Recent developments in three-way analysis: A showcase of examples

Kroonenberg, P.M.; Klar R., Opitz O.

In R. Klar & O. Opitz (Eds.)
classification and Knowledge Organisa

Recent Developments in Three-Way Data Analysis: A Showcase of Methods and Examples

P. M. Kroonenberg

Department of Education, Leiden University
Wassenaarseweg 52, 2333 AK Leiden, The Netherlands

Summary: In this paper an compact idiosyncratic overview will be provided of the areas into which three-way data analysis has expanded. The historical introduction will be followed by a scheme presenting an indication of the techniques involved. Then four condensed examples will give a feel of the scope of applications, while the final section is devoted to publicly available programs to perform the analyses.

1. Historical overview

Three-way analysis of continuous data originated in psychology during the *sixties*. Its founder Tucker (e.g. 1966) conceived the basic ideas for three-mode component and factor analysis models, developed algorithms to estimate the parameters, and published several, primarily psychological, applications. During the *seventies* three-way analysis expanded into multidimensional scaling primarily due to Carroll (e.g. Carroll and Chang (1970)), while Harshman (e.g. Harshman and Lundy (1984a,b)), working with similar models, extended the scope of component models. At the same time Bentler and co-workers (e.g. Bentler and Lee (1979)) developed structural-equation formulations for the three-mode common factor model. In the *eighties*, Kroonenberg and De Leeuw (1980) developed new algorithms for Tucker's models, and the former extended interpretational aspects of the technique and published several applications emphasizing interpretation, primarily within the social and behavioural sciences (e.g. Kroonenberg (1983a)). Flury's (1988) book summarises the development of his work on multi-sample common principal component methods. In France a new approach called STATIS (Structuration des Tableaux à Trois Indices de la Statistique) was developed by Escoufier and co-workers (especially, Lavit (1988)). Carroll, Arabie, DeSarbo and co-workers (e.g. Carroll and Arabie (1983), DeSarbo and Carroll (1985)) took three-way analysis to the clustering and unfolding domain, and Basford and McLachlan (e.g. 1985) introduced three-mode clustering. Carroll et al. (1980) also introduced constraints into three-way analysis. In the mean time the methods gradually filtered into other disciplines, like agriculture (e.g. Basford et al., 1991), ecology (e.g. Beffy (1992), spectroscopy (initially independently from the developments in psychology; e.g. Leurgans and Ross, 1992), and chemistry (e.g. Smilde, 1992). In the first part of this decade, much effort has gone into developing constrained three-way analyses especially from Kiers, Ten Berge and co-workers (e.g. Kiers, 1991, 1992; Krijnen, 1993). Franc (1992) extended French linear vector space thinking to the three-way area. Another strand is the expansion of the techniques into the analysis of three-way contingency tables (e.g. Carlier and Kroonenberg, 1996) and three-way analysis-of-variance designs (e.g. Van Eeuwijk and Kroonenberg, submitted).

The names mentioned above are an unfair selection of some of the protagonists, and many other persons have contributed to the developments in this area. An annotated bibliography upto 1983 is Kroonenberg (1983b) and the sequel to this ever-expanded bibliography is available from the author which covers virtually the whole field, both theory and applications, and everyone is invited to contribute papers, programs, and applications.

2. Three-way data and the method chart

2.1 Three-way data

In their basic form, most data seem to come in one of three broad classes, i.e. *profile data* (scores of subjects on variables), *similarity data* (judgements of similarity between two stimuli), and *preference data* (rankings of subjects on variables); here, the terms subjects, variables, and stimuli are used as generic terms. From these basic forms many other forms can be derived, such as means, covariances, frequency counts, etc., while data of one class can be transformed into another class, e.g., indirect similarities can be derived from profile data. Profile data are the bread-and-butter data in many substantive areas, similarity ratings are very prominent in several areas in psychology and marketing, while preference data enjoy limited popularity, probably because they are far more difficult to analyse and carry relatively less information. A further distinction is that between data with a *dependence structure* and those with an *interdependence structure*. With the former kind, the (mostly profile) data are split up in two (or more) groups, one of which is designated as the set of variables to be explained or predicted from the other set, while in the latter kind there is no such distinction, and the interrelationships between the varables is the object of study. A final distinction is that between 'true' three-mode data (or *repeated measures data*), which are fully crossed, barring missing data, and multiple-sets data. The former type of data are the 'true' or 'ordinary' three-mode data, because three different kinds of entities or modes are generally involved, one of which might be a time mode. When a *time mode* is involved, many techniques discussed here do not explicitly include this design feature in their analysis, but rather use time as an interpretational feature. True multivariate time series models within the Box-Jenkins approach (e.g. Chatfield, 1989), while falling essentially within the present framework, are not dealt with in this paper. *Multiple-sets data* are generally two-mode three-way data derived from raw data, rather than consisting of the raw data themselves. Typically one may think of cross-product matrices, covariance matrices, and often (dis)similarity matrices. Such data cannot always be analysed in their raw form with three-way methods, because they originate, for instance, from different samples, but after transformation to two-mode three-way data they can be analysed with such methods.

2.2 Three-way methods

With respect to methods, one may distinguish between those that explicitly use a *stochastic framework*, rely heavily on distributional assumptions, and therefore more or less by default treat subjects as replications. The other group consists of techniques which are primarily *data-analytic* in orientation, can be considered to address populations, and pay attention to *individual differences*. A further distinction that is often useful is that between *direct modelling* and *indirect modelling* techniques (e.g. Kruskal, 1984). The former techniques attempt to model the three-way data directly, while the latter techniques try to fit models to derived three-way matrices, such as covariance (correlation) matrices or cross-product matrices, and thus are often used for multiple-sets data.

2.3 Three-way method chart

In this section a three-part chart of methods is presented. Obviously a thorough discussion of this chart with proper references would require a full book. The chart is here presented as an appetiser of which topics one might come across in the three-way field. A more extended discussion of component models is contained in Kroonenberg (1992), while the other parts are still in my portfolio. In a way the Sage booklet by Arabie et al. (1987) can be seen as a treatment of Part III of the Method Chart.

3. Examples

In this section we will present four condensed examples to illustrate some more recent or less well-known techniques. The examples come from such diverse disciplines as agriculture (interactions from three-way ANOVA), demographics (three-way correspondence analysis), psychology (three-mode covariance structure analysis), and sensory perception (individual differences in orientation scaling).

3.1 Interactions from three-way ANOVA: Dutch maize data

In agriculture, three-way data summaries often take the form of three-way tables with one way consisting of different varieties of a crop, while the other two ways contain two environmental factors such as years and locations. This kind of summary is a direct consequence of the interpretation of the phenotype as the joint product of genotype and environment, where the environment encompasses everything that is non-genetic. Such three-way tables may be modelled with three-way analysis of variance, but due to the generally moderate to large number of entities in the ways, facilities to model two-way and three-way interactions multiplicatively should be considered.

METHOD CHART

Part I: Profile data

Dependence techniques: General linear model methods

two-block multiple regression, three-mode redundancy analysis;

Interdependence techniques: Component methods

three-mode component analysis, parallel factor analysis,
three-mode correspondence analysis, latent class analysis,
spatial evolution analysis;

Dependence and interdependence techniques

multi-set canonical correlation analysis, procrustes analysis,
multi-set discriminant analysis;

Clustering methods I

three-way mixture method clustering.

Part II: Covariance models for profile data

Stochastic covariance (or structural equations) models

Repeated-measures methods

invariant factor analysis,three-mode common factor analysis,
additive and multiplicative modelling of
multivariable-multioccasion matrices;

Cross-sectional methods

simultaneous factor analysis;

Exploratory covariance model methods

Repeated-measures methods

(quasi-) three-mode component analysis;

Cross-sectional methods

simultaneous component analysis,
indirect fitting with component analysis methods;

Part III: Similarity and preference data

Multidimensional scaling models

individual differences scaling, general euclidean models,
three-way multidimensional scaling;

Clustering methods II

individual differences clustering, three-way ultrametric trees,
synthesized clustering;

Unfolding models

three-way unfolding

3.1.1 Theory

In this section we will concentrate on modelling the three-way interaction with a three-way decomposition. There are various three-way generalizations of the two-way singular value decomposition, but the one most satisfactory for our purposes was that due to Tucker (1966). First we arrange the three-way ANOVA interaction parameters, $\alpha\beta\gamma_{ijk}$, in a three-way array, and decompose that array according to the Tucker3 model to give

$$\alpha\beta\gamma_{ijk} = \Sigma_p \Sigma_q \Sigma_r a_{ip} b_{jq} c_{kr} g_{pqr} + e_{ijk}$$

Only the most important multiplicative terms for each of the ways will generally be retained in the model. For testing multiplicative three-way interaction not very much theory has been developed. Boik (1990) presents a likelihood ratio test for the first term, including a table of critical values for comparatively small three-way tables. In cases covered by Boik, his test can be used otherwise we advise to use a procedure similar to one of the procedures for two-way tables, attributing *df* equal to the number of independent parameters to terms that stand out in the amount of three-way interaction described. This usually concerns only the first three-way component for each way. Occassionally second terms are involved.

For an $I \times J \times K$ three-way table of raw data, a three-way decomposition with P components for the first way, Q components for the second way, and R components for the third way, implies the estimation of $I \times P + J \times Q + K \times R$ scores plus $P \times Q \times R$ singular values. Owing to rotational invariance the number of constraints is equal to $P^2 + Q^2 + R^2$. The difference between parameters estimated and constraints imposed then results in $(I \times P + J \times Q + K \times R + P \times Q \times R) - (P^2 + Q^2 + R^2)$ df for the fitted model. When considering the three-way interaction instead of the raw data, I, J, and K should be reduced by one.

3.1.2 Example: Dutch maize variety trials

In the Netherlands there is an ongoing programme of testing maize varieties which aims to select the best varieties for cultivation and use under Dutch conditions. The present data set consists of 6 varieties (Brutus, Splenda, Markant, Vivia, Dorina, Irla) planted in four regions of the Netherlands characterised by the soil and location (Southern Sand, Central Sand, Northern Sand, and River Clay). Mean values on the attribute Percentage Dry Matter Content – DMC – for these varieties were available for the years 1980 through 1987 with the exclusion 1983 which was deleted due to missing data in one location. In other words, the data can be arranged as a 6 (varieties) by 4 (locations) by 7 (years) three-way table.

Table 1 shows the results of the straightforward three-way ANOVA for percentage dry matter content of the maize. The variance explained of the variety-by-year and the three-way interaction are roughly of the size of the explained variance by the varieties themselves, but they have, of course, much smaller mean squares due their larger degrees of freedom. One con-

Source	Degrees of Freedom	Sum of Squares	Mean Square
Variety	5	80	16.04
Site	3	962	320.73
Year	6	1195	199.11
Site × Year	18	851	47.30
Variety × Site	15	17	2.76
Variety × Year	30	65	2.16
Var. × Site × Year	90	80	0.89
1 × 1 × 1-solution	12	31	2.61
2 × 2 × 2-solution (given 1 × 1 × 1)	12	18	1.49
Deviations	66	31	0.47
Total	167	3520	

Table 1: Analysis of Variance for the Maize Data

clusion is that the first two terms of the decomposition of the three-way interaction are significant (for details see Van Eeuwijk and Kroonenberg, submitted).

With a three-way singular value decomposition of the three-way interaction effects table using two components for each of the ways, $P = Q = R = 2$, it was possible to identify the contrasts as responsible for the whole of the three-way interaction. The interaction patterns are summarised in Table 2.

The three-way interaction could be partially interpreted as a kind of correction on the Location by Year interaction. The three-way interactive pattern shows that the IJK-interaction is of a contrast type, i.e. it is caused by specific varieties reacting in a specific way at specific locations in contrast with the behaviour of some other varieties.

3.2 Three-way correspondence analysis: French canton data

Most three-way methods require data to be at least interval scaled. In this section we will analyse the situation where we have three categorical variables with more than a few categories. Carlier and Kroonenberg (1996) described in detail both the theory and the interpretation of three-way correspondence analysis (three-way CA). With the technique it is possible to produce measures and graphical displays of the dependence in a three-way

	Splenda			Brutus/Vivia			Dorina		
	South Sands	North Sands	Central Sands	South Sands	North Sands	Central Sands	South Sands	North Sands	Central Sands
1980	-	+		+	-				
1981				+	-	-	-		+
1985	+	-		-	+				
1986				-	+	+	+		-

Table 2: Interactive Patterns from Two-Dimensional Multiplicative Solution of Three-Way Interaction

tables, and as such it shares and extends many properties of ordinary (two-way) correspondence analysis.

3.2.1 Theory

The basic starting point is a three-way contingency table with orders I, J and K relative frequencies, p_{ijk}. To measure the deviations from the three-way independence model in such a table (thus taking into account all interactions), Pearson's *mean-square contingency coefficient*, Φ^2, or *Inertia* is an appropriate measure.

$$\Phi^2 = \frac{X^2}{n} = \sum_{i,j,k} \frac{(p_{ijk} - p_{i..}p_{.j.}p_{..k})^2}{p_{i..}p_{.j.}p_{..k}}. \tag{1}$$

It can be shown that the total inertia may be partitioned as follows

$$\begin{aligned}
\Phi^2 &= \sum_{ij} p_{i..}p_{.j.}\left(\frac{p_{ij.} - p_{i..}p_{.j.}}{p_{i..}p_{.j.}}\right)^2 + \sum_{ik} p_{i..}p_{..k}\left(\frac{p_{i.k} - p_{i..}p_{..k}}{p_{i..}p_{..k}}\right)^2 \\
&+ \sum_{jk} p_{.j.}p_{..k}\left(\frac{p_{.jk} - p_{.j.}p_{..k}}{p_{.j.}p_{..k}}\right)^2 + \sum_{ijk} p_{i..}p_{.j.}p_{..k}\left(\frac{p_{ijk} - {}_{\alpha}p_{ijk}}{p_{i..}p_{.j.}p_{..k}}\right)^2 \\
&= \Phi^2_{IJ} + \Phi^2_{JK} + \Phi^2_{IK} + \Phi^2_{IJK},
\end{aligned} \tag{2}$$

where ${}_{\alpha}p_{ijk}$ is implicitly defined. This is clearly an additive definition of the interaction in a three-way array. Equation (2) shows that the global measure of dependence, Φ^2, can be split into separate measures of dependence: there are three measures for the dependence due to each two-way margin which are identical to those used in two-way correspondence analysis, and one measure for the three-way interaction. Such a partitioning is the first step in the analysis of a three-way table.

The three-way analogue of the singular value decomposition, especially the Tucker3 model, will be used to model the dependence. There is a subtle difference in the present usage of the Tucker3 model in that, analogous to two-way correspondence analysis, orthonormality is defined with respect to *weighted* metrics defined by $\{p_{i..}\}$, $\{p_{.j.}\}$, and $\{p_{..k}\}$, respectively, moreover, a weighted least-squares criterion is used.

One of the attractive features of using the additive approach over a multiplicative (or loglinear modeling) one, is that one single decomposition of the global dependence is made, and that the marginal dependence can directly be modeled and assessed from the global decomposition. The contributions of the marginal dependences to the global dependence can be evaluated without having to construct special decompositions for lower-order interactions as was necessary in the previous example. Moreover, such interactions can be portrayed in the same plot as the global dependence. To portray the dependence we will again use *interactive biplots*, in which the markers of two of the three ways (here: (i, j))are combined and plotted in the same figure with the markers of the remaining way. Assuming j is an ordered mode, trajectories can be drawn in the biplot by connecting, for each i, the points (i, j) in their proper order. This will greatly facilitate interpretation, especially if j is a time mode.

3.2.2 Example: Changes over time in the Languedoc work force

During the census of 1954, 1962, 1968 and 1975, the people of 42 cantons in Languedoc-Roussillon (Southern France) were asked to state their profession. Their occupations could be grouped into seven major occupational classes: Farmers (*AF*), Agricultural labourers (*AL*), Owners of small and medium-sized businesses (*SB*), Professionals and senior managers (*PS*), Middle managers (*MM*), Employees (white-collar workers - *WC*), Labourers (blue-collar workers - *BC*), Employees in the service sector (*SE*), Other occupations (*OO*). Full details as well as the data themselves can be found in a special issue of *Statistique et Analyse des Données* (1985, *10* (1), especially p. 11-15).

To evaluate the results of the analysis we will first examine the table with the partitioning of the χ^2-variance to evaluate the sizes of the different interactions. Of the total χ^2-variance, the absolute largest amount is explained by canton-by-occupation interaction (57%) followed by the occupation-by-time interaction (22%). If the degrees-of-freedom are taken into account as well, the occupation-by-time interaction has by far the largest contribution per *df*, which indicates that the occupational distributions have undergone considerable changes over time. Also the canton-by-occupation has a sizeable contribution per *df* showing that there is considerable diversity among the cantons. The smaller cantons-by-time interaction contains the differential increase and decrease of the cantons, primarily a trek from the rural areas to the towns. The three-way interaction is not large, and it has by far the smallest contribution per *df*, and it will not be discussed further.

To describe the patterns in the data set, we have fitted a Tucker3 model with 5 components for the cantons, 5 components for the occupations and 2 components for the time mode. This model fits very well leaving only 8% unexplained, and Table 3 shows that the canton-by-occupation and the occupation-by-time interactions are very well explained (96% and 95%), re-

Source	df	X^2_{Tot}	% of X^2_{Tot}	X^2_{Tot}/df	Results Fitting 3-Way Model X^2_{Err}	% of X^2_{Err}	X^2_{Fit}/X^2_{Tot}
Main effects					3	≈0%	-
Cant. x Occup.	328	114307	57%	348	4571	29%	96%
Canton x Time	123	16278	8%	132	1793	11%	89%
Occup. x Time	24	43469	22%	1811	2191	14%	95%
C x O x T	984	25249	13%	26	7441	47%	60%
Total	1459	199303	100%	137	16001	100%	92%

Table 3: Cantons: Global and Marginal Quality Indexes: Chi-square Variances

spectively. The relatively unimportant three-way interaction has the smallest fit (60%). One of the strong points of the method proposed here is that it takes into account these two-way interactions within the framework of a single model which is fitted to the deviations from the three-way independence model.

To be able to display the results in interactive biplots we need to select of a reference mode. As we intend to study the changes in the distributions of the cantons over time, the occupations have to be chosen as reference mode. In the interactive biplot, the four canton-occasions points of each canton i (or interactive markers, $(i,k), k = 1,..,4$) are connected by a line ending in an arrow head for 1975; such lines are called *trajectories*. Trajectories can be interpreted in terms of the distributions of occupations in the cantons at each occasion. Here we will only look at the first two axes of the interactive biplot. The complete example is contained in Carlier and Kroonenberg (1996). The interactive plot of dimensions one and two (Figure 1) explains 71% of the inertia. In principle the biplot displays the complete global dependence, but it can also be used to study the three two-way interactions, because these interactions can be derived from the global dependence by (weighted) averaging of the coordinates. In other words, by computing weighted means of coordinates over one way, and displaying these means in the same figure as the global dependence figure, we can at the same time interpret the two-way interactions and assess what the global dependence contains over and above the two-way interactions. Here we will only look at the canton-by-occupation interaction.

In Figure 1 the centroids of the canton trajectories together with the occupation points indicate the canton-by-occupation interaction and they are displayed in the same graph as the global dependence itself by marking them with the abbreviations of the cantons on the trajectories. Moreover, the biplot with the canton centroids and the occupations can be interpreted exactly as the comparable two-way biplot from two-way correspondence analysis. Some of the more extreme features of the interaction as evident from

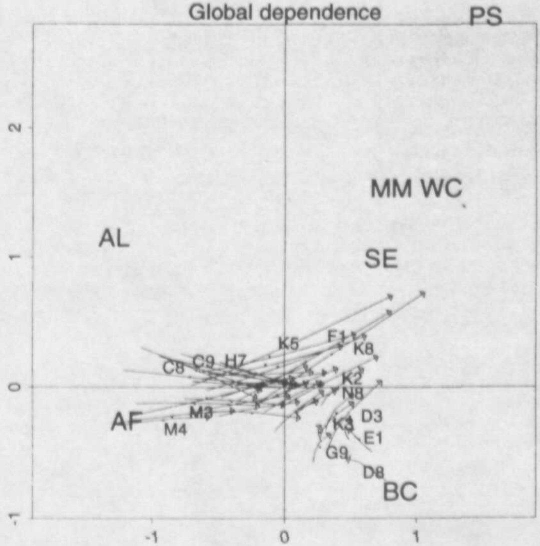

Figure 1: Interactive Biplot for Global Dependence of Cantons Data

Figure 1 are the following.

The cantons Fournels (M4) and Chateauneuf de Randon (M3) are very rural with a marked presence of independent farmers (AF). Similarly Lézignan-Corbières (C8), Narbonne (C9), and Capestang (H7) are very rural but with a predominance of agricultural labourers (AL) – possibly due to the viticulture in those areas. La Grand'Combe (D8), St Ambroix (E1), Ganges (K3), and Sumène (G9) are primarily industrial cantons with around 70 – 50%, of the work force occupied as blue collar workers, especially in the coal mines. Montpellier (K8), Nîmes (F1) and Les Matelles (K5) (a suburb of Montpellier) have a strong tertiary flavour with around 30% of the work force employed in the tertiary sector.

3.3 Exploratory three-mode covariance structure analysis: Personality-judgement data

In the literature, information in three-way data sets is often only reported in terms of covariances or correlations between the variables. A prevalent case is that of multitrait-multimethod (MTMM) matrices. There exists an

		Extraversion			Impulsivity			L.Anxiety			L.Motivation		
Methods	Ratings	P	T	S	P	T	S	P	T	S	P	T	S
	Peers	1.											
Extraversion	Teacher	.6	1.										
	Self	.5	.5	1.									
	Peers	.4	.3	.1	1.								
Impulsivity	Teacher	.4	.6	.4	.6	1.							
	Self	.3	.4	.4	.4	.5	1.						
	Peers	.4	.2	.1	.2	.2	.0	1.					
Lack of	Teacher	.3	.3	.1	.1	.1	-.1	.7	1.				
Test Anxiety	Self	-.0	-.1	.0	.1	.1	-.2	.4	.3	1.			
	Peers	.3	.1	-.1	.5	.2	.2	.5	.4	.1	1.		
Lack Academic	Teacher	.2	-.1	-.2	.3	.1	.1	.5	.6	.2	.7	1.	
Ach.Motivation	Self	.2	.0	.0	.4	.1	.4	.4	.4	.0	.6	.6	1.

Table 4: Correlations of Four Personality Variables Measured by Peers, Teachers, and Self Ratings

extensive literature on analyzing MTMM matrices following the stochastic line of three-way analysis, especially in the field of the analysis of covariance structures (e.g. Bentler and Lee, 1979; see Browne (1984) for a review). The procedures proposed are primarily confirmatory and statistical, and generally based upon specific distributional assumptions.

Kiers et al. (1992) developed a least-squares algorithm for Tucker's (1966) Method III, which allows an exploratory analysis of MTMM matrices by three-mode component analysis models. Just as in Bentler and Lee's analysis, subject component scores can no longer be found, and their role is taken over by what they call a (structured) loading matrix: the supermatrix containing the correlations between the individual differences components and the traits for each method.

As a (simplistic) illustration we will re-analyze an MTMM correlation matrix, analysed several times by Bentler et al. (e.g. 1979). As formulated in Bentler and Lee (1979, p. 93), correlations were available among four personality variables, each measured by Peers, Teachers and Self-ratings in a sample of 68 fifth-graders from two classrooms of a middle-class public elementary school (A sample size which is really too small for the stochastic approach). The traits measured were Extraversion (Extr), test anxiety, Impulsivity (Imp), and academic achievement motivation. For ease of interpretation we will reverse the scores on test anxiety and academic achievement motivation and label these variables Lack of test Anxiety (L.Anx) and Lack of Motivation (L.Mot), respectively. As a result, all sizeable correlations are positive, see Table 4 which contains a rearranged and rounded version of the original correlation matrix. The rearrangement was based on the results to be presented, and most of the large scale patterns can be seen from this table.

In an attempt to stay as close as possible to Bentler and Lee's original analyses, at first a solution was determined with 2 method components, 3 trait components, and 5 individual differences components (a $2 \times 3 \times 5$-solution). The solution accounted for 77% of the variability, but the fifth individual differences component accounted for only 5 percent, with an unclear pattern. Therefore, a $2 \times 3 \times 4$-solution was also determined. This solution explained 72% and will be presented here.

Table 5 contains the orthonormal components for the traits and the methods, the core matrix, and the loading matrices. The two method components indicate to what extent Teachers, Peers, and Self-ratings concur, and to what extent Teachers and Peers differ from the Self-ratings, respectively. The first method component also shows that the correlations tend to be somewhat smaller for the Self-ratings. The first two trait components show the classic pattern for generally correlated traits but divided into two blocks (here: Extraversion/Impulsivity and Lack of Test Anxiety/Lack of Academic Achievement Motivation) with lower correlations between blocks than within blocks (see also Table 4). The third trait component indicates that the situation is somewhat more complex, in particular that there is also some correlation between Lack of Test Anxiety and Extraversion, and between Impulsivity and Lack of Academic Achievement Motivation (again see Table 4).

From the size of the core elements, we see that all large core elements (1.93; 1.50; 1.90) refer to what the methods measure in the same way, as they pertain to the first method component. In other words, the major variability is due to traits rather than methods: Peer judgements, Teachers judgements, and subjects' Self-ratings concur on the general pattern among the four traits. Finally, the only other largish element (0.72) in the core matrix pertains to the Self-ratings versus Peer-and-Teacher contrast. It is rather difficult to see what this contrast precisely entails from just looking at the components, probably because the effect itself is rather small. Further information about this can be derived from the loading matrices. There we see in the fourth subject component the contrast between especially Peers and the Self-ratings with the Teachers siding largely with the Peers. However, an interpretation of the observed difference is not very clear, and the difference is also difficult to trace in the original correlation matrix.

Certain aspects of the interpretation would benefit from rotations, especially the trait components. The inverse transformations, however, immediately make the core matrix less interpretable because before rotation there are only four large elements in the core matrix while after rotation there will be hardly any large elements, but many medium-sized ones.

Clearly, the above description is a long way from a real substantive interpretation and carries little theoretical content. However, it may serve to illustrate that an exploratory analysis can be performed with the Tucker models. Furthermore, it should be noted that the outcomes of the analyses have a distinctly different flavour from Bentler and Lee's results. In particular, the question may be raised whether their consistent split between

A. Components

Trait	Trait Components T1	T2	T3	Method	Method Components M1	M2
Impulsivity	.61	.37	.34	Teacher	.61	-.24
Extraversion	.52	.47	-.49	Peers	.59	-.51
L. Ac.Ach. Motivation	*.50*	*-.53*	*.51*	*Self*	*.53*	*.83*
L. Test Anxiety	.33	-.61	-.62			
Proportion Variability	.37	.22	.13		.63	.09

B. Core Matrix

Method Components	Trait Components	Subject Components S1	S2	S3	S4
	T1 (Common)	1.93	.31	.01	-.02
M1	T2 (I&E vs. L.A&L.M)	-.39	1.50	.19	-.15
(Shared)	T3 (I&L.M vs. E&L.A)	-.04	-.26	1.09	-.01
	T1 (Common)	-.24	.23	-.06	.72
M2	*T2 (I&E vs. L.A&L.M)*	*.10*	*.15*	*-.01*	*.37*
(P&T vs. S)	T3 (I&L.M vs. E&L.A)	.32	.02	.27	.33
Proportion Variability Accounted for		.33	.21	.11	.06

C. Variable-Component Correlations

Method		Subject Components S1	S2	S3	S4			Subject Components S1	S2	S3	S4
	E	.7	.4	-.2	-.2		E	.5	.6	-.3	-.1
S1	I	.7	.1	.4	-.4	S2	I	.6	.5	.2	-.3
(Peers)	L.M	.7	-.4	.2	-.1	(Teacher)	L.M	.7	-.6	.2	.0
	L.A	.7	-.4	-.4	.1		L.A	.7	-.4	-.4	.1
	E	.3	.7	-.3	.3						
S3	I	.5	.5	.3	.5						
(Self)	L.M	.6	-.3	.4	.3						
	L.A	.2	-.4	-.6	-.1						

Table 5: Three-Mode Principal Component Solution with four Subject Components, three Trait Components, and two Method Components

Peers and Teacher judgements is adviseable, and our analysis suggests that a Peers/Teacher factor and a Self-ratings factor might be worth contemplating.

3.4 Individual differences in orientation scaling: Grigg's pain data

In the behavioural sciences it is not uncommon that data are collected as similarities. People are requested to express either directly or indirectly how similar two stimuli are. In the example, people were requested to indicate how similar certain sensations related to pain were. The question is whether the subjects perceived pain in a similar manner, and how the pain sensations grouped together. The unpublished data were kindly supplied by Dr. L. Grigg (University of Queensland, Australia).

3.4.1 Theory

In the present case these similarities are converted to dissimilarities. This makes the values comparable to distances, and in fact we will treat the dissimilarities as if they were *squared* distances. As is shown in the MDS-literature, double centring squared distances gives scalar products which can in turn analysed by scalar-product models, such as INDSCAL and IDIOSCAL (see Carroll and Chang (1970)). The INDSCAL model assumes that there exists a common stimulus configuration, which is shared by all judges (subjects), and that this configuration has the same structure for all judges, except that they may attach different importance (salience) to each of the (fixed) axes of the configuration. This results in some judges having configurations which are stretched out more along one of the axes. The IDIOSCAL model is similar except that each judge may put the axes of the common configuration under a different angle, and thus they may orient the common space in a different way.

Recently, Ten Berge et al. (1994) showed that the TUCKALS2 algorithm (see section 4.1.4) is an efficient way to estimate the IDIOSCAL model (see section 4.1.2). On the other hand, Arabie et al. (1987) indicate that this model "[..] has empirically yielded disappointing results in general" (p.45). In this section we will present an application of the IDIOSCAL model, but at the same time show that also in this data set interpreting the IDIOSCAL model is not worth the effort compared to interpreting the INDSCAL model.

3.4.2 Example: Grigg's pain data

From the original 41 subjects, sixteen were chosen for this example on the basis of the fit of the IDIOSCAL model to their data during a preliminary analysis. The analysis reported here is a two-component solution with a fit of 39.2%, indicating that the data are still very noisy. Figure 2 shows the stimulus space. Violent pains, such as shooting, burning, cramping, intense pain are in the same region of the space, less dramatic ones, such are mild, moderate, and annoying are also located near each other, as are tiring, miserable, and distressing. One identify directions in the space, for instance, by specifying an axis going from mild to intense and one from more short

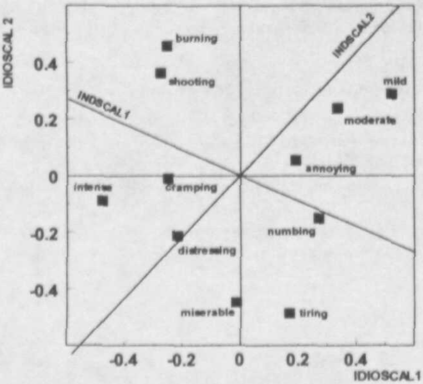

Figure 2: Grigg's Pain Data. IDIOSCAL Stimulus Space

sensations to longer lasting sensations.

The subject weights are shown in the left-hand panel of Table 6. The table provides the weights allocated to the first dimension and to the second dimension as well as the "individual orientation" expressed as a cosine between the two dimensions as well as the angle this represents. Clearly the subjects fall in two groups, those that put the dimensions under an acute angle and those that put them at an obtuse angle. Proof enough for an individual *differences of orientation scaling it seems. However, one problem is that for* identifiability of the model, we had to assume that the stimulus space was orthogonal. To check whether this was problematic we performed an INDSCAL analysis (with our TRILIN program (see section 4.1.4)). This analysis provided a fit of 38.3%, hardly worse than the previous analysis, and given that its interpretation is more straightforward it is clearly to be preferred. The additional complexity of the IDIOSCAL model was only apparent in this case and the results support Arabie et al.'s conclusion.

In Figure 2 we have drawn the orientation of the two INDSCAL axes, which have an inner product of -.31 and thus make an angle of 108 degrees. In the right hand panel of Table 6 we see the subject weights. Of course, they reflect the two groups found earlier. Staying with the basic INDSCAL interpretation we see that one group of subjects (1,2,3,10,11,12,13,14) tends to emphasize the axis of burning, shooting, intense, cramping pain in contrast with mild, numbing, and tiring. The other group of subjects (5,6,7,8,9,15,16) contrast mild and moderate pain with intense, tiring, distressing and miserable, and

Type of Subject	IDIOSCAL Subject Weights (1,1)	(2,2)	(1,2)	IDIOSCAL Cosines (1,2)	INDSCAL Subject Weights (1,1)	(2,2)
Control	.71	.15	-.27	-.82	.45	.03
Chronic Pain	.59	.20	-.22	-.64	.38	.05
Chronic Pain	.66	.25	-.18	-.46	.37	.13
Control	.53	.22	-.15	-.44	.31	.09
Control	.66	.09	-.03	-.13	.29	.19
Chronic Pain	.41	.23	-.24	-.79	.28	.02
Control	.50	.14	-.06	-.22	.24	.14
RSI Pain	.42	.32	.27	.75	.07	.42
RSI Pain	.53	.32	.19	.45	.16	.38
Chronic Pain	.34	.45	.17	.44	.09	.36
RSI Pain	.52	.24	.15	.44	.16	.33
RSI Pain	.45	.29	.12	.34	.15	.31
Chronic Pain	.35	.47	.08	.19	.14	.30
Control	.51	.30	.08	.22	.19	.30
Control	.25	.32	.08	.32	.08	.23
RSI Pain	.52	.09	.04	.16	.22	.19

Table 6: Grigg's Pain Data. IDIOSCAL Subject Weights and Cosines and INDSCAL Subject Weights (sorted with respect to INDSCAL weights)

place burning and shooting somewhere in the middle.

In the original design the subjects consisted of three groups: chronic pain sufferers, repetitive-strain-injury sufferers, and a control group. If the information available is correct then the division into two groups runs right through the design groups. Unfortunately, we have been unable to contact the original researcher to confirm the placement of the subjects in the groups or to request additional information on the subjects, which might shed light on the relationship between perceived and semantic pain expressions.

4. Software: 3WAYPACK (version 2)

The analyses presented in this paper have been carried out with the author's program package 3WAYPACK (Kroonenberg, 1994, 1996). This collection of programs has specifically designed for the analysis of three-way data. It consists of an menu-based interface, called INTERFACE3 (IF3), a preprocessing program, PREPROC3, three main analysis programs, TUCK-

ALS2 (Tucker2 model), TUCKALS3 (Tucker3 model), and TRILIN (Parafac model), and three output processing or postprocessing programs, ROTATE (rotating components), RESIDUAL (analysing residuals), and JOINTPLT (constructing joint (bi)plots).

References:

ARABIE, P.; CARROLL, J. D. and DESARBO, W. S. (1987): *Three-way scaling and clustering.* Sage, Beverly Hills.

BASFORD, K. E.; KROONENBERG, P. M. and DELACY, I. H. (1991): Three-way methods for multiattribute genotype by environment data: An illustrated partial survey. *Field Crops Research, 27, 131-157.*

BASFORD, K. E. and MCLACHLAN, G. J. (1985): The mixture method of clustering applied to three-way data. *Journal of Classification, 2, 109-125.*

BEFFY, J. L. (1992): Application de l'analyse en composantes principales à trois modes pour l'étude physico-chimique d'un écosystème lacustre d'altitude: Perspective en écologie. *Revue Statistique Appliquée, 40(1), 37-56.*

BENTLER, P. M. and LEE, S. Y. (1979): A statistical development of three-mode factor analysis. *British Journal of Mathematical and Statistical Psychology, 32, 87-104.*

BOIK, R. J. (1990): A likelihood ratio test for three-mode singular values: Upper percentiles and an application to three-way anova. *Computational Statistics and Data Analysis, 10, 1-9.*

BROWNE, M. W. (1984): The decomposition of multitrait-multimethod matrices. *British Journal of Mathematical and Statistical Psychology, 37, 1-21.*

CARLIER, A. and KROONENBERG, P. M. (1996): Decompositions and biplots in three-way correspondence analysis. *Psychometrika, 61, 355-373.*

CARROLL, J. D. and ARABIE, P. (1983): INDCLUS: an individual differences generalization of the ADCLUS model and the MAPCLUS algorithm. *Psychometrika, 48, 157-169.*

CARROLL, J. D. and CHANG, J. J. (1970): Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika, 35, 283-319.*

CARROLL, J. D.; PRUZANSKY, S. and KRUSKAL. J. B. (1980): CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika, 45, 3-24.*

CHATFIELD, C. (1989): The analysis of time series: An introduction (4th edition). Chapman and Hall, London.

DESARBO, W. S. and CARROLL, J. D. (1985): Three-way metric unfolding via alternating weighted least squares. *Psychometrika, 50, 275-300.*

FLURY, B. (1988): Common principal components and related multivariate models. Wiley, New York.

FRANC, A. (1992): Étude algebrique des multitableaux: Apports de l'algèbre tensorielle [An algebraic study of multi-way tables: Contributions of tensor algebra.] Unpublished doctoral thesis, Université de Montpellier II, France.

HARSHMAN, R. A. and LUNDY, M. E. (1984a): The PARAFAC model for three-way factor analysis and multidimensional scaling. In: H. G. Law, C. W. Snyder Jr., J. A. Hattie, a P. McDonald (eds): *Research methods for multimode data analysis. Praege New York, 122-215.*

HARSHMAN, R. A. and LUNDY, M. E. (1984b): Data preprocessing and the extended PARAFAC model. In: H. G. Law, C. W. Snyder Jr., J. A. Hattie, and R.P. McDonald (eds): *Research methods for multimode data analysis. Praeger, New York, 216-284.*

KIERS, H. A. L. (1991): Hierarchical relations among three-way methods. *Psychometrika, 56, 449-470.*

KIERS, H. A. L. (1992): TUCKALS core rotations and constrained TUCKALS modelling. *Statistica Appli ta, 4, 659-667.*

KIERS, H. A. L.; KROONENBERG, P. M. and TEN BERGE, J. M. F. (1992): An efficient algorithm for TUCKALS3 on data with large numbers of observation units. *Psychometrika, 57, 415-422.*

KRIJNEN, W. P. (1993): The analysis of three-way arrays by constrained parafac methods. DSWO Press, Leiden.

KROONENBERG, P. M. (1983a): Three-mode principal component analysis: Theory and applications. DSWO Press, Leiden.

KROONENBERG, P. M. (1983b): Annotated bibliography of three-mode factor analysis. *Britsh Journal of Mathematical and Statistical Psychology, 36, 81-113.*

KROONENBERG, P. M. (1992): Three-mode component models. *Statistica Applicata, 4, 619-634.*

KROONENBERG, P. M. (1994): The TUCKALS line: A suite of programs for three-way data analysis. *Computational Statistics and Data Analysis, 18, 73-96.*

KROONENBERG, P. M. (1996): 3WAYPACK User's manual (Version 2). Department of Education, Leiden University, Leiden.

KROONENBERG, P. M. and DE LEEUW, J. (1980): Principal component analysis of three-mode data by means of alternating least-squares algorithms. *Psychometrika, 45, 69-97.*

KRUSKAL, J. B. (1984): Multilinear methods. In: H. G. Law, C. W. Snyder Jr., J. A. Hattie, and R. P. McDonald (eds): *Research methods for multimode data analysis. Praeger, New York, 36-62.*

LAVIT, C. (1988): *Analyse conjointe de tableaux quantitatifs* [Simultaneous analysis of several quantitative matrices]. Masson, Paris.

LEURGANS, S. E. and ROSS, R. T. (1992): Multilinear models: Application in spectroscopy (with discussion): *Statistical Science*, 7, 289-319.

Statistique et Analyse des Données, 1985, *10* (1).

VAN EEUWIJK, F. A. and KROONENBERG, P. M. (Submitted): Multiplicative decompositions of interactions in three-way ANOVA, with applications to plant breeding. (*Biometrics*)

TEN BERGE, J. M. F.; BEKKER, P. A. and KIERS, H. A. L. (1994): Some clarifications of the TUCKALS2 algorithm applied to the IDIOSCAL problem. *Psychometrika, 59*, 193-201.

TUCKER, L. R. (1966): Some mathematical notes on three-mode factor analysis. *Psychometrika, 31, 279-311.*

A Hybrid Global Optimization Algorithm for Multidimensional Scaling

Rudolf Mathar

Institut für Statistik,
RWTH Aachen, D-52056 Aachen, Germany

Summary: Local search algorithms in Multidimensional Scaling (MDS), based on gradients or subgradients, often get stuck at local minima of STRESS, particularly if the underlying dissimilarity matrix is far from being Euclidean. However, in order to remove ambiguity from the model building process, it is of paramount interest to fit a suggested model best to a given data set. Hence, finding the global minimum of STRESS is very important for applications of MDS.
In this paper a hybrid iteration scheme is suggested consisting of a local optimization phase and a genetic type global optimization step. Local search is based on the simple and fast majorization approach. Extensive numerical testing shows that the presented method has a high success probability and clearly outperforms simple random multistart.

1. Why Global Optimization is Important

The purpose of MDS is to fit Euclidean interpoint distances to given dissimilarities. Depending on the underlying model corresponding benefit criteria can become quite complicated, for an overview see, e.g., de Leeuw and Heiser (1980) and Cox and Cox (1994). In its simplest and most intuitive form, however, the aim of MDS is to

$$\text{minimize}\left\{\sigma^2(\mathbf{X}) = \sum_{i<j} w_{ij}(\delta_{ij} - d_{ij}(\mathbf{X}))^2\right\} \quad \text{over} \quad \mathbf{X} \in \mathbb{R}^{n\times k}, \qquad (1)$$

where δ_{ij} are given symmetric dissimilarities between objects $\mathcal{O}_1, \ldots, \mathcal{O}_n$. $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$ denotes so called *configurations*, i.e., n points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^k$, and $d_{ij}(\mathbf{X}) = \|\mathbf{x}_i - \mathbf{x}_j\|$ is defined as the Euclidean distance between $\mathbf{x}_i$ and $\mathbf{x}_j$. $w_{ij} \geq 0$ are given weights, which by setting $w_{ij} = 0$ allows for including the case of missing values δ_{ij}. The benefit function of the weighted least squares scaling problem (1) is usually called STRESS. It is known to have a lot of stationary points which are not global minimizers. Fitting a low-dimensional ($k = 2$ or 3) metric coordinate model $\mathbf{X}$ to the dissimilarities allows for a graphical representation of the objects in a Euclidean space, and hence for visual inspection of conspicuous structures. Moreover, even positions of objects can be estimated if rotational and translational indeterminancy can be removed (see example 1 below).

Once having the model and the benefit function fixed, it is of paramount importance to fit the model to the data as close as possible. For MDS this means to determine the global minimum of STRESS. A lot of attempts has