Toward a Cultural Measure of Time: Remarks on Phasing CA-Derived Petrie-Matrices and the Use of DCA

Tim Kerig
Universität zu Köln - University of Cologne
Weyertal 125, 50923 Köln, Germany
kerigtim@hotmail.com

Abstract

Correspondence Analysis (CA) offers a proven algorithm for seriation. The eigenvalue-solution can be displayed in a coordinate system for showing the well known arch or parabola effect. There are serious obstacles for the use of these eigenvalues -or the differences between them- as arguments for phasing Petrie-matrices. Detrendet Correspondence Analysis (DCA) is an alternative algorithm, widely used in ecology, but not yet for seriation. DCA allows one to obtain an eigenvalue-solution different from that of CA, designed especially for eliminating the parabola effect. Displayed in a coordinate system the DCA eigenvalues allow a sounder procedure of phasing by measuring equally scaled distances. An example from the early Neolithic Mid-European Linearbandkeramik will show that these distances can be used as a measure of culturally represented time: The scaling of the DCA-solution leads to a general scale as an independent tool for quantifying changes in material culture by explicit archaeological means.

Relative chronology represents time - without any scale. Without external information we simply cannot measure the time depth between two strata in a stratigraphic context or between two graves in a seriation. We cannot even make a qualified guess. Here I will introduce a culturalistic concept of time and its formalized representation that allows an objective scaling of relative chronologies.

Seriation is the arrangement of combinations in terms of their elements along a gradient. Besides the strict definition (a matrix with the consecutive-1's property) I use the term Petrie-matrix in a broader sense: diagonalising a table with types of artefacts in the columns and finds containing the artefacts in the rows should come as close as possible to a matrix with a consecutive and diagonal block of numbers (cf. Kendall 1971).

Correspondence analysis (CA) is today the most common method for such a rearrangement (for example Baxter 1994:100-39, Müller and Zimmermann 1997). Before showing the advantages of an alternative algorithm, Detrended Correspondence Analysis (overview: Jongman 1995), it might be helpful to remember the principles of CA: The first step in CA is the conversion of the types counts into proportional data. A matrix of chi-squared distances is calculated. From this matrix eigenvectors are extracted by orthogonal regression, performed by singular value decomposition. There are as many eigenvectors computed as the smaller number of variables either in rows or columns, minus one. For the variables -both finds and types- the same principal component space is calculated. The first eigenvector accounts for most of the information (inertia) in the dataset; the second is situated to describe the variation of the remainder, and so forth. Let every eigenvector be an axis in an orthogonal coordinate system; hence the position of the variables -finds as well as types- can be determined by coordinates on the eigenvectors. The values on the first eigenvector - or first CA-axis - are of higher explanatory value than the values of all other single eigenvectors. If this main dimension of explanation shows characteristics that are assumed to be also the characteristics of our concept of relative chronology, then we can use the main dimension as gradient in seriation. One can display the existence of a dominant, single, and natural gradient by using the well-known Guttman effect (Greenacre 1984:226-232) for the so called parabola- or horseshoe-test: Here the values on the first eigenvector are plotted against the values on the second. A parabola forms, proving the existence of a single natural gradient. The seriation, the diagonalisation of the original contingency table, is quite simple: We have already identified the main dimension of explanation and equated it with time. The values on the first vector are -somehow!- representing time. The variables, the combinations in the rows as well as the types in the columns, are rearranged according to the ranks of their values.

In daily archaeological practice there is often a need for using phases rather than such ranks, e.g. if you want to date a handful of sherds from a surface collection. For such cases it is useful to determine typological phases. Phases may be defined as periods of hypothetical contemporaneity. There are different ways of phasing a rank-ordered Petrie-matrix: One may take the range of single dominant types for defining phases. This may be useful if there is already an established typological sequence and one wants to integrate the results of the seriation. There are also some examples in the literature of CA-derived Petrie-matrices, phased by Cluster-analysis of the CA-values. As the CA-solution is described as a set of coordinates, it is tempting to use these values as criteria for phasing. By using the coordinates in such a way one takes for granted that the computed CA-space allows the representation of an adequate concept of time. One would assume that the same absolute value in differences between the coordinates represent the same duration - that is not the case. The computed CA-space is not adequate for representations of our concept of absolute chronology. We assumed the CA-axes to be completely uncorrelated and we've therefore chosen an orthogonal coordinate system. There is no information about the absolute duration of anything at all. Expecting a chronology "a little bit more absolute" would be expecting the computer to do magic - doing the transubstantiation from relative to absolute chronology.

Detrended correspondence analysis (DCA) is an heuristic offspring of CA, especially designed to handle the problems in the relationship between the first and the second axis and to adjust the distortions already described. Software solutions are available in different Statistical Software Packages, such as CANOCO and the famous Bonn package WinBASP.

There are two principal methods of DCA: Detrending by polynomials and by segments. The first is more stable and considered more elegant. Nevertheless, detrending by segments gives a better first impression of the method. Imagine the solution of a CA displayed in a scatter plot. Now the first axis is divided into an arbitrary number of segments. These segments are moved in such a way that their mean on the second axis is approximately zero. Higher axes are treated in the same way with respect to each of the existing axes. In fact DCA by segments is not a method of rectifying CA-solutions; it is an algorithm of its own. Actually it is a CA that uses detrending instead of orthogonalisation.

DCA being without orthogonalisation offers the opportunity to go a step further: It is possible to scale the first axis in terms of the combination of variables, for example of artefacts of different types in graves. The appearance and disappearance of types along the first axis would be even - the rate of appearance along the time-representing gradient is held constant. Let the occurrence of artefacts of the same type be of normal distribution. As the standard deviation divides such curves in standardized areas, one can describe the phases in the use of artefacts of the same type in terms of standard deviation. By using the relation between the standard deviation and the variance - the first is the square root of the latter - one can equalize the curves' width: The variance of the optima of the types in a find should be an estimate of the average squared standard deviation of the types. For scaling one has to equalize the variances of the types-optima-in-the-finds as nearly as possible. Technically the first DCA-axis is divided in segments, and these segments are expanded if the finds in it show a low variance of the types' optima. Conversely, if the variance of the optima of the types is high, the segments are contracted. It is the positioning of the types that is rearranged. The finds are positioned in such a way that their coordinates are weighted means of the coordinates of the types that occur in them. As a standard scaling, the variance of the optima of the types in finds is set to 1. This method of scaling was introduced into ecology by Gauch and Whittaker 1972. They called the resulting unit Z. In 1980 Hill and Gauch described the scaling in combination with DCA. They named the unit standard deviation, sd.

Types situated at a distance larger than 4 sd will practically not occur together. The standardization allowed ecological sciences to describe the development of diversity along a gradient.

The use of standard deviations as units offers a wide range of possibilities. Distances between types or between finds can now be measured, and it will be possible to estimate contemporarity in terms of probability. Another important technical advantage is that one can define the length of the first axis as the range of the finds curves - if the range is too narrow we should check whether seriation is adequate. Finally, it allows the use of sd phasing: 1.) Perform the CA for identifying the main dimension of explanation; 2.) do the horseshoe-test for showing the qualities of the main dimension, and then 3.) calculate DCA and 4.) the scaling. If the first axis is of an adequate length, 5.) order the matrix using values as ranks, and 6.) finally show phases in units of standard deviations.

I will demonstrate the approach with an example (Fig.1): My data-set contains early Neolithic Bandkeramik from the region between the rivers Rhine, Main, and Weser. The contingency table contains settlement pits in the rows and types of ceramic decoration in the columns. The horseshoe-test shows the dominance of a single gradient. The ranks of the values of CA and DCA are the same and they remain practically the same after scaling. Earlier investigations made clear that on the first axis time is represented. On the second axis regional styles are displayed. The new scaling of the time axis confirmed the cultural sequence established by Meier-Arendt (1966). Meier-Arendt's sequence was recently attacked

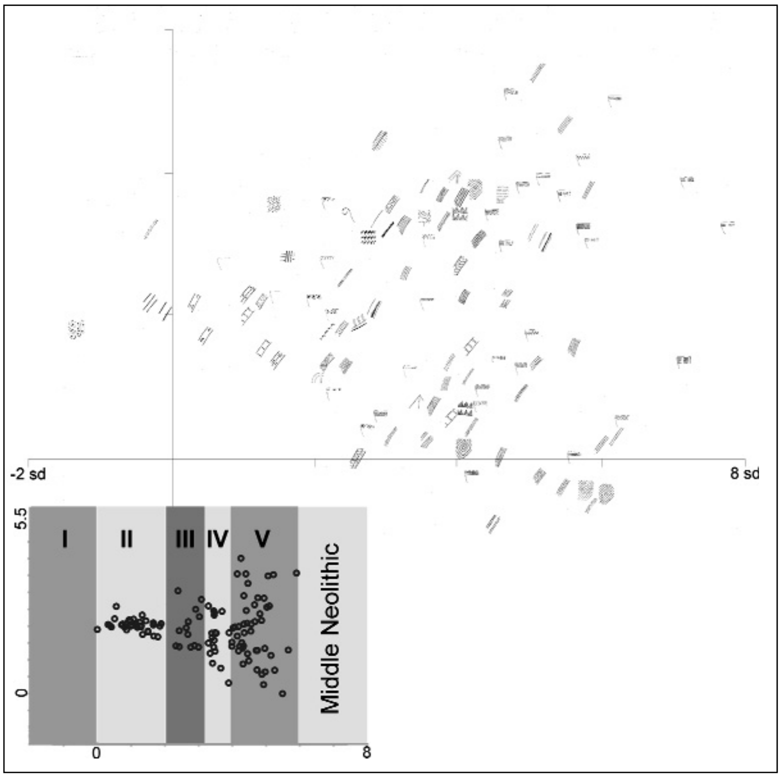

Figure 1 Early Neolithic Bandkeramik from the region between the rivers Rhine, Main and Weser. The types of ceramic decoration are shown in the plane of the first two DCA axes, scaled in sd units. The smaller box shows the pits as calculated in the same way and with the phases of Meier-Arendt inserted

(Kneipp 1998) because of its purely impressionistic foundation. Sd-units show the accuracy of Meier-Arendt's sense of style and at the same time they make it useable in an objectified way. Just an impression: Compare the marked groups of ornaments - the standing triangles and the rhomboids. Both groups are practically of the same relative age, but they are found in different regions. And please pay attention to the overall pattern. Displayed is the growing diversity from the very first Mid-European ceramic into regional groups. After a century of research, DCA allowed us for the first time to detect and show simultaneously the stylistic sequence on the first axis and the spatio-stylistic development of the neolithic groupings on the second (for further detail, Kerig, in press).

Scaling of relative chronologies in units of standard deviations has several attractive advantages on a technical level. Furthermore, scaling of relative sequences is a search for cultural traits. Imagine a large graveyard dated by seriation. The measurement allows us to compare and to scale intervals of change in different classes of artefacts. It must be meaningful, if, e.g., weaponry changes ten times faster than jewellery during a shortage of resources. It seems as if we could obtain a kind of an independent archaeological measurement for artefactual change - something like an archaeological unit.

References

BAXTER, M.J., 1994. Exploratory multivariate analysis in Archaeology. Edinburgh.

GAUCH, H.G. and WHITTAKER, R.H., 1972. Comparison of ordination techniques. Ecology 53:868-875.

GREENACRE, M.J., 1984. Theory and Applications of Correspondence Analysis. London.

HILL, M.O. and GAUCH, H.G., 1980. Detrended correspondence analysis: an improved ordination technique. Vegetatio 42:.47-58.

JONGMAN, R.H.G., 1995. Ordination. In Jongman, R.H.G., ter Braak, C.J.F. and van Tongeren, O.F.R. (eds.), Data analysis in community and landscape ecology, Cambridge:91-173.

KENDALL, D.G., 1971. Seriation from abundance matrices. In Hodson, F.R., Kendall, D.G. and Tautu, P. (eds.), Mathematics in the Archaeological and Historical Sciences, Edinburgh:215-252.

KERIG, T., 2002, in press. Zur relativen Chronologie der westdeutschen Bandkeramik. Paper presented at the Symposium "Die Bandkeramik im 21. Jahrhundert", Brauweiler, 16-20 September 2002.

KNEIPP, J., 1998. Bandkeramik zwischen Rhein, Weser und Main. Studien zu Stil und Chronologie der Keramik. UPA 47, Frankfurt.

MEIER-ARENDT, W., 1966. Die bandkeramische Kultur im Untermaingebiet. Amt Bodendenkmalpfl. Darmstadt 3, Bonn.

MÜLLER, J. and ZIMMERMANN, A. (eds.), 1997. Archäologie und Korrespondenzanalyse. Beispiele, Fragen, Perspektiven. Internat. Arch. 23, Espelkamp.