

May 1979

HOW TO USE SMACOF-III

A program for Metric Multidimensional Unfolding

Willem Heiser
Jan de Leeuw

DEPARTMENT OF DATA THEORY/FACULTY OF SOCIAL SCIENCES
UNIVERSITY OF LEYDEN-THE NETHERLANDS

SMACOF - III

Metric Multidimensional Unfolding.

1.0. Introduction.

In multidimensional unfolding (MDU) problems we consider a nonnegative datamatrix Δ of order $n \times m$, whose elements are interpreted as measures of dissimilarity between the n row-objects $\underline{R} = \{ r_1, r_2, \dots, r_i, \dots, r_n \}$ and the m column-objects $\underline{C} = \{ c_1, c_2, \dots, c_j, \dots, c_m \}$. Thus an element δ_{ij} of Δ gives the dissimilarity between objects r_i and c_j . In a psychological context the row-objects are often called subjects and the column-objects stimuli, with dissimilarities derived from preference judgements. Other possibilities are: rating scores of n concepts on m attributes, some transformation of response probabilities to n stimuli for m responses, or "pick any out of m " data from n subjects.

We want to represent both row- and column-objects as points in p -dimensional euclidean space. Thus, a representation of \underline{R} is the $n \times p$ configuration matrix X , with real elements x_{ia} , a representation of \underline{C} is the $m \times p$ configuration matrix Y , with real elements y_{ja} , and the interpoint distances $d_{ij}(X,Y)$ should be approximately equal to the interobject dissimilarities δ_{ij} . The euclidean distances are defined on the rows of X and Y by

$$d_{ij}(X,Y) = \left[\sum_{a=1}^p (x_{ia} - y_{ja})^2 \right]^{1/2} . \quad (1)$$

and to evaluate the badness-of-fit of a particular pair of configurations X,Y we use the loss function

$$\sigma(X,Y) = \sum_{i=1}^n \sum_{j=1}^m (\delta_{ij} - d_{ij}(X,Y))^2 . \quad (2)$$

We will consider two types of algorithms to solve the metric unfolding problem: one based on an algebraic analysis of (1), the other derived from the general SMACOF algorithm-model (cf. de Leeuw and Heiser, 1979), which tries to minimize (2) and takes the algebraic solution as a start. In both cases, we will sometimes utilize the additional requirement that each

column point should be equal to the centroid of certain specific row points; i.e., we require

$$Y = M^{-1}E'X, \quad (3)$$

where E is a $n \times m$ indicator matrix with elements ϵ_{ij} equal to zero or one and M a $m \times m$ diagonal matrix containing the marginal frequencies of E (i.e., the diagonal entries of M are $e'E$ where e is the unit vector). Although in general we may want to specify E independent of Δ (as in the case of an experiment where we conceive the column objects a priori as combinations of certain row objects), in the SMACOF - III program we derive E and M from Δ in the following way: for each column j we define

$$\epsilon_{ij} = 1 \quad \text{if} \quad \delta_{ij} = \min_k \{ \delta_{kj} \} \quad (4a)$$

$$\epsilon_{ij} = 0 \quad \text{if} \quad \delta_{ij} > \min_k \{ \delta_{kj} \} \quad (4b)$$

Constructed like this, E is called the matrix of first choices. On request of the user, the process of finding a minimum value in each column of Δ may be repeated discarding the earlier minimum, which would give us the matrix of first and second choices, and so on.

As will be shown in later sections, utilizing restrictions on the joint configuration does help a lot in avoiding "degenerate" solutions and local minima. Moreover, because it formalizes the idea that stimulus points should be "in between" the subjects who have chosen them first, the restricted metric unfolding problem has theoretical interest on its own.

1.1. Related work.

The one-dimensional unfolding model was proposed by Coombs (1950, 1964) as a consequence of his theory of preferential choice; in his terminology, the one-dimensional configuration Y is called the J-scale and each row of Δ an I-scale, which may be thought of as the J-scale folded at the ideal point x_i , with only the rank order of the stimuli given in order of increasing distance from the ideal point. The extension to multidimensional unfolding was made possible by the work of Bennett and Hays (1960, 1961; also see Coombs 1964). These older approaches are non-metric, not only in the sense that rankorders within rows of Δ are used, but they also end up with a represen-

tation consisting of p partial orders of projections on the axes of euclidean p -space.

The trick of specializing an iterative multidimensional scaling program such that it becomes an unfolding technique was developed by several people in the late sixties (cf. Kruskal and Carroll, 1969), but the problem of finding a good initial configuration for this special purpose seems to be a bit neglected. The algebraic approach goes back to Coombs and Kao (1960), who conjectured a connection with principal component analysis, Ross and Cliff (1964), who proved some theorems about this conjecture and Schönemann (1970), who proposed a method for recovering X and Y from the double centered squared distances (cf. Gold 1973). A critical review of these proposals can be found in Heiser and de Leeuw (1979).

2.0. Finding an initial configuration.

In this section we will discuss several approaches which try to fit the MDU model by what Kruskal (1977) has called "neclecting errors". They are based on finding equations that would correctly specify the desired solution in terms of the given data if there were no errors in the data. The first three of them directly start from the squares of the euclidean distances in (1); the next one incorporates the restrictions (3); the last two first estimate some or all missing inter-object dissimilarities utilizing the triangle inequality and then follow the Torgerson-Gower classical MDS approach.

In the case of errorless data, all of them are successfull; however, if the dissimilarities are partially unknown in the sense that only row-conditional rank numbers are known and/or corrupted by error, performance may differ a lot and some of them break down completely. An evaluation in terms of stress will be reported in section 3.3.

2.1. Neclecting errors; unrestricted model.

We suppose that the $n \times m$ matrix $D^{(2)} = \{ d_{ij}^2(X,Y) \}$ is given. According to the euclidean assumption,

$$\begin{aligned} d_{ij}^2(X,Y) &= \sum_{a=1}^p (x_{ia} - y_{ja})^2 \\ &= \sum_{a=1}^p x_{ia}^2 + \sum_{a=1}^p y_{ja}^2 - 2 \sum_{a=1}^p x_{ia} y_{ja} \end{aligned} \quad (5)$$

In matrix notation, this can be written as

$$D^{(2)} = e\beta' + \alpha e' - 2XY' \quad (6)$$

with e again the unit vector (its length will be obvious from the context), with the vector α defined by

$$\alpha_i = \sum_{a=1}^p x_{ia}^2, \quad (7a)$$

and a vector β defined by

$$\beta_j = \sum_{a=1}^p y_{ja}^2. \quad (7b)$$

In Heiser and de Leeuw (1979), three different ways to proceed are distinguished and discussed in great detail: no centering, single centering and double centering. We will call then here UNF1, UNF2 and UNF3, resp. Notice that (6) can be written as

$$D^{(2)} = \begin{bmatrix} \vdots & \vdots \\ e & \alpha & X \\ \vdots & \vdots \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} \beta' \\ \dots \dots \dots \\ e' \\ \dots \dots \dots \\ -2Y' \end{bmatrix} \quad (8)$$

In UNF1, we suppose that $n \geq m \geq p+2$ and $\text{rank}(D^{(2)}) = p+2$. A singular value decomposition of $D^{(2)}$ will give us, say

$$D^{(2)} = GH' \quad (9)$$

with $G'G = I$ and $H'H = \Lambda^2$, where Λ is the $(p+2) \times (p+2)$ diagonal matrix of singular values. It is well known that such a decomposition is unique up to a nonsingular transformation and we may employ this fact to find the special structure in (8). What we are looking for is a non-singular $(p+2) \times (p+2)$ matrix T such that

$$GT = \begin{bmatrix} \vdots & \vdots \\ e & \alpha & X \\ \vdots & \vdots \\ \vdots & \vdots \end{bmatrix} \quad (10)$$

and, correspondingly, the inverse transformation $S = (T^{-1})'$, such that

$$HS = \begin{bmatrix} \vdots & \vdots & \vdots \\ \beta & e & -2Y \\ \vdots & \vdots & \vdots \end{bmatrix} \quad (11)$$

The procedure for finding T and S is quite involved and it is difficult to ensure robust performance. The analysis may be simplified by removing the column means (and ignoring them altogether). Thus, in UNF2 we define

$$\dot{D}^{(2)} = J D^{(2)}, \quad (12)$$

where J is the centering operator

$$J = (I - \frac{ee'}{e'e}). \quad (13)$$

Now the transformed model equation becomes

$$\dot{D}^{(2)} = \dot{\alpha} e' - 2XY', \quad (14)$$

where $\dot{\alpha}$ and \dot{X} are centered versions of α and X , respectively. Again suppose

$$\dot{D}^{(2)} = GH' \quad (15)$$

as before, but now our task is to find a non-singular $(p+1) \times (p+1)$ matrix T such that

$$GT = \begin{bmatrix} \vdots & \vdots & \vdots \\ \alpha & X \\ \vdots & \vdots & \vdots \end{bmatrix} \quad (16)$$

and after that to find the inverse transformation. It is clear that "single centering" makes the MDU problem more simple, but it should be noted that

we use less information from the data. In the third approach, closely related to Schönemann (1970), we define

$$\ddot{D}^{(2)} = -\frac{1}{2}JD^{(2)}J, \quad (17)$$

which renders the transformed model equation

$$\ddot{D}^{(2)} = (X - e\mu')(Y - e\nu')', \quad (18)$$

where μ and ν are the centroids of X and Y , respectively. The non-uniqueness of the singular value decomposition

$$\ddot{D}^{(2)} = GH' \quad (19)$$

now amounts to finding a T , S , μ and ν such that

$$X = GT + e\mu', \quad (20a)$$

$$Y = HS + e\nu', \quad (20b)$$

$$TS' = I. \quad (20c)$$

To estimate these unknowns, a matrix F is defined as

$$F = D^{(2)} + 2GH', \quad (21)$$

which represents the (approximately) reconstructed row- and column means of $D^{(2)}$ and, in the perfect case, should be of the form

$$F = \phi e' + e\psi' + \chi ee', \quad (22)$$

where the unknown parameters are packed in the vectors ϕ and ψ (cf. Heiser and de Leeuw 1979), and can be estimated by standard methods.

2.2. Neglecting errors; restricted model.

We now want to incorporate the restrictions (3) into the model. Thus, we

assume

$$Y = M^{-1}E'X \quad (23a)$$

and require without loosing generality

$$JY = Y, \quad (23b)$$

i.e., the configuration of column points is centered. Furthermore, observe that the way in which E and M are constructed implies

$$M^{-1}E'e = e. \quad (23c)$$

As the first step in UNF⁴, we define

$$\bar{D}^{(2)} = M^{-1}E'D^{(2)}. \quad (24)$$

In terms of the parameters this implies, using (6) and (23):

$$\bar{D}^{(2)} = e\beta' + M^{-1}E'\alpha e' - 2YY'. \quad (25)$$

The next step is to define the double centered version $\bar{\bar{D}}^{(2)}$ of $\bar{D}^{(2)}$:

$$\bar{\bar{D}}^{(2)} = -\frac{1}{2}J\bar{D}^{(2)}J \quad (26)$$

and from (25) we get

$$\bar{\bar{D}}^{(2)} = YY'. \quad (27)$$

In the case of fallible data, $\bar{D}^{(2)}$ will not necessarily be symmetric and we use a truncated eigenvalue-eigenvector decomposition of $\frac{1}{2}(\bar{\bar{D}}^{(2)} + \bar{D}^{(2)})$ as the least squares estimate \hat{Y} of Y. The last step in UNF⁴ is to compute $\hat{\beta}$ from (7b) and the least squares estimate \hat{X} of X by

$$\hat{X} = -\frac{1}{2}(D^{(2)} - e\hat{\beta}')\hat{Y}(\hat{Y}'\hat{Y})^{-1}, \quad (28)$$

using (23b) and (6) again.

2.3. Utilizing the triangle inequality.

If we were able to find reasonable estimates of the "missing" inter-column-object and/or inter-row-object distances, we could use the classical Torgerson-Gower MDS approach for finding estimates of the coordinate values. Now, observe that the triangle inequality requires that the distance between any two column objects j and k satisfies

$$|d_{ij} - d_{ik}| \leq d_{jk} \leq d_{ij} + d_{ik} \quad (29)$$

for all row objects i . Consequently, d_{jk} must be somewhere in the intersection of these intervals:

$$\max_i |d_{ij} - d_{ik}| \leq d_{jk} \leq \min_i (d_{ij} + d_{ik}). \quad (30)$$

In fact, one of these inequalities becomes an equality if there is at least one row object located on the line passing through points j and k . As a reasonable estimate of the inter-column-object distance we may use:

$$\delta_{jk} = \frac{1}{2} \{ \max_i |\delta_{ij} - \delta_{ik}| + \min_i (\delta_{ij} + \delta_{ik}) \}, \quad (31)$$

the midpoint of the least upper bound and the greatest lower bound, derived from the data.

There are two possibilities now to proceed. In UNF5, (31) is used to estimate both inter-row-object and inter-column-object distances and the Torgerson-Gower procedure is used on the supplemented $(n+m) \times (n+m)$ symmetric dissimilarity matrix. In UNF6, (31) is used to estimate the inter-column-object distances only, then \hat{Y} is found with the T-G procedure ($m \times m$ problem) and finally \hat{X} is found using (28). Especially in large problems, this is a lot more efficient than UNF5.

3.0. SMACOF - III algorithm.

The SMACOF - III algorithm to minimize (2) is derived from the general algorithm-model described in de Leeuw and Heiser (1979), who also indicate a general way to handle restrictions on the configuration. Here, we only give the update formula's used in the program (for details of the derivation, see Heiser and de Leeuw 1979). Suppose X^μ and Y^μ are our current best solutions. First we compute

$$\dot{x}_{ia}^{(\mu)} = \left\{ \sum_{l=1}^m \frac{\delta_{il}}{d_{il}} \right\} x_{ia}^{(\mu)} - \sum_{j=1}^m \frac{\delta_{ij}}{d_{ij}} y_{ja}^{(\mu)} \quad (32a)$$

$$\dot{y}_{ja}^{(\mu)} = \left\{ \sum_{l=1}^n \frac{\delta_{lj}}{d_{lj}} \right\} y_{ja}^{(\mu)} - \sum_{i=1}^n \frac{\delta_{ij}}{d_{ij}} x_{ia}^{(\mu)} \quad (32b)$$

Then, the Guttman-transforms are

$$\bar{x}_{ia}^{(\mu)} = \frac{1}{m} (\dot{x}_{ia}^{(\mu)} - \frac{1}{n+m} \sum_{i=1}^n \dot{x}_{ia}^{(\mu)}) \quad (33a)$$

$$\bar{y}_{ja}^{(\mu)} = \frac{1}{n} (\dot{y}_{ja}^{(\mu)} - \frac{1}{n+m} \sum_{j=1}^m \dot{y}_{ja}^{(\mu)}) \quad (33b)$$

In the unrestricted algorithm, we simply set

$$X^{\mu+1} = \bar{X}^\mu \quad (34a)$$

$$Y^{\mu+1} = \bar{Y}^\mu \quad (34b)$$

and continue these iterations until the improvement of successive stress values is no longer appreciable.

3.1. Fitting the restrictions.

In each iteration, we have to solve the subproblem

$$\min_X \text{tr}(\bar{Z}^\mu - JHX)'V(\bar{Z}^\mu - JHX) \quad (35)$$

where we have used

$$\bar{Z}^{\mu} = \begin{bmatrix} \bar{X}^{\mu} \\ \dots \\ \bar{Y}^{\mu} \end{bmatrix} \quad (36a)$$

$$H = \begin{bmatrix} I \\ \dots \\ M^{-1}E' \end{bmatrix} \quad (36b)$$

$$V = \begin{bmatrix} mI & \vdots & -ee' \\ \dots & \ddots & \dots \\ -ee' & \vdots & nI \end{bmatrix} \quad (36c)$$

and J the centering operator. This is a least squares metric projection problem; its solution is

$$\hat{X}^{\mu} = (H'VH)^+ H'V\bar{Z}^{\mu} \quad (37)$$

and the updates satisfying the restrictions are

$$Z^{\mu+1} = JH\hat{X}^{\mu} \quad (38)$$

In actual computations, we utilize several simplifications implied by the partitionings in (36). Thus, for example,

$$(H'VH)^+ = \frac{1}{m} (J - C(I + C'C)^{-1}C') \quad (39)$$

with

$$C = \left(\frac{n}{m}\right)^{\frac{1}{2}} JEM^{-1} \quad (40)$$

3.2. Relaxation.

Very slow linear convergence is typical for MDS algorithms, particularly in the unfolding case. Following the line of reasoning developed in de Leeuw and Heiser (1979), it turns out that with the relaxed updates

$$X^{\mu+1} = 2\bar{X}^{\mu} - X^{\mu} \quad (41a)$$

$$Y^{\mu+1} = 2\bar{Y}^{\mu} - Y^{\mu} \quad (41b)$$

we attain convergence in approximately half the number of iterations without substantially larger costs.

3.3. Evaluation of different initialization techniques.

To evaluate the various approaches described in section 2 in terms of both the initial and the final stress they yield in the SMACOF-III program, we analysed the following 5 sets of data:

POWER: 17x8 problem, data from Gold (1958)

RANKO: 16x6 problem, rows being 16 permissible rankorders under the one-dimensional nonmetric unfolding model

IDEAL: 20x7 problem, data generated as distances between 27 known points

DIRIG: 48x13 problem, data from v.d. Geer (unpublished), being 48 rank-orders of 13 famous conductors

BINAR: 20x7 problem, distances from IDEAL dichotomized below and above their mean

In table 1, results are summarized for UNF1 - UNF6, which are APL programs implementing the corresponding techniques from section 2, and CHO1 - CHO3, which are FORTRAN implementations of UNF4 included in SMACOF-III. In CHO1, E is the matrix of first choices; in CHO2, second choices are added; in CHO3, third choices are added too. In UNF4, sometimes slightly different criteria are used to define E; moreover, only (27) and (28) are used, whereas in CHO1 - CHO3 the initial stress is computed after one projection (37) and (38) has been performed. All analyses were done with standard options in two dimensions (see note 1).

From table 1 it is clear that CHO1 outperforms all the others, with UNF5 also doing reasonably well (but this is a very expensive alternative). Note that in POWER UNF3 has a very good start, but it ends up in a local minimum (UNF4, UNF6, CHO2 and CHO3 arrive at the same configuration, up to rotations) with stress 0.0323. The final solutions of CHO1 and UNF5 are also nearly identical, with better stress (0.0300). RANKO only differentiates UNF1, UNF2 and UNF3 from the others; IDEAL is just for reassurance; DIRIG differentiates CHO1, CHO2 and CHO3 from the others. Finally, BINAR is intended to

	POWER	RANKO	IDEAL	DIRIG	BINAR
UNF1	.240675 /58/ .035540	a)	.000007 / 1/ .000004	b)	.461769 /18/ .186396
UNF2	.231916 /25/ .038786	a)	.000010 / 1/ .000004	b)	.422152 /17/ .208834
UNF3	.082794 /19/ .032328	a)	.000011 / 1/ .000006	b)	a)
UNF4	.203861 /69/ .032275	.043909 /13/ .006808	.059109 /12/ .030018	.288467 /83/ .079415	.136320 /43/ .025604
UNF5	.084674 /42/ .030425	.038558 /10/ .006719	.036276 / 9/ .000023	b)	.151032 /93/ .026887
UNF6	.186304 /72/ .032273	.056906 /18/ .006412	.046220 /10/ .000025	.274679 /66/ .088970	.198750 /99/ .037840
CHO1	.201425 /32/ .029936	.024774 /10/ .007009	.027755 /10/ .000021	.204352 /65/ .075408	.078762 /28/ .025576
CHO2	.205833 /28/ .032261	.027409 /10/ .006737	.018748 /10/ .000025	.229691 /53/ .079560	c)
CHO3	.177391 /65/ .032254	.043606 /15/ .006812	.027414 / 9/ .000013	.253044 /48/ .077647	c)

table 1. Initial stress values, number of iterations and final stress values for 5 problems with 9 different initial configurations. a) initial routine breaks down. b) problem too big under APL. c) does not apply.

be indicative for performance on "pick any out of m" data or "yes-no" responses (we know in this case that a reasonable fit is possible, because the data are a coarse but monotone transformation of the distances in IDEAL). It turns out that UNF1 and UNF2 are far away from a desirable solution, UNF3 cannot handle the situation, and UNF4, UNF5 but especially CHO1 perform well. In conclusion, then, CHO1 seems to be the most favourable initialization technique (and consequently has been chosen as the default option in the SMACOF-III program).

4.0. Program description.

The algorithm is implemented in the ANS-FORTRAN-IV program SMACOF-III. The logical steps of the program can be summarized as follows:

- start: read problem parameters; compute total array area and relative addresses; read and normalize datamatrix; set the dimensionality p equal to p_{\max} ; set the iteration counter μ equal to 0; compute or read initial configurations X_p^μ and Y_p^μ .
- step1: except in case that X_p^0 and Y_p^0 are provided by the user, apply restrictions $Y_p^0 = M^{-1} E' X_p^0$.
- step2: compute $d_{ij}(X_p^\mu, Y_p^\mu)$ and $\sigma(X_p^\mu, Y_p^\mu)$.
- step3: compute the Guttman transforms \bar{X}_p^μ and \bar{Y}_p^μ .
- step4: if requested, project \bar{X}_p^μ and \bar{Y}_p^μ on restriction space.
- step5: set $X_p^{\mu+1}$ and $Y_p^{\mu+1}$ equal to result of step3 or step4; or, if requested and $\mu > 5$, compute relaxed updates.

step6: compute $d_{ij}(X_p^{\mu+1}, Y_p^{\mu+1})$ and $\sigma(X_p^{\mu+1}, Y_p^{\mu+1})$;
 if $\sigma(X_p^{\mu}, Y_p^{\mu}) - \sigma(X_p^{\mu+1}, Y_p^{\mu+1}) < \text{CRIT}$ go to step7;
 else: increase μ by 1 and go to step3.
 step7: rotate $X_p^{\mu+1}$ and $Y_p^{\mu+1}$ to their joint principal axes;
 compute stress values for each row; take care of output options;
 decrease p by 1; if $p > p_{\min}$, set μ equal to 0 and go to step2;
 else: stop

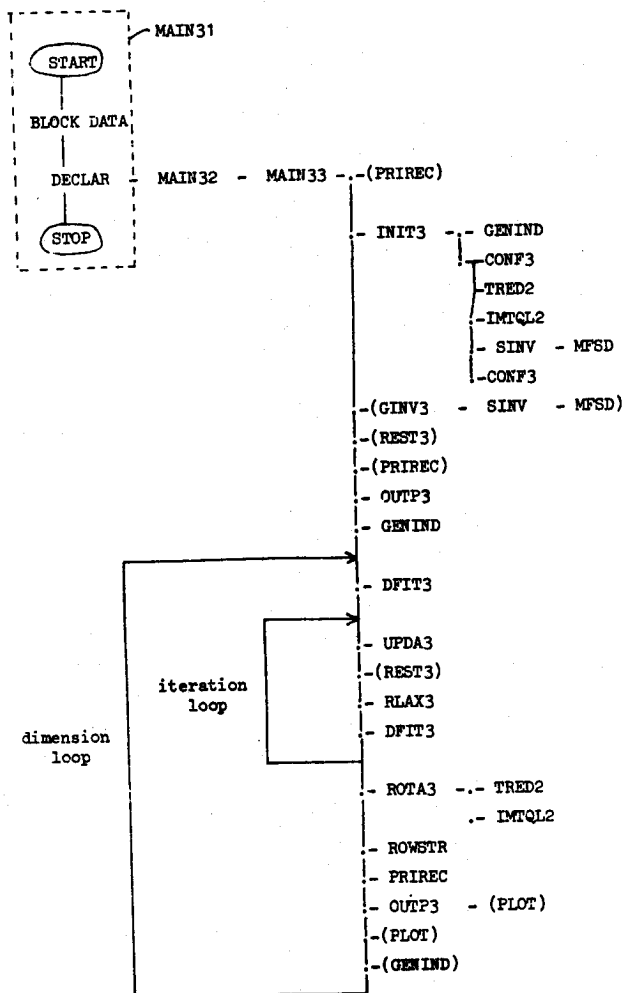


figure 1. Structure of the SMACOF-III program.

In terms of its subroutines, the program is structured as indicated in figure 1. Here successive subroutine calls are given from top to bottom, whereas further calls into a deeper level are pictured horizontally; mutually exclusive calls are indicated by forks. As many computational details are commented upon in the source program text, we only give a rough picture of the main features here. The subroutines DECLAR and MAIN32 compute the total array area and the relative addresses of all arrays, resp. If a genuine dynamic storage allocation facility is wanted, replace DECLAR with a machine compatible assembler routine. Subroutine MAIN33 controls the flow of the program. First, INIT3 is called which generates the matrix of first (second etc.) choices by calling GENIND and reads or computes initial configurations. Then output options are taken care of and step2 is performed by DFIT3; UPDA3, REST3, RLAX3 and DFIT3 take care of steps 3 to 6 and the remaining subroutines perform step7. The total number of source statements, excluding comments, is 1256.

Note 1. The stress-values reported in table 1 are computed according to formula (2). The actual program output will give the square root of these values, following a customary convention in the MDS literature.

Note 2. The row-stress-values R_i which are given on output are scaled such that they are related to the final stress value S as

$$S = \sqrt{\frac{1}{n+m} \sum_{i=1}^n R_i} .$$

5.0. SMACOF-III examples.

In this section we present results obtained with the SMACOF-III program for a wide variety of problems. All analyses were done in two dimensions and all initial configurations were computed with CHO1. We first give full sample output of a small social-psychological problem (POWER), then illustrate the effect of using restrictions in a bigger, politicological one (PARL68). Next, we will reanalyse data concerning preferences for family composition (Delbeke 1978) and attempt to assess the stability of the solution using a "bootstrap" technique (cf. Efron 1979). Finally, the unfolding approach to "seriation" or "sequencing" problems (arising frequently in the archeological and historical sciences) is illustrated with the well-known Münsingen-Rain data (Hodson 1968).

5.1. Power in the classroom.

Our first example concerns the evaluation of power characteristics according to eight different groups of middle-class american children (data reproduced in table 2, taken from Gold 1958). The 17 row-objects

	A	B	C	D	E	F	G	H
1.Smart at school	13.5	13	17	15	16	17	17	16
2.Good ideas how to have fun	1	17	13	6	6	10	9	4
3.Good at making things	13.5	6.5	12	15	13	12.5	15	14
4.Good at games with running and throwing	16.5	3	14	17	17	11	13	13
5.Knows how to fight	12	4	11	15	14	15	16	12
6.Strong	9.5	13	15	13	12	16	14	15
7.Acts friendly	2	15.5	3	3	3	5	2	2
8.A good person to do things with	9.5	1	4	11	9	6	6	9
9.Asks you to do things in a nice way	5.5	5	1	4	4	2	1	5
10.Doesn't start fights and doesn't tease	5.5	11	7.5	1	7	1	4	7
11.Knows how to act so people will like him	15	13	5	2	2	8	5	3
12.Plays with you a lot	3	8.5	9	10	8	9	11	6
13.Likes to do the same things you like to do	5.5	6.5	10	5	1	7	8	1
14.Nice looking	11	10	7.5	12	15	14	10	17
15.Has things you'd like to have	16.5	15.5	16	7	10	12.5	12	10
16.Gives you things	8	8.5	6	9	11	3	7	11
17.Does things for you	5.5	2	2	8	5	4	3	8

Table 2. Ranks of items by per cent of times they were rated "very important"; low value: most important. Taken from Gold (1958).

represent possible properties of children which, when valued highly in a particular group, supposedly contribute to the social power of children which possess them. The 8 column objects represent the eight groups of children, varying in age, sex and the instruction given. Full sample output of the SMACOF-III program is given on the next pages.

SMACOF - III

VERSION - 3

JULY 1974

METRIC MULTIDIMENSIONAL UNFOLDING

BILLEN MEISER

JAN DE LIGGHE

DEPT. OF DATA THEORY

THE UNIVERSITY OF LIEGEN

JOB-TITLE1-***** POWER IN THE CLASSROOM *****

PCRD OF PARAMETER CARDSET

17 8 2 2 0 0 8 88.0
1 2 2 1 1 0 9 7

DATA-SPECIFICATIONS:

THE NUMBER OF ROW POINTS IS 17
THE NUMBER OF COLUMN POINTS IS 8

ANALYSIS-SPECIFICATIONS:

THE MAXIMUM NUMBER OF DIMENSIONS IS 2
THE MINIMUM NUMBER OF DIMENSIONS IS 2
THE MAXIMUM NUMBER OF ITERATIONS IS 50
THE NUMBER OF DIFFERENT CHOICE VALUES IS 1
THE CONVERGENCE CRITERION IS 0.10E-04
RELAXED UPDATES ARE USED

DATA MATRIX (READ FROM UNIT NR 9 WITH FORMAT (8F6.2))

	1	2	3	4	5	6	7	8
1	13.500	13.000	17.000	15.000	16.000	17.000	17.000	16.000
2	11.000	17.000	13.000	6.000	6.000	12.500	9.000	4.000
3	13.500	6.500	12.000	15.000	13.000	12.500	15.000	14.000
4	16.500	3.000	14.000	17.000	17.000	11.000	13.000	13.000
5	12.000	4.000	11.000	15.000	14.000	15.000	16.000	12.000
6	9.500	13.000	15.000	13.000	13.000	14.000	14.000	15.000
7	2.000	15.500	3.000	3.000	3.000	5.000	2.000	2.000
8	9.500	1.000	4.000	11.000	9.000	6.000	6.000	9.000
9	5.500	5.000	1.000	4.000	4.000	2.000	1.000	5.000
10	6.500	11.000	7.000	1.000	2.000	4.000	4.000	7.000
11	15.000	13.000	5.000	2.000	2.000	8.000	5.000	3.000
12	3.000	8.500	4.000	10.000	8.000	9.000	11.000	6.000
13	5.500	6.500	10.000	5.000	1.000	7.000	8.000	1.000
14	11.000	10.000	7.500	12.000	13.000	14.000	10.000	17.000
15	16.500	15.500	19.000	7.000	10.000	12.500	12.000	10.000
16	8.000	8.500	6.000	9.000	11.000	3.000	7.000	11.000
17	5.500	2.000	2.000	8.000	5.000	4.000	3.000	8.000

INITIAL CONFIGURATION ***** POWER IN THE CLASSROOM *****

Row Potentials

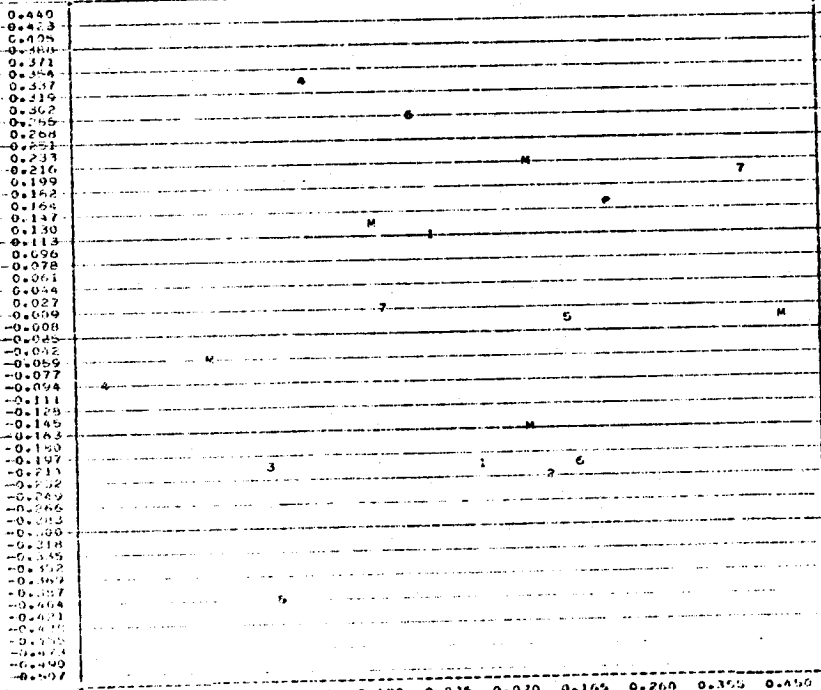
	1	2
1	0.037	-0.215
2	0.450	-0.010
3	-0.266	-0.207
4	-0.409	-0.098
5	-0.240	-0.312
6	0.172	-0.259
7	0.405	0.202
8	-0.339	-0.050
9	-0.113	0.132
10	-0.103	0.223
11	-0.078	0.155
12	0.126	-0.231
13	0.099	-0.165
14	-0.203	0.363
15	0.161	-0.054
16	-0.053	-0.280
17	-0.004	0.010

COLUMN POINTS

	1	2
1	0.450	-0.010
2	-0.339	-0.056
3	0.103	0.223
4	0.099	-0.155
5	0.103	0.223
6	0.126	-0.231
7	0.099	-0.165

2-DIMENSIONAL INITIAL SOLUTION

DIMENSION 1 (X-AXIS) VERSUS DIMENSION 2 (Y-AXIS)



-0.450 -0.350 -0.250 -0.150 -0.050 0.050 0.150 0.250 0.350 0.450

HISTORY OF COMPUTATION FOR ANALYSIS IN 2 DIMENSION(S)

***** POWER IN THE CLASSROOM *****

ITERATION NO.	STRESS	EPS
0	0.448901	0.448901
1	0.302034	0.146767
2	0.273328	0.028706
3	0.250323	0.023005
4	0.212916	0.037407
5	0.204391	0.008524
6	0.202189	0.002203
7	0.200548	0.001641
8	0.198875	0.001673
9	0.196436	0.002439
10	0.190391	0.006045
11	0.180388	0.010003
12	0.178090	0.002298
13	0.177243	0.000847
14	0.176686	0.000557
15	0.176279	0.000407
16	0.175906	0.000372
17	0.175596	0.000311
18	0.175293	0.000303
19	0.174949	0.000344
20	0.174603	0.000345
21	0.174228	0.000375
22	0.173853	0.000345
23	0.173644	0.000239
24	0.173452	0.000193
25	0.173321	0.000131
26	0.173229	0.000091
27	0.173162	0.000067
28	0.173120	0.000042
29	0.173088	0.000033
30	0.173055	0.000033
31	0.173025	0.000030
32	0.173014	0.000010
33	0.173000	0.000014
34	0.172964	0.000035

NO FURTHER IMPROVEMENT OF STRESS; FINAL VALUE: 0.172964

FINAL 2-DIMENSIONAL CONFIGURATION FOR ***** POWER IN THE CLASSROOM *****

17 POINTS

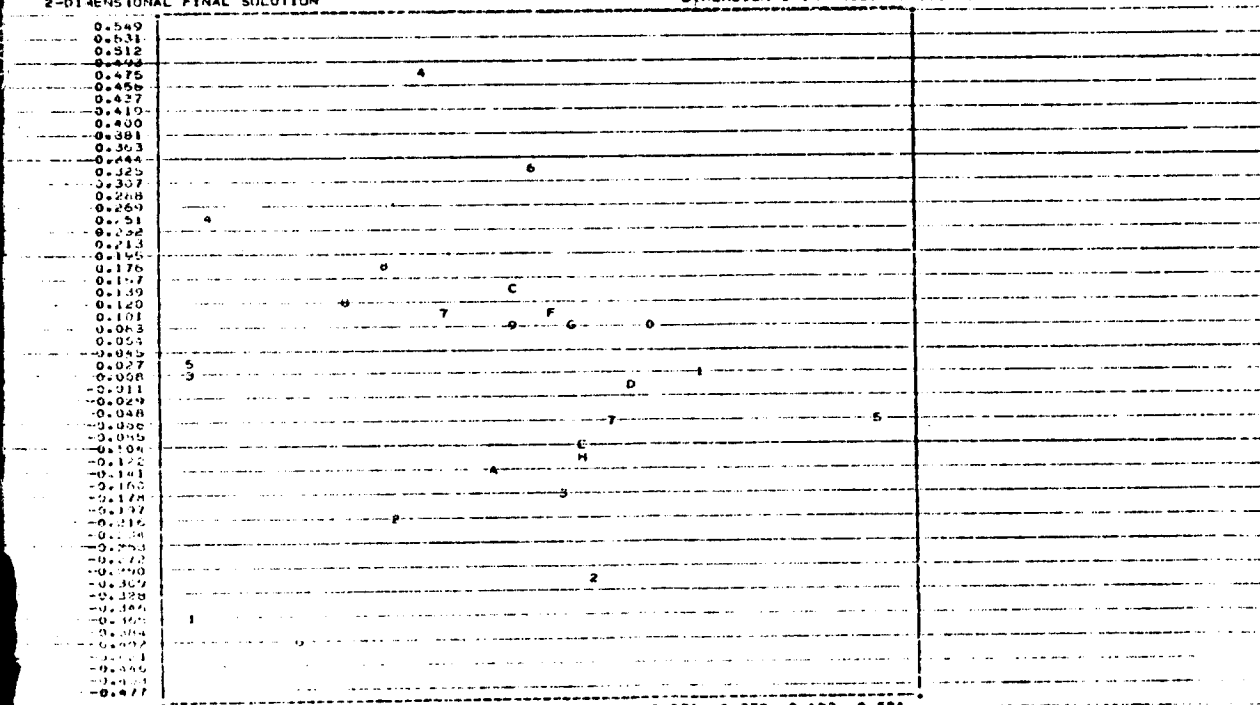
	DIMENSION 1	DIMENSION 2
1	0.0104003	-0.4401
2	0.0046667	-0.147
3	0.015036	-0.433
4	0.005100	-0.419
5	0.013312	-0.452
6	0.0050006	-0.278
7	0.004900	-0.202
8	0.005677	-0.145
9	0.005615	0.053
10	0.004387	0.059
11	0.0130268	0.334
12	0.010350	-0.123
13	0.002009	0.126
14	0.0071130	-0.005
15	0.0058671	0.504
16	0.017737	-0.073
17	0.017565	-0.056

COLUMN POINTS

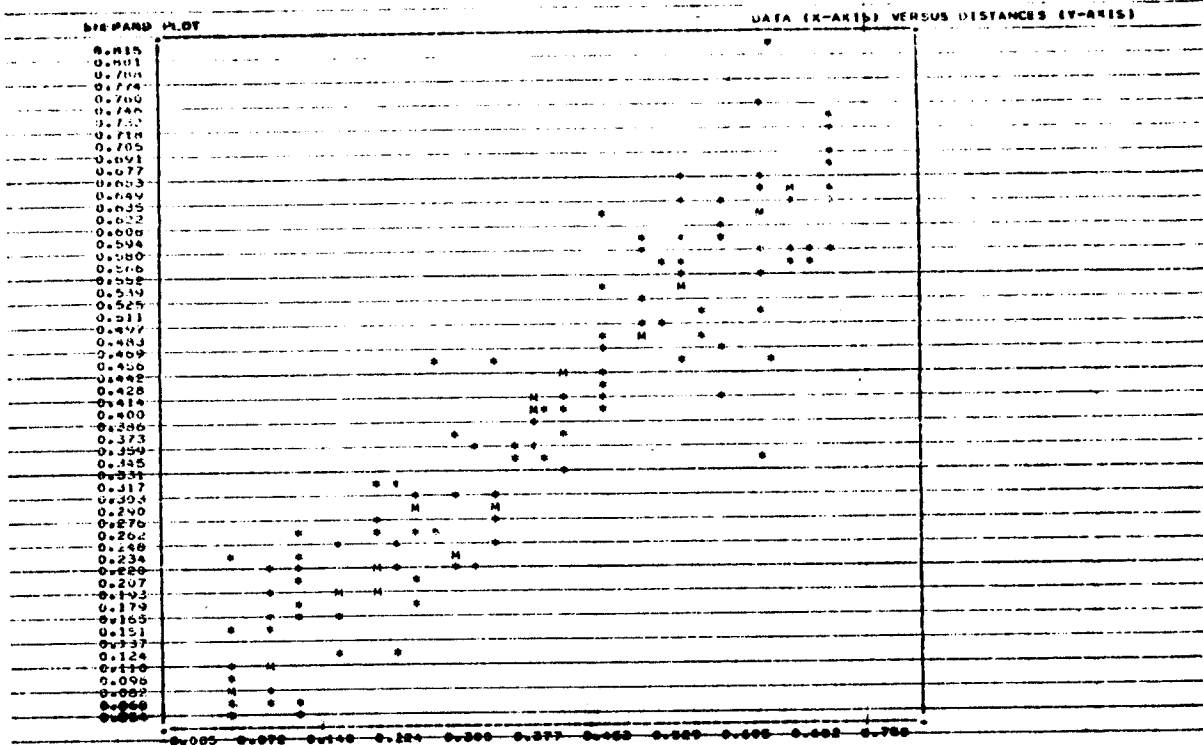
	1	2
1	0.021	-0.135
2	-0.207	0.127
3	-0.041	-0.136
4	0.220	-0.013
5	-0.149	-0.098
6	0.113	0.110
7	-0.129	-0.084
8	0.145	-0.116

2-DIMENSIONAL FINAL SOLUTION

DIMENSION 1 (X-AXIS) VERSUS DIMENSION 2 (Y-AXIS)



0.444 0.419 0.215 0.136 0.053 0.070 0.173 0.276 0.379 0.482 0.586



After heading, summary of options and print of the datamatrix we get the results of CH01, including a printplot of both row and column points (the first 9 integers and 0 are used for identifying the row points, so 11,21,31 get the same label as 1; likewise, points 2,12,22,32,... are labeled identically etc. The same labeling strategy is followed for the column points, this time using the characters A - J. When points coincide relative to the printing precision, the label M is used). Next, the history of computation and the final solution are given, including row-stress values (useful for identifying bad-fitting row objects such as 11), a printplot of row and column points and a "Shepard plot" of data values versus fitted distances, which might be inspected on derivation from linearity.

The solution presented here is different from those given in Heiser and de Leeuw (1979), and has lower final stress. Our experience with these data suggests, that points in the middle are quite stable, whereas points on the edge are not. The number of different local minima with about the same final stress is quite large.

5.2. Party preferences in the Dutch Parliament.

Our second example concerns the preferences of 141 members (out of 150) of Dutch Parliament for all 12 parties residing there at the time of the survey (1968). For earlier reports on these data, see Daalder and Rusk (1972), de Leeuw (1973) and Daalder and van de Geer (1977). In table 3, the names of the parties are given, along with their plottingcode and a short description. The number of respondents belonging to a particular party is

name	plotcode (col.obj.)	description	number of respondents	plotcode (row obj.)
CPN	CPN	communists	-	-
PSP	PSP	pacifists	4	P
PvdA	PVA	socialists	37	L
D'66	D66	democrats	7	6
PPR	PPR	radicals	-	-
KVP	KVP	catholics	42	K
ARP	ARP	anti-revolutionaries	15	A
CHU	CHU	christian-historicals	12	U
VVD	VVD	liberals	17	V
BP	BOP	peasants	4	B
SGP	SGP	dissident calvinists	2	S
GPV	GPV	dissident calvinist	1	G

table 3. Specifications of PARL68.

also listed, along with their plotting code. All subjects returned a complete rankorder, coded from 2 (own party) up to 13 (least preferred party). The result of the SMACOF-III analysis without imposing restrictions is plotted in figure 2 (stress = 0.1781). Basically, a curved left-right dimension is found, in particular for the row objects. The configuration of column objects is more differentiated and shows less 'degeneracy' than the Daalder and Rusk (non-metric) solution. Note that many respondents are rather far away from their own party (e.g. compare the V's with VVD). Figure 3 shows what

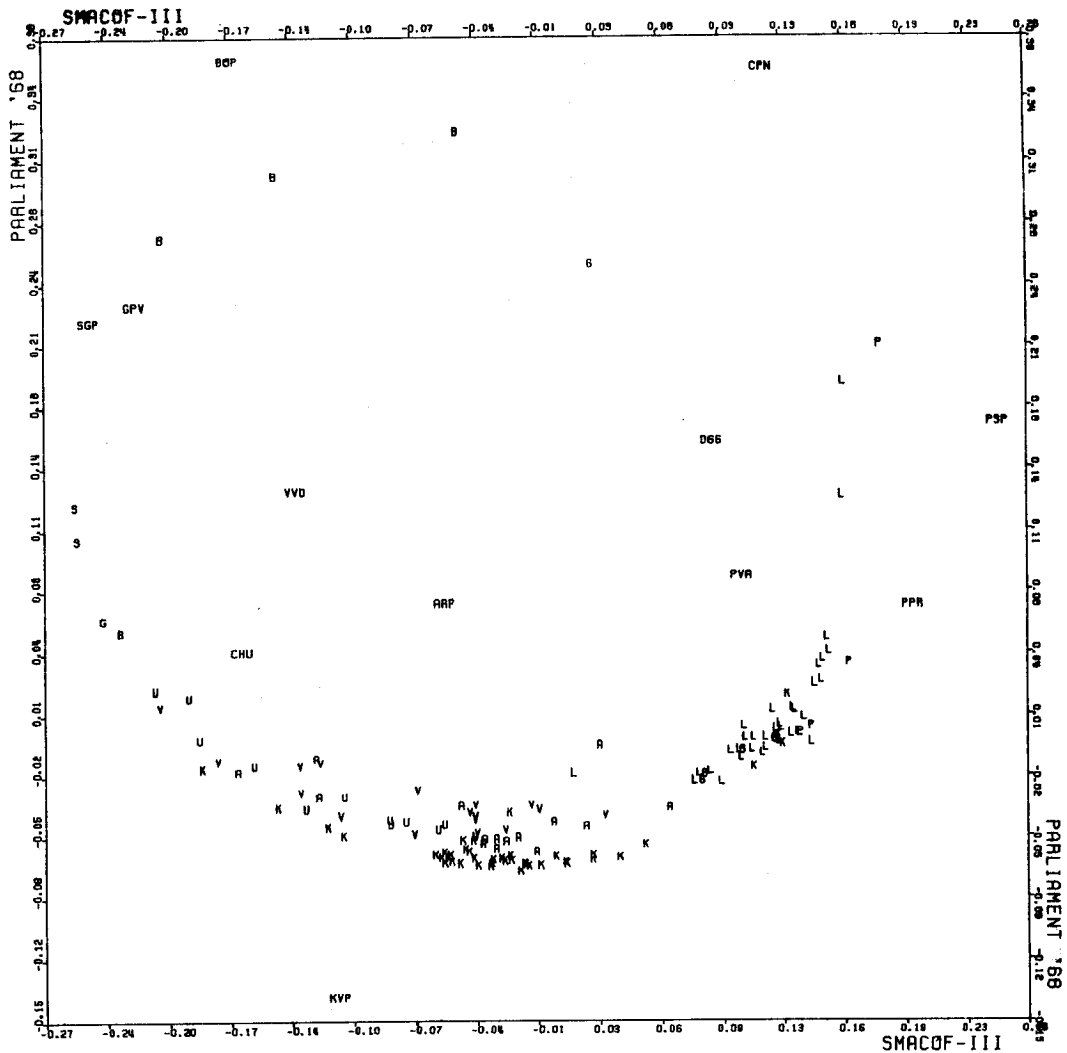


FIGURE 2. SHACOF-III; WITHOUT RESTRICTIONS; DUTCH PARLIAMENT 1968

happens if we use the first choices as restrictions during iteration. The stress (0.2276) is somewhat higher, not surprisingly, the party configuration hasn't changed very much, but the configuration of party-members has been spread in a more interesting way. The V's, for example, still are "right-wing", but differentiated in the second dimension from the christians (U, K, A) and some of them reach the 6's, which is consistent with politicological intuition.

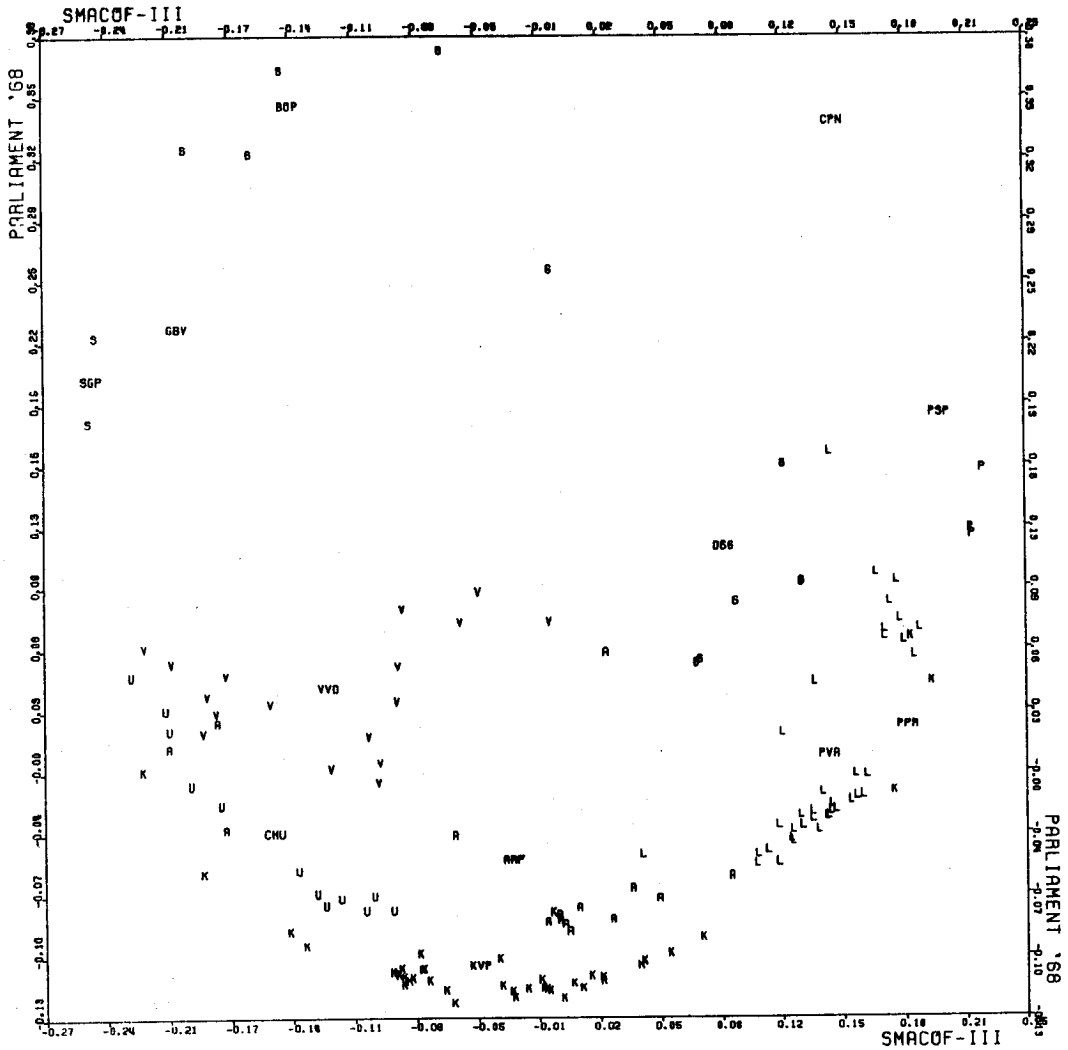


FIGURE 3. SMACOF-III; WITH RESTRICTIONS; DUTCH PARLIAMENT 1968

There is also a nice partition of L's, an interesting big spread of K's and A's, etc. The use of restrictions seems promising, it tends to emphasize good fit of the small distances and has a natural interpretation here ("each party is located in the middle of his own members").

5.3. Preferences for family composition.

We will now reanalyse a set of preferences kindly made available to us by Mr. Delbeke (who reported extensive analyses in Delbeke (1978)). The row

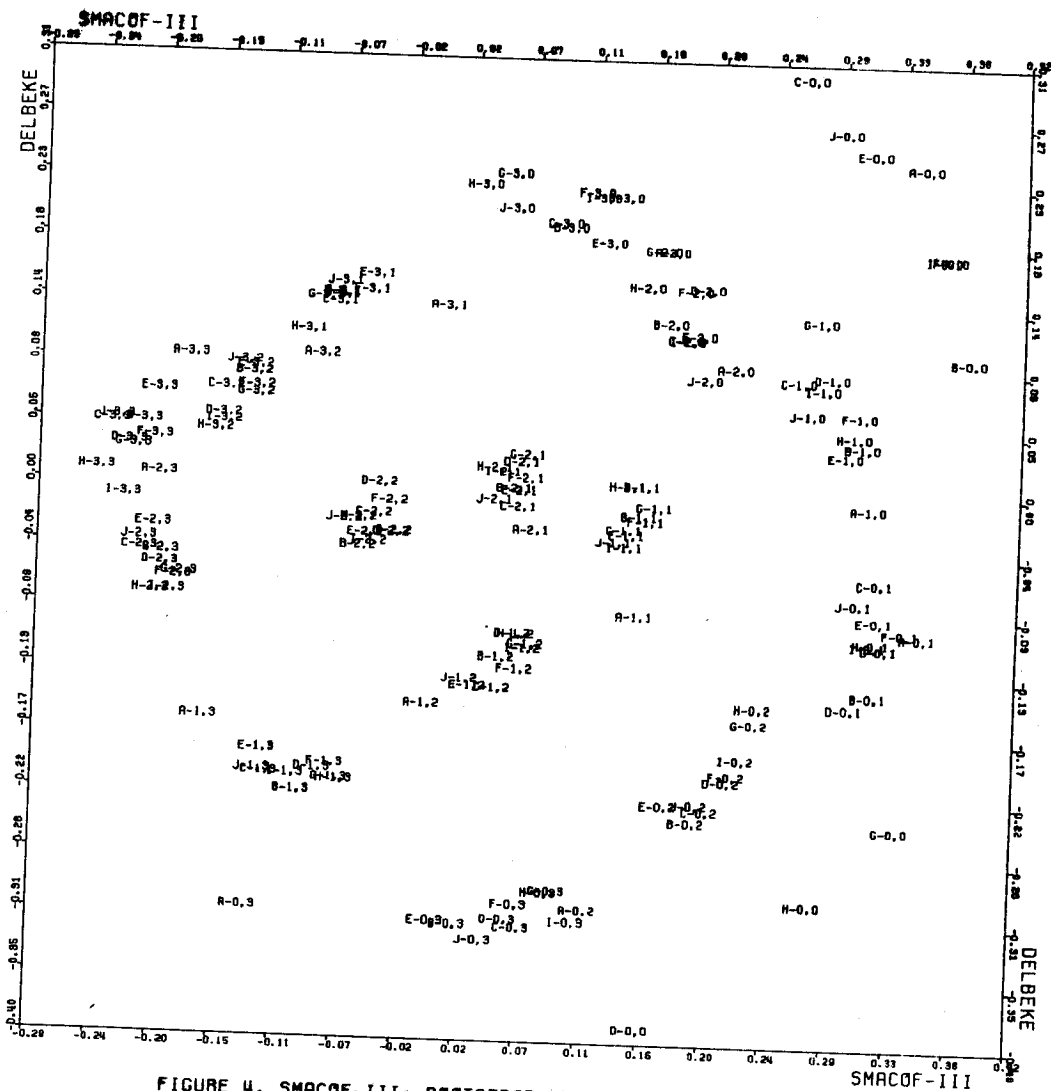


FIGURE 4. SMACOF-III; BOOTSTRAP DELBEKE DATA (10 SAMPLES)

objects are 82 psychology students at Leuven university, the column objects are 16 family types and the dissimilarities are again preference rank numbers of each student for all family types. The types were formed by taking all combinations of 'number of sons' and 'number of daughters' from 0 to 3. Thus, type 2,1 is a family with two sons and one daughter, type 0,3 one with no sons and three daughters, etc. In the standard SMACOF-III analysis, the column points form a nice grid and to convince ourselves of its stability, we applied a "bootstrap" technique. This technique, which can be seen as a

generalisation of the Quenouille-Tukey "jackknife", is a general non-parametric method for estimating the bias and variance of some statistic (Efron 1979). In our case, it simply amounts to taking N samples of size n (with replacement) from the rows of our datamatrix, and to perform N analyses (with identically the same options) on these N "new" datamatrices. Thus we "cross-validate" our results on the basis of one "real" sample only.

We have chosen $N = 10$ here and plotted all column point configurations on top of each other (figure 4). The stresses range from 0.1785 to 0.1990, a 80% confidence interval would be 0.1855 - 0.1961, which is rather narrow. The points in figure 4 are labeled according to the sample they come from (A - J) and familytype in an obvious way. The 10 configurations were not rotated or translated into best fit (only one of them had to be reflected). Still, most types are very neatly clustered, especially the more popular ones such as 2,2 and 2,1. The most impopular familytype 0,0 turns out to be very unstable; but it is always near persons with heavy sex-bias: "if I can't get two daughters, I would rather have no children at all".

5.4. The Münsingen-Rain data.

Our final example concerns the analysis of an incidence matrix, i.e. a binary matrix indicating whether or not to objects from different sets are related to each other. In particular, we have chosen an incidence matrix from archeology: the columns refer to 59 graves from the la Tène cemetery at Münsingen-Rain, Switzerland and the rows to 70 varieties of fibulae, anklets, bracelets etc. ('types'). If type i is found in grave j we set $\delta_{ij} = 0$, and $\delta_{ij} = 1$ otherwise. This incidence matrix has been referred to over and over again in Hodson, Kendall and Tautu (1971) and is widely used as a testcase for seriation techniques. The most obvious general technique to be used would be a type of generalized principal components analysis, like HOMALS (cf. van Rijkevorsel and de Leeuw 1978, who get satisfactory results if all $\delta_{ij} = 1$ are interpreted as missing), but metric MDU is a feasible alternative. Our first analysis, with standard options, gives figure 5 (stress = 0.2317); types are plotted as stars, graves are labeled according to Hodson's serial order (cf. Hodson 1968).

Note that almost all distances have been made approximately equal, which need not surprise us in view of the fact that 93,4% of the dissimilarities equal 1. The configuration of graves displays the expected horse-shoe form

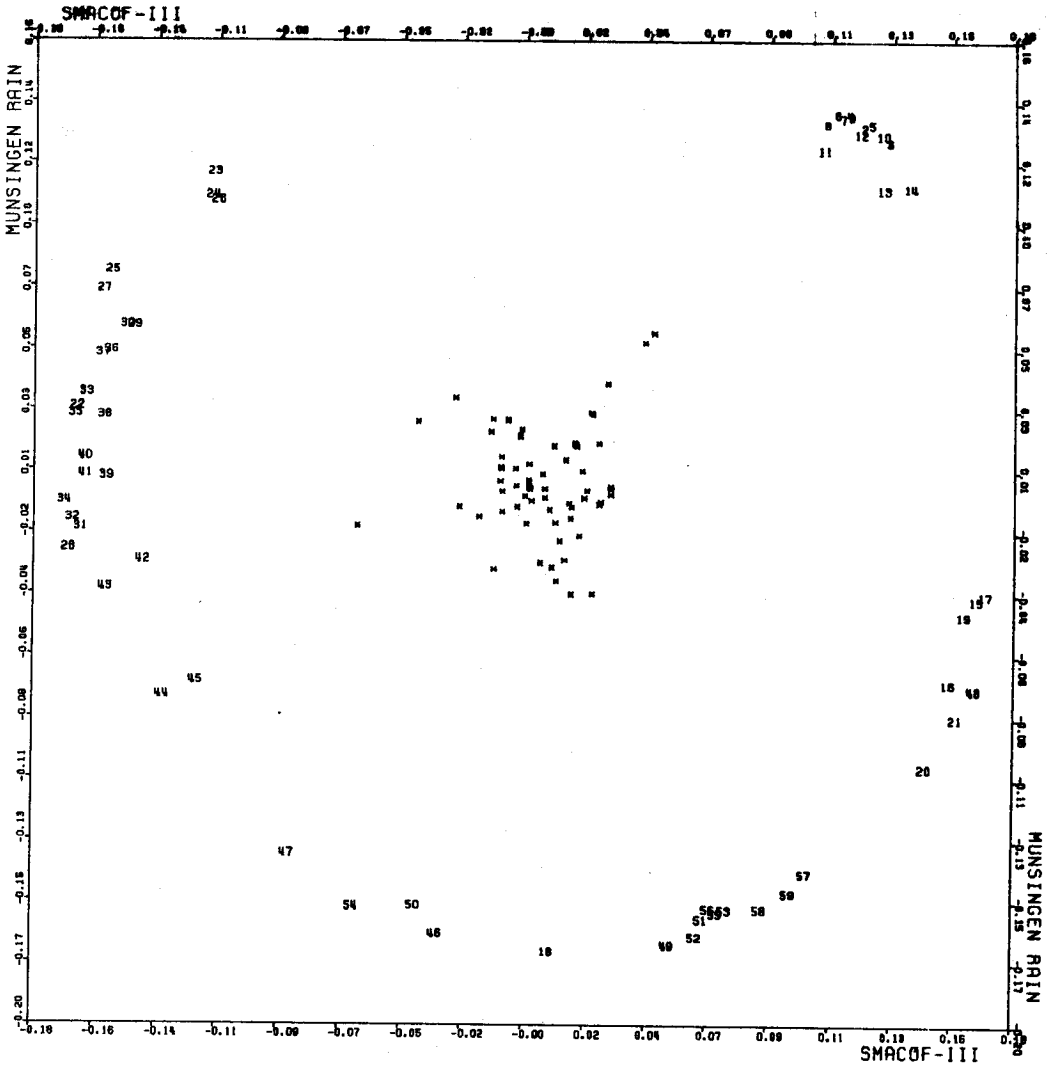


FIGURE 5. SMACOF-III; MUNSINGEN RAIN DATA (0,1)

(cf. Kendall 1971), but Hodson's order is seriously disturbed. In an attempt to remedy this, we define a "smoothed" version of the data as follows. If A is the original incidence matrix with elements

$$a_{ij} = 1 - \delta_{ij}, \quad (42)$$

then define the matrix B with elements b_{ij} as

$$B = AA'A \quad (43)$$

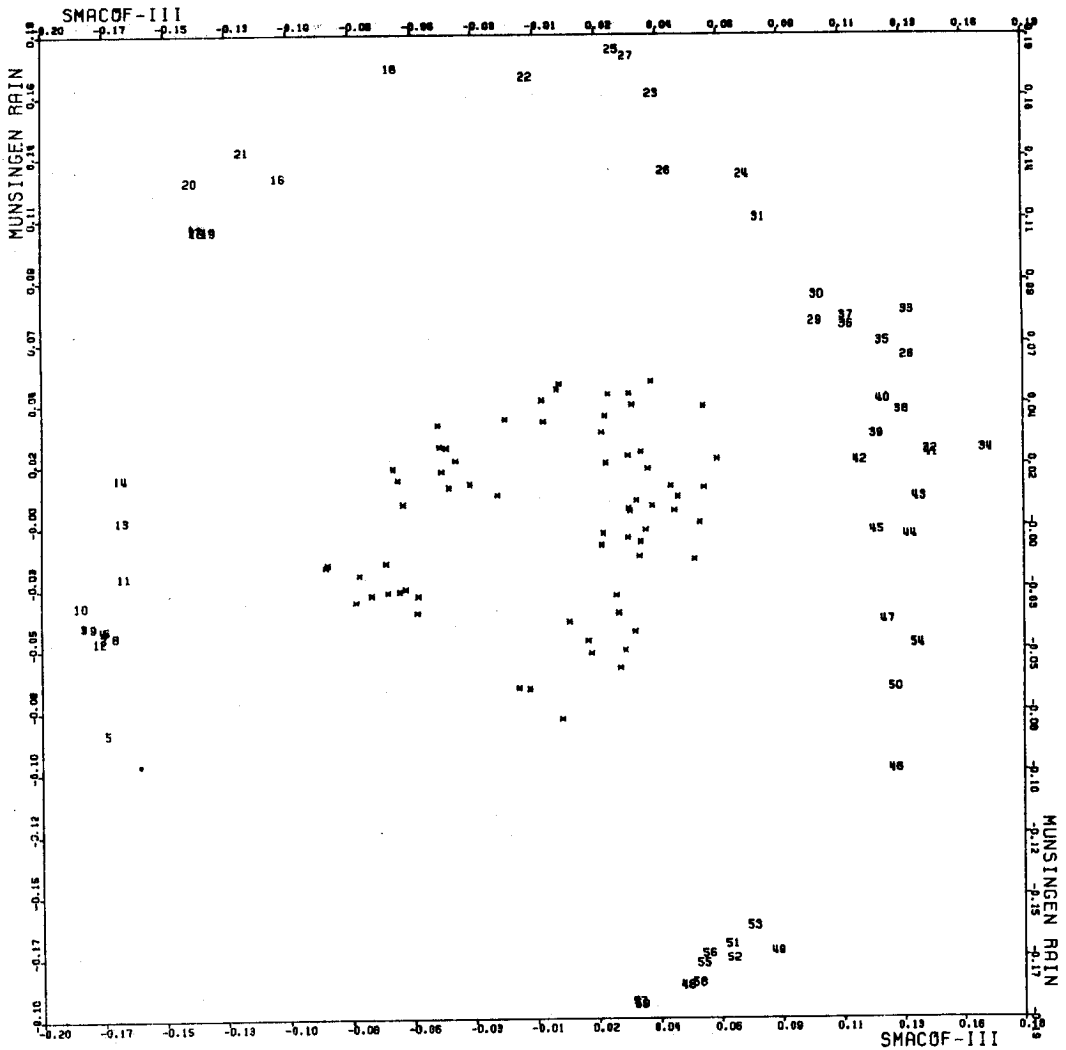


FIGURE 6. SMACOF-III; MUNSINGEN RAIN DATA SMOOTHED

and take as new dissimilarities some decreasing transformation of b_{ij} . The idea is, that A'A contains a measure of "overlap" between graves, and so the columns of B will represent the "direct" as well as the "indirect" similarity of grave j to type i . Still, 60,5% of the smoothed dissimilarities equal their maximal value. In figure 6, the obtained row and column configurations are given (stress = 0.1564); this time, two horseshoes appear and the Hodson order is globally recovered. Because the smoothing transformation is somewhat arbitrary, we use the original dissimilarities once again, put-

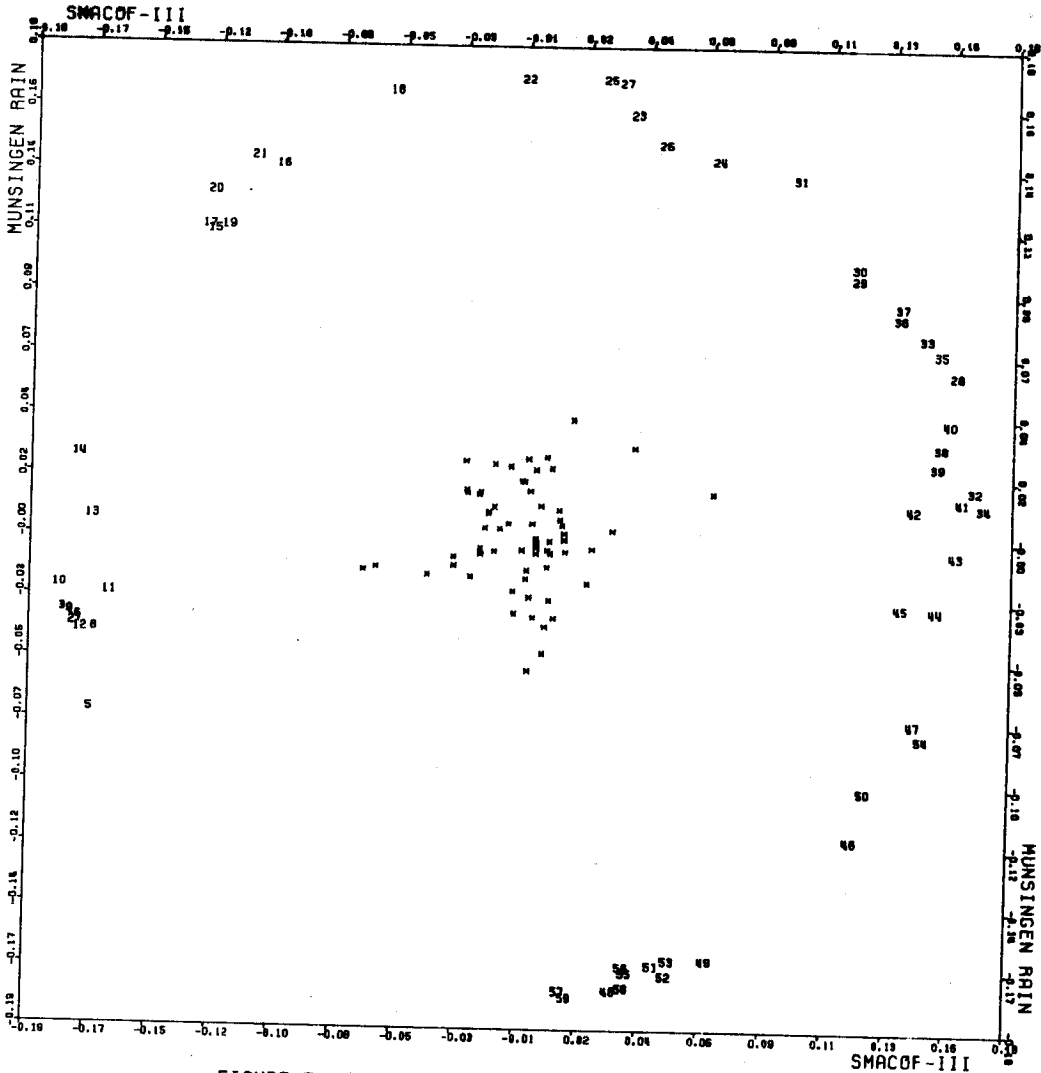


FIGURE 7. SMACOF-III; MUNSINGEN RAIN DATA (0.1)
WITH INITIAL CONFIGURATION

ting in as initial configuration the obtained results from the smoothed dissimilaties. What we get is figure 7 (stress = 0.2309). The picture has a big resemblance to figure 5 with virtually the same stress; the horseshoe of the graves, however, is nearly identical to the one in figure 6. This suggests, that many pictures like figure 7 may be obtained, with many orders on the horseshoe (depending upon the order in the initial configuration) and with about equal stresses. The non-uniqueness problem of unfolding is probably aggravated here by the non-uniqueness problem of seriation.

5.5. References.

- Bennett, J.F. and W.L.Hays, 1960. Multidimensional unfolding: determining the dimensionality of ranked preference data, PM, 25, 27 - 43.
- Coombs, C.H., 1950. Psychological scaling without a unit of measurement, Psych. Review, 57, 148 - 158.
- Coombs, C.H., 1964. A theory of data. Wiley, New York.
- Coombs, C.H., and R.C.Kao, 1960. On a connection between factor analysis and multidimensional unfolding, PM, 25, 219 - 231.
- Daalder, H. and J.P. van de Geer, 1977. Partijafstanden in de tweede Kamer der Staten-Generaal, Acta Politica, XII, 3, 289 - 345.
- Daalder, H. and J.G.Rusk, 1972. Perceptions of Party in the Dutch Parliament. In: S.C.Patterson and J.C.Wahlke (eds.) Comparative Legislative Behaviour: frontiers of research. Wiley, New York.
- Delbeke, L., 1978. Enkele analyses op voorkeursoordelen voor gezinssamenstellingen. Internal Report, Katholieke Universiteit Leuven.
- Efron, B., 1979. Bootstrap methods: another look at the Jackknife, Ann. Statist. 7, 1 - 26.
- Gold, E.M., 1973. Metric unfolding: Data requirement for unique solution and classification of Schönemann's algorithm, PM, 38, 555 - 569.
- Gold, M., 1958. Power in the classroom, Sociometry, 21, 50 - 60.
- Gower, J.C., 1966. Some distance properties of latent root and vector models used in multivariate analysis, Biometrika, 53, 325 - 338.
- Hays, W.L. and J.F.Bennett, 1961. Multidimensional unfolding: determining configuration from complete rank order preference data, PM, 26, 221 - 238.
- Heiser, W.J. and J. de Leeuw, 1979. Metric multidimensional unfolding, MDN, 4, 26 - 50.
- Hodson, F.R., 1968. The La Tène Cemetery at Münsingen-Rain. Berne: Stämpfi.
- Hodson, F.R., D.G.Kendall and P.Tăutu (Eds.), 1971. Mathematics in the Archeological and Historical Sciences. Edinburgh University Press, Edinburgh.
- Kendall, D.G., 1971. Seriation from abundance matrices. In: Hodson, F.R., D.G.Kendall and P.Tăutu (Eds.), 1971. Mathematics in the Archeological and Historical Sciences. Edinburgh University Press, Edinburgh.
- Kruskal, J.B., 1977. Multidimensional Scaling and Other Methods for Discovering Structure. In: Enslein, K., A.Ralston and H.S.Wilf, Statistical Methods for digital computers, vol.III, Wiley, New York.
- Kruskal, J.B. and J.D.Carroll, 1969. Geometrical models and badness-of-fit functions. In: P.R.Krishnaiah (Ed.) Multivariate Analysis II, 639 - 671 Academic Press, New York.

de Leeuw, J., 1973. Canonical Analysis of categorical data. Psychologisch Instituut, Leiden.

de Leeuw, J. and W.J.Heiser, 1979. Multidimensional Scaling with restrictions on the configuration. In: P.R.Krishnaiah (Ed.), Multivariate Analysis V, Academic Press, New York.

Rijckevorsel, J.van and J. de Leeuw, 1978. An outline to Homals-1. Dept. of Datatheory, Leiden.

Ross, J. and N.Cliff, 1964. A generalization of the interpoint distance model, PM, 35, 349 - 366.

Schönemann, P.H., 1970. On metric multidimensional unfolding, PM, 35, 349 - 366.

Torgerson, W.S., 1958. Theory and Methods of Scaling. Wiley, New York.

6.0. Job set-up.

CARD 1 : TITLE CARD

<u>col.</u>	<u>format</u>	<u>information</u>
1 - 80	20A4	any alphameric information to title the printout

CARD 2 : PROBLEM DEFINITION CARD

<u>col.</u>	<u>format</u>	<u>information</u>
1 - 5	I5	number of row points
6 - 10	I5	number of column points (less than number of row points)
11 - 15	I5	maximum number of dimensions (default = 2)
16 - 20	I5	minimum number of dimensions (default = 2)
21 - 25	I5	maximum number of iterations (default = 50)
26 - 30	I5	updating scheme 0 = relaxed updates 1 = Guttman transforms
31 - 35	I5	initial configurations 0 = compute 1 = provided by user
36 - 40	I5	restrictions; the absolute value of this parameter gives the number of different small distances in a column of the data- matrix which is to be used to restrict the column points in the initial solution (e.g. if a value of 1 is specified and the data are ranknumbers, each column point is res- tricted to be the centroid of all row points associated with the smallest rank number in that column). If the value of this parameter is negative, the restrictions are also used during iteration. (default = +1)
41 - 50	F10.8	convergence criterion (default = 0.00001)

CARD 3 : OUTPUT OPTIONS

<u>col.</u>	<u>format</u>	<u>information</u>
1 - 5	I5	printout of data 0 = no 1 = yes
6 - 10	I5	printout and printplot of initial configurations 0 = no print nor plot 1 = plot first dimension only (and print) 2 = plot first two dimensions (" ") k = plot (pairwise) k dimensions (" ")
11 - 15	I5	printplot of final configuration 0 = no plot 1 = plot first dimension only 2 = plot first two dimensions k = plot (pairwise) k dimensions
16 - 20	I5	history of computation 0 = no 1 = yes
21 - 25	I5	printplot of data versus distances (Shepard plot) 0 = no 1 = yes
26 - 30	I5	punch and store options (results are written to medium specified below) 0 = no punch/store 1 = final configuration punched/stored 2 = final distances punched/stored 3 = both final matrices punched/stored k = if the initial configuration should be punched/stored, add 10 to the options above
31 - 35	I5	reference number of the input medium for the data (default=5)
36 - 40	I5	reference number of the punch/store medium (default=7)

CARD 4 : DATA FORMAT

<u>col.</u>	<u>format</u>	<u>information</u>
1 - 80	20A4	FORTRAN format information for the data matrix (F-format)

FOLLOWING CARDS: DATA MATRIX

The data must appear here if the input medium is the same as that of the parameter cards, in the format specified on card 4.

FOLLOWING CARD: FORMAT OF THE INITIAL CONFIGURATIONS (optional)

<u>col.</u>	<u>format</u>	<u>information</u>
1 - 80	20A4	FORTRAN format information for both initial configurations (F-format; only if there is a 1 in columns 31-35 of card 2)

FOLLOWING CARDS: INITIAL CONFIGURATION MATRICES (optional)

If initial configurations are input, they should appear here, with the format specified above. The expected order is: first row points (size $n \times p_{\max}$), then column points (size $m \times p_{\max}$).

N.B.1 Cards 1 - 3 are read from input medium 5; card 4 and the following ones are read from the input medium specified in columns 31 - 35 of card 3 (default = 5).

N.B.2 To estimate the total number of words needed for the array area, use the formula $NWORDS = NROW \times (2 \times NDIM + (NROW + 5)/2) + NCOL \times (2 \times NDIM + 3) + NROW \times NCOL \times 3$. For problems where $NROW + NCOL$ is less than $NDIM \times NDIM$, add the difference to $NWORDS$. In the static allocation version, $NWORDS$ is fixed at 16500; this implies that analyses of 100×35 or 80×50 in 2 or 3 dimensions can be done within the standard area.