

# Accepted Manuscript

The goodness of fit and statistical significance of seriation solutions

Marko Porčić

PII: S0305-4403(13)00263-X

DOI: [10.1016/j.jas.2013.07.013](https://doi.org/10.1016/j.jas.2013.07.013)

Reference: YJASC 3762

To appear in: *Journal of Archaeological Science*

Received Date: 15 May 2013

Revised Date: 11 July 2013

Accepted Date: 12 July 2013

Please cite this article as: Porčić, M., The goodness of fit and statistical significance of seriation solutions, *Journal of Archaeological Science* (2013), doi: 10.1016/j.jas.2013.07.013.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



**Title:**

The goodness of fit and statistical significance of seriation solutions

**Author:**

Marko Porčić

Department of Archaeology

Faculty of Philosophy

University of Belgrade

email: [mporic@f.bg.ac.rs](mailto:mporic@f.bg.ac.rs)

Čika Ljubina 18-20

11000 Belgrade

Serbia

tel. +381113206238

**Abstract**

This paper presents a method for determining the statistical significance and goodness of fit of seriation solutions based on correspondence analysis. The basic assumption is that seriation solutions corresponding to relative chronological sequences should have unimodal distributions of types across assemblages in a sequence. In order to evaluate significance of chronological signal in data, the observed number of modes of a seriation solution based on correspondence analysis is compared to a distribution of number of modes generated by randomization of the original data table. A quantitative measure of the goodness of fit is presented. The method is tested on two data sets: 1) archaeological data with a known chronological ordering 2) non-archaeological data without a chronological patterning. The method successfully detected a significant chronological signal in the former, and failed to detect it in the latter.

**Keywords:** seriation; relative chronology; correspondence analysis; archaeological dating.

## 1. Introduction

### 1.1. Seriation in archaeology – theoretical background

Seriation is a method of relative dating. Its main purpose is to establish the relative chronology of archaeological units such as artifacts or assemblages. O'Brien and Lyman (1999) recognize three kinds of seriation: 1) phyletic seriation 2) occurrence seriation 3) frequency seriation. The most familiar kind of seriation is frequency seriation, invented by Alfred Kroeber (Kroeber, 1916a; 1916b; see also Lyman, et al., 1997; O'Brien and Lyman, 1999), where the aim is to order the units in such fashion that the relative frequencies of artifact types within units follow a unimodal curve across units, so that each type monotonically increases and then decreases in frequency after reaching a peak of popularity. (Dunnell, 1970). If this can be done, the resulting sequence is interpreted as a relative chronological sequence (Dunnell, 1970).

Visually, a good seriation solution should result in the “battle-ship” pattern of attribute/type frequencies. The particular sequence of units that produces this kind of result is interpreted as a relative chronological sequence. The underlying logic of the method is the *popularity principle* – each attribute/artefact type of a certain class of material culture (e.g. bowl shape type) first appears in small quantities, then rises in popularity, reaches a peak, declines and disappears afterwards.

On a more abstract level, seriation is a method based on similarities between archaeological units. Units can be artifacts or assemblages at any level. The aim of the procedure is to make such a sequence that the most similar units are adjacent to each other. The additional condition for a seriation solution to be considered successful is the condition of unimodality which states that the frequency of any attribute or type can have only one mode or local maximum (Dunnell, 1970).

For many decades seriation was successfully used in archaeology without a theoretical explanation of why it worked so well. The popularity principle only restated the empirical

fact in different terms. Only recently, evolutionary archaeology provided the theoretical explanation of why seriation worked (Lipo, et al., 1997; Neiman, 1995; O'Brien and Lyman, 2000). It was shown that the familiar seriation “battle-ship” pattern is a predictable consequence of the random cultural transmission or neutral evolution (Lipo, et al., 1997; Neiman, 1995).

### *1.2. Correspondence analysis and seriation*

Seriation is a general method, but there are many techniques that can be utilized to produce the seriation ordering (Dunnell, 1970; Marquardt, 1978). In the last two decades, correspondence analysis has become one of the most popular seriation techniques (Baxter, 1994; 2003; Shennan, 2004). The correspondence analysis (CA) is a multivariate statistical technique that reduces the dimensionality of contingency tables. It is analogous to Principal Component Analysis for data in the form of counts. This method allows the simultaneous representation of both cases (assemblages) and variables (types) in the two-dimensional space. Usually the results of CA are presented graphically as a scatter-plot with  $x$  and  $y$  axes being the first two CA dimensions which capture most of the variability (inertia in CA terminology) in the original contingency table. An additional criterion for a chronological interpretation of CA results is the presence of an arch effect (also known as a horseshoe or Guttman effect) – the tendency of data points (both assemblages and types) to form an arch on a two dimensional CA plot. If data points in some concrete CA analysis form this shape, the order of assemblages on the first dimension is interpreted as a relative chronological sequence, because arch effect appears as a consequence of the unimodality of type frequencies (Baxter, 1994; 2003; Baxter and Cool, 2010). The mathematical reason for the arch effect is the quadratic relationship between the first two axes.

CA is also used in ecology to identify gradients in the environment – e.g. spatial or environmental directions along which frequencies of individual species conform to unimodal curves (Ter Braak, 1985). The goodness of fit of the unimodal response model in ecology is then assessed by correlating the scores on the first CA axis with some

environmental variable (gradient). Ecologists are usually interested in removing the arch effect and expanding the distances between “compressed” data points located at the ends of the first CA axis, although these issues are widely debated (Hill and Gauch, 1980; Jackson and Somers, 1991; Peet, et al., 1988; Wartenberg, et al., 1987). The removal of the arch effect is called detrending (for applications of detrended CA in archaeology see Baxter and Cool, 2010; Lockyear, 2000). In general, the arch effect and the compression of points is not a problem if CA is used for relative dating, because archaeologists are usually interested only in the sequence of points (Baxter, 1994:120; Baxter and Cool, 2010).

### *1.3. Robustness and validity of seriation solutions*

The assessment of the fit of the data to the arch shape in archaeology is usually done visually without any formal criterion. In the CA literature a distinction is made between several aspects of CA solution properties (Greenacre, 2007:193): 1) stability of the solution 2) sampling variability 3) testing specific statistical hypotheses. In two recent studies the bootstrap method was used to assess the stability and sampling variability of the CA solution (Lockyear, in press; Peeples and Schachner, 2012; Ringrose, 1992), but this is not the same as the explicit test of significance of chronological signal. These studies were aimed primarily at evaluating the robustness of the resulting empirical solution in terms of the general structure of relationships between data points. They did not test for the stability and significance of any specific patterning such as arch effect or unimodality.

How can we know that the ordering we got by applying a seriation method was actually a chronological ordering, and if it is, how good is it? This would correspond to Greenacre's third aspect – testing of specific statistical hypotheses. Usually, hypotheses being tested in CA related research are the hypotheses about the independence/association of categorical variables making up the contingency table. Greenacre suggests that Monte Carlo and permutation test can be used as methods for testing specific hypotheses about the association of variables in CA (Greenacre, 2007:198-200). When CA is used as a

technique for seriation the specific hypothesis we wish to test is that the CA solution reflects relative chronology of seriated units. This may sound contradictory, because seriation should produce chronological patterning by definition. The contradiction is more apparent than real. Different seriation techniques will always produce output (seriation result) when applied to data. The key question is: is this output relevant as a chronological ordering?

Lipo and colleagues demonstrated that only those assemblages that were produced by communities that participated in the same cultural transmission network can be successfully seriated and used this implication to reconstruct cultural interaction zones in the Mississippi region (Lipo, et al., 1997; Lipo, 2001). One of the methodological problems these authors faced was how to discriminate between good and poor seriation solutions. The theoretical premise of unimodality served as a criterion – only seriation solutions with assemblages that could be ordered in a fashion that would not violate the unimodality principle were considered as valid solutions. Lipo and colleagues had a very specific research problem, but the issue they tackled in their paper is an example of a more general problem that exists in the archaeology – the problem of how to evaluate the goodness of fit of seriation results. They invented a technique named the deterministic seriation (Lipo, et al., 1997) or the iterative seriation method (Lipo, 2001:43-46) to manually compare frequencies of types between successive assemblages using confidence intervals. This is the only method that I am aware of for evaluating whether a set of assemblages conforms to a seriation model.

Given the theoretical and chronological significance of unimodality, it may be argued that only those CA seriation solutions that produce statistically robust unimodal patterns are good candidates for the construction of relative chronology (Dunnell, 1970). The method for evaluating statistical significance of unimodality in seriation solutions is presented in this paper. In archaeological practice, a unimodality of frequency distributions in seriation is taken as an indicator of chronological pattern. It should be emphasized that this is an inferential step that has to be justified independently. In general, unimodal distributions will arise whenever there is a gradient in data,

chronological or otherwise, so the equivalence of unimodality with chronological signal is inferential - it is justified only in cases when an analyst can be reasonably certain that other factors which may result in unimodal distribution of types within the assemblages can be excluded. This topic will be discussed in greater detail in Section 4.

The method and the specific technique is presented and then illustrated by applying it to the assemblage with a known chronological order and to the assemblage without an underlying chronological pattern. In the appendix to the paper the code written in R is provided so that the method can be readily applied and modified by the interested reader.

## **2. Method and data**

### *2.1. Evaluating statistical significance of seriation solutions*

In an ideal seriation model the total number of modes for an ordered data table should equal the total number of types. The existence of several modes per type and by implication the total number of modes that exceeds the number of types would violate the seriation assumption. Indeed, only three kinds of type frequency behavior are allowed in the ideal seriation model: 1) monotonic increase followed by monotonic decrease 2) monotonic increase 3) monotonic decrease. The departure from unimodality can be due to two reasons: 1) the effect of various biases that distort the underlying pattern 2) the true lack of unimodal behavior of frequencies. These two scenarios are what we wish to discriminate between. In reality we should expect to find departures from the ideal model due to sampling effects even if there is a real underlying chronological pattern in the data. Moreover, the number of possible modes increases as the number of assemblages (rows) increases (see 2.2.). The point is that it is very unlikely to find the perfect unimodality in real data – we can almost always expect to find more modes than types (or more than one mode per type to be precise). Therefore, the aim is to develop such method that would enable us to discriminate between the signal of unimodality (and by inference chronology) and random fluctuations.



The method developed in this paper presents one possible way of testing for the statistical significance of unimodal (chronological) signal in seriation solution. The procedure can be divided into 7 steps (Figure 1):

**Step 1.** The first step is to perform a CA analysis on a contingency table with  $i$  rows and  $j$  columns where each cell in the table represents the count of  $j$ th type items within the  $i$ th assemblage.

**Step 2.** The assemblages in the table should be reordered according to their order on the first CA axis.

**Step 3.** The third step is to transform the absolute counts in the reordered table into relative frequencies of artifact types within assemblages. Technically, this is done by dividing each cell with the corresponding row sum of absolute frequencies. This step has to be performed because assemblages usually differ in sizes so we need to normalize the absolute counts to make them comparable along the seriation sequence. We are interested in changes in relative frequency of types between assemblages along the seriation sequence.

**Step 4.** The next step is to count the number of modes within each type (column). It should be emphasized that the mode should not be understood in its strict sense (the most frequent value in a sample), but as a peak or local maximum. The local maximum for a given type is a relative frequency value that is greater than both the preceding and the following relative frequency. An illustration is shown in Figure 2. Horizontal axis presents the seriation sequence and the associated relative frequencies of Type 1 in assemblages from the Ezero data set (see Section 2.3.). In this example, there are 4 modes (marked by vertical bars in Figure 2). When the number of modes for each type is determined by counting the local maxima for that type, the total number of modes in ordered data is obtained by summing these counts. The observed total number of modes corresponds to the test statistic. This method of calculating the total number of modes is conservative in the sense that small deviations from unimodality are given the same

weight as great deviations. In the absence of theoretical criteria for weighting modes according to their prominence and intensity, it is safer to assume that all modes are of equal importance.

**Step 5.** In this particular context we wish to determine whether the observed pattern is strong enough to be considered significant. In order to evaluate the significance of the unimodal signal we create a distribution of the total number of modes from randomized assemblages. Randomization of assemblages is performed by interchanging (permuting) the observed frequencies of individual types (within columns) between different assemblages (rows) in the original data table. In this way, we are deliberately destroying any real patterning in the data and creating something that is for all practical purposes a random combination of type frequencies across assemblages. A large number of randomized (permuted) data tables should be generated. In this paper, 1000 randomized data sets are produced for each analysis.

**Step 6.** On each randomized data table, we perform CA, reorder the assemblages in the table in accord with the scores on the first CA axis, and then calculate the total number of modes using either the conservative or the weighted mode variant of Step 4. The total number of modes is recorded for each permutation of the original data. When this procedure is performed many times (1000 times in this case) the result is a distribution of 1000 randomized total number of modes values. This randomized distribution can be regarded as a null distribution – it gives us the probability of observing different values of the total number of modes when the frequencies of types are randomly distributed across the assemblages, when there is no real structure, chronological or otherwise in the frequency variability. The logic of this analysis is taken from the methodology used to test the strength of the phylogenetic signal in cladistics (Kitching, et al., 1998:122-124). This is no coincidence because there are important similarities in the theory behind the seriation method and the cladistic reconstruction of phylogenetic trees (O'Brien, et al., 2001; O'Brien and Lyman, 2003).

**Step 7.** The final step is to compare the test statistic (observed total number of modes) to this null distribution. If the observed value and some even more extreme value have a very low probability of occurring (e.g. below 0.05) in the distribution created by random data, then we have a relatively strong evidence against the null hypothesis of the non-unimodal structure in the data. Given that the unimodal pattern is usually an indicator of chronological differences, we can then argue that the observed chronological pattern is statistically significant.

## 2.2. Evaluating goodness of fit: the *S* coefficient

In addition to statistical significance, it would be useful to have some measure of goodness of fit that would indicate how close a particular seriation solution corresponds to the ideal seriation model. Such measure, named the seriation coefficient (*S*), is calculated in the following manner:

$$S = (\text{Max} - O) / (\text{Max} - E)$$

where *O* is the observed total number of modes, *E* is the expected total number of modes if all types had unimodal distributions, and *Max* is the maximum total number of modes which depends on the number of assemblages (rows) in the data matrix. For an even number of assemblages, the maximum total number of modes is equal to the number of types multiplied by the number of assemblages divided by 2. For an odd number of assemblages, the maximum total number of modes is equal to the product of the number of types on one side, and the number of assemblages plus 1 divided by 2 on the other. The number of assemblages has to be taken into account because data matrices with a low number of rows can have fewer departures from unimodality than matrices with many rows. Seriation coefficient can take values between 0 and 1 – values close to 1 indicate a strong fit to the seriation model, while values close to 0 indicate a poor fit. It should be noted that the formula for the seriation coefficient has some interesting structural similarities to the *retention index* formula from cladistics, which measures the

degree to which data fit the tree-like structure of the cladogram (Kitching, et al., 1998:97-99).

### 2.3. Data

In order to present and test this method in practice, it is applied first on an archaeological data set where the true chronological ordering of assemblages is known because assemblages come from stratified building horizons. This data set comes from the Early Bronze Age tell site of Ezero, Bulgaria (Георгиев, et al., 1979:Table 200). Table 1 shows the frequencies of relief decoration motifs within individual building horizons in Ezero. It was estimated that the duration of individual building horizons is roughly the same (Weninger, 1995). In this case the method is expected to detect a significant and strong chronological signal in the data.

The method is also applied on a contingency table data set without chronological signal or any unimodal signal whatsoever. This data was randomly generated. Each cell value in the data table (Table 2) is a random number taken from a uniform distribution of integers ranging between 0 and 50. In this way we can be certain that these are completely random data where no gradient, chronological or otherwise, is present. Therefore, the method should fail to detect the significant unimodal (chronological) signal when applied to this data that will serve as control for the method.

## 3. Results

### 3.1. Ezero data results

The results of CA analysis of the Ezero data are presented in Figure 4. CA summary information is given in Tables 3-4. Axis 1 and Axis 2 account for 47.07% and 13.97% of total inertia, respectively. The seriation diagram based on the sequence of assemblages along the first CA axis is presented in Figure 4. Apparently, there is a pronounced arch effect in the CA biplot and visual inspection of the seriation diagram shows that type

frequencies seem to follow a unimodal pattern in general. The seriation of assemblages by CA gives the correct chronological order for most of the assemblages. The positions of building horizons V and VIII are wrong, but the sequence in general is correct.

The total observed number of modes based on the CA seriation is 63. The histogram of the distribution of total number of modes for the conservative approach from 1000 permutations of the original data is shown in Figure 5. The observed value is smaller than the 5th percentile of the distribution of randomized samples which is 71. The value of the seriation coefficient is 0.63. Therefore, we can conclude that there is a relatively strong and significant signal of unimodality, and by inference chronological signal, in the data.

### *3.2. Control (random) data results*

The results of the CA analysis of the atemporal data set with the corresponding seriation diagram are shown in Figures 6 and 7, respectively. Axis 1 and Axis 2 account for 50.28% and 24.29% of total inertia, respectively. CA summary information is given in Tables 5-6. As expected, there is no arch effect, and the frequencies do not follow a unimodal pattern.

The observed total number of modes for this data is 31. From the histogram of randomized values we can see that the probability of obtaining the total number of modes value equal or less of the empirically observed value is rather high (Figure 8). The 5th percentile of the distribution is equal to 26 modes. Therefore, we do not have enough evidence to reject the null hypothesis of non-unimodality. The value of the seriation coefficient is 0.31. Therefore, we can conclude on the basis of the results that there is no unimodal signal in this data. Again, this is the correct conclusion since we know that there cannot be any meaningful unimodal signal in this data set.

## **4. Discussion**

### *4.1. Performance of the method*

The results clearly show that the method presented in this paper successfully detects the chronological signal when present, and fails to detect it when absent. Additionally it provides the measure of goodness of fit in the form of the seriation coefficient. Therefore, it can be used as means of determining the strength and the significance of the chronological signal in seriation solutions.

It is interesting to note that the value of the seriation coefficient in the case of the Ezero data set was over 0.5, but not so close to 1. This is good because it reflects the fact that the ordering recovered by the CA is not entirely correct. On the other hand, the fact that small deviations from unimodality are given the same weight as great deviations implies that S coefficient values are probably lower than they would be if we could *a priori* eliminate insignificant modes from the total number of observed modes. Therefore, even moderate values of S coefficient may reflect a significant unimodal pattern. It should also be noted that statistical properties of the S coefficient are unknown at the moment, so differences in S among seriations should be interpreted with caution.

It should also be emphasized that if there is a chronological signal in data, but this signal is not dominant, meaning that the greater portion of variation is accounted for by some other non-temporal factor, it is likely that the first CA axis will not represent chronology but this other dominant factor. Temporal dimension may be recovered on the second or the third CA axis. In this case, the presented technique will fail to detect the chronological signal even if there is one. Therefore, visual inspection of the CA plot of original data should always precede the formal analysis, because if indication of chronological patterning can be observed on the second CA axis, then the second axis can be used for seriation.

The reason why CA was used in this paper as a seriation technique for the initial ordination of assemblages is only due to its ubiquity in the literature. It is known that CA can often produce sequences that are correct in general, but not in details – it is not unusual to find some assemblages in the sequence with reversed position in relation to

their true seriation order (Baxter, 1994:121-123). This is what happened in the Ezero data set. Therefore, other seriation techniques can be used for the initial ordering of assemblages. For example, it is possible to construct a technique that would find the best seriation solution based solely on the criterion of unimodality. An algorithm can be devised that would try out all possible sequences of assemblages and keep only the sequence(s) where the departures from the unimodality criterion are the smallest. This would be a frequency seriation analogue for a technique that already exists for the occurrence seriation (O'Brien, et al., 2001). This best sequence can then be used to determine the observed total number of modes in the data. It should be noted that this would be very computationally intensive technique for larger datasets.

#### 4.2. Statistical considerations

Returning again to parallels with cladistics, a significant question arises: can we interpret the tail probability of obtaining the observed or smaller number of modes in respect to the randomized distribution as a statistical significance? Authors from the field of cladistics consider that this kind of permutation test can only offer a relative measure of the phylogenetic signal strength in the data which is not the same as the statistical significance (Kitching, et al., 1998:125-126). It is argued that the reason for this is that the phylogenetic hierarchical structure must be imposed on the randomized data which by definition have no such structure, so the comparison of the test statistic derived from real data to the distribution generated by the randomized data with no intrinsic structure is not justified (Carpenter, 1992).

While it may be argued that the traits of all biological species do have intrinsic phylogenetic hierarchy based on the theory of evolution (Carpenter, 1992), the same argument cannot be applied to archaeological assemblages *a priori*. There does not have to be an intrinsic chronological structure in all archaeological data set. For example, we could have several archaeological assemblages where differences in their content are caused by spatial or social factors rather than chronology. Additionally, we could have assemblages that do not belong to the same cultural tradition. By permuting the contents

of assemblages we are destroying a potential chronological signal in real data, if there is one, and we are simulating a range of unknown non-temporal processes that could have produced the data set with the same amount of variation.

The argument that “tail probabilities” resulting from permutation tests are not *Type I error* probabilities is true (Carpenter, 1992), but this is a wrong argument to use in this context. Indeed, statistical significance based on permutation, as well as statistical significance ( $p$ ) values derived by other means, are not the same thing as *Type I errors* developed in the Neiman-Pearson statistical testing framework (Hubbard and Lindsay, 2008). Statistical significance is a concept developed by Sir Ronald Fisher to measure relative strength of evidence against some null hypothesis. Strictly speaking  $p$  values are not absolute long-term probabilities of error (Hubbard and Lindsay, 2008). Therefore, it makes sense to interpret tail probabilities from randomized distributions as statistical significance in Fisher’s sense.

#### 4.3. Chronological validity of seriation

Finally, if the results of some concrete application of the test are statistically significant in the sense explained above, this does not have to mean that the detected signal of unimodality is necessarily chronological. It is only chronological by inference based on an *a priori* assumption that unimodal patterns in archaeology usually appear because of chronological differences.

Seriation is usually performed on the stylistic classes or attributes of material culture, variants of which are transmitted in a neutral fashion (Cochrane, 2001; Dunnell, 1978; Lipo and Madsen, 2001). For these classes of artifacts the neutral model predicts the unimodal distribution of type frequencies through time (Neiman, 1995). Therefore, the temporal gradient will be dominant in such cases, as long as all directions in the physical space are more or less equally likely for the cultural transmission to occur. Variants can have unimodal frequencies even if they have different adaptive values (if they are culturally selected) in the so called Red Queen environments – if there is a constant



pressure to develop better (in respect to some arbitrary criterion) variants (Shennan and Wilkinson, 2001). On the other hand, if different variants of the studied class of material culture have different adaptive values which are correlated with positions in physical or social space (e.g. different variants are advantageous in different ecological settings), then a seriation of assemblages will also reflect a spatial or social dimension. The most pragmatic way to proceed is to interpret the unimodal pattern in chronological terms if other explanations can be excluded as less probable. This can be achieved in practice by testing whether scores on the first CA axis or the seriation sequence order of assemblages correlate with, for example, environmental or spatial gradients. If there is no correlation then these factors can be excluded as factors standing behind the unimodal response of type frequencies as reflected by the seriation sequence.

## **5. Conclusion**

To summarize, the presented method should enable the researchers to formally evaluate the goodness of fit of their seriation solutions and to test for the significance of chronological signal in data. It is less sophisticated than the method developed by Lipo and colleagues (Lipo, et al., 1997; Lipo, 2001), because it does not include the manual fine tuning of the sequence by testing for significant differences between each pair of assemblages, but it is more practical given the fact that it can be done automatically and it provides a general goodness of fit indicator.

## **Acknowledgements**

This research was undertaken as a part of the project No. 177008 funded by the Ministry of Education and Science of the Republic of Serbia. I am most grateful to Professor Clive Orton and to Mladen Nikolić for their useful comments. I would also like to thank the anonymous reviewers for their detailed comments, good observations and constructive suggestions that significantly improved the paper. Of course, the responsibility for the remaining errors and omissions of any kind is exclusively mine.

## References

- Baxter, M.J., 1994. *Exploratory Multivariate Analysis in Archaeology*, Edinburgh University Press, Edinburgh.
- Baxter, M.J., 2003. *Statistics in Archaeology*, Hodder Arnold, London.
- Baxter, M.J., Cool, H.E.M., 2010. Correspondence analysis in R for archaeologists: an educational account, *Archeologia e Calcolatori* 21, 211-228.
- Carpenter, J.M., 1992. Random cladistics, *Cladistics* 8, 147-153.
- Cochrane, E.E., 2001. Style, Function, and Systematic Empiricism: The Conflation of Process and Pattern, in: Hurt, T.D., Raluta, G.F.M. (Eds.), *Conceptual Issues in Evolutionary Archaeology*, Bergin & Garvey, London, pp. 183-202.
- Dunnell, R.C., 1970. Seriation Method and Its Evaluation, *American Antiquity* 35, 305-319.
- Dunnell, R.C., 1978. Style and Function: A Fundamental Dichotomy, *American Antiquity* 43, 192-202.
- Greenacre, M., 2007. *Correspondence Analysis in Practice*, 2 ed., Chapman & Hall/CRC, London.
- Hill, M.O., Gauch, H.G., Jr., 1980. Detrended correspondence analysis: An improved ordination technique, *Vegetatio* 42, 47-58.
- Hubbard, R., Lindsay, R.M., 2008. Why P Values Are Not a Useful Measure of Evidence in Statistical Significance Testing, *Theory & Psychology* 18, 69-88.
- Jackson, D.A., Somers, K.M., 1991. Putting Things in Order: The Ups and Downs of Detrended Correspondence Analysis, *The American Naturalist* 137, 704-712.
- Kitching, I.J., Forey, P.L., Humphries, C.J., Williams, D.M., 1998. *Cladistics: The Theory and Practice of Parsimony Analysis*, Oxford University Press, Oxford.
- Kroeber, A.L., 1916a. Zuni Culture Sequences, *Proceedings of the National Academy of Sciences of the United States of America* 2, 42-45.
- Kroeber, A.L., 1916b. Zuñi Potsherds, *Anthropological Papers of the American Museum of Natural History*, 18, 1-37.
- Lipo, C., Madsen, M., 2001. Neutrality, "Style," and Drift: Building Methods for Studying Cultural Transmission in the Archaeological Record, in: Hurt, T.D., Rakita, G. (Eds.), *Style and function: conceptual issues in evolutionary archaeology*, Bergin & Garvey, Westport, Connecticut, pp. 91-118.
- Lipo, C.P., Madsen, M.E., Dunnell, R.C., Hunt, T., 1997. Population Structure, Cultural Transmission, and Frequency Seriation, *Journal of Anthropological Archaeology* 16, 301-333.
- Lipo, C.P., 2001. *Science, Style and the Study of Community Structure*, BAR, Oxford.
- Lockyear, K., 2000. Experiments with Detrended Correspondence Analysis, in: Lockyear, K., Sly, T.J.T., Mihailescu-Birliba, V. (Eds.), *Computer Applications and Quantitative Methods in Archaeology 1996*, British Archaeological Reports Oxford, pp. 9-17.
- Lockyear, K., in press. Applying bootstrapped Correspondence Analysis to archaeological data, *Journal of Archaeological Science*.
- Lyman, R.L., O'Brien, M.J., Dunnell, R.C., 1997. *The Rise and Fall of Culture History*, Plenum Press, New York.

- Marquardt, W.H., 1978. Advances in Archaeological Seriation, in: Schiffer, M.B. (Ed.), *Advances in Archaeological Method and Theory*, Academic Press, New York, pp. 257-314.
- Neiman, F.D., 1995. Stylistic Variation in Evolutionary Perspective: Inferences from Decorative Diversity and Interassemblage Distance in Illinois Woodland Ceramic Assemblages, *American Antiquity* 60, 7-36.
- O'Brien, M.J., Lyman, R.L., 1999. *Seriation, Stratigraphy, and Index Fossils: The Backbone of Archaeological Dating*, Kluwer Academic/Plenum, New York.
- O'Brien, M.J., Lyman, R.L., 2000. *Applying Evolutionary Archaeology*, Plenum Press, New York.
- O'Brien, M.J., Darwent, J., Lyman, R.L., 2001. Cladistics Is Useful for Reconstructing Archaeological Phylogenies: Palaeoindian Points from the Southeastern United States, *Journal of Archaeological Science* 28, 1115-1136.
- O'Brien, M.J., Lyman, R.L., 2003. *Cladistics and Archaeology*, The University of Utah Press, Salt Lake City.
- Peeples, M.A., Schachner, G., 2012. Refining correspondence analysis-based ceramic seriation of regional data sets, *Journal of Archaeological Science* 39, 2818-2827.
- Peet, R.K., Knox, R.G., Case, J.S., Allen, R.B., 1988. Putting Things in Order: The Advantages of Detrended Correspondence Analysis, *The American Naturalist* 131, 924-934.
- Ringrose, T.J., 1992. Bootstrapping and correspondence analysis in archaeology, *Journal of Archaeological Science* 19, 615-629.
- Shennan, S., 2004. *Quantifying Archaeology*, University of Edinburgh Press, Edinburgh.
- Shennan, S.J., Wilkinson, J.R., 2001. Ceramic Style Change and Neutral Evolution: A Case Study from Neolithic Europe, *American Antiquity* 66, 577-593.
- Ter Braak, C.J.F., 1985. Correspondence Analysis of Incidence and Abundance Data: Properties in Terms of a Unimodal Response Model, *Biometrics* 41, 859-873.
- Wartenberg, D., Ferson, S., Rohlf, F.J., 1987. Putting Things in Order: A Critique of Detrended Correspondence Analysis, *The American Naturalist* 129, 434-448.
- Weninger, B., 1995. Stratified <sup>14</sup>C dates and ceramic chronologies: case studies for the Early Bronze Age at Troy (Turkey) and Ezero (Bulgaria), *Radiocarbon* 37, 443-456.
- Георгиев, Г.И., Мерперт, Н.Я., Катинчаров, Р.В., Димитров, Д.Г., 1979. Езеро: раннобронзовото селище, Издателство на Българската Академия на Науките, София.

Figure captions:

Figure 1: Flow-chart of the significance testing procedure.

Figure 2: An illustration of the proportion of types and modes along the seriation sequence obtained by CA.

Figure 3: Correspondence analysis biplot of the Ezero data; building horizons labeled with Roman numerals, vertical handle types with Arabic numerals.

Figure 4: Seriation diagram of Ezero building horizons (in rows) ordered by the scores on the first CA axis. The seriation diagram was made using the Seriation Tool macro written by Tim Hunt and Carl Lipo (Lipo, et al., 1997) available on line at <http://www.lipolab.org/seriation.html>.

Figure 5: Distribution of total number modes based on 1000 randomizations of the Ezero data.

Figure 6: Correspondence analysis biplot of the control (random) data.

Figure 7: Seriation diagram of hypothetical assemblages from the Table 2 ordered by the scores on the first CA axis. The seriation diagram was made using the Seriation Tool macro written by Tim Hunt and Carl Lipo (Lipo, et al., 1997) available on line at <http://www.lipolab.org/seriation.html>.

Figure 8: Distribution of total number of modes based on 1000 randomizations of the control (random) data.

**Table 1. Ezero vertical handle data set (data from Георгиев, et al. 1979:Table 200). Building horizons in rows, type codes in columns.**

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
<b>II</b>	1	7	6	4	2	0	0	1	1	0	0	0	0	0	3	0	0	0	6	0	1	0
<b>III</b>	2	4	7	4	2	1	0	7	0	4	2	0	3	0	3	0	0	0	18	0	4	2
<b>IV</b>	5	10	3	8	4	3	0	5	1	5	4	0	3	1	5	0	0	0	14	0	5	3
<b>V</b>	0	3	2	15	7	1	0	5	1	3	5	0	3	0	6	0	0	0	11	1	5	3
<b>VI</b>	8	1	0	6	3	4	1	10	7	5	7	0	0	0	2	1	0	0	15	0	6	6
<b>VII</b>	24	5	1	1	3	16	0	14	9	2	6	0	0	0	2	0	0	0	18	1	12	12
<b>VIII</b>	11	0	0	0	0	6	0	7	4	0	2	0	0	0	6	0	0	0	8	1	0	3
<b>IX</b>	7	3	0	0	0	7	1	4	4	1	2	1	0	0	5	0	0	0	2	2	4	1
<b>X</b>	14	2	0	0	0	7	0	8	6	0	0	0	0	0	7	0	1	3	9	2	4	4
<b>XI</b>	32	3	0	0	0	11	0	8	0	1	0	0	0	0	0	0	5	3	12	6	6	5
<b>XII</b>	36	2	0	0	0	2	1	9	2	0	0	2	0	0	0	1	6	5	6	2	2	0
<b>XIII</b>	20	0	0	0	0	5	0	0	15	0	0	0	0	0	0	0	2	9	5	0	2	3

**Table 2. Randomly generated data set – control data.**

	A	B	C	D	E	F	G	
<b>Assem1</b>		44	50	6	13	24	6	28
<b>Assem2</b>		1	6	37	45	39	42	10
<b>Assem3</b>		17	25	14	37	35	22	41
<b>Assem4</b>		30	13	30	7	22	22	12
<b>Assem5</b>		26	37	25	23	30	23	20
<b>Assem6</b>		1	21	34	5	43	42	1
<b>Assem7</b>		44	16	25	7	32	5	47
<b>Assem8</b>		4	18	27	19	30	8	5
<b>Assem9</b>		36	38	42	3	37	37	26
<b>Assem10</b>		36	34	38	43	43	0	49
<b>Assem11</b>		37	12	43	48	47	43	6
<b>Assem12</b>		2	13	15	29	7	18	24

**Table 3. CA summary for columns (types) of Ezero data table (for details about the descriptive measures presented in the table see Shennan, 2004:325-327).**

Type	Mass	Quality	Inertia	Axis 1 score	Cor 1	Ctr 1	Axis 2 score	Cor 2	Ctr 2
1	0.19	0.956	0.123	0.695	0.885	0.232	0.197	0.071	0.063
2	0.048	0.61	0.056	-0.619	0.383	0.046	0.477	0.227	0.092
3	0.023	0.749	0.103	-1.313	0.448	0.098	1.077	0.301	0.223
4	0.045	0.731	0.108	-1.205	0.72	0.165	0.152	0.011	0.009
5	0.025	0.762	0.042	-1.035	0.761	0.067	0.021	0	0
6	0.075	0.642	0.038	0.361	0.304	0.025	-0.381	0.338	0.092
7	0.004	0.056	0.015	0.43	0.053	0.002	-0.108	0.003	0
8	0.093	0.14	0.017	0.029	0.005	0	-0.145	0.135	0.016
9	0.059	0.337	0.078	0.494	0.219	0.037	-0.361	0.117	0.066
10	0.025	0.568	0.031	-0.746	0.537	0.035	-0.178	0.031	0.007
11	0.033	0.764	0.032	-0.538	0.363	0.024	-0.566	0.401	0.091

12	0.004	0.289	0.026	1.008	0.168	0.009	0.855	0.121	0.022
13	0.011	0.612	0.038	-1.308	0.574	0.046	0.337	0.038	0.01
14	0.001	0.148	0.014	-1.15	0.137	0.004	0.323	0.011	0.001
15	0.046	0.246	0.048	-0.423	0.207	0.021	-0.184	0.039	0.013
16	0.002	0.072	0.012	0.492	0.056	0.001	0.261	0.016	0.001
17	0.017	0.853	0.05	1.149	0.519	0.055	0.922	0.334	0.12
18	0.024	0.586	0.084	1.204	0.488	0.087	0.542	0.099	0.059
19	0.147	0.517	0.025	-0.269	0.516	0.027	0.006	0	0
20	0.018	0.262	0.029	0.589	0.251	0.016	0.124	0.011	0.002
21	0.061	0.308	0.013	-0.117	0.076	0.002	-0.204	0.232	0.021
22	0.05	0.679	0.019	0.036	0.004	0	-0.46	0.674	0.09

**Table 4. CA summary for rows (building horizon assemblages) of Ezero data table (for details about the descriptive measures presented in the table see Shennan, 2004:325-327).**

Building horizon	Mass	Quality	Inertia	Axis 1 score	Cor <sup>a</sup> 1	Ctr <sup>b</sup> 1	Axis 2 score	Cor <sup>a</sup> 2	Ctr <sup>b</sup> 2
II	0.038	0.712	0.115	-1.036	0.423	0.103	0.858	0.29	0.238
III	0.075	0.697	0.089	-0.784	0.617	0.116	0.282	0.08	0.051
IV	0.094	0.748	0.08	-0.724	0.731	0.124	0.111	0.017	0.01
V	0.084	0.692	0.134	-0.963	0.691	0.197	-0.046	0.002	0.002
VI	0.098	0.492	0.052	-0.235	0.124	0.014	-0.406	0.369	0.137
VII	0.15	0.525	0.043	0.098	0.04	0.004	-0.342	0.485	0.149
VIII	0.057	0.364	0.035	0.229	-0.103	0.008	-0.365	0.261	0.065
IX	0.052	0.167	0.045	0.193	0.051	0.005	-0.29	0.116	0.037
X	0.08	0.436	0.03	0.342	0.368	0.024	-0.147	0.068	0.015
XI	0.109	0.558	0.083	0.569	0.506	0.089	0.182	0.052	0.031
XII	0.09	0.807	0.143	0.855	0.55	0.167	0.585	0.257	0.263
XIII	0.073	0.468	0.152	0.905	0.465	0.15	0.076	0.003	0.004

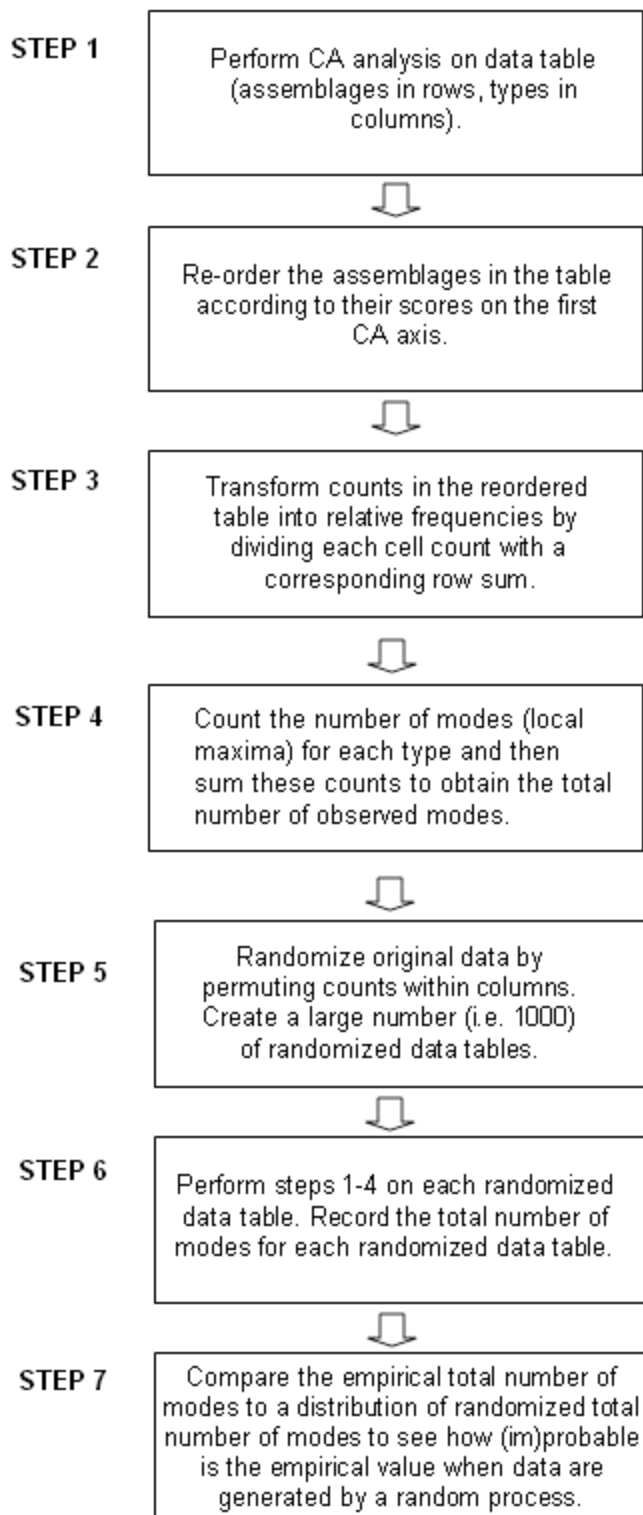
**Table 5. CA summary for columns (types) of the random data table (for details about the descriptive measures presented in the table see Shennan, 2004:325-327).**

Type	Mass	Quality	Inertia	Axis 1 score	Cor 1	Ctr 1	Axis 2 score	Cor 2	Ctr 2
A	0.132	0.772	0.189	-0.507	0.648	0.243	0.222	0.124	0.096
B	0.135	0.425	0.122	-0.289	0.331	0.08	0.154	0.094	0.047
C	0.16	0.581	0.075	0.251	0.48	0.072	0.115	0.101	0.031
D	0.133	0.919	0.187	0.239	0.146	0.054	-0.55	0.773	0.595
E	0.185	0.409	0.045	0.143	0.304	0.027	0.084	0.105	0.019
F	0.127	0.832	0.196	0.565	0.747	0.291	0.19	0.085	0.068
G	0.128	0.812	0.187	-0.504	0.626	0.232	-0.275	0.186	0.143

**Table 6. CA summary for rows (assemblages) of the random data table (for details about the descriptive measures presented in the table see Shennan, 2004:325-327).**

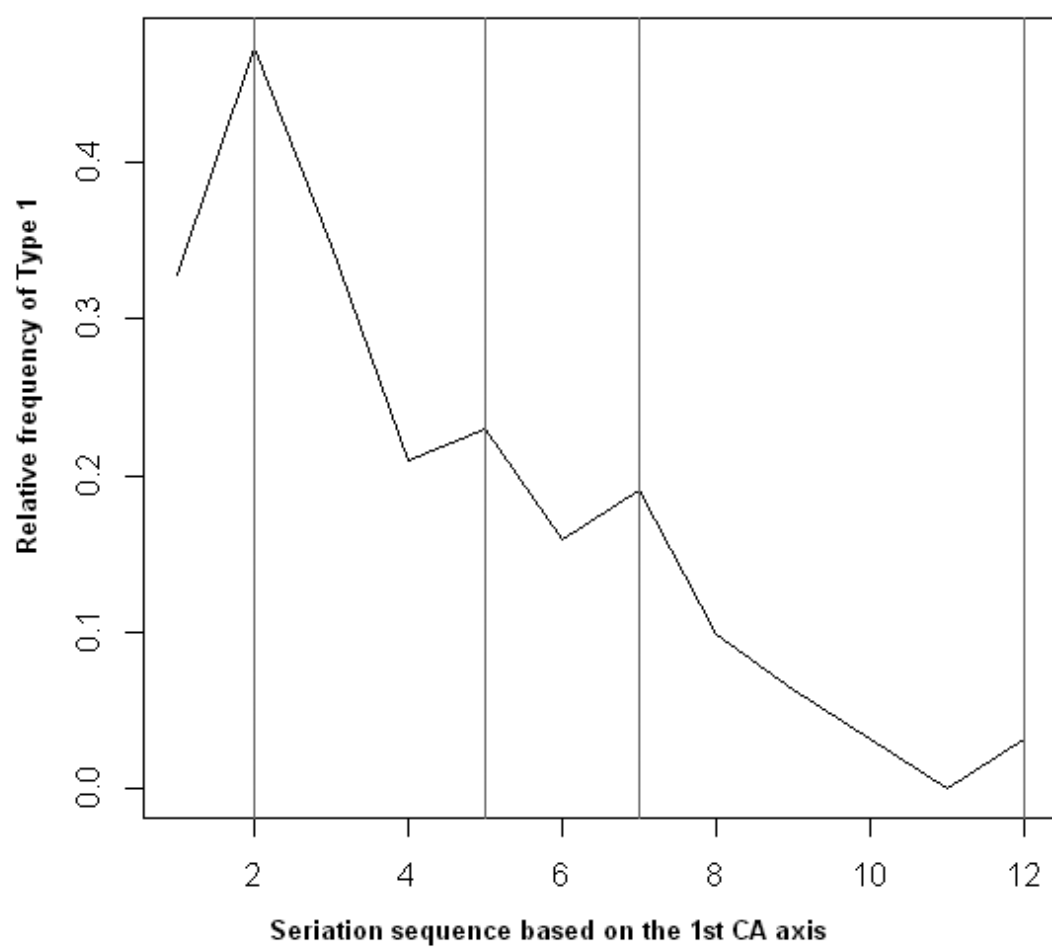
Assemblage	Mass	Quality	Inertia	Axis 1 score	Cor 1	Ctr 1	Axis 2 score	Cor 2	Ctr 2
Ass1	0.081	0.782	0.15	-0.616	0.741	0.221	0.145	0.041	0.025
Ass2	0.086	0.991	0.138	0.625	0.871	0.239	-0.232	0.119	0.068
Ass3	0.091	0.694	0.049	-0.094	0.059	0.006	-0.308	0.635	0.127
Ass4	0.065	0.576	0.039	-0.004	0	0	0.311	0.576	0.092
Ass5	0.088	0.212	0.014	-0.071	0.116	0.003	0.064	0.096	0.005
Ass6	0.07	0.877	0.15	0.592	0.586	0.175	0.417	0.291	0.18

Ass7	0.084	0.697	0.125	-0.536	0.693	0.172	0.043	0.004	0.002
Ass8	0.053	0.259	0.047	0.25	0.254	0.024	-0.036	0.005	0.001
Ass9	0.104	0.86	0.055	-0.059	0.024	0.003	0.351	0.836	0.19
Ass10	0.116	0.804	0.078	-0.295	0.464	0.072	-0.252	0.34	0.109
Ass11	0.112	0.496	0.082	0.318	0.495	0.081	-0.015	0.001	0
Ass12	0.051	0.693	0.074	0.124	0.038	0.006	-0.512	0.655	0.199

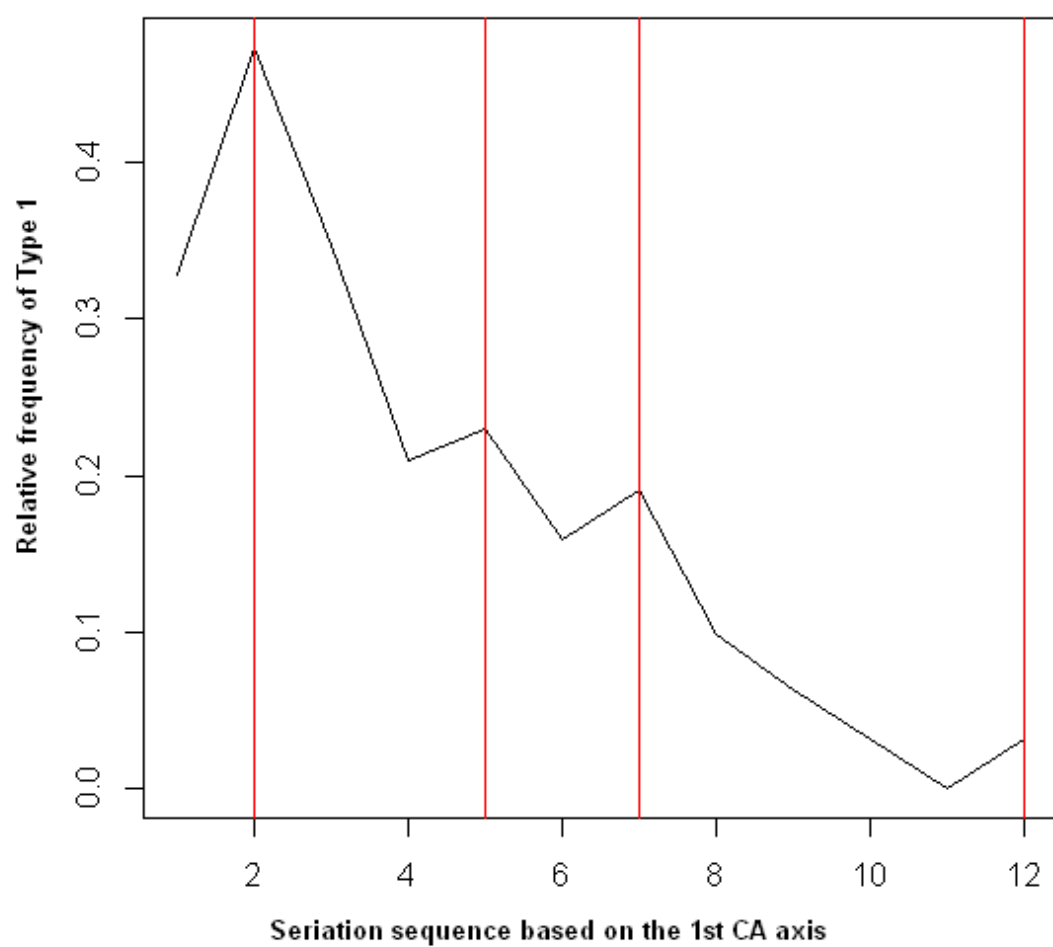


USCRIPT

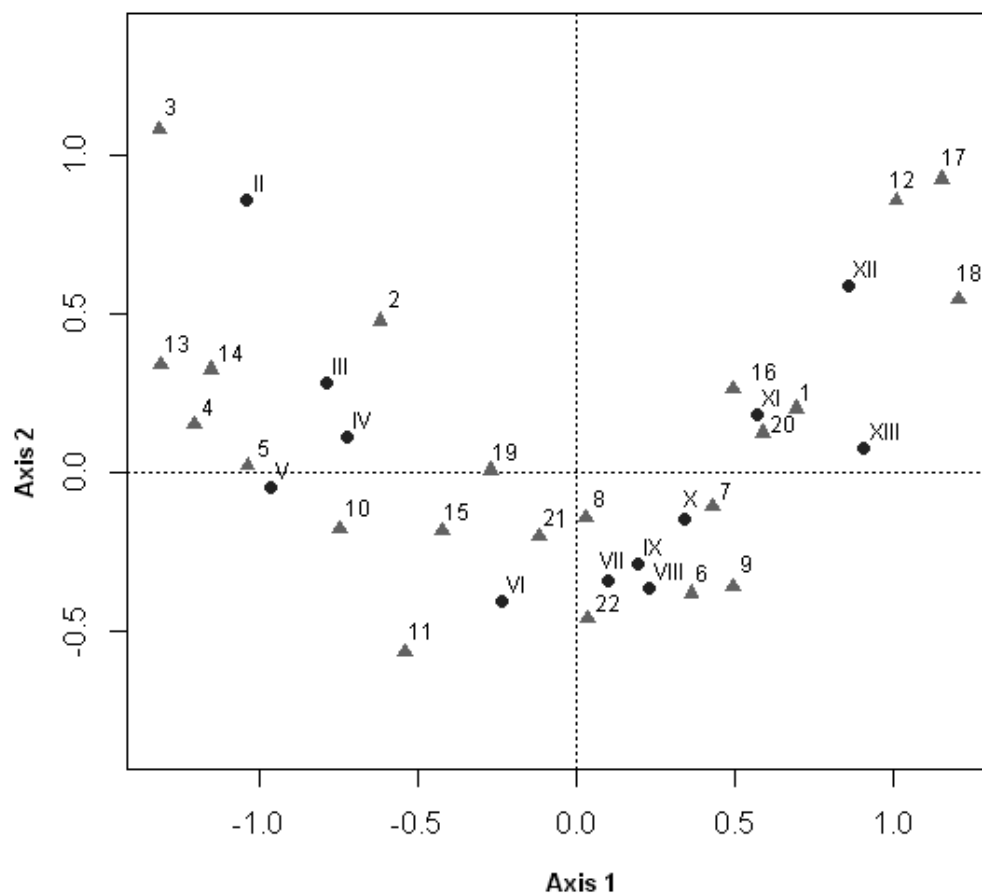


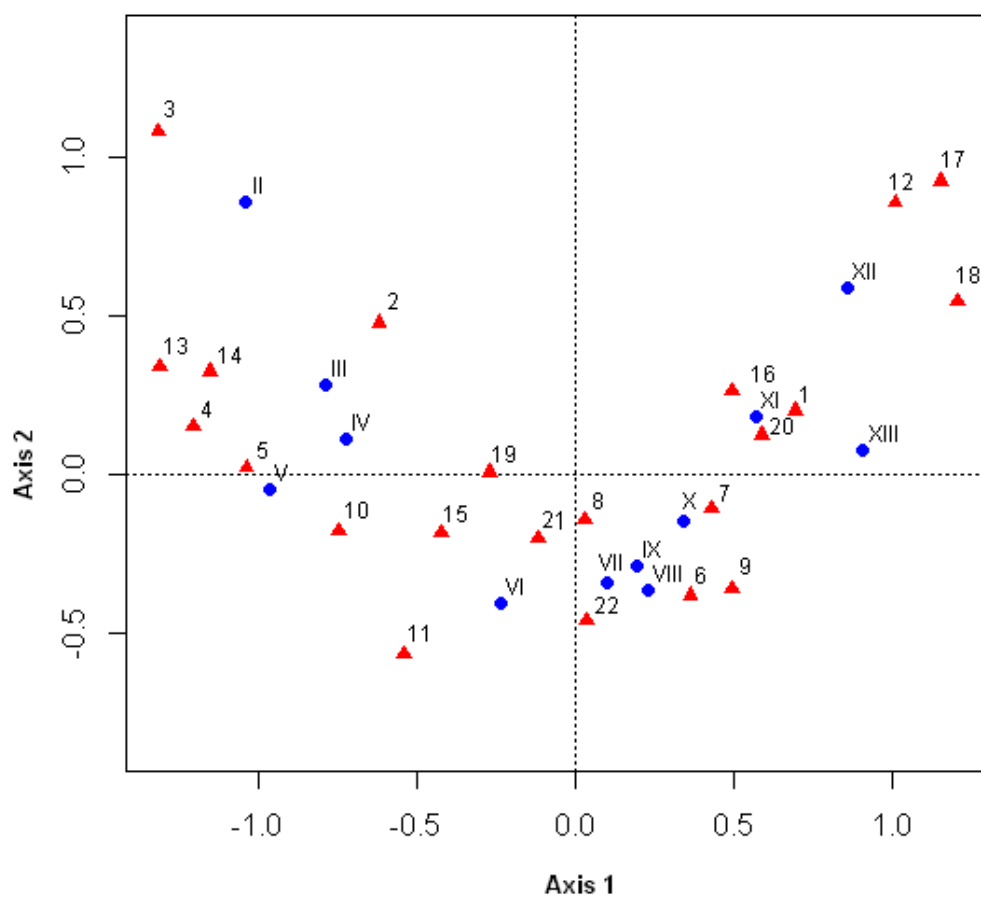


ACCEPTED



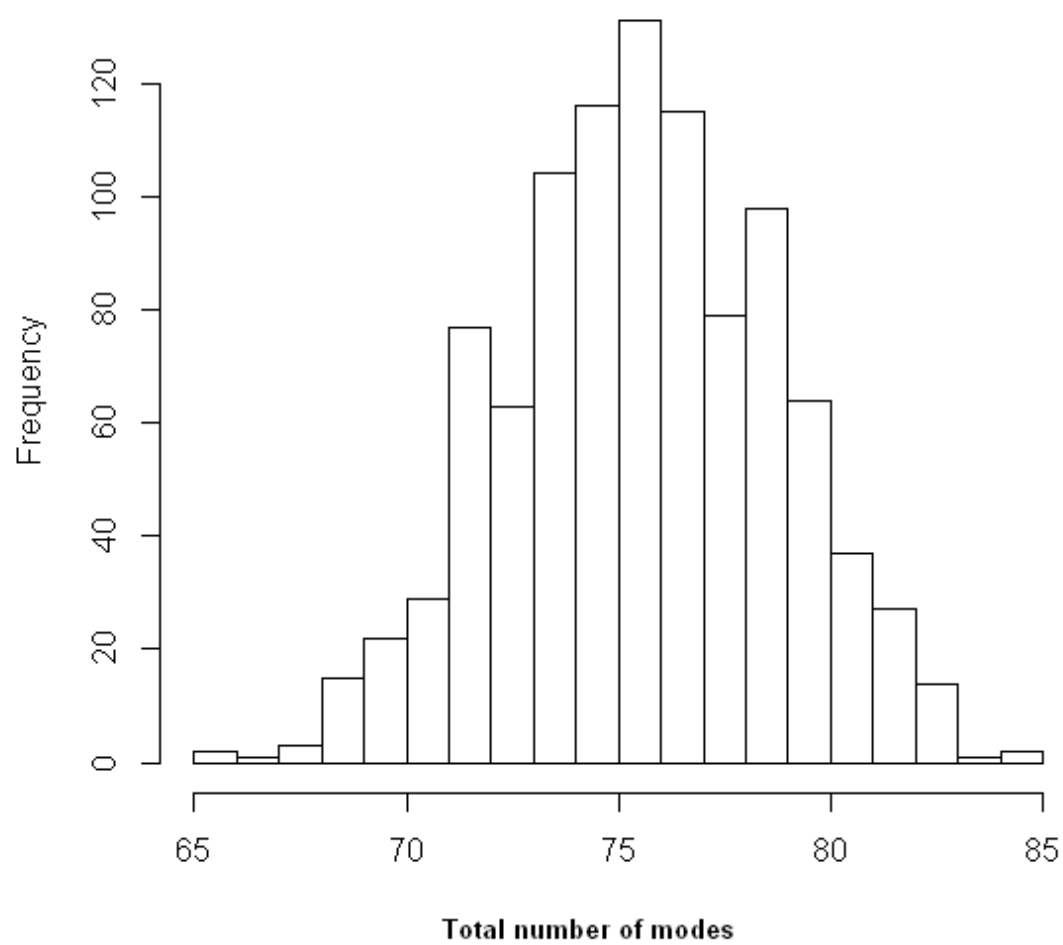
ACCEPTED



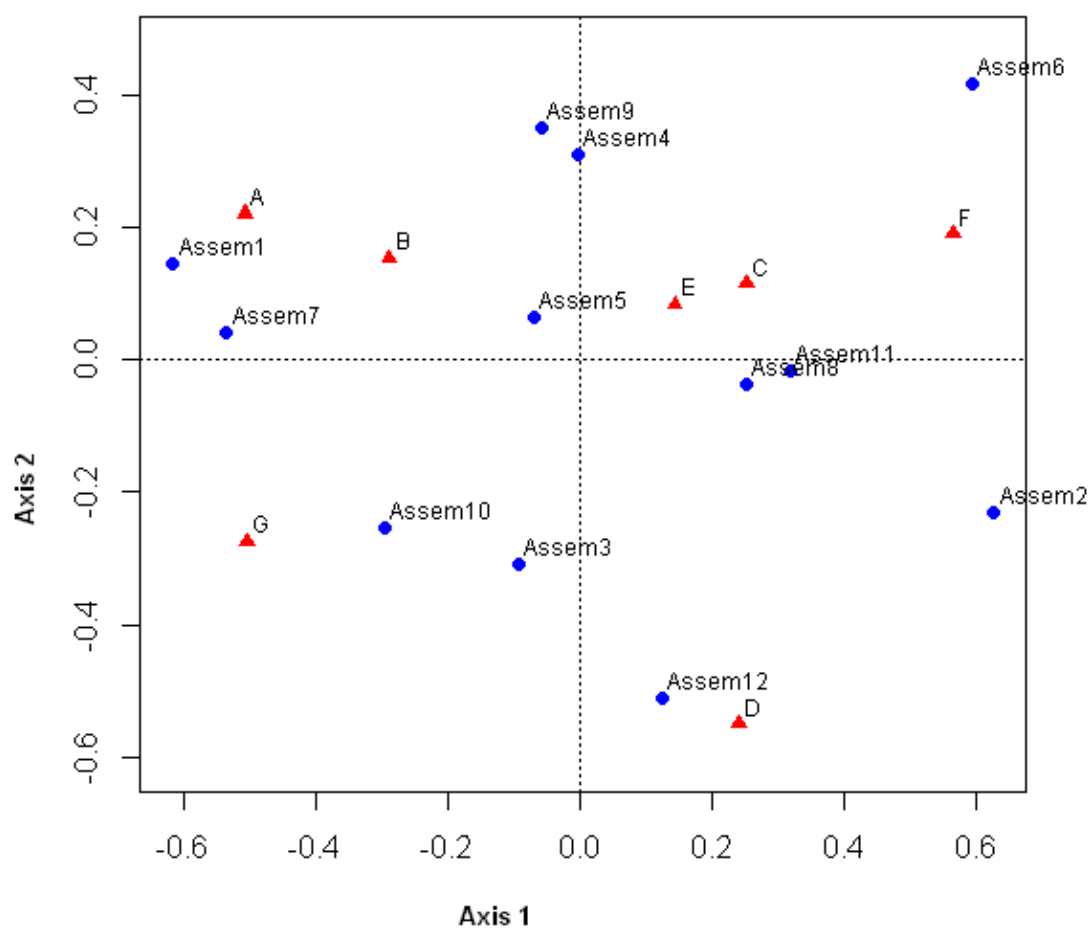


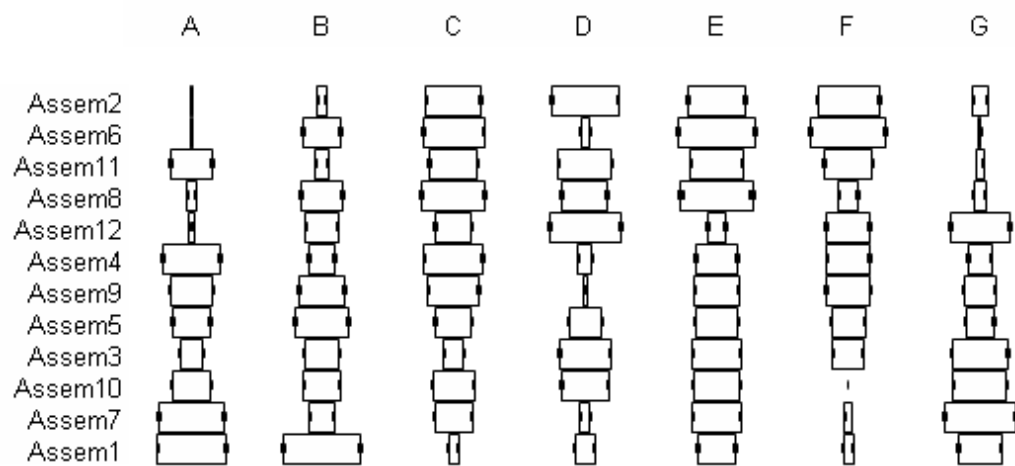
	Type
1	.....
2	.....
3	.....
4	.....
5	.....
6	.....
7	.....
8	.....
9	.....
10	.....
11	.....
12	.....
13	.....
14	.....
15	.....
16	.....
17	.....
18	.....
19	.....
20	.....
21	.....
22	.....

ACCEPTED MANUSCRIPT



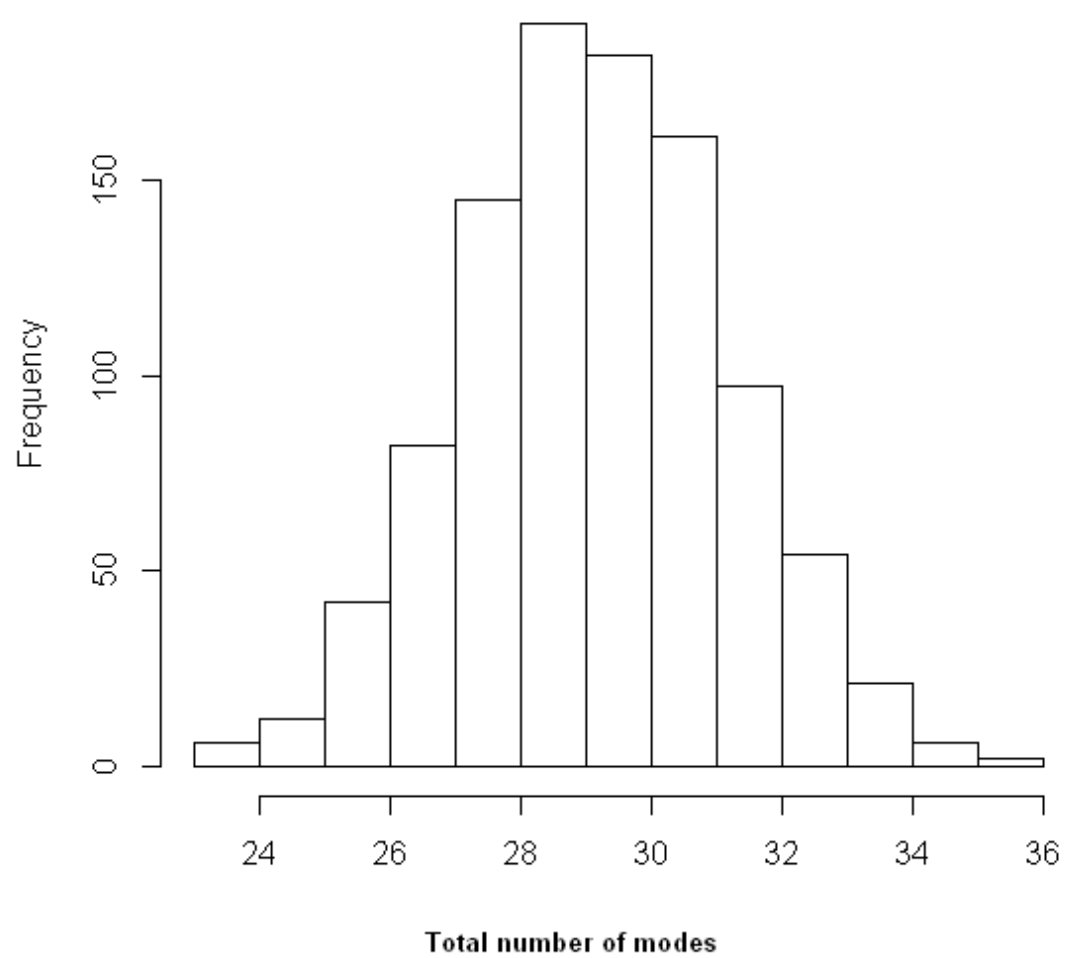
ACCEPTED





ACCEPTED MANUSCRIPT





ACCEPTED

Research highlights:

- A method for testing the significance of seriation solutions is presented
- Seriation coefficient is presented as a goodness of fit measure
- The method successfully detects chronological signal when present
- The method fails to detect chronological signal when absent

ACCEPTED MANUSCRIPT

## R CODE FOR PERFORMING THE PERMUTATION SIGNIFICANCE TEST

```

library(ca) #library ca has to be installed
library(plyr) #library plyr has to be installed
Perm <- 1000 # Perm sets the number of permutations for the
randomization test. The default is 1000 but it can be changed by the
user.
PERM <- c(1:Perm)

#Defining function for counting modes

localMaxima <- function(x) {
  # Use -Inf instead if x is numeric (non-integer)
  y <- diff(c(-.Machine$integer.max, x)) > 0L
  rle(y)$lengths
  y <- cumsum(rle(y)$lengths)
  y <- y[seq.int(1L, length(y), 2L)]
  if (x[[1]] == x[[2]]) {
    y <- y[-1]
  }
  y
}

#Reading data#

# Select and Copy a complete data table (in the form of Tables 1-2 in
the main text) from a spreadsheet

data <- read.table("clipboard") #after copying the data from the
spreadsheet, run this line

#Performing correspondence analysis (CA) on the data and calculating
the number of modes for the CA solution

M <- length(data[1,])
a <- c(1:M)
b <- c(1:M)
data2 <- as.matrix(data)
ord <- ca(data)$rowcoord[,1]
data <- as.matrix(data)
data1 <- cbind(ord, data)
data1 <- as.data.frame(data1)
G <- arrange(data1, desc(ord))
matr <- G[,2:(M+1)]/(apply(G[,2:(M+1)],1, sum))

for(j in 1:M) {

a[j] <- length(localMaxima(matr[,j]))

}

sum(a) #gives the observed total of modes

```

```
# Generating the distribution of total number of modes with randomized
data

for(i in 1:Perm) {
  for(j in 1:M) {
    data2[,j] <- sample(data[,j], replace = FALSE)}
    ord <- ca(data2[which(rowSums(data2)>0),,])$rowcoord[,1]
    data2 <- as.matrix(data2)
    data3 <- cbind(ord, data2[which(rowSums(data2)>0),,])
    data3 <- as.data.frame(data3)
    G <- arrange(data3, desc(ord))

    matr <- G[,2:(M+1)]/(apply(G[,2:(M+1)],1, sum))

    for(j in 1:M) {
      b[j] <- length(localMaxima(matr[,j]))}

    PERM[i] <- sum(b)}

hist(PERM)          # Draws a histogram of randomized total number of
modes

quantile(PERM, 0.05) # Gives the value of the 5th percentile of the
randomized distribution of total number of modes
```