Developments in Statistics and Methodology
A. Ferligoj and A. Kramberger (Editors)
Metodološki zvezki, 9, Ljubljana: FDV 1993

Effect of Configuration Type on Success of Recovery for Three Multidimensional Scaling Methods

K. Košmelj[1] and W. J. Krzanowski[2]

Abstract

The objective of this study was to assess quantitatively the success of three scaling methods in recovering a known multidimensional configuration. A range of configuration types was considered, from strictly clustered, through mixed, to purely random. The multidimensional scaling methods used were metric scaling, ordinal scaling and metric scaling on ranked input distances. The results indicate that previous recommendations, which were based on random configurations only, need to be modified in the presence of clustering.

Keywords: Analysis of variance; Configuration comparison; Multidimensional scaling methods; Procrustes analysis.

1 Introduction

Multidimensional scaling is a popular technique of multivariate analysis in many areas of application. It is appropriate whenever the available data are in the form of, or can be converted into, an (n x n) matrix of similarities or dissimilarities between every pair out of n stimuli or objects. The technique then aims to produce a low-dimensional configuration of n points in which the points represent the stimuli and the distance between any two points approximates the similarities or dissimilarities between the corresponding stimuli.

There are now many variants of the technique (see, e.g. Davidson, 1983), but two in particular have found favour among statisticians. The first of these is metric scaling

1 Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, 61000 Ljubljana, Slovenia
2 Mathematical Statistics and Operational Research Department, Exeter University, Streatham Court, Rennes Drive, Exeter EX 4PU, United Kingdom

(MS), also known as classical scaling (Torgerson, 1958) or the principal coordinate analysis (Gower, 1966); in this method the approximation between inter-point distances δ_{ij} and inter-object dissimilarities δ_{ij} is such that the value of $\Sigma\Sigma(\delta_{ij} - d_{ij})^2$ is minimised. Thus in this method the configuration of points is one in which the actual values δ_{ij} are recovered as well as possible. The second approach is non-metric scaling (Kruskal, 1964), also known as ordinal scaling (OS); in this method the configuration of points is the one in which the rank order of the d_{ij} matches that of the δ_{ij} as well as possible. Details of both methods can be found in most books on multivariate analysis (e.g. Chatfield and Collins, 1980).

In many applications, non-metric scaling is favoured over the metric version even when δ_{ij} are available in numerical form rather than merely as ranks, because a good representation can often be found in fewer dimensions than with metric scaling. However, it is also well known that if the objects or stimuli are clustered and within-cluster distances are smaller than the between-cluster distance, then non-metric scaling can lead to degenerate solutions whereas metric scaling is similarly not affected.

In an attempt to obtain the best of two worlds, Weeks and Bentler (1979) suggested carrying out metric scaling but on the ranked input distances (MRS), and justified their recommendation by means of a stimulation study. In this study they generated initial configuration from a uniform pseudorandom generator. They then calculated Euclidean distances d from these random configurations, added a random error e to each distance and distorted the result by a known function f, i.e. h = f(d + e). The following functions f were considered: identity, power, rank and absolute value; the first three are all monotone while the fourth is not. All their datasets were then scaled using the linear model (i.e. MS), the monotone model (i.e. OS), and the 'rank - linear model' (i.e. MRS). The authors concluded their study with the following remarks:

"The monotone model has the advantage of robustness over the linear model, since, in this study, it performed better than the linear model for all systematic and nonmonotone distortions. The linear model has the advantage of conceptual efficiency and avoids the danger of degeneracy. The rank- linear model appears to offer the advantages of both. The only computation it requires over the linear model is the initial ranking of data. The ranking eliminates all systematic monotone nonlinearities, whereas the linear analysis avoids the potential of degeneracy due to monotone regression."

It should be pointed out that only random configurations were included in their study. However, random configurations are not very likely in the analysis of real data. On the other hand degenerate solutions can be expected with clustered initial configurations, as these clusters may collapse into points in reconstructed configurations.

The objective of the present study was therefore to investigate the performance of MRS on other types of configurations and to compare it more fully with MS and OS. Does MRS 'offer the advantage of MS and OS' only for random input configurations? Configurations having particular characteristics were generated by computer and multidimensional scaling was done using each of three methods. The following configuration characteristics were varied: number of points, dimension of configuration and configuration type. Details of the study design are given in Section 2, results are presented in Section 3, and conclusions are summarised in Section 4.

2 Design of the study

2.1 Configuration Type

We wished to investigate the three multidimensional scaling methods on a range of configuration types that included both 'random' and 'clustered' configurations. In order to do so, we first set up the following definitions. Let a unit (hyper-)cube of given dimensionality r be divided into a grid of equal subcubes. A 'cluster' is defined to be a set of points allocated at random within one such subcube. For a fixed number k of clusters, a configuration of N points is called 'strictly clustered' if every point is allocated to a cluster and clusters lie within strictly non-adjacent subcubes in the grid.

Now suppose that a configuration contains N points. A continuum of configurations ranging from 'strictly clustered' to 'random' can be defined as follows. Let $(1 - \mu) \times N$ be the number of points that are allocated to a strictly clustered configuration, and let the remaining $\mu \times N$ points be allocated randomly within the whole unit cube. When $\mu = 0$ the configuration is strictly clustered, and when $\mu = 1$ the configuration is random. For $0 < \mu < 1$ it is a mixed configuration, the size of μ determining the amount of structure that it contains.

Figure 1 gives some examples of these configurations in 2-dimensional space.

2.2 Factors under study

In our study, the following factors were varied:

- N, the number of points (N = 30 points and N = 60 points);
- r, the dimensionality of the input configuration (r = 1, 2, 3, 4);
- μ, the type of input configuration (μ = 0, 1/3, 2/3, 1);
- the scaling method (MS, OS, MRS).

3 independent replications were generated for each combination of N, r and μ. Thus 2 x 4 x 4 x 3 = 96 configurations were processed by each of the three scaling methods.

2.3 Methods of assessment

Let us denote by δ_{ij} the Euclidean distance between the points i and j in the original configuration in r dimensions, and by d_{ij} the Euclidean distance between the corresponding points in the p-dimensional subspace generated by one of the multidimensional scaling methods ($p \leq r$). In order to assess quantitatively the success of this particular scaling method, a single numerical summary comparing the d_{ij} with δ_{ij} is inadequate; it is necessary to compare the two configurations themselves (Gower, 1971; Sibson et al, 1981). This can be done by Procrustes analysis (Gower, 1971; Sibson, 1978): the recovered configuration is translated, rotated and reflected so that it best fits the original configuration and the goodness-of-fit is then assessed by the sum of squared distances M^2 between corresponding points of the two configurations. This comparison could be done in r dimensions by appending (r - p) columns of zero coordinates to the (p-dimensional) recovered configuration, or in p dimensions by

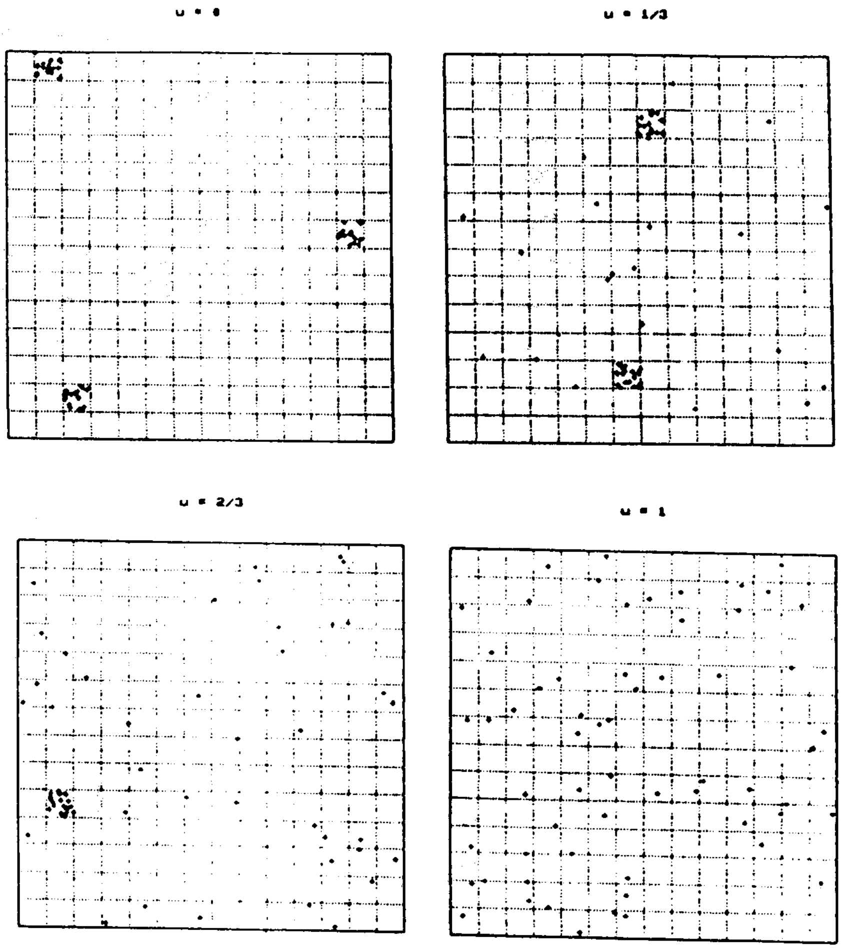

Figure 1:Examples of a strictly clustered configuration ($\mu = 0$), two mixed configurations ($\mu = 1/3$ and $\mu = 2/3$) and a random configuration ($\mu = 1$) in two-dimensional space. Number of points is 60, number of clusters 3, number of points per cluster 20.

taking the first p principal axes of the (r-dimensional) original configuration. The former method was adopted here, as it is in most practical applications.

This approach provides a natural way of comparing different methods of scaling across the different experimental conditions, providing that a suitable standardisation of data sets is applied so that values of M^2 can be compared across conditions. A simple normalisation is provided by rescaling each configuration so that its sum of squared coordinates is unity before the Procrustes analysis is conducted, and this was done in the present instance. Having ensured comparability of M^2 values, an approximate justification for analysing results via ANOVA has been provided by various authors (e.g. Gower, 1971).

3 Results

3.1 Output configurations

Each original r-dimensional configuration was reconstructed in p-dimensional subspaces for p = 1...r by each of three scaling methods. Thus, conditioning on r, 2 x 4 x 3 x 3 x r = 72 x r output configurations resulted for fixed r, yielding 720 output configurations in total. The multidimensional scalings were all conducted by using ALSCAL options within the SAS computer package, and subsequent Procrustes analysis were constructed using the ROTATE options within GENSTAT. OS failed to give an output configuration for 7 out of the 96 input configurations: 3 for dimension r = 3 and 4 for r = 4. Input configurations for 3 of these 7 degenerate solutions were strictly clustered, 3 were mixed and 1 was random. For all of them, however, reconstructed configurations in subspaces p = 1...r-1 existed. 2 out of 4 configurations for r = 4 (N = 60, μ = 0 and N = 30, μ = 0) gave perfect fit in 3 dimensions.

3.2 Analysis

For each true dimension r we had a 2 x 4 x 3 (i.e. N, μ, method) factorial experiment with 3 replications. As the output configurations resulted in subspaces of dimension p = 1...r, and the results for different p were correlated, a split-plot design with repeated measurement constraints was appropriate for the analysis. The Greenhouse and Geisser (1959) 'e-adjusted F-tests' were carried out in hypothesis testing of subplot interactions with dimension p (see also Kenward, 1981).

An analysis of variance table was obtained for each value of r using GENSTAT. Values of M^2 for degenerate solutions were treated as missing. Table 1 presents the analysis of variance of M^2 for r = 4.

Table 1: Analysis of variance of M^2 when r = 4

SOURCE OF VARIATION	DF	SS	SS%	MS	F
N	1	0.045243	0.22	0.045243	5.558
METHOD	2	0.052074	0.25	0.026037	3.199
CONFIG	3	3.470308	16.94	1.156769	142.105
N.METHOD	2	0.001459	0.01	0.000729	0.090
N.CONFIG	3	0.041949	0.20	0.013983	1.718
METHOD.CONFIG	6	0.069430	0.34	0.011572	1.422
N.METHOD.CONFIG	6	0.001302	0.01	0.000217	0.027
RESIDUAL1	48	0.390732	1.91	0.008140	
TOTAL1	71	4.072495	19.88	0.057359	
DIM	3	13.081988	63.86	4.360662	1632.188
N.DIM	3	0.032157	0.16	0.010719	4.012
METHOD.DIM	6	0.043029	0.21	0.007171	2.684
CONFIG.DIM	9	2.867780	14.00	0.318642	119.267
N.METHOD.DIM	6	0.000478	0.00	0.000080	0.030
N.CONFIG.DIM	9	0.071679	0.35	0.007964	2.981
METHOD.CONFIG.DIM	18	0.013124	0.06	0.000729	0.273
RESIDUAL2	158	0.422123	2.06	0.002672	
TOTAL2	212	16.53233	80.70	0.077983	
GRAND TOTAL	283	20.604828	100.58		

3.3 Results

The following features emerge from the analysis.

Main effects

The quality of recovery for 30 points is better than that for 60 points.

MRS performs significantly worse than MS or OS.

The more random the configuration, the worse is the recovery.

Interactions

OS and MS have the same profiles in p dimensions, for all p = 1...r: the quality of recovery worsens as μ increases. The more clustered is the configuration, the better are the results. For p = r, perfect fit is always obtained with MS, while for OS perfect fit or no fit are the two alternatives.

MRS reflects the same pattern for the subspaces of lowest dimensions, but as p increases the behaviour reverses: the more random the configuration the better are the results. It should be pointed out that in the subspace of true dimensionality very bad fit is obtained for strictly clustered and mixed configurations, while for random configurations the results of all three methods are very similar.

Figure 2 illustrates these interactions.

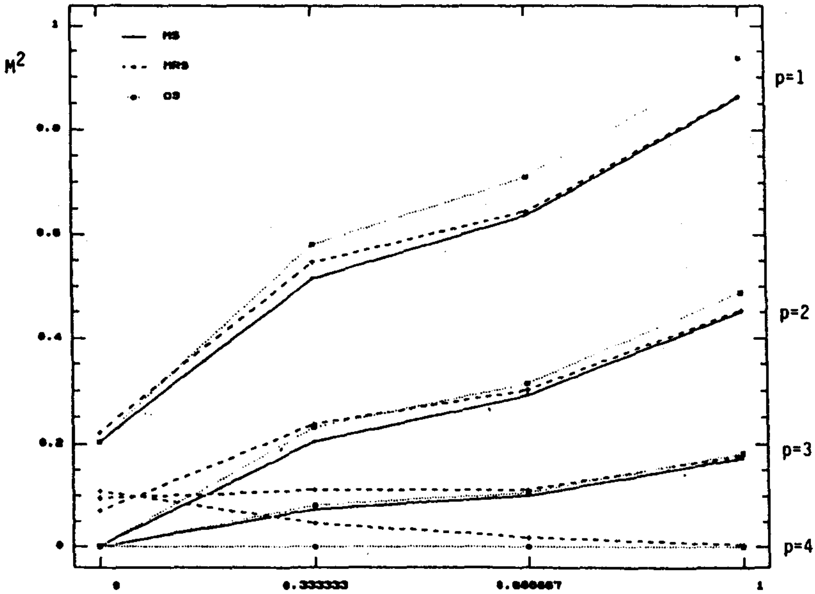

Figure 2: Mean values of M^2 plotted against μ, for $p = 1...4$ and for each of the scaling methods MS, MRS and OS. Missing values were eliminated in the calculation of the means.

4 Conclusions

Our study revealed that Weeks and Bentler's conclusions should be modified as follows: MRS inherits the advantages of MS and OS for random configurations, but the disadvantages of OS for strictly clustered and mixed configurations. The disadvantage of MRS becomes evident in the subspace of true dimensionality for strictly clustered and mixed configurations. The phenomenon is due to the fact that MRS inflates small differences in distances in clustered configurations by ranking them. Distances in random configurations are less affected by ranking, which explains the high quality of recovery.

References

[1] Chatfield C., and Collins A. J. (1980): *Introduction to multivariate analysis*. London: Chapman and Hall.

[2] Davidson M. L. (1983): *Multidimensional scaling*. New York: John Wiley.

[3] Gower J. C. (1966): Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, **53**, 325-338.

[4] Gower J. C. (1971): Statistical methods of comparing different multivariate analysis of the same data. In: Hodson, Kendall, Tautu (Eds.), *Mathematics in the archaeological and historical sciences*, 138-49. Edinburgh: University Press.

[5] Greenhouse S. W., and Geisser S. (1959): On methods in the analysis of profile data. *Psychometrika*, **24**, 95-112.

[6] Kenward M. G. (1981): *An investigation of certain methods for the analysis of repeated measurements*. PhD Thesis, University of Reading.

[7] Kruskal J. B. (1964): Multidimensional scaling by optimising goodness-of-fit to a nonmetric hypothesis. *Psychometrika*, **29**,1-27, 115-129.

[8] Sibson R. (1978): Studies in the robustness of multidimensional scaling. *J Royal Statistical Society Ser. B.*, **40**, 234-238.

[9] Sibson R., Bowyer A., and Osmond C. (1981): Studies in the robustness of multidimensional scaling: Euclidean models and simulation studies. *J Statist. Comput. Simul.*,**13**, 273-296.

[10] Torgerson W. S. (1958): *Theory and methods of scaling*. London: John Wiley.

[11] Weeks D. G., and Bentler P. M. (1979): A comparison of linear and monotone multidimensional scaling models. *Psychological Bulletin*, **86**, 349-354.