# A Monte Carlo Evaluation of Maximum Likelihood Multidimensional Scaling Methods

Bijmolt, T.H.A.; Wedel, M.

*Publication date:*
1996



*Citation for published version (APA):*
Bijmolt, T. H. A., & Wedel, M. (1996). *A Monte Carlo Evaluation of Maximum Likelihood Multidimensional Scaling Methods*. (FEW Research Memorandum; Vol. 725). Marketing.

# A Monte Carlo Evaluation of Maximum Likelihood Multidimensional Scaling Methods

Tammo H.A. Bijmolt (Tilburg University)

Michel Wedel (University of Groningen) [65281]

March 1996

[65281] This research was sponsored by the Economic Research Foundation, which is part of the Netherlands Organization for Scientific Research (NWO). Part of this study was performed while the first author was employed at the University of Groningen. The authors thank Ivo Molenaar for his helpful comments, and Martin Mellens for his help in the computations for this study.

Correspondence should be addressed to Tammo H.A. Bijmolt, Department of Business Administration, Tilburg University, P.O. Box 90153, 5000 LE, Tilburg, The Netherlands.

# A Monte Carlo Evaluation of Maximum Likelihood Multidimensional Scaling Methods

abstract

We compare three alternative Maximum Likelihood Multidimensional Scaling methods for pairwise dissimilarity ratings, namely MULTISCALE, MAXSCAL, and PROSCAL in a Monte Carlo study. The three MLMDS methods recover the true configurations very well. The recovery of the true dimensionality depends on the test criterion (likelihood ratio test, AIC, or CAIC), as well as on the MLMDS method. The three MLMDS methods fit the dissimilarity data equally well. The methods are relatively robust against violations of their distributional assumptions. MULTISCALE outperforms PROSCAL and MAXSCAL with respect to computation time. In a separate Monte Carlo study, it is shown that the MLMDS methods frequently converge to local optima, especially if a random start is used. Rational starts, however, turn out to provide a satisfactory solution for the local optima problem. Implications for researchers intending to apply MLMDS are provided.

# A Monte Carlo Evaluation of Maximum Likelihood Multidimensional Scaling Methods

Multidimensional scaling (MDS) has been used frequently by marketing researchers in science and practice (Greenberg, Goldstucker, and Bellenger 1977; Kinnear and Taylor 1991; Wind, Rao, and Green 1991) for a wide range of marketing issues (Cooper 1983; Green 1975). The number of MDS methods available and the complexity of these methods increases rapidly, however, which makes it difficult for a marketing researcher to choose between alternative MDS methods. The introduction of methods based on the maximum likelihood (ML) principle is considered to be one of the most important recent developments in MDS analysis of dissimilarity data. Davison (1983) stated '... the maximum likelihood approach may enable researchers to examine the fit of the model to their data more rigorously than has been possible with other approaches'. According to Young and Hamer (1987) 'The approach changes Multidimensional Scaling from a descriptive tool into an inferential tool'. Maximum Likelihood Multidimensional Scaling (MLMDS) offers a number of theoretical and practical advantages over classical MDS methods, such as ALSCAL (Young and Lewyckyj 1979) and KYST (Kruskal, Young, and Seery 1973). These advantages arise because MLMDS methods assume that the observed dissimilarity data are error-perturbed, and explicitly model the error component. Contrary to the classical methods, the ML approaches estimate uncertainty or variance parameters in addition to the stimulus coordinates. Furthermore, the ML approach enables the researcher to test between alternative models, for example in selecting the most appropriate dimensionality.

To date, three ML approaches to the MDS analysis of dissimilarity data have been developed: MULTISCALE (Ramsay 1977, 1978, 1982, and 1991a), MAXSCAL (Takane 1977, 1978a, 1979, 1981, 1982; Takane and Carroll 1979 and 1981), and PROSCAL (MacKay 1983, 1989; MacKay and Zinnes 1981, 1982, 1986, 1991; Zinnes and MacKay 1983). In applying MLMDS researchers have to choose one of these three alternative methods. Presently, however, there is only very limited knowledge of the relative performance of these methods under various conditions, such as different sample sizes, numbers of stimuli, numbers of dimensions, and error variance levels. As the conceptual and mathematical frameworks differ considerably, the relative performances of the three methods in a wide range of situations needs to be investigated (Spence 1983). Here, we will present a Monte Carlo simulation study comparing the three methods. Our study will focus on a comparison of the relative performance of the methods with respect to pairwise dissimilarity judgements made on rating scales, which is probably the most commonly used and the most appropriate method to collect dissimilarity data (Bijmolt and Wedel 1995). First, we describe the conceptual frameworks of the MLMDS methods. Second, we provide a review of previous Monte Carlo studies investigating the performance of the methods. Third, a description of the design of our main Monte Carlo study is given. Fourth, we present the results of this study. Next, we present a Monte Carlo study on the extent to which the three MLMDS methods suffer from the local optima problem. Finally, we discuss the results and provide implications for marketing researchers intending to apply MLMDS.

## *THE MLMDS MODEL*

In this section, a framework for MLMDS models will be provided and the three specific methods will be described. The MLMDS methods isolate the systematic variation in the data from the random variation. The models consist of three parts, namely a measurement model, a representation model, and an error model. Table 1 summarizes some similarities and differences between the three MLMDS methods with respect to their measurement, representation, and error models.

[ Insert Table 1 about here ]

We will use the following notation:

$i, j = 1, ..., I$ : indicate stimuli,

$m = 1, ..., M$ : indicates dimensions,

$n = 1, ..., N$ : indicates subjects,

$r = 1, ..., R$ : indicates replications within a subject,

$t = 1, ..., T$ : indicates categories of a rating scale,

$x_{im} =$ the coordinate of stimulus $i$ on dimension $m$,

$d_{ijn} =$ the errorless distance between stimuli $i$ and $j$ for subject $n$,

$\delta_{ijnr} =$ the observed dissimilarity between stimuli $i$ and $j$ for subject $n$ in replication $r$,

$e_{ijnr} =$ error in the observation of the dissimilarity between stimuli $i$ and $j$ for subject $n$ in replication $r$,

$w_{mn} =$ weight for dimension $m$ for subject $n$,

*Measurement model*

In MDS tasks subjects judge the dissimilarity relation between stimuli. The measurement model represents the judgment processes of subjects. MULTISCALE and PROSCAL assume the dissimilarity data to be metric. In the MULTISCALE analysis a data transformation can be specified for each subject, such as a linear, a power or a spline transformation. Contrary to MULTISCALE, PROSCAL assumes the metric dissimilarities to be observed directly, and no measurement model parameters have to be estimated. MAXSCAL assumes the dissimilarity data to be nonmetric. Various types of nonmetric dissimilarity data can be analyzed, among which: ranking of pairs, tetrads, triads, paired comparisons, same-different judgments, and conditional rankings. The methods for this wide range of dissimilarity data are incorporated in several versions of MAXSCAL (Takane 1979; Takane and Carroll 1979). In the present study focus is on pairwise dissimilarity judgments made on rating scales. MAXSCAL assumes that these scales have a small number of categories and are nonmetric. In MAXSCAL, the probability $p_{ijtn}$ that the dissimilarity between stimuli $i$ and $j$ for subject $n$ falls in category $t$ is given by

$$p_{ijtn} = P(b_{(t-1)n} < \delta_{ijn} < b_{tn}), \tag{1}$$

where $b_{tn}$ denotes the upper bound of category $t$ for subject $n$. Hence, in the measurement model of MAXSCAL the category boundaries $b_{tn}$ have to be estimated. In order to reduce the number of parameters to be estimated, the category bounds can be restricted to be identical across subjects, or linear constraints can be imposed upon them.

*Representation model*

The systematic component of the dissimilarities $\delta_{ijnr}$ is captured as Euclidean distances $d_{ijn}$ in the representation model, which specifies the distances as functions of the stimulus coordinates. The distance function in the representation model is defined as follows:

$$d_{ijn} = \sqrt{\sum_{m=1}^{M} w_{mn}(x_{im} - x_{jm})^2}. \tag{2}$$

Only MULTISCALE allows for individual weighting of the dimensions through the dimensional weights $w_{mn}$, as in INDSCAL (Carroll and Chang 1970). For MAXSCAL and PROSCAL, $w_{mn} = 1$ for all $m$ and $n$, whereby the representation model reduces to the simple Euclidean distance model.

*Error model*

The error model specifies the form of the error component in the relationship between $\delta_{ijnr}$ and $d_{ijn}$. In MULTISCALE and MAXSCAL, the distances are assumed to be error-perturbed, so that:

$$\delta_{ijnr} = f(d_{ijn}, e_{ijnr}), \tag{3}$$

where $f(.)$ is some function. The error component $e_{ijnr}$ is assumed to have a distribution $h(.)$ with zero mean and variance $\sigma_{ijnr}$. MULTISCALE and MAXSCAL contain two options for the distribution function, namely the normal and the lognormal distribution. These give rise to two different error models. The normal distribution implies that $f(.)$ is an additive function:

$$\delta_{ijnr} = d_{ijn} + e_{ijnr}, \tag{4}$$

while the lognormal distribution implies that $f(.)$ is a multiplicative function:

$$\delta_{ijnr} = d_{ijn} e_{ijnr}. \tag{5}$$

The variance of the error component $\sigma_{ijnr}$ can be assumed to equal $\sigma$ across all of $i, j, n$, and $r$. Alternatively, the variance can be assumed to differ between stimuli, between subjects, or between replications. Also, the variance component $\sigma_{ij}$ can be assumed to be related positively to the distance between the stimuli $i$ and $j$. MULTISCALE includes options to estimate error variance parameters differing between stimuli and subjects, and to relate the error variance to the distances. MAXSCAL allows the variance parameter to vary between subjects, but not between stimuli.

Contrary to MULTISCALE and MAXSCAL, PROSCAL is based on the assumption that normally distributed error is added to the stimulus coordinates $x_{im}$. The resulting distribution of the squared distances $d^2_{ijnr}$ is a function of the non-central chi-square distribution (Hefner 1958; Ramsay 1969). A consequence is that the error variance may differ between dimensions (MacKay 1989; MacKay and Dröge 1990). The PROSCAL error variance can be specified to differ between stimuli and dimensions. Furthermore, the error variance can be assumed to be related to the distances.

*Estimation and statistical inferences*

If it is assumed that the observed dissimilarities are independently distributed with a (discrete or continuous) density function given by $h(\delta_{ijnr} \mid \theta)$, where $\theta$ is a vector with the model parameters to be estimated, the log-likelihood function is given by

$$\log L = \sum_i \sum_j \sum_n \sum_r \log h(\delta_{ijnr} \mid \theta). \tag{6}$$

The MLMDS methods attempt to recover the true parameter values by maximizing the log-likelihood using iterative procedures. For this purpose MULTISCALE and MAXSCAL employ the Fisher's scoring method. PROSCAL offers two optimization methods, namely the Davidon-Fletcher-Powell method and a direct search method.

The ML approach to MDS permits statistical inferences to be made on alternative representations of the data, such as a representation in $M$ dimensions versus a representation in $M-1$ dimensions. Let $L_0$ and $L_1$ be the two likelihoods obtained under a null model and a more comprehensive alternative model respectively. If the null model is subsumed under the alternative model, the Likelihood Ratio statistic

$$LR = 2(\log L_1 - \log L_0) \tag{7}$$

is asymptotically chi-square distributed with degrees of freedom equal to the difference in the number of parameters in the two models. The chi-square distribution for the LR-test may not hold while testing for the dimensionality (Shapiro 1986). One of the regularity conditions for the asymptotic chi-square distribution of the LR-test is not satisfied in that case. This condition is a sufficient, but not necessary condition, so that its violation does not imply that the asymptotic chi-square distribution never holds. For those situations, or if the models are not nested, a number of information criteria are available. These criteria penalize the likelihood by the number of parameters estimated, and take the form

$$-2\log L + c\pi, \tag{8}$$

where $\pi$ is equal to the number of free parameters and $c$ the cost of adding a parameter to the model. The various model selection criteria differ in the extent to which they penalize each additional parameter in the model through the factor $c$. In Akaike's information criterion (AIC) (Akaike 1974) the $c$ is equal to 2. In the consistent version of the Akaike's criterion CAIC (Bozdogan 1987) $c$ is equal to $\log(S)+1$, where $S$ is the sample size. If pairwise dissimilarity data are available for $I$ stimuli from $N$ subjects without replications the total sample size $S$ equals $(NI(I-1)/2)$. The model for which the information criterion is lowest is selected as the best representation of the data. Because of the total number of dissimilarity data points in empirical MDS studies, $\log(S)$ is much larger than 2, and CAIC penalizes additional parameters more severely, which results in more parsimonious models compared to AIC.

## *PREVIOUS EVALUATION STUDIES*

In this section, we give a brief review of previous Monte Carlo studies on the performance of MLMDS methods.

Ramsay (1977) compared MULTISCALE to INDSCAL in an analysis of synthetic data. MULTISCALE outperformed INDSCAL in the recovery of both stimulus coordinates and dimensional weights. In addition Ramsay (1977) showed that MULTISCALE underestimated the error variance in the case of a single subject. Ramsay (1980a) further investigated the small sample properties of the estimates and of the LR-test for dimensionality. He used a single variance level across subjects. The distribution of the LR-statistic was found to deviate from the chi-square distribution, which led to the retention of too many dimensions. Therefore, Ramsay proposed a correction factor to adjust the LR-statistic. For data with a single subject Ramsay's corrected LR-test did not provide much power to detect the true number of dimensions. On the other hand, in the case of 5 subjects and 10 stimuli and in the case of 2 subjects and 15 stimuli a correct rejection rate of 70 % of 2 versus 3 dimensions was achieved. In general, the estimates of the standard errors contained little bias even when the dimensionality was misjudged. Weinberg, Carroll, and Cohen (1984) compared the confidence regions computed by MULTISCALE with those computed by INDSCAL using jackknife and bootstrap techniques. Though the shape of the confidence intervals seemed to be equal, the MULTISCALE regions were smaller. In small samples (less than 20 subjects) the MULTISCALE estimates of the standard errors tended to give an optimistic view of the reliability of the solution. Spence and Lewandowsky (1989) investigated the robustness against outliers of several MDS methods. Nonmetric MDS turned out to be more robust than metric MDS. MULTISCALE appeared to be the most robust metric method and almost as robust as nonmetric MDS. Recently, Storms

(1995) showed that MULTISCALE is fairly robust against violations of the assumption of the error distribution. Neither the recovery of the true distances, nor the fit of the dissimilarity data were seriously affected if the distribution was misspecified. Furthermore, Storms (1995) found the corrected chi-square test (Ramsay 1980a) to outperform the uncorrected chi-square test while testing for 3 versus 2 dimensions.

Takane (1978b) performed two Monte Carlo simulations in order to investigate various statistical and numerical properties of the MAXSCAL-1 algorithm for the analysis of tetrads and triads. In the first study, he did not consider violations of the distributional assumptions. Better recoveries of the coordinates were obtained when the number of subjects increased, the number of judgments per subject increased, and the magnitude of the error decreased. The estimated standard errors of the stimuli decreased, as expected, in the same direction. The AIC statistic (Akaike 1974) detected the correct dimensionality in almost all cases. In the second study, Takane examined situations in which the distributional assumptions were violated. The AIC statistic indicated that the additive error model with a constant variance fits the data best in each situation studied. However, the dimensionality was overestimated by the AIC statistic in most of the cases. Thus, the AIC statistic seems reliable if the distributional assumptions are correct, but if these assumptions are violated AIC may be very unreliable. Furthermore, Takane found in this second study, that the estimates of stimulus coordinates were robust against the violations of distributional assumptions, but that the goodness of fit statistics were not.

MacKay and Zinnes (1981) performed a small Monte Carlo simulation study in which they examined the advantage of PROSCAL over a nonmetric analysis with KYST (Kruskal, Young, and Seery 1973). They showed that PROSCAL outperformed KYST with respect to the recovery of the true distances. Büyükkurt and Büyükkurt (1990) reported two extensive simulation studies of PROSCAL dealing with small-sample properties and the robustness against violations of the assumptions concerning the error model. The recovery of true distances by PROSCAL turned out to be related positively to the number of stimuli and the number of subjects, and negatively to the error variance level. PROSCAL underestimated the standard deviation under all combinations of factors, with a mean negative bias of 10 percent. Furthermore, Büyükkurt and Büyükkurt (1990) investigated for what percentage of the data sets the AIC and the LR-test correctly selected the two-dimensional solution. Overall, they concluded that both criteria performed very well (over 90 percent), with exception of the cases with 7 stimuli and 1 subject. In those cases, the hit rate of both AIC and the LR-test dropped to little over 60 percent. In general, the percentage of correct identification was slightly higher for the LR-test than for the AIC. In addition to the factors mentioned above, effects of violations of the assumption of normality of the error term were considered, and KYST and PROSCAL were compared in terms of parameter recovery. The five error distributions investigated did not result in differences in the recovery of the distances nor in the estimates of the standard deviations. The goodness of fit of PROSCAL turned out to be superior to that of KYST, especially under conditions with a small number of stimuli and a high error variance level. The identification of the dimensionality with AIC as well as with the LR-test were hardly influenced by the violation of the normality assumption. In the analyses with KYST, the percentage of times the true dimensionality was correctly identified by the well known elbow heuristic was considerably less than that by the AIC and the LR-test in PROSCAL analyses.

As has been described above, some previous Monte Carlo research has been done on computational and statistical properties of the three MLMDS methods. Yet, no comparison of the relative performance of the three methods has been made. In this study the performance

of the methods is compared under a wide range of conditions.

## *DESIGN OF THE MONTE CARLO SIMULATION STUDY*

The major advantage of a Monte Carlo simulation over a comparison of alternative MLMDS methods on empirical data is that the true configuration is known (Spence 1983). In our Monte Carlo study, configurations are generated on the basis of various specified numbers of stimuli and dimensions. From these configurations distances between the stimuli are computed. Next, the distances are error-perturbed and the appropriate transformation is applied to the dissimilarity data. This procedure results in a data matrix with simulated dissimilarities for a specified number of subjects. Detailed information on the data generation process is provided in the appendix.

*The factors included*

We vary seven factors in order to evaluate the performance of the alternative methods. These seven factors are the number of subjects, number of stimuli, data type (metric versus nonmetric), number of dimensions, type of error distribution, the error variance level, and differences in error variance between subjects and stimuli. These factors are hypothized to affect the performance of the three methods. Below we explain the seven factors and their levels (see Table 2).

[ Insert Table 2 about here ]

A. Number of subjects

We assume that each subject yields only a single dissimilarity data matrix, and that there are no missing values. The number of subjects chosen is 2, 8, and 14. While this is smaller than the number encountered in some applications, we choose for 14 subjects or less, because previous studies (Büyükkurt and Büyükkurt 1990; Takane 1978b) indicated that no more subjects are needed for a good performance of the methods.

B. Number of stimuli

The number of stimuli is selected to be 9, 12, and 15. These numbers are chosen because of comparability with empirical applications.

C. Data type

One of the prominent conceptual differences between the three alternative MLMDS methods, is that MAXSCAL is developed for nonmetric data, while MULTISCALE and PROSCAL are developed for metric data. We include in our design both metric and nonmetric data, generated on the basis of transformations of the distances. Details of these transformations are provided in the appendix.

D. Number of dimensions

In the majority of empirical applications of MDS a dimensionality of 2 is selected, while solutions of a dimensionality higher than three are very rare (Shepard 1972). Therefore, we include 2 and 3 dimensions in our design.

E. Error distribution

In this study, we assume that the distances are error-perturbed. We include the normal, the noncentral chi-square, and the uniform distributions in our design. This enables a comparison of each of the methods under the condition that the actual error distribution optimally matches that assumed in the model. We select the uniform distribution, because it is basically different from the noncentral chi-squared, respectively normal and lognormal distributions, which are assumed in PROSCAL respectively MULTISCALE and MAXSCAL. Including these distributions allows for an investigation of the robustness of the methods under alternative distributions.

F. Error variance level

The error variance level is operationalized by the ratio of the standard deviation of the error distribution and the standard deviation of the error free distances. We specify two levels of this ratio, namely a low error variance level (25 %) and a high level (75 %), which correspond to levels used in previous studies (e.g. Weeks and Bentler 1979).

G. Error variance differences

The three MLMDS methods differ in options of allowing the error variance to vary between stimuli and between subjects (Table 1). In the design, we specify a factor that represents these differences in error variance (see appendix).

*Analyses of the data sets*

In this study the data sets will be analyzed with the following versions of the MLMDS programs: MULTISCALE II (Ramsay 1982 and 1991a), MAXSCAL-2.1 (Takane 1979 and 1981), and the updated version of PROSCAL (MacKay and Zinnes 1991). In the remainder of this paper these will be referred to simply as MULTISCALE, MAXSCAL, and PROSCAL.

[ Insert Table 3 about here ]

In the MULTISCALE, MAXSCAL, and PROSCAL analyses, those program options are used that optimally correspond to the conditions under which the data are generated (Table 3). In PROSCAL we applied the Davidon-Fletcher-Powell optimization method as it is faster than the direct search method and produces reasonably good estimates (MacKay and Zinnes 1982). In MULTISCALE a power transformation of the data is estimated for each subject, whereas for MAXSCAL and PROSCAL no transformation is estimated. In the MAXSCAL analyses, the metric data sets are analyzed using a linear constraint on the category boundaries, and the nonmetric data sets without such a constraint. In both cases the boundaries are constrained to be equal across subjects. With respect to the error distribution, MULTISCALE and MAXSCAL options of a normal distribution are used if the true distribution is normal, and lognormal if the true distribution is non-central chi-square or uniform. PROSCAL assumes a non-central chi-square distribution for each data set. In the design of the study the error variance level is either equal across stimuli and subjects, or varies across stimuli, or varies across subjects. As MULTISCALE includes each of these options, we use the options that perfectly match the structure of the data sets in this respect. PROSCAL allows for error variance differences between stimuli but not between subjects, hence only the former option has been used. Though MAXSCAL allows for error variance differences between subjects, we decided not use this option as it results in a considerable increase in the number of parameters to be estimated, which turned out to cause insurmountable computational problems. Hence, to reduce the sparsity of the discrete data, in the MAXSCAL analyses no individual specific parameters are estimated, neither with respect to the category boundaries nor the error variances. Consequently, MAXSCAL treats the data matrices of the subjects as mere replications (Takane 1981). In addition, when individual parameters are present the asymptotic properties of the maximum likelihood estimates for any of the models concerned do not hold.

Analyses are performed for each data set for 1 to 4 dimensions. For several data sets, however, MAXSCAL turns out have problems to derive a solution with a single dimension. After a few iterations the program terminates the iteration process without giving a final configuration as output. In those situations, MAXSCAL was run again but now for 2 to 4 dimensions. These problems can be explained from the findings of Hubert and Arabie (1986, 1988), who showed that gradient based methods generally perform worse in case of a city-block distance model. As for Minkowski power distance models a unidimensional solution is equivalent to the city-block metric, the MLMDS are expected to encounter difficulties to derive satisfactory unidimensional solutions. In order to equate the iteration process across the three MLMDS methods, we standardized the maximum number of iterations to 200. If an analysis did not converge within 200 iterations, it was checked whether the final solution is satisfactory or an increase in the maximum number of iterations is needed. The methods differ in the definition of the convergence criterion for the improvement of the likelihood (Table 3). In MULTISCALE a fixed small value can be given for the improvement, in MAXSCAL the threshold value is specified relative to the value of the likelihood, and in PROSCAL the threshold value is related in a complex manner to various parameters of the model at hand. Since the formulation of the likelihoods also differs considerably between the three methods, equating these threshold values is next to impossible, and for each MLMDS method the default values are used in this study. Table 3 also presents equations for the respective numbers of parameters estimated for the three procedures.

*Evaluation criteria*

The performance of the three MLMDS methods is evaluated using a number of criteria that are clustered in three groups, namely according to whether they assess parameter recovery, goodness of fit of the model, or computational effort required.

Parameter recovery entails both the recovery of the coordinates and the identification of the correct dimensionality. The configurations derived by MLMDS with the simple Euclidean distance function are invariant under rotations. Therefore, recovery of the coordinates has to be assessed after rotating the original configuration and the derived configuration to maximum congruence (Peay 1988; Ten Berge 1977). Here we assess the equivalent recovery of the distances instead of the recovery of the coordinates. We use the correlation between the $I(I-1)/2 \times 1$ vectors of true distances and estimated distances for each subject, where the estimated distances are calculated on the basis of the true dimensionality.

For the selection of the correct dimensionality, a number of heuristics are available. In this study, we evaluate the Likelihood Ratio chi-square test (LR-test), the minimum AIC rule (Akaike 1974), and the minimum CAIC rule (Bozdogan 1987) for each of the three MLMDS methods. Likelihood ratio tests ($\alpha = 0.01$) are performed to test between the $M$-dimensional solution versus the $(M-1)$-dimensional solution. For PROSCAL, additional analyses are required to identify the dimensionality as the $M$-dimensional solution and the $(M-1)$-dimensional solution are not nested. This is due to fact that the noncentrality parameter of the non-central chi-square distribution is equal to the dimensionality of the solution, hence the likelihoods for different dimensionalities are based on the different density functions and are therefore not comparable. To compute correct likelihood ratio tests, the $(M-1)$-dimensional solution has to be specified through constraints on the $M$-dimensional solution. Hence, in PROSCAL, the minimum AIC or CAIC rule has to be applied for each successive pair of constrained and unconstrained solutions. For MULTISCALE and MAXSCAL, the AIC or CAIC are computed for each solution, and that solution is selected that yields the smallest value for AIC or CAIC. In addition, for MULTISCALE we apply the Likelihood ratio test with the correction factor for the chi-square value (CLR-test), as proposed by Ramsay (1980a). The percentage of times the dimensionality is correctly identified, is calculated for each method and each statistic at the various levels of the factors in the design (Table 2). As noted above, the conditions needed for these statistics to have their asymptotic properties may not hold in testing for $M$ versus $M+1$ dimensions.

We use the correlation between the $NI(I-1)/2 \times 1$ vectors of observed dissimilarities and estimated distances, calculated from the true dimensionality, as a measure of the goodness-of-fit of the MLMDS solutions. The latter vector is computed as $I_N \otimes Vec(\mathbf{D})$, where $I_N$ is a $N \times 1$ unit vector and $\mathbf{D}$ is a lower-diagonal matrix with estimated distances.

The computational effort required is measured by the CPU time needed (on a 486DX 33Mhz PC).

*Experimental design*

A full factorial design based on the factors presented in Table 2 would require $2^3 3^4 = 648$ data sets. By assuming interactions of three factors or more to be negligible, we can apply a fractional factorial design to reduce the number of data sets required, while retaining a high power to detect main effects and 2-factor interactions, as will be shown below. We adopt a fractional design of 162 data sets from which all main-effects and two-factor interactions can be estimated (Connor and Young 1961). Each of these data sets is analyzed with the three MLMDS methods. This results in 486 solutions for which the above mentioned evaluation criteria are calculated. Repeated measurement analyses of variance are used to test the effects of the seven factors and MLMDS methods on the evaluation criteria (three repeated measures for each of the 162 data sets). Partial omega squared values (Keren and Lewis 1979) are reported to indicate effect size.

On the basis of Cohen (1988) we assess the power of the F-tests in the ANOVA's. We want to detect at least medium-sized effects (Cohen 1988), corresponding to about 6 percent of the total variance accounted for, at a significance level of $\alpha = 0.01$. We assume that second and higher order interactions are negligible, and correct for the fact that the F-tests are part of an ANOVA

model with 87 degrees of freedom in total. Tests for main effects of factors with 2 or 3 levels have a power of 0.99 respectively 0.98 in our design. An F-test of a 2-factor interaction (both factors at 3 levels) has in a power of 0.95. We consider these levels of power of detecting the above-mentioned effects to be highly satisfactory.

## *RESULTS*

*Recovery of the distances*

The results of the ANOVA of the correlations between true and estimated distances (after a Fisher transformation $F(\rho) = \frac{1}{2}(\log(1+\rho) - \log(1-\rho)))$ are presented in Table 4. For each of the three MLMDS methods, the average correlations of the true distances with the estimated distances for each factor level are given in Table 5.

[ Insert Table 4 about here ]

[ Insert Table 5 about here ]

The first observation to be made from Tables 4 and 5, is the fact that all three MLMDS methods recover the true distances very well under all circumstances, as all mean correlations are above 0.90. Furthermore, the variance between data sets is much larger than the variance due to the difference between MLMDS methods. The mean correlations do not differ significantly between the MLMDS methods, but some of the interactions between MLMDS methods and the design factors are significant. The 3-factor interactions between the MLMDS method, the number of subjects, and the number of stimuli, respectively between the MLMDS method, the data type, and the error level arise because MAXSCAL recovers the distances less well in case of large data sets with 14 subjects and 15 stimuli (average correlation of 0.949), and ordinal data with a high error level (average correlation of 0.916). The performance of MULTISCALE and MAXSCAL is relatively better when the error distribution is normal, whereas PROSCAL performs relatively better when the error distribution is uniform. This latter result is unexpected, but may be caused by the bounded range of the uniform distribution. Hence, the MLMDS methods tend to perform somewhat better when assumption of the error distribution is valid, though they are rather robust against violations of these assumptions. Whereas PROSCAL performs best when there are differences in error variance, MULTISCALE and MAXSCAL perform best when there are differences in error variance between stimuli. For MAXSCAL this result is somewhat surprising, since such differences are not accommodated for in the analyses. MULTISCALE recovers the true distances less well when the error variance differs between subjects, even though the model accommodates these differences. The same holds for PROSCAL's performance when the error variance differs between stimuli. These results may be attributed to a larger number of parameters that needs to be estimated in these situations (Table 3).

The number of subjects, the data type, the number of dimensions, and the error variance turn out to have significant and substantial main effects on the recovery of the true distances (Tables 4 and 5). The total number of data points available is expected to have a positive effect on distance recovery. It turns out that the number of subjects indeed does have this effect, whereas the number of stimuli does not. This might be due to the fact that an increase in the number of stimuli not only causes an increase in the amount of data but also an increase in the number of parameters to be estimated, which tends to deteriorate the performance of the algorithms. Configurations of dimensionality 2 are better recovered than those of dimensionality 3, which is probably also related to an increase in number of parameters to be estimated. Furthermore, the three MLMDS methods recover the true distances better from metric data than from nonmetric data, which is most apparent in case of a non-central chi-square distribution. An increase of the number of subjects, however, considerably decreases this difference between the recovery from metric and non-metric data. Hence, information on a larger sample of subjects diminishes the negative effect of the ordinal nature of dissimilarity data.

As expected, a higher error variance level results in a decrease in the recovery of the distances. Furthermore, error variance differences between subjects turn out to reduce recovery, whereas differences between stimuli do not have such an effect. This effect should be seen in the context of the different performance of the three methods in this respect. The interactive effects of the error distribution and the error variance differences respectively the dimensionality, are caused by a relatively good distance recovery for the combination of a non-central chi-square distribution and error variance differences between stimuli and the absence of the dimensionality and differences in error variance effects in the case of a uniform distribution. Note that the main effect of the error distribution is insignificant and extremely small.

*Recovery of the dimensionality*

The results of a logistic regression analysis to analyze the effects of the factors in our design, the MLMDS methods, and the selection criteria (omitting the CLR-test) on the extent to which the true dimensionality is correctly recovered is presented in Table 6. Here, a zero indicates incorrect and a one indicates correct recovery of the dimensionality. Table 7 summarizes the extent to which the true dimensionality is underestimated, estimated correctly, or overestimated by each of the criteria for each of the three MLMDS methods under each of the conditions in the design.

[ Insert Table 6 about here ]

[ Insert Table 7 about here ]

As shown in Tables 6 and 7, there are significant differences between the MLMDS methods with respect to the recovery of the true number of dimensions. Across the selection criteria, for MAXSCAL and PROSCAL the true number of dimensions is identified somewhat more frequently than for MULTISCALE. However, there is an interaction effect between the MLMDS methods and the selection criteria: MULTISCALE performs best with the conservative CAIC heuristic, whereas MAXSCAL and PROSCAL perform better with the less conservative criteria: the LR-test and AIC. As expected, for MULTISCALE and MAXSCAL the LR-test and AIC tend to overestimate and CAIC tends to underestimate the dimensionality. Apparently, the penalty imposed on the log-likelihood by the CAIC-statistic tends to be too severe for these models. The penalty imposed by the AIC appears to be insufficient, resulting in overestimation of the dimensionality, For MULTISCALE, the correction factor for the LR-test as suggested by Ramsay (1980a) corrects for the tendency of the LR-test to overestimate dimensionality, and the recovery of the dimensionality with this CLR-test is even slightly better than with CAIC. The CLR-test does not have a serious bias towards either overestimation or underestimation. The recovery of the dimensionality for MAXSCAL is best by the LR-test. Finally, all three selection criteria tend to underestimate the dimensionality for PROSCAL, especially if the true number of dimensions is three. Due to the problems with the one-dimensional solution mentioned previously, the three MLMDS methods, and especially MAXSCAL, tend not to underestimate the true number of two dimensions. As can be derived by comparing the percentage of underestimation for a true dimensionality of 2 to that of 3, it indeed rarely happens that the MLMDS methods identify the unidimensional solution as the most appropriate one. In several cases, MAXSCAL did not provide a unidimensional solution at all. Since the true dimensionality is 2 or 3, this somewhat inflates the percentages of correct recovery. For PROSCAL, the LR-test and AIC perform about equally well and clearly better than CAIC, which imposes too severe penalties on the likelihood.

As expected, an increase in the number of subjects has a positive effect on the recovery of the true number of dimensions. For all MLMDS methods and criteria this increase is substantial for 2 to 8 subjects, but the effects marginalizes for a further increase to 14 subjects. The main effect of the number of stimuli is not significant, but as can be derived from Table 7, a higher number of stimuli results in general in a higher dimensionality being indicated by the criteria, regardless of the true number of dimensions, which causes the interactive effect with the true number of dimensions to be significant. When the number of stimuli increases, the true number of dimensions is more frequently indicated for PROSCAL, but less frequent for MAXSCAL, while for MULTISCALE such effects are absent. MULTISCALE recovers the true dimensionality clearly better from metric data than from nonmetric data, whereas for PROSCAL and MAXSCAL no differences are found.

The main effect of distribution of the error is not significant, but the interaction with the MLMDS methods is. In accordance with the theoretical framework of the methods, for PROSCAL the true dimensionality is identified most often if the error is chi-square distributed, and while for MAXSCAL (and to a lesser extent MULTISCALE) this is the case with normal distributed error. As expected, the true number of dimensions is recovered most frequently in case of low error variance level and in case of no error variance differences between subjects or stimuli, with minor differences across methods.

*Fit of the dissimilarities*

The fit of the models is studied by correlating the vector of the dissimilarity data with the vector of derived distances. As measure of the recovery of true distances, mean correlations are computed and an ANOVA is performed on the Fisher transformation of the correlations. Table 8 presents the ANOVA results, and Table 9 the mean correlations.

[ Insert Table 8 about here ]

[ Insert Table 9 about here ]

An increase in the number of subjects or the number of stimuli results in a decrease in the fit of the models to the data. This is due to the fact that in such situations there is less opportunity to adjust the model parameters to individual data points. Especially if the error level is high, an increase in the number of subjects causes a decrease in fit. As shown in Table 8, the type of data is one of the most important factors explaining the fit. The mean correlation for metric data is substantially higher than the mean correlation for nonmetric data. Two interactive effects involving the type of data are substantial and significant. The goodness-of-fit statistic is relatively low (average correlation of 0.693) under the condition of high error variance and nonmetric data, that is if the error component is relatively large compared to the information in the data. The fit is relatively good (average correlation of 0.864) under the conditions of metric data and error variance differences between stimuli. The number of dimensions turns out to have virtually no effect on the goodness-of-fit, which may be caused by the fact that here the fitted number of dimensions is equal to the true number of dimensions.

The factors pertaining to the error model have large effects on the goodness-of-fit. Especially an increase in error variance level affects the correlations between the data and the derived distances negatively. The effects of the error distribution and the error variance differences are also considerable. Data with a non-central chi-square distribution or with error variance differences between stimuli are represented best. Moreover, the interaction effect between these two factors consists mainly of an additional positive effect of the combination of the non-central chi-square distribution with error variance differences between stimuli (average correlation of 0.875). These may be caused by the skewness of the non-central chi-square distribution, where the extreme values that arise inflate the fit measures, and by the congruence of the non-central chi-square distribution and error variance differences between stimuli.

The differences between the MLMDS methods with respect to the fit are very small, especially compared to the variance due to the between data sets design factors. The MLMDS methods main effect is not significant, though there are some significant, but relatively small, interactions with the number of subjects and error variance differences. MAXSCAL performs relatively well with a small number of subjects, except when there are error variance differences between these subjects. This is expected, since we did not accommodate such differences in the MAXSCAL analyses. Moreover, MULTISCALE and MAXSCAL provide relatively good fit when there are error variance differences between stimuli. This is consistent with the finding that the models recover the true distances well when the error variance differs between stimuli, although a theoretical explanation is lacking. We conclude that all three methods derived solutions that fit the data reasonably well, that is with an average correlation between the derived distances and the dissimilarity data of 0.790.

*Computational effort*

The computational effort required is measured by the CPU time needed (on a 486DX 33Mhz PC). The results of ANOVA of the CPU times are given in Table 10. For each MLMDS program, the mean CPU time according to the factor levels is presented in Table 11.

[ Insert Table 10 about here ]

[ Insert Table 11 about here ]

The most prominent result from Tables 10 and 11 is that there are substantial differences between the MLMDS programs with respect to the CPU time. MULTISCALE takes on average about 30 seconds, which is much less than PROSCAL (average of 3.54 minutes) and MAXSCAL (average of 6.45 minutes). Besides this main effect, their are some interesting interactions of the MLMDS methods with other factors in the design. The CPU time used by the MLMDS methods depends highly on the number of stimuli. For MAXSCAL this effect is largest; an increase from 9 to 15 stimuli results in an increase in CPU time from 1.62 to 12.70 minutes on average. For PROSCAL and to a lesser extent for MULTISCALE, an increase in the number of subjects causes an increase in the CPU time. MAXSCAL requires most CPU time in the case of a small number of subjects. The effect of the measurement scale of the data is most apparent for MAXSCAL. As expected, MAXSCAL takes more time analyzing metric data, because it has to evaluate equation (1) for 50 categories. MULTISCALE takes more time analyzing nonmetric data. Estimating parameters related to error variance differences between stimuli or subjects in MULTISCALE and PROSCAL causes the CPU time to increase, especially when allowing for difference between stimuli in MULTISCALE.

## *EXAMINING THE LOCAL OPTIMA PROBLEM*

For each of the MLMDS methods convergence to a global optimum is not guaranteed. A solution that is generally used is to start the iteration process with a rational start, e.g. a metric decomposition of the average dissimilarity matrix. In the Monte Carlo study presented in the previous sections, such rational starts for the stimulus coordinates have been used for each of the methods. The extent to which the MLMDS methods may converge to local optima is investigated in a separate Monte Carlo study presented in this section.

To examine the convergence to local optima for the MLMDS methods, synthetic data sets are generated for the following fixed factor levels: nonmetric (7-point) dissimilarities, two dimensions, a uniform error distribution, high error variance, and no error variance differences between subjects or stimuli. The data sets vary, however, in two factors that may affect convergence to local optima since asymptotically the problem of local optima vanishes: the number of subjects (2, 8, 14) and the number of stimuli (9, 12, 15). With two replications in each cell, this 3×3 design results in 18 data sets (Table 12). Each of these data sets has been analyzed using 12 sets of starting values for the stimulus coordinates, namely rational starting values from the defaults in each of the three MLMDS programs, the true coordinates, and 10 different sets of randomly generated coordinates. In the case of MAXSCAL, however, several analyses with random starts failed to begin the iterative process because of a log domain error. For each analysis the value of the likelihood function at convergence is examined. If the difference in the likelihood at convergence between different analyses of the same data set is within the pre-specified level of convergence (Table 3), these analyses are reported to result in the same solution. Table 12 presents for each MLMDS method the extent to which the 12 analyses of the 18 data sets result in the same solution.

[ Insert Table 12 about here ]

The MLMDS methods frequently converge to local optima (Table 12), especially if random starts are used. Rational starting values, however, provide a good (though not perfect) solution to the local optima problem. Across all methods and data sets for 9 out of 54 analyses the rational start is outperformed by the true start and/or one or more random starts, indicating that the global optimum is not reached. This number slightly differs between the three MLMDS methods (MULTISCALE: 3; MAXSCAL: 5; PROSCAL: 1). This result can be explained partly by the fact that the convergence criterion for the improvement of the likelihood value is much smaller for MAXSCAL than for MULTISCALE and PROSCAL (Table 3). For those cases where the rational start did not result in the global optimum, the extent to which the rational start is outperformed in terms of the final log-likelihood is extremely small. The maximum relative improvement of the solution from the rational start by the best solution is only 0.12 % for MULTISCALE, 0.21 % for MAXSCAL and 0.24 % for PROSCAL. It is interesting to note that a global optimum was not found starting the algorithms from the true parameter values in 15 out of 54 analyses (MULTISCALE: 2; MAXSCAL: 6; PROSCAL: 7). In addition, it was expected that the number of stimuli and/or the number of subject affects the local optima problem, but these two factors do not seem to have an effect in this study (Table 12). We attribute these two results to the uniform error distribution, which results in misspecification for each of the three models.

## *CONCLUSIONS*

The three MLMDS methods perform very well with respect to recovering the true distances, even if the number of subjects is as small as two. This finding confirms results of previous Monte Carlo studies (Büyükkurt and Büyükkurt 1990; MacKay and Zinnes 1981; Ramsay 1977; Storms, 1995; Takane 1978b). Additionally, this study shows that MULTISCALE, MAXSCAL, and PROSCAL hardly differ in the extent to which the true distances are recovered. For each of the MLMDS methods, the recovery of the true distances increases with an increase of the number of subjects, is better for metric data than for nonmetric data, decreases with an increase of the dimensionality of the true configuration, and decreases with an increase of the error variance.

The results of the fit to the dissimilarity data are consistent with those on distance recovery: the differences between the three MLMDS methods are small. As the number of data points increases, due to an increase in the number of subjects and/or the number of stimuli, the MLMDS methods have less opportunity to adjust the model parameters to each individual data point. This results in lower goodness-of-fit of the model to the entire data set. As expected, low error variance levels result in better model fit as compared to a higher error variance level.

The recovery of the true dimensionality differs between the three MLMDS methods. For MAXSCAL and PROSCAL the true dimensionality was identified more frequently than for MULTISCALE, across the criteria investigated (LR-test, AIC, and CAIC). For MULTISCALE and MAXSCAL, the LR-test and AIC tend to overestimate the dimensionality, while CAIC tends to underestimate the dimensionality. For PROSCAL all three criteria tend to underestimate the dimensionality. Hence, the penalty imposed on the likelihood by AIC is generally insufficient, while that imposed by CAIC is generally too severe. Both AIC and CAIC can only be used as heuristics for selecting the appropriate number of dimensions, since they depend upon the same regularity conditions needed for the likelihood ratio test to have its asymptotic distribution under the null hypothesis. These conditions may not hold when testing for $M$ versus $M+1$ dimensions. For MULTISCALE, the corrected LR-test proposed by Ramsay (1980a) appears to correct for overestimation of the number of dimensions, although not completely (see also Storms, 1995). The properties of the criteria appear to improve when the number of subjects increases, when the underlying distributional assumptions are correct, and when the error variance is relatively low. Because the asymptotic properties of the statistics are lacking, their usefulness depends on the specific model, the number of parameters estimated, and the data at hand: clearly an undesirable situation.

Though CPU time becomes a less important criterion as computers become faster, it may remain important especially if a substantial number of analyses have to be performed. Considering the CPU time, MULTISCALE is much faster than PROSCAL and MAXSCAL. Especially with metric data, with a large number of scale categories for which MAXSCAL was not designed, MAXSCAL requires significantly more CPU time.

The most frequent criticism of MLMDS, is the fact that assumptions have to be made about the error model (see e.g. the discussion of Ramsay (1982) or Carroll and Arabie (1980)). The proposition made by skeptics is that the performance of the MLMDS methods will critically depend on the correctness of the distributional assumptions. This study supports Büyükkurt and Büyükkurt (1990), Storms (1995), Storms and Delbeke (1991), and Takane (1978b) and shows that for each of the three MLMDS methods neither the recovery of the true configuration nor the fit of the model to the dissimilarity data is seriously influenced by violations of distributional assumptions. However, for MAXSCAL and

PROSCAL the identification of the true number of dimensions appears to dependent more on the correctness of the assumptions, whereas for MULTISCALE this is less apparent. This finding supports and supplements the findings of Takane (1978b) and Storms (1995), but is contrary to the results of Büyükkurt and Büyükkurt (1990).

Traditional MDS methods have been developed for both metric data (Torgerson 1952 and 1958; Carroll and Chang 1970) and nonmetric data (Kruskal 1964; Shepard 1962). Within the MLMDS framework MULTISCALE and PROSCAL are tailored to metric data, whereas MAXSCAL assumes the data to be nonmetric. Nonmetric data contain in general less information about the true configuration as compared to metric data. The deterioration of the recovery of true distances and lower goodness-of-fit of nonmetric data as compared metric data revealed in this study are therefore expected. The metric methods MULTISCALE and PROSCAL perform similar to the nonmetric method MAXSCAL in case of nonmetric data. Hence, in accordance with previous research on metric quality of ordered scales in general (Srinivasan and Basu 1989), the conclusion seems justified that pairwise dissimilarity judgments on 7-point scales may very well be treated as metric.

The three MLMDS methods suffer from the danger of converging to local optima. So far, little research has addressed the extent to which these methods converge to local optima in finite samples. This paper shows that if random starts are used, the MLMDS methods converge frequently to local optima for sample sizes used in this study. Using rational starting values, however, provides a reasonably good solution to the problem of local optima. In about 80 % of the cases the solution from rational starting values could not be outperformed by 10 solutions from random starting values and a solution from the true values as starting values. Furthermore, in those cases where the solution from a rational start is outperformed the difference is extremely small in terms of the relative improvement of the likelihood. Moreover, if the rational starts used are informative, their selection rather than the use of random starts may guard against incorrect inferences in samples with limited information. Rational starts restrict the estimated solution to a region of the parameter space that on substantive grounds is more likely to contain the true solution. Hence, by employing a rational start the probability that the true maximum is in an a-priori unacceptable region of the parameter space is decreased (Manton, Woodburry, and Tolley 1995, p. 72). We therefore argue that rational starts are preferable, both from a theoretical and empirical perspective. These rational starts were used throughout our study. The findings reported are most likely not due to difference in the tendency to converge to local optima among the methods.

## *DISCUSSION*

Marketing researchers who intend to apply MLMDS may base their choice between the alternative methods on the results presented in this paper. We found that all three MLMDS methods perform extremely well on two important criteria, namely recovering true distances and fitting error-perturbed dissimilarities. The differences between the methods with respect to these criteria are negligible. MAXSCAL and PROSCAL recover the true dimensionality somewhat more often as compared with MULTISCALE. These differences are rather small, however, and depend heavily on the selection criterion used. MULTI-SCALE, on the other hand, takes much less CPU time. PROSCAL outperforms the other two methods with respect to avoiding local optima. MAXSCAL is ranked last concerning both CPU time and avoiding local optima. Hence, none of the three MLMDS methods clearly dominates the other methods with respect to all evaluation criteria.

Generally, the selection of an MLMDS method may involve more than the results of Monte Carlo simulation studies. User-friendliness of the programs, for example, may affect the choice made. Our experience with the use of the MLMDS programs is that all three are relatively easy to use. The programs use control files and are processed in batch mode. Well-written manuals are available for each program, where the MULTISCALE manual is most comprehensive. The MULTISCALE and PROSCAL programs are still regularly updated (MacKay and Zinnes 1991; Ramsay 1991a). Furthermore, each of the three MLMDS methods has a number of unique features. Some of these features have been omitted in this Monte Carlo evaluation in order to make the methods comparable, which may have biased our study towards converging results for the three different MLMDS methods. Some types of data not considered, such as triadic combinations or conditional rankings, require analysis with MAXSCAL. MULTISCALE, on the other hand, is the only MLMDS method that allows for individual weighting of the dimensions. Furthermore, Ramsay (1980b, 1991b) extended MULTISCALE to the simultaneous analysis of attribute ratings, preferences, and dissimilarities. Within the framework of PROSCAL, it is possible also to analyze preference data (MacKay and Zinnes 1986). Moreover, in situations where it is necessary to assume stimuli to be error-perturbed and the error variance to differ between dimensions, for example while studying newly developed products with well-known but also some new attributes, PROSCAL has to be applied. Hence, specific characteristics of a study may direct a marketing researcher towards the use of one of the three MLMDS methods.

Considering both the results of the Monte Carlo studies presented and more practical and study-specific issues raised above, we tend to favor PROSCAL and MULTISCALE over MAXSCAL for application by marketing researchers.

Though this paper provides insight in the relative performance of the three MLMDS methods under a wide range of circumstances, a number of issues remain to be investigated more closely. The sensitivity of the methods to convergence to local optima could be addressed in a more extensive Monte Carlo study to complement the results presented in this paper. Additionally, whether metric MLMDS methods perform well on nonmetric data may critically depend on the number of scale categories. In this paper, it is shown that a 7-point scale contains enough information to warrant metric analyses, but future studies may focus on the relative performance of the methods when the number of categories is for example 3 or 5. Finally, research into alternative criteria for selecting the dimensionality, such as Monte Carlo tests and modified AIC criteria and the criteria used in this study, is needed.

# APPENDIX. THE DATA GENERATION PROCESS

For each of the 162 data sets to be generated, a 3-step procedure is followed.

In the first step, a configuration is generated for a given number of stimuli and dimensions. The coordinates of the stimuli for each dimension are drawn from a uniform distribution on the interval (0, 10). From the resulting configuration error-free distances $d_{ij}$ between the stimuli are computed according to equation (2). We set $w_{mn} = 1$ for all $m$ and $n$, as only MULTISCALE allows for estimation of dimensional weights.

In the second step, the error-free configuration from the previous step is error-perturbed in accordance with the the three factors pertaining to the error model, namely the error distribution, the error variance level, and whether or not there are error differences between stimuli or subjects. The error component $e_{ijn}$ is generated applying the following formula:

$$e_{ijn} = u_n(v_i + v_j)\eta_{ijn}, \tag{9}$$

where $\eta_{ijn}$ is drawn from the appropriate distribution with variance level $\sigma^2$, $u_n$ is either drawn from a uniform distribution on the interval (0.5, 1.5) or $u_n = 1$ (error variance differences between subjects or not), and $v_i$ and $v_j$ are either drawn from a uniform distribution on the interval (0, 1) or $v_i = v_j = 0.5$ (error variance differences between stimuli or not). The standard deviation of the error $\sigma$ is computed as 0.25 or 0.75 times the standard deviation of the error-free distances (error variance level low or high). In case of the normal and uniform distributions, the error term $e_{ijn}$ is added directly to the error-free distances $d_{ij}$. To generate non-central chi-square distributed error the stimulus coordinates are error-perturbed using a normal distribution. The error variance is equated among the three error distributions by relating the error variance in coordinates to the error variance in distances using equations (3) and (4) from Zinnes and MacKay (1983). The error variance level for a stimulus is approximated by imputing the mean and variance of the true distances, multiplied by $u_n(v_i + v_j)$, into the equation relating the error in coordinates and the error in distances.

In the third step, the appropriate transformation is applied to obtain the final dissimilarity data. First, error-perturbed distances smaller than 0 were replaced by a small positive value. In case of metric similarity data no further transformation is applied. For the analysis with MAXSCAL, however, the metric data were transformed to correspond to 50-point metric scales. A linear transformation is used, which made the maximum distance in each data set equal to 50, and the minimum distance equal to 1. All other distances are rounded off to the nearest integer. In order to obtain nonmetric data we transform the data to correspond to dissimilarity ratings on 7-point scales. This is done for each subject separately, by drawing 6 random numbers from a lognormal distribution with mean and variance equal to the mean and variance of the error-perturbed distances. These 6 numbers are ordered and form the bounds of 7 categories as in equation (1). Next, each observed distance is assigned to one of the categories and replaced by the code of the appropriate category. This procedure results in a data matrix with simulated dissimilarities among stimuli for each of a specified number of subjects.

*REFERENCES*

Akaike, Hirotugu (1974), "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, AC-19, 716-723.

Bijmolt, Tammo H.A. and Michel Wedel (1995), "The effects of alternative methods of collecting similarity data for multidimensional scaling," *International Journal of Research in Marketing*, 12 (4), 363-371.

Bozdogan, Hamparsum (1987), "Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions," *Psychometrika*, 52 (September), 345-370.

Büyükkurt, B. Kemal and Meral Demirbag Büyükkurt (1990), "Robustness and small-sample properties of the estimators of probabilistic multidimensional scaling (PROSCAL)," *Journal of Marketing Research*, 27 (May), 139-149.

Carroll, J. Douglas and Phipps Arabie (1980), "Multidimensional scaling," *Annual Review of Psychology*, 31, 607-649.

__________ and Jih-Jie Chang (1970), "Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition," *Psychometrika*, 35 (September), 283-319.

Cohen, Jacob (1988), *Statistical power analysis for the behavioral sciences.* second edition. Hillsdale: Lawrence Erlbaum.

Connor, W.S. and S. Young (1961), *Fractional factorial designs for experiments with factors at two and three levels.* National Bureau of Standards, Applied Mathematics Series, 58.

Cooper, Lee G. (1983), "A review of multidimensional scaling in marketing research," *Applied Psychological Measurement*, 7 (Fall), 427-450.

Davison, Mark L. (1983), *Multidimensional scaling.* New York: Wiley.

Green, Paul E. (1975), "Marketing applications of MDS: Assessment and outlook," *Journal of Marketing*, 39 (January), 24-31.

Greenberg, Barnett A., Jac L. Goldstucker, and Danny N. Bellenger (1977), "What techniques are used by marketing researchers in business?," *Journal of Marketing*, 41 (April), 64-65.

Hefner, R. (1958), *Extensions of the law of comparative judgment to discriminable and multidimensional stimuli.* Dissertation, Department of Psychology, University of Michigan.

Hubert, Lawrence and Phipps Arabie (1986), "Unidimensional scaling and combinatorial minimization." In: De Leeuw, et al., editors, *Multidimensional data analysis.* Leiden: DSWO Press, 181-196.

__________ and __________ (1986), "Relying on necessary conditions for optimization: unidimensional scaling and some extensions." In: Bock, editor, *Classification and Related Methods of Data Analysis.* Amsterdam: Elsevier, 463-472.

Keren, Gideon and Charles Lewis (1979), "Partial omega squared for ANOVA designs," *Educational and Psychological Measurement*, 39 (Spring), 119-128.

Kinnear, Thomas C. and James R. Taylor (1991), *Marketing research: An applied approach*, Fourth edition. New York: McGraw-Hill.

Kruskal, Joseph B. (1964), "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, 29 (March), 1-27.

__________, Forest W. Young, and Judith B. Seery (1973), *How to use KYST, a very flexible program to do multidimensional scaling and unfolding.* Unpublished manuscript, Bell Laboratories.

MacKay, David B. (1983), "Alternative probabilistic scaling models for spatial data," *Geographical Analysis*, 15 (July), 173-186.

__________ (1989), "Probabilistic multidimensional scaling: an anisotropic model for distance judgments," *Journal of Mathematical Psychology*, 33 (June), 187-205.

__________ and Cornelia Dröge (1990), "Extensions of probabilistic perceptual maps with implications for competitive positioning and choice," *International Journal of Research in Marketing*, 7 (4), 265-282.

__________ and Joseph L. Zinnes (1981), "Probabilistic scaling of spatial distance judgments," *Geographical Analysis*, 13 (January), 21-37.

__________ and __________ (1982), *PROSCAL: a program for Probabilistic scaling.* Discussion paper 218, School of Business, Indiana University.

__________ and __________ (1986), "A probabilistic model for the multidimensional scaling of proximity and preference data," *Marketing Science*, 5 (Fall), 325-34.

__________ and __________ (1991), *PROSCAL update - June, 1991.* Unpublished manuscript, Indiana University.

Manton, K.G., M.A. Woodburry, and D.H. Tolley (1995), *Statistical applications using fuzzy sets.* New York: Wiley.

Peay, E.R. (1988), "Multidimensional rotation and scaling of configurations to optimal agreement," *Psychometrika*, 53 (June), 199-208.

Ramsay, James O. (1969), "Some statistical considerations in multidimensional scaling," *Psychometrika*, 34 (June), 167-182.

__________ (1977), "Maximum likelihood estimation in multidimensional scaling," *Psychometrika*, 42 (June), 241-266.

__________ (1978a), "Confidence regions for multidimensional scaling analysis," *Psychometrika*, 43 (June), 145-160.

__________ (1978b), *MULTISCALE: four programs for multidimensional scaling by the method of maximum likelihood.* National Educational Resources.

__________ (1980a), "Some small sample results for multidimensional scaling," *Psychometrika*, 45 (March), 139-144.

__________ (1980b), "The joint analysis of direct ratings, pairwise preferences, and dissimilarities," *Psychometrika*, 45 (June), 149-165.

__________ (1982), "Some statistical approaches to multidimensional scaling (and discussion), *Journal of the Royal Statistical Society*, A 145 (3), 285-312.

__________ (1991a), *MULTISCALE manual.* Unpublished manuscript, McGill University, August 16, 1991.

__________ (1991b), *MULTISCALE manual (Extended version).* Unpublished manuscript, McGill University, September 24, 1991.

Shapiro, Alexander (1986), "Asymptotic theory of overparameterized structural models," *Journal of the American Statistical Association*, 81, 142-149.

Shepard, Roger N. (1962), "The analysis of proximities: Multidimensional scaling with an unknown distance function, I and II," *Psychometrika*, 27, (June), 125-140 and (September), 219-246.

__________(1972), "Introduction to volume I." In: Shepard, Romney, and Nerlove, editors, *Multidimensional scaling: I. Theory.* New York: Seminar Press, 1-20.

Spence, Ian (1983), "Monte Carlo simulation studies, *Applied Psychological Measurement*, 7 (fall), 405-425.

__________and Stephan Lewandowsky (1989), "Robust multidimensional scaling," *Psychometrika*, 54 (September), 501-513.

Srinivasan, V. and Amiya K. Basu (1989), "The metric quality of ordered categorical data," *Marketing Science*, 8 (Summer), 205-230.

Storms, Gert (1995), "On the robustness of Maximum Likelihood scaling for violations of the error model," *Psychometrika*, 60 (June), 247-258.

__________and Luc Delbeke (1992), "The irrelevance of distributional assumptions on reaction times in multidimensional scaling of same/different judgment tasks," *Psychometrika*, 57 (December), 599-614.

Takane, Yoshio (1977), *Statistical procedures for nonmetric multidimensional scaling.* Dissertation, Department of Psychology, University of North Carolina.

__________(1978a), "A maximum likelihood method for non-metric multidimensional scaling: I. the case in which all empirical pairwise orderings are independent - theory," *Japanese Psychological Research*, 20 (1), 7-17.

__________(1978b), "A maximum likelihood method for non-metric multidimensional scaling: I. the case in which all empirical pairwise orderings are independent - evaluations, *Japanese Psychological Research*, 20 (3), 105-114.

__________(1979), *How to use MAXSCAL-2.1.* Unpublished manuscript, McGill University.

__________(1981), "Multidimensional successive categories scaling: a maximum likelihood method," *Psychometrika*, 46 (March), 9-28.

__________(1982), "The method of triadic combinations: a new treatment and its applications," *Behaviormetrika*, 11, 37- 48.

__________and J. Douglas Carroll (1979), *How to use MAXSCAL-4.1.* Unpublished manuscript, Bell Laboratories.

__________and J. Douglas Carroll (1981), "Nonmetric maximum likelihood multidimensional scaling from directional rankings of similarities," *Psychometrika*, 46 (December), 389-405.

Ten Berge, Jos M.F. (1977), "Orthogonal procrustes rotation for two or more matrices," *Psychometrika*, 42 (March), 267-276.

Thurstone, L.L. (1927), "A law of comparative judgment," *Psychological Review*, 34, 273-286.

Torgerson, Warren S. (1952), "Multidimensional scaling: I. theory and method," *Psychometrika*, 17, 401-419.

__________(1958), *Theory and methods of scaling.* New York: Wiley.

Weinberg, Susan L., J. Douglas Carroll, and Harvey S. Cohen (1984), "Confidence regions for INDSCAL using the jackknife and bootstrap techniques," *Psychometrika*, 49 (December), 475-491.

Weeks, David G. and Peter M. Bentler (1979), "A comparison of linear and monotone multidimensional scaling models," *Psychological Bulletin*, 86 (March), 349-354.

Wind, Jerry, Vithala R. Rao, and Paul E. Green (1991), "Behavioral methods," in: *Handbook of consumer behavior*, Thomas S. Robertson and Harold H. Kassarjian, eds. Englewood Cliffs: Prentice-Hall. 507-532.

Young, Forest W. and Robert M. Hamer (1987), *Multidimensional scaling: history, theory, and applications.* Hillsdale: Lawrence Erlbaum.

__________and Rostyslaw Lewyckyj (1979), *ALSCAL-4 user's guide.* Unpublished manuscript, University of North Carolina.

Zinnes, Joseph L. and David B. MacKay (1983), "Probabilistic multidimensional scaling: complete and incomplete data," *Psychometrika*, 48 (March), 27-48.

TABLE 1

The Conceptual Framework of the MLMDS Methods

| | MULTISCALE | MAXSCAL | PROSCAL |
|---|---|---|---|
| *Measurement Model* | | | |
| Scale Level of the Data | metric | nonmetric | metric |
| Transformations | scale, power, or spline | none | none |
| *Representation Model* | | | |
| Distance Metric | Euclidean | Euclidean | Euclidean |
| Individual Differences | weighted or unweighted | unweighted | unweighted |
| *Error Model* | | | |
| Error Basis | distances | distances | coordinates |
| Distribution of Distances | normal, lognormal | normal, lognormal | noncentral chi-square |
| Options to Differ Error Variance Level | | | |
| between Stimuli | yes | no | yes |
| between Subjects | yes | yes | no |
| between Dimensions | no | no | yes |

TABLE 2

Factors in the Monte Carlo Simulation

| Factor | Levels |
|---|---|
| *Measurement Model* | |
| A. Number of Subjects | 2 |
| | 8 |
| | 14 |
| B. Number of Stimuli | 9 |
| | 12 |
| | 15 |
| C. Data Type | metric |
| | nonmetric |
| *Representation Model* | |
| D. Number of Dimensions | 2 |
| | 3 |
| *Error Model* | |
| E. Error Distribution | normal |
| | noncentral chi-square |
| | uniform |
| F. Error Variance Level | 25 % |
| | 75 % |
| G. Error Variance Differences | no differences |
| | between stimuli |
| | between subjects |

TABLE 3

Program Options Used

| Factor | MULTISCALE | MAXSCAL | PROSCAL |
|---|---|---|---|
| Data Type | | | |
| Metric Data | metric, power transformation | linear, equal | metric |
| Nonmetric Data | metric, power transformation | ordinal, equal | metric |
| Error Distribution | | | |
| Normal | normal | normal | non-central chi-square |
| Noncentral Chi-square | lognormal | lognormal | non-central chi-square |
| Uniform | lognormal | lognormal | non-central chi-square |
| Error Variance Differences | | | |
| None | no differences | no differences | no differences |
| Between Stimuli | between stimuli | no differences | between stimuli |
| Between Subjects | between subjects | no differences | no differences |
| Algorithm Specifications | | | |
| Optimization method | Fisher's scoring method | Fisher's scoring method | Davidon-Fletcher-Powell |
| Convergence Threshold | $\Delta(\log L) < 0.05$ | relative $\Delta(\log L) < 0.0000025$ | variable ($\approx \Delta(\log L) < \frac{\log L}{1000}$) |
| Maximum Number of Iterations | 200 | 200 | 200 |
| Starting Procedure | rational | rational | rational |
| Number of Parameters * | $(I-2)M + 2N +$ | $(I-2)M + 2\delta_{metric} +$ | $(I-2)M +$ |
| | $\delta_{stim}(I-1) + \delta_{subj}(N-1)$ | $6(1-\delta_{metric})$ | $\delta_{stim}(I-1)$ |

* $\delta_{metric}$ equals 1 for metric data and 0 for nonmetric data

$\delta_{stim}$ equals 1 for error variance differences between stimuli, 0 otherwise

$\delta_{subj}$ equals 1 for error variance differences between subjects, 0 otherwise

TABLE 4

ANOVA of Distance Recovery

| Source * | Sum of Squares | d.f. | F | p | partial $\omega^2$ |
|---|---|---|---|---|---|
| Between Data Sets | | | | | |
| Within + Residual | 10.99 | 99 | | | |
| A. Number of Subjects | 67.08 | 2 | 302.20 | $< 0.001$ | 0.859 |
| B. Number of Stimuli | 0.39 | 2 | 1.75 | 0.180 | 0.034 |
| C. Data Type | 14.37 | 1 | 129.46 | $< 0.001$ | 0.567 |
| D. Number of Dimensions | 3.10 | 1 | 27.95 | $< 0.001$ | 0.220 |
| E. Error Distribution | 0.36 | 2 | 1.60 | 0.206 | 0.031 |
| F. Error Variance Level | 26.59 | 1 | 239.61 | $< 0.001$ | 0.708 |
| G. Error Variance Differences | 0.81 | 2 | 3.66 | 0.029 | 0.069 |
| A $\times$ C | 1.90 | 2 | 8.55 | $< 0.001$ | 0.147 |
| C $\times$ E | 1.56 | 2 | 7.04 | 0.001 | 0.125 |
| D $\times$ E | 1.31 | 2 | 5.90 | 0.004 | 0.106 |
| E $\times$ G | 9.54 | 4 | 21.49 | $< 0.001$ | 0.465 |
| Within Data Sets | | | | | |
| Within + Residual | 6.91 | 198 | | | |
| H. MLMDS Method | 0.07 | 2 | 0.94 | 0.390 | 0.009 |
| E $\times$ H | 1.33 | 4 | 9.53 | $< 0.001$ | 0.161 |
| G $\times$ H | 1.14 | 4 | 8.14 | $< 0.001$ | 0.141 |
| A $\times$ B $\times$ H | 1.06 | 8 | 3.80 | $< 0.001$ | 0.133 |
| C $\times$ F $\times$ H | 0.82 | 2 | 11.80 | $< 0.001$ | 0.106 |

* Only interactions significant at $\alpha = .01$ and with partial $\omega^2 > 0.10$ are reported.

TABLE 5

Mean Correlations between True Distances and Estimated Distances

| Factor | Level | MULTISCALE | MAXSCAL | PROSCAL | Row Mean |
|---|---|---|---|---|---|
| A. Number of Subjects | 2 | .902 | .910 | .918 | .910 |
| | 8 | .974 | .974 | .973 | .974 |
| | 14 | .984 | .973 | .981 | .979 |
| B. Number of Stimuli | 9 | .951 | .953 | .955 | .953 |
| | 12 | .950 | .955 | .957 | .954 |
| | 15 | .959 | .949 | .960 | .956 |
| C. Data Type | metric | .964 | .965 | .970 | .968 |
| | nonmetric | .943 | .940 | .945 | .943 |
| D. Number of Dimensions | 2 | .961 | .963 | .967 | .964 |
| | 3 | .946 | .942 | .947 | .945 |
| E. Error Distribution | normal | .958 | .955 | .956 | .956 |
| | noncentral chi-square | .949 | .950 | .956 | .952 |
| | uniform | .953 | .952 | .961 | .955 |
| F. Error Variance Level | 25 % | .975 | .975 | .973 | .974 |
| | 75 % | .931 | .930 | .941 | .934 |
| G. Error Variance Differences | no differences | .956 | .953 | .965 | .958 |
| | between stimuli | .960 | .962 | .953 | .958 |
| | between subjects | .944 | .942 | .954 | .947 |
| Column Mean | | .953 | .952 | .957 | .954 |

TABLE 6

Logistic regression of Selecting the Correct Dimensionality

| Source * | Wald | d.f. | p |
|---|---|---|---|
| A. Number of Subjects | 82.94 | 2 | $< 0.001$ |
| B. Number of Stimuli | 0.29 | 2 | 0.864 |
| C. Data Type | 21.60 | 1 | $< 0.001$ |
| D. Number of Dimensions | 70.99 | 1 | $< 0.001$ |
| E. Error Distribution | 2.97 | 2 | 0.227 |
| F. Error Variance Level | 53.23 | 1 | $< 0.001$ |
| G. Error Variance Differences | 30.56 | 2 | $< 0.001$ |
| H. MLMDS Method | 17.17 | 2 | $< 0.001$ |
| I. Selection Criteria | 3.18 | 2 | 0.204 |
| $A \times B$ | 16.00 | 4 | 0.003 |
| $A \times C$ | 24.52 | 2 | $< 0.001$ |
| $A \times D$ | 18.21 | 2 | $< 0.001$ |
| $A \times G$ | 15.45 | 4 | 0.004 |
| $B \times D$ | 36.48 | 4 | $< 0.001$ |
| $B \times H$ | 24.52 | 4 | $< 0.001$ |
| $B \times I$ | 18.94 | 4 | $< 0.001$ |
| $C \times D$ | 8.97 | 1 | 0.003 |
| $C \times F$ | 14.81 | 1 | $< 0.001$ |
| $C \times H$ | 28.70 | 1 | $< 0.001$ |
| $D \times F$ | 14.19 | 1 | $< 0.001$ |
| $D \times H$ | 24.74 | 2 | $< 0.001$ |
| $D \times I$ | 63.79 | 2 | $< 0.001$ |
| $E \times F$ | 10.43 | 2 | 0.005 |
| $E \times H$ | 32.70 | 4 | $< 0.001$ |
| $G \times H$ | 25.13 | 4 | $< 0.001$ |
| $H \times I$ | 27.67 | 4 | $< 0.001$ |

* Only interactions significant at $\alpha = .01$ are reported.

TABLE 7

Percentages of Selecting the Correct Dimensionality

| MLMDS methods | MULTISCALE | | | | MAXSCAL | | | PROSCAL | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Selection Criteria | LR-test | CLR-test | AIC | CAIC | LR-test | AIC | CAIC | LR-test | AIC | CAIC |
| A. Number of Subjects | | | | | | | | | | |
| 2 | 13 37 50 | 37 59 9 | 4 30 67 | 43 48 9 | 25 57 17 | 17 46 37 | 44 54 2 | 44 50 6 | 39 54 7 | 61 37 2 |
| 8 | 4 61 35 | 17 72 11 | 2 56 43 | 35 63 2 | 4 76 20 | 2 72 26 | 35 65 0 | 26 69 6 | 20 72 7 | 35 59 6 |
| 14 | 4 61 35 | 7 78 15 | 2 50 48 | 20 76 4 | 4 82 15 | 0 70 30 | 19 74 7 | 17 78 6 | 15 78 7 | 32 65 4 |
| B. Number of Stimuli | | | | | | | | | | |
| 9 | 11 65 24 | 35 63 2 | 6 52 43 | 46 50 4 | 15 82 4 | 9 74 17 | 39 61 0 | 39 57 4 | 35 61 4 | 52 46 2 |
| 12 | 4 54 43 | 19 74 7 | 0 46 54 | 30 67 4 | 11 74 15 | 6 61 33 | 35 63 2 | 30 63 7 | 22 70 7 | 41 54 6 |
| 15 | 6 41 54 | 7 67 26 | 2 37 61 | 22 70 7 | 7 59 33 | 4 54 43 | 24 69 7 | 19 76 6 | 17 72 11 | 35 61 4 |
| C. Data Type | | | | | | | | | | |
| metric | 7 72 21 | 19 79 3 | 1 65 33 | 30 69 1 | 12 73 15 | 5 67 28 | 27 73 0 | 31 64 5 | 28 65 6 | 33 62 5 |
| nonmetric | 6 35 59 | 22 57 21 | 4 25 72 | 36 56 9 | 10 70 20 | 7 59 33 | 38 56 6 | 27 67 6 | 21 70 9 | 52 46 3 |
| D. Number of Dimensions | | | | | | | | | | |
| 2 | 0 54 46 | 7 79 14 | 0 43 57 | 11 80 9 | 0 73 27 | 0 63 37 | 0 94 6 | 9 82 10 | 6 80 14 | 17 77 6 |
| 3 | 14 52 35 | 33 57 10 | 5 47 48 | 54 44 1 | 22 70 7 | 12 63 25 | 65 35 0 | 49 49 1 | 43 56 1 | 68 31 1 |
| E. Error Distribution | | | | | | | | | | |
| normal | 7 54 39 | 17 72 11 | 2 44 54 | 32 65 4 | 13 82 6 | 6 76 19 | 33 65 2 | 30 65 6 | 24 69 7 | 48 48 4 |
| noncentral chi-square | 6 44 50 | 19 63 19 | 2 35 63 | 33 59 7 | 6 63 32 | 4 44 52 | 26 69 6 | 20 74 6 | 15 76 9 | 30 67 4 |
| uniform | 7 61 32 | 26 69 6 | 4 56 41 | 33 63 4 | 15 70 15 | 9 69 22 | 39 59 2 | 37 57 6 | 35 59 6 | 50 46 4 |
| F. Error Variance Level | | | | | | | | | | |
| 25 % | 3 59 38 | 9 77 15 | 0 53 47 | 16 78 6 | 5 77 19 | 3 70 27 | 22 74 4 | 17 77 6 | 16 77 7 | 28 68 4 |
| 75 % | 11 47 42 | 32 59 9 | 5 37 58 | 49 47 4 | 17 67 16 | 10 56 35 | 43 54 3 | 41 54 5 | 33 59 7 | 57 40 4 |
| G. Error Variance Differences | | | | | | | | | | |
| no differences | 7 65 28 | 19 76 6 | 2 57 41 | 37 63 0 | 6 74 20 | 2 67 32 | 32 67 2 | 19 82 0 | 15 83 2 | 37 63 0 |
| between stimuli | 9 50 41 | 20 59 20 | 6 43 52 | 26 61 13 | 13 72 15 | 11 61 28 | 32 65 4 | 39 46 15 | 37 48 15 | 44 44 11 |
| between subjects | 4 44 52 | 22 69 9 | 0 35 65 | 35 63 2 | 15 69 17 | 6 61 33 | 35 61 4 | 30 69 2 | 22 72 6 | 46 54 0 |
| Column Mean | 7 53 40 | 20 68 12 | 3 45 53 | 33 62 5 | 11 72 17 | 6 63 31 | 33 64 3 | 29 65 6 | 25 68 7 | 43 54 4 |

Each cell contains the percentages of underestimation, correct estimation, and overestimation respectively

TABLE 8

ANOVA of Goodness of Fit

| Source * | Sum of Squares | d.f. | F | p | partial $\omega^2$ |
|---|---|---|---|---|---|
| Between data sets | | | | | |
| Within + Residual | 3.24 | 99 | | | |
| A. Number of Subjects | 0.64 | 2 | 9.75 | $< 0.001$ | 0.165 |
| B. Number of Stimuli | 0.34 | 2 | 5.18 | 0.007 | 0.095 |
| C. Data Type | 8.40 | 1 | 256.18 | $< 0.001$ | 0.721 |
| D. Number of Dimensions | $< 0.01$ | 1 | 0.11 | 0.737 | 0.001 |
| E. Error Distribution | 1.07 | 2 | 16.39 | $< 0.001$ | 0.249 |
| F. Error Variance Level | 14.42 | 1 | 440.07 | $< 0.001$ | 0.816 |
| G. Error Variance Differences | 2.04 | 2 | 31.11 | $< 0.001$ | 0.386 |
| A $\times$ F | 0.40 | 2 | 6.12 | 0.003 | 0.110 |
| C $\times$ F | 1.21 | 1 | 36.94 | $< 0.001$ | 0.272 |
| C $\times$ G | 0.49 | 2 | 7.53 | 0.001 | 0.132 |
| E $\times$ G | 4.17 | 4 | 31.82 | $< 0.001$ | 0.562 |
| Within Data Sets | | | | | |
| Within + Residual | 0.25 | 198 | | | |
| H. MLMDS Method | $< 0.01$ | 2 | 0.13 | 0.879 | 0.001 |
| A $\times$ H | 0.03 | 4 | 5.85 | $< 0.001$ | 0.106 |
| G $\times$ H | 0.07 | 4 | 13.02 | $< 0.001$ | 0.208 |
| A $\times$ G $\times$ H | 0.05 | 8 | 4.55 | $< 0.001$ | 0.155 |

* Only interactions significant at $\alpha = .01$ and with partial $\omega^2 > 0.10$ are reported.

TABLE 9

Mean Correlations between Dissimilarity Data and Estimated Distances

| Factor | Level | MULTISCALE | MAXSCAL | PROSCAL | Row Mean |
|---|---|---|---|---|---|
| A. Number of Subjects | 2 | .805 | .815 | .812 | .811 |
| | 8 | .786 | .785 | .784 | .785 |
| | 14 | .778 | .770 | .776 | .775 |
| B. Number of Stimuli | 9 | .800 | .805 | .801 | .802 |
| | 12 | .782 | .785 | .782 | .783 |
| | 15 | .787 | .780 | .787 | .785 |
| C. Data Type | metric | .835 | .839 | .839 | .838 |
| | nonmetric | .744 | .741 | .742 | .742 |
| D. Number of Dimensions | 2 | .799 | .802 | .803 | .801 |
| | 3 | .781 | .778 | .778 | .779 |
| E. Error Distribution | normal | .789 | .786 | .783 | .786 |
| | noncentral chi-square | .801 | .804 | .806 | .804 |
| | uniform | .778 | .780 | .782 | .780 |
| F. Error Variance Level | 25 % | .856 | .854 | .854 | .855 |
| | 75 % | .723 | .726 | .726 | .725 |
| G. Error Variance Differences | no differences | .790 | .785 | .792 | .789 |
| | between stimuli | .817 | .816 | .806 | .813 |
| | between subjects | .762 | .770 | .773 | .768 |
| column mean | | .790 | .790 | .790 | .790 |

TABLE 10

ANOVA of CPU Time

| Source * | Sum of Squares | d.f. | F | p | partial $\omega^2$ |
|---|---|---|---|---|---|
| Between Data Sets | | | | | |
| Within + Residual | 480.53 | 99 | | | |
| A. Number of Subjects | 118.93 | 2 | 12.25 | $< 0.001$ | 0.198 |
| B. Number of Stimuli | 1757.41 | 2 | 181.03 | $< 0.001$ | 0.785 |
| C. Data Type | 601.09 | 1 | 123.84 | $< 0.001$ | 0.556 |
| D. Number of Dimensions | 0.30 | 1 | 0.06 | 0.805 | 0.001 |
| E. Error Distribution | 41.70 | 2 | 4.30 | 0.016 | 0.080 |
| F. Error Variance Level | 0.08 | 1 | 0.02 | 0.895 | $< 0.001$ |
| G. Error Variance Differences | 10.53 | 2 | 1.09 | 0.342 | 0.021 |
| B $\times$ C | 247.95 | 2 | 25.54 | $< 0.001$ | 0.340 |
| B $\times$ G | 93.53 | 4 | 4.82 | 0.001 | 0.163 |
| C $\times$ E | 60.82 | 2 | 6.27 | 0.003 | 0.112 |
| Within Data Sets | | | | | |
| Within + Residual | 992.29 | 198 | | | |
| H. MLMDS Method | 2769.60 | 2 | 276.32 | $< 0.001$ | 0.736 |
| A $\times$ H | 804.40 | 4 | 40.13 | $< 0.001$ | 0.448 |
| B $\times$ H | 1880.67 | 4 | 93.82 | $< 0.001$ | 0.655 |
| C $\times$ H | 1357.57 | 2 | 135.44 | $< 0.001$ | 0.578 |
| A $\times$ B $\times$ H | 188.09 | 8 | 4.69 | $< 0.001$ | 0.159 |
| B $\times$ C $\times$ H | 547.67 | 4 | 27.32 | $< 0.001$ | 0.356 |
| B $\times$ G $\times$ H | 193.03 | 8 | 4.81 | $< 0.001$ | 0.163 |
| C $\times$ E $\times$ H | 122.77 | 2 | 6.12 | $< 0.001$ | 0.110 |

* Only interactions significant at $\alpha = .01$ and with partial $\omega^2 > 0.10$ are reported.

TABLE 11

Computational Effort: Mean CPU Times (in minutes)

| Factor | Level | MULTISCALE | MAXSCAL | PROSCAL | Row Mean |
|---|---|---|---|---|---|
| A. Number of Subjects | 2 | 0.42 | 8.40 | 1.22 | 3.34 |
| | 8 | 0.44 | 4.92 | 3.53 | 2.96 |
| | 14 | 0.58 | 6.02 | 5.86 | 4.15 |
| B. Number of Stimuli | 9 | 0.39 | 1.62 | 2.32 | 1.44 |
| | 12 | 0.40 | 5.03 | 3.55 | 2.99 |
| | 15 | 0.63 | 12.70 | 4.74 | 6.02 |
| C. Data Type | metric | 0.37 | 9.95 | 3.52 | 4.62 |
| | nonmetric | 0.58 | 2.94 | 3.55 | 2.36 |
| D. Number of Dimensions | 2 | 0.54 | 6.85 | 3.52 | 3.64 |
| | 3 | 0.42 | 6.04 | 3.55 | 3.33 |
| E. Error Distribution | normal | 0.33 | 5.42 | 3.46 | 3.07 |
| | noncentral chi-square | 0.55 | 6.83 | 3.59 | 3.66 |
| | uniform | 0.55 | 7.09 | 3.55 | 3.73 |
| F. Error Variance Level | 25 % | 0.48 | 6.74 | 3.66 | 3.63 |
| | 75 % | 0.47 | 6.16 | 3.41 | 3.35 |
| G. Error Variance Differences | no differences | 0.24 | 7.28 | 3.29 | 3.60 |
| | between stimuli | 0.93 | 5.78 | 4.02 | 3.58 |
| | between subjects | 0.26 | 6.27 | 3.30 | 3.28 |
| Column Mean | | 0.48 | 6.45 | 3.54 | 3.49 |

TABLE 12

The Convergence of the Algorithms from Various Starting Values to Local Optima

| Data set | Number of subjects | Number of stimuli | MULTISCALE | MAXSCAL | PROSCAL |
|---|---|---|---|---|---|
| 1 | 2 | 9 | $R = T = 3 \succ 7$ | $2 \succ T \succ R \succ 2$ | $R = 6 \succ T = 1 \succ 3$ |
| 2 | 2 | 9 | $R = T = 10$ | $R = T = 4$ | $R = 6 \succ 3 \succ T \succ 1$ |
| 3 | 2 | 12 | $R = T = 5 \succ 5$ | $R = T \succ 2$ | $T = 4 \succ R = 1 \succ 5$ |
| 4 | 2 | 12 | $R = T = 8 \succ 2$ | $R \succ T \succ 3$ | $R = 8 \succ 2 \succ T$ |
| 5 | 2 | 15 | $R = T = 9 \succ 1$ | $T \succ R \succ 2$ | $R = T = 4 \succ 6$ |
| 6 | 2 | 15 | $R = 1 \succ 2 \succ T = 6 \succ 1$ | $1 \succ R \succ 5 \succ T$ | $R = 6 \succ T = 2 \succ 2$ |
| 7 | 8 | 9 | $5 \succ R = T = 5$ | $R = T = 4 \succ 1$ | $R = T = 2 \succ 8$ |
| 8 | 8 | 9 | $R = T = 7 \succ 3$ | $R = T = 1 \succ 3$ | $R = T = 2 \succ 8$ |
| 9 | 8 | 12 | $T = 4 \succ R = 6$ | $R \succ 1 \succ T \succ 4$ | $R = 2 \succ 3 \succ T \succ 5$ |
| 10 | 8 | 12 | $R = T = 6 \succ 4$ | $R = 1 \succ 4 \succ T \succ 2$ | $R = 2 \succ T = 2 \succ 6$ |
| 11 | 8 | 15 | $R = T = 10$ | $R = 2 \succ 2 \succ T \succ 2$ | $R = T \succ 10$ |
| 12 | 8 | 15 | $T = 5 \succ R = 1 \succ 4$ | $2 \succ R \succ T \succ 1$ | $R = T = 7 \succ 3$ |
| 13 | 14 | 9 | $R = T = 9 \succ 1$ | $R = T = 3 \succ 1$ | $R = T = 4 \succ 6$ |
| 14 | 14 | 9 | $R = T = 10$ | $R = T = 3 \succ 1$ | $R = T = 5 \succ 5$ |
| 15 | 14 | 12 | $R = T = 9 \succ 1$ | $R = T = 1 \succ 5$ | $R = T = 4 \succ 6$ |
| 16 | 14 | 12 | $R = T = 7 \succ 3$ | $T = 1 \succ R \succ 4$ | $R = 4 \succ 4 \succ T \succ 2$ |
| 17 | 14 | 15 | $R = T = 4 \succ 6$ | $R = T \succ 5$ | $R = T = 1 \succ 9$ |
| 18 | 14 | 15 | $R = T = 3 \succ 7$ | $R = T \succ 4$ | $R = T = 3 \succ 7$ |

$R$ and $T$ indicate solutions with rational and true coordinates as starting values, respectively

1, 2, .., 10 indicate the number of solutions with randomly generated coordinates as starting values

$=$: solutions equal within the convergence range

$\succ$: the former solutions outperform the latter solutions