27

Analysis of multidimensional matrices for archaeological data

John Wilcock

(School of Computing, Staffordshire University, Stafford, UK)

27.1. Introduction

Seriation has a long history, going back to Flinders Petrie (1899), who worked on pre-Dynastic Egyptian cemeteries. He studied the pottery and other finds from over 900 graves at Naqada, in the valley of the Nile, and attempted to put them into chronological order using the distribution of pottery types alone. The result was eighteen groups of graves, called "sequence dates", which formed an ordered list of relative dates. Petrie did all this using slips of paper, which he ordered on the floor of a warehouse, effectively inventing the "incidence matrix" and a matrix re-ordering methodology which was later to be employed in the days of computers. Petrie's methodology depended on what Kendall (1963; 1971) calls the "Concentration Principle":

> '*if* the typology is "chronologically significant", and when the graves have been correctly ordered (or anti-ordered), then the "sequence-date" *ranges* for the individual types will be found to have been individually *or in some communal way* minimised' (Kendall's italics).

This is based on the general principle that each pottery type is manufactured over a restricted period of time, before and after which it is not made, but during which it is commonly available. Graves containing each type of pottery should therefore be placed in the narrowest possible date sequence, i.e. the sequence-date range (the last "date" minus the first "date") for the pottery type must be minimised, and this must be done for all the pottery types as far as possible. A matrix which has been re-ordered in this form has been referred to by Kendall as being in *Petrie form*, or *petrifiable* (Kendall 1969) and the operation itself as "Petrifaction". Kendall developed the HORSHU algorithm from this idea, so-called because the La Tène Iron Age cemetery test data from Münsingen-Rain (Switzerland) were arranged in a coordinate plot by MDSCAL which was horseshoe-shaped, and the seriation appeared linearly along the horseshoe from one end to the other (Kendall 1971). The starting point was an "abundance matrix", an incidence matrix of grave number *v.* artefact type number where all the elements (artefact counts) in the matrix are known. Kendall and Wilkinson (Wilkinson 1971; 1974) later showed that the horseshoe could be "unbent" to a straight line seriation using the algorithm DIM/DROP, a procedure they referred to as "Operation Speckled Band", a literary reference to the well-known story by Conan Doyle in which Dr Grimesby Roylott bends a poker into a curve, with a threatening manner, and Sherlock Holmes straightens it out again.

Seriation studies were at the forefront of quantitative archaeological research in the 1950s and early 1960s. The incidence matrix and its derived symmetric similarity matrices have been employed commonly for the analysis of archaeological data since the times of Brainerd and Robinson (Brainerd 1951; Robinson 1951; Brainerd & Robinson 1952), when Robinson, a statistician, was asked by Brainerd, an archaeologist, to provide a numerical method for the seriation of pottery assemblages. There is an extensive literature on the subject of seriation (see for example Goldmann 1971; 1972; Cowgill 1972; Doran & Hodson 1975; Marquardt 1978; Ester 1981). Doubts about the validity of certain seriation methods have led to various simulation studies in more recent years (see for example Graham *et al.* 1976; Guio & Secco 1984; Laxton 1987; and Herzog & Scollar 1988, during work on the Bonn Seriation and Clustering Package). Extensive trials with Hallstatt Iron Age material has helped in the design of the IAGRAVES seriation package at the Institute of Archaeology, University of London (Hodson 1988).

27.2. Matrix analysis

The term "matrix analysis" has become associated in archaeology with a procedure first suggested by Tugby (1958), a development of the work of Brainerd and Robinson, and used by Clarke (1962; 1966; 1970), for example, in the seriation of British Beaker pottery. The method forces a linear seriation of the two types of entities being employed, which could be sites *v.* assemblages, for instance, or artefacts *v.* properties.

Let us suppose there are n superior entities *v.* m inferior entities in the incidence matrix. The similarity study which then compares the superior entities with each other, producing an $n \times n$ symmetric matrix, is referred to as the Q technique, while that which compares the inferior entities with each other, producing an $m \times m$ symmetric matrix, is referred to as the R technique. Whichever is employed, the entities are seriated linearly, and the matrix is re-ordered in both rows and columns according to this seriation (naturally, if the data do not fit a linear seriation, e.g. if there is a branched minimum spanning tree, the results will be less than perfect). A coincidental by-product of this seriation is that all the highest similarity coefficients migrate towards the central diagonal.

However, little attention has been paid to the meaning of this process in an archaeological sense, beyond the obvious one of dating objects relatively, or trying to spot an evolutionary sequence. Nor has much attention been paid to objective measures of perfection for estimating the success of the matrix reordering process, and matrices of three or more dimensions have rarely been studied.

This paper proposes several new methods for:

- matrix reduction
- matrix seriation where the criterion for reordering is not similarity measures *per se*, but a positive de-

sire to place the largest entry values in the matrix near to the central diagonal
- measuring the success of the matrix reordering ("Perfection Coefficient")
- extending the methods to three or more dimensions in the incidence matrix, yielding a generic technique.

27.3. Matrix reduction

Storage required for matrix analysis in general mounts as N^D where N is the size of the largest dimension and there are D dimensions. For large matrices any computer may run out of memory, thus the incidence matrix, which is a non-symmetric matrix of raw data, may have to be reduced before it can be analysed in the available memory space.

For two dimensions the incidence matrix typically expresses items in terms of properties, but it is intended that all the algorithms described in this paper should be applicable to any number of dimensions greater than or equal to two. For a 2-dimensional matrix, rows or columns may be treated as sub-matrices. For a 3-dimensional matrix, the sub-matrices are 2-dimensional planes, and so on. For D dimensions, each sub-matrix will have (D – 1) dimensions. A suitable algorithm is proposed as follows:

1. Choose dimension to be reduced
2. Count the number of non-blank entries in each sub-matrix, and record the minimum entry count
3. If any sub-matrices have zero entry counts, delete these sub-matrices, then repeat from step 1 above
4. Ask if sub-matrices with the minimum non-zero count are to be deleted; if YES, delete these sub-matrices and repeat from step 1; if NO, terminate the procedure.
5. Deleted sub-matrices should be identified by a print-out or other indication to the user. For a 2-dimensional matrix it would be usual to reduce alternate dimensions. For three or more dimensions, any possible sequence of dimension reductions should be allowed, to suit the data. Having reduced one dimension, the algorithm would normally continue until all the remaining dimensions have been tested, and there has been a positive decision *not* to reduce them.

There is, however, a danger that the "boy scout and potato" problem will arise, i.e. the boy scout with muddy fingers peels all the mud off the potato until nothing is left. By deleting sub-matrices which are below the selected minimum count in one dimension, sub-matrices in another dimension may be brought below the minimum count and deleted, and so on until the whole matrix disappears. This will be more applicable to very sparse matrices, and may not arise significantly for abundance matrices, but nevertheless the problem must be recognised and guarded against.

27.4. Matrix seriation

27.4.1. 2-dimensional matrices

The "essence" of the data may be revealed by seriating the incidence matrix, so that in a 2-dimensional matrix similar higher order entities (items) will occur close together on one axis, and similar lower order entities (properties) will also occur close together on the second axis. In a 2-dimensional matrix this involves reordering the rows, followed by reordering the columns, or vice versa, in a repeating sequence (see example in Appendix 27.1). This is best done on the non-symmetric incidence matrix, but for some reason several workers have not done this on the raw data but have operated on either the Q or R derived square similarity/ dissimilarity matrices, thereby discarding already some information which is relevant to the seriation.

The procedure of re-ordering a matrix is well-known, and has been embodied in computer programs by Ascher & Ascher (1963), Kuzara, Mead & Dixon (1966), Hole & Shaw (1967) and Craytor & Johnson (1968). The main problems with these algorithms is that, unless there is a perfect linear seriation inherent in the data, the outcome of the procedure depends on:

1. the initial starting position of the data (for this reason several runs may be necessary with different randomised starting orders)
2. whether the rows of the matrix or the columns of the matrix are re-ordered first (the outcome is often different)
3. whether there are local maxima (optima) in which the sequence can become trapped
4. whether there is a cycle, i.e. a return to a repetitive sequence of transitions — in this case the place to stop must be decided.

For these reasons the use of seriation has become more cautious, as rarely in archaeological problems does the data exhibit a perfect linear seriation. Non-metric Multidimensional Scaling and Principal Components Analysis, which do not suffer from these problems, have therefore tended to replace Seriation in archaeological multivariate analysis.

A form of iterative heuristic search is often employed in seriation studies, with an initial re-ordering of the rows and columns according to selected pseudo-random numbers. For successive runs, small changes are then made in the initial order, as dictated by a number of pre-selected sequence modification operators, and the final results are compared. Any change which results in a better seriation is retained, otherwise it is discarded. The complete set of operators is repeatedly scanned until there is no improvement, and the final result is retained as a local maximum. This complete procedure is repeated a number of times with different random start configurations. Finally the best local maximum is returned as the result — there is no guarantee that this solution will really be the best possible seriation, since the choice of operators is arbitrary. One reason for poor success would of course be that the data does not really fit a linear seriation, and should instead be represented by some form of minimum spanning tree.

27.4.2. 3-dimensional matrices

There has not been much consideration in previous work of incidence matrices of three or more dimensions — these higher dimensions of data have been treated by Non-Metric Multidimensional Scaling and Principal Components Analysis.

In a 3-dimensional matrix the dimensions may be considered as planes, rows and columns. In place of the alternate re-ordering of rows (1) followed by re-ordering of columns (2) in a 2-dimensional matrix, the planes (1), rows (2) and columns (3) are re-ordered in turn. However, where there are only the 1-2 or 2-1 sequences possible for two dimensions, there are six possible sequences for 3 dimensions: 1-2-3, 1-3-2, 2-1-3, 2-3-1, 3-1-2 and 3-2-1. For a general n-dimensional matrix all the *n!* sequences of repeated re-ordering must be examined in turn, to see which gives the best final result. The high values which migrate to the central diagonal (top left to bottom right of a 2-dimensional matrix) will have been replaced by high values on a central diagonal between opposite corners of a cuboid for a 3-dimensional matrix. The diagonal is impossible to perceive for four or more dimensions, since it is in hyperspace, but the generic method remains valid for these higher number of dimensions. The interesting feature is that all the dimensions have been seriated simultaneously by the method.

The proposed algorithm is as follows:

1. Determine by user parameter whether zero entries (as opposed to blank entries) are to be ignored, i.e. treated as blanks, or are to be treated as valid. This will affect the re-ordering process
2. Calculate the data cell positions on the central diagonal; the number of such cell positions N_p is:

$$\text{INT}(1 + \sqrt{\ } ((N_1 - 1)^2 + (N_2 - 1)^2 + (N_3 - 1)^2 + ...))$$

where N_1 is the size of dimension 1, N_2 is the size of dimension 2, N_3 is the size of dimension 3, ... and so on up to the number of required dimensions.

3. The first cell position on the central diagonal has coordinates (1, 1, 1, ...). Each subsequent central diagonal position numbered D_p (range 2 to N_p) has a coordinate in each dimension of:

$$\text{INT}(1.5 + ((D_p - 1)/(N_p - 1)) \times (N - 1))$$

where N is the size of the dimension. Note that the expression gives rounding to the nearest integer cell coordinate, with resultant integer values from 1 to N. There must be no gaps between diagonal cells, and this formula guarantees sufficient overlap.

4. Next, the nearest central diagonal cell to each of the remaining cells must be found. A true Euclidean distance is employed. If (C_1, C_2, C_3, ...) are the coordinates of a cell and (P_1, P_2, P_3, ...) the coordinates of a central diagonal cell under consideration, the Euclidean distance is:

$$\sqrt{\ }((C_1 - P_1)^2 + (C_2 - P_2)^2 + (C_3 - P_3)^2 + ...)$$

The minimum such distance is taken, i.e. each matrix cell is related to its nearest central diagonal cell. If two or more central diagonal cells give the same minimum distance, the first (i.e. lowest-numbered) such cell is taken. A typical example is given in Appendix 27.2.

5. The tree route from each cell to the nearest central diagonal cell is determined. The route will be via a tree of other cells which are connected to the same central diagonal cell. This is best done by using a stack organisation, with a stack pointer SP and a memory structure ST(SP).
   - i) We commence with SP=0.
   - ii) Starting at each central diagonal position in turn, set current coordinates to those of the central diagonal cell.
   - iii) Each current cell has in general $(3^D - 1)$ adjacent cells to consider, where D is the number of dimensions, e.g. a 2-dimensional cell has $(3^2 - 1)$ or 8 nearest neighbour cells, a 3-dimensional cell has $(3^3 - 1)$ or 26 nearest neighbours, while a 4-dimensional cell has no less than $(3^4 - 1)$ or 80 nearest neighbours.
   - iv) In each dimension coordinates may vary unit distance 0, +1 or –1 from the current cell coordinates. Edge effects must be detected, i.e. if a current coordinate in some dimension is 1, then the cell at (coordinate – 1) does not exist, while if a current coordinate is at the maximum N for some dimension, then the cell at (coordinate + 1) does not exist. Considering each possible adjacent cell in turn, if that cell has been recorded as being connected to the current central diagonal cell, then:
     - Record route for this cell as being from the current cell (C_1, C_2, C_3, ...)
     - Increment SP
     - Stack (push) adjacent cell coordinates for further consideration into ST(SP)
   - v) When all possible adjacent cells have been considered:
     If SP=0 then repeat for next central diagonal cell, ELSE
     - Unstack (pop) a set of coordinates from the stack and set the current coordinates to them
     - Decrement SP
     - Repeat consideration of adjacent cells for the new current coordinates
   - vi) In this way, a tree is built up from each central diagonal cell to all the cells which are nearest to it. This route is of importance for the "perfection coefficient" calculation proposed below, and a typical example is illustrated in Appendix 3.
6. All possible reordering regimes for the matrix must then be considered. In general there are D! possible reordering sequences for the D dimensions. All must be considered, each running to a stable configuration or to a cycle which is detected. The re-ordering algorithm for the current dimension is as follows:

i) Clear a sum and a count to zero for each sub-matrix in the current dimension (for D dimensions, each sub-matrix has (D – 1) dimensions). Blank value entries are ignored as irrelevant in the following calculation. It will have been already decided whether zero entries are relevant: if they are relevant, they are treated like any other value number, otherwise they are ignored, i.e. treated as blanks.

ii) For each sub-matrix, form a product sum which consists of the relevant value entries multiplied by weights for the position in each remaining dimension, e.g. for an entry in row 3, column 5 of a 2-dimensional sub-matrix, the product will be (value of entry)×3×5. There will be (D – 1) such weights in the product. The product is added to the sum, and the count is increased by (value of entry). When all relevant entries have had weighted products calculated and summed, a quotient is formed for each sub-matrix, calculated as (sum / count).

iii) Use the quotients to reorder the complete sub-matrices. If no reordering is necessary, the current reordering sequence terminates, otherwise the procedure continues for the next dimension in the current sequence.

iv) When a particular sequence is terminated, the next sequence is processed, until all D! sequences have been run.

27.5. How good is the result?

27.5.1. A previous method: the Concentration Principle

An objective measure of the success of the reordering of the matrix is required. A simple way to express the Concentration Principle as an objective measure is to sum the ranges of positions of valid entries in each column (or each row). These two results will, however, be different for a non-symmetric incidence matrix; only for a symmetrical similarity matrix will it not matter whether the columns or the rows are considered for this range test. Thus if the function to be minimised is the sum of ranges over the columns, the expression is:

$$S_{cols} R_i$$

where R_i is the difference in row positions for the first and last valid entries in a column.

A more complex mathematical formulation of the Concentration Principle was derived by Kendall (1963) on grounds of probability and maximum likelihood. The criterion is:

$$S_{cols} N_i \log R_i$$

where R_i is the range of entries in column i, and N_i is the number of artefacts of type i which are recorded in the matrix.

27.5.2. The "Perfection Coefficient"

In this paper a new objective method is proposed where the criterion for determining success in the seriation is the proximity of the largest entry values to the central diagonal. The measure is termed the "Perfection Coefficient". The method consists of moving from each cell of the matrix to its next tree connection cell towards the central diagonal, and noting what happens to the entry values in the two cells. If the entry value does not decrease, i.e. if the entry value of the cell nearer to the central diagonal, along the connected tree to the nearest central diagonal cell, is greater than or equal to the entry value of the cell further away from the central diagonal, then a count is increased by one. This is done for every cell apart from the central diagonal cells themselves. If C is the count of the transitions where there is no decrease, then the perfection coefficient is calculated as:

$$PC = 100 \times C / (\text{total number of transitions})$$

27.6. More than three dimensions

Problems with four or more dimensions may still be handled by the proposed algorithms, which are completely generic in nature. The configuration may no longer be represented as a geometrical model, since it is in hyperspace, but the methodology may still be applied, and the interesting point to note is that all the dimensions are seriated simultaneously.

In fact, problems with more than two dimensions are commonplace in archaeology. We need only to add the *time* dimension to see this. An object may have a different "value" in each time period, and what we then have is a "time sequence collection", typically with observations of values at discrete irregular intervals. In fact the variable being measured may be "sampled" continuously, or at regular or irregular discrete intervals, and the variable itself may exhibit continuously smooth changes, or stepwise changes in value.

A hierarchy of typical dimensions applicable in archaeology is:

- Time
- Culture
- Site
- Phase
- Assemblage
- Artefact
- Property

i.e. at least seven dimensions could be considered. Each of these dimensions requires a framework which typically includes the dimension on each side of it in the hierarchy, e.g. a culture is meaningless without being defined in terms of some time period and some geographical location (sites or areas). For example, the Palaeolithic culture, often thought of as being in the remote past, still exists in the present in some regions of the Amazonian rain forest and in Australia. Even the 2-dimensional studies which have dominated the literature are in reality 3-dimensional, for they have considered a particular culture such as the British Beaker period or the La Tène Iron Age period in addition to the pottery *v.* properties or graves *v.* artefacts.

27.7. Testing the methodology

The methods have been tested on numerous randomly-generated matrices of more than two dimensions. As typical archaeological information the Münsingen-Rain data (Hodson 1968) was employed. It was easy to make three dimensions from this, using Hodson's "horizons" (phases), in addition to the graves and artefacts. The results, with a horizon in each plane, and a final central diagonal from the earliest grave of the first horizon to the latest grave of the last horizon, were found to be comparable with Hodson's original sequence.

27.8. Conclusions

The paper has presented a new suggested methodology for seriating any number of dimensions of information, specifically more than two dimensions. The algorithms presented are for:

- matrix reduction
- matrix seriation
- an objective "Perfection Coefficient" to measure the success of the reordering of the matrix

The methods therefore constitute a set of generic techniques for any number of dimensions, and they have been tested on typical archaeological data sets.

Appendix 27.1. Reordering a typical incidence matrix

Consider the following incidence matrix, for a small set of data (5 items, 4 properties):

		Items					Row sums
		A	B	C	D	E	
Properties	P	5	1	3	2	4	44/15=2.93
	Q	4	2	1	3	5	48/15=3.2
	R	3	5	2	4	1	40/15=2.67
	S	2	4	5	1	3	44/15=2.93
Col. wts.		1	2	3	4	5	

For reordering the rows first, each column is given a column weight, and the row sums are calculated as:

$$\frac{\sum_{i=1}^{c} v_i w_i}{\sum_{i=1}^{c} v_i}$$

where values v are multiplied by column weights w for all the columns. Rows are then re-ordered according to these sums:

		Items					Row weights
		A	B	C	D	E	
Properties	R	3	5	2	4	1	1
	P	5	1	3	2	4	2
	S	2	4	5	1	3	3
	Q	4	2	1	3	5	4
Column sums		35/14	27/12	27/11	23/10	38/13	
		=2.5	=2.25	=2.45	=2.3	=2.92	

Row weights have now been allocated, and column sums calculated in a similar fashion. Columns are then re-ordered according to these sums:

		Items					Row sums
		B	D	C	A	E	
Properties	R	5	4	2	3	1	36/15=2.4
	P	1	2	3	5	4	54/15=3.6
	S	4	1	5	2	3	44/15=2.93
	Q	2	3	1	4	5	52/15=3.47
Column weights		1	2	3	4	5	

Rows are re-ordered:

		Items					Row weights
		B	D	C	A	E	
Properties	R	5	4	2	3	1	1
	S	4	1	5	2	3	2
	Q	2	3	1	4	5	3
	P	1	2	3	5	4	4
Column sums		23/12	23/10	27/11	39/14	38/13	
		=1.92	=2.3	=2.45	=2.79	=2.92	

The algorithm terminates here, as the column sums are in the correct order. Note that the higher values have migrated towards the central diagonal.

However, reordering the columns first from the original incidence matrix gives:

		Items					Row weights
		A	B	C	D	E	
Properties	P	5	1	3	2	4	1
	Q	4	2	1	3	5	2
	R	3	5	2	4	1	3
	S	2	4	5	1	3	4
Column sums		30/14	36/12	31/11	24/10	29/13	
		=2.14	=3.0	=2.81	=2.4	=2.23	

Reordering the columns gives:

		Items					Row sums
		A	E	D	C	B	
Properties	P	5	4	2	3	1	36/15=2.4
	Q	4	5	3	1	2	37/15=2.47
	R	3	1	4	2	5	50/15=3.33
	S	2	3	1	5	4	51/15=3.4
Column weights		1	2	3	4	5	

The algorithm terminates here, as the row sums are in the correct order. Note that the higher values have migrated towards the central diagonal. However, this is a different result to that obtained by reordering rows first. The outcome of the procedure depends on:

1. the initial starting position of the data (different randomised starting positions should be tried for both columns and rows). In the example only one starting position has been shown.
2. whether the rows of the matrix, or the columns of the matrix are re-ordered first (the outcome has been shown to be different for the data above)

3. whether there are local maxima (optima) in which the sequence can become trapped. The two results above may or may not be local maxima.
4. whether there is a cycle, i.e. a return to a repetitive sequence of transitions. In that event the place to stop must be decided. No such cycle has occurred for the data above. It is clear that a form of perfection coefficient is necessary to choose between the different results that will be obtained from different starting configurations and different sequences of dimension reordering.

Appendix 27.2. A typical 3-dimensional matrix

This is a typical example of a 3-dimensional matrix of 4 planes, 5 rows and 6 columns. The four separate planes are shown below, with the central diagonal cells outlined heavily in black. The remaining cells have two numbers, the first giving the central diagonal cell to which that cell is nearest, and the second giving the coordinates of the cell to which the cell is connected along the tree to the central diagonal cell (in terms of plane, row, column).

Number of cells on central diagonal

$$= \text{INT}(1 + \sqrt{(4-1)^2 + (5-1)^2 + (6-1)^2})$$

$$= \text{INT}(1 + \sqrt{9+16+25})$$

$$= \text{INT}(1 + \sqrt{50})$$

$$= 8$$

The first central diagonal cell is 111. Each subsequent central diagonal cell has a coordinate in each dimension

$$= \text{INT}(1.5 + ((D_p - 1)/(N_p - 1)) \times (N - 1))$$

where D_p is the number of the central diagonal cell, from 2 to 8, N_p is 8, and N is the size of the dimension. This results in the following coordinates:

1	111	2	122	3	222	4	233
5	334	6	345	7	445	8	456

1	1 111	2 122	2 113	4 125	5 226
1 111	2	2 122	4 233	4 135	5 226
2 122	2 122	4 233	4 233	4 144	6 145
2 131	4 233	4 233	4 233	6 256	6 256
4 152	4 253	4 253	4 253	6 256	6 256

Plane 1

1 111	3 222	3 222	3 313	5 316	5 316
3 222	3	3 222	4 233	5 334	5 316
3 222	3 222	4	4 233	5 334	6 345
3 331	4 233	4 233	4 233	6 345	6 345
4 342	4 342	4 342	6 345	6 345	6 345

Plane 2

3 222	3 222	3 222	5 425	5 425	5 425
3 222	3 222	3 222	5 334	5 334	5 425
3 222	3 222	4 233	5	5 334	6 345
3 331	4 233	4 233	5 334	6	6 345
4 342	4 342	4 342	6 345	6 345	8 456

Plane 3

3 422	3 422	3 422	5 425	5 425	5 425
3 331	3 331	5 334	5 334	5 334	7 436
3 331	3 331	5 334	5 334	7 445	7 445
3 331	4 342	5 334	7 445	7	7 445
4 342	4 342	7 454	7 445	7 445	8

Plane 4

Appendix 27.3. A typical Perfection Coefficient calculation (2D)

This is a typical Perfection Coefficient calculation for a simple 2-dimensional matrix with 4 rows and 5 columns.

Number of central diagonal cells:

$$= \text{INT}(1 + \sqrt{(4-1)^2 + (5-1)^2})$$

$$= \text{INT}(1 + \sqrt{9+16})$$

$$= \text{INT}(1 + \sqrt{25})$$

$$= 6$$

The first cell position on the central diagonal is 11. Each subsequent central diagonal cell position has a coordinate in each dimension of

$$\text{INT}(1.5 + ((D_p - 1)/(N_p - 1)) \times (N - 1))$$

where D_p increments from 2 to 6, $N_p = 6$, and N is the size of the dimension. The central diagonal cell coordinates are:

1	11	2	22	3	23
4	33	5	34	6	45

The matrix below shows the central diagonal cells outlined heavily. The remaining cells have two numbers giving the central diagonal cell which is nearest, and the tree route which is necessary for the calculation of the Perfection Coefficient:

1	1 11	3 23	3 23	3 14
1 11	2	3	3 23	5 34
2 22	2 22	4	5	5 34
2 31	4 33	4 33	5 34	6

If the values in the seriated matrix are as follows:

5	4	2	3	1
4	5	3	1	2
3	1	4	2	5 **
2	3	1	5 **	4

then only the double-asterisked cells have a higher value than the cell to which the tree directs them. The total number of transitions on the tree is 13, hence the Perfection Coefficient

= 100 × (number of transitions where the value does not decrease) / (total number of transitions)

= 100×11/13 = 84.6%

Bibliography

ASCHER, R. & M. ASCHER 1963. "Chronological ordering by computer", *American Anthropology* 65: 1045–1052.

BRAINERD, G. W. 1951. "The place of chronological ordering in archaeological analysis", *American Antiquity* 16: 301–313.

CLARKE, D. L. 1962. "Matrix analysis and archaeology with particular reference to British Beaker pottery", *Proceedings of the Prehistoric Society* 28: 371–382.

CLARKE, D. L. 1966. "A tentative reclassification of British Beaker pottery in the light of recent research", *Palaeohistoria* 12: 179–198.

CLARKE, D. L. 1970. *Beaker potter of Great Britain and Ireland.* Cambridge University Press, Cambridge.

COWGILL, G. L. 1972. "Models, methods and techniques for seriation", *in* Clarke, D. L. (ed.), *Models in archaeology*, pp. 381–424. Methuen, London.

CRAYTOR, W. B. & L. JOHNSON 1968. *Refinements in computerized item seriation, Bulletin* 10, Museum of Natural History, University of Oregon.

DORAN, J. E. & F. R. HODSON 1975. *Mathematics and computers in archaeology*. Edinburgh University Press, Edinburgh.

ESTER, M. 1981. "A column-wise approach to seriation", *American Antiquity* 46: 496–512.

GOLDMANN, K. 1971. "Some archaeological criteria for chronological seriation", *in* Hodson, F. R., D. G. Kendall & P. Tautu (eds), *Mathematics in the Archaeological and Historical Sciences*, pp. 202–208. Edinburgh University Press, Edinburgh.

GOLDMANN, K. 1972. "Zwei Methoden chronologischer Gruppierung", *Acta Praehistorica et Archaeologica* 3: 1–34.

GRAHAM, I., P. GALLOWAY & I. SCOLLAR 1976. "Model studies in computer seriation", *Journal of Archaeological Science* 3: 1–30.

GUIO, A. D. & G. SECCO 1984. "A new computer seriation algorithm", *in* Laflin, S. (ed.), *Computer Applications in Archaeology 1984*, pp. 199–209. University of Birmingham.

HERZOG, I. & I. SCOLLAR 1988. "A mathematical basis for simulation of seriatable data", *in* Rahtz, S. P. Q. (ed.), *Computer and Quantitative Methods in Archaeology 1988*, pp. 53–62 , British Archaeological Reports International Series 446, Oxford.

HODSON, F. R. 1968. *The La Tène cemetery at Münsingen-Rain*. Acta Bernensia V, Stämpfli, Bern.

HODSON, F. R. 1988. *Hallstatt, the Ramsauer graves: Quantification and analysis*. Society of Antiquaries, London.

HOLE, F. & M. SHAW 1967. "Computer analysis of chronological seriation", *Rice University Studies* 53(3): 1–166. Rice University, Houston.

KENDALL, D. G. 1963. "A statistical approach to Flinders Petrie's sequence-dating", *Bull. I.S.I. 34th session, Ottawa*, pp. 657–680.

KENDALL, D. G. 1969. "Some problems and methods in statistical archaeology", *World Archaeology* 1(1): 68–76.

KENDALL, D. G. 1971. "Seriation from abundance matrices", *in* Hodson, F. R., D. G. Kendall & P. Tautu (eds), *Mathematics in the Archaeological and Historical Sciences*, pp. 215–252. Edinburgh University Press, Edinburgh,

KUZARA, R. S., G. R. MEAD & K. A. DIXON 1966. "Seriation of anthropological data: a computer program for matrix ordering", *American Anthropologist* 68(6): 1442–1455.

LAXTON, R. R. 1987. "Some results on mathematical seriation with applications", *in* Ruggles, C. L. N. & S. P. Q. Rahtz (eds), *Computer and Quantitative Methods in Archaeology 1987*, pp. 39–44. British Archaeological Reports International Series 393, Oxford.

MARQUARDT, W. 1978. "Advances in archaeological seriation", *in* Schiffer, M. (ed.), *Advances in archaeological method and theory* 1, pp. 257–314. Academic Press, New York.

PETRIE, W. M. F. 1899. "Sequences in prehistoric remains", *Journal of the Anthropological Institute* 29: 295–301.

ROBINSON, W. S. 1951. "A method for chronologically ordering archaeological deposits", *American Antiquity* 16: 293–301.

ROBINSON, W. S. & G. W. BRAINERD 1952. "Robinson's coefficient of agreement — a rejoinder", *American Antiquity* 18: 60–61.

TUGBY, D. J. 1958. "A typological analysis of axes and choppers from southeast Australia", *American Antiquity* 24: 24–33.

WILKINSON, E. M. 1971. "Archaeological seriation and the travelling salesman problem", *in* Hodson, F. R., D. G. Kendall & P. Tautu (eds), *Mathematics in the Archaeological and Historical Sciences*, pp. 276–284. Edinburgh University Press, Edinburgh.

WILKINSON, E. M. 1974. "Techniques of data analysis: seriation theory", *Archaeo-Physika* 5: 31–42.

Dr J. D. Wilcock,
Staffordshire University,
School of Computing,
The Octagon,
Beaconside,
GB-ST18 0AD Stafford

cmtjdw @ staffs.ac.uk