removing the horseshoe curve and counteracting the bunching effect discussed above. The analysis does raise some questions, which in light of Greenacre's comments regarding 'artifacts' in the results, must be examined. These are:

1. In what way are Şeica Mică and Bran Poartă (SEI & BPT) similar to each other?
2. In what way are they different from the other hoards including those in their own group?
3. Is the division within group *b* real, and if so what is it?

Five years (species) have extreme positive values on the second axis: 29, 31, 37, 207 and 65. Şeica Mică has one coin from each of these years. Of the other hoards, only Cosa (COS) has a coin of 207 BC and Carbonara (CR2) has a coin of 65 BC; no other hoards have coins of 29, 31 or 37. This explains the position of Şeica Mică on the sample map, but not Bran Poartă. $Dmax_{obs}$ for these two hoards is 11.1% and the application of the two sample Kolmogorov-Smirnov test shows no significant difference at the 0.01% level. Plotting these two hoards as cumulative frequency curves reveals their inherent similarity over the whole of the curve which, when allowing for variation due to Bran Poartă's small size, results in the two hoards appearing on the plot in much the same location — see Fig. 2.8, *cf.* Fig. 2.7. This clearly demonstrates the danger of interpreting symmetric maps without reference to the original data or diagnostic statistics. As Greenacre (1984, p. 65) states:

> The display of each cloud of points indicates the nature of the similarities and dispersion within the cloud, while the joint display indicates the correspondence between the clouds. Notice, however, that we should avoid the danger of interpreting distances between points of different clouds, since no such differences have been explicitly defined.

This analysis is a classic example of the problem, and is one reason why I prefer to present CA maps as two separate figures rather than one joint map.

The second question is how do these two hoards vary from the others in group *f*? Fig. 2.8 shows that the Italian hoards have relatively more new coin and relatively less old coin. The maximum cumulative difference being reached in 74 BC. The extremely jagged nature of the maximum difference line is mainly due to the small sample size of Bran Poartă.

The last question is, are the two sub-groups of *b* really different? To examine this problem Fig. 2.9 plots eight of the hoards from that group. As can be seen, the Romanian hoards have relatively more old coin when compared to the Italian hoards which have relatively more new coin. This is despite the fact that three of the Italian hoards have early closing dates. Carovilli (CAR) has no coin until 136 BC but closes first in 82 BC. This hoard is, however, the smallest in the analysis and it is not surprising that it has little of the older coinage. The maximum difference between hoards is reached in *c.* 117 BC and remains relatively level until hoards start closing in the 70s.

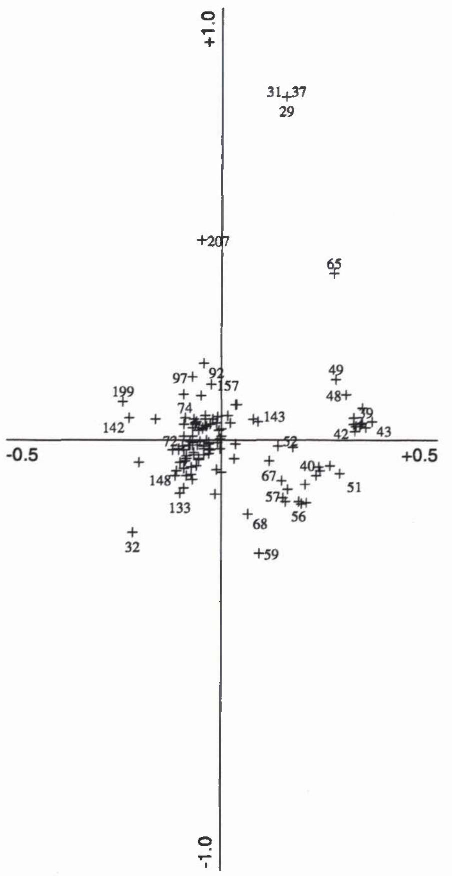

Figure 2.6: Species map from CA detrended by third order polynomials. Data points are years. First (horizontal) and second axes of inertia.

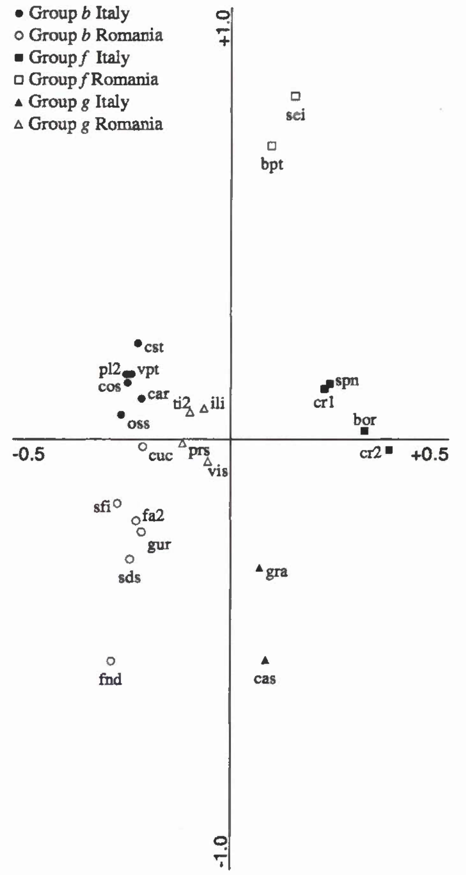

Figure 2.7: Sample map from CA detrended by third order polynomials. Data points are hoards. First (horizontal) and second axes of inertia.

2.2.3 Analysis three — the full data set

The full data set of 241 hoards was re-analysed using DCA and the maps presented in Figures 2.10–2.11. The first two axes accounted for 26.9% of the variation in the data set.

The initial impression is that this analysis is little more use than the original one. The data points of the right side of the maps are plotted on the first axis in a tight sequence. On the left side of the map the distribution on the second axis is extremely spread out. Examining the species map (Fig. 2.10) we can see that the early issues/hoards are on the right-hand side of the map (*e.g.*, Petacciato, PET, which closes in 141 BC). On the other hand, most of the years in the top left quadrant of the map post-date the battle of Actium, *i.e.*, are issues of Augustus. The issues at the extreme of the distribution in the bottom left quadrant are those of the 40s BC, *i.e.*, during the Civil Wars. The hoards labelled in the top-left quadrant are from across Europe: Penamacor (PEN) in Portugal, Bourguiel (BOU) in France, Breaza (BRZ) in Romania and Zara (ZAR) in Italy. The hoards labelled in the bottom-left quadrant were identified as having exceptional quantities of coins from the 40s BC.

These maps are capable of historical interpretation. Hoards from *c.* 150 BC to *c.* 55 BC come principally from three regions, Italy, Spain and Romania. The Romanian hoards, on the whole, resemble Italian hoards of the 70s BC as discussed previously; Spanish hoards are not hugely different from the Italian ones. Towards the end of the Republic coinage started to be struck at a wider variety of locations by the various protagonists in the Civil Wars. This leads to wider differences between hoards. Under Augustus, the Roman coinage system is imposed on the western provinces, while at the same time the incidence of hoarding in Italy decreases quite dramatically. The more widely dispersed hoards and the lack of a strong Italian 'benchmark' leads to very complicated patterning within these later hoards that makes the interpretation of the results of CA difficult (*e.g.*, Lockyear 1996b, pp. 240–246).

2.3 Conclusions

How much use is DCA? The analyses presented here are often, with hindsight, relatively predictable when one compares them with 'ordinary' CA. The method does, however, sometimes reveal aspects of the data

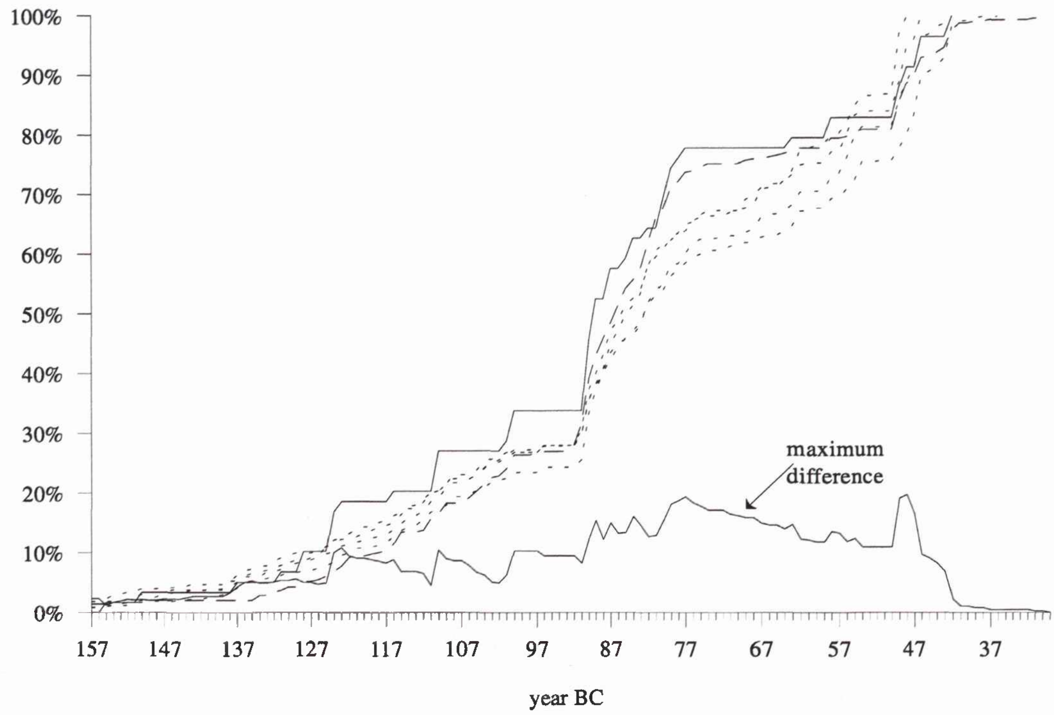

Figure 2.8: Cumulative percentage curves for BPT (upper solid line), SEI (dashed line), SPN, CR1, BOR, & CR2 (dotted lines). Bottom solid line is the maximum difference between hoards.

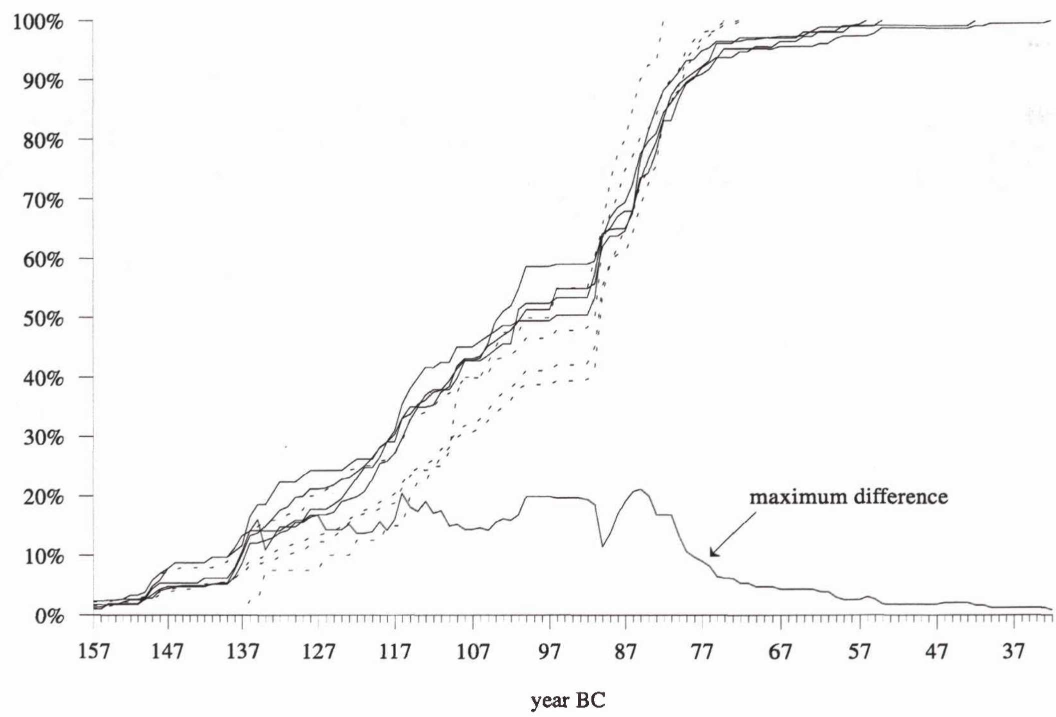

Figure 2.9: Cumulative percentage curves for eight hoards from cluster group *b* used in detrended correspondence analysis. CAR, COS, CST and OSS from Italy (dotted line) and FA2, GUR, FND and SDS from Romania (solid line). Bottom solid line is the maximum difference between hoards.

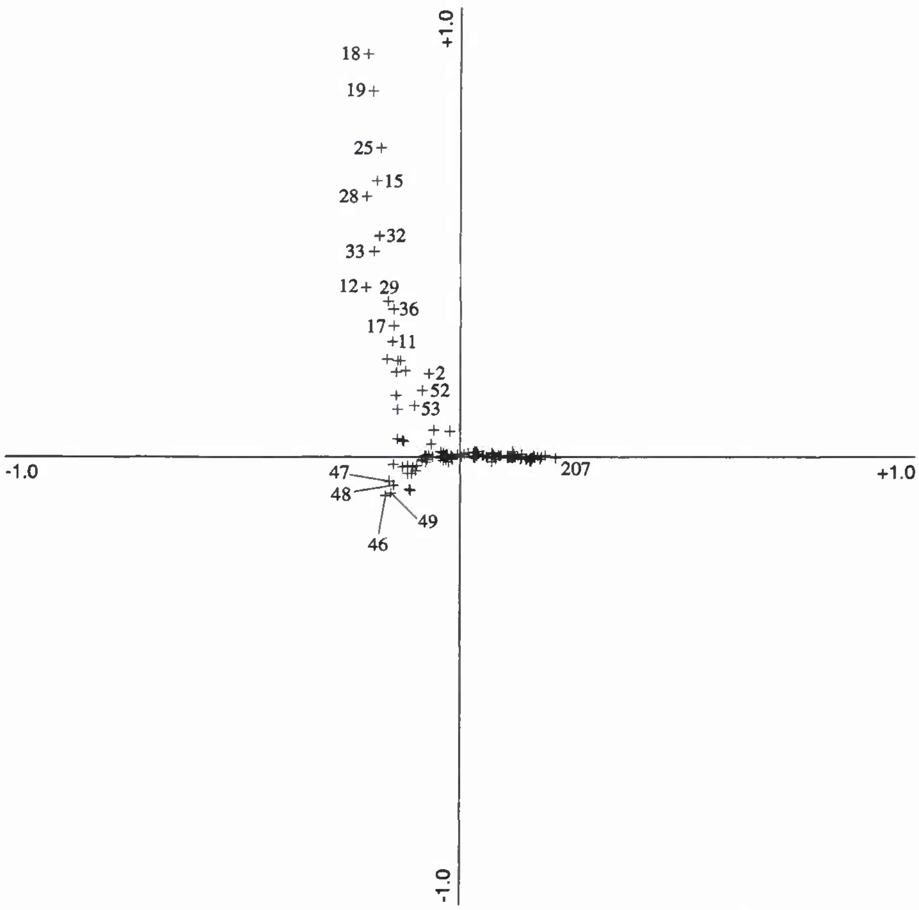

Figure 2.10: Species map from CA detrended by second order polynomials. Data points are years of issue BC; *cf.* Fig. 2.1.

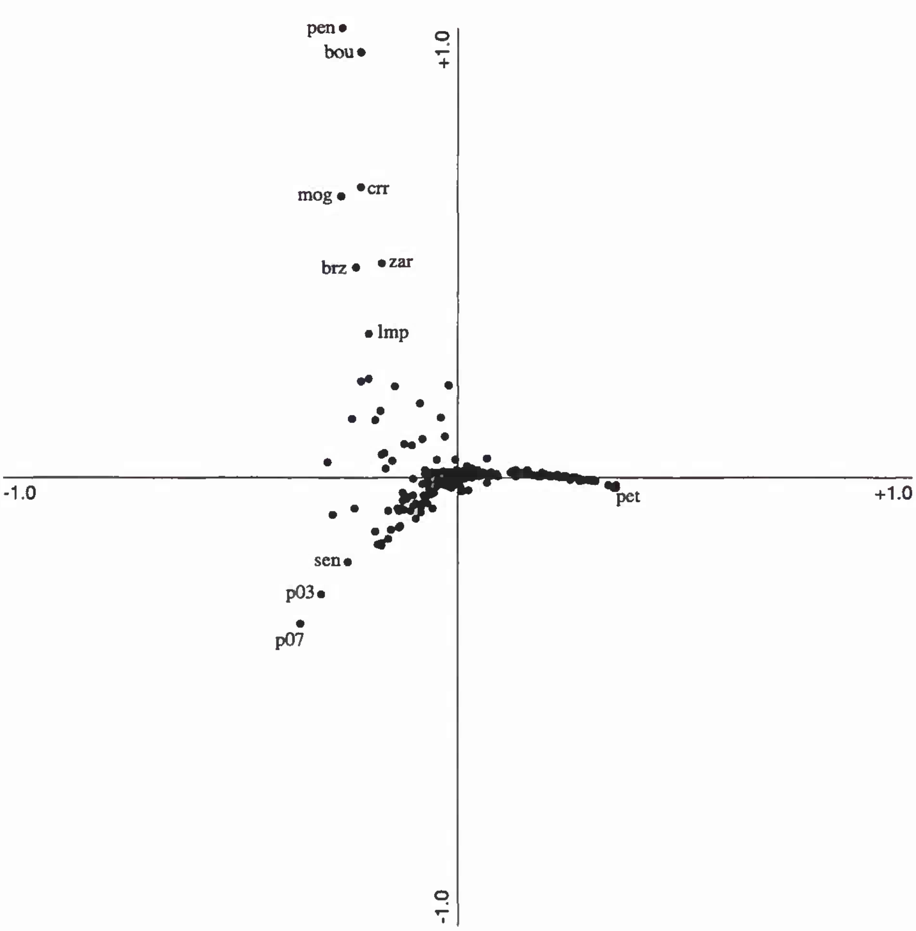

Figure 2.11: Sample map from CA detrended by second order polynomials. Data points are 241 hoards; *cf.* Fig. 2.2.

not visible on the original maps, such as the division of group *b* in Figure 2.7, and perhaps makes one look more closely at features such as the Şeica Mică and Bran Poartă hoards (*cf.* Figs. 2.5 with 2.7). the technique can show some other aspects of very large data sets, such as the 241 hoards presented here, but can still suffer from simply being too big.

Although DCA requires more careful interpretation than CA, it does seem worthwhile to try the technique if a data set exhibits a strong horseshoe curve, and to then compare and contrast the results of the two analyses. I would not recommend the use of the method without prior analysis by ordinary CA, and careful reference back to the original data is also a necessity.

Notes

1. *i.e.*, the date of newest coin in the hoard.
2. This type of analysis was originally provided by the program DECORANA but has been re-implemented in CANOCO (ter Braak 1987–1992) and WIN-BASP. CANOCO was used in the analyses here and I would like to thank the Dept. of Archaeology, University of Southampton for allowing me access to this program while I was working on my Ph.D.
3. Shennan also used the package CANOCO.
4. In this paper I have adopted the term 'clusters' for the groups derived from the cluster analysis to differentiate them from 'groupings' seen on the CA maps.
5. For the difference between symmetric and asymmetric maps see Greenacre (1993) or Shennan (1997).
6. Detrending by second order polynomials produces very similar results to those presented here. Fourth order polynomials pull in the more extreme negative values on the second axis. Other experiments not presented here were also undertaken (Lockyear 1996b, 307–311).

References

BAXTER, M. J. 1994. *Exploratory Multivariate Analysis in Archaeology.* Edinburgh University Press, Edinburgh.

BECK, C. W. & S. J. SHENNAN 1991. *Amber in British Prehistory.* Oxbow Monographs 8. Oxbow, Oxford.

GREENACRE, M. J. 1984. *Theory and Applications of Correspondence Analysis.* Academic Press, London.

GREENACRE, M. J. 1993. *Correspondence Analysis in Practice.* Academic Press, London.

HILL, M. O. & H. G. GAUCH 1980. 'Detrended correspondence analysis: an improved ordination technique.' *Vegetatio* 42: 47–58.

LOCKYEAR, K. 1995. 'The supply of Roman Republican *denarii* to Romania.' *Studii şi Cercetări de Numismatică* 11: 85–102. Published 1997.

LOCKYEAR, K. 1996a. 'Dmax based cluster analysis and the supply of coinage to Iron Age Dacia.' *In* H. Kamermans & K. Fennema (eds.), *Computer Applications and Quantitative Methods in Archaeology CAA95*, pp. 165–178. Institute of Prehistory, University of Leiden, Leiden. Analecta Praehistorica Leidensia 28.

LOCKYEAR, K. 1996b. *Multivariate Money. A statistical analysis of Roman Republican coin hoards with special reference to material from Romania.* Ph.D. thesis, Institute of Archaeology, University College London.

RUGGLES, C. L. N. & S. P. Q. RAHTZ (eds.) 1988. *Computer and Quantitative Methods in Archaeology 1987*, Oxford. British Archaeological Reports International Series No. 393.

SHENNAN, S. J. 1997. *Quantifying Archaeology.* Edinburgh University Press, Edinburgh, second edition.

TER BRAAK, C. J. F. 1987–1992. CANOCO — *a* FORTRAN *program for Canonical Community Ordination.* Microcomputer Power, Ithaca, New York.

Kris Lockyear
Institute of Archaeology, University College London
31–34 Gordon Square
London WC1H 0PY
k.lockyear@ucl.ac.uk