A three-way clusterwise multidimensional unfolding procedure for the spatial representation of context dependent preferences

Wayne S. DeSarbo, A. Selin Atalay, Simon J. Blanchard

Abstract

Various deterministic and latent structure approaches for combining forms of multidimensional scaling and cluster analysis have been previously discussed. A new clusterwise three-way unfolding methodology for the analysis of two-way or three-way metric dominance/preference data is proposed. The purpose of this proposed methodology is to simultaneously estimate a joint space of stimuli and cluster ideal point representations, as well as the clusters themselves, such that the geometry underlying the clusterwise model renders some indication of the underlying structure in the data. In the three-way case, it is shown how multiple ideal points can represent preference change over contexts or situations. Partitions, overlapping clusters, stationary and context dependent preference representations are allowed. After a literature review of related methodological research, the technical details of the proposed three-way clusterwise spatial unfolding model are presented in terms of modeling context/situational dependent preferences (i.e., preferences for various stimuli collected over the same set of subjects over time, situation, etc.). The psychological basis for the models is provided in terms of the extensive behavioral decision theory and consumer psychology literature on contextual preferences and situational effects. An application to a data set exploring preferences for breakfast/snack food data over a number of different usage situations is then presented, followed by a discussion on future potential research directions.

1. Introduction

The classification and psychometric literature abounds with clusterwise procedures that simultaneously perform some form of multivariate analysis and classification in an attempt to represent sample heterogeneity. Clusterwise linear regression, originally proposed by Späth (1979, 1981, 1982), is a deterministic method originally devised for simultaneous regression and classification. In a traditional linear regression context, Späth's procedure clusters subjects non-hierarchically in such a manner that the fit of the regression within the simultaneously derived clusters is optimized. Efforts to extend this clusterwise framework to simultaneous multidimensional scaling (MDS) and classification has occurred on two major fronts (broadly defined, multidimensional scaling comprises a family of geometric models for the multidimensional representation of data and a corresponding set of methods for fitting such models to actual data (from Carroll and Arabie (1980, p. 608))). One front deals with parametric finite mixture or latent class MDS models (LCMDS) (cf., DeSarbo et al. (1994)). In particular, LCMDS models for the analysis of preference/dominance data have been proposed by a number of different authors over...
the past fifteen years employing either a scalar products/vector (Tucker, 1960; Slater, 1960) or unfolding (Coombs, 1964)
representations for two-way preference/dominance data. In such LCMDS models, vectors or ideal points of clusters are
estimated in place of individual subjects. Thus, the number of parameters is significantly reduced relative to individual level
models. LCMDS models are traditionally estimated using the method of maximum likelihood (E-M algorithms are typically
employed) for two-way data. For example, DeSarbo et al. (1999) developed a LCMDS vector model for normally distributed
data. DeSarbo et al. (1991) extended this model to a weighted ideal point model. De Soete and Winsberg (1993) and De
Soete and Heiser (1993), respectively, extended these two models by accommodating linear restrictions on the stimulus
coordinates. Böckenholt and Böckenholt (1991) developed simple and weighted ideal-point models for binary data. DeSarbo
et al. (1995) developed a vector LCMDS model for Dirichlet-distributed data, and Chintagunta (1994) developed a LCMDS
vector model for multinomial data. Wedel and DeSarbo (1996) extended the entire family of exponential distributions to
such LCMDS models for two-way preference/dominance data.

The second front of such combined spatial and classification methods dominance/preference type data involve
deterministic clusterwise approaches. Several authors have proposed such deterministic approaches for the simultaneous
classification and reduction of two-way dominance data in the context of MDS (Heiser and Groenen, 1997; Kiers et al., 2005),
optimal scaling (Van Buuren and Heiser, 1989), and principal components analysis (PCA) (De Soete and Carroll, 1994; Vichi
and Kiers, 2001). In the case of three-way dominance data, some literature exists for various methodologies which integrate
MDS and clustering (Meulman and Verboon, 1993; Carroll and Chaturuvedi, 1995), and PCA and clustering (Roccì and Vichi,
2005). No such clusterwise procedure exists for multidimensional unfolding analysis and clustering.

Our objective here is to devise a new deterministic clusterwise procedure for the analysis of two or three-way
preference/dominance data vis-à-vis a spatial unfolding, ideal point representation. Like the deterministic clusterwise
approaches mentioned above, our method does not require parametric assumptions as do LCMDS procedures. In addition,
the proposed clusterwise multidimensional unfolding (MDU) model permits either non-overlapping or overlapping cluster
solutions (unlike the fuzzy empirical classifications derived from LCMDS methods), and options for stationary or context
varying ideal point representations. Finally, our procedure is formulated to accommodate two or three-way dominance data.

In the three-way case, we present a “floating preference model” spatial representation which permits cluster ideal points
to vary or float over the third way of the array (consumptive situations in our application), i.e., accommodating multiple
ideal points per derived cluster (c.f., DeSarbo et al. (2007)). This provides a concise spatial representation for the analysis of
contextual preferences as shown in the application to follow.

The next section provides the model description of the proposed clusterwise spatial methodology for unfolding models,

as well as an efficient alternating least-squares estimation framework. We position this new methodology to be a clusterwise
extension of the DeSarbo and Carroll (1985) metric three-way unfolding model. We next provide substantive background
from the psychology and behavioral decision theory literatures concerning situation dependent, contextual, and dynamic
preferences. An application to situational effects on preferences for breakfast/snack foods is described including the study,
data description, methodological findings, and resulting implications. Finally, potential extensions for future research are
discussed.

2. The proposed clusterwise MDU methodology

The concept of unfolding was initially developed by Coombs (1950), and later generalized to the multidimensional case
by Bennett and Hays (1960) (see Borg and Groenen (2005) for an excellent recent description of the nature and historical
development of the literature on MDU). The goal of traditional MDU is to uncover a joint dimensional space of stimulus
coordinates and subjects' ideal points such that the distance between these two entities recovers some aspect of the input
(dis)preference judgments. In the traditional notion of two-way MDU, one typically estimates an ideal point per subject, as
well as a point location for each stimulus (as well as other model specific parameters depending upon the specific model
and properties of the data). As such, large numbers of parameters are typically estimated whose order typically varies
with the rows and columns of the input data matrix. Our objective is to provide a clusterwise unfolding methodology that
simultaneously estimates stimulus locations, ideal points by derived cluster, as well as the cluster memberships themselves,
with the same data as one would utilize in traditional MDU.

2.1. The clusterwise MDU model structure

Given that the application we will be discussing concerns the analysis of consumer preference judgments for various
snack/breakfast food brands over various consumptive contexts/situations/goals, we will use this scenario to technically
define the parameters of the proposed methodology below. Note, however, that this proposed methodology can be utilized
for the analysis of any type of two or three-way metric dominance data. We describe the fully parameterized model and
later discuss various model options.

Let:

\[ i = 1, \ldots, N \] consumers;

\[ j = 1, \ldots, J \] brands;

\[ t = 1, \ldots, T \] consumptive situations (e.g., time, usage occasion, goal, etc.);
s = 1, . . . , 5 market segments/clusters (unknown);
\( r = 1, . . . , R \) dimensions (unknown);
\( \Delta_{jt} \) is the dispreference for brand \( j \) given by consumer \( i \) in situation \( t \).

The full model can be expressed as:

\[
\Delta_{jt} = \sum_{s=1}^{S} p_{is} \sum_{r=1}^{R} w_{rt}(x_{jr} - y_{st})^2 + b_t + \epsilon_{ijt},
\]

(1)

where:

\( x_{jr} \) = the \( r \)-th coordinate for brand \( j \);
\( y_{st} \) = the \( r \)-th coordinate of the ideal point for market segment \( s \) in situation \( t \);
\( w_{rt} \) = stretching/shrinking parameters for dimension \( r \) in situation \( t \);
\( b_t \) = an additive constant for situation \( t \);
\( p_{is} \) = 1 if consumer \( i \) is classified into market segment \( s \), 0 else;

where:

\( p_{is} \in \{0, 1\} \),
\[ \sum_{s=1}^{S} p_{is} = 1 \forall i \) (for non-overlapping segments),

or

\[ 0 < \sum_{s=1}^{S} p_{is} \leq S \) (for overlapping segments),

\( \epsilon_{ijt} \) = error (deterministic).

Note, the weighted squared distance model in expression (1) is analogous to the DeSarbo and Carroll (1985) three-way metric unfolding model generalized to a clusterwise formulation. As noted by DeSarbo and Carroll (1985) and Carroll (1980), the unfolding model accommodates the vector model as a special case in individual-level preference spatial models.

2.2. Estimation

Given \( \Delta = ((\Delta_{ji})) \) and values of \( S \) and \( R \), we wish to estimate \( W = ((w_{rt})) \), \( b = (b_t) \), \( P = ((p_{is})) \), \( X = ((x_{jr})) \), and \( Y = ((y_{st})) \) so as to minimize:

\[
\Phi = \sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{t=1}^{T} q_{ijt} \left[ \Delta_{ji} - \hat{\Delta}_{ji} \right]^2,
\]

(2)

where:

\[
\hat{\Delta}_{ji} = \sum_{s=1}^{S} p_{is} \sum_{r=1}^{R} w_{rt}(x_{jr} - y_{st})^2 + b_t,
\]

(3)

and

\( q_{ij} \) = a pre-specified weighting function of the data.

As introduced in DeSarbo and Rao (1984) and DeSarbo and Carroll (1985), \( q_{ij} \) is a user specified weighting function that is adapted to prevent degenerate solutions as often encountered in MDU practice with real data (for alternative approaches to resolve degenerate solutions in traditional MDU see also Heiser (1981), Kim et al. (1999), Busing et al. (2005) and Van Deun et al. (2005)). Its function is to weight different data values differentially depending upon the goals of the analysis, type of data collected, and beliefs concerning the reliability of the data. Setting \( q_{ij} = 1 \), as an option across all \( i, j, t \), provides equal weighting to all data values. If one were most concerned with recovering the most preferred data values best (e.g., in marketing research studies concerned with top consumer choices and brand market share; c.f., DeSarbo and Rao (1986), then one option could be to set \( q_{ij} = (1/\Delta_{ij})^c \), assuming \( \Delta_{ij} > 0 \) and \( c > 1 \). Or, many studies suggest that subjects display more reliability in their preference judgments for higher preferred and higher dispreferred stimuli (i.e., subjects can reliably identify what stimuli they most prefer and most dislike in a reliable manner; c.f., Ben-Akiva et al. (1992)), in which case a \( U \)-shaped \( q_{ij} \) can be specified. Preferences have not been found to be equally stable over repeated measurements (Ben-Akiva et al., 1992; Kivetz et al., 2004). Consider for example, Person A, who is lactose intolerant. This person has a rigid constraint that makes her not consider meal options that include any dairy products. All dairy products in the consideration set are then going to be very stable in being the least preferred options because of this individual restriction that is never going to change. Now consider Person B, who is a strong variety seeker. Variety seekers may have multiple needs and are intrinsically motivated to explore other options (c.f., Van Trijp et al. (1996)). For this person, one would observe less stability for the most preferred options, as the rank order preference between the most preferred options changes frequently. Finally, consider Person C who is strongly loyal to his favorite coffee brand. This person may have a biased high attitude towards the brand and this attitude would culminate in multiple repurchases (Dick and Basu, 1994; Jacoby and Chestnut, 1978). His preference
for his favorite coffee brand would therefore be very stable. Note, the \( q_{jt} \) weights are also utilized to accommodate missing data by setting them equal to zero for missing entries in \( \Delta_{jt} \).

Visually, in \( R = 2 \) dimensions with \( J = 10 \) stimuli, \( T = 3 \) situations, and \( S = 3 \) clusters, Fig. 1 illustrates a hypothetical spatial solution derived by our proposed clusterwise MDU methodology, where \( y_{jt} \) labels the associated ideal point for the \( s \)-th cluster in the \( t \)-th situation, and the letters A–J label the hypothetical stimuli. As shown in this illustration, Cluster 1’s three ideal points all fall into the first quadrant and do not vary all that much over the three situations with respect to their location. Preference appears to be highest for stimuli B, G, and F depending somewhat upon the situation. Notice the differences when comparing Cluster 1 here to the remaining two clusters. Cluster 2 has its three ideal points located in quadrants 2, 3, and 4, as does Cluster 3. For the members of Cluster 2, stimuli A and E are most preferred in situation 1, stimuli D and I in situation 2, and stimuli C, H, and J in situation 3. For Cluster 3, stimuli C, H, and J are preferred in situation 1, stimuli A and E in situations 2, and stimuli E and D in situation 3.

We devise an alternating least-squares iterative estimation procedure involving four distinct stages. Given the non-linear nature of the underlying model and objective function, convergence to globally optimum solutions is not guaranteed as with all other MDU approaches. These estimation steps are summarized below:

**Stage 1: Estimate \( \mathbf{W} \) and \( \mathbf{b} \).**

We first define:

\[
\hat{\Delta}_{jt}^* = \sum_{s=1}^{S} \nu_s (x_{jt} - y_{srt})^2; \tag{4}
\]

\( \mathbf{L} \): the \( NJ \times R \) matrix of concatenated elements of \( \hat{\Delta}_{jt}^* \), for \( t = 1, \ldots, T; \)

\( \mathbf{K}_t = (\mathbf{1}_L \times (\mathbf{Q}_s))^{1/2} \): where \( \mathbf{Q}_s \) represents the \( NJ \times (R - 1) \) matrix where each column is the concatenated vector whose elements are \( q_{jt} \); \n
\( \mathbf{1}_L = (1, 1, \ldots, 1) \): an \( NJ \times 1 \) vector of 1’s;

\( \mathbf{m}_t = \mathbf{1}_L \times \mathbf{m} \): the \( NJ \times 1 \) vector of concatenated elements of \( [\Delta_{jt} \times (q_{jt})^{1/2}] \), for \( t = 1, \ldots, T. \)

Akin to DeSarbo and Carroll (1985) for the three-way metric unfolding case, the following formulation:

\[
\left( \begin{array}{c}
\hat{b}_t \\
\hat{w}_{jt}
\end{array} \right) = (\mathbf{K}_t^T \mathbf{K}_t)^{-1} \mathbf{K}_t^T \mathbf{m}_t, \tag{5}
\]

can be gainfully employed to obtain current estimates of the additive constant and weights vector by situation, given fixed current values of \( \mathbf{P}, \mathbf{X} \) and \( \mathbf{Y} \). A constrained weighted least-squares procedure (cf., Lawson and Hanson (1995)) is utilized as an option for constraining the entries in \( \mathbf{W} \) to be positive given the difficulties in interpreting negative weights in three-way unfolding models.

**Stage 2: Estimate \( \mathbf{X} \) and \( \mathbf{Y} \).**

Now, holding fixed current values of \( \mathbf{b}, \mathbf{W} \), and \( \mathbf{P} \), we estimate \( \mathbf{X} \) and \( \mathbf{Y} \) via a conjugate gradient procedure (see DeSarbo and Carroll (1985, Appendix B) and Miller (1999)). Here, the respective partial derivatives of \( \phi \) with respect to \( x_{jt} \) and \( y_{srt} \) are:

\[
\frac{\partial \phi}{\partial x_{jt}} = -4 \sum_{i=1}^{J} \sum_{t=1}^{T} q_{ijt} (\Delta_{ijt} - \hat{\Delta}_{ijt}) \left[ \sum_s \nu_s w_{rt} (x_{jt} - y_{srt}) \right] \tag{6}
\]

and

\[
\frac{\partial \phi}{\partial y_{srt}} = +4 \sum_{i=1}^{J} \sum_{t=1}^{T} q_{ijt} (\Delta_{ijt} - \hat{\Delta}_{ijt}) \left[ \nu_s w_{rt} (x_{jt} - y_{srt}) \right]. \tag{7}
\]

This conjugate gradient search (with automatic restarts) progresses for a preset number of iterations or up to the total number of free parameters being estimated in \( X \) and \( Y \) for which orthogonal directions can be generated. No iteration is accepted that does not improve the objective function.

Stage 3: Estimate \( P \).

For estimating the membership matrix \( P \), note that the error sums of squares objective function minimized in (2) is calculated as a sum of squared discrepancies over contexts (\( t \)), brands (\( j \)), and consumers (\( i \)). One can thus rewrite:

\[
\Phi = \sum_{i=1}^{N} \Phi_i,
\]

where:

\[
\Phi_i = \sum_{j=1}^{J} \sum_{s=1}^{S} q_{ijt} \left( \Delta_{ijt} - \sum_{r=1}^{R} p_{ir} \sum_{t=1}^{T} w_{rt} (x_{ijt} - \hat{y}_{ijt})^2 - b_t \right)^2.
\]

I.e., with respect to the subject index \((i)\), (2) is separable with respect to that mode of the array. That is, \( p_i \) and \( \Delta_i \) are the only two entities indexed by \( i \) that affect \( \Phi \) in (2), and thus one could perform the subsequent optimization by subject in minimizing (9) for each consumer \( i \). Using this separable property, the proposed procedure performs a complete enumeration over \( 2^S - 1 \) (ignoring the 0 solution so that each consumer is assigned to a cluster) options for the overlapping cluster case per consumer (\( S \) options for the non-overlapping case) to obtain a globally optimum solution for \( P \) conditioned on current values of \( X, Y, W, \) and \( b \).

Stage 4: Convergence tests

Thus, the proposed alternating least-squares estimation procedure iteratively cycles through (5)–(7), and the \( N \) complete enumerations to minimize the error sums of squares in (2). Stage 2 involving the conjugate gradient procedures provide at least locally optimum estimates of the associated set of parameters \( (X, Y) \) estimated conditional on the remaining sets of parameters held at their current values. Stages 1 and 3 provide conditionally globally optimum estimates of \((W, b)\) and \( P \) holding all other parameters fixed at their current values. This entire process obviously does not guarantee a final globally optimum result and thus multiple runs (at least five per R,S solution) are recommended for each solution in order to detect holding all other parameters fixed at their current values. Stages 1 and 3 provide

For estimating the membership matrix \( P \), note that the error sums of squares objective function minimized in (2) is calculated as a sum of squared discrepancies over contexts (\( t \)), brands (\( j \)), and consumers (\( i \)). One can thus rewrite:

\[
\Phi = \sum_{i=1}^{N} \Phi_i,
\]

where:

\[
\Phi_i = \sum_{j=1}^{J} \sum_{s=1}^{S} q_{ijt} \left( \Delta_{ijt} - \sum_{r=1}^{R} p_{ir} \sum_{t=1}^{T} w_{rt} (x_{ijt} - \hat{y}_{ijt})^2 - b_t \right)^2.
\]

I.e., with respect to the subject index \((i)\), (2) is separable with respect to that mode of the array. That is, \( p_i \) and \( \Delta_i \) are the only two entities indexed by \( i \) that affect \( \Phi \) in (2), and thus one could perform the subsequent optimization by subject in minimizing (9) for each consumer \( i \). Using this separable property, the proposed procedure performs a complete enumeration over \( 2^S - 1 \) (ignoring the 0 solution so that each consumer is assigned to a cluster) options for the overlapping cluster case per consumer (\( S \) options for the non-overlapping case) to obtain a globally optimum solution for \( P \) conditioned on current values of \( X, Y, W, \) and \( b \).

Stage 4: Convergence tests

Thus, the proposed alternating least-squares estimation procedure iteratively cycles through (5)–(7), and the \( N \) complete enumerations to minimize the error sums of squares in (2). Stage 2 involving the conjugate gradient procedures provide at least locally optimum estimates of the associated set of parameters \( (X, Y) \) estimated conditional on the remaining sets of parameters held at their current values. Stages 1 and 3 provide conditionally globally optimum estimates of \((W, b)\) and \( P \) holding all other parameters fixed at their current values. This entire process obviously does not guarantee a final globally optimum result and thus multiple runs (at least five per R,S solution) are recommended for each solution in order to detect holding all other parameters fixed at their current values. Stages 1 and 3 provide

Given that the overall magnitude of the sums-of-squared error objective function in (2) is contingent on such things as the size of the data array \((I, J, T)\), the scale of the preference ratings collected, etc., two normalized goodness of fit measures selected on the basis of scree plots (VAF vs. \( S \) and \( R \)), interpretation of the derived solutions, parsimony, etc. Given that the overall magnitude of the sums-of-squared error objective function in (2) is contingent on such things as the size of the data array \((I, J, T)\), the scale of the preference ratings collected, etc., two normalized goodness of fit measures selected on the basis of scree plots (VAF vs. \( S \) and \( R \)), interpretation of the derived solutions, parsimony, etc. Given that the overall magnitude of the sums-of-squared error objective function in (2) is contingent on such things as the size of the data array \((I, J, T)\), the scale of the preference ratings collected, etc., two normalized goodness of fit measures selected on the basis of scree plots (VAF vs. \( S \) and \( R \)), interpretation of the derived solutions, parsimony, etc. Given that the overall magnitude of the sums-of-squared error objective function in (2) is contingent on such things as the size of the data array \((I, J, T)\), the scale of the preference ratings collected, etc., two normalized goodness of fit measures selected on the basis of scree plots (VAF vs. \( S \) and \( R \)), interpretation of the derived solutions, parsimony, etc. Given that the overall magnitude of the sums-of-squared error objective function in (2) is contingent on such things as the size of the data array \((I, J, T)\), the scale of the preference ratings collected, etc., two normalized goodness of fit measures selected on the basis of scree plots (VAF vs. \( S \) and \( R \)), interpretation of the derived solutions, parsimony, etc.
Table 1  
Factors and levels for the Monte Carlo simulation experiment

<table>
<thead>
<tr>
<th>#</th>
<th>Factor</th>
<th>Levels (Codes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Number of Subjects I</td>
<td>50 (0), 100 (1)</td>
</tr>
<tr>
<td>2</td>
<td>Number of Brands J</td>
<td>10 (0), 20 (1)</td>
</tr>
<tr>
<td>3</td>
<td>Number of Situations T</td>
<td>3 (0), 6 (1)</td>
</tr>
<tr>
<td>4</td>
<td>Number of Clusters S</td>
<td>3 (0), 5 (1)</td>
</tr>
<tr>
<td>5</td>
<td>Number of Dimensions R</td>
<td>2 (0), 3 (1)</td>
</tr>
<tr>
<td>6</td>
<td>Errora</td>
<td>10% (0), 20% (1)</td>
</tr>
<tr>
<td>7</td>
<td>Ideal Points</td>
<td>Stationary: Y (0), Overlapping (1)</td>
</tr>
<tr>
<td>8</td>
<td>Type of clustering</td>
<td>Non-overlapping (0), Overlapping (1)</td>
</tr>
<tr>
<td>9</td>
<td>Start</td>
<td>Random (0), Rational (1)</td>
</tr>
</tbody>
</table>

Dependent measures:
1 V.A.F.
2 Number of major iterations (full cycles of expressions 5, 6, 7 and estimating P).
3 CPU time, in seconds.
4 Recovery of the joint space (X, Y).
5 Recovery of the weights (W).
6 Recovery of the additive constant (b).
7 Recovery of the additive constant (b).

2.3. Clusterwise MDU model options

The proposed clusterwise MDU methodology allows the user to select from a number of model options that provides additional flexibility in the analysis of such dominance data. One, either two-way or three-way analyses are accommodated by this procedure. The weights W matrix reduces to a vector of multiplicative constants for one context/situation in the two-way case, and is therefore not identifiable since its values can be embedded directly into X and Y. Two, we can estimate either stationary ideal points yr or context dependent ideal points ysr. Three, we can estimate either non-overlapping clusters or overlapping clusters. Four, we can impose constraints on W (whose entries are all assumed to be positive) including positivity constraints or equality constraints wr = wt for all r (this later constraint allows for a simple unfolding clusterwise model). Five, we can perform either internal or external analyses regarding the stimulus coordinates X, where X can be given and fixed throughout the iterations as well as in external analyses or estimated freely as in internal analyses. Finally, one can select from a variety of starting options including given starting values, random starting values, or rational starting values where the data is utilized in obtaining starting estimates for X (use singular value decomposition on the averaged over situations/contexts preference data). P (use a quick K-Means analysis on either the averaged preference data or individual starting ideal points), and Y (use quadratic regression procedure as in PREFMAP2 external analysis to estimate ideal points by subject given X as in Carroll (1980)).

2.4. Monte Carlo simulation analysis

Given the plethora of clusterwise MDU model variants that can be accommodated by the proposed methodology, we sought to explore the performance of the proposed clusterwise MDU methodology across a number of data, model, and error possibilities in an experimental setting with synthetically constructed data whose structure is known. Some nine experimental independent factors, each at two levels, were generated in this endeavor and are listed in Table 1. Common with DeSarbo and Carroll (1985) and a host of other methodological articles published previously in the classification and psychometric literature, a fractional factorial design (allowing the estimation of main effects only) was generated for a 29 design involving 32 trials/experiments. This fractionated design appears in Table 2. Each row in the orthogonal array provides the particular design utilized in the data creation, model estimation, and error for a particular experimental trial. Various dependent measures were measured to assess some four major performance factors: recovery of the input data, recovery of the known model parameters utilized to construct the data, computational intensity, and recovery of the classifications. These various dependent measures are also listed in Table 1. Recovery of the clusters was analyzed by calculating the differences between each cell of the original and the obtained cluster matrices (a matching coefficient representing the number of matches divided by the total number of entities in P), after any necessary column permutations (one could also have used an adjusted Rand Index; see Hubert and Arabie (1985)). For the weights, recovery is measured by calculating the average SSSAF between the original and obtained matrices after any necessary column permutations. To measure the recovery of the joint space (X, Y), we calculate the average VAF between the original and obtained joint spaces after centering and any necessary column permutations. The recovery of the vector of additive constants b is obtained by calculating the root mean square error (RMSE) between actual and recovered values. Finally, for computational intensity, we report the number of major iterations as well as CPU time (on an AMD Athlon64 2.64 GHz computer).

To generate the overlapping clusters’ membership, individuals were assigned to cluster s = 1…5 following a Bernoulli distribution (p = 0.5). For non-overlapping clusters, 5 uniform random variables were drawn for each individual and the
Table 2

$2^9$ fraction factorial design for the Monte Carlo experiment

<table>
<thead>
<tr>
<th>Factor</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
<th>Trial 4</th>
<th>Trial 5</th>
<th>Trial 6</th>
<th>Trial 7</th>
<th>Trial 8</th>
<th>Trial 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

individual was assigned to the cluster where the draw was the highest. The brand coordinates were drawn from a normal distribution (0, 1) and then column standardized. The ideal points (per cluster and per context) were drawn from a normal distribution (0, 1). The weights/multiplicative constants were drawn from a uniform distribution (4, 7) and the additive constants from a normal distribution (0, 1). Using all these draws, dispreference scores were then calculated. Normal error was then added by varying the variance parameter of a normal distribution as to obtain the desired ratio of added variance over total variance.

Table 3 presents the various regression analysis results (equivalent to ANOVA's per dependent variable) for each of the dependent measures based on the nine factors previously discussed. Each column represents the regression for a particular dependent measure, and the content of the cells indicate the main effects for the factors used as independent variables as well as their significance levels. The means and standard deviations of each dependent measure computed across all 32 experimental trials are also presented at the lower portion of the table. With respect to the joint space recovery, we see consistent fitting across all of the nine independent factors as nothing is significant, including the overall regression function. However, we do see a smaller overall mean fit but with a large variance over the 32 trials. The lack of significant findings is also witnessed with respect to the additive constant. However, we notice that recovery of the $W$ matrix is significantly impacted by data with larger numbers of contexts/slices, additional dimensions, and, to a lesser extent, estimating context dependent ideal points. It appears that the more parameters one estimates, the worse the recovery of $W$. With respect to the cluster membership recovery, nothing is significant indicating robust recovery of the $P$ matrix. With respect to the two measures of computational time, a very significant increase in computational effort occurs with larger data sets involving higher numbers of contexts/slices and subjects. To a lesser degree, estimating overlapping clusters and higher numbers of clusters also significantly increases computational effort. This is to be expected since the enumeration method utilized has to search over additional options for the overlapping case, as well as for the case with larger number of clusters. Finally, VAF data recovery is significantly worse in the high error condition and significantly better when employing rational (as opposed to random) starts. Here too, this result is as expected.

In summary, the overall results of this modest Monte Carlo simulation are encouraging. Somewhat robust recovery over the independent variable levels tested is observed with respect to the additive constants, the cluster membership, and the joint space (although the latter is associated with somewhat poorer mean recovery). We do notice a slight problem with respect to $W$ recovery as this particular aspect does seem somewhat sensitive to the number of parameters being estimated. Data recovery (VAF) is naturally affected by the amount of error in the data as well as using a rational start.
Finally, computational intensity is affected by the size of the data set and the number of parameters being estimated as anticipated.

3. Contextual, situational, and dynamic preferences

The existent psychology literature strongly supports the contention that preferences are context dependent. The same individual may have different preferences in different contexts or situations. The study of the dynamic nature of preference and the contextual influences on behavior and decision making dates back over five decades. Lewin (1951) argued that behavior depends both on the situation and the overall attitudes of the decision maker; \( B = f(P,E) \). This interaction between the person and the environment and its impact on behavior is well documented, especially in the behavioral decision making and consumer psychology literature (e.g., Belk (1974, 1975), Puto (1987) and Payne et al. (1992)). Such studies suggest that preference is dynamic since the context influences the judgment of the individual (Payne, 1982; Tversky and Simonson, 1993). DeSarbo et al. (2005) recently surveyed aspects of this dynamic preference literature and proposed Bayesian evolutionary utility function estimation procedures for repeated-measures, stated/revealed preference studies.

“Context” is the composition of the observable, aggregate characteristics of the time and place of observation which are not related to personal and stimulus attributes (Belk, 1974; Tversky and Simonson, 1993). The environment in which the behavior happens can be divided into two components: the situation and the object. Belk (1975, p. 159) presents some five groups or types of situational characteristics:

1. Physical surroundings—these include geographical and institutional location, décor, sounds, aroma, lighting, weather, etc.;
2. Social surroundings—involve other people present, their characteristics, their apparent roles, interpersonal interactions, etc.;
3. Temporal perspectives—time of the day, season of the year, time constraints imposed by prior or standing commitments, time relative to some previous phenomena, etc.;
4. Task definition—in a consumer psychology context, this may include an intent or requirement to select, shop for, or obtain information about a general or specific purpose. It may reflect different buyer and user roles or usage occasions anticipated by the individual;
5. Antecedent states—momentary moods (anxiety, pleasantness, hostility, excitation, etc.) or momentary conditions (e.g., cash on hand, fatigue, illness, etc.) as opposed to chronic individual traits.

In a study of the interaction between the consumer and the situation, Belk (1974) asked participants to evaluate a list of 10 snack food alternatives to serve or consume in 10 different situations. Participants were asked to indicate how likely they were to choose each product in each of the situations. The same participants also responded to the same questions in a meat product domain. In the meat product survey, the participants were asked to evaluate a list of 11 meat product alternatives to serve or consume in nine different situations. These participants also indicated their likelihood of choosing each meat product for each situation. Based on subsequent multivariate statistical analysis of the data, his study provided substantial evidence that food preferences are heavily influenced by situational factors. According to Belk (1974), there were three clusters or market segments of consumers that formed with respect to how their preferences shift as the food consumption situations changes.

### Table 3

<table>
<thead>
<tr>
<th></th>
<th>X, Y</th>
<th>b</th>
<th>W</th>
<th>P</th>
<th>IT</th>
<th>CPU time</th>
<th>VAF</th>
</tr>
</thead>
<tbody>
<tr>
<td>I = 100</td>
<td>0.02</td>
<td>−1.81</td>
<td>0.08</td>
<td>0.04</td>
<td>6.25</td>
<td>36.32**</td>
<td>0.00</td>
</tr>
<tr>
<td>J = 20</td>
<td>0.09</td>
<td>2.15</td>
<td>0.03</td>
<td>0.00</td>
<td>−4.50</td>
<td>19.86</td>
<td>0.00</td>
</tr>
<tr>
<td>T = 6</td>
<td>−0.02</td>
<td>−0.05</td>
<td>−0.14**</td>
<td>−0.03</td>
<td>11.50**</td>
<td>41.07**</td>
<td>−0.01</td>
</tr>
<tr>
<td>S = 5</td>
<td>−0.02</td>
<td>−3.27</td>
<td>0.00</td>
<td>−0.03</td>
<td>4.88</td>
<td>26.76</td>
<td>0.01</td>
</tr>
<tr>
<td>R = 3</td>
<td>−0.10</td>
<td>−0.30</td>
<td>−0.15**</td>
<td>−0.03</td>
<td>2.88</td>
<td>21.83</td>
<td>0.01</td>
</tr>
<tr>
<td>20% error</td>
<td>0.04</td>
<td>3.28</td>
<td>−0.06</td>
<td>−0.02</td>
<td>2.50</td>
<td>9.06</td>
<td>−0.10**</td>
</tr>
<tr>
<td>Context dependent ideal points</td>
<td>0.01</td>
<td>−2.06</td>
<td>−0.10*</td>
<td>0.04</td>
<td>0.87</td>
<td>11.57</td>
<td>−0.01</td>
</tr>
<tr>
<td>Overlapping clusters</td>
<td>0.05</td>
<td>2.40</td>
<td>−0.02</td>
<td>−0.03</td>
<td>10.50*</td>
<td>24.03</td>
<td>0.00</td>
</tr>
<tr>
<td>Rational start</td>
<td>0.06</td>
<td>−3.27</td>
<td>0.04</td>
<td>0.03</td>
<td>−7.00</td>
<td>2.61</td>
<td>0.02</td>
</tr>
<tr>
<td>S.E.</td>
<td>0.17</td>
<td>7.43</td>
<td>0.13</td>
<td>0.06</td>
<td>10.77</td>
<td>0.52</td>
<td>0.02</td>
</tr>
<tr>
<td>( \beta^2 )</td>
<td>0.24</td>
<td>0.24</td>
<td>0.59</td>
<td>0.43</td>
<td>0.55</td>
<td>0.81</td>
<td>0.92</td>
</tr>
<tr>
<td>Aij ( \hat{\beta}^2 )</td>
<td>−0.06</td>
<td>−0.07</td>
<td>0.42</td>
<td>0.19</td>
<td>0.36</td>
<td>0.66</td>
<td>0.88</td>
</tr>
<tr>
<td>F</td>
<td>0.79</td>
<td>0.78</td>
<td>3.50**</td>
<td>1.82</td>
<td>2.99</td>
<td>4.76**</td>
<td>27.26**</td>
</tr>
<tr>
<td>Mean</td>
<td>0.55</td>
<td>5.81</td>
<td>0.83</td>
<td>0.97</td>
<td>25.31</td>
<td>39.29</td>
<td>0.83</td>
</tr>
<tr>
<td>S.D.</td>
<td>0.17</td>
<td>7.30</td>
<td>0.17</td>
<td>0.07</td>
<td>13.52</td>
<td>45.90</td>
<td>0.06</td>
</tr>
</tbody>
</table>

* \( p < 0.05 \)
** \( p < 0.01 \)
Table 4
Breakfast item labels and attributes

<table>
<thead>
<tr>
<th>Attribute vectors key</th>
<th>Attribute vectors key</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>Ease of preparation</td>
</tr>
<tr>
<td>R3</td>
<td>Adult vs. kids</td>
</tr>
<tr>
<td>R5</td>
<td>Flavor</td>
</tr>
<tr>
<td>R7</td>
<td>Cost</td>
</tr>
<tr>
<td>R9</td>
<td>How eaten</td>
</tr>
<tr>
<td>R11</td>
<td>Toasted</td>
</tr>
<tr>
<td>R2</td>
<td>Flavor</td>
</tr>
<tr>
<td>R4</td>
<td>Calories</td>
</tr>
<tr>
<td>R6</td>
<td>Texture</td>
</tr>
<tr>
<td>R8</td>
<td>Fillingness</td>
</tr>
<tr>
<td>R10</td>
<td>Shape</td>
</tr>
</tbody>
</table>

Product key

<table>
<thead>
<tr>
<th>Product key</th>
<th>Breakfast item labels and attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>Toast pop-up</td>
</tr>
<tr>
<td>BT</td>
<td>Buttered toast</td>
</tr>
<tr>
<td>JD</td>
<td>Jelly donut</td>
</tr>
<tr>
<td>CT</td>
<td>Cinnamon toast</td>
</tr>
<tr>
<td>BMM</td>
<td>Blueberry muffin and margarine</td>
</tr>
<tr>
<td>HR8</td>
<td>Hard rolls and butter</td>
</tr>
<tr>
<td>TMD</td>
<td>Toast and marmalade</td>
</tr>
<tr>
<td>EMM</td>
<td>English muffin and margarine</td>
</tr>
<tr>
<td>BTJ</td>
<td>Buttered toast and jelly</td>
</tr>
<tr>
<td>TMn</td>
<td>Toast and margarine</td>
</tr>
<tr>
<td>CB</td>
<td>Cinnamon bun</td>
</tr>
<tr>
<td>DP</td>
<td>Danish pastry</td>
</tr>
<tr>
<td>CC</td>
<td>Coffee cake</td>
</tr>
<tr>
<td>CMR</td>
<td>Corn muffin and butter</td>
</tr>
<tr>
<td>GD</td>
<td>Glazed Donut</td>
</tr>
</tbody>
</table>

4. The application: Breakfast/snack food preferences

4.1. The study

The data for this study was collected by Green and Rao (1972, pp. 167–180) at the Wharton School of the University of Pennsylvania from 21 Wharton School MBA students and their wives. The 42 respondents were asked to consider six preference situations and indicate their preferences for 15 breakfast food items: toast pop-up, buttered toast, English muffin and margarine, jelly donut, cinnamon toast, blueberry muffin and margarine, hard rolls and butter, toast and marmalade, buttered toast and jelly, toast and margarine, cinnamon bun, Danish pastry, glazed donut, coffee cake, and corn muffin and butter. The preference scenarios included their overall preferences between the 15 breakfast food items as well as five other breakfast and snack situation contexts reflecting Belk's (1975) Task Definition: 1. “Overall preference”, 2. “When I’m having a breakfast consisting of juice, bacon and eggs, and beverage”, 3. “When I’m having a breakfast consisting of juice, cold cereal, and beverage”, 4. “When I’m having a breakfast consisting of juice, pancakes, sausage, and beverage”, 5. “Breakfast with beverage only”, 6. “At snack time with beverage only”. Respondent background information was also collected. Respondents indicated their age, sex, marital status, number of children they had, if they considered themselves overweight and how much overweight they considered themselves, their typical week day and weekend breakfasts, and their coffee, tea, and milk drinking habits. Finally, seven point semantic differential ratings were collected by subject for each of the 15 snack food items on some ten perceptual attributes: Ease of preparation, Flavor, Adult orientation, Calories, Artificial–Natural, Texture, Expense, Fillingness, Context of consumption, and Shape. We added an 11th attribute representing whether or not the item was typically toasted or not based on previous MDS analyses performed on various portions of this data. The plotting labels to be utilized for these snack food items and attributes are presented in Table 4.

4.2. The derived spatial configuration of 15 breakfast/snack food items

These conditional rank order preference data (especially overall preference) have been analyzed extensively (both as metric and non-metric) with a variety of traditional multidimensional scaling procedures. Typically, two or three dimensions such as sweetness, toasted-non-toasted, and consistency have been derived through the use of such two-way spatial procedures (see Green and Rao (1972)). DeSarbo and Carroll (1981) applied their three-way metric unfolding procedure to the complete data array and also found two dimensions (sweetness and consistency), but derived negative weights for one of the six situations which proved problematic interpretation wise.

We performed our analysis on the entire centered three-way data (i.e., we subtracted the grand mean) with the proposed clusterwise spatial procedures for \( s = 1 \ldots 5 \) clusters and \( r = 1 \ldots 5 \) dimensions across the many different possible models varying by type of clustering (partitions vs. overlapping clusters), weighted vs. simple unfolding models, constrained vs. unconstrained weights models, preference representation (context dependent vs. stationary), etc. Five runs per model and value of R and S with rational and different random starts were performed, and the best solution was selected (Model 2, \( S = 3, R = 2 \)).

To investigate for local optima with respect to the application, 100 runs were run using different random starts. The VAF = 0.345 solution presented here, encountered among the 5 initial runs, was still the best solution. With all 100 VAFs for Model 2 (\( S = 3, R = 2 \)) found being between 0.336 and 0.345, all were within 3% of the optimal solution. Additionally, no loss function weights appeared necessary for this application (i.e., \( q_{ik} \) were all set equal to 1) since degenerate solutions did not appear problematic. Table 5 presents the VAF goodness-of-fit values R and S values for the most complex (weighted, unconstrained, context dependent, overlapping—\( M_5 \)) vs. least complex (simple unfolding, stationary, partitions, constrained—\( M_1 \)) clusterwise MDU models as defined in terms of number of free parameters. We
do see a noticeable difference in fit values across comparable (R, S) solutions between these two classes of clusterwise MDU models (\(M_1\) vs. \(M_3\)). Based on comparable fit values across a range of other potential clusterwise MDU models fit, the non-overlapping, context dependent, unweighted, and unconstrained model \(M_2\) appears to fit the data best in terms of parsimony given the number of parameters and corresponding fit values. The corresponding goodness-of-fit values are also displayed in Table 5 for this type \(M_2\) of clusterwise MDU model. One immediately notices how comparable such solutions are to the most complex model solutions in terms of overall fit suggesting here that the preferences are quite context dependent. In particular, the \(S = 3\), \(R = 2\) context dependent, unweighted unfolding model with non-overlapping clusters solution appears most parsimonious with \(SSE = 45\, 431.33\) and \(VAF = 0.345\) fit values observed. Fig. 2 presents the estimated stimulus space alone for the 15 breakfast/snack food items. Note, in order to better interpret the corresponding space, we property fit (via linear regression procedures as suggested in Green and Rao (1972, pp. 66–72)) the 10 attributes that were evaluated by each of the respondents for each of the breakfast/snack food items plus our 11th attribute reflecting toasted. This property fitting procedure finds the best direction in the brand space for each attribute so that the projections of the brands onto the fitted attribute vector maximally correlate linearly with the original attribute rating. This method amounts to a multiple regression of the resulting brand configuration \(X\) with each attribute scale serving, in turn, as the criterion variable. We normalize the length of the associated vectors to be equal for convenience and show these vectors in the resulting figures. Table 6 presents the list of attributes with their correlations with each of the two derived dimensions which further aids in the interpretation. Dimension I (horizontal) appears to separate the sweeter and more filling items on the left hand side from the less sweet, toasted, and less filling items on the right hand side. Dimension II (vertical) appears to further aid in the interpretation. Dimension I (horizontal) appears to separate the sweeter and more filling items on the left hand side from the less sweet, toasted, and less filling items on the right hand side. Dimension II (vertical) appears to separate items that are easier to prepare and mainly for children (top) vs. those items that are mainly for adults and harder to prepare.

Our clusterwise MDU model solution of the preferences elicited by these individuals across six breakfast and snack time contexts simultaneously revealed three clusters or market segments of individuals for the 15 breakfast food items on the two dimensions discussed, and the shifts from their overall preferences (one can treat this as a base condition) across different situations. We now examine Fig. 3(a)–(c) that represent the floating/multiple ideal point model solution respectively for Clusters/Segments 1, 2, and 3. We use Fig. 2 as a base and add in the situation specific ideal points by derived cluster/segment.
4.3. The three derived clusters/segments

The three clusters/segments are different from each other in terms of the overall preferences (base condition or situation 1) of the individuals within the same cluster/segment as well as the ways the individuals in these clusters/segments shift from their overall preferences with varying contexts (situations 2–6). As shown in Fig. 3(a), Cluster/Segment 1 consists of individuals who have an overall preference (situation 1) for mainly toasted breakfast products. The individuals in this cluster seem to prefer toasted breakfast products across most contexts. Only when they are considering a snack time option do they change their preference and indicate a preference for muffins, which are the more healthy of the sweeter items. These individuals do not engage in much variety seeking as indicated by the relative homogeneity in these preference ideal point orientations over all six situations. The preferences of the individuals in this cluster/segment suggest that these individuals have a tendency to avoid variety across situations, and even when the context changes, these individuals do not venture too far from their overall preferences. It is also clear that the members of this cluster/segment tend to avoid the sweeter, high caloric, more filling items. Of the three derived clusters/segment, this groups seemingly appears most Health Conscious.

The overall preferences of the individuals in Cluster/Segment 2 in Fig. 3(b) are quite different in comparison to Cluster/Segment 1. These individuals have an overall preference toward sweeter and high-caloric breakfast products such as coffee cakes and Danish pastries. These individuals also seem to have higher variety seeking tendencies across contexts compared to the individuals in Cluster/Segment 1. These individuals seem to seek sweeter options when they are having a light breakfast consisting of a single breakfast option and a beverage. When they are having heavier breakfasts consisting of high proteins and fats they seem to seek out options that contain fewer calories such as buttered toast. This preference to consume sweeter options only when having lighter breakfasts suggests a Diet Balance attitude. The individuals in Cluster/Segment 2 seem to be striving to maintain a balanced diet or a certain total caloric intake. For instance they choose heavier more substantial options (i.e. Danish pastries) only when the item is the only item that they are consuming along with a beverage (“Breakfast with beverage only” or “At snack time with beverage only”). The avoidance of heavy options when consuming more items than just a beverage further demonstrates that these individuals are spending effort to balance the food that they consume.

Cluster/Segment 3 in Fig. 3(c) consists of individuals with overall preferences \( Y_1 \) similar to those individuals in Cluster/Segment 2. However, unlike the members of Cluster/Segment 2, the individuals in Cluster/Segment 3 do not appear...
to be very high variety seekers across contexts. These individuals have a high preference for Sweet breakfast/snack products across most contexts. They indicate a different preference only when they are considering a situation where they are consuming a heavier breakfast composed of high proteins and fats ("When I’m having breakfast consisting of juice, bacon and eggs and beverage"). In this heavy breakfast situation, they prefer the lighter toasted options. This implies that this cluster of individuals does not tend to seek variety across contexts, and they have somewhat stable or enduring preferences or tendencies to prefer a particular kind of food (Sweet items).

Can we gain any further insight as to what other characteristics the members of these three derived clusters/segments share? Recall, \( P \) is the estimated binary indicator matrix of cluster/segment membership for the 42 subjects in the study. Based on the \( P \) column summations, 11 (26.2%), 17 (40.5%), and 14 (33.3%) subjects were allocated to each of the three derived clusters/segments respectively. In relating this binary membership matrix to collected demographic variables, we can observe the following demographic tendencies for the derived clusters/segments. Cluster/Segment 1 (Health Conscious) is typically male, older, has a family with children, and a coffee drinker. Cluster/Segment 2 (Diet Balancers) is typically female, younger, not married, not considered over-weight, a tea drinker, and tends to eat lighter breakfasts. Cluster/Segment 3 (Sweet Eaters) is typically male, younger, married, overweight, a milk and coffee drinker, and eats heavier breakfasts.

Thus, it appears that the individuals in Clusters/Segments 1 and 3 have somewhat more stable preferences and the different contexts/situations may not activate widely different goals for them. The different breakfast contexts appear to activate balanced nutrition goals for the individuals in Cluster/Segment 2. These individuals seem to make more different choices across situations and shift from their overall preferences to be able to meet their goal of having balanced nutrition. However, the various patterns of ideal point configurations across contexts are dramatically different for the three derived

---

Fig. 3. Cluster estimated joint spaces.
5. Discussion

We have thus proposed a new spatial clusterwise MDU approach for the analysis of two or three-way dominance data. Technical descriptions of the proposed model, its various model options, and alternating least-squares estimation procedures have been provided. The area of context/situational dependent or dynamic preferences has been the focal point of application where the pertinent psychological and behavioral decision theory literature was reviewed documenting the existence of changing preferences in response to differing contexts, situations, time, etc. The Green and Rao (1972) three-way breakfast/snack food study was presented, including a brief history of previous MDS results with this data (mostly involving the overall preference two-way data). In applying the proposed methodology to the entire three-way data, two dimensions were estimated summarizing the sweetness, toasted, and adult orientation attributes underlying these 15 breakfast/snack foods. In addition, three non-overlapping clusters/segments of subjects were estimated whose preference ideal points over the six situations were quite different from one another. In summary, the first (Health Conscious) and the third (Sweet Eaters) clusters/segments of subjects have somewhat more consistent preferences over situations compared to Cluster/Segment 2 (Diet Balancers). The second cluster of subjects (mostly young females) appears to have balanced nutrition motives. The pattern of ideal points across contexts was drastically different by derived cluster/segment.

A number of fertile avenues for future research exist in this particular area. One, extending the models to accommodate linear restrictions/reparameterization of the stimuli akin to CANDELINC (Carroll et al., 1980) and Desarbo and Rao (1984) would prove useful in aiding the interpretation of the derived space. Such estimated coefficients for the specified stimulus attributes could be jointly plotted as vectors in the derived space, similar to what was done with respect to our standard property fitting methods using ordinary regression procedures.

Two, extending the methodology to explicitly accommodate non-metric preference data may prove to be a useful research direction. We note that one of the limitations of the existing manuscript has been the application of a basically metric procedure to non-metric rank order preference data. While there are several documented applications in the two-way case where the corresponding results do not differ all that much between the two types of procedures (metric vs. non-metric MDS), one could synthetically create a data set where conditional monotone non-linear transformations of the underlying preferences would render meaningless results should a metric analysis be performed. (See Busing et al. (2005) for approaches for preventing degenerate solutions for the non-metric case).

Three, more work needs to be performed in such deterministic spatial models concerning model selection. Like all such spatial deterministic models, one is relegated to inspect a scree plot and/or rely on interpretation for model selection. Unlike many stochastic procedures, there are no information heuristics, Bayes factors, etc. to rely upon (although these heuristics have their own limitations as well) for selecting a particular solution in terms of S and R.

Finally, the proposed procedure requires significant computational effort and is therefore limited with respect to the size of the applications considered. It is clear that the proposed clusterwise MDU methodology (like the vast majority of other MDS procedures) is not appropriate for large scale data mining applications. More efficient estimation procedures should be explored in an effort to reduce the computational intensity required for estimation, as well as to resolve potential problems with local optimum solutions as is commonplace with many alternating least-squares MDS estimation algorithms.

References


