Simulation, Seriation and the Dating of Roman Republican Coins

KRIS LOCKYEAR

RESEARCH ARTICLE

]u[ubiquity press

ABSTRACT

Seriation was one of the earliest applications of computers to an archaeological problem. Despite the origins of the technique in numismatics, the vast majority of coinage studies manually sequence coin hoards and issues. For many periods, the coin designs or legends can be used to provide a date. For the Republican series, however, detailed sequences rely on the use of coin hoard data. In recent years, Correspondence Analysis has become the *de facto* seriation routine of choice. For coinage studies, however, where the period of manufacture was very short, a successful seriation would leave one triangle empty in the re-arranged matrix rather than concentrating the largest values on the diagonal as would be the aim in other situations. The aim of this paper is to assess the effectiveness of various easily-available off-the-shelf open-source seriation routines that have been used in archaeology for the analysis of this type of data. Given that we know a great deal about the pattern of production of Republican coinage, it is possible to create simulated coin hoard assemblages to test the various seriation routines and assess which technique is likely to provide the most successful results. This paper presents the results of applying 14 seriation methods to 27 simulated coin hoard data sets, and discusses the results.

CORRESPONDING AUTHOR:

Kris Lockyear

Institute of Archaeology,
University College London,
31–34 Gordon Square, London,
WC1H 0PY, UK

noviodunum@hotmail.com

KEYWORDS:

seriation; simulation;
Correspondence Analysis;
Roman Republican coin
hoards; numismatics

TO CITE THIS ARTICLE:

Lockyear, K. 2022. Simulation, Seriation and the Dating of Roman Republican Coins. *Journal of Computer Applications in Archaeology*, 5(1), pp. 1–18. DOI: <https://doi.org/10.5334/jcaa.57>

This paper is dedicated to the memory of Mike Baxter in grateful thanks for all his help and friendship over thirty years. He is much missed.

1 INTRODUCTION

Dating artefacts, and by extension contexts, layers and assemblages, is a fundamental concern for all archaeologists. Relative dating by typology is one of the oldest techniques in the archaeologist's toolbox, if one fraught with uncertainty as it relies so heavily on expert opinion. Petrie (1899), in his rightly famous paper, outlined a method which we now call *seriation*, that is, putting a series of assemblages into a sequence on the basis of the presence and absence of different types of artefact within them. Numismatists had, however, been using this concept for quite some time previously (Crawford 1990) and it seems likely that Petrie, who had quite a substantial coin collection, developed his idea from the methods used in numismatics.

The process of ordering the rows and/or columns of a table on the basis of values within them is one which attracted statistical attention from quite early on (e.g., Brainerd 1951; Robinson 1951), and many papers were published during the 1950s and 1960s, especially in the journal *American Antiquity*. Seriation was also an early application of computing to an archaeological problem (e.g., Ascher & Ascher 1963; Hole & Shaw 1967; Kuzara et al. 1966). O'Brien & Lyman (1999) provide a detailed overview of the subject. With the wider adoption of Correspondence Analysis in the 1980s (e.g., Madsen 1988; 1989), this has become the *de facto* method for seriation, with some authors — erroneously in my view — appearing to consider this the only rôle of CA (e.g., Siegmund 2015). One of the more sophisticated studies which used CA for seriation was an analysis of early Anglo-Saxon graves in England (Bayliss et al. 2013. See also the reviews by Baxter 2014a, b).

O'Brien & Lyman (1999) usefully divide seriation methods into three categories: phyletic, incidence and frequency. The first is essentially an ordering based on the principles of evolutionary archaeology, the latter two based on the presence/absence or the frequency of artefacts in closed assemblages, be they archaeological contexts, graves, or as is the case here, hoards. This study exclusively examines methods for frequency seriation.

Despite the origins of seriation techniques within numismatics, there has been almost no cross-fertilisation between the two disciplines. Numismatists have created quite complex sequences entirely manually (e.g., Crawford 1969, 1974) and I am only aware of one attempt to use a statistical method to seriate coin types (Backendorf & Zimmermann 1997). Conversely, archaeologists are generally blissfully unaware of the complexities of dating coins, despite often relying on them for dating their sites (Lockyear 2012). How accurately coins can be assigned a

date of manufacture varies considerably from one series of coins to another. For Roman Imperial coins, the date is often given by the portrait and name of the Emperor, and refined by the titles awarded to the Emperor and commemorated on the coin. So for example, a *denarius*¹ of Trajan (RIC 238) proclaims IMP TRAIANO AVG GER DAC PM TR P COS VI PP SPQR OPTIMO PRINC celebrating, among other things, his victories over the Germans and the Dacians, as well as being made Consul for the sixth time. From the historical sources we can date that coin to AD 112–114.

For the Roman Republican series, *denarii* are generally entirely devoid of direct evidence of date. There is a general stylistic development of designs going from completely anonymous, to anonymous plus symbols, then short sequences of letters, and eventually to longer legends. Even these longer legends, however, often only refer to one of the magistrates in charge of coinage for a year, who maybe otherwise unknown. Some types can be better dated, such as the *denarius* that has on its reverse two daggers and the legend EID MAR, which must post-date the assassination of Julius Caesar on the Ides of March 44 BC. The dates provided by the standard catalogue of these coins — *Roman Republican Coinage* (Crawford 1974) — are much disputed and alternative arrangements have been proposed for parts of the series (e.g., Mattingly 2004, chapter 13).

Following presentations of my work analysing Roman Republican coin hoards, the question has been often asked if Correspondence Analyses could contribute to the problem of dating (e.g., Melville-Jones 2009). I have always been somewhat cautious about this (Lockyear 2007, pp. 177–8). Most seriation techniques, including CA, seek to maximise abundance along the diagonal of the reorganised data matrix. In the case of using coin hoards, however, the aim is to have the one triangle of the matrix entirely empty (see [Figure 1](#) for an example). Discussions with statistical colleagues over thirty years have failed to provide a definitive answer to the question as to the effectiveness of CA in this specific situation. This raises the question as to which of the many methods currently available would be most appropriate for the seriation of Roman Republican coinage? Carlson (2017, chapter 17) has recently reviewed some of the methods available, and Hahsler et al. (2018, 2008) have conveniently provided many of these in a package for the R statistical system (R Core Team 2018). One aim of this current paper, therefore, is to explicitly test the effectiveness of these routines for the seriation of Republican coinage. I have deliberately restricted the analyses to easily-available off-the-shelf open-source packages. The simulation routines developed are available via the UCL research data repository, as are the data sets created for this paper should future researchers wish to test new routines.²

The essential problem in assessing the various methods is that we do not know what the correct sequence is. There have been studies using simulated data sets to assess the results of Correspondence Analysis (Kjeld Jensen &

	Lucoli \$	Zastok	Gerenzago	Terranova di Sicilia	Ancona	Bevacqua	Maddaloni \$	Fosoblanco	Monte Carotto \$	Taranto	La Barroca	Baresa	Segaro	El Centenillo	Córdoba \$	Sierra Morena	Torre de Juan Abad	San Lorenzo \$
MN.ACILI BALBVS	2	1	2	2	1	15	14	3	1	2	1	?	4	1	3	9	11	7
Q.FABI LABEO		1	2	5	1	11	10		2	3	1	?	3	3	11	16	24	4
C.CATO		3	2	2		16	9	2	3	1	2	3	4	11	14	11	7	
M.FAN C.F		1	1	3		6	5		1	3	?	?	1	2	5	11	7	4
M.CARBO		3	2			6	3		1	1	4	?	2	5	10	7	6	
Q.MINV RVF						3	5	1	2	1	3	1	1	2	3	6	11	1
C.PLVTI		1	3			4	3	2	1			?		2	11	3	5	
CARB		1			1	4	5	4		1	1			2	1	4	6	5
M.TVLLI		1	1	2	1	1	1			1				3	4	21	15	13
M.FOVRI L.FPHILI			3	1	1	18	13	16	6	5						4	1	1
L.PORCI LICI, L.LIC, CN.DOM						1	1	1		1	1	?				1		3
L.COSCO M.F. L.LIC, CN.DOM						3	2	5				1		1		1		2
L.POMPONI CN.F. L.LIC, CN.DOM					1	10	3											
M.AVRELI SCAVRI, L.LIC, CN.DOM					1	2	3											
C.MALLE C.F L.LIC, CN.DOM					1	3	2		1							2		
Q.MAR, C.F., L.R.						1	2									1		
M.CALID, Q.MET, CN. FOVL					1	6										2		
CN.DOMI						1										3		6
Q.CVRT, M.SILA										1	1	?				3		5
M.SERGI SILVS Q							8	1		1	2	1	3	1	1	9	5	2
Anonymous with Roma, wolf and twins								1		2	1	3	1	1	5	17	18	8
M.CIPI M.F																1	8	3
C.FONT										3	2	5	4	2	6	5	20	4
MN.AEMILIO LEP										1	2	2	3			5	3	5
CN.BLASIO CN.F									2	4	3			1	1	15	14	13
TI.Q.											2				2	2	3	4
L.CAESI																3	2	6
P.NERVA																2	3	3
L.TORQVA Q												7	2			4	5	4
L.PHILIPPVS																		1
T.DEIDI														1		6	2	4
AP.CL, T.MAL, Q.VR														1				2
C.PVLCHER														1	1	3	2	2
P.LAECA														2	2	4	5	4
A.MANLI Q.F SER																1		4

Figure 1 An extract from *Roman Republican Coin Hoards* (Table XI) showing part of the arrangement of coinage issues by Michael Crawford. Reprinted with permission of the Royal Numismatic Society.

Høilund Nielsen 1997) but these have, of necessity, been somewhat artificial. For later Roman Republican coinage, unlike many classes of archaeological evidence, we have an excellent understanding of the pattern of production:

- Three magistrates were elected each year to oversee the production of coinage. For most of the period in which we are interested, this leads to three different coin types being struck each year.
- Each coin type is struck for *only* one year.
- The size of coinage issues were highly variable. Although we do not know their absolute size, we can see their relative sizes via the hoard evidence.

Previously, it was found that including the whole series of late Roman Republican coin types in one analysis was unhelpful due to the complex mixture of underlying patterns (Lockyear 2007, pp. 44–6). The majority of my previous analyses have, therefore, minimised the time gradient by analysing hoards from as short a period as possible, occasionally a single year (e.g., 74 BC; Lockyear 2007, section 5.4.10), whilst still having a reasonable sized data set. In one case, however, the time sequence was of interest as there appeared to be a relationship between date and find spot (Lockyear 2007, section 5.4.5). After some experimentation, the following was found to be effective:

1. using a minimum hoard size of 30 well-identified coins;

2. using the dates of coin types from the standard catalogue, hoards closing³ over a period of twenty years were included;
3. also using the dates from the standard catalogue, all issues ten or more years older than the closing date of the oldest hoard were omitted.

In this way, it was possible to reduce the influence of early issues prior to the closing date of the earliest hoard without reducing the size of those hoards to the extent that they would not be viable in the analysis.

Using this experience as a template, and our knowledge of the pattern of coinage production, it is possible to create simulated data sets that have the same characteristics as real data, but for which we know the correct sequence of hoards and issues.⁴ By running these simulated data sets through the various seriation routines available, we can compare the true sequence with that suggested by the analyses, and are thus able to assess which technique is the most successful.

2 THE SIMULATION PROGRAM

The simulation program has been written in the statistical software system R (R Core Team 2018). In its simplest form, the program takes a single parameter which is the number of hoards required in the new data set. **Figure 2** shows the series of steps taken in a simplified form. Of the optional parameters, *seed* allows the user to manually

Figure 2 A simplified flow diagram showing the sequence used in the simulation. Of the inputs shown on the left, only the number of hoards required is compulsory. The remainder have default values which can be over-ridden if desired and are shown on the right.

set the seed for R's pseudo-random number generator, either as an integer, or as a string. For this study 27 data sets were created, nine each with 10, 20 and 50 hoards respectively. Five out of each nine were created using a short string as a seed, the remainder were run without manually setting a seed. In the following discussion, assemblages created using a string as a seed are referred to by that string, the remainder as `Rand7.50` where 7.50 represents the 7th data set of 50 hoards.

The simulation runs through a series of steps:

Create a list of the years of issue. Each simulated data set has coins from 91 issues: a year zero ('000'), which represents coinage from before the 30 year time span covered by the rest of the data set, and a further 90 issues, three per year. These are labelled 01a, 01b, 01c, 02a *et cetera*.

Create the hoard list. A list of hoard names is created of the form H01, H02, H03... until the requested total number of hoards has been reached.

Set the size of each issue. Estimates have been made for the size of issues during the Roman Republic by extrapolation from die counts (Crawford 1974). Although these estimates have been controversial (for example Buttrey 1993, 1994), it

has been shown that they are a good indication of the relative size of the issues (Lockyear 1999), and can be regarded as *Relative Issue Size Coefficients* (RISC). The distribution of RISC values over the whole of the Republic is approximately log-normal and so 90 RISC figures are created by sampling from a log-normal distribution with the same characteristics: a log-mean of 4.32 and a log-standard deviation of 1.18. The size of the issues is capped at 994. One can optionally pass a vector of issue sizes to the program via the parameter `mydies`.

Set size of the early issues. The size of the early issues (year 000) is calculated as a percentage of the sum of the RISC figures for the first 30 issues. This defaults to 40% based on the observed figure in real data sets, but alternatives can be set using the parameter `early`.

Create the coin populations. For each issue the appropriate RISC figure is entered into a matrix for the appropriate year. The rate of coin loss per annum was originally estimated at 2% by Thordeman (1948) for Swedish silver coinage, and a similar figure has been estimated by Patterson (1972) for 19th century US silver coinage. Although the use of 2% for Roman Republican silver *denarii*

has also proved controversial, it has been shown to be a reasonable estimate (Lockyear 1999). The RISC figures are, therefore, subjected to a compound depreciation of 0.02 using the formula $A = P(1 - r)^n$ where P is the RISC figure for that year, r is the depreciation ('decay') rate, here set to 0.02, and n is the number of years since that issue of coinage was struck. Having calculated the depreciation for each issue, we can sum the depreciated RISC values for each issue in an individual year to obtain the annual coinage pool. Alternative depreciation rates can be set using the parameter `decay`.

Set the hoard parameters. For the number of hoards requested, random closing dates between year 11 and year 30 are created. These are then sorted and assigned to hoard names so that H01 is always the earliest hoard.

The size of real hoards also follows a log-normal distribution, so the size of the simulated hoards is created by sampling from a log-normal distribution with similar characteristics: a mean of 4.36 and a standard deviation of 1.44, restricted to hoard sizes between 30 and 1500. The minimum figure is that used by the author in his analyses of real data. Alternatively, a vector of hoard sizes can be passed to the program using the parameter `totals`.

Create the simulated hoards. For each hoard, a random sample of coins of the size set in the previous step is selected from the coinage population for the appropriate year determined by the closing date. The sampling uses the profile of RISC figures for that year converted to proportions as a vector of probabilities. The program then returns the simulated data set as a matrix with the first column representing the closing date of that hoard, and the remainder of the columns representing the numbers of coins per issue in each hoard.

3 SERIATION ROUTINES

Fourteen seriation techniques were tested with the simulated data sets. All the analyses were conducted within the R Statistical System (R Core Team 2018). An extended batch file was written which created the required number of simulated data sets, and then applied the 14 seriation techniques to that data, and finally compared the results to the 'correct' sequence as discussed below. In each case, the early issues (year 000) were omitted from the analyses leaving a maximum of 90 variables representing three issues per year over a thirty year period, by 10, 20 or 50 hoards.

Correspondence Analysis (CA) The sequence of issues/hoards on the first axis of inertia. (Package `ca`; Nenadic & Greenacre 2007).

Detrended CA (DCA) Mainly used by ecologists to 'straighten out' horseshoe curves often seen in the CA of data with an underlying gradient. Originally proposed by Hill & Gauch (1980), but rarely used by archaeologists (Lockyear 2000). Sequence taken from the 'straightened' first axis. (Command `decorana` in the `vegan` package; Oksanen *et al.* 2017).

The remaining methods listed below were applied to data sets in three forms. For the PCA and PCA Angle methods, raw counts, counts standardised by types, or as percentages by hoards were used. As the Bond Energy Algorithm and the BEA Travelling Salesman Problem require non-negative data (Marquardt 1978, p. 294), these were tested using the raw counts, percentages and the chord transform using the function provided by the `BiodiversityR` library (Kindt & Coe 2005).⁵

Although CA has become the preferred method for analysing count data since the late 1980s, PCA can be used in the analysis of artefact counts (Baxter 1994, pp. 61–2), and has been used for the analysis of coinage assemblages (Ryan 1982, Ryan 1988, p. 72–87, Creighton 1992, 33–5, *c.f.* Lockyear 1996, sections 3.12.2 and 8.2.5, Lockyear 2007, pp. 60–4). I have, therefore, included this method in those tested here.

Thus, three options were tested with all four methods.

Principal Components Analysis (PCA) The sequence of issues/hoards on the first axis. (Command `seriate(x,method='PCA')` from the `seriate` package; Hahsler *et al.* 2018.)

PCA Angle (PCAang) Uses the angle from the origin to the data point from a plot of the first two principal components. Uses the largest angle between two data points to determine the ends of the sequence. (Command `seriate(x,method='PCA_ANG')` from the `seriate` package; Hahsler *et al.* 2018; Carlson 2017, p. 384.)

Bond Energy Algorithm (BEA) Based on a method originally suggested by McCormick Jr. *et al.* (1972). Starting from a random point, the algorithm arranges the rows and columns to minimise the measure of effectiveness. As the results are dependent on the starting point, the routine is run 25 times and the optimal result kept. Note that because of this, multiple runs of this method may not produce identical results. (Command `{seriate(x,method='BEA',control=list(rep=25))}` from the `seriate` package; Hahsler *et al.* 2018.)

BEA Travelling Salesman Problem (BEA TSP) As for the previous method but using the travelling salesman solver as a measure of effectiveness. (Command `seriate(x,method='BEA_TSP',control=list(rep=25))` from the `seriate` package; Hahsler *et al.* 2018.)

4 RESULTS

To compare the results of the seriation routines with the original data a Spearman's Rank Correlation coefficient (ρ) was calculated between the correct order of the issues/hoards and the seriated sequence, a method used effectively in other studies (van de Velden *et al.* 2009, p. 3135). In order to avoid negative correlation coefficients where the technique has reversed the sequence, the absolute value has been reported in all cases.⁶

For the issues, the first ten years of issue are, essentially, a randomising element as the earliest possible closing date for the hoards is year 11. I have, therefore, calculated $\hat{\rho}$ for both the whole data set, and for a trimmed data set omitting years 1–10. I have presented both values because in real data sets we cannot be sure which issues pre-date the earliest hoard.⁷ Some issues may be missing from the simulated data sets entirely if they are rare, and so the total number of issues involved varies slightly from data set to data set. In all cases, $\hat{\rho}$ is a conservative figure because each group of three issues per year could be

represented in the seriated sequence in any order but still be correct.

Issues which only occur in a single hoard present an additional problem. If the issues are large, and only occur in one hoard, we could probably be confident that they belong at the end of the sequence. A rare issue might, however, occur a few times in an especially large hoard, or a common issue only once in a small late hoard. Rather than introduce a measure of subjectivity in deciding what to exclude or include, I have run the analyses with all the data. This, again, might result in a conservative estimate of $\hat{\rho}$ but as all methods are treated equally, this is not a problem. With an analysis of real data, one would consider quite carefully what to exclude.

Figures 3–5 show the results graphically. Details of the results are presented in Appendix A, Tables 2–4. The tables show the results for the issues and hoards for all the methods along with the lowest, median, and highest values of $\hat{\rho}$ for the nine sets of simulated hoards.

Figure 3 and Table 2 show the results for the small assemblages of only ten hoards. This is very small: the CHRR

Figure 3 Results of the simulations with 10 hoards.

database (Lockyear 2016) has at least 10 hoards per decade from 130 BC onwards. Even with this small a data set, CA has mainly performed well with only data sets 7.10 and 9.10 being especially poor. The results for the hoards are better than the issues but this is to be expected as there are only ten hoards to sequence correctly compared to 90 issues, or 60 when trimmed. DCA has performed similarly well to CA for issues although marginally more poorly for hoards. The remaining methods were all out-performed by CA in most cases. CA/DCA was the most consistent technique with values of $\hat{\rho}$ between 0.65 and 0.89 for issues (0.70–0.90 trimmed). Whereas the other techniques may sometimes do as well, such as the PCA angle method used on standardised data, they have a much greater range of values of $\hat{\rho}$. For example,

the PCA angle method for issues returned $\hat{\rho}$ figures between 0.29 and 0.89 (0.27–0.86 trimmed).

The increase in sample size to 20 hoards improves the results slightly for CA/DCA (Figure 4 and Table 3). The remaining methods, however, still exhibit a wide range of values for $\hat{\rho}$ for both issues and hoards. Increasing the sample size to 50 hoards for CA/DCA of issues improves the results markedly (Figure 5 and Table 4). The figures for hoards drop slightly from the smaller data sets simply because we now have 50 hoards closing over a 20 year period leading to many having identical closing dates. CA/DCA remains the most consistently successful technique with values for the issues of $\hat{\rho}$ in the range of 0.84–0.95 (0.94–0.98 trimmed).

Figure 4 Results of the simulations with 20 hoards.

Figure 5 Results of the simulations with 50 hoards.

To examine the results more fully, let us look at one data set — *Kevin* — in detail ([Table 1](#)). One way to easily visualise the data is via a Bertin plot using the facilities provided by the *seriate* package. [Figure 6](#) presents the original data in the ‘correct’ order. Four issues dominate the assemblage: 06c (8.7%), 09c (7.6%), 15c (10.0%) and 19c (4.5%). Two of these issues predate the earliest hoard. Hoard H10 contains no coins of year 20 despite being selected from the coinage pool for that year. For real data, this hoard would have been given a closing date of year 19. Similarly, hoard H16 has no coins of year 25 and would be dated to year 24. I have modelled the impact of the composition of the coinage pool on dating elsewhere (Lockyear 2012, pp. 203–7).

[Figure 7](#) presents the data in the sequence as suggested by the first axis derived from CA. First impressions show a largely empty upper triangle as wanted. The lower-left of the graph is, however, much sparser than the original data and is the result of the method, in effect, trying to concentrate larger values on the diagonal. Most surprising, however, is the incorrect placement of issue 15c before issue 09c. This is, however, a pretty successful result with $\hat{\rho}$ being 0.80 for the issues (0.95 trimmed) and 0.95 for the hoards. Examination of the ordination maps for this data set ([Figure 8](#)) show a pretty typical result compared to real hoard data sets which are characterised by the clustering of early issues and the spread-out sequence of the later issues (see, for example Lockyear 2007, Fig. 5.68)

We can compare this to the result for the PCA angle method with standardised data (Figure 9). In this case, it was marginally more successful than CA if one uses the whole data set (issues: $\hat{\rho} = 0.83$) but marginally less successful if compared just using the trimmed data ($\hat{\rho} = 0.93$). This is not, however, universal with some data sets having very poor results for this method (e.g., *Kim*:

$\hat{\rho} = 0.16$, *Rand6.20*: $\hat{\rho} = 0.12$). With the *Kevin* data set, the technique has mixed the sequence of three of the largest issues: 15c, 06c and 09c. In general, it is more successful with the later hoards and issues.

In contrast, the results from the BEA TSP method on chord transformed data are less than optimal (Figure 10). The order for the issues has a low value for $\hat{\rho}$ of only 0.08 and the plot shows a concentration of values at the centre of the plot, and sparse areas at either side. The later hoards, H17–H20, are placed in the middle of the hoard sequence.

At present, just the first axis of inertia from the CA is being used to give the sequence. It might be preferable to be able to use the first two axes and to take the sequence around the horseshoe curve (Baxter 2003, p. 204). Principal curves (Hastie & Stuetzle 1989) is one possibility that has been suggested (Carlson 2017, pp. 390–6) and is available via the R-package *princurve*. Given the nature of the data distribution, fitting a curve unaided provides an unhelpful result, illustrated here using the *Kevin* data set (Figure 11a). It is possible, however, to include some points in the calculations to improve the fit (Carlson 2017, pp. 391–2) as shown in Figure 11b. Unfortunately, it would be difficult to automate this, especially with more problematic data sets (see below). For the *Kevin* data set, using the sequence around the curve improved the results slightly for issues ($\hat{\rho} = 0.83$).

An alternative method of ‘straightening out’ the curve is detrended CA, as applied in the above tests. Detrended CA has not met with universal acceptance (see references cited by ter Braak & Šmilauer 2015, p. 684) but is relatively common in the field of community ecology. The process of detrending was developed because the ‘arch effect’, often seen in archaeology as a sign of a successful seriation (Baxter 2003, p. 139), is argued to be the result of ‘spurious polynomial axes’ (ter Braak & Šmilauer 2015, p. 688) especially when a single gradient is required as here. Normally, the detrending is calculated ‘by segments’ as originally proposed by Hill & Gauch (1980). Detrending by polynomials seems

HOARD	CLOSING DATE	TOTAL	TOTAL EXCL. YEAR ZERO
H01	12	68	53
H02	14	33	23
H03	15	84	66
H04	15	895	730
H05	16	42	36
H06	16	650	521
H07	16	387	309
H08	17	201	165
H09	18	191	155
H10	20	37	33
H11	20	126	112
H12	21	1069	911
H13	22	225	193
H14	23	239	202
H15	24	145	126
H16	25	76	64
H17	27	245	216
H18	28	148	138
H19	28	77	65
H20	29	744	666
Totals		5682	4784

Table 1 Details of the hoards in data set *Kevin*.

Figure 6 The original data in hoard *Kevin* as a Bertin plot.

Figure 7 The data in hoard Kevin as a Bertin plot after seriation using CA.

Figure 8 Ordination maps from CA of data set Kevin.

Figure 9 The data in hoard Kevin as a Bertin plot after seriation using the PCA angle method on standardised data.

Figure 10 The data in hoard Kevin as a Bertin plot after seriation using the BEA TSP method on chord transformed data.

to only be available through CANOCO (Lepš & Šmilauer 2003, pp. 52–4). As an alternative to principal curves detrending by segments was included in the trials. In only one of 27 cases, however, was there a difference greater than 0.05 between the $\hat{\rho}$ figures for the CA and DCA. The one exception was data set Rand9.10 which had a low score of $\hat{\rho} = 0.6$ for CA but an even lower score of 0.43 for DCA. Despite the reservations in literature about DCA, it would appear no better or worse in this scenario than ordinary CA.

One weakness of the automated approach adopted in this study compared to using CA to analyse real data sets is that problematic data which would ordinarily be dealt with by the analyst, are left ‘as is’. As an example, the CA of data set Lyn is clearly dominated by two issues on the second axis of inertia (Figure 12a). Problems such as these are not uncommon in real data (Lockyear 2007, section 5.2.5). They are usually created by rare issues in small hoards. In the case of the Lyn data set, there are only three coins of issues 29a–b which occur in the only hoard to close in years 29 or 30, H20. H20 is a small hoard of only 32 coins of which only 29 are included in the analysis. Data set Lyn returned poor results from the seriation routines and so to test whether omitting issues 29a–b improved the results, the CA and DCA was re-run omitting

those years. The result of the CA is shown in Figure 12b and, in terms that would normally be applied to a CA, is much more satisfactory. Testing the sequence on the first axis, however, showed almost no difference between the CA of the full data set and that omitting years 29a–b.⁸

5 DEALING WITH UNCERTAINTY

With real data we do not, of course, have the ‘correct’ answer to compare to the results of a CA. One thing is clear from the simulations is that even with a large set of ‘hoards’ which are truly random selections from the coinage pools — something we cannot be sure of with real data sets — the results are never perfect. Although numismatists frequently attempt to assign precise dates to issues, and suggest sequences into which they all fit, we must take account of uncertainty. Clearly, small issues will be harder to place within a sequence than larger issues. One method by which we can assess uncertainty would be to use bootstrapped CA (Lockyear 2013).

In order to look at how we might use such an analysis, I have subjected a trimmed-down version of a previously analysed data set of hoards from Italy and the Iberian peninsula at the end of the second century BC (Lockyear

Figure 11 Principal curves through the issues from the CA of data set Kevin.

2018) to a bootstrapped CA using the *cabootcrs* package written by Ringrose (2012). Hoards which are known to be somewhat unreliable and those with a poor ‘quality’ in the original analysis were dropped from the data set leaving 41 coin hoards containing 65 different coinage issues. Early issues from ten years or more prior to the closing date of the earliest hoard, using the dating scheme proposed by Crawford (1974), were omitted from the analysis. As a result, the hoards as analysed contained between 10 and 664 coins, and the issues were represented by between six and 239 examples. The maps from this analysis are shown in [Figure 13](#).

With a large analysis like this, assessing overlaps between individual issues is difficult from maps like [Figure 13b](#) and so I have plotted each issue’s score on the first axis, along with a line representing approximately 95% of the bootstrapped scores for that issue ([Figure 14](#)).⁹ In that figure we can see that 21 early issues, shown by the topmost 21 lines, have a very similar score on the first axis and probably represent those struck before the earliest hoards’ closing dates. These are the equivalent of years 01a–10c in the simulated hoards, and cannot be sequenced from this data set. From the line for RRC 282 onwards, however, the scores change more

Figure 12 Symmetric maps from CA of data set Lyn. Filled circles are hoards, open circles are issues.

quickly. Some issues, however, have a very wide range of possibilities. These are generally the rarer issues such as RRC 295 which only has 14 examples in the data set. Larger issues, such as RRC 289 (239 examples), RRC 302 (183 examples) or RRC 299 (138 examples) are more securely sequenced. In [Figure 14](#) the uppermost 95% limit for RRC 302 is shown as a vertical dashed line. From this we can see that there is a fairly high probability that it dates to before issues RRC 323, 325, 320 and so on, as there is no overlap between their 95% lines.

Numismatically speaking, the result for RRC 299 is one of the more interesting as it is placed much later in the sequence than Crawford had placed it.

6 CONCLUSIONS

Despite CA generally concentrating the largest values on the diagonal of a re-organised data matrix, it represents the best method, of those tested here, to seriate coin

Figure 13 Ordination maps from the bootstrapped CA of hoards from Italy and Spain at the end of the second century BC. Ellipses are 95% confidence regions derived from the bootstrapping.

hoard data as shown by the trials using the simulated data. Although detrended CA may help with the issue of ‘stretching’ where the later hoards are more spread-out on the first axis from a CA than the earlier ones — as discussed in detail elsewhere (Lockyear 2000) — the concern that the process destroys the geometry of the method (Greenacre 1984, p. 232) seems to outweigh any minor benefits it might have.

The basic premise of using simulated data sets to test statistical methods has long been known. One of the principal issues, however, is knowing enough about the material to create realistic data sets. For this particular coinage series, we have a very good understanding of this thanks to several hundred years of research. Unlike pottery, for example, we do not need to take into account variable use-lives as, with one exception,¹⁰ these

Figure 14 The circles plot score for each coin issue on the first axis of inertia from a CA. The lines represent approximately 95% of the values from the bootstrapped replicates calculated from the reported variances. The issue numbers are those from Crawford (1974). The vertical dashed line shows the upper bound for issue RRC 302.

are constant (cf. van de Velden *et al.* 2009). Similarly, the majority of issues are struck in Rome and circulate quite quickly minimising the problems of distance between production sites and hoard locations. In analysing real data one would take the patterns of supply and production shown in my previous work (Lockyear 2007) into account. For example, hoards from Romania that include unrecognised contemporary copies and where the pattern of supply is erratic (Lockyear 2008) would need to be omitted from the analyses. We do not, however, lack for data. As of 1st January 2022, the CHRR database contains information on 974 hoards, of which detailed data exists for 585 (mainly dating to 157–2 BC) that contain 111,846 well-identified coins.

An appreciation of the variable degrees of uncertainty in the sequence can be obtained by bootstrapping. In general, we have to accept that despite there being a ‘correct’ answer for the striking of Republican issues, it is very unlikely we will be able to unequivocally recover that sequence. Rare issues are always going to be especially problematic. We should be prepared to think in terms of probabilities.

CA remains, however, somewhat of a blunt instrument. There are other sources of data which could contribute to the correct sequencing of issues such as die-links. Coin dies have a limited life, and where two issues can be shown to share dies, it is almost certain they were struck in the same year, or perhaps at most a year or two apart. Historical information on the coins, such as that mentioned in the introduction, could also contribute to refining the model.

One approach would be to use a modified version of the constrained CA method (Groenen & Poblome 2003; Poblome & Groenen 2003). Alternatively, a Bayesian approach might be adopted such as that proposed by Halekoh & Vach (2004). If we could include prior knowledge noted above, we could improve the sequences we obtain, and could even start to test and compare competing hypotheses as to the correct date sequence. Ideally, such an approach should be capable of being generalised to other archaeological situations. This is, however, non-trivial and beyond the scope of the current study.

NOTES

- 1 The standard Roman silver coin in circulation between 211 BC and the mid-third century AD.
- 2 The program code is available at DOI [10.5522/04/12076530](https://doi.org/10.5522/04/12076530) and the data sets at DOI [10.5522/04/12074604](https://doi.org/10.5522/04/12074604).
- 3 The closing date of a hoard is given by the date of the latest coin included within it.
- 4 For earlier studies using simulated hoard data sets, see Lockyear (1991).
- 5 I would like to thank the anonymous referee for the reference and suggesting the use of the chord transform.
- 6 Although the precise sequence for Roman Republican coins is unknown, the approximate sequence can be determined via historical and stylistic criteria and it is, therefore, possible to unambiguously identify the start and end of a sequence suggested by seriation.
- 7 As noted above, a moderate number of issues from before the closing date of the earliest hoard have to be included to avoid having hoards which, after trimming of the earlier issues, contain almost no coins.
- 8 CA full data set: $\hat{\rho} = 0.76$, trimmed $\hat{\rho} = 0.83$; omitting years 29a–b: $\hat{\rho} = 0.74$, trimmed $\hat{\rho} = 0.82$. DCA results are identical.
- 9 The 95% estimate has been calculated using the variances in the scores from the bootstrapped replicates for each issue reported by the cabooters package.
- 10 The debased legionary *denarii* of Mark Antony (RRC 544) remained in circulation much longer than other Republican issues which were removed from the coinage pool in the early second century.

ADDITIONAL FILE

The additional file for this article can be found as follows:

- **Appendix A.** Detailed results (Tables 2–4). DOI: <https://doi.org/10.5334/jcaa.57.s1>

ACKNOWLEDGEMENTS

I would like to thank Trevor Ringrose for his extremely helpful comments on an earlier draft of this paper, and for his help with bootstrapped CA.

COMPETING INTERESTS

The author has no competing interests to declare.

AUTHOR AFFILIATIONS

Kris Lockyear orcid.org/0000-0003-2508-168X
Institute of Archaeology, University College London, 31–34
Gordon Square, London, WC1H 0PY, UK

REFERENCES

- Ascher, M** and **Ascher, R.** 1963. Chronological Ordering by Computer. *American Anthropologist*, 65(5): 1045–1052. DOI: <https://doi.org/10.1525/aa.1963.65.5.02a00040>
- Backendorf, D** and **Zimmermann, A.** 1997. Bemerkungen zur Seriation römisch-republikanischer Münztypen. In:
- Müller, J and Zimmermann, A (eds.). *Archäologie und Korrespondenzanalyse. Beispiele, Fragen, Perspektiven*. Internationale Archäologie. Espelkamp: Verlag Marie Ledorf GmbH, pp. 175–178.
- Baxter, MJ.** 1994. *Exploratory Multivariate Analysis in Archaeology*. Edinburgh: Edinburgh University Press.
- Baxter, MJ.** 2003. *Statistics in Archaeology*. London: Arnold.
- Baxter, MJ.** 2014a. *Anglo-Saxon Chronology I — the male graves. A commentary on Chapter 6 of Anglo-Saxon Graves and Grave Goods of the 6th and 7th Centuries AD: A Chronological Framework*. Available at: https://www.academia.edu/5990242/Anglo-Saxon_Chronology_I_the_male_graves_A_commentary_on_Chapter_6_of_Anglo-Saxon_Graves_and_Grave_Goods_of_the_6th_and_7th_Centuries_AD_A_Chronological_Framework [last accessed 18 January 2022].
- Baxter, MJ.** 2014b. *Anglo-Saxon Chronology II — the female graves. A commentary on Chapter 6 of Anglo-Saxon Graves and Grave Goods of the 6th and 7th Centuries AD: A Chronological Framework*. Available at: https://www.academia.edu/5990358/Anglo-Saxon_Chronology_II_-_the_female_graves_A_commentary_on_Chapter_7_of_Anglo-Saxon_Graves_and_Grave_Goods_of_the_6th_and_7th_centuries_AD_A_Chronological_Framework [last accessed 18 January 2022].
- Bayliss, A, Hines, J, Højlund Nielsen, K, McCormac, G** and **Scull, C.** 2013. *Anglo-Saxon Graves and Grave Goods of the 6th and 7th centuries AD: a chronological framework*. London: Society for Medieval Archaeology.
- ter Braak, CJF** and **Šmilauer, P.** 2015. Topics in constrained and unconstrained ordination. *Plant Ecology*, 216(5): 683–696. DOI: <https://doi.org/10.1007/s11258-014-0356-5>
- Brainerd, GW.** 1951. The Place of Chronological Ordering in Archaeological Analysis. *American Antiquity*, 16(4): 301–313. DOI: <https://doi.org/10.2307/276979>
- Buttrey, TV.** 1993. Calculating Ancient Coin Production: Facts and Fantasies. *Numismatic Chronicle*, 153: 335–351.
- Buttrey, TV.** 1994. Calculating Ancient Coin Production II: Why it Cannot be Done. *Numismatic Chronicle*, 154: 341–352.
- Carlson, DL.** 2017. *Quantitative Methods in Archaeology using R*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/9781139628730>
- Crawford, MH.** 1969. *Roman Republican Coin Hoards*. RNS Special Publication 4. London: Royal Numismatic Society.
- Crawford, MH.** 1974. *Roman Republican Coinage*. Cambridge: Cambridge University Press.
- Crawford, MH.** 1990. From Borghesi to Mommsen: The Creation of an Exact Science. In: Crawford, MH, Ligota, CR and Trapp, JB (eds.). *Medals and Coins from Budé to Mommsen*. Warburg Institute Surveys and Texts XXI. London: The Warburg Institute, pp. 125–32.
- Creighton, JD.** 1992. *The Circulation of Money in Roman Britain from the First to the Third Century*. PhD Thesis. Durham: University of Durham.
- Greenacre, MJ.** 1984. *Theory and Applications of Correspondence Analysis*. London: Academic Press.

- Groenen, PJF and Poblome, J.** 2003. Constrained Correspondence Analysis for Seriation in Archaeology Applied to Sagalassos Ceramic Tablewares. In: Schwaiger, M and Opitz, O (eds.). *Exploratory Data Analysis in Empirical Research. Studies in Classification, Data Analysis, and Knowledge Organization*. Berlin and Heidelberg: Springer, pp. 90–97. DOI: https://doi.org/10.1007/978-3-642-55721-7_10
- Hahsler, M, Buchta, C and Hornik, K.** 2018. *seriation: Infrastructure for Ordering Objects Using Seriation*, version 1.2-3. Available at <https://CRAN.R-project.org/package=seriation> [last accessed 18 January 2022]. DOI: <https://doi.org/10.1002/9781119188230.saseas0531>
- Hahsler, M, Hornik, K and Buchta, C.** 2008. Getting things in order: An introduction to the R package seriation. *Journal of Statistical Software*, 25(3): 1–34. DOI: <https://doi.org/10.18637/jss.v025.i03>
- Halekoh, U and Vach, W.** 2004. A Bayesian approach to seriation problems in archaeology. *Computational Statistics and Data Analysis*, 45: 651–73. DOI: [https://doi.org/10.1016/S0167-9473\(03\)00034-3](https://doi.org/10.1016/S0167-9473(03)00034-3)
- Hastie, T and Stuetzle, W.** 1989. Principal Curves. *Journal of the American Statistical Association*, 84(406): 502–16. DOI: <https://doi.org/10.2307/2289936>
- Hill, MO and Gauch, HG.** 1980. Detrended Correspondence Analysis: an Improved Ordination Technique. *Vegetatio*, 42: 47–58. DOI: <https://doi.org/10.1007/BF00048870>
- Hole, F and Shaw, M.** 1967. *Computer Analysis of Chronological Seriation*. Rice University Studies 3. Houston: Rice University.
- Kindt, R and Coe, R.** 2005. *Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies*. World Agroforestry Centre (ICRAF), Nairobi. ISBN 92-9059-179-X.
- Kjeld Jensen, C and Høilund Nielsen, K.** 1997. Burial Data and Correspondence Analysis. In: Kjeld Jensen, C and Høilund Nielsen, K (eds.). *Burial and Society. The chronological and social analysis of archaeological burial data*. Aarhus: Aarhus University Press, pp. 29–61.
- Kuzara, RS, Mead, GR and Dixon, KA.** 1966. Seriation of Anthropological Data: A Computer Program for Matrix-Ordering. *American Anthropologist*, 68(6): 1442–1455. DOI: <https://doi.org/10.1525/aa.1966.68.6.02a00050>
- Lepš, J and Šmilauer, P.** 2003. *Multivariate Analysis of Ecological Data using Canoco*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511615146>
- Lockyear, K.** 1991. Simulating coin hoard formation. In: Lockyear, K and Rahtz, SPQ (eds.). *Computer Applications and Quantitative Methods in Archaeology 1990*. BAR International Series 565. Oxford: Tempus Reparatum, pp. 195–206. DOI: <https://doi.org/10.30861/9780860547136>
- Lockyear, K.** 1996. *Multivariate Money. A statistical analysis of Roman Republican coin hoards with special reference to material from Romania*. PhD Thesis. Institute of Archaeology, University College London.
- Lockyear, K.** 1999. Hoard Structure and Coin Production in Antiquity — an Empirical Investigation. *Numismatic Chronicle*, 159: 215–243.
- Lockyear, K.** 2000. Experiments with Detrended Correspondence Analysis. In: Lockyear, K, Sly, TJT, and Mihăilescu-Bîrliba, V (eds.). *Computer Applications and Quantitative Methods in Archaeology 1996*. BAR International Series 845. Oxford: Archaeopress, pp. 9–17. DOI: <https://doi.org/10.30861/9781841710495>
- Lockyear, K.** 2007. *Patterns and Process in Late Roman Republican Coin Hoards 157–2 BC*. BAR International Series 1733. Oxford: Archaeopress. DOI: <https://doi.org/10.30861/9781407301648>
- Lockyear, K.** 2008. Aspects of Roman Republican coins found in late Iron Age Dacia. In: Spinei, V and Munteanu, L (eds.). *Miscellanea numismatica Antiquitatis. In honorem septagenarii magistri Virgilii Mihăilescu-Bîrliba oblata*. Honoraria. Bucharest: Editura Academiei Române, pp. 147–176.
- Lockyear, K.** 2012. Dating coins, dating with coins. *Oxford Journal of Archaeology*, 31(2): 191–211. DOI: <https://doi.org/10.1111/j.1468-0092.2012.00385.x>
- Lockyear, K.** 2013. Applying bootstrapped Correspondence Analysis to archaeological data. *Journal of Archaeological Science*, 40(12): 4744–4753. DOI: <https://doi.org/10.1016/j.jas.2012.08.035>
- Lockyear, K.** 2016. The Coin Hoards of the Roman Republic Database: The History, the Data, and the Potential. *American Journal of Numismatics*, 28: 157–182.
- Lockyear, K.** 2018. Mind the Gap! Roman Republican Coin Hoards from Italy and Iberia at the End of the Second Century BC. *Numismatic Chronicle*, 178: 123–164.
- Madsen, T (ed.).** 1988. *Multivariate Archaeology. Numerical Approaches to Scandinavian Archaeology*. Moesgård: Jutland Archaeological Society.
- Madsen, T.** 1989. Seriation and multivariate statistics. In: Rahtz, SPQ and Richards, JD (eds.). *Computer Applications and Quantitative Methods in Archaeology 1989*. BAR International Series 548. Oxford: British Archaeological Reports, pp. 205–214.
- Marquardt, WH.** 1978. Advances in Archaeological Seriation. In: Schiffer, MB (ed.). *Advances in Archaeological Method and Theory*. New York: Academic Press, pp. 257–314. DOI: <https://doi.org/10.1016/B978-0-12-003101-6.50014-7>
- Mattingly, HB.** 2004. *From Coins to History: Selected Numismatic Studies*. Ann Arbor: University of Michigan Press. DOI: <https://doi.org/10.3998/mpub.17766>
- McCormick, Jr, WT, Sehweitzer, PJ and White, TW.** 1972. Problem Decomposition and Data Reorganization by a Clustering Technique. *Operations Research*, 20(5): 993–1009. DOI: <https://doi.org/10.1287/opre.20.5.993>
- Melville-Jones, J.** 2009. Review of Patterns and Process in Late Roman Republican Coin hoards 157–2 BC. *Journal of Archaeological Science*, 36(3): 933–934. DOI: <https://doi.org/10.1016/j.jas.2008.11.011>
- Nenadic, O and Greenacre, MJ.** 2007. Correspondence Analysis in R, with two- and three-dimensional graphics: The ca

- package. *Journal of Statistical Software*, 20(3): 1–13. DOI: <https://doi.org/10.18637/jss.v020.i03>
- O'Brien, MJ and Lyman, L.** 1999. *Seriation, stratigraphy, and index fossils: the backbone of archaeological dating*. New York and London: Kluwer Academic/Plenum Publishers.
- Oksanen, J, Blanchet, FG, Friendly, M, Kindt, R, Legendre, P, McGlinn, D, Minchin, PR, O'Hara, RB, Simpson, GL, Solymos, P, Stevens, MHH, Szoecs, E and Wagner, H.** 2017. *vegan: Community Ecology Package*, version 2.4-2. Available at <https://CRAN.R-project.org/package=vegan> [last accessed 18 January 2022].
- Patterson, CC.** 1972. Silver stocks and losses in ancient and modern times. *Economic History Review*, 25: 207–210. DOI: <https://doi.org/10.2307/2593904>
- Petrie, WMF.** 1899. Sequences in Prehistoric Remains. *Journal of the Anthropological Institute*, 29: 295–301. DOI: <https://doi.org/10.2307/2843012>
- Poblome, J and Groenen, PJF.** 2003. Constrained Correspondence Analysis for Seriation of Sagalassos tablewares. In: Doerr, M and Sarris, A (eds.). *The Digital Heritage of Archaeology. CAA2002. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 30th CAA Conference*, Heraklion, Crete, April 2002. Athens: Hellenic Ministry of Culture, pp. 301–306.
- R Core Team.** 2018. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Ringrose, TJ.** 2012. Bootstrap confidence regions for Correspondence Analysis. *Journal of Statistical Computation and Simulation*, 82(10): 1397–1413. DOI: <https://doi.org/10.1080/00949655.2011.579968>
- Robinson, WS.** 1951. A Method for Chronologically Ordering Archaeological Deposits. *American Antiquity*, 16(4): 293–301. DOI: <https://doi.org/10.2307/276978>
- Ryan, NS.** 1982. Characterising fourth century coin loss: an application of Principal Components Analysis to archaeological time-series data. *Science and Archaeology*, 24: 25–32.
- Ryan, NS.** 1988. *Fourth Century Coin Finds in Roman Britain: a Computer Analysis*. BAR British Series 183. Oxford: B.A.R. DOI: <https://doi.org/10.30861/9780860545224>
- Siegmund, F.** 2015. *How to Perform a Correspondence Analysis: a short guide to archaeological practice*. CreateSpace Independent Publishing Platform.
- Thordeman, B.** 1948. The Lohe hoard: a contribution to the methodology of numismatics. *Numismatic Chronicle* 108: 188–204.
- van de Velden, M, Groenen, PJF and Poblome, J.** 2009. Seriation by Constrained Correspondence Analysis: A simulation study. *Computational Statistics and Data Analysis*, 53: 3129–3138. DOI: <https://doi.org/10.1016/j.csda.2008.08.020>

TO CITE THIS ARTICLE:

Lockyear, K. 2022. Simulation, Seriation and the Dating of Roman Republican Coins. *Journal of Computer Applications in Archaeology*, 5(1), pp. 1–18. DOI: <https://doi.org/10.5334/jcaa.57>

Submitted: 01 April 2020 Accepted: 16 January 2022 Published: 08 February 2022

COPYRIGHT:

© 2022 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Journal of Computer Applications in Archaeology is a peer-reviewed open access journal published by Ubiquity Press.