Review: [untitled]
Author(s): George L. Cowgill
Reviewed work(s):
   Computer Analysis of Chronological Seriation by Frank Hole ; Mary Shaw
Published by: Society for American Archaeology
Stable URL: http://www.jstor.org/stable/278612
Accessed: 09/11/2010 11:45

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=sam.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

This is one of the most interesting and important accounts of experiments with computers and archaeological data yet to have been published. It is relatively free of defects which mar many other publications in the field: needlessly esoteric jargon, inflated pretensions, or shallow grasp of the statistical manipulations set loose on one’s data by the great machine. Hole and Shaw are generally in control of their data, techniques, and theory; and their style shows an admirable interplay of technical sophistication and calm good sense. It turns out that there are some surprising implications of their data they have missed, but this does not lessen the importance of the work.

Two major topics are presented. There is a discussion and complete listing of PHOENIX-II, a program system written in ALGOL-20 designed to accept data in a variety of forms, compute matrices of various kinds of agreement coefficients, seriate the matrices by a variety of methods, print the resulting best sequence and measures of its goodness (as well as other details of the computations, if desired), and produce contour displays of the matrices and frequency plots of the seriated categories, all according to a number of easily specified options. What the program will do is explained at some length and should be comprehensible to anyone who has ever seriously worked at the Brainerd-Robinson seriation method.

The other important topic is an extensive discussion and review of seriation methods and theory, including the Brainerd-Robinson method, Meighan’s 3-pole graph technique, the Dempsey-Baumhoff approach using presence-absence matrices and “contextual analysis,” and the Aschers’ computer method for Brainerd-Robinson seriation (Ford’s manual technique is dismissed as “time-consuming, tedious, and frequently inaccurate” (p. 8)). These methods are applied separately to ceramics, flints, and other artifacts from known sequences in several Iranian sites occupied between 3000 and 3500 B.C., and the results are compared and discussed. This application of several techniques to the same set of real data, and of the same technique to several sets of data, gives results which are far more convincing, and far more enlightening, than many times the worldage devoted to polemics for or against any single technique. Hole and Shaw point out that such large-scale testing would have been impossible without computers. However, everyone doing seriations or at all dependent on results of seriations, whether by computer or not, should study Hole and Shaw’s work with great care.

One defect is inadequate discussion, and sometimes plain misrepresentation, of another recent excellent paper on computer seriation (Kuzara, Mead, and Dixon, “Seriation of Anthropological Data: a Computer Program for Matrix-Ordering,” American Anthropologist, Vol. 68, pp. 1442–55. For brevity, I will hereafter refer to this as KMD, to Hole and Shaw as HS, to Brainerd-Robinson as BR, and to Dempsey-Baumhoff as DB). Important here are two different views about the central task of seriation, which are (1) to deduce as nearly as possible the correct chronological sequence of a set of archaeological units in cases where stratigraphy, radiocarbon, inscriptions, or other dating methods are not sufficient, or (2) to arrange a set of units in that sequence which comes closest to being such that (in terms of some sensible and clearly defined measure of similarity) similarity of a unit to another unit invariably decreases as we compare it with units farther away in the sequence. For task 2 the most similar units are (ideally) always closest together and the most different ones farthest apart, and for a BR matrix this ideal means that similarities invariably decrease as one moves away from the principal diagonal.

The connection between task 1 and task 2 is, of course, that if we are satisfied that each unit consistently shows less and less similarity to other units as we compare it with units more distant in time, then achievement of task 2 will mean that task 1 is also achieved.

HS explicitly take task 1 as their main objective, with matrix ordering simply one important method whose value they wish to explore. This leads them to believe that (except for testing or comparing methods) knowledge about correct chronology should be used to constrain the arrangements of units that will be considered (p. 24). Also, while the possibility that results can reflect different activities rather than chronology is recognized, on the whole factors other than time are regarded as undesirable “noise,” and the emphasis is on seriating assemblages which are from reasonably homogeneous cultures. KMD, on the other hand, regard task 2 as their main objective, with task 1 only one important kind of interpretation which task 2 may help one to make. They explicitly recognize that matrix-ordering can have other objectives (such as expression of intercultural relationships), and their position is that if the resulting sequence disagrees with known chronology, this itself is an important result.

Neither view is right or wrong. Depending on circumstances, either task 1 or task 2 can be most important. In the former case it will make sense to automatically exclude arrangements known to be chronologically wrong, but if one’s interests are more general these arrangements should not be excluded.

A related matter is the choice of a “goodness of fit” criterion or “norm” for evaluating different sequences when no perfect sequence is possible. For a matrix of agreement coefficients, and with respect to task 2, three obvious ways are to simply count the coefficients which do not fit the ideal model, to take the total of the absolute values of all differences which go the wrong way, or to take the total of the squares of such differences. With all these methods, the best “norm” is the smallest. There is a consensus that the first method is too simple and counts minor deviations from the ideal equally with major ones. HS at one point (pp. 28–9) argue for the second choice and criticize KMD for using the third choice. Except that it requires less computing, HS’ reasons for preferring the
second choice are not very convincing to me, and elsewhere (p. 82) they "take the position that good data should give identical or almost identical orderings under all reasonable norms and that variability of results produced by switching norms probably indicates poor data." Finally HS say (p. 95) that KMD "say they used two norms, but report the results of just one." In fact, KMD use essentially the same "sum of absolute values of deviations" criterion used by HS, but do explicitly report some results of using the "sum of squares" choice, and state that in other cases both choices gave the same results (KMD, p. 1450). More precisely, KMD mainly use 2(norm)/Nδ, where "norm" is the coefficient used by HS and N is the number of units. HS (p. 30) note one might divide by Np, without mentioning that KMD in fact do so, but HS feel that this only partially compensates for differences between different matrices. I feel that although it does not solve all problems, it makes much more sense than HS allow, although (N-1)(N-2) rather than Np is the maximum number of wrong-way differences in the worst possible matrix, and seems a better denominator.

HS also discuss the problem of a goodness of fit measure with respect to what I call task I, which is of course possible to compute only when one happens to know what the true chronological sequence is. Of several possibilities considered, they make a good case for preferring an "error coefficient" defined as the sum of all units placed between units which in fact should have been adjacent (p. 35).

An important point made by both KMD and HS is the astonishing rate at which the number of possible orders increases as the number of units being seriated increases. For six units, there are 720 possible orders, for 12 units there are more than a trillion, and for 20 units more than two million million million. It would take years or millennia for any foreseeable computer to try every order for a dozen or more units. Most seriation methods depend on examining a limited number of possibilities, finding successive improvements, and stopping with a sequence which cannot be improved in any obvious way. Unless the data are relatively well behaved, one usually cannot be positive one has found the best possible sequence. PHOENIX-II features a system called Permutation Search which involves successive applications of "pairwise interchange" (interchanging each different pair of units) and "successive rotation" (successively moving each unit through the otherwise fixed sequence in search of the place it fits best; this is the only technique used by KMD) until no further cycles lead to reduction of the "norm." Unfortunately this method depends on the starting order of the units and can fail to reach the best sequence if the data are not well behaved. The Ford hand method cannot avoid this problem either, though it can make it easier to avoid facing up to it. HS stress the need for more work on search techniques.

If there is one thing above all others that archaeologists should learn from HS' results, it is that relying on simple presence-vs-absence of categories does not reduce statistical problems; on the contrary it increases them, and it often wastes important and perfectly good information. This is obvious on statistical theoretical grounds but it is much more convincing for the nonstatistician to see how consistently HS get poorer results with their real data; in 36 comparisons, use of percentages leads to a lower error coefficient in 21 cases, presence-absence wins in 8 cases, and there are 7 ties. One trouble with presence-absence, as pointed out by HS and by Lipe (American Antiquity, Vol. 30, pp. 103-4) is that even slight mixing between units will have serious results. I would add that the chances of some sparsely present category being absent in a collection depend very much on the size of the collection. Often the distinction between "absent or sparse" (in the collection) and "abundant" is more meaningful than sheer "present" or "absent," and it is also much less sensitive to total sample size. Presence-absence makes sense only when we feel sure that absence in the collection means that the trait was really absent (not just rare) in whatever entity the collection has sampled.

For task 2, HS get their best results (as measured by their "norm") from straight permutation search in 11 out of 16 cases. The DB "contextual analysis" method is better once, ties in 2 cases, and is worse in 13 cases (in 8 of which it gives the worst norm of all methods compared). HS are rightly very critical of contextual analysis, and rightly say that DB (American Antiquity, Vol. 28, pp. 496-509) are wrong in claiming that "pattern III" frequency distributions create a problem for BR matrix methods. By the "norm" criterion, the Meighan 3-pole graph technique and the Ascher computer program are also much poorer than permutation search. HS claim that either contextual analysis or the 3-pole method are useful for pre-ordering a sequence of units whose order can then be improved by permutation search. Permutation search applied to a sequence pre-ordered by contextual analysis or 3-pole does lead about as often as not in their examples to a better norm than permutation search applied (apparently) directly to the known correct chronological sequence of units.

The real bombshell in this study comes when one looks more closely than HS themselves seem to have at what happens with respect to task I, actually achieving the correct chronology, as measured by the "error coefficient." Here it turns out that, for the nine cases where it was applied, the simple 3-pole graph method (though never perfect) leads to better results than straight permutation search in four cases, is tied twice, and is worse three times. Contextual analysis is better than either permutation search or 3-pole in one of six cases, is tied with permutation search in one case, and is worse than either in four cases. Permutation search applied to a matrix pre-ordered by contextual analysis or 3-pole tends to make the error coefficient worse (while improving the norm). The Ascher program again gives very bad results. It was a pioneering approach to computer seriation which is now outmoded.

Thus there are two jolting implications, as I read HS' results. Firstly, when all one wants to do is just get the
right chronological sequence, there is nothing that yet demonstrates that any other method studied by HS works any better than the very easy 3-pole graph technique of Meighan (American Antiquity, Vol. 25, pp. 203–11). Secondly, the reason that permutation search (which tends to give the best results for task 2) doesn’t work any better than 3-pole for chronology, is that the best ordering of units according to their similarity (task 2) is not generally the best chronological ordering (task 1). With hindsight, it is glaringly clear that in 15 of 16 cases, the known correct order (which automatically has an error coefficient of zero) has a much worse norm than many alternative sequences have.

HS’ results should do much to dampen naive and sometimes dogmatic overconfidence in seriation as a way of deriving chronology, but it would be wrong to conclude that it really discredits the approach. Instead, we should look more closely at why the best matrix orderings are not the best chronologies. Broadly there are three kinds of reasons: sampling errors, poor definition of categories, and sources of significant variation other than time. HS recognize these problems, and they are well aware that their ceramics seriate fairly well and the flints rather badly, with other artifacts somewhere in between. They feel, rightly I think, that much of their trouble with ceramics comes from some poorly chosen categories, especially plain wares which wax, wane, and wax again in popularity as preference for decoration first increases, then decreases. They interpret the bad behavior of flints as evidence that frequency changes in these categories reflect changes in activities more than style trends, although they do note that a “problem worth mentioning is that there are relatively few flints in the Tepe Sabz components . . .” (p. 61), and they did try to pick types “that had neither unusual abundance or scarcity” (p. 59). Their interpretation of the flint data looks plausible and important, but the raw data (which HS conscientiously provide) show that they have grossly underrated the effects of sample size. Even for their selected types, only the seven earliest units have samples of well over 100 objects, while the others include one sample of 73, eight ranging from 9 to 28 in size, and one consisting of only 2 flints! The earlier seven units seriate fairly well; most of the trouble is in the later ten. With such small samples, it is surprising that they seriated as well as they did. In this case, HS’ sophistication and good sense failed badly, and their inadequate handling of sample size should be a lesson for everyone.

A few other lesser criticisms can be made. I do not agree with HS’ statement (p. 85) that the BR model tacitly assumes a closed cultural system. It assumes only that (a) each category will show only a single important peak in popularity, (b) components earlier than the peak (if any) and components later than the peak (if any) will show a reasonably consistent tendency for declining popularity of the category with increasing time distance from the peak, and (c) change will never be abrupt enough to prevent some persistence of categories from one component to the next. This is not inconsistent with substantial outside influences, although it is inconsistent with factors which make the culture inhomogeneous, whether external or internal in origin.

A persistent minor irritation is the use of “correlation” as synonymous with “agreement.” It would be far better to restrict correlation to Pearsonian r and closely related statistics. Robinson’s agreement coefficient is quite different.

On page 12 HS say that in presence-absence matrices the maximum possible agreement coefficient is the number of components minus one; actually it is the number of categories minus one. On page 30, line 7, the words “the sums of” should be deleted. On the same page, Table 3-1 was very puzzling until I realized that the frequencies given are percentages of total values within specified ranges; fuller labelling would have helped. The computer contour plots were hard to read until I colored in high and low value regions. A choice of symbols making high value areas much darker than low regions would be better.

Finally, I want to suggest that HS and KMD have perhaps carried the BR and related techniques about as far as is profitable. These are inherently limited to finding the best arrangement of units along some single axis. Where there is any reason to think there may be more than one source of significant variation, multidimensional scaling or nonmetric factor analysis would tell much more. And, even when there is only a single major factor (such as time), programs for multidimensional scaling are probably more efficient and less troubled by (though not free from) the problem of stopping with a less-than-optimum sequence which cannot be improved on except by starting over with a new sequence.

GEORGE L. COWGILL
Brandeis University
Waltham, Massachusetts


Mr. Deuel has assembled 42 selected writings, articles, or essays, on the archaeology of pre-Columbian America. These are grouped into regional or topical sections (11 on Andean South America; 2 on Lower Central America; 15 on Mesoamerica; 6 on the North American “Mound Builders” and the Southwest; 6 dealing with Early Man; and 3 on the Arctic). In each section the writings are arranged more or less historically, following dates of original publication. The period covered is the last 150 years or so, and many of the selections are well-known “classics.” Von Humboldt, Squier, John Lloyd Stephens, and Thomas Jefferson lead off, in their respective places, to be followed by F. W. Putnam, S. G. Morley, and so on, down to our contemporary colleagues. Each selection is preceded by an introductory statement by Deuel which tells us something of the particular author and his times.