9
Models, methods and techniques for seriation

G. L. COWGILL
Brandeis University, U.S.A.

The object of this paper is partly to present as clearly as possible my understandings about the assumptions and models, implicit and explicit, which underly seriation as a method for chronological ordering in archaeology. Also important is a new and highly economical computer technique for seriation.

On the level of method and general assumptions, recent publications of importance include Rowe (1961), Kuzara, Mead and Dixon (1966), Dethlefsen and Deetz (1966), Hole and Shaw (1967), Rouse (1967), Cowgill (1968) and Dunnell (1970). For techniques, important recent work includes Ascher and Ascher (1963), Dempsey and Baughman (1963), Johnson (1968), Bordaz and Bordax (1970) and Renfrew and Sterud (1969). Especially important is recent work by Kendall (1969) and Gelfand (1971).

By seriation as a formal technique, I refer to any technique for arranging a set of entities into a sequence such that, starting from any specific entity, the other entities most similar to it are closest to it in the sequence, and similarity decreases monotonically as one compares entities progressively more distant in the sequence. Monotonic decrease means that as one moves away from the starting point, similarities may stay the same, decrease a little, or decrease sharply, but will in no case increase. A monotonically decreasing function can be graphed by a line which may remain horizontal, drop gradually or steeply, drop a little or a lot, but in no case rises. Sometimes this ideal arrangement of entities can be fully achieved, but for many sets of entities the mutual similarities are such that no completely ideal sequence is possible. That is, it may be that whenever the entities are arranged so that the criterion of monotonic decrease in similarity is met with reference to one entity, there is always at least one other entity in the set for which the criterion is not met. For this reason, seriation also includes techniques for obtaining and evaluating the best possible approaches to the ideal sequence.
Models in archaeology

It is important to notice that one thing not included in this definition of seriation as a formal technique is any notion of arranging entities with regard to any assumptions about development or derivation; that is, there is no attempt to consider whether the nature of the similarities are such that one entity is more likely to have developed from or been derived from this or that other entity. This exclusion is not because such approaches are not important or valuable. On the contrary, I think there is great value in further development and testing of such methods both for chronology and for processual understandings of the past (Clarke, 1968, makes some important observations on these matters). I exclude them here simply because they would introduce too much to deal with at one time. Seriation as I have defined it is, as a formal technique, a relatively well-defined and self-contained set of problems, and a considerable amount of work has been done within this general framework. One important and often-mentioned consequence of this limitation to some general concept of overall similarity, without regard for formal features that might suggest that one entity is derived from another, is that all solutions to the seriation problem come in pairs; for any specific sequence of entities, another sequence that is its exact reverse is an equally good seriation.

Considering seriation as a formal technique, there are three broad kinds of problems. Central, of course, is the task of finding, as expeditiously as possible, a sequence which is either ideal or else makes a good case for being as close to ideal as is possible. Second, there is the general topic of deciding between variation 'badness of fit' criteria when no ideal sequence can be found; that is, deciding how to rate various alternative non-ideal sequences. Finally, there is the matter of precisely how similarity is to be measured or evaluated. This itself divides into two topics: decisions about the features of the entities that are to be observed and recorded, and decisions about what mathematical functions of these observations should be used for computing coefficients of similarity.

In the 'graphical' seriation technique of Ford (1962) the solutions adopted for all these problems can remain somewhat vague, ambiguous, intuitive, and possibly inconsistent. Whenever techniques are used that require specific numerical values for similarity coefficients, the answers to all the questions posed above must be made explicit. This is not, perhaps, as inherently damaging a criticism of the graphical technique as one might think; I agree with Dunnell (1970) that there has been a little too much tendency to downgrade it simply because it is more simple-minded than more mathematical techniques. As a technique for seriation, the graphical way is all right if it is not too hard to do, and if different workers can agree (with unimportant
Models, methods and techniques for seriation

differences) on the same 'best' sequence. The troubles with the technique
seem to be that too often it does involve a great deal of work and that there can
be serious differences of opinion about the best sequence. And where these
troubles arise there is surely a need to look for easier techniques and more
explicitly defined concepts.

None of the formal problems can be usefully discussed if we continue to
think of seriation only as a formal technique, without relating it to the prob­
lems and objectives of archaeology. Its relationship to models of human be­

For practical purposes, the only archaeological use for seriation is for
chronological ordering of units in cases where other techniques (such as
stratigraphy, dated inscriptions, firm cross-ties with established sequences, or
radiocarbon or other physical or chemical techniques) are inapplicable or
inadequate. The units to be seriated are typically either grave lots, caches or
hoards; or else archaeological assemblages. In both cases, as Rouse (1967, p.
158) has emphasized, the units of real interest are events. In the case of a grave
lot, the event would be the bringing together and deposition of grave offerings.
For a unit that is an assemblage, the 'event' is the occupation of a site or site
segment during the period that saw the creation of the depositional units
assigned to one assemblage by the archaeologist. In any case, when I refer to
the duration of a unit, what I mean is the time span of the event that the unit
reflects. The assumption is that differences between units mainly reflect
differences in time, and that the seriation sequence is a good approximation
to the time sequence of the units. At least two units need to be datable by
some other technique, so that the proper time direction of the seriation se­
quence can be established; additional chronological information of course
serves as a test or corroboration of the seriation sequence.

In principle, as Kuzara, Mead and Dixon (1966) especially have pointed
out, seriations of archaeological units could arrange them according to many
different possible kinds of differences: relative to status, activities, cultural
tradition, and so on. In practice, there is little or nothing to recommend
seriation as a technique for investigating such differences, because the seria­
tion model only allows for arrangement of units along some single axis or
dimension, and this makes sense only if the whole pattern of relationships
between all the entities can be tolerably well related to some single 'under­
lying' variable. Whenever one is looking for contemporaneous variations
among units, it generally seems likely from the beginning that several factors
may be operating at once, and one should reject a seriation model in favour of
explicitly multivariate techniques, such as factor analysis, multidimensional
Models in archaeology

scaling, discriminant functions, or the like. In fact, there is not much to recommend any inherently ‘single axis’ technique even for chronological ordering, but at least it appears that fairly often one really finds important sets of archaeological units where change over time is far greater than are differences due to other factors. Of course, this is also frequently not true, and, as Dunnell (1970) emphasizes, it always needs to be tested.

A very important feature of the technique described in the latter part of this paper is that it uses a multidimensional scaling in two or more dimensions. This amounts to a test of the ‘one axis’ hypothesis. The data may or may not fit well into an essentially one-dimensional pattern. If they do, and if chronological data available from other sources are consistent with this one axis being a time axis, then it makes sense to go ahead with a pure seriation technique. Otherwise, one will require some sort of multivariate method which allows for time as one among several axes of change. It should be added that it is hardly necessary to demonstrate that there is no source of consistent variation other than time-related change; it is only necessary that such other sources of variation, within the particular set of units being studied, make for differences that are small relative to differences reflecting the smallest time intervals one hopes to reliably distinguish.

Besides the assumption that changes over time are the only important source of differences between units, three other assumptions relating the behaviour of ancient peoples to formal models need discussion. First, there is the view, which I think is a misconception, that seriation models imply relatively gradual and continuous cultural change, a view particularly advocated by James Ford (1962). On the contrary, for seriation to be useful as a basis for chronology, it seems to me that all that is required is that there never, among the set of units being seriated, be a break in the sequence so abrupt and catastrophic that units immediately following the break bear no (or only accidental) resemblance to units before the break. If this were to happen we could still divide all the units into two subsets and do a formally adequate seriation within each subset, but we would not know which end of one sequence to join to which end of the other sequence. If, instead, some units at one end of one sequence show even moderately more resemblance to units at one end of the other sequence than to those at the other end, then it is possible to seriate the whole set of units. In other words, a relatively abrupt and drastic change in material culture is no bar to good seriation, provided that it does not result in total obliteration of earlier forms, and provided that all other conditions required for valid seriation are met.

A second requirement, mentioned by Rouse (1967, p. 162) is that the dura-
Models, methods and techniques for seriation

tions of all units being seriated be roughly comparable. I suggest that this one aspect of a somewhat more general requirement – namely, that for any two units whose chronological positions are claimed to be reliably or usefully distinguished, both have durations that are at least not much greater than their time difference, and preferably durations that are considerably smaller than their time difference. To illustrate, consider a situation where this is not so. This would mean that there is considerable overlap in the durations of the two units. What we mean by saying that one such unit seriates ‘earlier’ than the other must be that some middle time or time of maximal activity in one unit is earlier than a comparable middle or maximal time in the other unit, but not by very much, relative to the total duration of at least one of the two units. This would mean that much, if not all, of the activity responsible for one unit was actually contemporary with activity responsible for the other unit. Being able to say that ‘on the average’ the activity in one unit was a little later than activity in the other unit would be of dubious value. Furthermore, in such a situation even the judgement that ‘on the average’ one unit is a little earlier than the other may depend a good deal on sampling accidents, and may not be very reliable.

In a seriation there is nothing wrong with some units having considerable overlap in their duration; to admit such subsets of units is just to admit that there can be ties or near ties in the overall sequence. That is, there can be units whose chronological separation is relatively slight, and little stretches of the overall sequence within which the exact subsequence is not very reliably determined. This is all right as long as it is recognized that what one claims is that units are in the right general position in the total sequence, not that the sequence is exactly right in detail. What is important, if the seriation is to be of any practical value, is that there should be a fair proportion of pairs of units for which their time separations are large relative to their durations. One consequence is that, even if there may be a number of near ties in the sequence, no unit should have a duration that is any longer than the smallest distinguishable interval to which it is assigned.

The third and most basic requirement is that, in terms of the criteria of similarity used, trends of increasing dissimilarity over time indeed never reverse themselves. In fact it is very easy to develop criteria of similarity that do show reverse trends, so that units toward the beginning and end of some chronologically known sequence rate higher in similarity to one another than to units in the middle of the sequence. Particularly, this can be due to picking too few features to observe, e.g. by using the relative abundance of painted versus plain pottery, where a plain-painted–plain sequence of styles will make
earlier units seem like late units, even though they may have little in common except shared absence of painted pottery. But these kinds of reversals are not likely to happen over any substantial time spans if one uses much judgement in selecting the features on which to base indices of similarity. Over short time spans, small-scale reversals may well occur, and this is another reason why one should not put too much faith in the fine details of a seriation chronology.

Commonly the technique for evaluating similarities is begun by sorting pottery or other artefacts into categories (‘types’), whose occurrence or relative abundance in each unit is noted; or by noting occurrences or relative abundances of distinctive features (‘modes’) on objects pertaining to each unit. Typological operations of this general kind are used for many purposes besides seriation and involve a number of methodological issues discussed elsewhere in this volume. For seriation per se, I suspect that we have worried too much about these problems. If we really have no other objective than the approximately correct time ordering of a number of units, then we should be able to do this with no more than a tiny fraction of the information potentially available. At least for the sake of argument, I suggest that one needs to use very little of the total richness of the data (much of it of prime importance for other questions) in order to do a good seriation. A principal lesson from the good results often obtained from Meighan’s technique of computing percentages of only three well-chosen types (Meighan, 1959; Ascher, 1959; Hole and Shaw, 1967; Cowgill, 1968; Rattray, 1971) is this: often one can throw away most of one’s information and still get a good chronological ordering. It is true that much depends on which small fraction of the data one retains for seriation purposes, but I suspect that rather than it being a very tricky business to get good types or modes for chronological ordering, it may be difficult to go very far wrong, provided a few general rules are followed.

The general rules for chronological types that I have in mind are: first, that one will look for types that occur in more than a very few but less than half of the total set of units to be seriated; second, one will try to avoid redundant types that nearly always are present or absent in the same units (if presence or absence of one type can be predicted with high success from the presence or absence of some other type, then one adds little information not already provided by the other); and, finally, one will look for types which, on the basis of whatever other chronological information one has, and on the basis of preliminary seriation trials, seem to occur rarely or never before some point in time, then rise fairly monotonically to a single peak frequency of occurrence or peak relative abundance, and then decrease fairly monotonically to the
Models, methods and techniques for seriation

level of few or no occurrences. Of course, it is not necessary that every type exhibit the whole cycle within the set of units being seriated; many types may be ‘caught’ only on the descending or the ascending segments of their cycles, and the essential thing is that trends up to, or down from, a single peak be fairly monotonic.

Another important point which should be made here is that if an ideal seriation of a data set exists, it is probably easy to find. At any rate, Gelfand (1971) has shown that, for seriations based on relative abundances, if there exists an ideal sequence one can find it simply by beginning with the highest observed similarities, and selecting as nearest neighbours for each unit the other two most similar units, unless one of these is unavailable because it already has two nearest neighbours or unless this would mean linking the head of a chain of units to the tail (thus forming a closed loop). In either of these cases, the most similar units that are not ineligible are picked as nearest neighbours to a given unit. With ideal data, this seems to be a very easy and effective technique. For instance, Kendall (1969) gives a sample set of six units for which he obtained an ideal sequence through multidimensional scaling. He suggests that the reader will find it instructive to try seriating his data set by hand. Indeed it is. Using Gelfand’s technique it took me less than two minutes with pencil and paper to obtain the same sequence that Kendall got through multidimensional scaling.

If a set of units cannot be made to fit the seriation ideal perfectly, but something very close to an ideal sequence is possible, Gelfand’s technique should enable one to obtain easily a sequence that is at least close to the best possible sequence. On the other hand, if nothing at all close to an ideal sequence is possible, one should not, as I have said before, be trying to do a seriation anyhow. The main justification for techniques more elaborate than Gelfand’s seems to be that situations arise, probably rather frequently, where the best possible sequence is far enough from the ideal to require something more than his technique alone, yet not so far from ideal that seriation is inappropriate. In such cases it is still important to look for the best and easiest techniques.

It is to some of these considerations of technique that the rest of this paper is devoted. First, some definitions are in order. Many of these come from Kendall (1969), whose terms I have adopted unless there seemed to be a good reason for change.

It is very useful to think of the data for a seriation as arranged in an array of rows and columns, a matrix, in which each row represents a unit and each column pertains to a type (what Kendall calls a variety; I prefer ‘type’ because it is shorter and perhaps carries fewer confusing connotations of the
Models in archaeology

'type variety' terminology of some American archaeologists). Such a matrix can be read as a data table, in which units are listed from top to bottom along the left margin, as row headings, and types are listed from left to right across the top, as column headings. It is useful to distinguish two kinds of data matrices. An occurrence matrix is one that merely records presence or absence of types in each unit. A '1' in column 5 of row 8, then, indicates that Type 5 is present in unit 8; while a zero in column 3 of row 12 indicates that Type 3 is absent in unit 12. In technical terms, component $a_{ij}$ of an occurrence matrix $A$ is unity if type $j$ occurs in unit $i$; otherwise it is zero. Components of matrices are conventionally represented by two subscripts, the first of which designates the row and the second the column, so that $a_{24}$ stands for the number in row 2 and column 4 of the array.

Kendall uses the term 'incidence matrix' for what I call an occurrence matrix, a term I find more congenial, which is used by Dunnell (1970).

The other important kind of data matrix also represents types as columns and units as rows; the only difference from an occurrence matrix is that the values shown as per cents or relative proportions of the count for a given type, relative to the total count of all specimens of all types exhibited by a given unit. In such a matrix, $B$, element $b_{ij}$ is the percentage of type $j$ in unit $i$, relative to all the relevant material in unit $i$. For any one row, the percentages must add to 100%. This is the kind of data matrix used as a starting point both for Ford's graphical technique and for Brainerd's (1951) and Robinson's (1951) matrix technique. It is often preferable to an occurrence matrix, especially when the units are sites or site components, where (1) types are commonly represented by numerous examples, (2) many types may occur in at least small amounts in all units, so that an occurrence matrix would be nearly all 1's and very few 0's, and (3) there may be considerable danger of at least slight admixture of earlier material or later material than that pertaining to the unit one hopes to date (a danger discussed by Lipe, 1964). A percentage matrix presents more information than an occurrence matrix, but occurrence matrices are simpler and may be preferable wherever the additional information that a percentage matrix provides does not do much to improve the chronological ordering. Occurrence matrices make most sense when the units are grave lots or other closed finds, where types that occur are rarely represented by more than one example in a given unit, mixture of material that pertains to some other unit is not likely to be unrecognized or unavoidable, and any specific type is absent in a good proportion of all units. Table 9.1 and Figs. 9.4, 9.8, 9.9, 9.11, 9.14 and 9.15 are examples of occurrence matrices — although all the figures are unconventional in replacing zeros by blanks.

For an occurrence matrix, the task of seeking a rear column are bunched in such a sequence is a change occurred in every unit reappeared. If we define occur in both units, so exhibits the largest number of seriation is equivalent. For percentage matrices, percentage for any one type value) to a maximum, the value). Although they an ocurrence to percentage matrices.

Kendall (1969) introduced for an occurrence matrix column consecutive. He all matrix multiplication of the technical terms, $V = A^T A$.

where $g$ is the total number of any pair of types $i$ and $j$, the occurrence matrix, multiply rows. If either or both the $i$ only for those rows where $b_{ij}$ (this corresponds to the unit represented by this row). To obtained, summing over all $j$, types $i$ and $j$ both occur. The $V$ of fairly simple functions of similarities in the occurrence matrix.

As a simple example, Table 9.2 shows the $V$ matrix.
Models, methods and techniques for seriation

For an occurrence matrix the formal job of seriation can be phrased as the task of seeking a rearrangement of the rows such that all the 1's in each column are bunched into a consecutive run, uninterrupted by any zeros. If such a sequence is a chronology, it indicates that each type, once it appeared, occurred in every unit until some time of obsolescence, after which it never reappeared. If we define similarity between units as the number of types that occur in both units, so that a unit is most similar to whatever other unit exhibits the largest number of the types that occur in unit, then this definition of seriation is equivalent to the definition given at the beginning of this paper. For percentage matrices, seriation amounts to seeking a sequence such that the percentage for any one type rises monotonically from zero (or some minimum value) to a maximum, then decreases monotonically to zero (or some minimum value). Although they are important, I will make only passing further reference to percentage matrices, and will devote the rest of this paper to occurrence matrices.

Kendall (1969) introduces the term Petrie Matrix (after Flinders Petrie) for an occurrence matrix that is in the ideal pattern, with all the 1's in each column consecutive. He also defines two other matrices that can be formed by matrix multiplication of the occurrence matrix: A and its transpose, A^T. In technical terms, V = A^T A is a matrix whose (i, j)th component is

\[ v_{ij} = \sum_{k=1}^{g} a_{ik} a_{kj} \]

where g is the total number of units being studied. What this says is that, for any pair of types i and j, one moves down column i and column j of the occurrence matrix, multiplying the numbers that occur together in the same rows. If either or both the numbers is a zero, the product will be zero, and only for those rows where both numbers are 1's will the product also be unity (this corresponds to the situation where types i and j both occur in the unit represented by this row). The component v_{ij} is the sum of all the 1's thus obtained, summing over all units, and is thus the number of units in which types i and j both occur. The components of the V matrix, as well as a number of fairly simple functions of its components, are coefficients reflecting similarities in the occurrence patterns of pairs of types.

As a simple example, Table 9.1 shows an occurrence matrix, A, with four columns (types) and six rows (units). Table 9.2 shows the V matrix, with four rows and columns, derived from A.
Models in archaeology

If, instead, $A$ and $A^T$ are multiplied in the opposite sequence, we get $G = AA^T$, whose $(i, j)$th component is

$$g_{ij} = \sum_{k=1}^{v} a_{ik}a_{jk}$$

where $v$ is the total number of types. Here what is done is, for any pair of units $i$ and $j$, to move across row $i$ and row $j$, multiplying the numbers that occur together in the same columns. Again, the products will be zero except for those columns where the numbers in both rows are 1's, corresponding to the situation where units $i$ and $j$ both exhibit the type corresponding to that column. The component $g_{ij}$ is the sum of all the 1's thus obtained, summing over all

**TABLE 9.1. Sample occurrence matrix**

\[
\begin{pmatrix}
1 & 0 & 1 & 1 \\
1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 \\
0 & 1 & 0 & 0 \\
\end{pmatrix}
\]

never clear from a seriation of a sequence. If one were to think of $G$ as analogous to $G$ would not be analogous to $g_{ij}$.

For percentage matrices and is another example of the Brainard–Robinson

$$g_{ij} = \sum_{k=1}^{\nu} a_{ik}a_{jk}$$

that is, for units $i$ and $j$, the $i$th occurrence matrix, finding for in percentages of the type $f$ positive values, adding the the sum from 200 (which difference for two units in $w$ in the other; that is, for un

I mention the Brainard– what $V$ and $G$ are, and how 9.2) the number of rows and these types are represented $i$ and from left to right for $c$ with column $1$, of row $2$ to

**TABLE 9.2. Sample V matrix, derived from the A matrix of Table 9.1**

\[
\begin{pmatrix}
3 & 0 & 2 & 2 \\
0 & 3 & 1 & 0 \\
2 & 1 & 3 & 1 \\
2 & 0 & 1 & 2 \\
\end{pmatrix}
\]
Models, methods and techniques for seriation

get \( G = AA^T \), the intersection of the \( j \)th row and \( i \)th column. This is, in effect, another way of saying that the similarity of unit (or type) \( i \) to unit (or type) \( j \) is the same as the similarity of unit (or type) \( j \) to unit (or type) \( i \), and this is the reason why it is

<table>
<thead>
<tr>
<th>TABLE 9.3. Sample ( G ) matrix, derived from the ( A ) matrix of Table 9.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(</td>
</tr>
<tr>
<td>( \quad 2 2 1 0 0 0 )</td>
</tr>
<tr>
<td>( \quad 2 1 2 0 1 0 )</td>
</tr>
<tr>
<td>( \quad 0 0 0 1 1 1 )</td>
</tr>
<tr>
<td>( \quad 1 0 1 1 2 1 )</td>
</tr>
<tr>
<td>( \quad 0 0 0 1 1 1 )</td>
</tr>
</tbody>
</table>

never clear from a seriation alone which is the early and which is the late end of a sequence. If one were to develop a formal method that made use of notions about what was likely to be derived from what, then the matrix analogous to \( G \) would not be symmetrical; in general \( g_{ij} \) would not be equal to \( g_{ji} \).

For percentage matrices, the Brainerd–Robinson matrix is analogous to \( G \), and is another example of a square symmetric matrix. In this case, if the components of the initial percentage matrix \( B \) are \( b_{ij} \), the \((i,j)\)th component of the Brainerd–Robinson matrix \( R \) is

\[
    r_{ij} = 200 - \sum_{k=1}^{v} |b_{ik} - b_{jk}|
\]

That is, for units \( i \) and \( j \), \( r_{ij} \) is found by moving across rows \( i \) and \( j \) of the percentage matrix, finding for each column (that is, for each type) the difference in percentages of the type for the two units, converting all these differences to positive values, adding the values obtained for all columns, and subtracting the sum from 200 (which is the maximum possible cumulative percentage difference for two units in which none of the types that occur in one unit occur in the other; that is, for units that show the greatest possible difference).

I mention the Brainerd–Robinson matrix partly to give readers a feel for what \( V \) and \( G \) are, and how they compare with \( R \). In the case of \( V \) (see Table 9.2) the number of rows and columns is equal to the number of types, and these types are represented in the same sequence from top to bottom for rows, and from left to right for columns. It follows that the intersections of row 1 with column 1, of row 2 with column 2, and so on, along the 'principal
diagonal' from the upper left to the lower right corner of the matrix, contain coefficients indicating similarity in occurrence of each type with itself, and these are necessarily maximal values for each row and column. For $G$ or $R$, the number of rows and columns is equal to the number of units, and the principal diagonal contains values indicating the similarity of each unit to itself, which again are maximal values.

Two important mathematical results have recently been obtained for occurrence matrices. Fulkerson and Gross (1965) have shown that all the information needed to decide whether an occurrence matrix $A$ can be petrified (that is, whether there exists a sequence of the rows such that all the 1's in each column are consecutive; the term was introduced by Kendall) is contained in $V$, and they give a graph-theoretical algorithm for answering this question. Kendall (1969) has complemented this by showing that if $A$ is indeed petrifiable, then all the information needed to rearrange $A$ into the Petrie form is contained in $G$. He does this by proving that $A$ will be petrified when the rows are rearranged so that $G$ is in what he calls the Robinson form, meaning by this a condition in which all components monotonically decrease as one moves away from the principal diagonal along any row or column.

This point, that a data set is ideally seriated if and only if a certain matrix derived from it is in the Robinson form, is of course just what Robinson (1951) had in mind for his matrix $R$, derived from a percentage matrix. What is new about Kendall's work is that he proves it also for the different matrix $G$, derived in a different way from an occurrence matrix.

These results are nice, but, as Kendall recognizes, of somewhat limited use, since (1) for nearly all real archaeological data sets perfect petrification is demonstrably impossible, and (2) even where it is possible, these results do not tell one how to obtain the Petrie sequence. It is here that Gelfand's (1971) work is so valuable, for he shows that if it is possible to put a matrix into the Robinson form at all, it can be done very easily by rearranging rows and columns so that each entity has as its nearest neighbours the other two entities (not already assigned to two nearest neighbours and not already linked through an intermediate chain) with which it shares the highest coefficients. Operating on $R$, this technique will seriate an ideal percentage matrix and, operating on $G$, it will seriate an ideal occurrence matrix. For matrices that cannot quite be put into the Robinson form, this technique should provide a very useful preliminary ordering, which one might then try to improve by other means.

Before moving on to the question of what to do when the data are not close enough to the ideal for Gelfand's technique to be sufficient, I would like to add a few comments of my own about some logical properties of petrifiable occurrence matrices. It appears that the largest number of distinguishable types recognized is $2v$. It seems that no $v$ types could occur, no matter how many there are that the occurrence matrix is in the negative bunches, and suppose that the sequence is different from the beginning and can be, then, at most $v$ different beginning points, or $2v$ altogether. But new combinations can occur in a Petrie matrix.

Since the combinations of types in a sequence in which the rows of the occurrence matrix is easy partial test of whether or not consists of counting the number actually occur. If it is greater than possibly be petrified. Presumably the combinations to $2v$ will give some data could be made to approximate if the number of observed type is certain that the matrix can be petrified, happen to have identical end points, fewer than $2v$ distinct type combination or fewer in a non-petrifiable matrix type combinations would strongly suggest nearly petrifiable.

A different practical implication of occurrence matrix is distinctly informative it can provide. If there number of units is more than twice the $2v$ must exhibit a combination of types in some other unit, and hence any other or more other units, and chronology information additional to that in the number $2v$ assumes that we include. Realistically, we would hardly try to at least one of the set of types under means that at most $2v - 1$ periods occurrence matrix is petrifiable.
If the matrix, containing a type with itself, and columns for $G$ or $R$, the number of units, and the direction of each unit to
be petrifiable (that is, if all the 1's in each column obtained for occurrence matrices. It appears that, if there are $v$ types in such a matrix, then the largest number of distinguishable intervals that could conceivably be recognized is $2^v$. It seems that not more than $2^v$ different combinations of types could occur, no matter how many units there are. To see this, suppose that the occurrence matrix is in the Petrie form, so that all 1's are in consecutive bunches, and suppose that the beginning and ending point for every type is different from the beginning and ending point of every other type. There can be, then, at most $v$ different beginning points, and $v$ different ending points, or $2^v$ altogether. But new combinations of types can occur only when at least one type begins or ends, so it appears that at most $2^v$ different combinations can occur in a Petrie matrix.

Since the combinations of types that occur are the same, regardless of the sequence in which the rows of the occurrence matrix happen to be written, an easy partial test of whether or not an occurrence matrix may be petrifiable consists of counting the number of different combinations of types that actually occur. If it is greater than $2^v$, it seems clear that the matrix cannot possibly be petrified. Presumably the ratio of the observed number of type combinations to $2^v$ will give some useful information about how nearly the data could be made to approximate the ideal Petrie form. On the other hand, if the number of observed type combinations is $2^v$ or less, one cannot be certain that the matrix can be petrified. For example, if two or more types happen to have identical end points for their ranges, then there could be fewer than $2^v$ distinct type combinations in a petrifiable matrix, or exactly $2^v$ or fewer in a non-petrifiable matrix. However, the occurrence of $2^v$ or fewer type combinations would strongly suggest that the matrix is probably at least nearly petrifiable.

A different practical implication of this result is that an ideal, petrifiable occurrence matrix is distinctly limited in the amount of chronological information it can provide. If there are more than $2^v$ units — that is, if the number of units is more than twice the number of types — all the units in excess of $2^v$ must exhibit a combination of types which is identical to those that occur in some other unit, and hence any of these excess units must be tied with one or more other units, and chronologically indistinguishable (unless, of course, information additional to that in the occurrence matrix is used). Actually, the number $2^v$ assumes that we include units in which none of the types occur. Realistically, we would hardly try to fit any unit into the sequence unless at least one of the set of types under consideration occurs in that unit, and this means that at most $2^v - 1$ periods can be usefully distinguished if the occurrence matrix is petrifiable.
But if the occurrence matrix is not petrifiable, then the conceivable number of different type combinations that could occur in different units is $2^v$ (or $2^v - 1$, if we exclude the case where no types occur in a unit). If, for example, we were using twelve types, then a petrifiable occurrence matrix could not usefully distinguish more than twenty-three distinct intervals, but these same twelve types could occur in 4095 combinations, if every combination other than total absence of all twelve types occurred.

This is not to say that an occurrence matrix consisting of 4095 units, each with a different combination of one or more of the twelve types would enable us to distinguish 4095 intervals and arrange them in a convincing sequence; on the contrary, such an occurrence matrix would provide no chronological information at all and would be profoundly unsatisfactory. But it does seem to be true, paradoxically, that a less than ideal occurrence matrix, with say thirty or forty rather than twenty-three type combinations occurring, would enable one to distinguish (and arrange in a plausible order) more intervals than would an ideal, petrifiable matrix. At any rate, with many more units than types, which is often the case when the units are graves, the only way to avoid having a relatively large number of tied units, in which identical type combinations occur, is by having an occurrence matrix that cannot be petrified.

To understand more about this, it is useful to think of the Petrie model in terms of probabilities of occurrences of types. The Petrie model, as formalized by Kendall, assumes that in the ideal case, for any given type, its probability of occurrence during any finite interval is zero before a certain point in time, changes instantaneously and discontinuously to unity at this point, remains unity until a second point in time, and then drops back to zero in a second discontinuous instantaneous step. It is because the probability of occurrence has only two values, and changes value only twice, that, if all types really behaved in this ideal way, any one type's occurrence or non-occurrence in a given unit could only enable one to judge whether the unit belonged to an interval somewhere between the beginning and ending points for that type, or to an interval somewhere outside this range, but would not enable one to make any judgements (based on this one type alone) about where within or outside this range the unit belongs.

On the other hand, models of human behaviour that seem more realistic to me will assume that when a type first occurs, it may not be very popular, so that its probability of occurrence rises more or less gradually from zero to some maximum value, then declines more or less gradually to zero again. The implication of this model is that, if we cannot find an arrangement of the occurrence matrix that does not leave some gaps in the occurrences of some types, then an arrangement near the beginning and end points of type ranges (alternative model). Here, he becomes of some importance in deciding what is the best assumption about how the occurrence of types was distributed within the sequence; but, in order intervals has a general high assumption about how the time.

Actually, the only kinds of the monotonic increase from zero to unity, or less gradually to zero again. The implication of this model is that, if we cannot find an arrangement of the occurrence matrix that does not leave some gaps in the occurrences of some types, then an arrangement near the beginning and end points of type ranges is not a very good time indicator. It seems, then, for to eliminate types that seem to be relatively infrequent types are to a relatively brief time span.

Within these limits, different occurrence probably do not and compare different models.

Notice that this general matrices has some resemblance to a relatively brief time span.

Within these limits, different occurrence probably do not and compare different models.
the conceivable number different units is $2^n$ (or a unit). If, for example, a matrix could not intervals, the same other than the only way to avoid identical type combinations other than the precise form of the probability function will lead to somewhat different criteria for deciding what is the best seriation – that is, for deciding how best to allocate the unavoidable gaps in occurrences of some types. In short, a non-petrifiable occurrence matrix admits the possibility of more different observed type combinations, and hence the possibility of distinguishing more intervals in the sequence; but, in order to claim that a particular sequence of these intervals has a generally high validity, one has to justify a more complicated assumption about how the probability of occurrence of a type changes over time.

Actually, the only kinds of assumptions that seem reasonable are variations of the monotonic increase from zero to some maximum, followed by monotonic decrease back to zero, but there does remain room for substantial differences within this general family of models. For example, if a type occurs in only 10 out of 100 units, is it more likely that it occurred intermittently over a fairly long time span (its maximum probability of occurrence always remaining well below unity), or that it occurred with high frequency over a very limited span of time? Either case, of course, is quite possible, but the former situation means that the type (whatever its importance for other purposes) is not a very good time indicator. Sporadic occurrence over long time ranges should manifest itself in preliminary work by occurrence of the type together with other types which themselves seem to have little or no association, and types that have this sporadic occurrence pattern should be eliminated or redefined. It seems, then, for formal seriation algorithms, that we should first eliminate types that seem to occur sporadically, and then assume that other relatively infrequent types are infrequent because their occurrence was limited to a relatively brief time span.

Within these limits, different reasonable models of the probabilities of occurrence probably do not differ very much, but it is important to try out and compare different models.

Notice that this general kind of model for non-petrifiable occurrence matrices has some resemblances to the Brainerd–Robinson model for percentage matrices. In the latter case, it is the observed percentage of each type that is expected to show monotonic increase and decrease to and from a single

---

Models, methods and techniques for seriation

---

395
peak; for occurrence matrices it is then probability of occurrence that is thought to exhibit this pattern, and the probabilities will be reflected in the matrix itself by a tendency for more or longer gaps in occurrences to appear the nearer one comes to the ends of each type's range. There is the important difference that this assumption provides a unique best sequence for a percentage matrix for which the assumption is true, while it does not, by itself, tell one how to find the best sequence for a non-petrifiable occurrence matrix. For a percentage matrix, putting the Robinson matrix into the Robinson form will yield the ideal sequence, but, for an occurrence matrix, as we have seen, it will be impossible to put the analogous G matrix into the Robinson form if the occurrence matrix is not petrifiable.

How, then, should one proceed for occurrence matrices which are not petrifiable? The brute force method would be to try every possible sequence, but it turns out that this would mean, for eight units, over 40,000 possible sequences; nearly 4 million for ten units; and over 20 million million for sixteen units. Of course, since for every sequence there is another that is its exact reverse and therefore equivalent, only half these need be considered, but it helps little to know one only needs to deal with 10 million million possibilities. The requirements quickly overwhelm computers as well as people.

Computer programs such as that of Hole and Shaw (1967) employ various stratagems to seek for relatively good sequences, then seek improvements in these sequences. But this still involves a good deal of trial and error and a limited sampling of all the possibilities. A much more powerful and attractive approach (and probably less costly in computer time) is the multidimensional scaling technique of Shepard (1962) and Kruskal (1964a, 1964b; Kruskal and Carmone, 1969; Kendall, 1969). In particular, Kruskal's MDSAL program has been in use, in several versions, for about seven years, and considerable practical experience with it has been gained. I have found it, with a little practice, easy, fast and reliable. Its availability in an already developed state is a major factor in making seriations that use it relatively easy.

Suffice it to say, here, that multidimensional scaling is a way of seeking configurations of points, one point representing one entity, such that the rank order of distances between points is the inverse of the rank order of similarities between the entities represented by the points. That is, if we had four entities, A, B, C and D, of which the most similar pair were A and C, followed by A and D, B and D, C and D, B and C, and A and B, in that order; then the configuration of points sought would be one in which points A and C have the smallest distance, points A and D are next closest to one another, and so on, with points A and B the most distant from one another. Or, if relationships between entities are expressed in a computer will seek a configuration is the same as the rank of the most dissimilar entities, MDSAL, it is only necessary to the matrix of similarity or dissimilarity.

It may be impossible to satisfy all the points are embedded in there are entities being compared; but, if one is entitled to think that the best possible configuration in the stress, a measure of the distance from the ideal of correctly ranked entities. Normally one computes the optimal configuration of points in few dimensions spaces of progressively fewer dimensions needed can be judged to be manageable, and or by a sudden rise in process a few times with different starting points there is a possibility of stopping and be obtained without some manner of very best possible. If one repeatedly runs it is almost surely the best thing. If the entities are such that into the Robinson form, then stress solution in one dimension, that it is better to seek the best possible. In a 'local minimum with low stress, or if they do not appear one-dimensional (either an arc or somewhat sinusoidal), the factor underlying variation in a seriation, at least not of the tests, the appropriateness of
Models, methods and techniques for seriation

between entities are expressed in terms of differences or dissimilarities, the computer will seek a configuration of points such that the rank order of distances is the same as the rank order of dissimilarities; the points corresponding to the most dissimilar entities being the most distant from one another. In MDSCAL, it is only necessary to inform the machine whether the input is a matrix of similarity or dissimilarity coefficients.

It may be impossible to satisfy the condition sought by this technique unless the points are embedded in a space of nearly as many dimensions as there are entities being compared. In other cases, however, the condition may be met, or nearly met, in a space with much fewer dimensions. In such a case, one is entitled to think that this is because the total pattern of similarities between the entities is mainly due to no more independent factors than the minimum number of dimensions needed. In MDSCAL, the computer finds the best possible configuration in a specified number of dimensions, in terms of the stress, a measure of the extent to which the point configuration departs from the ideal of correctly representing all rank orders of resemblances between entities. Normally one begins with a space of several dimensions, computes the optimal configuration and the corresponding stress, moves to a space of one fewer dimension and repeats the process, and continues with spaces of progressively fewer dimensions. The least number of dimensions needed can be judged by the point where the stress is judged to be unacceptable, and/or by a sudden rise in the stress. In practice, it is well to repeat the process a few times with different arbitrary starting configurations, since there is a possibility of stopping with a configuration that is the best that can be obtained without some major rearrangements of points, but may not be the very best possible. If one repeatedly obtains virtually the same configuration, it is almost surely the best there is (Fig. 8.20).

If the entities are such that their matrix of similarity coefficients can be put into the Robinson form, then multidimensional scaling should lead to a low-stress solution in one dimension. In practice, I strongly agree with Kendall that it is better to seek the best two-dimensional solution. Even if the data fit well into one dimension, there is less danger in two dimensions of getting caught in a 'local minimum'. If the entities do not fit into two dimensions with low stress, or if they do but the configuration is not elongated and essentially one-dimensional (either a nearly straight-line pattern, or at least close to an arc or somewhat sinuous line), then evidently there is more than one major factor underlying variation among the entities, and it is not sensible to attempt a seriation, at least not of the whole set. In this way, multidimensional scaling tests the appropriateness of a seriation model for a given data set.
Another advantage of multidimensional scaling is that, even though it works with ranks rather than metric values, if the number of entities is substantially more than the number of dimensions the result provides a good deal of metric information. Thus, relative distances between points will reflect relative differences. Widely spaced points in the sequence indicate relatively large differences between consecutive units, which could mean either a relatively large time interval, or a period of relatively rapid change. Clusters of points will indicate groups of rather similar units. To some extent, as Hole and Shaw (1967) show, one can get some of this information by looking at the contours of a Robinson matrix, but the spacings provided by a MDSCAL configuration show this more clearly.

The most direct way to use MDSCAL for seriation of an occurrence matrix is to use units as entities, and use the $G$ matrix or some closely related matrix of coefficients of similarities between pairs of units as the input. This is the technique discussed by Kendall (1969). This is excellent, as long as the number of units is not too large. In its present form, MDSCAL cannot handle over about sixty entities, and, while it would certainly be possible to expand this number somewhat, it remains true that the size and expense of the computing job goes up rapidly (from several dollars to several tens of dollars or more) as the number of entities increases. In a symmetric square matrix of similarity coefficients with $n$ rows and $n$ columns there will be $n^2$ components, but the $n$ components along the principal diagonal need not be considered (since they express similarities of entities to themselves) and of the remaining $(n^2 - n)$ coefficients, the half above and to the right of the principal diagonal merely duplicate the half below and to the left. Therefore MDSCAL need only consider $(n^2 - n)/2$ coefficients in order to do its job. But if $n$ is large, this means that the rate of increase is almost proportional to $n^2$.

If the number of types is considerably smaller than the number of units, it would be much more economical to use MDSCAL on $V$, or some other matrix of similarities between pairs of types, rather than pairs of units. Kendall (1969) quite correctly points out that this is not an equivalent operation. Units are thought of as essentially points in time (or at any rate have durations that do not overlap very much), while on the contrary types have ranges that cover appreciable parts of the whole sequence, and one ought to think of them as represented by lines in one dimension, by disks in two dimensions (and, presumably, by hyper-ellipsoids in $k$-dimensional spaces), so that their similarities might be expressed by the extent of overlap between the geometric figures corresponding to the different entities. However, no formal method of doing this has yet been developed. Kendall therefore suggests that for seriating more than 50 or a set of units up in chronological intervals block separately, the operation 'Kendall' (Kendall 1969).

The essential concern to show a way of saving in computer time. As an example, using MDSCAL on a matrix of types is more than one gap type. If one assumes units, such a less-than-good seriation, one ought to think of the more-row types. In seriating types, similarity coefficients of the MDSCAL is petrifiable. We sho in the one-dimension determined by projection on to the major or major points in the stretch occurrence matrix, with think rather concretely so that each row has then the vertical distance of consecutive 1's in any of consecutive 1's in another, which could be...
even though it works
entities is substantially
a good deal of metric
will reflect relative
licate relatively large
an either a relatively
turts is substantially
an either a relatively
ge. Clusters of points
Hole and Shaw
looking at the contours
MDS CAL configuration
an occurrence matrix
closely related matrix
the input. This is the
it, as long as the num-
AL cannot handle over
possible to expand this
ense of the computing
doors or more) as
re matrix of similarity
ponents, but the n
considered (since they
the remaining \((n^2 - n)\)
icipal diagonal merely
DS CAL need only con-
n is large, this means
he number of units, it
, or some other matrix
its Kendall (1969)
 operation. Units are
e have durations that
have ranges that cover
ht to think of them as
dimensions (and, pre-
so that their similari-
etween the geometric
t, no formal method of
suggests that for seriating
more than 50 or at most 100 units the best method would be to break the full
set of units up into interpenetrating blocks (presumably on the basis of
chronological information available from other methods) 'and to treat each
block separately, fitting together the seriations thus obtained in a subsequent
operation' (Kendall, 1969, p. 76).

The essential contribution I have to offer here, on the level of technique, is
to show a way of overcoming much of the difficulty pointed out by Kendall in
using MDS CAL on types rather than units. This means that, where the number
of types is more than a little smaller than the number of units, a substantial
saving in computer costs can be effected. And, especially, the technique I will
describe makes it feasible to seriate several hundred units in a single operation.

As an example of the economies offered, suppose we have 100 units
and 20 types. This gives us a choice of dealing with 4950 pairs of units, or 190
pairs of types. Incidentally, if at least one type occurs in all units, and if no
two units are identical, this implies that at most there could be a subset of
39 of the 100 units which, by themselves, could be petrified. Each of the other
61 units must, wherever it is fitted in, introduce at least one gap. While I have
not made a thorough analysis of this situation, I believe there need not be
more than one gap for each of these units, for an average of three gaps per
type. If one assumes that each type occurs in not less than 10 or more than 50
units, such a less-than-ideal occurrence matrix still seems capable of a quite
good seriation, especially since there is no reason why more gaps cannot be
allocated to the more often occurring types, and fewer to the less often occur­
ring types.

In seriating types rather than units, the first thing one must do is find a
similarity coefficient that will make possible a definite and sensible inter­
pretation of the MDS CAL results if applied to an occurrence matrix which in fact
is petrifiable. We should try to get a coefficient that makes sense if the points
in the one-dimensional MDS CAL configuration (or, more exactly, the points
determined by projecting the points of the two-dimensional MDS CAL result
on to the major or main axis of their configuration) are regarded as the middle
points in the stretches of consecutive 1's in the corresponding column of the
occurrence matrix, when it is ideally seriated and in the Petrie form. If we
think rather concretely of the occurrence matrix as drawn to a uniform scale,
so that each row has the same physical width (distance from top to bottom),
then the vertical distance from the level of the middle point of the stretch of
consecutive 1's in any column, to the level of the middle point of the stretch
of consecutive 1's in any other column, will be some quite specific scalar dis­
tance, which could be measured in centimetres or inches. Of course, this does
not mean that the petrified occurrence matrix provides a metric time scale, since what we are measuring are distances between rank positions of different units. Nevertheless, it gives us a definite idea of a distance which the coefficients of the input matrix for MDSCAL, might represent. Rather than using \( V \) as the input for MDSCAL, we should derive a different matrix, whose components are more appropriate coefficients for expressing distances between midpoints of type occurrence stretches.

It is useful, without any loss of precision, to play a little more with our concrete picture of an occurrence matrix, by (1) replacing the zeros by blanks, and (2) drawing each 1 as an unadorned vertical line, stretching from the very bottom to the very top of its row. This means that a bunch of consecutive 1's will simply be represented by an unbroken vertical line, and, in a Petrie matrix, there will be only a single uninterrupted line in each column. Notice also that the total line length in any column is a constant, proportional to the number of units in which the type represented by that column occurs, and is unchanged by any permutation of the rows of the occurrence matrix. In fact, if we make each row of unit width, then the total line length in each column will just equal the number of units in which the given type occurs.

With respect to any two types, \( i \) and \( j \), whether the matrix is petrifiable or not, all units must fall into one of four categories: \( i \) and \( j \) both present, \( i \) but not \( j \), \( j \) but not \( i \), or neither \( i \) nor \( j \). We can denote the number of units where both \( i \) and \( j \) occur by \( a \); the number with \( i \) but not \( j \) by \( b \); the number with \( j \) but not \( i \) by \( c \); and the number with neither \( i \) nor \( j \) by \( d \). The sum of \( a + b + c + d \) must equal \( g \), the total number of units, and is the distance from top to bottom of the occurrence matrix, if all rows have unit width. Notice that \( a \), \( b \), \( c \) and \( d \) are not changed by permutations of the rows of the occurrence matrix, and can be determined quite easily regardless of the order in which the units are listed.

Now, if the occurrence matrix is petrifiable, there are three possibilities for the ranges of types \( i \) and \( j \). Either they partially overlap one another, as in Fig. 9.1, they do not overlap at all, as in Fig. 9.2, or the range or span of one type completely includes the range of the other type, as in Fig. 9.3. Each of these cases needs to be looked at more closely.

If types \( i \) and \( j \) overlap and, let us say, \( i \) begins earlier than \( j \), as in Fig. 9.1, it means that in some units both types occur, in other units \( i \) occurs but not \( j \), and in others \( j \) occurs but not \( i \). That is, \( a \), \( b \) and \( c \) are all greater than zero (\( d \) may or may not be greater than zero; its value is not critical). The distance that we want to derive is \( x \), the vertical distance between the midpoints of the ranges of the two types. In this situation, \( x \) can be found by reasoning that the...
vides a metric time scale, en rank positions of differa of a distance which the
represent. Rather than a different matrix, whose pressing distances between
a little more with our consid­
ing the zeros by blanks, e, stretching from the very a bunch of consecutive 1's cal line, and, in a Petrie he in each column. Notice instant, proportional to the that column occurs, and is occurrence matrix. In fact, ine length in each column even type occurs.
the matrix is petrifiable or i and j both present, i but the number of units where j by b; the number with j i d. The sum of $a + b + c + d$ distance from top to bottom h. Notice that $a, b, c$ and d he occurrence matrix, and order in which the units are are three possibilities for overlap one another, as in r the range or span of one pe, as in Fig. 9.3. Each of earlier than j, as in Fig. 9.1, her units i occurs but not j, are all greater than zero ($d$ not critical). The distance tween the midpoints of the found by reasoning that the
Models in archaeology

total distance from the lower endpoint of type $i$ to the upper endpoint of type $j$ is $a + b + c$. The total range of type $i$ is $a + b$; therefore the distance from the lower endpoint of type $i$ to its midpoint is $(a + b)/2$. By a similar argument, the distance from the upper endpoint of type $j$ to its midpoint is $(a + c)/2$. The distance between the two midpoints, $x$, is then $a + b + c - (a + b)/2 - (a + c)/2$, which simplifies to

$$x = (b + c)/2 \quad (1)$$

In the case where there is no overlap at all in the ranges of types $i$ and $j$, we know this only because there is no unit in which $i$ and $j$ both occur, so that $a$ is zero. We have no way of knowing, on the basis only of information about the occurrences of these two types, whether $i$ and $j$ barely fail to overlap (as in the example on the left side of Fig. 9.2) or whether they occur at opposite extremes of the total sequence (as in the example on the right side of Fig. 9.2). The shortest possible distance between their midpoints is $(b + c)/2$, while the greatest possible distance is $g -(b + c)/2$. If one reasons that any distance within this range is as probable as any other, then the statistically expected value, the expected long-run average for a large number of cases of non-overlapping pairs, is the midpoint of the range of possible distances, which is $[(b + c)/2 + g -(b + c)/2]/2$, which simplifies to $g/2$. Somewhat surprisingly, this expected value is independent of $b$ and $c$; that is, it does not depend on the number of units in which either type occurs, so long as the types never occur together. Now in the earlier case, where there was some overlap in the type ranges, the largest possible value for $x$ would occur if $a = 1$ and $d = 0$; that is, if one or the other of the types occurs in all units, but the two types occur together in just one unit. In this case, $b + c = g - 1$, and, from Equation 1, the value of $x$ will be $(g - 1)/2$.

Thus, the expected average distance between midpoints of any two types that do not overlap is greater than the maximum possible distance between midpoints of any types that do overlap. This suggests that all type pairs that do not overlap be given distance coefficients which are tied for some maximum value. Since MDSCAL operates on the ranks of the coefficients, it does not matter what this maximum value is, provided it is equal to $g/2$ or greater. However, there is more than one option in MDSCAL for treatment of ties, and it seems best to choose the option that permits the computer to seek a configuration in which distances between points corresponding to tied coefficients have maximum freedom to be unequal, so long as rank order is not violated for any non-tied coefficients. Non-overlapping type pairs are given tied coefficients not because we know distances between midpoints of their ranges are all the same, but because we do not have any information on which to judge that some are greater than others.

This approach amounts to the recognition, independent of specific instances where either $a$ or $b$ might be greater than zero, and $b$ would be between midpoints of $a/2$, or simply $b/2$. The expected distance between midpoints of any type is $b/4$, for $b$ between midpoints of $a/2$, or simply $b/2$. Thus, all distances within the expected distance, for $b/4$ for types $i$ and $j$, (use subscripts to link certain) be smaller than any overlapping pair

$$b_{ij}/4$$

where $b_{ii}$ and $c_{ij}$ must be zero. There is no guarantee that if a distance formula $x = (b + c)/2$, my procedure has been satisfied, that is, whenever one member of one other type also occurs in a unit, the distance less than the
The upper endpoint of type 'i' is the distance from the endpoint of type 'j' by a similar argument, the midpoint is \((a + c)/2\). The midpoint of \((b + c)/2 - (a + c)/2\) is

\[ (a + c)/2 \] (1)

The ranges of types 'i' and 'j' both occur, so that only information about 'i' barely fail to overlap (as they occur at opposite the right side of Fig. 9.2). (b + c)/2, while the minimum distance that any distance with statistical expected value, of non-overlapping types, which is \((b + c)/2 + g\) – assuming this expected value and on the number of units occur together. Now in the type ranges, the largest that is, if one or the other of them are both greater than zero, but 'c' is zero (if the range of 'j' completely included that of type 'i', then of course 'a' and 'c' would be greater than zero, and 'b' would be zero). In this case, the maximum possible distance between midpoints of the two types, when the matrix is petrified, is \((a + b)/2 - a/2\), or simply \(b/2\). The shortest possible distance is zero. Again assuming that all distances within these limits are equally probable, the statistically expected distance, for the average of such cases, is \(b/4\). The condition that \(b_{ij}/4\) for types 'i' and 'j' (where 'i' includes the range of 'j', and where we must now use subscripts to link different variables to the pairs of types to which they pertain) be smaller than the distance \(x_{mn}\) between midpoints of the ranges of any overlapping pair of types 'm' and 'n', is that

\[ b_{ij} < 2(b_{mn} + c_{mn}) \] (2)

where \(b_{mn}\) and \(c_{mn}\) must both be integers greater than or equal to unity. There is no guarantee that this condition will always be met, and perhaps the distance formula \(x = (b + c)/4\) should be used whenever 'b' or 'c' is zero. However, my procedure has been the simpler one of setting \(x = 0\) in all these cases. That is, whenever one member of a pair of types occurs only in units where the other type also occurs, they are treated as tied with all other such pairs for a distance less than the distances between midpoints of any pairs which overlap.
Models in archaeology

(such pairs must have a distance between range midpoints of at least unity). This, again, has given good results.

As a matter of efficient design, in order to get a maximum amount of chronological information for a maximum number of units for a minimum number of types, it would be good to avoid including more than one member of any pair of types for which \( a \) is large relative to \( b + c \); that is, where the number of units in which both types occur is large relative to the number in which either type occurs without the other, since in such cases the types are relatively redundant. One would also want to avoid using any type that occurs in only a very small proportion of the total set of units since such a scarce type (however significant it might be in marking off those few units as different from others) will not help to seriate many units. Thus, in a well-designed study there should be few or no cases where the range of one type is completely included in the range of another, and also few cases where the condition specified by Equation 2 is not satisfied. Also, in any real data set, almost surely not petrifiable, it is not very likely that any type will never occur except in units where some other type occurs. For all these reasons, the choice between setting \( x = 0 \) or \( x = (b + c) / 4 \) when \( b \) or \( c \) is zero does not seem to constitute a very important practical problem.

In summary, a distance (or dissimilarity) coefficient has been developed which is very easy to compute, and which expresses distances between midpoints of the ranges of types, for petrified occurrence matrices, and which is untroubled by the fact that the ranges themselves are stretches of finite length, rather than points. It is convenient to standardize the distances by dividing by \( g / 2 \), and, since \( x / g \) will always be less than \( 1 / 2 \) for any overlapping pairs of types, it is also convenient to set the distance equal to unity for non-overlapping type pairs. All the distance coefficients between pairs of types can then be arranged in a square symmetric matrix, \( W \), analogous to \( V \), whose \((i,j)\)th component is

\[
\begin{align*}
\varpi_{ij} &= 0 \text{ if either } b_{ij} \text{ or } c_{ij} \text{ is zero} \\
\varpi_{ij} &= 1 \text{ if } a_{ij} \text{ is zero} \\
\varpi_{ij} &= (b_{ij} + c_{ij}) / g \text{ otherwise}
\end{align*}
\]

Incidentally, if one works out \( a_{ij}, b_{ij}, c_{ij} \) and \( d_{ij} \) as functions of \( v_{ij} \) (the \((i,j)\)th component of Kendall’s \( V \) matrix), it turns out that: since \( a_{ij} = v_{ij}, b_{ij} = v_{ii} - v_{ij}, \) and \( c_{ij} = v_{jj} - v_{ij}, \) the above three equations could be expressed as

\[
\begin{align*}
\varpi_{ij} &= 0 \text{ if either } v_{ii} = v_{jj} \text{ or } v_{ij} = v_{jj} \\
\varpi_{ij} &= 1 \text{ if } v_{ij} \text{ is zero} \\
\varpi_{ij} &= (v_{ii} + v_{jj} - 2v_{ij}) / g \text{ otherwise}
\end{align*}
\]
Models, methods and techniques for seriation

Suppose we have an occurrence matrix in the Petrie form, compute the $W$ matrix of distance coefficients according to the formulae given above, and also measure with a ruler the vertical distances between the midpoints of the consecutive bunches of 1's in each column. The computed distance coefficients will have the same rank order as the actually measured distances, subject to the limitations that we may have given slightly too low a rank to the distances between midpoints of some of the pairs where one type includes the range of the other; that we may have given slightly too high a rank to the distances between midpoints for some of the pairs where there is no overlap at all; and that in either of these cases we have had to treat as ties some distances between which we may be able to observe differences when we actually measure the occurrence matrix.

This means that, with these minor exceptions, the distances measured by our set of coefficients are monotonically related to actual distances measured along a single dimension; namely, up and down, along columns and spanning rows of the occurrence matrix. Therefore a multidimensional scaling of the $w_{ij}$ should result in a low stress configuration in one dimension. Moreover, the measured distance between two points in the MDSCAL output configuration should be directly proportional to the number of rows separating midpoints of the ranges of the corresponding types in the petrified occurrence matrix.

The essential point, of course, is that we do not need to use any particular order of the occurrence matrix in order to compute the $w_{ij}$; which are uniquely determined by Equations $3a$, $3b$ and $3c$, independently of any particular sequence of units. Hence multidimensional scaling of a $W$ matrix which results in a good one-dimensional solution not only tells us that the occurrence matrix $A$ is petrifiable (or nearly so), it also provides a great deal of useful information about the best sequence of rows for $A$.

One could perfectly well compute the $W$ matrix defined by Equations $3a$, $3b$ and $3c$ and use it as the input for MDSCAL. Actually, I computed a different matrix, $S$, whose $(i,j)$th component, $s_{ij}$, is $1 - w_{ij}$. The rules for computing $s_{ij}$ are then

$$s_{ij} = 0 \text{ if } a_{ij} \text{ is zero} \quad (4a)$$

$$s_{ij} = 1 \text{ if either } b_{ij} \text{ or } c_{ij} \text{ is zero} \quad (4b)$$

$$s_{ij} = (a_{ij} + d_{ij}) / g \text{ otherwise} \quad (4c)$$

It is worth noting that in the general case, where neither $a$, $b$ nor $c$ is zero—that is, where types $i$ and $j$ partially overlap—the coefficient $s_{ij}$ is just the simple matching coefficient widely used by numerical taxonomists to compare...
Models in archaeology

units (not types) (Sokal and Sneath, 1963, p. 133). In all cases, the $s_{ij}$ are similarities, rather than differences, which means only that MDSCAL should attempt to put the pairs with highest $r$'s closest together rather than furthest apart.

I should add that limited experimentation with some other coefficients for which there is a less clear rationale, including what Sokal and Sneath call $S_j$, the Jaccard coefficient (which counts only positive matches), suggests that they also may give fairly good results. At any rate, with reasonable data sets they seem to yield low stress one-dimensional MDSCAL solutions. Notably bad results have been obtained from Pearson’s $r$, which would be suitable if the probability of occurrence of one type were a linear function of the probability of occurrence of another type. For the seriation models being used here, however, $r$ is highly inappropriate.

Granted that Equations 3a, 3b and 3c or 4a, 4b and 4c should give good results when the occurrence matrix is in principle petrifiable, the practical question is how it works when the matrix cannot quite be petrified, or when the matrix is rather far from petrifiable. I have not made a formal analysis of the consequences, but have experimented with non-petrifiable sets of test data. The results presented below indicate that if the matrix is near enough to being petrifiable to justify seriation at all, very useful results are still obtained by this technique.

To complete description of the technique, it is necessary to explain how information from the MDSCAL result from the $S$ matrix can be used to seriate $A$. This is best illustrated by an example. Fig. 9.4 shows a set of invented data consisting of twenty-five units and ten types (I's indicate occurrences; non-occurrences are indicated by blanks), arranged in what by fiat I declare to be their 'true' sequence. As can be seen, this occurrence matrix is petrified. Note that the order in which types appear is not the same as that in which they disappear; some types appear earlier but disappear later than some other types.

The first step is to compute the matrix of similarity coefficients. For types $A$ and $B$, for example, all units with $A$ also have $B$, so $s_{AB}$ is unity. Type $A$ occurs in none of the same units as type $C$, so $s_{AC}$ is zero. Type $I$ and type $J$ occur together in two units and are both absent in sixteen, so $s_{IJ}$ is $(2 + 16) / 25$, or $0.72$. In all there are forty-five different pairs of types here, and all forty-five coefficients can be computed fairly quickly by hand, though for larger data sets a very simple computer program to do this is convenient. Remember that it does not matter what order the units are listed in; any order leads to the identical $S$ matrix. A change in the order in which types are listed would still lead to the same forty-five coefficients, but the arrangement of the rows and columns of the best two-dimensional essentially one-dimensional MDSCAL configur (Fig. 9.6). A refinement if anything worth the trouble is to use the central axis in numbers of units, present and point then point $A$ should be point $J$ should be matrix. The distance 25 to $(1+3.5) or 2$ points predicted for
Models, methods and techniques for seriation

133). In all cases, the $s_{ij}$ are means only that MDSCAL should it together rather than furthest with some other coefficients for what Sokal and Sneath call $S_{ij}$, positive matches), suggests that rate, with reasonable data sets DSCAL solutions. Notably bad which would be suitable if the function of the probability models being used here, how-

134 and 4c should give good re- place petrifiable, the practical quite be petrified, or when ot made a formal analysis of non-petrifiable sets of test the matrix is near enough to ful results are still obtained necessary to explain how in- rix can be used to seriate $A$. rows a set of invented data indicate occurrences; non-what by flat I declare to be ce matrix is petrified. Note me as that in which they ater than some other types. nitity coefficients. For types $\beta$, so $s_{AB}$ is unity. Type $A$ reo. Type $I$ and type $J$ oc- ren, so $s_{ij}$ is $(2 + 16)/25$, or es here, and all forty-five d, though for larger data avement. Remember that i; any order leads to the pse are listed would still gement of the rows and

columns of the $S$ matrix will be different. If random starting configurations are used for MDSCAL, changes in the rows and columns of $S$ will not matter. The resulting $S$ matrix was used as input for MDSCAL, and the resulting best two-dimensional configuration is shown in Fig. 9.5. It is elongated and essentially one-dimensional, and has a very low stress value, 0.022, which indicates an excellent fit.

In order to scale more precisely the predicted midpoint for the range of occurrence for each type, a line was fitted by eye to the longest axis of the MDSCAL configuration, and perpendiculars were drawn to the individual points (Fig. 9.6). A refinement would be to fit this line by least squares, but I doubt if anything worth the trouble would be gained. The distances were read off the central axis in centimetres, and multiplied by a scale factor to convert to numbers of units. The idea here is that since there are two units with type $A$ present and point $A$ of Fig. 9.6) is supposed to be the midpoint for type $A$, then point $A$ should be one unit above the lower endpoint of the optimally seriated matrix. Similarly, since there are seven units in which type $J$ occurs, point $J$ should be 3.5 units below the upper endpoint of the optimally seriated matrix. The distance between points $A$ and $J$, then, should correspond to $25 - (1 + 3.5)$ or 20 units. Applying the resulting scale factor to all the midpoints predicted by MDSCAL, the predicted midpoints can be compared with

FIG. 9.4. Occurrence matrix for invented ideal data in 'true' sequence of units. Rows represent twenty-five units; columns A to J represent ten types. Occurrences are consecutive for all types.

![Occurrence matrix](https://via.placeholder.com/150)

![Predicted midpoints](https://via.placeholder.com/150)
the true midpoints found from measuring Fig. 9.4. The results, given in Fig. 9.7, show a quite good correspondence and perfect rank ordering, except for reversal of the nearly tied types I and H. Incidentally, this is a good example of how MDSCAL, in spite of working with just rank orders of distances, can recover a good degree of localization of points than dimerization of units as in a separated context proportional to the number of points.

The most obvious dimerization of units is a case of the nearly tied types I and H. Incidentally, this is a good example of how MDSCAL, in spite of working with just rank orders of distances, can recover a good degree of localization of points than dimerization of units as in a separated context proportional to the number of points.
Models, methods and techniques for seriation

recover a good deal of metric information, if there are substantially more points than dimensions.

The most obvious way to use these results about types in order to obtain a seriation of units is to draw every predicted midpoint at the predicted height in a separated column, and draw to scale a vertical line whose length is proportional to the number of units in which the corresponding type occurs. The result looks like Figs. 9.1, 9.2 or 9.3, except that there are lines representing all the types, rather than a single pair. If one then lays a straight edge horizontally across the diagram, it will cross some specific combination of vertical lines, corresponding to some specific combination of types that are predicted to occur together. By keeping the straight edge horizontal and sliding it from the bottom edge of the figure to the top edge, noting in order each different predicted combination of types that appears, one can derive a predicted sequence of occurrence of predicted combinations. This sequence can be compared with the observed combinations, which can now be listed in their predicted sequence. For the ideal data of Fig. 9.4, this method led to perfect recovery of the true sequence, except that the first unit, with only type B present, was moved up to join the fourth unit, which also has just type B.

It is important to dwell on this little error for a moment, because it has nothing to do with the specific seriation technique used, and would have been an inevitable consequence of any attempt to seriate the occurrence matrix of

FIG. 9.7. Predicted occurrence midpoints derived from Fig. 9.6, plotted against true midpoints derived from Fig. 9.4.
Models in archaeology

Fig. 9.4. The point is that, at least in terms of the types being used, units 1 and 4 are identical, and hence no seriation technique could possibly lead to putting them into different intervals of the sequence. Units 1 and 4 are only one example here of tied units; others are units 2 and 3, or units 21–25. Examination of Fig. 9.4 shows that only thirteen different type combinations occur, so that the best any possible seriation technique could do with this occurrence matrix would be to order the twenty-five units in terms of thirteen distinguishable intervals. Furthermore, units 1 and 4 here give an example of identical units which are not actually consecutive (even though in Fig. 9.4 the matrix is petrified), and must necessarily be re-ordered so that they are consecutive by any seriation technique. This should remind one that it is possible to find a seriation sequence, sometimes, which ‘improves’ on the true sequence, and that, although the results are likely to be correct in their main outlines, one should not put too much trust in the fine details.

A much more serious difficulty is that, with real data, a good many of the type combinations actually exhibited by one or more units will not correspond to any of the predicted combinations. One can use the predicted sequence of combinations as a sort of skeleton matrix (a term I have taken from Bordaz and Bordaz, 1970), and fit the non-predicted but observed combinations in to the places where one judges they look best. I did this by hand in a day or so for the 109 different observed type combinations in the real data set I will discuss below, and obtained a seriation which looked very good in terms of Kendall’s ‘concentration principle’. That is, there were relatively few zeros interspersed among the 1’s in each column of the matrix, and most zeros occurred near the ends of the maximum range of occurrence for each type. The effort was very much less than would have been required without the skeleton matrix derived from the MDSCAL results. Nevertheless, it seemed desirable to do this part of the seriation by computer also.

In order to do this, one must formalize the judgement procedures used for fitting observed type combinations into the predicted sequence. This brings us back to the subject of goodness of fit criteria, for the problem is where best to fit units that do not fit into an idealized sequence. We must invoke some definite (if simple) model of cultural change in order to decide how best to handle these units.

A number of models might be tried, but I began with the relatively simple assumption that the probability of occurrence of a given type in a given unit is specified by a normal curve, whose mean is centred on the predicted midpoint of the type’s range of occurrence, and whose standard deviation is proportional to the number of units in which this type occurs. A normal curve
Types being used, units 1 and 4 possibly lead to putting units 1 and 4 are only one except units 21–25. Examination of the occurrence matrix of thirteen distinguishable examples of identical units. In Fig. 9.4, the matrix is in its occurrence or non-occurrence of any other types; it may be that contingencies of this kind are sometimes important, but they would lead to far more complex models and, furthermore, would suggest that it was false to assume that change over time was the only major factor accounting for differences between units, and would therefore suggest that seriation as such was an inadequate way for dealing with the occurrence matrix.

But if the probability of each type’s occurrence in a given unit depends only on that unit’s distance from the occurrence midpoint for the type, then the probability of the joint occurrence of a given combination of types in a given unit will be just the product of the probabilities of each individual type’s occurrence in that unit. This follows from the basic rule of probability calculus that the probability of the joint occurrence of two or more independent events can be found by multiplying all the probabilities of the occurrences of the events themselves. The most probable place for a given unit in the sequence, then, will be that place which maximizes the joint probability of occurrence for all the types which occur in the unit. This joint probability can be found by multiplying together the normal curve functions corresponding to each type which occurs in the unit (each function will have different parameters corresponding to the different predicted midpoints and different numbers of units in which each type occurs). The location corresponding to the maximum joint probability can then be found by differential calculus, by taking the derivative of the joint probability function, and solving for zero.

The result turns out to be very simple. To find the predicted occurrence point for some unit $U$, sum predicted midpoints divided by the square of the number of units in which the type occurs, for each type, and divide by the sum of the squared reciprocals of the numbers of units in which each type occurs. If $L_u$ is the predicted location for unit $U$, $m_i$ is the predicted midpoint for type $i$ (obtained from the MDSCAL configuration), and $g_i$ is the number of units in which type $i$ occurs (that is, component $g_{ii}$ of Kendall’s G matrix), then

$$L_u = \frac{\sum_i m_i g_i^2}{\sum_i g_i^2}$$

where the summation is over all types that occur in unit $U$. Notice that if type $A$ is the only type that occurs in unit $U$, then the predicted location for unit $U$ is just the predicted midpoint for type $A$. If all the types that occur in unit
Models in archaeology

\( U \) occur in the same total number of other units, so that \( g_i \) is the same for all the types, then the predicted midpoint for unit \( U \) is the arithmetic mean of the predicted midpoints of all types that occur in unit \( U \). In general, where the \( g_i \)'s will be different, \( L_U \) is a kind of weighted mean.

Also, notice that the result does not depend on what values we assume for the constant of proportionality that relates the standard deviations of the 'probability of occurrence' functions to the numbers of units in which each type occurs; it only requires that we assume that the proportionality factor is the same for all types. That is, it turns out that it does not matter whether we assume that, relative to the ranges that would be observed if all occurrences of a type were to be consecutive, the probability of occurrence rises and declines gradually over a much longer time interval, or whether we assume that the probability peaks much more sharply over a relatively short interval.

As long as we assume that for all types the probability curves are normal, and that for all types the proportionality factor relating the spread of the curve to the number of units in which the type occurs is the same, we do not need to make any specific assumption at all about what the proportionality factor is, in order to get specific and unique predicted locations for all the units. This means that one can make a distinctly weaker assumption about the details of culture change than one might expect, and still get a specific result for seriation.

However, if Equation 5 is used to compute the predicted occurrence point of a unit in which one relatively common type and one relatively scarce type occur, one finds that the predicted point is relatively close to the predicted midpoint of the scarcer type. It seems as if concentration or bunching of the occurrences of scarce types is given too much weight, at the expense of looseness or intermittence in occurrences of more common types.

A more intuitively satisfying prediction equation comes from simply summing the ratios of predicted midpoints to numbers of occurrences, and dividing by the summed reciprocals of numbers of occurrences, without squaring anything, thus,

\[
L'_U = \frac{\sum_i m_i / g_i}{\sum_i 1 / g_i} \quad (6)
\]

Again, \( L'_U \) is just the arithmetic mean of the predicted midpoints of all types that occur in unit \( U \), if all these types occur in the same number of units. If they do not, \( L'_U \) is a somewhat different weighted mean than \( L_U \): a weighted mean that gives less emphasis to concentration for scarce types. An appealing feature of \( L'_U \) is that if types \( A \) and \( B \) are the only types in unit \( U \), and if the predicted ranges of \( A \) and \( B \) are conceptually the same, the predicted occurrence point for \( U \) is midway between the immediately preceding and succeeding occurrence values of \( A \) and \( B \), and the predicted occurrence point for \( U \) will be different in the absence of \( A \) and \( B \).

One could consider

\[
\text{Equation 5} \quad \text{for a moderate} = \text{MDSCAL and the same time the}
\]

predicted range where the parameter with other probability function of all the types that all the types that maximize this is considerably more

Notice that predicted midpoint of each type occurrence, nor on occurrence matrix, a desirable that the predicted common be a right columns of ranking units in to infer that some terms of the series wider seriations.

Equation 6 is for a moderate space have written a single row by row in a MDS and that the same time the
and on what values we assume for the standard deviations of the numbers of units in which each type occurs, the proportionality factor could be observed if all occurrences of a type were observed over a relatively short interval. Probability curves are normal, and the spread of the curve to a given point is the same, we do not need to decide the proportionality factor is, in general, where the mean of the arithmetic mean of the number of units in which each type occurs is derived.

Equation 5 comes from simply stating the predicted occurrence point of a type as close to the predicted occurrence point of another type as the weight, or the expected number of occurrences, of the least common type.

Equation 6 is so simple that it would not be too difficult to compute by hand for a moderate number of units, but since the object is to make life easier, I have written a short computer program which reads in the occurrence matrix row by row in any sequence, reads in the predicted midpoints derived from MDSCAL and the numbers of units in which each type occurs (obtained at the same time the S matrix is computed for MDSCAL input), uses Equation 6

\[
L(k)_U = \frac{\sum m_i k^k}{\sum 1/k_i^k}
\]

where the parameter \( k \) happens to be 2 or 1. It would be of interest to experiment with other values for \( k \). Another variation would be to make the joint probability function not only the product of the probability of occurrence of all the types that occur in a unit, but also the probability of non-occurrence of all the types that fail to occur in the unit, and then seek the location that would maximize this joint probability. But the mathematics of doing this seem considerably more difficult, and it may not be worthwhile.

Notice that Equation 7 and its special cases, 5 and 6, require only the predicted midpoints derived from MDSCAL, plus the number of units in which each type occurs. They do not depend at all on any particular starting sequence, nor on any trial assumptions about the best ordering of the occurrence matrix, nor on any iterative or successive approximations method. It is desirable that the predicted midpoints be scaled in terms of numbers of units, so that the predicted occurrence points will also be in terms of units. They will commonly be expressed in terms of fractional parts of a unit (as in the far right columns of Figs. 9.8, 9.10 and 9.15), but this causes no difficulties for ranking units in their predicted sequence. In fact, these values could be used to infer that some units are nearly contemporaneous, or at least nearly tied in terms of the seriation, while other consecutive units may be separated by much wider seriational gaps.

Equation 6 is so simple that it would not be too difficult to compute by hand for a moderate number of units, but since the object is to make life easier, I have written a short computer program which reads in the occurrence matrix row by row in any sequence, reads in the predicted midpoints derived from MDSCAL and the numbers of units in which each type occurs (obtained at the same time the S matrix is computed for MDSCAL input), uses Equation 6

\[
L(k)_U = \frac{\sum m_i k^k}{\sum 1/k_i^k}
\]
Models in archaeology

to compute a predicted occurrence point for each unit, sorts the units according to this predicted order of occurrence, and prints this sorted sequence as a seriation of the occurrence matrix. Judged by the numbers of zeros interspersed in the occurrence ranges of types, the results have not been quite as good as the best results from fitting units by judgement to the skeleton plot derived from the same MDS CAL results, but they are nearly as good; the differences are because the computer is actually trying to satisfy a slightly different criterion, and the program takes almost no time. On a rather slow machine (IBM 1130) it took only a few seconds to seriate the same 129 units that took me more than a day to do by the skeleton plot technique.

Fig. 9.8 shows a randomized input sequence for the occurrence matrix of Fig. 9.4, plus the seriation produced by the computer program. It is identical to the postulated true sequence of Fig. 9.4, except that (1) for reasons already discussed, unit 1 is moved next to unit 4, and (2) units 5 and 6 are reversed. This second change is slight in terms of the total sequence, but highly significant. It is a clear violation of the concentration principle; a non-petrified sequence is preferred even though a petrified sequence is in fact possible. The model behind Equation 6 leads to a criterion that is somewhat different from the Petrie criterion. The Petrie model, as formalized by Kendall, requires only that each unit should preferably occur somewhere within the ranges of consecutive occurrences of all the types which occur in the unit. The model represented by Equation 6 puts emphasis on a unit’s occurring relatively near the midpoints of the ranges of the types that occur in the unit, and finds a compromise value between the contradictory demands of the different midpoints that is weighted most heavily for scarcest types. This can lead to a preference for sequences that differ a little from the Petrie ideal. It is important to realize that seriations can be judged by criteria other than the concentration criterion; in what ways, and in what circumstances these other criteria may lead to seriations which are likely to be closer to true chronologies is a matter that needs more exploration.

For a second test of the technique, I invented another occurrence matrix, consisting of fifty units and ten types, for which the ‘true’ sequence (Fig. 9.9) is not a Petrie matrix, but contains some zeros (blanks) interspersed in the occurrence ranges of the types. Here the same steps were followed; a matrix of similarity coefficients was computed, and a two-dimensional MDS CAL configuration was obtained and inspected. It was again linear, with low stress (0.034). Distances of predicted midpoints were scaled from the long axis of the MDS CAL configuration, and plotted against the true mean points of type occurrences, derived from Fig. 9.9. The resulting agreement (Fig. 9.10) was
For each unit, sorts the units according to the numbers of zeros interspersed in the results have been not quite as good by judgement to the skeleton plot derived they are nearly as good; the differences are trying to satisfy a slightly different crit-

criterion to the skeleton plot derived 

nearly as good; the differences are trying to satisfy a slightly different crit-

not time. On a rather slow machine is to seriate the same 129 units that took 

sequence for the occurrence matrix of 

he computer program. It is identical to Fig. 9.9, except that (1) for reasons already 

total sequence, but highly significa-

tion that is somewhat different from 

ormalized by Kendall, requires only 

somewhere within the ranges of 

to occur in the unit. The model 
at occur in the unit, and finds a 

7 demands of the different mid-

est types. This can lead to a 

from the Petrie ideal. It is impor-

criteria other than the concent-

circumstances these other criteria 

closer to true chronologies is a 

of another occurrence matrix, 

true' sequence (Fig. 9.9) (blanks) interspersed in the 

 were followed; a matrix 

dimensional MDSAL contain 

linear, with low stress 

d from the long axis of 

true mean points of type 

agreement (Fig. 9.10) was 

<table>
<thead>
<tr>
<th>INPUT SEQUENCE</th>
<th>TYPES</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BEST FIT FOR OBSERVED COMBINATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SERIATION RESULT</th>
<th>TYPES</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BEST FIT FOR OBSERVED COMBINATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

---

Fig. 9.8. Randomized rearrangement of the sequence of units (rows) in the occurrence matrix of Fig. 9.4, and result obtained from re-ordering the input sequence according to the predicted midpoints from Fig. 9.6, using Equation 6 and sorting units in order of predicted occurrence points.
Fig. 9.9. Occurrence matrix for invented non-ideal data in ‘true’ sequence of units, for fifty units and ten types.
Occurrences are not entirely consecutive for any type.

actually better than for the ‘ideal’ data of the first example. The same computer program as before was used to compute a predicted point of occurrence for each unit and to sort a randomized input matrix into this predicted sequence. The result (Fig. 9.11) is generally very similar to the ‘true’ sequence (Fig. 9.9) with two differences. First, all units which exhibit identical combinations of types are tied, and are listed consecutively (an inevitable consequence of any seriation technique). Second, by the Petrie criterion (as well as
Models, methods and techniques for seriation

by the criterion of Equation 6), the computer result is a distinct improvement over the 'true' sequence. That is, the occurrences of types are more nearly consecutive in Fig. 9.11 than in Fig. 9.9. The technique has been a little too successful.

Experiences like these should encourage a healthy scepticism about the detailed accuracy of any chronology based on a seriation, but they are hardly a reason for discrediting the method altogether. The sequences of Figs. 9.9 and 9.11 do not differ very much, and if one were presented with the jumble of Fig. 9.11 (top) it would be a major step toward the true chronology to

![Fig. 9.10. Predicted occurrence midpoints, derived from MDSCAL results for the data of Fig. 9.9, plotted against true mean points derived directly from Fig. 9.9.](image)

obtain the sequence of Fig. 9.11. Furthermore, if one were presented with Fig. 9.9 and not told that this in fact was the true sequence, there would be nothing in it to lead one to think it a more plausible or probable sequence than Fig. 9.11. In a situation of ignorance, it would be better to bet on Fig. 9.11.

We are simply confronted with the fact that the betting odds favour the most probable sequence (assuming our probability model is a valid representation of culture change); it would not make sense to arbitrarily favour some less probable sequence; and yet we can be sure that in the real world it is not always the most probable event that occurs; so our most probable sequence is very apt to be untrue in some details.

Finally, how does the technique fare when confronted with some real data? I have tested it by applying it to an occurrence matrix representing fifteen
Fig. 9.11. Randomized rearrangement of the sequence of units (rows) in the occurrence matrix of Fig. 9.9, and result obtained from reordering the input sequence according to the predicted midpoints derived from MDSCAL, using Equation 6 and sorting units in order of predicted occurrence points.
pottery types in some 554 graves from the Hualfin valley of north-west Argentina. I am very much indebted to Dr Alberto Rex Gonzalez, who was responsible for gathering the data, doing the ceramic analyses and tabulations of materials in each grave, and keypunching of the data. I should add that since I have no expert knowledge of this material, and have not yet discussed

the results in detail with Dr Gonzalez, my present results are only a test of the technique, and not a contribution to Andean prehistory (Dr Gonzalez and I plan to publish such a contribution elsewhere).

The **MDSCAL** configuration for the fifteen Hualfin types is shown in Fig. 9.12. It is again quite linear, although the stress is rather high (about 0.30).
Models in archaeology

The line fitted to these points as a basis for scaling predicted midpoints is shown in Fig. 9.13. In the 554 graves 109 different type combinations occurred, resulting in 109 potentially seriatable intervals. The seriation derived from Equation 6 looks very good, but it includes a number of graves in which only one of the fifteen types occurs. For example, there are ninety-eight graves in which only type $F$ is found. To suppose that these ninety-eight graves represent ninety-eight consecutive burials, only before which and after which people were willing to include other types along with $F$, is so unrealistic that I prefer not to publish the results. I have also decided, somewhat arbitrarily, to exclude from this illustration graves in which only two types occurred, since it also seems unrealistic to think that all these were consecutive. Instead, I have shown results from the 129 graves with three or more types present, representing sixty different type combinations. I should emphasize that the reason for showing results for 129 units rather than 554 is not that the latter were hard to handle or that the results were not good in a formal sense, but is solely because it is unrealistic to suppose that any seriation techniques can specify very precisely the dates of graves with only one or two types present. Although the minimum advisable number of types deserves further study, it seems better to seriate only units with perhaps about three or more types, than satisfy oneself by assigning poorer units to the broad intervals in which their types occur in the seriation of the richer units.

For the 129 graves with three or more types present, Fig. 9.14 shows the randomized input sequence, and Fig. 9.15 shows the result of seriation by Equation 6. In general it looks good, though not so tidy, as the results with invented data. Many of the bad-looking cases seem unavoidable. Types $H$ and $K$, for example, both show single isolated early occurrences, but these are both in graves where the early type $C$ makes its latest appearances, and to place these graves later would balance the improvement for types $H$ and $K$ by a worsening for type $C$. Still, a compromise that made $C$ somewhat worse and $H$ and $K$ somewhat better might be more attractive. The essential point is that, however we deal with it, we have a pair of graves here in which anomalous combinations of types occur. Perhaps some of the types have been misidentified or should be redefined, possibly a keypunch error is involved, or perhaps it is simply true that events which we think improbable nevertheless occurred.

At any rate, taking the input data as given, there is an anomaly here which no seriation technique could resolve very well, and which probably accounts in part for the fairly high stress of the MDS solution (this high stress of course also suggests that, aside from using these data as a somewhat severe test
Fig. 9.14. Occurrence matrix model for the Hualfin data (129 graves with three or more of the fifteen types present), in random order of units (rows).

Fig. 9.15. Results obtained from reordering the sequence of Fig. 9.14, according to the predicted midpoints derived from Fig. 9.13, using Equation 6 and sorting units in order of predicted occurrence points.
Models in archaeology

of a seriation technique, we should examine the data for systematic variation relative to other variables besides time.

The same trade-off between improving the sequence for one type, at the expense of worsening it for another, applies to most other examples of spotty occurrences in Fig. 9.15. There are some exceptions; the latest occurrence of type G could be brought much closer to the main bunch without appreciably loosening any other sequences, although the bad appearance of type G is caused mainly by several blocks of graves with identical combinations of three or four types, which introduce ties. If Fig. 9.15 showed only the sixty intervals defined by the sixty type combinations that occur, the gaps would be much smaller. But I do not claim that this technique always finds the arrangement with the fewest zeros (or blanks) interspersed among the 1's. It does provide an extremely quick and easy way for deriving a sequence which is at least close to an optimal seriation. One is then in a good position to see if, by inspection or other means, some further improvement can be obtained. Readers may like to see how much they can improve on the sequence of Fig. 9.15.

In the course of this discussion of so many aspects of seriation, on the level of both method and technique, it may not seem as if the technique I have described is actually easy. I should stress that it is, first because it is built around multidimensional scaling, which is a laborious operation made easy because there exist good programs for doing it (such as MDS CAL), and second because computation of the matrix of similarity coefficients and seriation by Equation 6 are very easy operations, for which simple computer programs can be written. To repeat, one begins by computing similarity coefficients (from Equations 4a, 4b and 4c); uses these as input to MDS CAL; judges whether the resulting two-dimensional configuration looks good enough to justify proceeding with a seriation; and, if so, scales the predicted midpoints of type ranges from the MDS CAL result and uses Equation 6 to get predicted occurrence points for all the units. The result can be inspected for possible further improvements in the sequence, or perhaps for suggestions that some types be excluded or redefined, in which case the whole process might be repeated with a different occurrence matrix for the same units.

Notes

1 My work on the Hualf'in data was supported in part by an allocation from Brandeis University of funds from National Science Foundation Institutional Grant GU2839, and in part by a grant from the Milton Fund of Harvard University. Gordon R. Willey provided valuable encouragement and assistance.
References


KRUSKAL, J. B. and CARMONE, F. (1969) How to use M-D-SCAL (Version 5M) and Other Useful Information. MS., Bell Telephone Laboratories, Murray Hill, N.J.
Models in archaeology


1 A shorter version of some sections of this paper, entitled "Problems and Some Solutions in Seriating Large Data Sets", was presented and made available for distribution at the Annual Meeting of the Society for American Archaeology, Norman, Oklahoma, May 1971.

2 Gelfand (1971) also describes an "averaging" technique which has worked well with non-ideal data sets. Taking each unit in turn as a reference unit, all other units are put into a sequence such that similarity to the reference unit decreases monotonically. A final estimated sequence is obtained by giving each unit a rank that is, in effect, the arithmetic mean of its ranks in all the individual sequences. This technique has the advantage of taking into account information about similarities between all pairs of units (not merely those most similar), while at the same time being independent of any particular starting sequence, and avoiding the iterative or trial and error procedures that have been so time-consuming in most seriation techniques. It is simple enough that it should be feasible for a computer even for sizable numbers of units. Although, as Gelfand points out, this technique does not provide the same kind of information that multidimensional scaling does, it does seem to be one of the very best techniques now available for sets of units where the suitability of any one-dimensional sequence has been established.

3 From Figure 4 one can see that, if one somehow already knew the "true" sequence, type B would have been used, since it is the type which truly occurs at one extreme of the sequence. But, from the information in the S matrix and in Figure 6 alone, there is no way to know this.

4 If \( m_i \) is the predicted midpoint for type \( i \), \( g_i \) the number of units in which type \( i \) occurs, and \( k \) is the coefficient of proportionality relating the standard deviation of the probability function to the number of units in which any type occurs, then the probability density function for type \( i \), \( P_i(x) \), is given by

\[
P_i(x) = \frac{1}{k g_i \sqrt{2\pi}} e^{-\frac{(x-m_i)^2}{2k^2 g_i}}
\]

where \( x \) is the distance along the scale derived from the long axis of the MDSCAL configuration, as in Figure 7. If in unit \( U \), the number of different types that occur is \( u \),
our assumption is that \( F_U(x) \), the probability density function for unit U, is just the product of the \( P_i(x) \) for all the types that occur in unit U. This means that

\[
F_U(x) = \frac{1}{k^u(2\pi)^{u/2} \prod g_i} e^{-(1/2k^2)\sum (x-m_i)^2/g_i^2}
\]

where the product and sum are over all types present in unit U. Since \( F_U(x) \) is a continuous function of \( x \), with a continuous first derivative, we can find the points which might be maxima of \( F_U \) by setting the derivative equal to zero.

\[
\frac{dF_U(x)}{dx} = -\frac{2(x \sum 1/g_i^2 - \sum m_i/g_i)}{k^{u+2}(2\pi)^{u/2} \prod g_i} e^{-(1/2k^2)\sum (x-m_i)^2/g_i^2}
\]

\[= 0\]

Since \( e^{-z} \) is zero only for \( z \) infinite, which would correspond to zero (minimal) probability that unit U is either infinitely early or infinitely late, and since the denominator consists entirely of terms that are greater than zero, \( F_U(x) \) can be a maximum only if

\[
x \sum 1/g_i^2 - \sum m_i/g_i^2 = 0
\]

that is, if

\[
x = \frac{\sum m_i/g_i^2}{\sum 1/g_i^2}.
\]