Available online at www.sciencedirect.com
SCIENCE DIRECT®
Food Quality and Preference 15 (2004) 195–208
Food Quality and Preference
www.elsevier.com/locate/foodqual

Outlier and group detection in sensory panels using hierarchical cluster analysis with the Procrustes distance

Tobias Dahl[a,*], Tormod Næs[b,c]

[a]Department of Informatics, University of Oslo, PO Box 1080, N-0316 Blindern, Norway
[b]Norwegian Food Research Institute, Ås, Norway
[c]Department of Mathematics, University of Oslo, Blindern, Norway

Received 14 August 2002; received in revised form 18 March 2003; accepted 12 April 2003

Abstract

Generalised Procrustes analysis (GPA) is a much-used method for analysing sensory profile data. In this paper, hierarchical clustering using the Procrustes distance is proposed for situations where the data profiles are believed to come from a non-homogeneous group. This new approach to sensory panel analysis may be used at an exploratory stage, in combination with GPA, to gain insight into the structures of the data set. It can help the researcher detect outliers and subgroups, help him/her make decisions regarding further analysis, and reduce the risk of erroneous inference about the data.

Keywords: Procrustes distance; Generalised Procrustes analysis; Dendrograms; Sensory analysis; Shape analysis

1. Introduction

In descriptive sensory analysis it is important to be able to define meaningful "average" profiles for the sensory panel to be used either for direct interpretation or for further statistical analyses of the data. The easiest way of doing this is to use the regular raw averages, but there are some obvious problems with this approach:

- there may be confusion about the use of terms (for instance saltiness and bitterness);
- there may be differences in the scaling used by the assessors; and
- the signal to noise ratios may be different.

Generalised Procrustes Analysis (GPA) is a technique frequently used to handle some of these problems. It is based on standardising profiles with respect to rotation/reflection, isotropic scaling and translation, in order to provide a better average, a so-called consensus configuration (Arnold & Williams, 1985).

Procrustes methods were first introduced in psychometrics, an important branch of multivariate statistical analysis. Useful references include Cliff (1966), Gower (1975), Green (1952), Gruvaeus (1970), Kristof and Wingersky (1971), Langron and Collins (1985), Mosier (1939), Schønemann (1966, 1968), Schønemann and Carroll (1970), Sibson (1978, 1979) and ten Berge (1977). Since the middle of the 1980s the method has been used as a standard tool in sensory analysis, due to important contributions by Arnold and Williams (1985), Dijksterhuis (1996), Qannari, MacFie, and Courcoux (1999), Qannari, Vigneau, Luscan, Lefebvre, and Vey (1997), Wakeling, Raats, and MacFie (1992) and Wu, Guo, deJong, and Massert (2002) among others. Procrustes methods have also found useful applications in another completely different branch of statistics, namely statistical shape analysis (see Dryden & Mardia, 1997; Goodall, 1991; Kendall, 1984, 1989). In this area, several interesting results have been developed, one of them being a new distance measure known as the Procrustes distance.

Even though Procrustes analysis can be a very useful tool, it has some shortcomings. Fig. 1 illustrates a situation where the Procrustes transformation is not very meaningful. Working with two or three-dimensional data, such problems can sometimes be spotted by simple

* Corresponding author.
E-mail addresses: tobias@ifi.uio.no (T. Dahl), tormod.naes@matforsk.no (T. Næs).

0950-3293/03/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0950-3293(03)00058-2

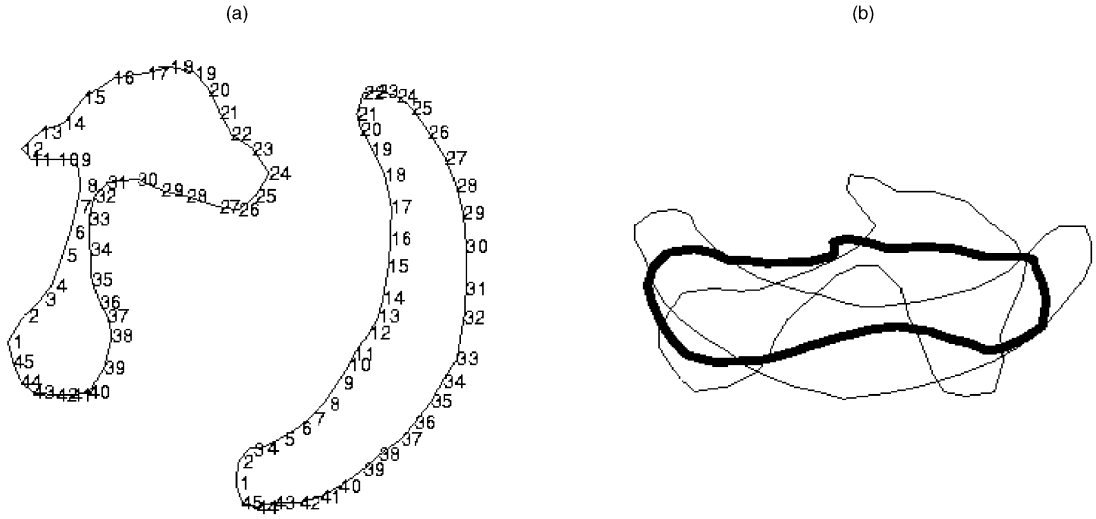

Fig. 1. An illustration of two configurations whose difference does not fit to a Procrustes model, i.e a model with translation, isotropic scaling and rotation. In (a) is presented two different configurations before tranformation and in (b) after transformation. The solid line in (b) is the consensus.

plotting techniques, but for higher dimensions, other techniques may be required. The present paper proposes a method inspired by Procrustes analysis that can be used to handle such problems. The method is based on grouping the assessors into clusters using a combination of hierarchical cluster analysis (HCA) and the Procrustes distance mentioned above. Using this approach, one can check the adequacy of the Procrustes model, detect possible groups of assessors and also obtain diagnostic information about assessors who are totally different from the rest. HCA with the Procrustes distance is primarily intended to be used in an informal and explorative way, at an early stage of an investigation. Questions such as determining the number of clusters have been considered by other authors (see e.g. Gordon, 1999), but will not be central in this paper. For other ideas concerning clustering for three-way data, see Carroll and Arabie (1983), Gordon and Vichi (1999), Krieger and Green (1999) and Vichi (1999).

Other approaches to the same problem can be found in Wakeling et al. (1992) and in Qannari, Courcuox, and Vigneau (2001). In the former paper a randomisation test was proposed for the hypothesis that there is no clear consensus among the profiles. The *P*-value of this randomisation test may give indications about possible problems with the overall consensus and may then point to the need for a method like the one proposed here. The second of the two papers proposes a method of clustering of data tables based on their agreement on a one-dimensional representation space. It may be possible to extend this method to something which is similar to the procedure presented here.

Two sensory data sets will be used to illustrate the use of the method proposed here. One of them is based on sensory analysis of sausages. The data will be manipulated in various ways for the purpose of illustrating some properties of the method. The other example is from sensory analysis of green peas and will be studied as a real life example. For the latter data set, near infrared (NIR) spectral data are also available and these will be used for comparison, i.e. for external validation. To illustrate the gains of this new approach, existing GPA diagnostics will be computed and compared with information obtained by the new method.

2. Methodology

2.1. Cluster analysis

The purpose of cluster analysis is to group N objects into G groups or clusters. The objects in each cluster should be ”similar” in some sense, making the clusters more homogeneous than the full set of data. There are different approaches to clustering, both hierarchical and criterion based ones (see e.g. Bezdec, Coray, Gunderson, & Watson, 1981). In this paper, the focus will be on the former. The results from such analyses are usually presented graphically in tree-structures called dendrograms. Even though all objects will, in such a process, end up in the same group, the grouping process as visualised by the dendrogram is interesting in itself. For instance, the length of the edges connecting the nodes, give information about the degree of dissimilarity between the clusters.

A hierarchical clustering process is based on distances or dissimilarities between the objects. In addition to these basic distances, one will during the clustering process need a measure of the distance between clusters G_i and G_j. This can be constructed in a number of different ways, depending on the nature of the objects and the purpose of the clustering. If the objects to be clustered

are vectors, a frequently used distance function between two clusters G_i and G_j is

$$d(G_i, G_j) = \min_{z \in G_i, w \in G_j} \|z - w\| \quad (1)$$

where $\| \ \|$ is the Euclidean norm and the z and w are symbols for members of the two clusters. Using this distance function leads to so-called single linkage clustering. Other variants will be described in Section 2.4, together with a distance function for matrices rather than vectors. For more information on clustering, see Gordon (1999), Kaufman and Rouseeuw (1990) and Mardia, Kent, and Bibby (1979).

2.2. Procrustes analysis and the Procrustes distance

If $\|.\|_F$ denotes the Frobenius norm of a matrix, the Procrustes transformation for a matrix X_1 to match with X_2 (both Nxp) is found by solving

$$\min_T \|T(X_1) - X_2\|_F = d(X_1, X_2) \quad (2)$$

with the requirement that the transformation T is composed of a rotation/reflection matrix P, $(P^T P = I)$ and an isotropic scaling factor c, i.e.

$$T(X_1) = X_1 c P. \quad (3)$$

The solution to the transformation problem can be written as

$$c = \frac{\mathrm{tr}(X_1^T X_2)}{\|X_1\|_F^2}, \quad P = UV^T \quad (4)$$

where $X_1^T X_2 = USV^T$ is the singular value decomposition of $X_1^T X_2$. This can easily be derived using standard results from linear algebra (Cliff, 1966 or Mardia et al., 1979).

The dissimilarity measure $\mathrm{d}(X_1, X_2)$ measures the degree of dissimilarity between two matrices after rotational and scaling effects have been removed, i.e. after they have been made as "similar" as possible. It is, however, not symmetric. Generally

$$\mathrm{d}(X_1, X_2) \neq \mathrm{d}(X_2, X_1) \quad (5)$$

This makes clustering difficult to interpret. If, however, the X_1 and X_2 are scaled to have the same variance (after centring), $\|X\|_F^2 = K$(usually K is set equal to 1), the distance becomes symmetric. The new distance is called the full Procrustes distance and can be written as

$$\min_T \left\| T\left(\frac{X_1}{\|X_1\|}\right) - \frac{X_2}{\|X_2\|} \right\|_F = d_F(X_1, X_2) \quad (6)$$

with the same requirement on T as above. A proof for the symmetry of d can be found in Dryden and Mardia (1997).

Translation is usually applied together with scaling and rotation. It can be proven that the optimal way of translating the point-sets is by column-centring the profiles. Throughout the paper, it is assumed that all profiles (shapes) are pre-processed in this way.

2.3. Generalized Procrustes analysis

Generalised Procrustes analysis (GPA) does more or less the same for several matrices as Procrustes Analysis does for two. It is based on computing an iteratively updated average or consensus Z, and provides together with the consensus Z, a set of rotated matrices as similar with Z as possible (Gower, 1975).

If T_i denotes the optimal (and implicitly defined) transformation of the ith profile X_i, the GPA is formally defined as minimisation of

$$\mathrm{g}(X_1, X_2, \ldots X_Q) = \sum_{i=1}^{Q} \|T_i(X_i) - Z\|_F^2 \quad (7)$$

with respect to Z and transformations T_i. The Q is here the number of matrices.

2.4. HCA with the Procrustes distance

If the profile matrices (X_1,...,X_Q) are different in a way that can not be accounted for by scaling, translation and rotation, it is difficult to justify the use of GPA and the consensus Z as a natural representation of all the data (see Fig. 1). There is obviously a need for diagnostic tools that can help obtaining this type of insight. Simple and obvious candidates for this are the individual contributions to the sum in (7), i.e.

$$g_i = \|T_i(X_i) - Z\|_F^2, \quad i = 1, \ldots Q. \quad (8)$$

Scatter plots of the different rotated assessors along principal components of the consensus can also be envisioned as potentially useful tools: There are, however, some intrinsic problems related to these methods.

- If for instance one profile is an outlier, it has already been allowed to influence the consensus Z when g_i is calculated. Thus, an outlier will seem less of an outlier after the iterative procedure.
- If there are several groups of profiles that are homogeneous within but not between groups, the g_i diagnostics may all be comparable in size and an uninteresting consensus may pass unnoticed.
- If regular two-dimensional plots of all profiles (after rotation and scaling) are used for a large number of assessors (for instance about 10) and objects (for instance 30), there will be very many points to represent and interpretation may be quite difficult.

As will be shown below, cluster analysis with the use of the Procrustes distance, is an important and simple helping tool for revealing group structures as well as detecting individuals who do not fit well with others. Visualisation by the use of dendrograms can help the researcher to detect when the averaging process "breaks down" due to the influence of profiles very different from the rest, as well as situations where subgroups of profiles differ substantially from one another. In the following we will describe in some more detail four different clustering strategies that will be tested below. Three of them have a common structure and will be discussed first.

2.4.1. Single, complete, average linkage

Let the sensory profile matrices $X_1,\ldots,X_Q$ be the objects to be grouped in a cluster analysis. At each step in the clustering process, each cluster G_i contains one or more objects (assessors). When there are more than one object/assessor in a cluster, it is as mentioned above not obvious how to define a distance from the object to the cluster. The following three candidates are much used in practice.

$$\mathrm{d_S}(G_i, G_j) = \min_{X\in G_i, Y\in G_j} \mathrm{d_F}(X, Y) \tag{9}$$

$$\mathrm{d_C}(G_i, G_j) = \max_{X\in G_i, Y\in G_j} \mathrm{d_F}(X, Y) \tag{10}$$

$$\mathrm{d_M}(G_i, G_j) = \operatorname*{average}_{X\in G_i, Y\in G_j} \mathrm{d_F}(X, Y) \tag{11}$$

They correspond to single, complete and average linkage respectively. From these three distances, three different HCA variants are derived.

Let $\{G_1,\ldots,G_Q\} = \{X_1,\ldots,X_Q\}$ be the initial clusters. A distance matrix $D = \{d_{ij}\}$, where $d_{ij} = \mathrm{d}(G_i, G_j)$ is used as a starting point. The minimum element value d_{ij} of the matrix D is identified and the corresponding two matrices are joined to a cluster. The distance from each matrix to the new cluster is formed using one of the three distances above. The procedure is repeated until there is only one cluster left, or until a stopping criterion determines the end of the process.

2.4.2. Centroid linkage and GPA

Centroid linkage reflects some intrinsic ideas of GPA more clearly than the other three methods. Rather than computing distances (minimum, maximum or average) between elements or clusters, each cluster can be represented by an average element, and the distance between clusters can be measured in terms of the full Procrustes distance between the average elements. A natural choice of average element could be the GPA consensus Z_i of the matrices in each cluster G_i,

$$d_{ij} = \mathrm{d_F}(Z_i, Z_j) \tag{12}$$

At the final step, this process is equivalent to GPA, because all matrices are members of the same cluster, and the average matrix computed for this cluster is the ordinary GPA consensus.

The main drawback with this so-called centroid method is that it fails to preserve a basic clustering property, the continuous increase of distance levels through the stages of the clustering process. Let d^s denote the distance function used to group the two clusters when there are $(g-s)$ clusters left. We should expect that

$$d^1 \leqslant d^2 \leqslant d^3 \ldots \leqslant d^{q-1} \tag{13}$$

This property can easily be derived for single, complete and average linkage. But see Fig. 2c, for an example where the centroid method fails to fulfil (13). This is commonly called "inversion" in hierarchical cluster analysis, and is typical when centroid linkage is used.

2.4.3. Ward's method

Another important clustering rule is Wards method. Two variants of Wards method can be envisioned in the setting of GPA and the Procrustes Distance, which are otherwise equivalent for Euclidean distances.

2.4.3.1. Distances between profiles and the GPA average. When two clusters A and B are considered for joining, the criterion used for this method is whether the average squared distance between the points in the common cluster (A,B) and their average element is larger than for another coupling, say (A,C). This average distance can be written as

$$E = \frac{1}{n}\sum_{i=1}^{n} \mathrm{d_F}\left(X_i^{\mathrm{AB}}, Y^{\mathrm{AB}}\right)^2 \tag{14}$$

where X_i^{AB}, Y^{AB} denote the members and the GPA average of the common cluster (A,B) respectively and n is the number of elements in the common cluster (A,B).

2.4.3.2. Distances between individual profile pairs. In the other approach the average distance between any two members of the joined cluster (A,B) is used as criterion, i.e.

$$E = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{d_F}\left(X_i^{\mathrm{AB}}, X_j^{\mathrm{AB}}\right)^2 \tag{15}$$

It is well known from multivariate analysis that the squared sum of distances to the mean is proportional to the squared sum of distances between the individual pairs. However, this only applies when the mean can be defined in a straightforward manner, e.g. by taking the normal average.

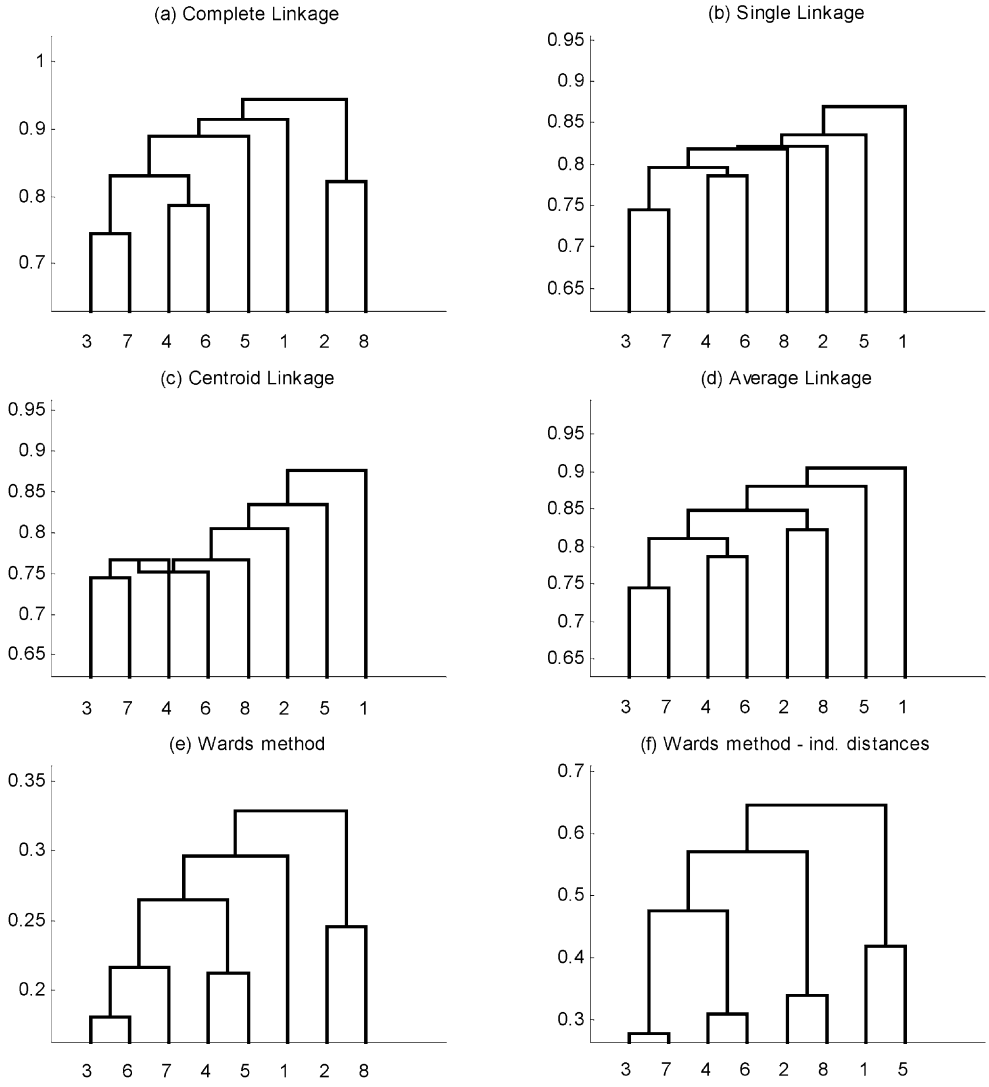

Fig. 2. The dendrograms (for all the four clustering methods) for basic study of the sausage data. No manipulated data points.

3. Example 1: detecting group structure and outliers in sausage data

The first data set considered is based on testing sensory properties of a number of different sausages made from different recipes (based on an experimental design). The original data are described in (Baardseth et al., 1992) and contain measurements on $N=60$ different sausages made by $Q=8$ assessors using $p=8$ sensory attributes (whiteness, colour intensity, smokiness, off-taste, rancidity, firmness, fatness and stickiness). Each individual matrix was column-centred and scaled to have unit variance (or unit Frobenius norm) before analysis. For the purpose of the present paper the data set was manipulated in various ways in order to highlight important properties of the method proposed.

3.1. Basic study

The four clustering variants were first tested on the original profiles $X_1,\ldots,X_8$ as they are. The resulting dendrograms can be found in Fig. 2. As can be seen, the different methods give very similar results. Some of the observed features are related to basic properties of the different algorithms, for instance:

- The complete linkage tends to have longer edges in the tree than the others. This is because it joins cluster elements of maximum distance, and thus takes longer before connecting single objects with large clusters. The chance that one object in a big cluster is far away from a single object outside the cluster is usually large. Rather than grouping

objects quickly, it tends to pair up objects, then pairs of pairs, and so on.
- The single linkage method, on the contrary, easily allocates new (single) objects to clusters with many members. It is here much easier to find one within-cluster object close to the new object.
- The centroid method has the "inversion", which is typical for centroid methods, and to be expected. This is due to the calculation of a cluster representative (centroid) which does not conform with the principle of an increasing sequence of distances.
- The average linkage seems, not surprisingly, to be an in-between solution of single and complete linkage.
- Both variants of Wards method produce tighter clusters, which is in accordance with the known properties of this clustering rule.

As can be seen, the complete linkage method differs from the others in that it connects object 2 and 8 at an early stage. From the figures, one can also see that assessors 4, 6, 3 and 7 are very similar to each others and there is an indication that assessor 1 is different from most of the other assessors in some way. The GPA distances are

Table 1
GPA distances for the basic study of the sausage data (eight assessors) (normalised by dividing each value on the largest)

Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6	Assessor 7	Assessor 8
1.00	0.96	0.89	0.89	0.96	0.91	0.90	0.92

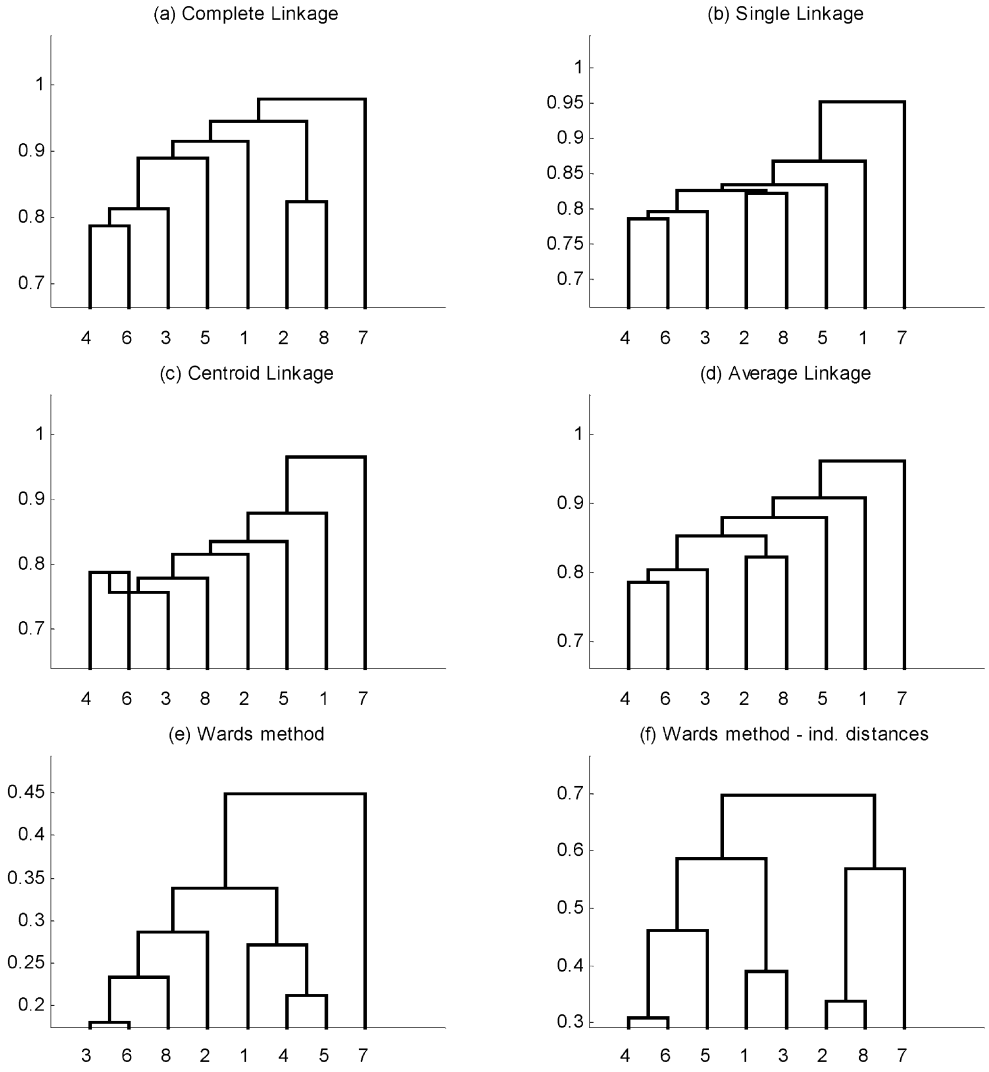

Fig. 3. The dendrograms (for all the four clustering methods) for the sausage data. Assessor 7 manipulated to be an outlier.

given in Table 1. They are quite similar in size and there is no particular feature to comment further on.

3.2. Outlier detection

The next step is to illustrate the method's ability to detect outliers in the data. The data matrix for assessor number 7 was turned into an outlier by replacing his/her scores by random noise in the same interval as the original measurements. The resulting dendrograms are found in Fig. 3. As can be seen, all variants of the clustering methods, except Ward's method based on the individual distances, correctly identify the outlier, but the complete and average linkage are less clear than the centroid and single linkage. This corresponds well with general properties of the methods (see basic study above). The problem with the Ward's method here could be its tendency to produce tight clusters—it joins the outlier with another single object, rather than grouping existing clusters (pairs) into larger clusters.

For comparison we also computed the individual GPA values g_i. Each element was divided by the largest element giving the maximum value equal to 1. From Table 2 it is clear that these values give no indication of any problem with assessor 7.

Table 2
GPA distances for the sausage data with one manipulated outlier (eight assessors, assessor 7 is an outlier) (normalised by dividing each value on the largest)

Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6	Assessor 7	Assessor 8
0.97	0.98	1.00	1.00	0.98	0.99	0.94	0.99

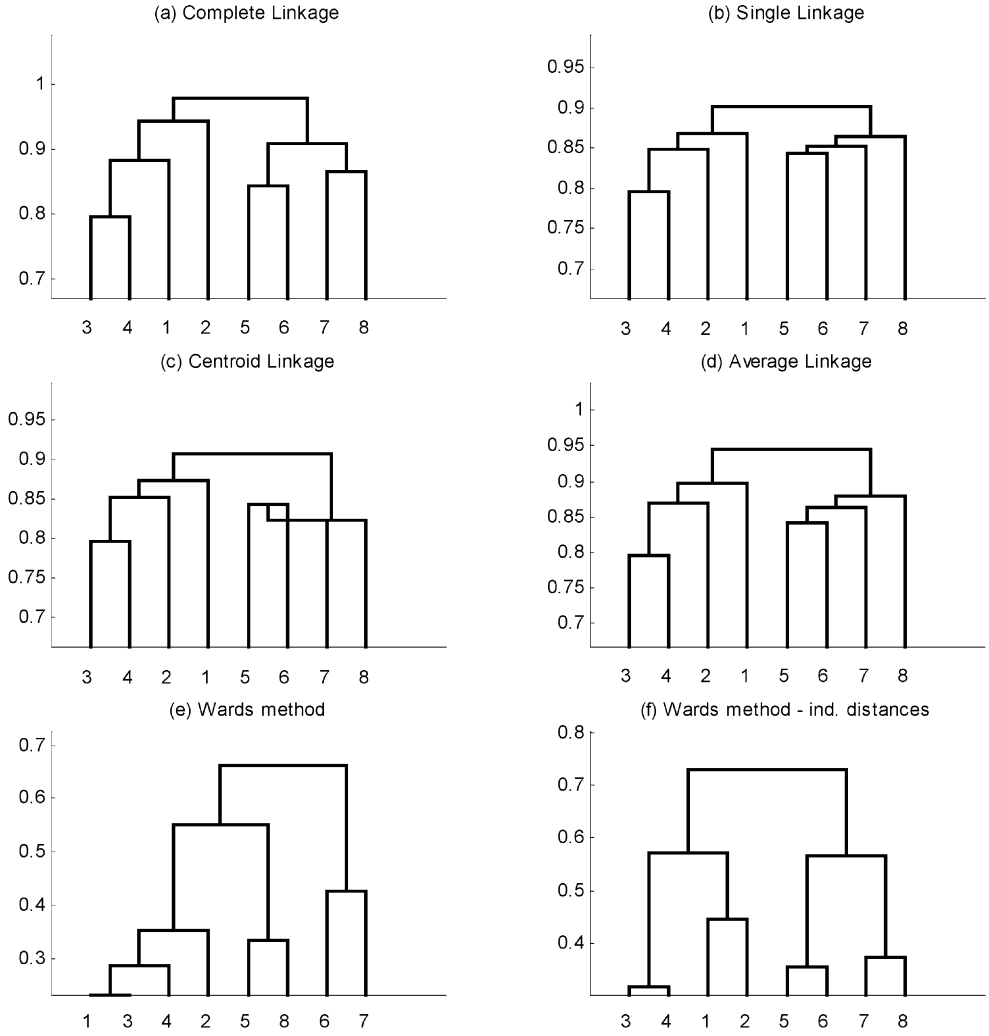

Fig. 4. The dendrograms (for all the four clustering methods) for the sausage data. Here four assessors ar manipulated in order to create two distinct groups. The groups are (1, 2, 3, 4) and (5, 6, 7, 8).

3.3. Detecting two distinct classes

In order to illustrate what happens in a situation with two clearly distinct subgroups, the original data were manipulated the following way: The original profiles were first transformed according to regular generalised Procrustes analysis. Then, the singular values for the two first components of the consensus profile were switched. New matrices for four of the assessors (5, 6, 7, 8) were then created by using the same transformation matrix as used for switching the singular values of the consensus. The other four were kept as they are. See the Appendix for details.

The clustering results from the analysis of these data can be seen in Fig. 4. The two groups of assessors are clearly visible in the dendrogram for each of the clustering methods, except for the Ward's method based on distances to the centre.

The Procrustes fits are given in Table 3. As above, no indication is given on the clustering into two groups. The other classical tool mentioned above for possible detection of outliers and group structures was also tested here; a plot of the projections of all rotated assessors onto the two first components of the consensus matrix is presented in Fig. 5. As can be seen, it is impossible to see any tendencies or structures in the data.

Table 3
GPA distances for the sausage data with two distinct groups [eight assessors in two groups (1, 2, 3, 4) and (5, 6, 7, 8)] (normalised by dividing each value on the largest)

Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6	Assessor 7	Assessor 8
0.89	0.91	0.93	1.00	0.94	0.94	0.97	0.94

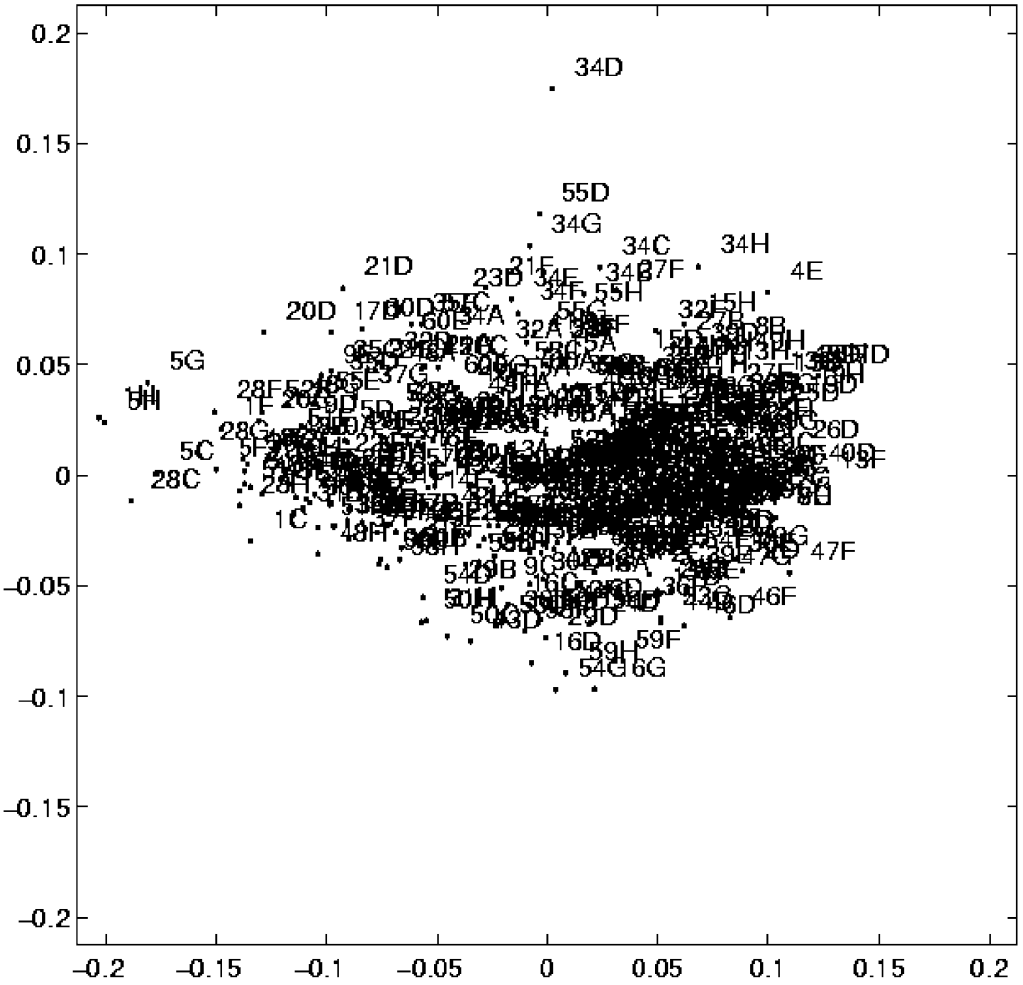

Fig. 5. Projections of all rotated assessors onto the first two components of the consensus matrix. The same data as used for Fig. 4 with two manipulated groups of assessors. Symbols 1–60 represent samples and A–H represent assessors.

3.4. Detecting two distinct classes and one outlier

Finally, the two situations above were combined. From the results in Fig. 6 we see that the clustering methods single linkage, centroid linkage and the average linkage were able to detect the essential structure in the data. The complete linkage failed to detect the outlier. The two variants of the Ward's method also failed, but in different ways. The "individual distances" Ward's method fails in the same way as it did in Fig. 3 whereas the other one commits a less serious error: It joins the outlier with one of the clusters before the final step where all clusters are joined. Note that the two clusters should have more in common (e.g. the principal directions) than an outlier should have with any of them.

Again the regular GPA method (Table 4) was unable to detect the essential structure of the data.

4. Example 2: analysis of descriptive sensory data from peas

The pea data used for this example have previously been analysed by Næs and Kowalski (1989) and the

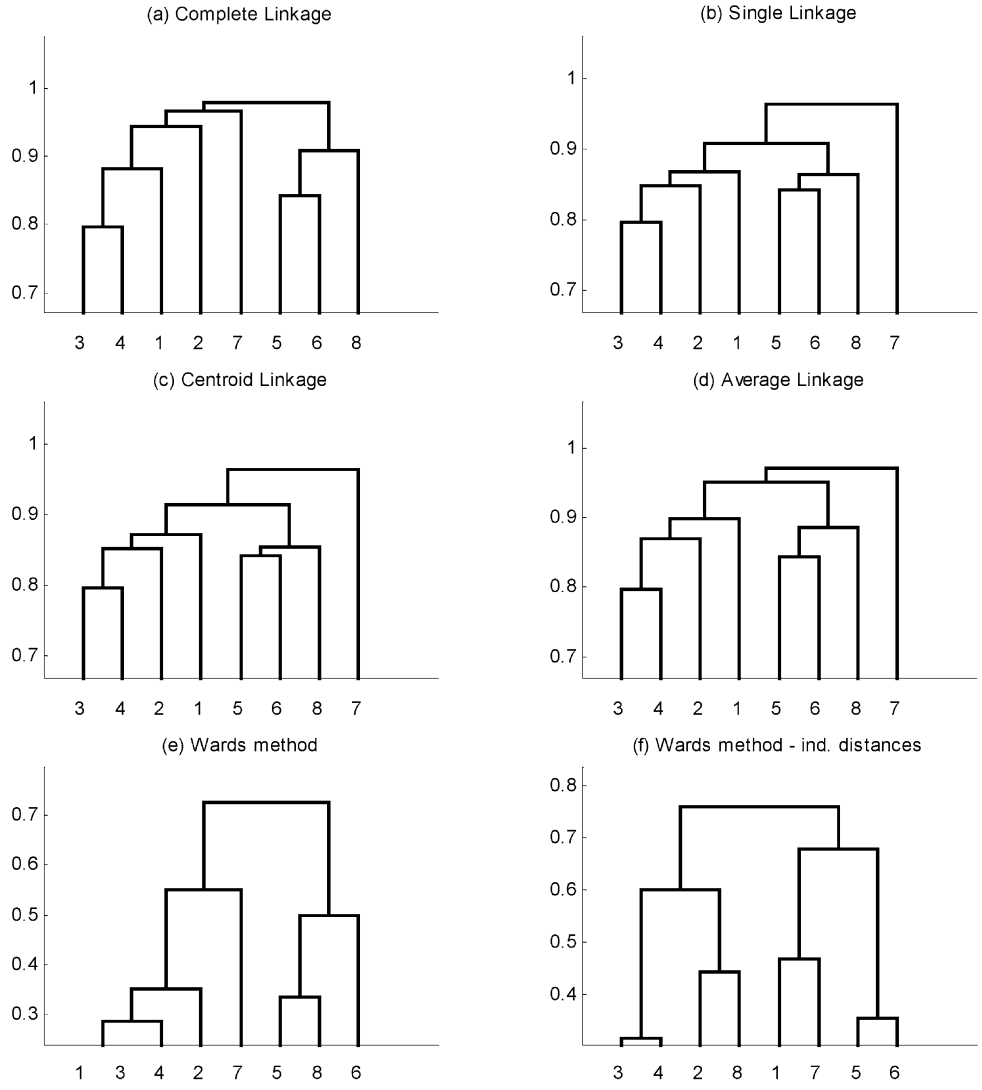

Fig. 6. The dendrograms (for all the four clustering methods) for the sausage data. In this case, ther is one manipulated outlier (assessor 7) and two manipulated distinct groups (1, 2, 3, 4) and (5, 6, 8).

Table 4
GPA distances for the sausage data with one outlier and two distinct groups [eight assessors, assessor 8 is an outlier, the two groups are (1, 2, 3, 4) and (5, 6, 8)] (normalised by dividing each value on the largest)

Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6	Assessor 7	Assessor 8
0.84	0.88	0.89	1.00	0.85	0.85	0.68	0.85

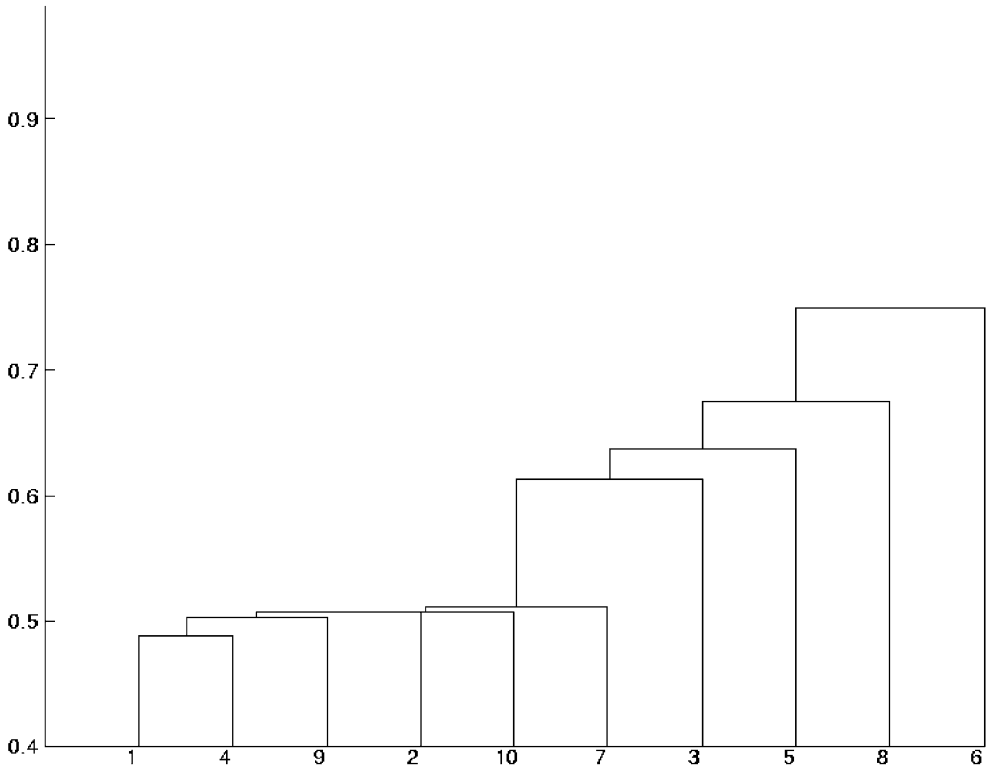

Fig. 7. The dendrograms for the pea data (single linkage).

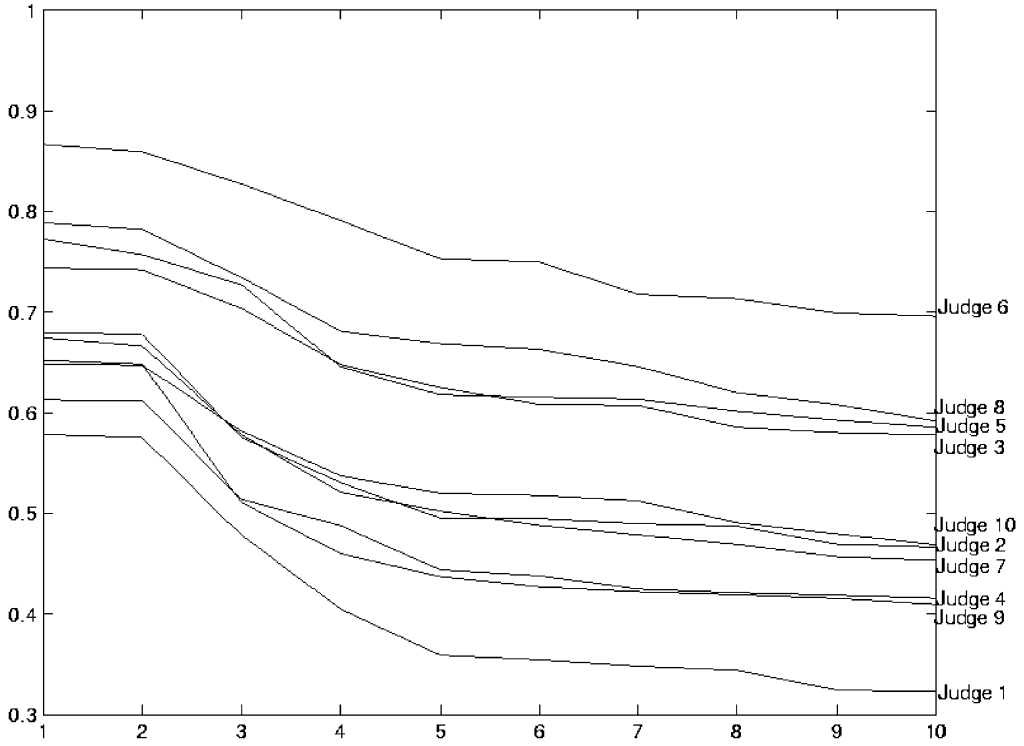

Fig. 8. MSEP curves (as functions of the number of principal components) for prediction of sensory data from NIR data. Each curve corresponds to an individual assessor.

reader is referred to that paper for details. The data contain sensory measurements made by $Q = 10$ assessors for $N = 60$ different samples of peas (different varieties and different degree of maturity). 10 sensory attributes were measured, but in this paper we only consider $p = 4$ of them (pea flavour, sweetness, off-flavour and mealiness). There were two replicates for each sample and these were averaged before statistical analysis. For each sample, a near infrared (NIR) spectrum was also available. The NIR data in this case contained absorbance readings at 116 different wavelengths. Below we will relate these Near Infrared (NIR) spectra to the sensory data for comparison. These computations will also illustrate the use of external information for validation.

For this data set we confine ourselves to single linkage only. The dendrogram for the pea data using this technique is shown in Fig. 7. It gives a clear idea about the similarity and differences among the assessors. First of all, there is a clear group of 6 assessors, 1,2,4,7,9, and 10, who are very similar to each other. The (full) Procrustes distance (vertical axis) between assessor 1 and 4 is only slightly smaller than between 1 and 7, which is the last one joined to this cluster. The next assessor to be joined is number 3, which is considerably further away. Assessor number 6 seems quite different from the rest in this study.

In order to verify these conclusions, the sensory profiles were related to principal components of the NIR data. First of all, each individual assessor was considered separately (Fig. 8). As can be seen, after 4-5 principal components of the NIR data the prediction ability is reasonably good for some of the assessors. The prediction error is here defined as the sum of the prediction errors (MSEPs) of all the attributes. In all cases the sensory variables are scaled using the Frobenius norm before computations. It is also clear that the assessors in the group that was determined to be rather homogeneous (1,2,4,7,9 and 10), are the ones that are easiest to predict. Assessor number 6 is clearly the one with the least clear relationship between sensory and NIR data. The assessor 3,5 and

Table 5
GPA distance for the pea data (ten assessors) (normalised by dividing each value on the largest)

Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6	Assessor 7	Assessor 8	Assessor 9	Assessor 10
0.39	0.54	0.61	0.47	0.79	1.00	0.55	0.78	0.50	0.49

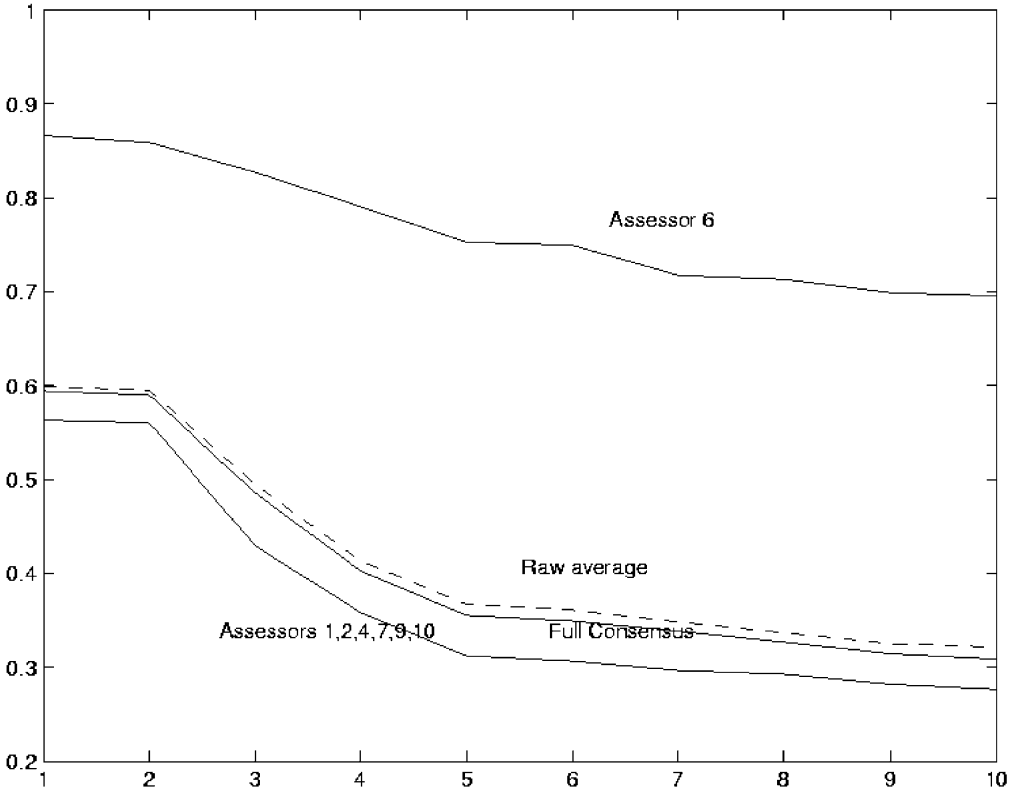

Fig. 9. MSEP curves (as functions of the number of principal components) for prediction of sensory data from NIR data. The lower line correspond to the consensus for the assessors 1, 2, 4, 7, 9 and 10.

8 come in an intermediate position. Fig. 9 presents similar prediction results for the consensus of the six similar assessors, for the full panel and for the raw average of the profiles. Scaling was done as above. Assessor 6 is also plotted for comparison. The results show that the consensus from the six similar ones is clearly easier to relate to the NIR data than the full consensus and the raw average.

These results together clearly indicate that assessors 1, 2, 4,7, 9 and 10 are similar and have a simpler and more predictable relationship to NIR than the other four. Then there is a gap to the next group 3, 5, 8 before a new gap separates assessor 6 from the rest. A possible and quite tempting explanation for this is that 1, 2, 4, 7, 9 and 10 are simply more reliable than the rest of the assessors in this case.

The GPA fit results are given in Table 5. For this case, it was quite clear that the assessor number 6 fits less well to the consensus than the others and that the assessors 5 and 8 were also less easy to fit. No clear indication is, however, given for assessor 3.

5. Discussion

5.1. Other distance functions

One of the main criticisms against GPA is the fact that it uses only rigid transformations (rotation, isotropic scaling and translation) to compensate for systematic differences between judges. However, there is no clear reason to believe that there does not exist any more subtle difference among assessors. There are methods that handle such differences better than GPA, for instance generalised canonical analysis (GCA), Tucker-2, Tucker-3 (Tucker, 1966) or PARAFAC (Harshman & Lundy, 1984). Since the new method does not actually carry out a regular GPA, one could group the objects in more flexible ways by using alternative dissimilarity measures. These could be constructed to detect similarity with respect to specific non-rigid transformations, such as for instance affine transformations, or thin plate splines for shape analysis. It might also be possible to design dissimilarity measures based on entropy measures (from information theory). In this case, the dissimilarity between two matrices would be determined from their joint entropy. In that case, one does not need to find an optimal mapping from one profile to another, it suffices to measure the degree of common information, which helps determining whether there is an optimal, possibly non-linear mapping (see e.g. Hyvärinen, 1999).

5.1.1. Alternative ways of studying profile data

In this paper, iterative averaging (GPA) and HCA with the Procrustes Distance have been investigated. These are, however, only two possible ways of exploring profile data. Another approach is to study profiles from the perspective of minimum spanning trees, which is strongly related to single linkage clustering. This technique is used in for instance botany (see e.g. Dahl, 1982; and Gauslaa, 1985) and computer networking. It has recently been employed to connect multiple PCA and PLS-models in chemometrics (unpublished work by Martens, Anderssen, & Høy, 2000). Minimum spanning trees generated by the Procrustes distance could be used to create a map (a graph) to see how judges relate.

6. Conclusions

The computations and simulations presented above illustrate a number of situations were HCA could be used in combination with GPA. The single linkage and the centroid linkage seem to be slightly better suited for identifying group structures and outliers than the other methods tested. The centroid variant has two drawbacks; it has the ''inversions'' which are generally considered inappropriate, and it requires the computation of a GPA consensus at each step, making it rather computer-intensive. The Ward's method did not seem to have as attractive properties as the other methods. Based on these results, single linkage HCA, using the Procrustes distance as dissimilarity measure, seems to be the method to recommend for sensory analysts. Since data sets may be quite different in structure, this aspect has to be studied in more detail before a more definite conclusion can be drawn. Through simulations, it has been demonstrated that two-group situations and outliers can be detected using this approach. In a real-life experiment, HCA was demonstrated as a first natural step when examining the panel data. The method needs further investigations to explore its more general properties and ability to detect groups and outliers.

Appendix. Manipulation of profiles to produce group structures (sausage data)

In order to to test the performance of the proposed method for group detection, we modified the profiles so as to produce group tendencies.

One intuitive way of producing groups would be to reorder the variables in one group, while keeping the ordering in the other fixed. (e.g. to permute the columns of some profiles, and not of others). However, this would not work, since the optimal rotation found by the Procrustes method would compensate for this reordering. Thus, a more sophisticated way of manipulation must be considered.

A common assumption in sensory experiments, is that there exists some common underlying tasting experi-

ence, and that this is reflected in the profiles. However, mapping from the "true underlying variables" to the observed variables will differ from judge to judge. Thus, the manipulation should happen at the level of the PCA model, and not on the level of the observed profiles.

To manipulate the data, we assume that there is an underlying PCA model. We rotate the profiles to agreement with GPA. Given this average, it is easy to construct a mapping that would reorder the importance of the two first principal components. We base this on the SVD of the consensus Y which is a real N by p matrix,

$$Y = USV^T \tag{A1}$$

where U (N by p) and V (p by p) are orthogonal matrices of left and right side singular vectors, and $S = \mathrm{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_p\}$ is a diagonal matrix containing the singular values in the descending order. If we construct a matrix W

$$W = VDV^T \tag{A2}$$

where

$$D = \mathrm{diag}\left\{\frac{\sigma_2}{\sigma_1}, \frac{\sigma_1}{\sigma_2}, \sigma_3, \ldots, \sigma_p\right\} \tag{A3}$$

Then, postmultiplying Y by W has the effect of interchanging the "importance" (associated variance) of the two top principal components. In other words, this interchanges the *strength of perception* associated with the two leading variables in the underlying PCA model. However, rather than applying this mapping W to the consensus Y, we apply it to a subset of the profiles that are rotated to match one another (and Y) by GPA. Thus, a group of profiles is created that has a different ordering of the variables (principal components) in the underlying PCA model.

References

Arnold, G. M., & Williams, A. A. (1985). The use of generalized Procrustes Techniques in sensory analysis. In J. R. Piggott (Ed.), *Statistical procedures in food research* (pp. 233–253). London: Elsevier.

Baardseth, P., Næs, T., Mielnik, J., Skrede, G., Hølland, S., & Eide, O. (1992). Dairy ingredients effects on sausage sensory properties studied by principal component analysis. *Journal of Food Science*, *57*(4), 822–828.

Bezdek, J. C., Coray, C., Gunderson, R., & Watson, J. (1981). Detection and characterization of cluster substructure. I: Linear structure: fuzzy c-lines. *SIAM J. Appl. Math.*, *40*, 339–357.

Cliff, N. (1966). Orthogonal rotation to congruence. *Psykometrika*, *31*, 33–42.

Carroll, J. D., & Arabie, P. (1983). INDCLUS: an individual differences generalization of the ADCLUS model and the MAPCLUS algorithm. *Psychometrika*, *48*, 157–169.

Dahl, E. (1982). *Ordinasjonsanalyse. Analyser av tilfeldig valgte prøveflater* (Ordination analysis. Analysis of randomly selected test areas). Unpublished work, NLH—Norwegian Agricultural University.

Dijksterhuis, G. (1996). Procrustes analysis in sensory research. In T. Næs, & E. Risvik (Eds.), *Multivariate analysis of data in sensory science*. Elsevier Science Publishers.

Dryden, I. L., & Mardia, K. V. (1997). *Statistical shape analysis*. UK: J. Wiley and Sons.

Gauslaa, Y. (1985). The ecology of Lobarion Pulmonariae and Parmelion Capetarae in Quercus dominated forests in south-west Norway. *Lichenologist*, *17*(2), 117–140.

Gordon, A. D. (1999). *Classification. Monographs on Statistics and Applied Probability* (2nd ed.). Chapman and Hall.

Gordon, A. D., & Vichi, M. (1999). Partitions of partitions. *Journal of Classification*, *15*, 265–285.

Goodall, C. (1991). Procrustes methods in the statistical analysis of shape (with discussion). *Journal of Roy. Stat. Soc. B*, *53*, 285–339.

Harsman, R. A., & Lundy, M. E. (1984). The PARAFAC model for three-way factor analysis and multidimensional scaling. In H. G. Law, C. W. Sbyder Jr., J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multi-mode data analysis* (pp. 122–215). NY: Praeger.

Krieger, A. M., & Green, P. E. (1999). A generalized Rand-Index method for consensus clustering of separate partitions of the same data base. *Journal of Classification*, *16*, 63–89.

Gower, J. C. (1975). Generalized Procrustes analysis. *Psychometrika*, *40*, 33–51.

Green, B. F. (1952). The orthogonal approximation of an oblique structure in factor analysis. *Psychometrika*, *17*, 429–440.

Gruvaeus, G. T. (1970). A general approach to Procrustes pattern rotation. *Psychometrika*, *35*, 493–505.

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. on Neural Networks*, *10*(3), 626–634.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data*. New York, USA: John Wiley and Sons.

Kendall, D. G. (1984). Shape manifold, Procrustean metrics and complex projective spaces. *Bulletin of the London Mathematical Association*, *16*, 81–121.

Kendall, D. G. (1989). A survey of the statistical theory of shape. *Statistical Science*, *4*, 87–120.

Kristof, W., & Wingersky, B. (1971). Generalizations of the orthogonal Procrustes rotation procedure to more than two matrices. *Proceedings of the 79th Annual Convention of the American Psychological Association*, *6*, 81–90.

Langron, S. P., & Collins, A. J. (1985). Perturbation theory for generalized Procrustes analysis. *Journal Roy. Stat. Soc.*, *47*, 277–284.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. Academic Press.

Martens, H., Anderssen, E., & Høy, M. (2000). *Unpublished work on minimum spanning trees for PCA and PLS models*. openhagen, Denmark: KVL.

Mosier, C. I. (1939). Determining a simple structure when loadings for certain tests are known. *Psychometrika*, 149–162.

Næs, T., & Kowalski, B. (1989). Predicting sensory profiles from external instrumental measurements. *Food Quality and Preference*, *4*/*5*, 135–147.

Qannari, E. M., Vigneau, E., Luscan, P., Lefebvre, A. C., & Vey, F. (1997). Clustering of variables, application in consumer and sensory studies. *Food Quality and Preference*, *5*/*6*, 423–428.

Qannari, E. M., MacFie, H. J. H., & Courcoux, P. (1999). Performance indices and isotropic scaling factors in sensory profiling. *Food Quality and Preference*, *10*, 17–21.

Qannari, E. M., Courcoux, P., & Vigneau, E. (2001). Common components and specific weights analysis performed on preference data. *Food Quality and Preference*, *12*, 365–368.

Sibson, R. (1978). Studies in the robustness of multidimensional scaling: Procrustes statistics. *Journal Roy. Stat. Soc. B*, *40*, 234–238.

Sibson, R. (1979). Studies in the robustness of multidimensional

scaling: perturbation analysis of classic scaling. *Journal Roy. Stat. Soc., B*, *41*, 217–229.

Schønemann, P. H. (1966). A generalized solution to the orthogonal Procrustes problem. *Psychometrika*, 1–10.

Schønemann, P. H. (1968). On two sided orthogonal Procrustes problems. *Psychometrika*, 19–33.

Schønemann, P. H., & Carroll, R. M. (1970). Fitting one matrix to another under choice of central dilation and rigid motion. *Psychometrika*, 245–255.

ten Berge, J. M. F. (1977). Orthogonal Procrustes rotation for two or more matrices. *Psychometrika*, *42*, 267–276.

Tucker, L. R. (1966). Some mathematical notes on three-mode factors analysis. *Psychometrika*, *31*, 279–311.

Vichi (1999). One-mode classification of a three way data matrix. *Journal of Classification*, *16*, 27–44.

Wakeling, I. N., Raats, M. N., & MacFie, H. J. H. (1992). A new significanec test for consensus in generlised Procrustes analysis. *J. Sensory Studies*, *7*, 91–96.

Wu, W., Guo, Q., deJong, S., & Massart, D. L. (2002). Randomisation test for the number of dimensions of the groups average space in generalised procrustes analysis. *Food Quality and Preference*, *13*, 191–200.