

22

Seriation by similarity and consistency

R. R. Laxton*

J. Restorick†

22.1 Introduction

There are two main theoretical problems concerning the seriation of abundance or incidence matrices whose data are associated with changes in use or fashion of artifact types over a period of time (cf. Doran & Hodson 1975, p. 267–284, Kendall 1969, Kendall 1971, Orton 1980, p. 65–105, Laxton 1976, Laxton 1987).

1. To give criteria which determines if such a matrix can be seriated, i.e., if it is pre- Q .
2. To obtain methods which seriate rapidly a pre- Q matrix, i.e., which permutes its rows so that the resulting matrix is a Q -matrix (cf. Fig. 22.1).

Furthermore, any method of seriation should be able to deal effectively with *real* archaeological data associated with chronology. That is, provided the data approximates the model for changes in artifact types over a period of time and has not been too disturbed subsequently, the seriation should produce a matrix which is recognisable as approximating a Q -matrix. The resulting order of its rows will then still be a candidate for the chronological order of the corresponding provinces. Furthermore, as Kendall has remarked (1971), the method should be allowed to fail with data which are not associated with chronological ordering or, if the 'noise' in the data is too great, to get even an approximate answer.

There are, of course, several well-known method of seriation—multidimensional scaling (e.g. Kendall 1971) and correspondence analysis—to name but two (for a fairly up-to-date set of applications to seriation and an extensive bibliography, see Madsen 1988; for the theory see Greenacre 1984). But these use standard metrics, which may not be appropriate in particular cases, piecewise linearization and data compression for the interpretation. Thus in the latter, the data are often represented as points in a 2-dimensional space and the ordering is obtained from this pictorial representation. Both these methods have, of course, wide applicability in multidimensional analysis and not just to seriation (cf. Djindjian 1988). But if the ultimate aim is to identify the chronological influence on the data—especially from other influences such as wealth or status, stochastic movement under gravity, etc.,—then it seems necessary to have a method of seriation which adheres rigidly to the model of chronological change. This is the aim here; to present a rapid method of seriation based strictly on this model of change. (When it comes to considering other influences on the data, these too must be modelled appropriately, together with their interactions; we hope to turn to this problem shortly.) Perhaps the rapid methods of Gelfand (1971) are more in the spirit of the method presented here.

* Department of Mathematics
University of Nottingham
Nottingham NG7 2RD

† Department of Education
University of Nottingham
Nottingham NG7 2RD

Level 1 (top)	4	6	86	0	4
2	6	14	76	0	4
3	8	19	70	1	2
4	18	49	30	3	0
5	23	54	20	3	0
6	32	49	5	14	0
7	39	43	0	18	0
8	49	30	0	21	0
9	62	20	0	18	0
10	87	3	0	10	0

Table 22.1: An abundance Q -matrix of the ceramic contents in levels in stratigraphic order from a trash mound at Awatovi, Northeastern Arizona (Burgh 1959). In no column do the values strictly decrease and then strictly increase again. Entries in each row are percentages of totals in the corresponding level.

	1	2	3	4	5	6	7	8	9	10
1	100									
2	89	99								
3	83	93	100							
4	40	50	58	99						
5	30	40	48	89	100					
6	15	25	33	74	79	99				
7	10	20	28	64	69	89	100			
8	10	20	28	51	56	76	87	100		
9	10	20	28	41	46	66	77	87	100	
10	7	9	12	24	29	45	52	62	75	100

Table 22.2: The similarity R -matrix of the abundance matrix of Table 22.1. The coefficient of similarity used is Kendall's (1971); the common content measure. The similarity values increase to the diagonal and then decrease again (the matrix is symmetric).

22.2 Some results and examples

We say that a set of rows a_i, \dots, a_j of an abundance/incidence matrix A is *consistent* in this order if they form a Q -matrix. The method of seriation offered here depends on the following observations.

Result 1. *If $a_u, \dots, a_i, \dots, a_j, \dots, a_v$ are consistent in this order and the rows a_i, \dots, a_j are replaced by their average $a_{(i,j)}$, then the rows $a_u, \dots, a_{i-1}, a_{(i,j)}, a_{j+1}, \dots, a_v$ are also consistent in this order.*

Example In the 10×5 Awatovi data matrix, if rows 5, 6 and 7 are replaced by their average (31.3, 48.3, 8.3, 11.7, 0) thus,

1	4	6	86	0	4
2	6	14	75	0	4
3	8	19	71	1	2
4	18	48	30	3	0
(5, 6, 7)	31.3	48.3	8.3	11.7	0
8	49	30	0	21	0
9	62	20	0	18	0
10	87	3	0	10	0

then this matrix is also a Q -matrix.

Result 2. More generally, if $a_u, \dots, a_{i-1}, a_i, \dots, a_j, a_{j+1}, \dots, a_v$ are consistent in this order and the rows a_u, \dots, a_{i-1} are replaced by their average b_1 , a_i, \dots, a_j by their average b_2 and a_{j+1}, \dots, a_v by their average b_3 , then b_1, b_2, b_3 is also a consistent set in this order. x

Example The average of the rows 1, 2, 3 and 4 of the Awatovi data matrix is (9.0, 13.0, 77.3, 0.3, 3.3) and that of the rows 8, 9 and 10 is (66.0, 17.7, 0, 16.3, 0). Then the 3×5 matrix

(1, 2, 3, 4)	9.0	13.0	77.3	0.3	3.3
(5, 6, 7)	31.3	48.3	8.3	11.7	0
(8, 9, 10)	66.0	17.7	0	16.3	0

is also a Q -matrix. The following result concerns the similarities between the rows of an abundance/incidence matrix. The similarity measure is that of Kendall, the common content of the two rows. We know from Kendall's work (1971) that if a_i, a_j, a_k is a consistent set of rows in this order, then

$$S(a_i, a_k) \leq S(a_i, a_j) \text{ and } S(a_j, a_k);$$

i.e., the similarity between the two outside rows is not greater than that between them and the row between in the consistent order. This is, of course, why the similarity matrix in Fig. 22.2 of the Q -matrix in Fig. 22.1 is an R -matrix. It is important to recall that if the similarities satisfy this inequality, then it does *not* necessarily follow that the order between the rows is consistent; the condition that they are consistent in some order is needed. Thus if an abundance or incidence matrix is a Q -matrix its similarity matrix is an R -matrix but not vice-versa. However, if the similarity matrix is an R -matrix *and* every set of three rows of the abundance matrix is consistent in some order, then the abundance matrix is a Q -matrix (Laxton 1987). *This shows that in general to deduce chronological order from an order of rows the method of ordering must involve testing for consistency in some way and not be by similarity/distance methods alone.*

However,

Results 3. If $S(a_i, a_j)$ is the strict maximum among all similarities $S(a_u, a_v)$, $u \neq v$, then a_i and a_j must be adjacent to each other in any consistent order of the rows of A .

Example $S(a_2, a_3) = 93$ is the maximum similarity among all similarities between the ten rows of A (see Fig. 22.2) and so these two rows must be next to each other in any consistent order of the rows of A (see Fig. 22.1).

22.3 Method of seriation

The method of rapid seriation proposed here is based on these three results. It is assumed that the rows a_1, \dots, a_n of A are consistent in some order, i.e., A is pre- Q

(Laxton 1976). Hence by Kendall's result (1971) we may use the similarity criteria alone to seriate.

Step 1. Choose a_i, a_j with $S(a_i, a_j)$ the (strict) maximum among all the S -values between the rows of A . By Result 3 these must be adjacent to each in any consistent order of the rows of A .

Replace a_i, a_j by their average and label all other rows of A so that at the end there are $(n - 1)$ rows

$$A_1 = \{b_1, \dots, b_{n-1}\}.$$

Step l . This step begins with the set of rows from Step $(l - 1)$

$$A_{n-l+1} = \{b'_1, \dots, b'_{n-l+1}\},$$

where each b'_i is an average of a consistent groups of rows, a_{i+1}, \dots, a_{i+u} , say, and it is assumed that these rows are adjacent to each other in *any* consistent order of the rows of A .

Choose b'_i, b'_j with $S(b'_i, b'_j)$ the (strict) maximum among all the S -values between the b'_1, \dots, b'_{n-l+1} . Then b'_i, b'_j must be adjacent to each other in any consistent order of A_{n-l+1} , by Result 3.

But any consistent order of the rows of A leads to a consistent order of A_{n-l+1} by Results 1 and 2 and the assumption that the rows a_{i+1}, \dots, a_i forming the average b'_i are next to each other in any consistent order of A .

Let b'_j be the average of the rows a_{j+1}, \dots, b_{j+v} , which are consistent in this order. Split each of these two sets of rows into two 'halves',

$$\{a_{i+1}, \dots, a_{i+\lfloor \frac{1}{2}u \rfloor}\}, \{a_{i+\lfloor \frac{1}{2}u \rfloor+1}, \dots, a_{i+u}\}$$

and

$$\{a_{j+1}, \dots, a_{j+\lfloor \frac{1}{2}v \rfloor}\}, \{a_{j+\lfloor \frac{1}{2}v \rfloor+1}, \dots, a_{j+v}\}, \tag{22.1}$$

where $\lfloor \frac{1}{2}u \rfloor$ is the greatest integer less than or equal to $\frac{1}{2}u$ and similarly for $\lfloor \frac{1}{2}v \rfloor$. Let the averages of these four 'halves' be b_{i1}, b_{i2}, b_{j1} and b_{j2} , respectively.

Assume without loss of generality that $S(b_{i2}, b_{j1})$ is the (strict) maximum among all the S -values $S(b_{iu}, b_{i+ju})$, $u, v = 1, 2$.

We claim that the set

$$a_{i+1}, \dots, a_{i+u}, a_{j+1}, \dots, a_{j+v}$$

is consistent in this order and that they must be next to each other in this order in any consistent order of the rows of A .

Proof We know by assumption at the beginning of Step l that the rows in each of the four consistent groups of (1) are next to each to other in any consistent order of A .

Also, since any consistent order of the rows of A gives rise to a consistent order for the rows b'_1, \dots, b'_{n-l+1} by Results 1 and 2 and, as $S(b'_i, b'_j)$ is maximum, then b'_i, b'_j are next to each other in any consistent order of b'_1, \dots, b'_{n-l+1} by Result 3. It follows that the rows

$$\{a_{i+1}, \dots, a_{i+u}\} \text{ and } \{a_{j+1}, \dots, a_{j+v}\}$$

must be next to each other in any consistent order of the rows of A .

There are, then, four possible arrangements in A :

$$\{a_{i+1}, \dots, a_{i+\lfloor \frac{1}{2}u \rfloor}\}, \{a_{i+\lfloor \frac{1}{2}u \rfloor+1}, \dots, a_{i+u}\}, \{a_{j+1}, \dots, a_{j+\lfloor \frac{1}{2}v \rfloor}\}, \{a_{j+\lfloor \frac{1}{2}v \rfloor+1}, \dots, a_{j+v}\} \tag{22.2}$$

or

$$\{a_{i+1}, \dots, a_{i+\lfloor \frac{1}{2}u \rfloor}\}, \{a_{i+\lfloor \frac{1}{2}u \rfloor+1}, \dots, a_{i+u}\}, \{a_{j+\lfloor \frac{1}{2}v \rfloor+1}, \dots, a_{j+v}\}, \{a_{j+1}, \dots, a_{j+\lfloor \frac{1}{2}v \rfloor}\} \tag{22.3}$$

or

$$\{a_{i+[\frac{1}{2}u]+1}, \dots, a_{i+u}\}, \{a_{i+1}, \dots, a_{i+[\frac{1}{2}u]}\}, \{a_{j+1}, \dots, a_{j+[\frac{1}{2}v]}\}, \{a_{j+[\frac{1}{2}v]+1}, \dots, a_{j+v}\} \quad (22.4)$$

or

$$\{a_{i+[\frac{1}{2}u]+1}, \dots, a_{i+u}\}, \{a_{i+1}, \dots, a_{i+[\frac{1}{2}u]}\}, \{a_{j+[\frac{1}{2}v]+1}, \dots, a_{j+v}\}, \{a_{j+1}, \dots, a_{j+[\frac{1}{2}v]}\}. \quad (22.5)$$

But the last three of these arrangements will give rise to the orders (by Results 1 and 2) $b_{i1}, b_{i2}, b_{j2}, b_{j1}$; $b_{i2}, b_{i1}, b_{j1}, b_{j2}$ and $b_{i2}, b_{i1}, b_{j2}, b_{j1}$ which are not consistent by Result 3 since by assumption $S(b_{i2}, b_{j1})$ is the (strict) maximum similarity. Hence only the order in (2) is possible, the order claimed. (That is, the order in (2) has been checked for consistency.)

At the end of Step l replace the sets in the consistent orders a_{i+1}, \dots, a_{i+u} and a_{j+1}, \dots, a_{j+v} by the set in consistent order $a_{i+1}, \dots, a_{i+u}, a_{j+1}, \dots, a_{j+v}$. These have to be adjacent to each other in any consistent ordering of the rows of A . Replace b'_i and b'_j by the single row which is the average of these $u + v$ rows of A to get a set

$$A_{n-l} = \{b''_1, \dots, b''_{n-l}\}$$

where the remaining b'_k have been relabelled only.

This proves by induction that the method leads to a consistent ordering of the rows of A .

Of course, this is based on the assumption that A is pre-Q, ie. has a consistent order.

22.4 Example

Step 1 The maximum similarity is 93 between rows 2 and 3. So these are averaged to give a row

$$(2, 3) = (7.00, 16.50, 73.00, 0.50, 3.00)$$

and the nine rows $A_1 = \{1, (2, 3), 4, 5, 6, 7, 8, 9, 10\}$. The new similarity matrix is compiled between these nine rows.

Step 2/3 The maximum similarity for the set of nine rows of A_1 is 89 (not strict, but this does not effect the result) between rows 4 and 5 and between 6 and 7. These are averaged to give the two new rows

$$(4, 5) = (20.50, 51.00, 25.00, 3.00, 0.00)$$

$$(6, 7) = (35.50, 45.50, 2.50, 16.00, 0.00)$$

and the seven rows $A_3 = \{1, (2, 3), (4, 5), (6, 7), 8, 9, 10\}$. The new similarity matrix is compiled between these seven rows.

Step 4 The maximum similarity for the set of seven rows of A_3 is 87 between rows 8 and 9. The average is formed to give the new row

$$(8, 9) = (55.50, 25.00, 0.00, 19.50, 0.00)$$

and the six rows $A_4 = \{1, (2, 3), (4, 5), (6, 7), (8, 9), 10\}$. The new similarity matrix is compiled between these six rows.

Step 5 The maximum similarity for the set of six rows of A_4 is 86 between the average of (2, 3) and 1. Since $S(1, 2) = 89 > S(1, 3) = 83$, they are consistent in the order 1, 2, 3. Form the average

$$(1, 2, 3) = (6.00, 13.00, 77.33, 0.33, 3.33)$$

and the five rows $A_5 = \{(1, 2, 3), (4, 5), (6, 7), (8, 9), 10\}$. The new similarity matrix is compiled between these five rows.

Step 6 The maximum similarity for the set of five rows of A_5 is 76.5 between the two averages of (6, 7) and (8, 9). Now since $S(7, 8) = 87$ is the maximum among all the similarities between 6 and 8, 6 and 9 and 7 and 9, i.e., between the 'halves', it follows that 6, 7, 8, 9 is a consistent order. Its average is

$$(6, 7, 8, 9) = (45.50, 35.25, 1.25, 17.75, 0.00)$$

and the four rows $A_6 = \{(1, 2, 3), (4, 5), (6, 7, 8, 9), 10\}$. The new similarity matrix is compiled between these four rows.

Step 7 The maximum similarity for the set of four rows of A_6 is 60.0 between the two averages (4, 5) and (6, 7, 8, 9). The averages of the 'halves' are

$$(6, 7) = (35.50, 45.50, 2.50, 16.00, 0.00)$$

$$(8, 9) = (55.50, 25.00, 0.00, 19.50, 0.00)$$

and it is clear that

4	18	48	30	3	0.0
5	23	54	20	3	0.0
(6, 7)	35.5	45.5	2.5	16.0	0.0
(8, 9)	55.5	25.0	0.0	19.5	0.0

is consistent for the 'halves' in this order since the similarity $SL(5, (6, 7)) = 74.0$ is a maximum among all S -values between these four rows. We take the consistent order (4, 5, 6, 7, 8, 9) in what follows and the three rows $A_7 = \{(1, 2, 3), (4, 5, 6, 7, 8, 9), 10\}$. The new similarity matrix is compiled between these three rows.

Step 8 The maximum similarity for this set of three rows of A_7 is 50.17 between the average of (4, 5, 6, 7, 8, 9) and 10. A check on the similarities between their 'halves' shows these are consistent when in order 4, 5, 6, 7, 8, 9, 10 with an average

$$(4, 5, 6, 7, 8, 9, 10) = (44.29, 35.14, 7.86, 12.43, 0.00).$$

Step 9 The similarity between the averages of (1, 2, 3) and (4, 5, 6, 7, 8, 9, 10) is 27.19. By comparing the similarities between the two 'halves' 1, and (2, 3) and the two 'halves' (4, 5, 6) and (7, 8, 9, 10), and determining the maximum among them, the final consistent order for A is seen to be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10—as required.

22.5 Seriating actual data

There are three parts to this method of seriation.

1. The maximum $S(b_i, b_j)$ of all similarities is determined at each stage.
2. Sets of rows are formed which are consistent.
3. The average of each consistent set of rows is formed (together with the averages of their 'halves').

It has been shown above that if the rows of A are consistent in some order (i.e., A is a pre- Q matrix), then the process described leads to a consistent order of the rows of A (i.e., permutes the rows so that A becomes a Q -matrix) and that (ii) is implied by (i) by Kendall's result (1969, 1971)). But the method has advantages in seriating a data set even if the set can only *approximate* a consistent set in some order. By

starting the process with the highest similarity values, those that are most similar in content are placed next or at least close to each in the ultimate order. Consistency is demanded at each stage since consistency is not now guaranteed by the similarity criteria alone (Laxton 1987). Thus condition (ii) is now independent of condition (i) and has to be checked in some form or other if chronological order is to be deduced from the seriation (Note, a pair of rows is always consistent). By averaging at each stage it is to be hoped, and expected when the 'noise' element is not too great, that the set of average rows gradually approximates more closely the seriation model and therefore that consistency is possible between them at the later stages as the similarity values drop.

In the method of seriation described above for ideal data consistency was demanded at all stages between the two 'halves' of any two averages being placed next to each other in the ordering. If 'noise' is present in the data, then this test will in general fail and the seriation come to a stop. For this reason we only demand consistency between the 'halves' if an average has at least a certain number, say c , of constituent rows. There are two extreme cases: if $c = 2$, then consistency is demanded at all stages and the method is the same as that described in the previous section for ideal data, whereas, if c is set to be greater than the total number of rows, then no consistency is demanded at all! The process to be described below starts with $c = 2$ when the similarity is high and then c increases gradually to avoid the process stopping. (If the rows can be permuted into a Q -matrix, then $c = 2$ would be retained automatically throughout.)

With this in mind the following method of seriation has been adopted.

Stage 1, $c = 2$. The similarity level is high, say at least 90 when percentages are used. Here the process is that described in the previous section and so *will lead to a consistent order for the rows of A , if such an order exists*. Recall that at each step an ordered group of $c \leq 2$ rows is checked for consistency. In practice, and depending on how far the data set departs from the seriation model, the stage is likely to end with several consistent groups of two or more elements each and with high similarity between the rows in each consistent group; similarity between the groups and the single rows left at the end of the stage is usually lower than those employed to form the consistent groups.

Stage 2, $c = 3$. The similarity level is lowered, say to 80, or lower if there are no similarities this high. Averages of all the consistent groups formed in Stage 1 are taken together with the averages of the two 'halves' of each group. Again groups are formed among these *average* rows using the highest similarities first, but for any group with at least three rows in it, consistency is demanded between the two sets of 'halves'.

The process is repeated until it is no longer possible to form consistent groups at this level of similarity.

Stage 3, $c = 4$. The similarity is lowered, say to 70. The averages of all consistent groups formed in Stage 2 are taken together with the averages of the two 'halves' of each group. Again groups are formed among these average rows using the highest similarities first, but for any group with at least four rows in it, consistency is demanded between the two sets of 'halves'. The process is repeated until it is no longer possible to form consistent groups at this level of similarity.

Later stages The value of c is gradually increased and the value of similarity is gradually decreased in stages largely depending on the data size and circumstances. It is good practice to increase and decrease as slowly as possible. The process must end in one group at or before $c =$ number of rows in A , since at this latter value consistency is not asked for. At this stage also, the similarity values dealt with are likely to be low.

In practice, if no new groups are formed at a stage with $c = x$, then c is raised in value to one more than the size of the smallest group larger than x . If all the rows are ordered into one group at a relatively low value of c , then the data can be considered to be well seriated. On the other hand, if groups of rows are formed at a value of c and stay so for several more higher values, then these groups can be considered separately for their chronological order and a chronological order for the whole set might be rejected.

22.6 Example (One room in the mound at Awatovi)

Level 1 (top)	2	0	88	0	11
2	2	4	82	0	12
3	17	60	19	4	0
4	22	60	14	3	1
6	9	59	27	5	0
7	35	41	0	23	0
8	29	44	0	27	0
9	35	48	0	18	0
10	44	31	2	24	0

(For illustrative purposes an anomalous level 5 is omitted.)

Stage 1, $c = 2$

At similarity 90 and above the consistent groups (1, 2), (6, 3, 4) and (9, 7) are formed. The following averages are now formed

(1, 2)	2.00	2.00	85.00	0.00	11.50
(6, 3, 4)	16.00	59.67	20.00	4.00	0.33
8	29.00	44.00	0.00	27.00	0.00
(9, 7)	35.00	44.50	0.00	20.50	0.00
10	44.00	31.00	2.00	24.00	0.00

with similarity matrix

	(1, 2)	(6, 3, 4)	8	(9, 7)	10
(1, 2)					
(6, 3, 4)	24.33				
8	4.00	64.00			
(9, 7)	4.00	64.00	93.50		
10	4.00	53.00	84.00	86.5	

No more consistent groups can be formed at this stage with similarity at least 90. (Note. The similarities between 8 and 7 and between 8 and 9 are greater than 90 but rows 7, 8 and 9 do not form a consistent group in any order. Hence they are not grouped at this stage with $c = 2$.)

Stage 2, $c = 3$. The similarity level is lowered to 80. Now the similarity between the average (9, 7) and 8 is the maximum (93.5) and, as a pair, they form a consistent group (recall that any pair form a consistent group). The similarities between 8 and 9 and 7 are, respectively, 91 and 93. Hence the order given to the group of rows is (9,7,8) since 8 is closer to 7 than to 9. The average

(9, 7, 8)	33.00	44.33	0.00	22.67	0.00
-----------	-------	-------	------	-------	------

is formed. The similarity between this and 10 is 86.67, the maximum similarity. Since the former group now involves three rows, consistency between the 'halves' 9, (7,8) and 10 must be checked before a larger group can be formed. They are consistent in the order 9, (7,8), 10. Hence the average (9,7,8,10) is formed and the group ordered in this way.

The final result is

(1, 2)	2.00	2.00	85.00	0.00	11.50
(6, 3, 4)	16.00	59.87	20.00	4.00	0.33
(9, 7, 8, 10)	38.80	38.80	4.20	18.60	0.00

with similarity matrix

	(1, 2)	(6, 3, 4)	(9, 7, 8, 10)
(1, 2)			
(6, 3, 4)	24.33		
(9, 7, 8, 10)	8.20	63.0	

The similarity table between the 'halves' of the different groups formed so far is

	1	2	6	(3, 4)	(9, 7)	(8, 10)
1						
2						
6	29	33				
(3, 4)	19	23				
(9, 7)	2	6	58.5	67.5		
(8, 10)	3	7	52.5	61.5		

There are no consistent groups among the 'halves' of groups with three or more rows, and so the process will stop even if the level of similarity had been as low as 60.

Stage 3, $c = 4$. The similarity level is now set at 60.

The maximum similarity is 63 between (6,3,4) and (9,7,8,10). The 'halves' to check for consistency are (6,3,4), (9,7) and (8,10), since the first has less than four rows in it. These are not consistent in any order and so amalgamation does not take place at this stage with $c = 4$.

Stage 4, $c = 5$. The similarity level remains at 60. The two averages just mentioned do not now have to be checked for consistency since they have less than 5 rows each. From the similarity table for the 'halves' of these groups it is evident that they are grouped in the order 6,3,4,9,7,8,10 and the average is

$$(6, 3, 4, 9, 7, 8, 10) = 27.3 \quad 49.0 \quad 8.6 \quad 14.9 \quad 0.1$$

This ordered group has seven rows in it. Hence for further ordering to take place at this stage, consistency between the 'halves' is required. But the 'halves' (1,2), (6,3,4) and (9,7,8,10) are consistent in this order, as the above table shows. Finally, since the similarities between the 'halves' 1 and (6,3,4), 1 and (9,7,8,10), 2 and (6,3,4) and between 2 and (9,7,8,10) are 22.33, 6.2, 26.33 and 10.2, respectively, it is clear that the final order of the rows is

$$1, 2, 6, 3, 4, 9, 7, 8, 10$$

This is a slightly different order from the stratigraphic one.

The data matrix with rows in this order becomes

1 (top)	2	0	88	0	11
2	2	4	82	0	12
6	9	59	27	5	0
3	17	60	19	4	0
4	22	60	14	3	1
9	35	48	0	18	0
7	35	41	0	23	0
8	29	44	0	27	0
10	44	31	2	24	0

This, of course, is not a Q-matrix, though it clearly approximates one.

22.7 Conclusions

The method, as far as it goes, has been described in some detail and applied to some simple examples to illustrate its workings. It has been used to seriate a number of well-known data sets and gives similar, though not identical ones to those obtained by multivariate methods. One feature of the method is that it does produce ordered groups in general, and not necessarily a complete ordering of the data, in cases where the data as a whole does not approximate well the seriational model for artifact change with time.

It is clear from an analysis of the results on these well-known data sets that much of the seriation is carried out by similarity alone, or almost alone, at the lower values of similarity. But in theory the consistency check should be especially important at these low values of similarity and this may point to a weakness in current practice. This may not be important if it is known *a priori* that time is the only factor effecting the data, but if this is not known or other factors are involved, then some sort of check on consistency is called for at the lower levels of similarity if chronological is to be inferred. What this seems to imply for the present method is that some other, possibly slightly weaker method of checking consistency to the seriational model is needed, eg. some form of fitting.

Notwithstanding this criticism, a jack-knifing procedure is being programmed. This will remove in a systematic way or to place elsewhere, if possible, individual rows which have not been fitted into the rest of the ordered group in the best possible way during the above process. This, together with other modifications will be presented later.

Bibliography

BURGH, R. F. 1959. "Ceramic profiles in the western mound at Awatovi, North-Eastern Arizona", *American Antiquity*, 25: 184-202.

DJINDJIAN, F. 1988. "Improvements in intra-site spacial analysis techniques". in Rahtz, S. P. Q., (ed.), *Computer and Quantitative Methods in Archaeology 1988*, International Series 446, pp. 95-105. British Archaeological Reports, Oxford.

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

GELFAND, A. E. 1971. "Rapid seriation methods with archaeological applications". In Hodson *et al.* 1971, pp. 186-201. (Proc. Anglo-Romanian Conference, Mamia 1970).

GREENACRE, M. J. 1984. *Theory and applications of correspondence analysis*. Academic Press, London.

HODSON, F. R., D. G. KENDALL, & P. TAUTU 1971. *Mathematics in the archaeological and historical sciences*. Edinburgh University Press, Edinburgh. (Proc. Anglo-Romanian Conference, Mamaia 1970).

KENDALL, D. G. 1969. "Incidence matrices, interval graphs and seriation in Archaeology", *Pacific J. Math.*, 28: 565-570.

KENDALL, D. G. 1971. "Seriation from abundance matrices". In Hodson *et al.* 1971, pp. 215-222. (Proc. Anglo-Romanian Conference, Mamaia 1970).

LAXTON, R. R. 1976. "A measure of pre-Q-ness with applications to archaeology", *J. Arch. Sci.*, 3: 43-54.

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

MADSEN, T., (ed.) 1988. *Multivariate Archaeology, Numerical Approaches to Scandinavian Archaeology*. Jutland Archaeological Society Publications XXI. Aarhus University Press, Aarhus.

ORTON, C. 1980. *Mathematics in Archaeology*. Collins, London.