6 Programs For Analysing Two-Way Data

Now, in perusing what follows, the reader should bear in mind not only the general circuit as adumbrated above, with many sidetrips and tourist traps, secondary circles and skittish deviations but also the fact that far from being an indolent partie de plaisir, *our tour was a hard, twisted teleological growth whose sole* raison d'être *(these French clichés are symptomatic) was to keep my companion in passable humour from kiss to kiss.*

V. V. Sirin, *Taina* (p. 260)
translated from the Russian by
Vivian Darkbloom as *Tatyana* (p. 19)

In this chapter programs implementing models and transformations appropriate for analysing two-way data are described (further details are contained in the *MDS(X) User's Manual* Reports).

The defining characteristics of the relevant MDS(X) programs are described in Table 6.1, and the sections of this chapter follow the sequence indicated in the final column of the table. For two-way data, the basic distinction is whether the solution represents the relationship between *one* set of objects, the stimuli (one-mode) or between *two* sets of objects, the subjects and the stimuli (two-mode). This distinction is used to organise this chapter.

The chapter is written in such a way that the user can either read it through sequentially to get an overview of all the relevant programs, or else go directly to the section (program) which is of most immediate interest. Each section is therefore largely self-contained and refers to a single program or two highly similar programs.

The structure of each section is also the same and consists of:

a brief definition of the program, listing its specific characteristics:
an extended description of the program and its uses: followed usually by
an example of its application.

6.1 One-mode Data

A mode refers to a single class of entities, which could in fact be stimuli, subjects, test items, occasions, geographical sites, botanic species, lexical items etc. (Carroll and Arabie 1980). Two-way single-mode data consist typically of a matrix giving (dis)similarity or other proximity measure between each pair of objects taken from the class of entities. Such data are certainly the most common form of data.

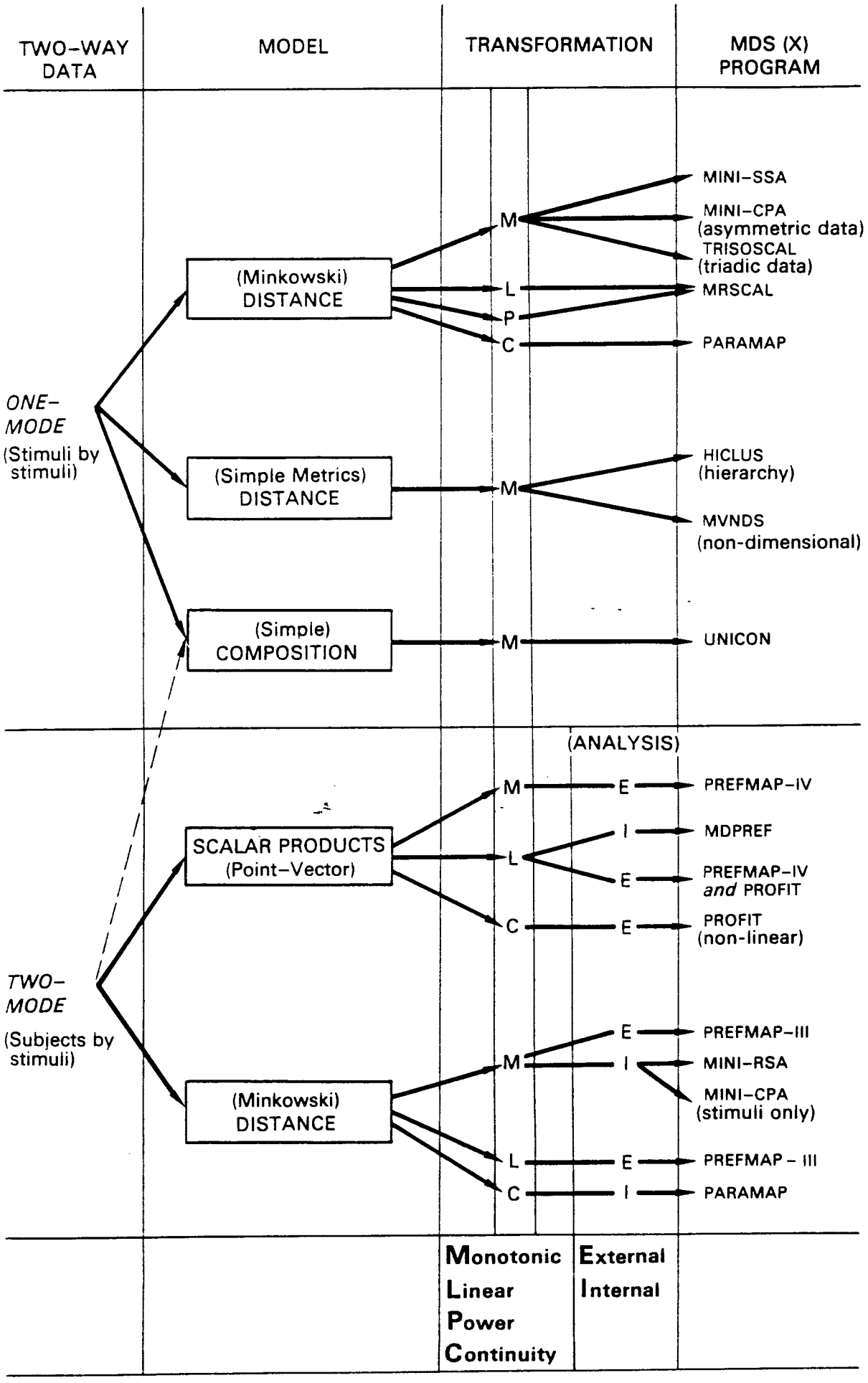


Table 6.1 *Analysis of 2-way data by MDS(X) programs*

6.1.1 The basic non-metric model (MINISSA(N) and SSA(M))

Concisely: MINISSA (Michigan-Israel-Nijmegen Integrated Smallest Space Analysis) in its *Nijmegen* and *Michigan* versions provides

internal analysis of a two-way symmetric matrix of (dis)similarities
by a Euclidean distance model
using a monotone transformation of the data.

The form of this basic non-metric model has been described fully in 3.5 and 3.6. The simplest and most efficient implementation is the Nijmegen version (MINISSA) and there is also the more extended Michigan version (SSA(M)) which contains parameters for permitting a wider range of alternatives:

Data Missing data, fixing part of the configuration (in effect, external scaling) and adding points

Transformation Local or global monotonicity, Kruskal or Guttman's minimisation procedure.

6.1.2 Conditional Proximity Analysis (MINICPA)

Concisely: MINICPA (Conditional Proximity Analysis) provides:

internal analysis of two way (dis)similarity data in a row-conditional format
by a Euclidean distance model
using a monotone transformation of the data.

The MINICPA program is designed to implement the third approach to scaling asymmetric data described in 5.1.1.1. It differs from MINISSA only in that each row of the square data matrix is assumed to be conditional, i.e. as a rank order in its own right. Usually, then, δ_{jk} will not normally equal δ_{kj} and each of the two dissimilarities are fit by a separate disparity value. There is no separate representation of the row and column elements: row j and column j are both represented by a single stimulus point j. The solution configuration provided by CPA will consequently be a compromise between the upper and lower triangular elements, since the diagonal elements are ignored.

This program is particularly well adapted to the analysis of data collected by the method of conditional rank order, or where the user wishes to ignore asymmetries and obtain a single solution for the row and column elements. A typical example occurs for sociometric data, where each individual in a group is asked to provide a rank order of those closest or most similar to him or herself. Collected together, such data form a square, row-conditional set of rankings whose diagonal entries are irrelevant. For example, Gleason (1969) analyses an aggregated set of sociometric rankings obtained from the last four weeks of Newcomb's (1961) study of the process of friendship formation in an experimentally-monitored men's dormitory at the University of Michigan.

In the original analysis, Gleason analyses the matrix as two-mode data (cf. section 6.2.3) with the individual-as-row element (the giver of the friendship choice) being represented as one point and the (same) individual-as-column element (the receiver of the friendship choice) as a distinct point. Considerable attention is paid

not only to the inter-personal distances but also to the 'intrapersonal separations' of the individuals' two 'ideal points'. As might be expected, the intra-personal distances are small compared to the interpersonal ones, but appear to be otherwise uninterpretable. Gleason (ibid. p. 118) coyly concludes, 'Unfortunately, very little else can be said about these distances at this time', and indeed he himself symmetrises the data for performing hierarchical clustering. Such considerations strongly indicate the use of a one-mode conditional proximity approach of MINICPA.

Whilst this example is more useful as an example of what can be done than providing a good illustration of its utility, further procedures for scaling such sociometric data are given in Breiger, Boorman and Arabie (1975), and MINICPA provides a useful addition to the methods discussed there. The Newcomb data form the test data for the MINICPA program in the MDS(X) version.

6.1.3 Triadic data analysis (TRISOSCAL)

Concisely: TRISOSCAL (TRIadic Similarities Ordinal SCALing) provides:

internal analysis of a set of triadic (dis)similarity measures
by a Minkowski distance model
using a local or global monotonic transformation of the data.

Triadic data are collected especially by psychologists, sociologists and anthropologists wishing to elicit the constructs which subjects use in making judgments of similarity. The basic idea is that the subject is asked to consider groups of three objects at a time, taken from the full set. Two forms are in common use:

partial triadic data, where from each presentation of the three objects the subject is asked to pick out only the single most similar pair; and

full triadic data where both the most similar pair and the least similar pair are picked out. The intermediate pair can then simply be inferred.

Although the method of triads is a useful technique for data collection, the number of triads increases very rapidly with the number of objects (for $p = 5$, 10, 15 and 20 the number of triads is 10, 120, 455 and 1140 respectively). Obviously, beyond about $p = 8$, the presentation of the full set of triads becomes totally unfeasible and very taxing on the subject. Burton and Nerlove (1976) give a full description and discussion of experimental designs for minimising the number of triads presented whilst maximising the information gained.

The most important advantage which triadic data possess compared to simpler forms is that contextual effects of judgment can be directly examined—by examining whether the similarity between two objects remains the same when the third element is changed.* It is therefore unfortunate when triadic data are turned into 'vote-count' data before scaling, since the effect is to obliterate the triadic information. (In brief, the vote-count method consists of counting the number of

*The assumption that the judgment remains unchanged in the presence of irrelevant alternatives is important in a number of theories of choice, preference and social welfare: see Rescher (1969).

times that object j is judged more similar than object k in the data.) Roskam (1970, p. 406) has shown that such a procedure badly misrepresents the order information in the data and often results in ill-fitting scaling solutions.

The TRISOSCAL program provides a direct method for scaling triadic data non-metrically by a distance model. It differs from Roskam's original MINITRI program in allowing the user to decide either that the order information implied across *all* the triads be fitted systematically ('global stress'), or only that the separate orders *within* each triad be fitted.

The distinction can be illustrated as follows. Suppose triadic data have been collected from a group of individuals. That being so, it is quite likely that when presented with objects (A, B, C) one subject will decide that (AB) is most similar and (AC) least similar (implying that $d(A, B) \leqslant d(B, C) \leqslant d(A, C)$). Another subject when presented with the same triad, may decide just the opposite—that the pair (AC) is most similar, and (AB) least similar (implying that $d(A, C) \leqslant d(B, C) \leqslant d(A, B)$). Both agree, by implication, that $d(B, C)$ is intermediate—but how shall the conflicting information concerning $d(A, C)$ and $d(A, B)$ be fitted? The answer proposed by Roskam's 'local stress' approach is: treat each subject's triad, that is, judgment, as a distinct entity, fit each of the three distances within a triad separately and then define 'the' overall fitting value as the average of the different disparity values. Hence there will be as many disparity or fitting values (d^0_{jk}) as occurrences of the pair (j, k) in the triads data. In this instance, there will be two distinct values of d^0_{ab} and d^0_{ac}, but their respective arithmetic average (denoted $\bar{d}^0_{ac}$) will represent 'the' fitting value for the pair concerned. Roskam suggests that the form of raw stress to be minimised should be:

Triads: 'Local' Stress

$$\text{Stress}_0 = \sum n_{jk}(d_{jk} - \bar{d}^0_{jk})^2$$

where n_{jk} is the number of times that the pair (jk) occurs in the set of triads to be analysed.

Roskam's approach obviously tolerates a good deal of inconsistent data, but is a sensible strategy when data from a set of individuals is combined, or where the number of objects is large.

Prentice (described in Coxon and Jones 1979, p. 49 et seq.) suggests a more restrictive and stringent approach: to require total consistency *between* triads and to count each and every infraction of transitivity in the stress value rather than averaging. Consider the following two pieces of triadic data:

	Triad	Most similar (MS) pair	Least similar (LS) pair	Dissimilarity data implied
1	(A, B, C)	(AB)	(BC)	$(AB) < (AC) < (BC)$
2	(B, C, D)	(BC)	(CD)	$(BC) < (BD) < (CD)$

Taken together, the information from both triads is consistent (transitive) and implies the following order of data:

$$AB < AC < BC < BD < CD.$$

Prentice's global approach requires that the data be fitted by disparity values in the same order, and in particular he requires that the dissimilarity (BC) be fitted by the *same* disparity value in both Triad 1 and in Triad 2.

The two approaches and the different fitting values which result are illustrated in Table 6.2 for the set of data just given. Assume that there is a configuration of four points whose interpoint distances are as in the second column of the table. Column 1 gives the pairs in the two triads in their correct order according to the data, and column 2 gives the corresponding distances in the current configuration (note that allowance has to be made for two fitting values for (BC), although there can obviously be only one such distance in the configuration).

How well does the current configuration match the data according to the criteria of local and global stress? In the next three columns (3–5) the necessary fitting values are calculated using monotone regression.

Column 3 Local stress allows the fitting value for (BC) in the first triad to be different to the fitting value for (BC) in the second triad, so the d values are calculated separately within each triad, yielding $(BC)_1 = 2\frac{1}{2}$ (the block average of 3 and 2) and $(BC)_2 = 1\frac{1}{2}$ (the block average of 2 and 1).

Column 4 'The' fitting value for (BC) according to Roskam's local stress is defined as the average of its two appearances, i.e.

$$\bar{d}^0_{bc} = (d^0_{(bc_1)} + d^0_{(bc_2)})/2 = (2\tfrac{1}{2} + 1\tfrac{1}{2})/2 = 2$$

Column 5 The fitting values according to Prentice's global stress approach, which requires weak monotonicity over all the pairs, produce a *single* fitting value for (BC), which by block-averaging over all but the first and last pair, gives a value

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	DATA	DISTANCE	FITTING VALUE DISPARITIES			RESIDUALS	
	Pair (ij)	d_{ij}	d^0 (local)	$\bar{d}^0$ (local)	$\hat{d}$ (global)	$(d - d^0)^2$ (local)	$(d - \hat{d})^2$ (global)
Triad	(AB)	1	1	1	1	0	0
1	(AC)	3	$2\frac{1}{2}$	$2\frac{1}{2}$	2	$\frac{1}{4}$	1
	$(BC)_1$	2	$2\frac{1}{2}$	2	2	0	0
Triad	$(BC)_2$		$1\frac{1}{2}$	2		0	0
2	(BD)	1	$1\frac{1}{2}$	$1\frac{1}{2}$	2	$\frac{1}{4}$	1
	(CD)	4	4	4	4	0	0
					Total	$\frac{1}{2}$	2

Raw Stress (local) $= \frac{1}{2}$
Raw Stress (global) $= 2$

Table 6.2 *Triadic data: illustration of local and global stress*

of 2. (It so happens that in this example both local and global approaches give the same fitting value for d_{bc}; this will not often be the case.)

The squared discrepancy values contributing to raw stress are given in column 6 for local and in column 7 for global stress, and the stress values are given at the foot of the table. Note that, as expected, requiring a complete ordered fit over all the pairs of both triads increases the badness-of-fit. Note also that global stress assumes the same value (2) for all but two pairs, but *is* weakly monotone with the order implied by the data. Local stress, by contrast, not only fits the same distance (BC) by two distinct values but also averages them to obtain the 'overall fitting value' which (as in this case) usually will *not* be in the same order as that implied by the triads:

$$\begin{array}{rccccccccc} \text{data:} & AB & < & AC & < & BC & < & BD & < & CD \\ \hat{d}^0\text{:} & 1 & < & 2\tfrac{1}{2} & \not< & 2 & \not< & 1\tfrac{1}{2} & < & 4 \end{array}$$

In the case of inconsistent sets of triads, even greater inversions occur. When triadic information comes from a number of subjects or sources, highly inconsistent (intransitive) data often result and the user is faced with a difficult choice: either to choose global stress and risk technically degenerate solutions (since the inconsistencies can only be dealt with by imposing the same fitting value on a large number of pairs, capitalising on weak monotonicity) or to choose local stress, and lose all information about the order across triads as well as drowning same-pair inconsistencies by fitting averaged disparities. At least the Prentice approach signals the inconsistencies and global intransitivities by a high stress value—though at the disadvantage of increased computing costs.*

In such situations the user is advised to scale at least a random subset of data (to save computing time), using the global stress approach, and to compare resulting stress values and configurations with those obtained using local stress. Whichever stress option is chosen, triadic data are scaled in their integrity, and as far as possible consistency is kept within triads in the configuration. But only the global stress approach tries to represent information *between* the triads.

In some applications it is advantageous to insist upon between-triad comparisons by choosing the global approach, as when context effects are to be minimised. But by the same token context effects can best be studied by collecting triadic data and the user can then examine whether a given pair of objects tends to be fitted by approximately the same value. This can only be done by choosing the local stress approach.

6.1.4 The basic metric model (MRSCAL)

Concisely: MRSCAL (MetRic SCALing) provides:

internal analysis of two-way data of a lower triangle format of a (dis)similarity measure
by a Minkowski distance function,
using a linear and/or logarithmic transformation of the data.

*In Coxon and Jones 1978a, global stress$_2$ values of over 0.95 were frequently observed for such heterogeneous data sets, with corresponding local stress$_1$ values of around 0.20. In one case, 169 triadic comparisons of 13 occupations made by a set of policemen produced a global stress$_2$ value of 0.960 when scaled, and 75 out of the 78 global $\hat{d}$ values had the same value!

As we have seen, the assumption that data dissimilarities are a linear function of the distances of the solution historically precedes the monotonic assumption, and the earliest computational techniques for distance model scaling all assumed a linear (or ratio) transformation. Nowadays it is sensible to begin by scaling one's data by the non-metric model. thus making more defensible assumptions about the level of measurement of one's data. But since regular functions are special cases of the general monotonic family of transformations, the Shepard diagram obtained from non-metric scaling should be inspected to determine whether a more regular relationship is discernible between the data and the solution. If so, it makes eminent sense to go on to submit the data to a program such as MRSCAL, which implements the more regular linear and power transformations described in section 5.2.3. The examples in Chapter 3 provide illuminating illustrations of Shepard diagrams where a more regular relationship is evident. For the Scottish mileage data. the Shepard diagram (Figure 3.4) suggests an S-shaped or sigmoid power function*, although in the main range of distances (0.05 to 2.00) the relationship to the data is linear. In the small illustrative example of the similarity of eight crimes (Figure 3.2) the perfect strong monotone function is very close to being linear ($r = 0.97$). By contrast, in the case of the 'real data' example of scaling the co-occurrence frequency of occupational titles. the final Shepard diagram at iteration 23 (Figure 3.14b) shows a distinctly J-shaped (downwardly concave) form. again suggesting the use of a power function (in this case. a negatively accelerated exponential decay function).

As a matter of interest, the 2-dimensional configuration and its associated Shepard diagram obtained from a *linear* scaling of the seriousness of offences data are presented in Figures 6.1 and 6.2. Because the data are very well fit, the equation of the line relating the data to the solution distances is of interest, and attention focuses chiefly upon the slope, as in any other case of linear regression. The linear scaling transformation can be interpreted as indicating that, if the data are increased throughout by an additive constant of 0.251, then the distance between the points will on average be one quarter (0.241) of the difference between those saying they are 'unalike in their seriousness'. The advantage of a regular transformation function, then, is that it is possible to extrapolate beyond the original data (and in this sense 'predict' further data) if the model is correct. (In this particular example, we should not expect the transformation to be linear throughout its range, since there is an upper limit of 1.00 on the data values.)

The MRSCAL program in the MDS(X) series implements the linear and power models outlined in Roskam (1972) and contains options for implementing the city block and Euclidean distance metrics, among others.

6.1.5 Parametric mapping (PARAMAP)

Concisely: PARAMAP (PARAmetric MAPping) provides:

internal analysis of either a rectangular or a square symmetric two-way data matrix

by a distance model which maximises continuity or local monotonicity.

*In fact, a logistic function would best fit these data, but an ordinary logarithmic transform closely approximates its form.

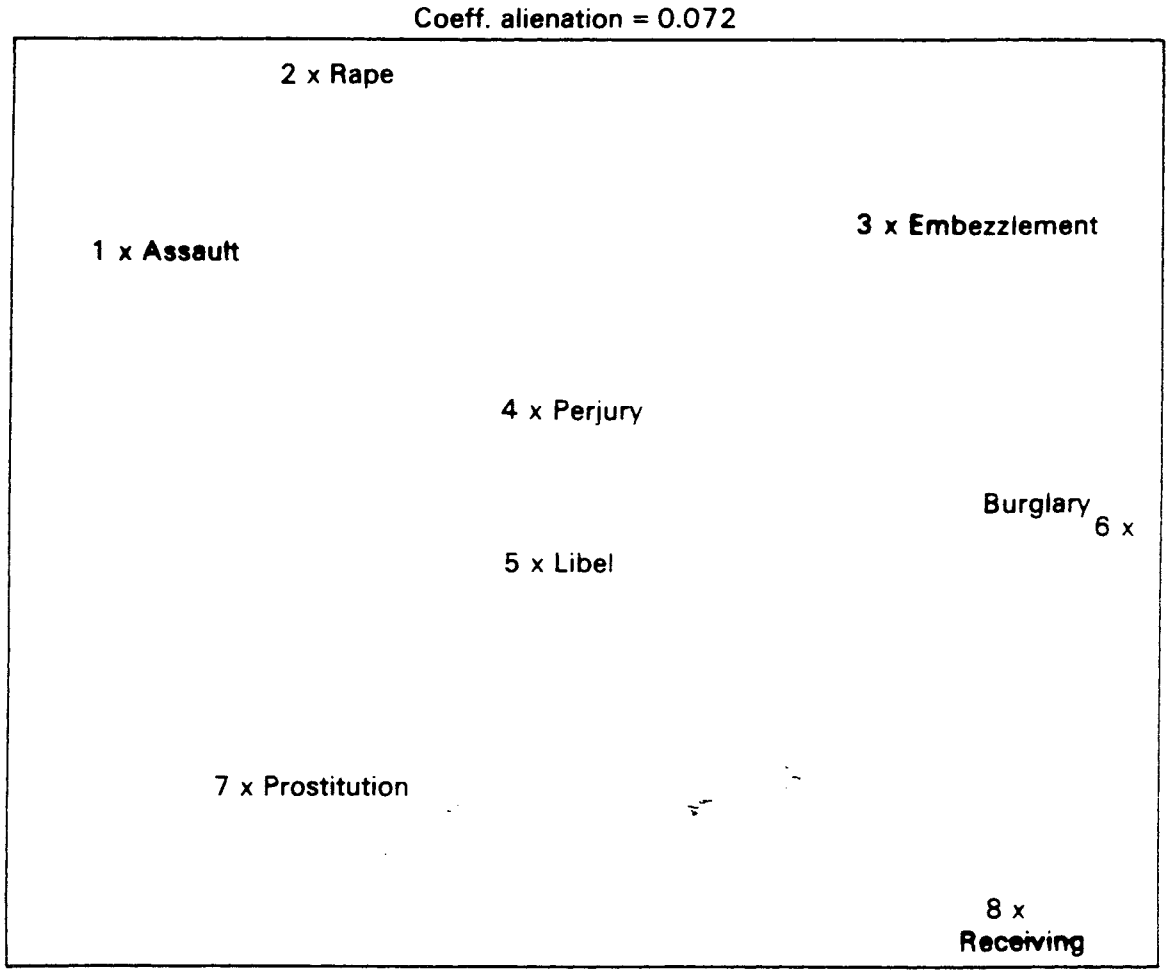


Figure 6.1 *MRSCAL 2-D solutions to data of Table 3.1*

The PARAMAP program accepts data either in the form of symmetric lower-triangular distances *or* in rectangular form consisting of profile values or spatial co-ordinates, but it only represents the objects (rows) in the latter case. The 'smoothness' or continuity transformation used in PARAMAP is the kappa family of continuity indices described in detail in 5.2.2.1 and in Appendix A5.1. The default values of the program parameters produce the simple 'normalised kappa' index (Appendix A5.1, equation 5), and the effect of these and other variants on the representation of the data are discussed in some detail at that point.

The main distinguishing characteristics of parametric mapping are:

(i) the faithful preservation of the *local* information (small distances) around each point, virtually ignoring large distances (the user is thus given control of the degree of local monotonicity); and

(ii) the 'flattening' of configurations into as small a dimensionality as possible, by increasing the size or variance of the largest distances.

The two are related: the price paid for preserving local information in a low dimensionality is the considerable distortion of global information. This fact should be borne in mind when interpreting PARAMAP solutions, since it contradicts the usual MDS maxim that configurations are globally stable, in the sense that the main features are reliably fixed but are locally unstable in that the points in a configuration can be moved around slightly without any major change in stress.

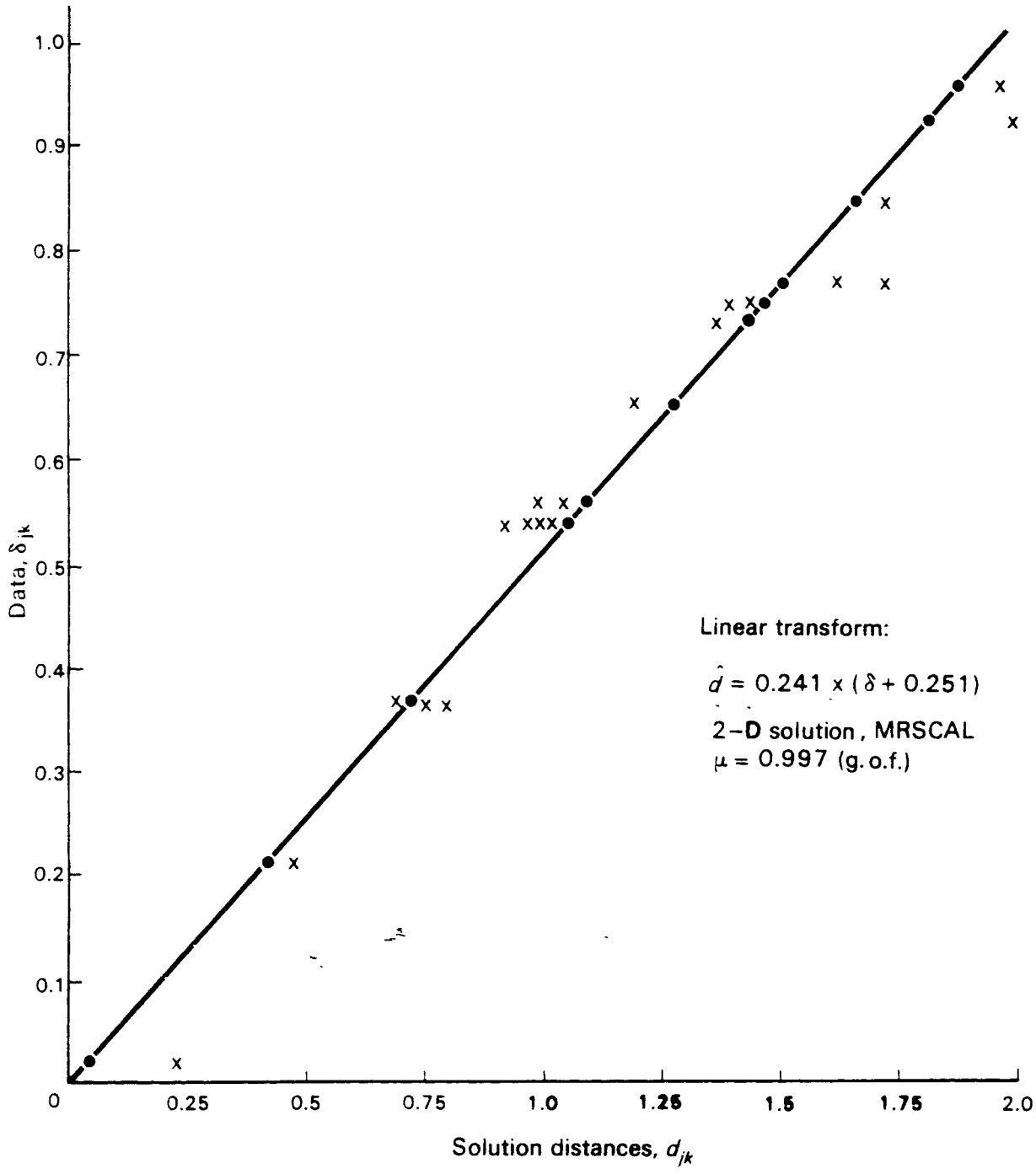


Figure 6.2 *Shepard diagram metric (linear) solutions to data of Table 3.1*

Moreover, because of the considerable non-linearity of the continuity rescaling transformation, any external properties the user wishes to represent should be mapped into a configuration using the *non-linear* option in PROFIT, which optimises an index akin to kappa. If hierarchical clustering is used to interpret a PARAMAP solution, only the initial stages of the clustering should be mapped into the configuration, because the largest distances are bound to be badly represented as a result of the continuity criterion.

The most dramatic examples of parametric mapping, which show these properties most obviously, occur for the mathematical structures of points defined at regular intervals on the circle (mapped into the line), on the sphere and on the torus, when mapped down into two dimensions (see Shepard and Carroll 1966,

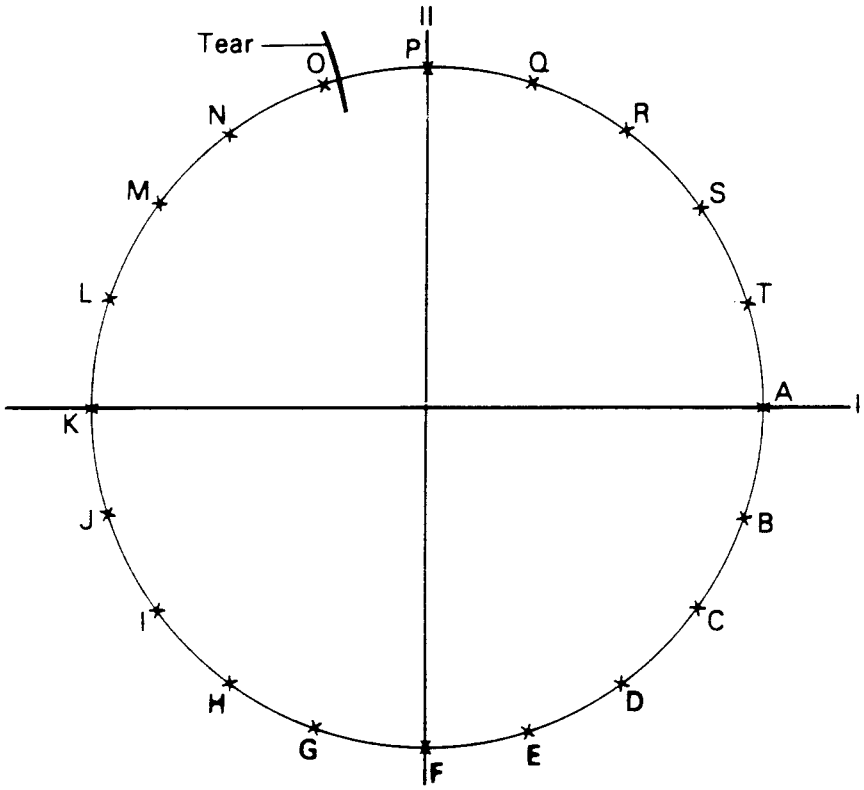
section 3.7). These form the test data for the MDS(X) program. The mapping of 20 points on a circle down onto a line is illustrated in Figure 6.3. In the original 2-dimensional space (the circle) the points are equally spaced: a continuity mapping attempts to preserve this property in the 1-space. But since the program is required to represent it in one dimension, this can best be done (producing only one major discontinuity) by tearing the circle at some point—indeed, at *any* one point. At the first iteration the adjacent points are equally spaced (except the two end points) but the ordering is entirely wrong—they are about four positions away from where they should be. By the 15th iteration the fit (κ) is nearly perfect and by the 30th the sequence of equally-spaced points is entirely correct except for the 'tear' in the circle between O and P. The Shepard diagram is not shown in this example (it forms a perfect quadratic function, as it should!) but the usual shape in the case of a PARAMAP solution is that of a fan (see Figures 5.3 and 6.4), with the larger dissimilarities fit by a wide range of distances (the broad end) and the smallest values fit extremely well (the narrow end).

Parametric mapping is defined and illustrated by a range of examples in Shepard and Carroll (1966), but the number of other applications in the published literature is disappointingly small. It is an especially useful procedure both for the analysis of two-mode profile data, when only the objects/stimuli are to be represented (ibid., section 2.1) and for mapping an already obtained configuration or set of distances into a yet lower-dimensional space. The former type of application is illustrated by Coxon and Jones (1979, pp. 140 et seq. and 1979, section U5.1), where subjects judged how appropriate 50 descriptions (such as 'A—would have a boring repetitive job') were of 20 occupations. Each occupation was then described by a 50-description profile, consisting of the percentage of subjects who considered that a given description was 'always' appropriate, and the resulting set of 20 profiles were then submitted to PARAMAP. Since PARAMAP is prone to produce local minima, and is also generally slow to converge, three different analyses were run, and the best yielded a kappa value of 1.09 in two dimensions. Figure 5.3 is the resulting Shepard diagram. The resulting configuration bore a good resemblance to the three-dimensional INDSCAL solution of similar data (see Coxon and Jones 1979, U5.2 for a discussion of the comparison), although only one dimension was identifiably the same—in effect, two of the INDSCAL solutions had been merged by the local monotonicity constraints.

The second type of application is illustrated by using PARAMAP to attempt to 'unbend' the highly non-linear horseshoe sequence of occupations described in 4.6 into a one-dimensional continuum. (Devotees of Conan Doyle will recognise that Holmes performed a similar feat, recalled in Kendall 1971a, p. 231.) The three-dimensional configuration was submitted to a one-dimensional parametric mapping, producing the instructive result illustrated in Figure 6.4.

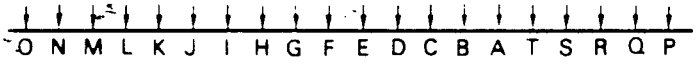
It will be recalled from section 4.6 that graphical interpretation of a 3-D scaling configuration of occupational similarities yielded a 'horseshoe', bent at the end into the third dimension, representing skill (see Figure 4.5). How well can PARAMAP project this 3-space into what is believed to be a non-linear one-dimensional sequence? Using the default options for kappa, the result is given in Figure 6.4. The one-dimensional sequence does a good job in tracing out the horseshoe sequence, with the highly proximate occupations kept close along the line—as should be the

1. Input data:
   20 points on
   a circle



2. Solution
   (1-dimension)
   Iteration 1:

   Iteration 30



3. Fit index

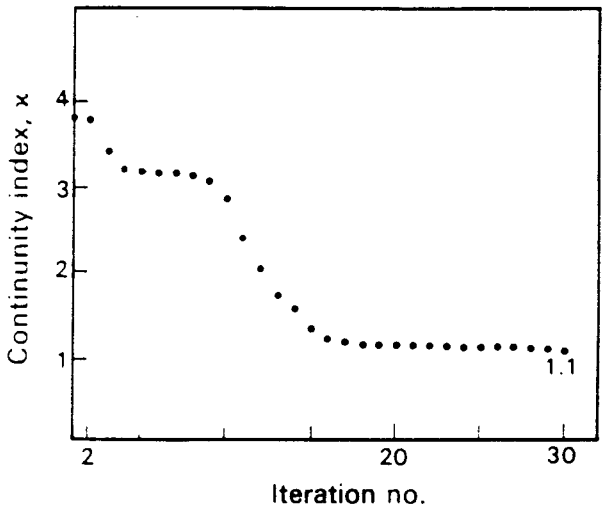
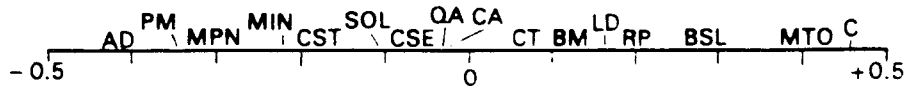
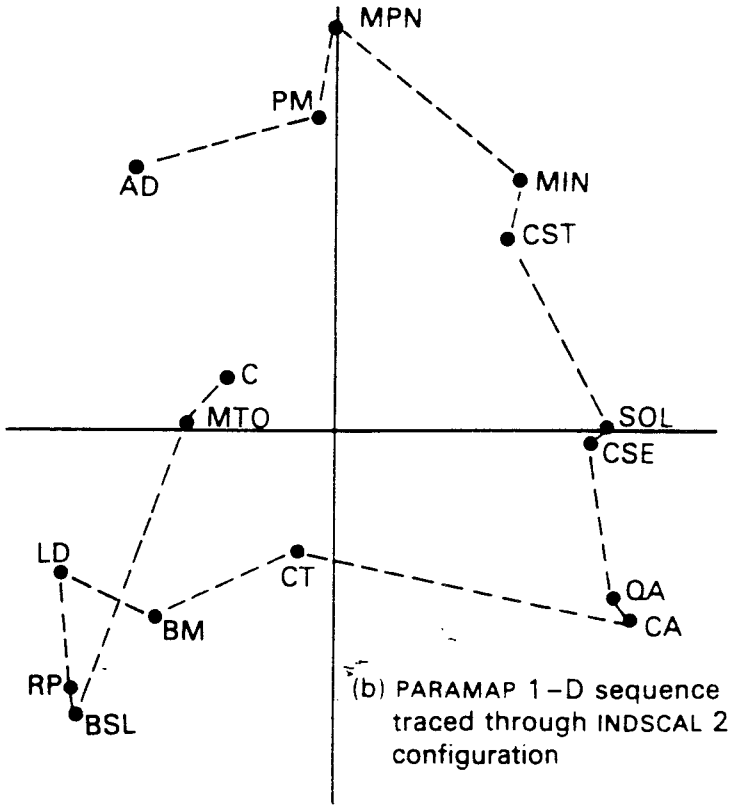


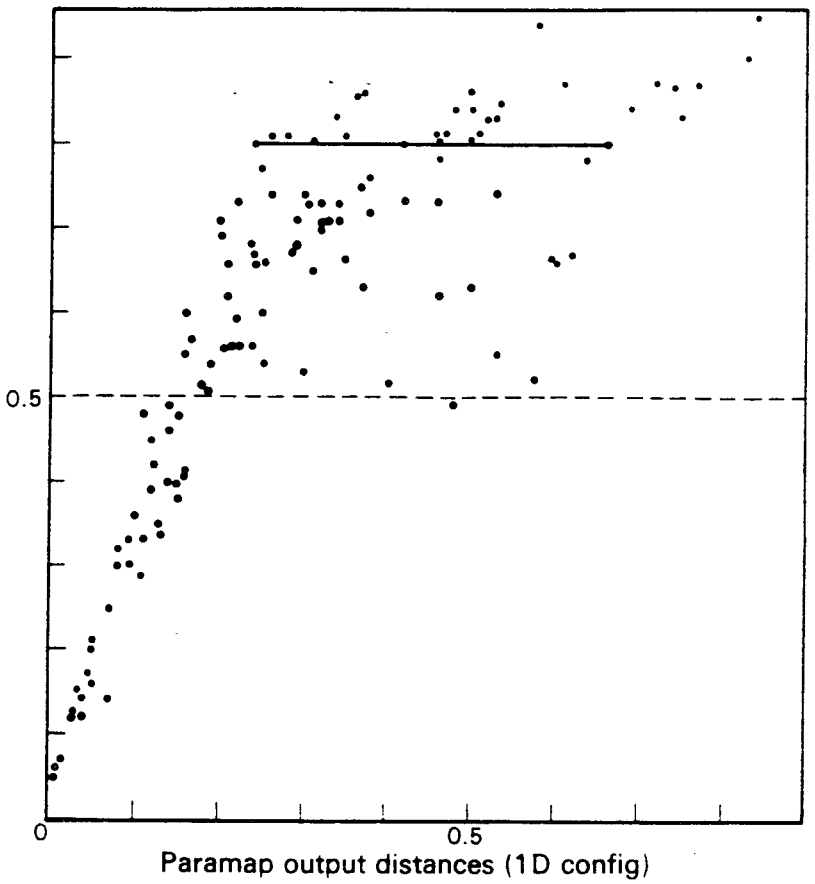
Figure 6.3 *PARAMAP maps a circle into a line*



(a) PARAMAP 1–D mapping ($x = 1.07$, with $a = 1$, $b = 2$, $c = 1$), iter = 30



(b) PARAMAP 1–D sequence traced through INDSCAL 2–D configuration



Input data (INDSCAL 3–D distances)

(c) Shepard diagram

Figure 6.4 *PARAMAP 1-D sequencing*

case with kappa giving so much weight to local monotonicity. When the PARAMAP sequence, but not the actual intervals, is mapped back into the original configuration (Figure 6.4), it can be seen that the 'cut' in the horseshoe sequence has been made between AD and C and not, as in the original interpretation, between BM and CA. Is the PARAMAP sequence wrong? Not really. The default values used make the solution virtually ignore long distances, so we should not expect that they will be faithfully reproduced. Look at the enormous variation in the PARAMAP distances of 0.5 and above for original configuration distances indicated above the line in Figure 6.4. For instance, input distances of 0.8 have output distances ranging from 0.24 to 0.66—almost half the total range! In fact a reasonable job is done in reproducing even the larger distances—after all, the ambulance driver-lorry driver distance, which is an important part of the original interpretation, is actually greater than the chartered accountant-commercial traveller interval which PARAMAP links.

The analysis followed in this example—preliminary scaling in multidimensional space, detection of non-linear sequences by graphical procedures and subsequent mapping a configuration down into a one-dimensional scale—may seem a cumbersome and complex procedure, but it is salutary to remember that the sequence would almost certainly *not* have been detected if we had proceeded immediately to a uni-dimensional scaling of the data. However, users should also be aware that the parametric mapping procedure assumes that the input data already represent distances in a space, and caution should be exercised before mistakingly submitting *any* dissimilarity data as input to PARAMAP. A useful initial step would be to make preliminary use of a program such as MVNDS (see 6.1.7) to produce a 'better behaved' set of distances from the original data.

6.1.6 Hierarchical clustering schemes (HICLUS)

Concisely: HICLUS (HIerarchical CLUStering) provides:

internal analysis of two-way (dis)similarity data
by means of a hierarchical clustering model,
using a monotonic transformation of the data.

Hierarchical agglomerative clustering is an unusual inclusion in a set of programs for MDS, since it really belongs in the family of discontinuous clustering ('taxonomic') models for the analysis of dissimilarity data, and the solution consists of a discrete set of groups or classification,* rather than continuous spatial models of MDS. Nonetheless, HICLUS has so frequently been used in conjunction with MDS solutions that its inclusion as a utility is virtually mandatory.

In the language of cluster analysis, HICLUS includes two agglomerative procedures (that is, methods which begin by merging two most similar points into a single grouping and continue to merge points into successively more inclusive groupings). These two methods are:

single linkage (or nearest neighbour), referred to as the 'connectedness' or minimum method in HICLUS; and

*An elementary introduction to cluster analysis is given in Everitt (1974) and a more advanced mathematical treatment is contained in Jardine and Sibson (1971). See also 8.4.2.

complete linkage (or furthest neighbour or diameter method) referred to as the 'diameter' maximum method in HICLUS.

Since the two methods define 'the' distance between a cluster and another point as the minimum or the maximum intra-cluster distance respectively, they represent two extreme ways of representing the data and neither will be changed if the data are monotonically transformed. A set of data which can be perfectly represented by an HCS will obey not only the triangle inequality requirement of any distance measure (see A2.1.1), but also the more restrictive 'ultra-metric inequality', namely:

Ultra-metric Inequality
For all triples of distinct points (i, j, k),
$d(i, k) \leqslant \max \{d(i, j), d(j, k)\}$

In the case of perfect data, both the 'connectedness' and the 'diameter' methods give rise to the same hierarchical clustering. The HICLUS procedure and its applications in MDS were described in detail in 4.3.3.1.

6.1.7 Maximum variance non-dimensional scaling (MVNDS)

Concisely: MVNDS (Maximum Variance Non-Dimensional Scaling) provides:

internal analysis of two-way data in a lower-triangle format of a (dis)similarity measure
by a simple distance model
using a locally monotonic and variance maximising transformation of the data guaranteed to satisfy the triangle inequality criterion.

This program and its implementation are described in detail in Cunningham and Shepard (1974). MVNDS is like all the other programs in the MDS(X) series in that it rescales the data (dis)similarities into a set of corresponding distances. It differs from them in *not* producing a configuration of points in a continuous space to which the distances correspond. Indeed, the solution distances may well not be capable of being represented spatially at all. All the MVNDS scaling procedure requires is that the solution distances have two properties. They must

(i) satisfy the metric axioms, especially the triangle inequality; and
(ii) be as close as possible to being a monotonic function of the data.

The basic distance model rescales the data into a set of disparities which are *as close as possible* to being distances. In any imperfect solution it is very likely that some triples of the disparities will *not* satisfy the triangle inequality, and this is one reason why they are sometimes termed pseudo-distances. In MVNDS the scaling problem is dealt with in a different way—by severing the link to a Euclidean (or other similar) space and requiring that the actual fitting values (the rescaled data) satisfy only the triangle inequality.

As they stand, these two requirements are not sufficient to obtain a (non-trivial) solution but need to be supplemented by a third, which serves the same purpose as seeking a solution in as low a dimensionality as possible in SSA:

(iii) to maximise the variance of the distances.*

The MVNDS program maximises a goodness-of-fit index between the data and solution distances which has three component weights of the form:

$$\text{Index} = W1 - W2 - W3.$$

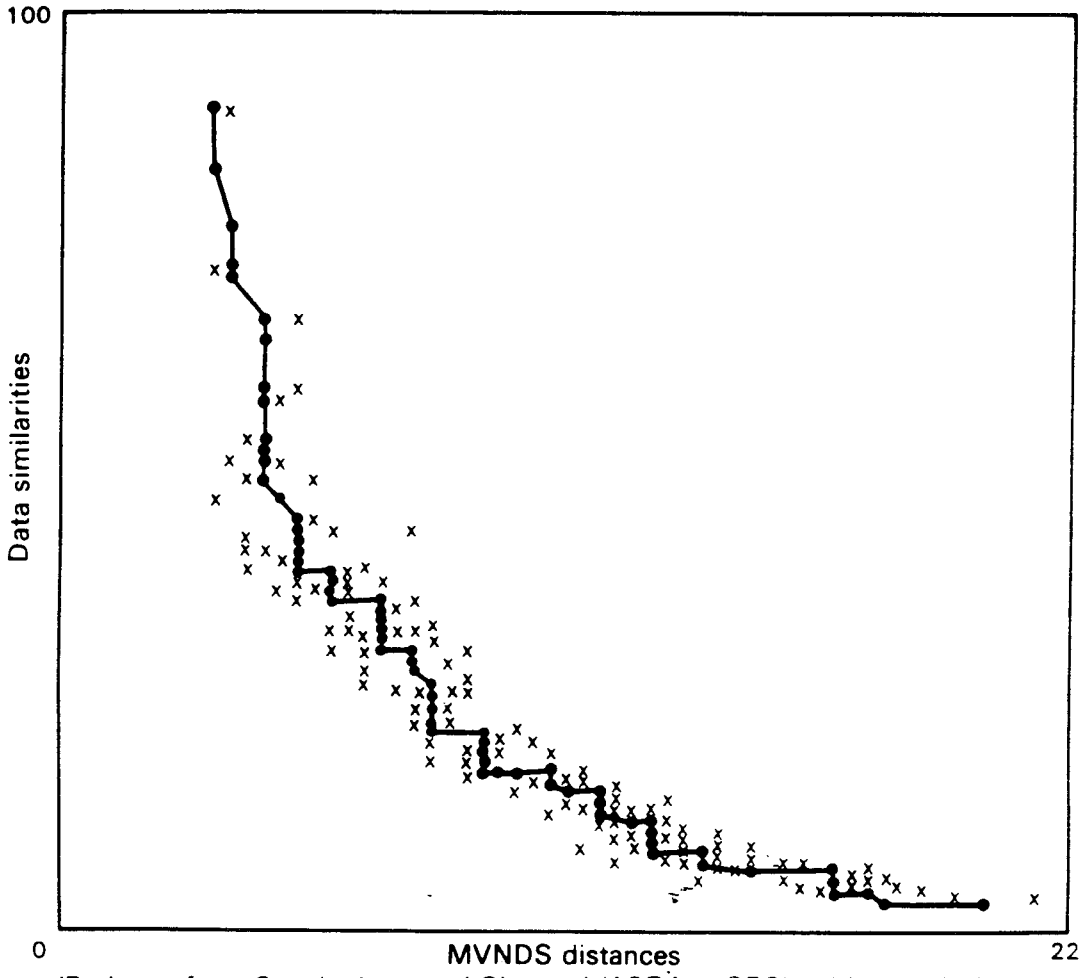
W1 represents the variance of the distances. W2 represents departure from weak monotonicity (raw stress) and W3 represents violations among the distances of the triangle inequality. (The default values give W3 thirty times the weight of W2 and one hundred times that of W1.) As the number of iterations increases, violations of the triangle inequality dramatically decrease, usually producing a set of rescaled quantities which perfectly satisfy the triangle inequality and other distance measure axioms.

From simulation studies it turns out that MVNDS is very well adapted to recovering distances derived from non-spatial structures (such as the length of paths through a graph or tree) to recovering non-Euclidean distances and to recovering distances which have been subject to a wide range of distorting transformations. (Incidentally, it does far better in this than the basic non-metric MDS model.)

The claims of the MVNDS model thus appear to be well-founded: MVNDS does a very good job at recovering a wide class of distances very accurately and is an excellent choice when the user is primarily interested in rescaling data rather than in producing a spatial configuration to mirror the data. A good deal of evidence suggests that many cognitive and semantic structures (networks, generative structures, associative graphs) and many aspects of organisations, life-histories, trade-flows, social networks etc., are better represented by non-continuous structures than by spatial models. Use of MVNDS as a preliminary step either to spatial or non-spatial representation is therefore a sensible choice in these applications.

An interesting application of MVNDS is the analysis of Rothkopf's experimental data based upon judgments of perceived similarity between 36 morse code (dot and dash) signals. He presented the 600 subjects with pairs of auditory signals (separated by a 1.4-second interval, with a gap of 3 seconds between the pairs) and asked them to say whether each pair was 'the same' or 'different'. These observations were converted into a matrix of confusion measures whose entries δ_{jk} gave the proportion of subjects saying that signal j and signal k were the 'same'. The data are presented in Shepard (1963) together with a non-metric MDS two-dimensional solution, readily interpretable in terms of (1) the number of dots and dashes, and (2) the predominance of dots over dashes. A subset of these data, restricted to 21 of the 36 signals, was re-analysed in Cunningham and Shepard (1974, p. 354 et seq.) according to the MVNDS model, with weights which ensured

*This corresponds to forcing a solution down into a lower dimensionality whilst still preserving the monotonic constraints of the data (cf. Shepard 1962). Intuitively, it helps to consider a triangle with unequal sides. Keeping the order of the distances, the variance will increase as the longest side is lengthened until it equals the sum of the other two sides—thus lying down in a one-space. In general 'when variance is forced upward by ... stretching large distances and shrinking small distances, the dimensionality of the least space that will adequately fit a set of points tends quite generally to decrease' (Cunningham and Shepard 1974, p. 341).



(Redrawn from Cunningham and Shepard (1974, p 356), with permission

Figure 6.5 *MVNDS solution to Rothkopf's morse-coded confusion data*

that the solution distances conformed to the triangle inequality, but which did not emphasise conformity to monotonicity very strongly, since these were known to be errorful data. The resulting Shepard diagram is presented in Figure 6.5. (A smaller subset of these data provide the test data for the MVNDS program in the MDS(X) series.) It is instructive to note that the monotone function is very much like a negative exponential 'decay' curve—a relationship expected on theoretical grounds and frequently found for such data (Shepard 1980, p. 391). It also turns out that the MVNDS 'distances' are in fact very close to Euclidean distances of the 2-dimensional non-metric MDS solution (correlating at 0.954).

When MVNDS is used with empirical data, it is advisable to set W3 at a high level to ensure conformity to the metric axioms, and then seek a compromise between the maximum variance (W1) and monotonicity (W2) conditions, with W2 set in the vicinity of 2.0 (see Cunningham and Shepard 1974, pp. 357–8 for discussion of suitable strategies). In any event, users are strongly recommended to obtain more than one solution, using different values of W2.

6.1.8 Simple conjoint composition models (UNICON)

Concisely: UNICON (UNIdimensional CONjoint Measurement) provides:

> internal analysis of a table of values representing the composite or conjoint effect of up to five variables (or facets) on a single dependent variable by either additive, subtractive, multiplicative (and more complex) models, using a monotonic transformation of the data.

The additive sub-model of UNICON is the one most commonly used and will form the basis of presentation here. In brief, the purpose is to find *uni-dimensional* scale values, or 'utilities', for each category of a table of two or more nominal variables ('ways' or 'facets' in Guttman's terminology) which, when combined by addition, best reproduce the rank order of the entry in the table.*

UNICON is most useful when the researcher's primary concern is to estimate the quantitative effect which each of a number of 'events' (nominal variable categories) has upon a dependent variable, and/or investigate in what way such effects combine to produce the dependent variable value. This model has a lot in common with a number of currently popular and frequently-used methods for analysis of two-way (and higher) tables of data, whose entries may be values (mean, median) of a quantitative dependent variable (e.g. the expenditure on food in households where one parent is from class x and the other from class y) or else a percentage frequency, as in the usual contingency table. Such methods include conventional analysis of variance (ANOVA) and Tukey's related, but more robust and resistant, 'median polishing' technique (Mosteller and Tukey 1977, p. 186 et seq.)† in the first case, and log-linear analysis of contingency tables in the case of percentaged data.

ADDIT can perform both types of analysis, but seeks the less restrictive *ordinal* rescaling of the table entries. It is therefore especially useful when the researcher has used a rating scale method to collect data and/or wishes to make rather modest measurement assumptions about her data. Note especially that UNICON solution scale values are uni-dimensional (one uni-dimensional scale for each way of the data)‡ and are only interval-level values, i.e. unique up to a linear transformation. In both these respects UNICON differs from other MDS programs.

Very commonly, when two-way data are analysed by conventional ANOVA, the additive model has to be extended to include an 'interaction term' to allow a separate effect for each entry in the table over and above the purely additive effect of each row and column. In one sense, the interaction effect is evidence that the simple additive model does not hold, and statisticians have frequently looked for a transformation of the data which will reduce or remove the interaction effects. Since considerable interaction effects are frequently found in empirical research (and especially in social psychological studies based upon rating data whose measurement status is obscure) it is always worth at least entertaining the hypothesis that interaction effects observed from conventional ANOVA may in fact arise from the somewhat arbitrary measurement scale of one's data. In at least one instance, ordinal additive analysis can be shown to render data perfectly additive when the authors initially claimed the reverse. Sidowski and Anderson (1967) had subjects rate the attractiveness of working in certain professions in cities of varying sorts (e.g. *doctor* in a city chosen to be *moderately high* in attractiveness as a place

*Following Roskam's convention this additive model and its iterative implementation is referred to as ADDIT, which was the name given the program in an earlier release of MDS(X). The measurement theory underlying the model (conjoint measurement) is discussed in detail in Krantz et al. (1971) and a simple overview is presented in Coombs et al. (1970, pp. 25–30).

†See Coxon and Jones (1978b, pp. 66–86) for an example of the use of median polishing in conjunction with MDS scaling procedures.

‡Strictly, this makes UNICON a method for analysing multi-mode, multi-way data, but in practice several ways refer to the same mode.

to live) and a conventional ANOVA yielded a highly significant interaction term. Krantz et al. (1977, pp. 445–7) show that an ordinal rescaling of the data yields a perfect additive fit and that:

> the interaction between city and occupation, therefore, is attributable to the nature of the rating scale, because it can be eliminated by appropriate rescaling. (ibid. p. 446)

This conclusion has been challenged in Birnbaum, 1974.

The algorithm used in ADDIT and UNICON is described in the MDS(X) documentation and the additive case is discussed in some detail in Roskam (1968, ch. VI) and Kruskal (1965). Basically, the initial scale values are provided by pseudo-random numbers, and Kruskal's weak monotone regression used to obtain fitting values. Stress_2 is then used to estimate fit, and the values are improved by a gradient procedure similar to that used in the basic non-metric model (see Kruskal 1965, pp. 261–2). Users should be on their guard against degenerate (artificially low stress) solutions which capitalise upon the *weak* monotonicity of Kruskal's procedure. The principal means of detecting such degeneracy is to compare the number of distinct values (equivalence classes) in the data, in the disparities and in the solution values. If there are significantly fewer distinct values in the disparities it is advisable to check where these occur, and, especially whether they are concentrated in any particular category of the table variables. In any event, it is probably sensible to err on the side of caution if ties occur in the data and use the secondary approach to ties.

An example of the use of the ADDIT sub-model of UNICON is given by the data on mean fertility of 1,893 married couples in a national (England and Wales) sample in 1949, originally collected by Berent (1952) and subsequently analysed by Blau and Duncan (1967) and Hope (1972). Average fertility is assumed to be the additive effect of the typical fertility behaviour of (i) the social origin class of the husband, and (ii) of his current status. In the first version, the class fertility effects are assumed to be different, so the model takes the form

$$x_{ij} = \mu + a_i + b_j$$

where μ is the overall mean fertility, a_i is the effect of the row (social origin) class category i, b_j is the (possibly different) effect of the column (present status) class category j, and x_{ij} is the predicted mean fertility of couples whose husbands came from class i and are presently in class j. (This model is coded in UNICON simply as: MODEL $A + B$.) If an additive model holds well and there are no significant interaction effects it can be asserted that the couple's fertility is simply attributable to the influence of the husband's (though why consider only the husband's?) origin and destination status and is not separately affected by distinct social mobility effects (represented by particular combinations, or moves between, class categories). Obviously, the analysis of residuals, here representing the interactions, will be central to the analysis.

The second variant (often referred to as the 'halfway hypothesis') simply assumes that the fertility effect of the classes is the same, whether they feature as the origin or the destination statuses. (In terms of UNICON, this simply constrains the a_i and b_i effects to be the same, and is simply coded as: MODEL $A + A$). In this

case, unlike the first hypothesis, the effects will be symmetric—the predicted fertility will be the same whether the husband has moved from i to j or from j to i.

Using conventional ANOVA, HOPE (1972, p. 98) concludes:

> Both analyses result in the rejection of the halfway hypothesis and in the acceptance of an alternative which involves an effect of direction of social mobility on fertility, the downwardly mobile being more fertile than the upwardly mobile, with the non-movers in the middle.

If the fertility data are assumed to be ordinal (hence leading to a more resistant analysis where more extreme values have relatively little effect), do the same conclusions emerge?

The scale values for the two models and the tables of residuals are given in Table 6.3. The scale values are given in the row and column stubs, and the entries within the table give the residuals $(d - \hat{d})$ from the monotone additive model.

I MODEL: SAME EFFECTS
('Halfway' or symmetric mobility effects on fertility)
(MODEL: $A + B$, with $A = B$ $\text{Stress}_2(\hat{d}) = 0.263$)

SOCIAL CLASS OF:		DESTINATION I	II	III	IV	*Row Effects* a_i
ORIGIN						
	I	0	−0.857	0	−0.141	*−0.252*
	II	0	+0.138	+0.589	0	*1.153*
	III	0	0	0	0	*1.194*
	IV	+0.267	0	−0.542	+0.541	*2.277*
Column Effects:	b_j	*−0.252*	*1.153*	*1.194*	*2.277*	

II MODEL: DIFFERENT EFFECTS
(Asymmetric mobility effects on fertility)
(MODEL: $A + B$, with $A = B$: $\text{Stress}_2(\hat{d}) = 0.160$)

		DESTINATION I	II	III	IV	Row *Effects* a_i
ORIGIN						
	I	0	−0.140	+0.088	0	*0.504*
	II	0	0	+0.140	−0.165	*0.814*
	III	−0.084	+0.028	+0.008	+0.069	*1.355*
	IV	−0.003	+0.118	−0.219	+0.150	*1.435*
Column Effects:	b_j	*−0.141*	*0.911*	*0.881*	*1.250*	

Table 6.3 UNICON *(monotone additive) analysis of Berent's fertility data* (Solution scale values are given in the row and column stubs. Entries in the tables are the residuals $(d - \hat{d})$ from the monotone additive model.)

Compared to Hope's linear analysis of the same data (see especially Hope 1972, pp. 88 and 98–9), the conclusions from the monotone analysis presented here are markedly different. The order of the scale values is the same, as we should expect, and the linear correlation of scale values from the linear and monotone analyses is quite high ($r = 0.90$), but analysis of the residuals shows a rather different structure. In particular, there is no consistent pattern in the sign of the residual values according to whether the subjects are upwardly or downwardly mobile (Hope expects positive residuals for the latter and negative values for the former) and the sources of significant interaction are different.

In this example the differences between a linear and monotone analysis of such data should not be taken too seriously—after all, a linear analysis makes obvious sense for a variable such as the average number of children. But when the dependent variable is based on ratings data, as is usually the case, the effects due to the variable and the purely artefactual effects due to the arbitrary nature of the rating scale are likely to be confounded. In such instances it is eminently sensible to rely primarily on a monotonic analysis which makes no assumptions about the interval nature of the data.

In addition to ADDIT analysis, UNICON provides for subtractive models (of the form: $X = A - B$) and multiplicative models (of the form: $X = A \times B$), and more complex mixed models, such as: $X = (A \times B) + C$.

6.2 Two-Mode Data

In two-mode data the input data matrix represents *two* (usually distinct) sets of objects, each of which is represented separately in the solution. The data matrix is normally rectangular (since the number of row and column elements is usually different), asymmetric, and row-conditional, in the sense that each row of the data is fitted separately and data comparison between rows is considered illegitimate.

In the case of simple composition models, the UNICON program discussed in the previous section is the appropriate model.

In other cases, (vector or distance models) a further distinction must be made between whether the analysis is internal or external (see 5.1.2). In most applications, the column elements represent the objects or stimuli and the row elements represent individual sources of data (often subjects), and the number of columns is much less than the number of rows. Consequently the positioning of the objects will depend very heavily upon satisfying the constraints of as many individuals' data as possible according to the assumptions of the model. Therefore great care should be taken to interpret the stimulus configuration with explicit reference to the location of subject points (or vectors). This matter is taken up below.

By contrast, in an *external* analysis, one part of the configuration—usually that part representing the objects (columns)—is provided by the user and (usually) remains fixed during the analysis. The data are then used to locate just the subject points within the stimulus configuration.

6.2.1 External mapping by the vector model (PREFMAP IV *and* PROFIT)

Concisely: PREFMAP (PREFerence MAPping) (Phase IV) provides:

external analysis of two-way, row-conditional data
by a scalar products model,
using either a monotonic or a linear transformation of the data.

Concisely: PROFIT (PROperty FITting) provides:

external analysis of a configuration
by a set of property ratings or rankings in row-conditional format
by a scalar products model,
using either a linear or continuity transformation of the data.

Given

(i) a fixed configuration of p stimulus points in a specific number of dimensions, and

(ii) a rectangular data matrix of N rows (one for each subject or property) and p columns (identical to the configuration points),

these two programs map each of the N subjects/properties into the stimulus configuration as a vector, pointing in the direction in which the data values—preferences, property values, similarities—are increasing. For convenience, all the vectors pass through the centroid (centre of gravity) of the stimulus configuration. (Vector representation was discussed earlier in 4.4.1.)

What differentiates the two programs is simply the rescaling transformation option which can be requested:

(1) PREFMAP-IV

The rescaling transformation can either be *linear* or *monotonic* (strictly speaking, quasi-monotonic, meaning that the procedure first performs a linear transformation and then goes on to a set of iterations using monotone regression). If the monotonic option is chosen, three further options exist, depending on whether ties exist in a subject's row or data:

(a) no ties exist (FIT(1)).
(b) ties are treated as equalities (secondary approach, FIT(2)).
(c) ties are treated as indeterminate (primary approach, FIT(3)).

(2) PROFIT

The rescaling transformation can be either *linear* or a *simple continuity* (kappa) function (see 5.2.2.1). In the former case, linear PROFIT is formally equivalent to PREFMAP-IV with linear fitting, although the algorithm used to produce the solution differs somewhat.

External mapping by a vector model is extensively used as a means of interpreting configurations and identifying dimensions; these uses were discussed in section 4.4.1 at some length.

External mapping is also often used to map subjects' preferences into an already obtained stimulus space. There is much to be said for analysing preference data in this way, since the evaluation of things may often be quite distinct from the cognition or adjudged similarity of things. We may agree entirely on the

characteristics and relative similarity of a set of regimes, people, foods, books, but disagree violently on their merits or over which we prefer.

A full discussion of such analyses and a number of examples of applications of this sort is provided in Carroll (1972, especially pp. 130–46).

External mapping is also particularly useful when the researcher wishes to represent information within an already known or physical configuration—such as a geographical map or the plan of a machine assembly. In this way, the directions can be estimated in which a particular plant species increases in a botanical area, or social deprivation increases in a city, or strains increase within a machine, or travel priorities lie in a country, can be readily assessed. It is also possible, using the other distance sub-models of PREFMAP, to see whether these data would be better represented as a point—an 'ideal point'—and if so, what sort of more complex distance might best fit the data.

The significant information which should be attended to in an external vector analysis is, first, how well a particular row of data (subject, property) fits in the configuration. (The goodness of fit depends, of course, on the transformation chosen: a linear relation is more restrictive than a monotonic one and therefore a vector of data is bound to fit better—or at least no worse—monotonically than linearly.) Given acceptable fit, the most important information conveyed by a vector is simply its direction (in none of these programs is the length of the vector relevant) and in comparing vectors, the angle of separation is crucial and represents the correlation between the subjects' data. Communality or concentration in a set of vectors, akin to the cluster in distance representations, will show as a tightly bound sheaf, and divergence or difference is signalled by empty sectors.

A brief illustration of the results of using PREFMAP-IV for the analysis of the occupational data described above is given in Figure 6.6. The ratings of (a) social usefulness and (b) earnings averaged within nine sub-groups of subjects were mapped into the configuration (see Figure 4.6) obtained from scaling the similarities data. (In each case, the analysis of variance in directions of the vectors gives a significant difference between the groups of less than one per cent: see Table 4.16 and 4.20 in Coxon and Jones 1978a.) Note, first, that the vectors are heavily concentrated, in each case, within a small sector. The average 'social usefulness' vector is located in a NNE direction (at 62°) and the variability in the groups' vectors is contained within 34°; nine-tenths of the unit circle is empty. In this direction, the people-oriented, but low paid occupations ('vocations') are located, followed by less educated/lower paid and more educated/better paid through to the least socially useful. Notice the high degree of agreement between disparate groups—student teachers, engineers and clergy agreeing almost entirely (at least insofar as group averages go). By contrast, the average 'earnings' vector orders the occupations in an entirely independent manner (separated by just over 90°, representing a correlation of virtually zero) in a SE direction, but again with little spread, the extreme vectors subtending an angle of merely 30°. However, in this instance there is a detectable difference between the professionals' (more accurate) estimate and the working class groups' estimates, which tend to underestimate the professional occupations' remuneration compared to their own. It should be emphasised that present data are averages: the same conclusions do not necessarily hold when examining individual data.

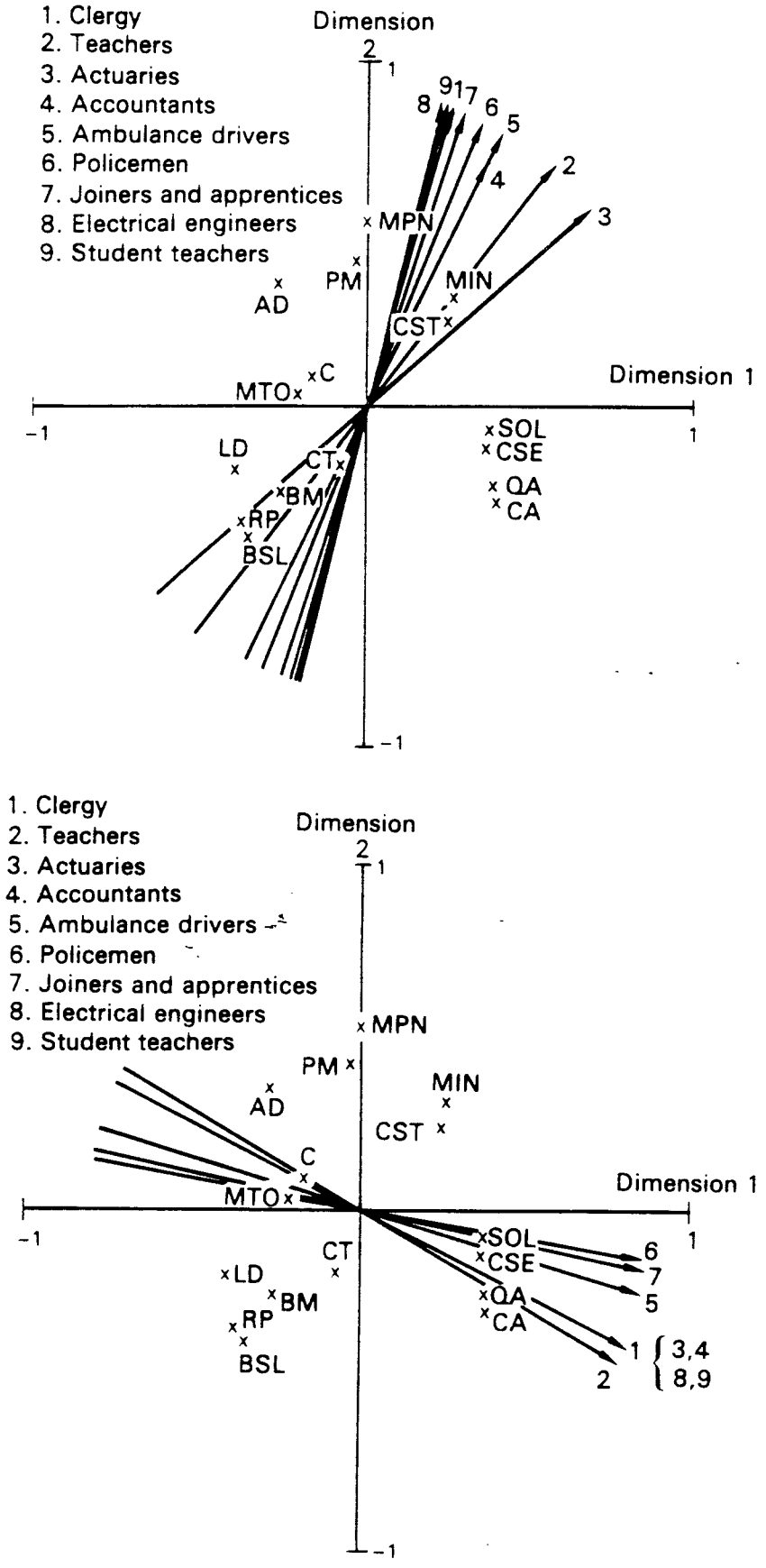


Figure 6.6 *PREFMAP-IV: external mapping of two criteria into INDSCAL configuration*

6.2.2 Internal mapping by the point-vector model (MDPREF)

Concisely: MDPREF (MultiDimensional PREFerence Scaling) provides:

internal analysis of two-way preference data in the form either of a row-conditional matrix or of a set of paired comparisons matrices
by a scalar products (point-vector) model,
using a linear transformation of the data.

Note that MDPREF is an *internal* form of analysis, positioning stimuli points and subject vectors simultaneously from the data, and is a *linear* procedure: data are assumed to be at the interval level of measurement. As a form of two-mode factor analysis, MDPREF is becoming increasingly popular for analysing preference data, personal constructs rankings* and semantic differential ratings data. Since the solution is analytic rather than iterative, it is a computationally cheap and efficient procedure and the results are often a good deal more stable than for other internal two-mode models such as multidimensional unfolding (MINIRSA, q.v.).

From the user's point of view the main difference between the external (PREFMAP-IV) and internal (MDPREF) vector models is in the way of interpreting the solutions. In the external case, it will be recalled, subjects are located within a *fixed* reference configuration and the location of subjects' vectors could with some confidence be referred to or interpreted in terms of the stimulus locations. For internal analysis this is not true: the stimulus points are located in such a way that as many as possible of the subjects' data are fit well and the stimulus configuration can only be 'read' by direct reference to the location of the subject vectors.

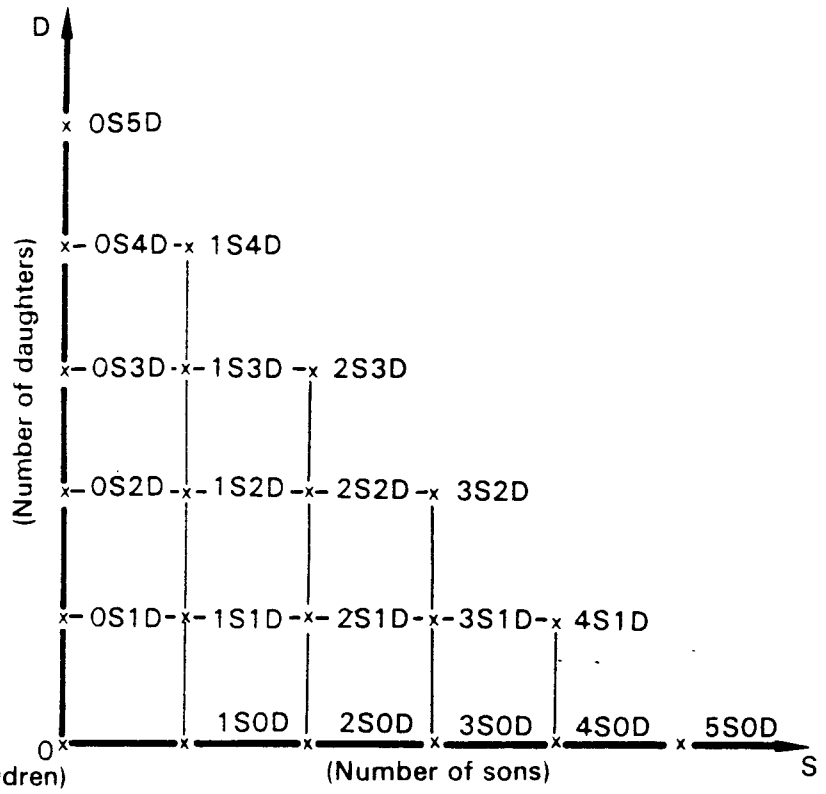
The MDPREF algorithm is described in the MDS(X) documentation and in Carroll (1972, pp. 123–9). Basically, the algorithm forms the major and minor product moment matrices from the original rectangular data matrix (called the 'first score matrix' in the MDPREF terminology) and obtains the latent roots of those matrices. (These give a good estimate of the 'true'—or at least the lowest acceptable—dimensionality of the data.) The location of the stimulus points and subject vectors are then found by producing a factoring or decomposition which gives a 'second score' matrix which best fits the data in the number of dimensions chosen by the user. The model has already been discussed above in 5.3.2.

Issues in interpretation and application of the MDPREF program are best illustrated by reference to what is now a quite well-known data set: the Bollen-Delbeke data on family size and composition preferences. It will also be used to illustrate the corresponding distance model analysis (MINIRSA).

In 1960 Bollen collected data from psychology students at the Catholic University of Louvain on their preferences for families, which differed in terms of the number and sex of the children. In all, 21 such stimuli, (family size compositions) were defined—all possible compositions from no children up to families of size five. These are illustrated in Figure 6.7. Each subject was then given each of the 210 pairs of stimuli, e.g. (3 sons, 2 daughters) *vs* (1 son, 1 daughter), and asked which he or she preferred. The data for 80 subjects (40 male, 40 female)

*In many ways MDPREF resembles (and is indeed superior to) the INGRID program developed by Slater (1960) for the analysis of repertory grid data, frequently used in personal constructs analysis in psychology and sociology. See Tagg (1979).

formed the basis for subsequent analysis, and were analysed by Bollen and later by Delbeke (1968) and Coxon (1974). The data exist both in the original form as a set of 80 pair-comparison dominance (0, 1) matrices and as a set of 80 preference



(a)

Summary information on rank scores for different family size compositions*

Stimuli	Code	Range	Mean	Variance
0 children		0–4	0.2	0.4
1 son	1S0D	0–12	3.6	7.3
2 sons	2S0D	2–17	8.1	13.1
3 sons	3S0D	5–16	9.7	6.6
4 sons	4S0D	4–18	10.0	9.6
5 sons	5S0D	0–20	8.7	20.6
1 daughter	0S1D	1–11	2.6	4.4
2 daughters	0S2D	2–13	5.6	7.7
3 daughters	0S3D	2–12	6.2	5.0
4 daughters	0S4D	1–13	5.8	5.9
5 daughters	0S5D	0–13	4.8	11.8
1 son, 1 daughter	1S1D	4–19	11.0	17.5
2 sons, 1 daughter	2S1D	10–20	15.1	6.1
3 sons, 1 daughter	3S1D	10–19	15.9	3.5
4 sons, 1 daughter	4S1D	5–20	14.8	12.6
1 son, 2 daughters	1S2D	6–20	12.5	7.5
2 sons, 2 daughters	2S2D	13–20	17.9	2.3
3 sons, 2 daughters	3S2D	12–20	18.4	4.6
1 son, 3 daughters	1S3D	2–18	12.0	8.7
2 sons, 3 daughters	2S3D	9–20	17.1	5.6
1 son, 4 daughters	1S4D	1–18	10.1	17.5

(b)

*Highest preference is a rank of 20, and lowest preference has a rank of 0.

Figure 6.7 *Delbeke family size and composition data: summary*

rankings, formed by summarising the rows of each dominance matrix to produce a 'vote count' preference order, since the subjects were remarkably consistent.* A basic summary information on the rankings data is provided in Figure 6.7b. Several points should be noted. First, the 'no children' stimulus had so little variation (virtually everyone preferred it least) that it was removed from subsequent analysis. Its inclusion led to the structure in scaling solutions being distorted, and the universal dislike of no children meant that its location was highly unstable, varying from run to run—being located at any position so long as it was maximally distant from ideal points in the distance model, and maximally opposite in direction to the preferred regions in the vector case.

Secondly, there is marked preference for large, mixed families, with the composition of 3 sons, 2 daughters having highest overall preference. The data, it should be remembered, were collected from unmarried Catholic students before the changes in attitude to and practice of birth control following Vatican II. Other characteristics are also evident, and are recognisable in the scaling solutions in differing ways, depending on the model.

(i) For each given single-sex family size, all-boy families are preferred to all-girl families, and the difference in average preference of boys increases systematically with the overall size of family.

(ii) For every family size, a mixed-sex composition is preferred.

(iii) A preponderance of boys is preferred in mixed-composition families.

MDPREF was applied to both the preference scores and the pair-comparison data. Solutions were sought in two and three dimensions, for men and women subjects separately, though only that for the male subjects is reported here. The preference score and pair comparisons data produce almost identical solutions in each case. Inspection of the roots of the first-score matrices strongly suggests that a two-dimensional solution is adequate—the percentage of variation accounted for by the first four dimensions is 75 per cent, 14 per cent, 6 per cent, 1 per cent, so 89 per cent is concentrated in the first two dimensions and little is gained by adding any subsequent dimensions. The two-dimensional configuration for males is presented in Figure 6.8. (The female data form the test data for the MDPREF program in the MDS(X) version.)

First, let us concentrate on the location of subject vectors, recalling that for a two-dimensional MDPREF solution their termini (end points) are normalised to unit length and will therefore lie on a circle. However, virtually all subjects' vectors occupy just over one-quarter of the circle, indicating a quite high degree of consensus in their preferences. In internal analysis the position of subject vectors should take priority in interpreting the joint space. It is often helpful to begin by locating an average subject and reading back along from the vector end, through the origin of the space to the other side of the circle, noting how the stimuli project onto it. (The location of the dimensions is of course arbitrary in a vector model so they do not need interpretation, but the origin is significant as the centroid of the stimulus points and the point through which all subject vectors pass.) In this case

*See Coxon (1974, p. 197, ftn 11) for details of tests of consistency. The average coefficient of consistence was 0.94.

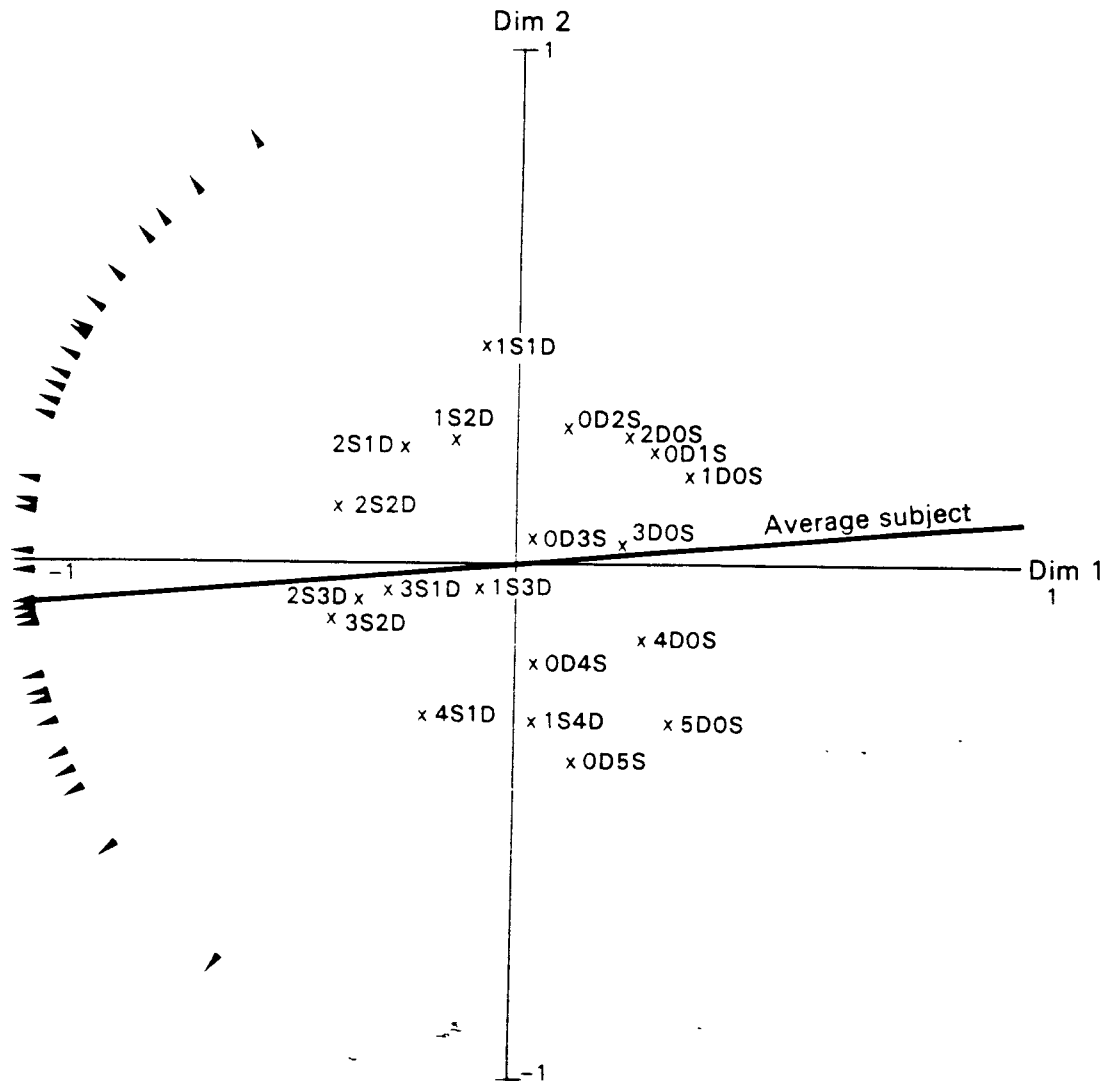


Figure 6.8 *MDPREF analysis of family size and composition data 2-D, males only*

the most popular stimulus (3S2D) is the first, most preferred one, followed closely by other fairly large, mixed compositions, then into larger single-sex families, and finishing with the least preferred single son and single daughter compositions. (The projections mirror fairly accurately the mean ranks given in Figure 6.7b.)

Secondly, what do the differences in vectors signify? The subject at one extreme (located in the NNW direction) clearly prefers much smaller families and is not by any means as concerned with the balance between sexes. By contrast, the subject at the other extreme (in the SSW direction) is greatly in favour of very large family size whether mixed in composition or not. On the whole, virtually everyone prefers mixed to unmixed composition, and the greatest variation is on the size of the family composition.

The structure of the stimulus configuration, then, should be interpreted by reference to the subjects' vectors in internal analysis and it is important to realise that the 'natural' lattice-like structure of the stimuli (see Figure 6.7a) cannot be

discerned in the stimulus configuration. (Try connecting the sons' axis, denoted on Figure 6.8 as OD1S, OD2S, OD3S, OD4S, OD5S, and the daughters' axis and you will note that they follow a roughly parallel and non-linear sequence, with a major shift in direction at OD2S and 2DOS. In no way can the lattice structure be discerned in this configuration.) Nonetheless, a very coherent structure *is* evident, which happens to be part of a radex (see 4.5), and this is illustrated in Figure 6.9. A semi-circle can be drawn which divides the *mixed* from the single sex or *unmixed* family compositions, and a set of lines can be drawn emanating from the 'centre' which divide the space into sectors corresponding exactly in this case to the overall *family size*. Another way of describing the radex (Shepard 1978, p. 57 et seq.) is to view the structure as having polar co-ordinates, with latitude (distance from the centre of the radex located in an approximate way in Figure 6.9) corresponding to the degree of mixedness, and incidentally to the degree of preference, and the angular position around the perimeter corresponding to the size of the family. Moreover, the three characteristics noted above from the initial data analysis can

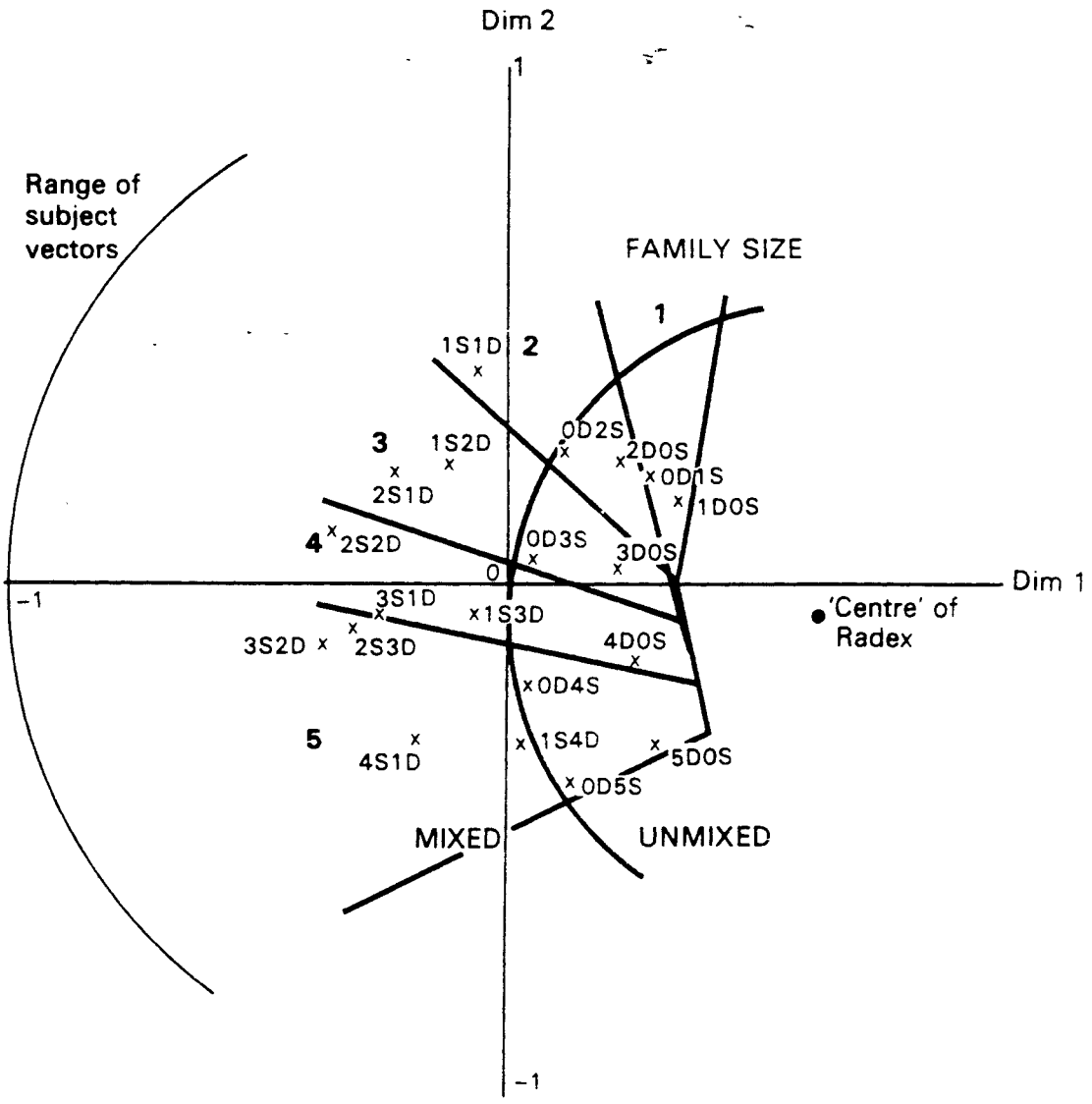


Figure 6.9 *Radex structure of family preference MDPREF solution*

in most cases also be 'read out' in a manner consistent with the form of the vector model:

(i) *All-boy families are preferred to all-girl families, for given single-sex family-sizes.* Concentrating on single-sex pairs of fixed family-size, the points representing all-boy families are systematically to the left of (more preferred than) all-girl families of the same size.

(ii) *A mixed sex-composition is preferred with a given family-size.* Within each given family-size the points representing mixed families are consistently to the left of (preferred to) those representing unmixed family composition.

(iii) However, the inference that *within* a given type of mixed composition family, a preponderance of boys is preferred does not seem to be detectable.

(The same characteristics will be recognisable, but in a different and apparently unrelated representation, when the distance (unfolding) MDS solution is given later in section 6.2.3.)

A final caution in the use of MDPREF. The program allows users to remove either the row effects (individual response-style) and or the column effects (removing the 'mean utility' or consensus) of the first-score data matrix. The removal of row effects rarely has important consequences, but the removal of column effects will produce a much wider dispersal of subject vectors, and individual differences then become the major focus of the analysis (see Heiser and de Leeuw 1979, p. 28 et seq.). On the other hand, *double*-centring—removing both column and row effects—actually turns the vector model into a distance model and typically leads to an over-estimation of the appropriate dimensionality (see Carroll, 1970, p. 278). It should be used with considerable caution, if at all.

6.2.3 *Internal mapping by the distance model* (MINIRSA)

Concisely: Multidimensional Unfolding provides: MINIRSA (Rectangular Space Analysis) or

internal analysis of two-way data in a row-conditional format of a (dis)similarity measure
by a Euclidean distance model
using a monotonic transformation of the data.

The basic multidimensional point-point distance ('unfolding') model was described in 5.3.3.1, and the algorithm is described in the MDS(X) documentation and in Roskam (1979, pp. 300–4). MINIRSA positions each subject *as a point* (the 'ideal point' or single point of maximum preference) in a joint space with the stimuli, also represented as points, such that the rank order of the distances from the ideal point to each of the stimuli is as close as possible to being in the same rank order as the subject's preferences (or other similar) data.

In the current MDS(X) series a metric variant of multidimensional unfolding is possible using the 'quasi-internal' options* of PREFMAP-III.

*INIT (0), FIT (0) and S-PHASE (3). See 7.5.5.2.

The *non-metric* internal analysis of rectangular data by multidimensional unfolding is implemented by the program MINIRSA. When used with few data or on data with little variation, however, the solution is not likely to be well constrained, and the algorithm is particularly subject to local minima. It also tends to be expensive in terms of computer time, requiring a large number of iterations to achieve satisfactory improvement. Despite these inherent problems of non-metric unfolding, MINIRSA is generally a useful program so long as the data are sufficient (at least 30 subjects and 8 stimuli is a rule of thumb for a 2-dimensional solution), with variation in the ranks of each stimulus.

Once again, the program can be illustrated by reference to the Bollen-Delbeke data. MINIRSA minimises $stress_2$ and the overall values for 3 dimensions was 0.069, and 0.124 for 2 dimensions. The 2-dimensional configuration is presented in Figure 6.10.

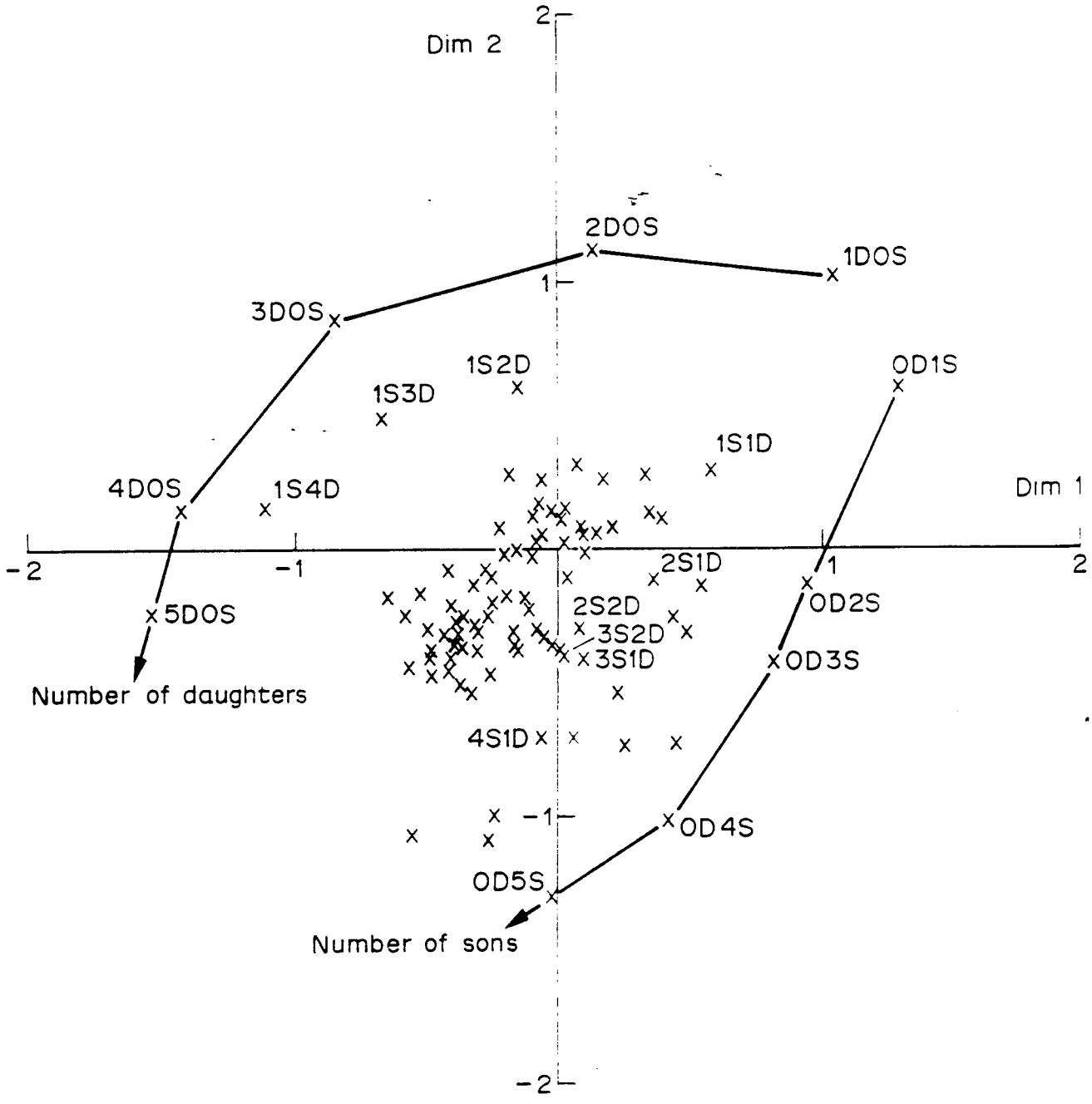


Figure 6.10 *MINIRSA analysis of family size and composition data*

Because it is an internal analysis, it is important to begin by examining the subject points—located in a main cluster at the centre of the configuration (representing preference for largish mixed families), with stragglers to the SE (representing those who prefer a preponderance of boys).

Significant information is additionally conveyed by the occurrence of empty spaces in a configuration. In the vector model, we saw that the fact that parts of the unit circle (or sphere) do not have subject vector ends indicates that no preferences increase in these directions. In the distance case, we need to take note of the *regions* within which ideal points do not occur. This can be done explicitly by constructing the isotonic regions from the stimulus configuration (see Figure 5.5) and then looking at the regions in which ideal points are concentrated and at the regions which do not contain any ideal points.

The stimulus configuration for these data is quite different from the MDPREF one, but can be interpreted in a similar manner. First, the stimuli do form a distorted version of the defining lattice of Figure 6.7a, though the component dimensions (drawn in the figure) are not quite linear or at 90 (uncorrelated); rather, they are pulled in towards one another to enfold the ideal points. Nonetheless, a linear PROFIT fits both 'number of sons' and 'number of daughters' properties with a correlation of around 0.93. The main distortion in the configuration is the way in which the larger family-sizes equally mixed in composition virtually collapse onto one point, and the program deals with them by locating them as close to as many ideal points as possible. Here is a further example of how internal analysis is likely to contort a stimulus structure to satisfy the subjects' data better.

What of the three characteristics of the data? How can these be read out of the configuration?

(i) *All-boy families are preferred to all-girl families.* This is evident in the way in which a number of ideal points congregate closely to the Number of Sons line, whilst none are located very close to the Number of Daughters line.

(ii) *Mixed sex composition is preferred.* The main concentration within the swarm of subject points lies almost exactly at 45° counter-clockwise inclination, which represents the mixed composition line of the stimuli.

(iii) *Within a given family size, a preponderance of boys is preferred.* This also is discernible in the location close to the Number of Sons line.

Although the form of representation is different, the same content can be read out of both the vector and the point representation, but as we have seen, the researcher has to pay especial attention to the *joint* space and the assumptions of the model: the stimulus configuration is not likely to be accurately recovered where there is not enough variation in the data. In any event, it would be advisable to obtain a separate estimate of the stimulus configuration and then map preferences into it, particularly in the case of data based on human judgments of this sort.

6.2.4 *External mapping by the distance model* (PREFMAP-III)

Concisely: PREFMAP (PREFerence MAPping) (Phase III) provides:
external analysis of two-way, row-conditional data
by a simple Euclidean distance model,
using either a monotonic or a linear transformation of the data.

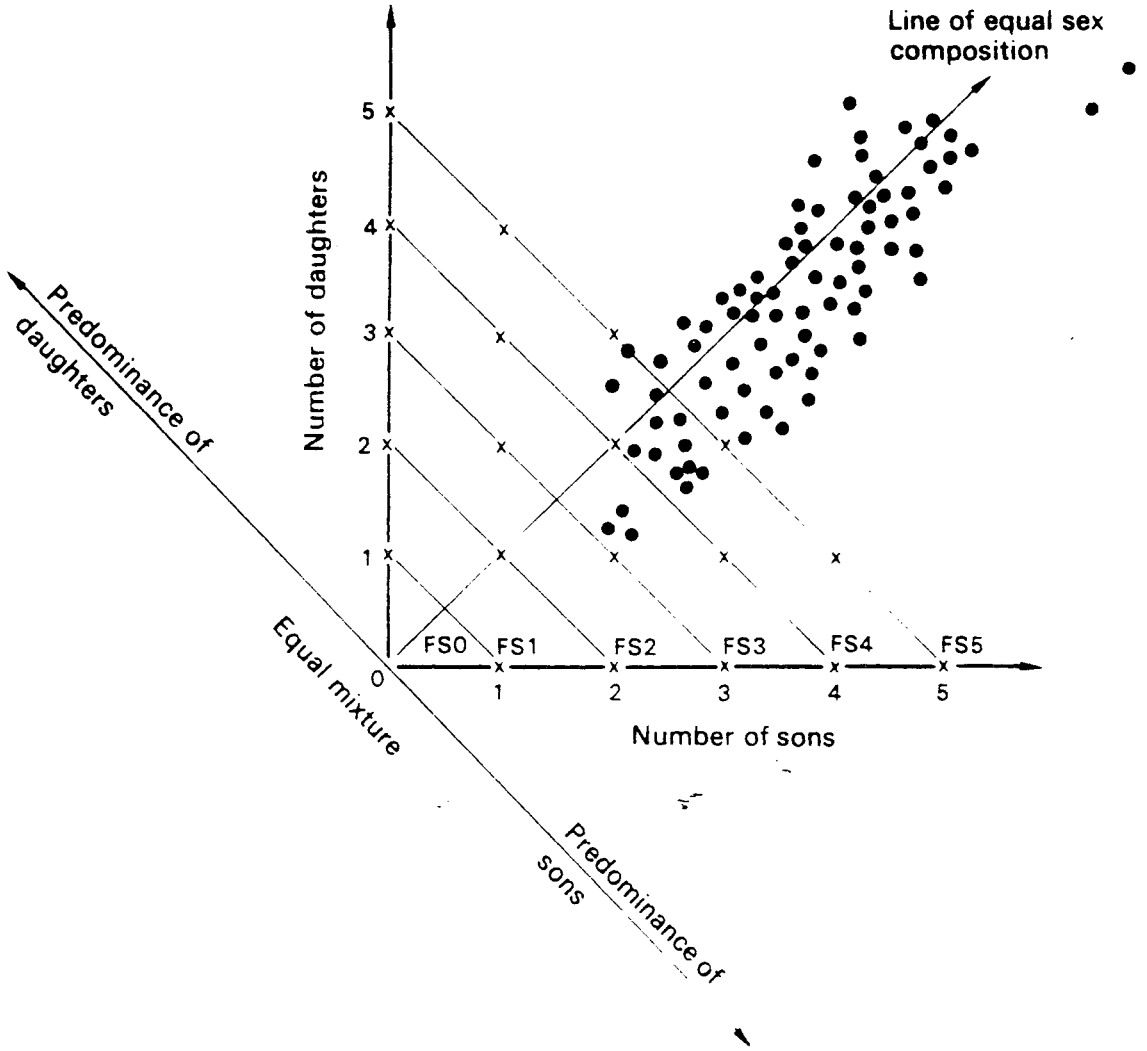


Figure 6.11 *PREFMAP-III (external distance model) analysis of family preference data*

In section 4.4.2 we introduced the idea of representing a ranking (or set of ratings) as a *point* in a configuration in such a way that the relative distances from the new point to the existing ones correspond to the data ranking. Such data might consist of a set of preference rankings or ratings, which are embedded as a set of points (of maximum preference) in an already existing configuration, obtained, say, by scaling an independently-obtained set of data on the similarities between the same objects. This model is implemented in PREFMAP-III and allows for either linear or monotonic transformation of the data. It is also an external analogue of Coombs' multidimensional unfolding model (Coombs 1964, pp. 140–9) discussed in the last section.* As in other instances of external scaling, the fixed input configuration could come from any source: a configuration obtained in previous research (to make findings compatible); a configuration obtained from a *sample* of

*Yet more general distance models, forming PREFMAP-I and PREFMAP-II, are discussed in the next chapter. Since the PREFMAP program implements an hierarchical family of four models of decreasing generality, users may wish to map their data externally according to more than one model. This is effected very simply by choosing a starting phase or model by the S-PHASE parameter and the ending phase by F-PHASE. The program includes approximate ANOVA tests to help the user decide on the appropriate level of complexity. If the user decides to start at a higher level than III, she should be aware that points may have a 'pessimal' or least preferred value on one or more dimension. The meaning of such negative weights is discussed in Carroll (1972, pp. 121–3).

one's subjects (to allow a larger number of subjects' data to be scaled); a 'rational configuration' implicit in the stimuli (as in the Bollen-Delbeke data), and so on. In this case, the current example can be continued, using the configuration of Figure 6.7a as the framework for analysis. Here the prime intervals (difference between N and $N + 1$ sons/daughters) are treated as equal in size and the stimuli are spaced so that both sons and daughters have the same interval width. The resulting mapping of the full 80 subjects is presented in Figure 6.11.

As can readily be seen, the vast majority of subjects prefer families of mixed (and almost equal) sex composition, with differences in preference for overall size of family. This we already know: what the rational configuration helps us see in addition (because of the equal interval spacing) is that there is far greater variation in family *size* preference (the line of equal sex composition) than in the *composition* of the family (the line of mixture/predominance). The fit of the data is also very good. For this metric solution, the overall root mean square correlation between data preferences and distances of the solution is 0.94 (compared to 0.64 for the metric PREFMAP-IV external vector model). There can be little doubt that the simple distance model fits the data well.