**A comparison of Generalised Procrustes Analysis and Multiple Factors Analysis for projective mapping data**

Tomic, O¤, Berget, I\* and Næs, T+\*.

¤ Norwegian Knowledge Centre for the Health Services, Pilestredet Park 7, P.O. Box 7004 St. Olavs plass, 0130 Oslo, Norway

\*Nofima AS, Oslovegen 1, P.O. Box 210, N-1431 Ås, Norway

+University of Copenhagen, Faculty of Life Sciences, Department Food Science, Rolighedsvej 30, 1958 Fredriksberg Copenhagen, Denmark

# Abstract

Generalised procrustes analysis and multiple factor analysis are multivariate statistical methods that belong to the family of multiblock methods. Both methods are often used for analysis of data from projective mapping (a.k.a. Napping). In this study, generalised procrustes analysis and multiple factor analysis are compared for a number of simulated and real data sets. The type of data used in this study were (I) random data from Monte Carlo simulations; (II) constructed data that were manipulated according to some specific criteria; (III) real data from nine Napping experiments. Focus will be on similarities of the consensus solutions and the danger of overfitting and overinterpretation. In addition we will consider interpretation of the RV coefficient and individual differences between assessors.

Keywords: projective mapping, Napping, generalised procrustes analysis, GPA, multiple factor analysis, MFA, consumer test, multiblock method, RV coefficient

# Introduction

In recent years, rapid sensory methods have gained a lot of interest in the field of sensory science (Dehlholm et al., 2012, Valentin et al., 2012, Varela and Ares, 2012). Among the advantages of these methods is their simplicity in use, that they can be used by untrained assessors and that the analysis can often be carried out quickly. One of the best-known and most used methods in the category is projective mapping (Risvik et al., 1994), also later known as Napping (Pages, 2005). With this method, a number of individuals (typically between 10 and 100) are asked to place a number of products on a two-dimensional sheet according to how similar or dissimilar they consider the products to be, using their own criteria. Despite being documented to be less precise than descriptive sensory analysis (Valentin et al., 2012), projective mapping has gained much popularity especially within the food industry because of the advantages listed above. It should also be mentioned that the method sometimes can, due to its holistic character, provide additional information (ref) as compared to standard attribute based sensory methods.

By placing products on a sheet, each individual generates a two-dimensional data matrix representing the coordinates of all the placed products. These data need to be analysed with a suitable statistical method in order to extract information about the tested products, which can be utilised for further product development or product optimisation. The two most established methods for analysing projective mapping data are generalised procrustes analysis (GPA) (Gower, 1975) and multiple factor analysis (MFA) (Escofier and Pages, 1994). Even though both GPA and MFA are conceptually very different, both belong to the family of the so-called multiblock methods (Abdi et al., 2013). They provide information about the “consensus” product configuration, which in practical terms represents the “mean” product configuration across all individuals and which gives important insight into the overall perception of the products. Although several other methods, for instance INDSCAL (Carrol and Chang, 1970), STATIS (Schlich, 1996) and the different Tucker methods, (Tucker, 1964) can be envisioned for handling this type of data, it is of interest to compare the two because of their frequent use.

To the authors’ knowledge there exist only one study that in some detail discusses the differences and similarities between the two methods applied to the same set of projective mapping data. This is the study by . Nestrud and Lawless (2008) which reports that both methods have been tested on the same data set and that results were very similar. In that study, GPA and MFA were applied to data that were generated from a single experiment where 13 citrus juices were evaluated by a group of experienced chefs and a group of untrained consumers.

The present study attempts to provide more insight into differences and similarities between results acquired with the methods GPA and MFA in the context of projective mapping. A secondary objective is to discuss the use of the RV coefficient (ref) which is used frequently in the area for comparing data sets and consensus solutions. For these purposes we will use (I) random data in Monte Carlo simulations; (II) constructed data that were manipulated according to some specific criteria; (III) real data from nine Napping experiments. In particular, the following points will be highlighted

* The importance of proper validation of the consensus solution.
* The importance of using simple computer simulations in order to understand differences better.
* The importance of looking at individual differences between assessors for obtaining information about validity and stability.
* The importance of accompanying the RV coefficient by graphical displays of the data.
* The possibility of extending the focus to more than two components.

# 2. Methods

## 2.1 Projective Mapping

Projective mapping is a method where individuals evaluate the overall perception of a number of products and place them on a sheet according to the products’ similarities or dissimilarities (Risvik et al., 1994, Pages, 2005). Placement can be done either by putting products directly on a sheet of paper or by indicating their position on a computer screen. Individuals are instructed to place similar products close to each other using their own criteria or criteria given by the instructor. Other than that, individuals are generally not given further directions. If the placement of the products needs to be refined, the individuals may taste the products again until placement is considered to be satisfactory.

Optionally, individuals may be asked to write down sensory descriptors on the sheet close to the tested products, that best describe each group of products. By doing so, the projective map is turned into an Ultra Flash Profile method as described previously (Perrin et al., 2008, see also Williams and Arnold, 1985 for other situations where free assignment of words is relevant ). In this study, however, focus will be only on the product coordinates derived from the positions of the products on the sheet or on the computer screen (two-dimensional data blocks in form of x- and y-coordinates).

A well know critique regarding projective mapping worth mentioning is that complex multidimensional products may be difficult to place on the two-dimensional sheet since the two dimensions of the sheet may not be enough to distinguish the products properly and may then leave the user with a non-satisfying placement of the products. Recent research (Nestrud and Lawless, 2011), however, refutes this criticism by claiming that important components and configurations could be recovered using MFA and multidimensional scaling. Since the two first components are the dominating ones and also those that are given main attention in the literature, main focus will here be on these two components. We will, however, also discuss briefly the possibility of interpreting more components than two.

## 2.2 General Structure of Projective Mapping Data

Every individual taking part in the projective mapping trial is supposed to place a number of products on a projective mapping sheet resulting in individual data blocks that are of dimension with . Here  represents the number of objects or products tested by the individuals.

## 2.3 Generalised Procrustes Analysis (GPA)

GPA (Gower, 1975, Dijksterhuis, 1996, Gower and Dijksterhuis, 2009) is a multivariate statistical method that is applied for multiple data blocks. The main goal is to acquire a consensus from the blocks after they have undergone Procrustes transformations that reduce individual differences by means of translation, rotation and reflection as well as isotropic scaling. GPA is therefore well suited for analysis of projective mapping data given our goal to find a consensus product configuration across all individuals who take part in the mapping. Note that GPA consists of two steps: (A) Procrustes transformation followed by (B) Principal Component Analysis on the transformed data blocks (optional). Since in our case the consensus is two-dimensional, the PCA only represents a rotation of the original axes found by the Procrustes transformations, which means that the latter step only represents an improved interpretation possibility.

Clearly, there will always be variations in how the individuals place the products on the sheet. The variation between the data blocks comes from different perception of products, and because of the more or less arbitrary ways of using the directions on the mapping sheet. Regarding the former, these are the sensory differences that are relevant for computation of the consensus product configuration. One would, however, like to eliminate the latter since this is generally not product related.

In more detail, the Procrustes transformation (A) itself consists of three steps that can be summarised in the following way: (A.1) translation, meaning that all individual configurations are moved to the middle of the mapping sheet. In statistical terms, this corresponds to a mean centring of the x- and y-coordinates; (A.2) rotation and reflection of individual configurations until they are in best possible agreement with one another (see equation (2)).; (A.3) isotropic scaling, i.e. shrinking or stretching of individual configurations until they are as alike as possible but without changing the relative distances between the products in each configuration. Since the mean, scaling and rotation are related to individual differences of minor value for the interpretation of the Napping data, the Procrustes method is very well suited for the situation. It preserves relative distances between objects (see criterion below), which may be seen as an advantage. Mathematically, the three steps of the Procrustes transformation may be summarised in the following way

 **(1)**

where represents the Procrustes transformation of block . The is the matrix of translation constants (step (A.1)) which is easily handled by simply subtracting the mean. The represents the rotation matrix (step (A.2)) and represents the scalar from isotropic scaling (step (A.3)). Note that is an orthogonal matrix; **.**

Translation can be removed from Equation 1 by centring of each variable first. The and of each data block are then obtained by minimising:

  (2)

where represents the mean or so-called consensus matrix across all transformed blocks. The is of dimension , i.e. exactly the same dimension as the individual data blocks . As a final step is then analysed with PCA (i.e. rotated) where the scores plot represents the final consensus sensory map. The final consensus configuration is denoted **.** Note that the final PCA transform is optional.

Note that since is of dimension , only two principal components (PC) may be extracted from the data. As a consequence, all information in the resulting consensus product configuration , which is also of dimension , will be contained in the space spanned by these two PC’s.

## 2.4 Multiple Factor Analysis (MFA)

There are several ways to describe mathematically how MFA (Escofier and Pages, 1994) works. A thorough review of the alternatives is provided elsewhere (Abdi et al., 2013). To keep this section brief MFA is presented as a ‘simple PCA’ of a concatenated matrix consisting of all the original data blocks.

In MFA the consensus is computed by the following steps

1. SVD of each single block (centred) and dividing each by its first singular value , obtaining

 (3)

1. All are concatenated horizontally, obtaining

 (4)

1. PCA is applied on which results in a consensus product configuration .

Note that more than 2 components can be extracted for visualisation of the consensus product configuration (see also Nestrud and Lawless(2008)). The will thus have dimension where represents the number of PC’s extracted by PCA. Typically in practice one looks at only the two first components which have the highest eigenvalue. However, one can as indicated also investigate component 3 and further. One possible argument for considering more than two components is that assessors may use different criteria for their sensory assessments and that this information may possibly be made visible if more than two components are considered. This aspect is discussed further when analysing the real data sets below.

It is important to note that MFA is essentially a multi-block PCA of concatenated matrices. There are several names for this in the literature, for instance Tucker-1 (Tucker, 1964), CPCA and Sum-PCA (Smilde et al., 2003). Before doing a concatenated PCA it is a reasonable practice to give the individual blocks the same weight to correct for individual differences in the use of the scale. The MFA is just one such possibility, an even more common approach for other types of data is to just divide each block by the square root of the sum of squares (after centring which is always done).

In MFA, the consensus is obtained as a linear combination of the original data sets and is in this sense more flexible than GPA which only accounts for translation, scaling and rotation differences. An important question is whether it is natural and useful to change object distances in this way (which is not done in GPA, see also above).

**2.5. Standardisation of variables**

In multivariate analysis of standard descriptive sensory analysis, each variable is always mean-centred. There is, however, always a discussion of whether one should standardise the variables or not prior to PCA. There are arguments for both strategies and there is no obvious unique solution to it. The results should, however, be interpreted according to which approach is used (Næs et al., 2010). Also for Napping data, the means are subtracted and the same discussion about standardisation can be raised. One can argue that the natural solution is to use mean centred data as they are, since distances in two orthogonal directions are equally important to distances along the same direction. In this paper we will therefore focus on non-standardised data, but for comparison purposes also present some plots for both possibilities in order to evaluate the effect of standardisation.

**2.6. Validation of the methods.**

Since both methods will always give a solution, it is always a good practice to do some type of validation of the results. This means that one should always put some emphasis on testing whether the consensus makes sense in describing the original data.

the most common procedure is probably the permutation test proposed by (). This is based on the The rationale for the criterion is that if the consensus does not describe a substantial part of the original data, it does not represent a good description of the data. The actual testing is done by repeating a large number of permutations and then comparing the observed proportion with the permutation distribution. For an overview of permutation testing we refer to Dijksterhuis and Heiser (2995).

For the MFA, the most used procedures are probably the ones based on bootstrap confidence ellipses around each of the points in the consensus configuration. There are different ways of constructing these ellipses based on bootstrapping either raw data or on projections as described in for instance (ref). If the ellipses overlap a lot, there is reason to question the validity of the consensus solution. Another possibility is to use regular cross-validation as discussed in for instance Martens and Næs (1989). One eliminates one sample at a time, projects the eliminated sample down on the solution obtained by the rest, and calculates the explained validation variance the normal way.

A possibility which can be useful for both approaches is to simply look at the individual differences as plotted in the way described above. If there is large individual variation around the consensus points, there is reason to question the validity of the solution obtained. This will be illustrated below.

## 2.7 RV coefficient

As for the analysis of the data themselves, there are several ways of comparing matrices (Ramsey et al., 1984, Gower and Dijksterhuis, 2009). In the area of napping, however, the RV coefficient (Robert and Escoufier, 1976) has obtained a status as the standard method. For the same reasons as discussed above for the choice of analysis methods, the RV coefficient will be the method to be studied here.

The similarity of two object configurations and , can be measured with the RV coefficient (Robert and Escoufier, 1976} which can be computed as follows:

 (5)

Note that both and are here assumed to be column centred (see also description of GPA and MFA). The RV coefficient is a scalar that varies between 0 and , with 1 corresponding to exact equality. Important properties of the RV coefficient are scale and rotation invariance, which is very convenient when analysing data from projective mapping where two product configurations often have different orientation, centre and span of axes. In general, when computing the RV coefficient for two data matrices and the number of variables in each matrix may be different. In this study, the RV coefficient will be used for comparing consensus configurations with the same number of variables.

Several studies report relatively high RV coefficient values for two data matrices (for instance RV > 0.75) suggesting that there are high levels of repeatability and reproducibility for their respective tasks (Lawless and Glatter, 2010, Kennedy 2010, Vidal et al., 2014). There is, however, an increasing awareness (Ares et al., 2014, Garbez et al., 2014) that one should be careful when interpreting RV values. This is because the value of the RV coefficient depends on the number of objects and variables in and (Smilde et al., 2009) and that it may be subject to a centring effect (Tomic et al., 2013). Another and equally important aspect is that the RV coefficient puts most emphasis on the first principal component. This can be seen clearly if each of the input matrices is substituted by its singular value decompositions as done in for instance Ramsey et al. (1984).The RV coefficient is namely a function of the singular values in such a way that it is clearly dominated by the largest ones. If for instance the first component has a much larger explained variance than the second, an apparent similarity between and as measured by the RV can thus possibly be a result only of similarity along the first component (Ramsey et al., 1984).

## 2.8 Similarity ratio for projected individual data

The main result from GPA and MFA is a consensus product configuration where c represents either GPA or MFA. The validity of the consensus product configuration can in addition to the tools mentioned above be evaluated by measuring how well represents each individual product configuration. This is basically done by “projecting” the individual configurations onto the subspace spanned by the corresponding loadings. For GPA this is done by applying the PCA transform (optional, see above) obtained for the consensus to the rotated and scaled data for each individual. The new individual scores acquired in this way may then be plotted together with the consensus. For MFA the individual scores from each block can be obtained by multiplying the individual napping data with the corresponding loadings (properly scaled, Abdi et al, 2013, see also Xlstat. Again, these can be plotted in the consensus space even though they essentially belong to individual/different subspaces. Here we propose to measure how well a particular individual is represented by the consensus, by computing the similarity ratio : Nevne Leanie.

 (6)

where represents the individuals participating in the projective mapping; represents the method that the ratio is computed for; represents projected scores of an individual and the norm is the Frobenious norm, i.e. the square root of the sum of squares of all elements . Computing the similarity ratio in this way is convenient since it is independent of the scores units (i.e. one can multiply one of the solutions by a constant without changing the value). This makes comparisons across the two methods possible. From Eq. 6 it can be seen that the larger the difference between the projected scores and the consensus product configuration is, the higher is for that particular individual. There is no upper limit for the similarity ratio, but values higher than 1 mean that the differences between the actual individuals and the consensus are larger than the variability itself, which is clearly an indication of no fit/similarity at all. If the projected scores are exactly the same as the consensus scores then will be zero since the nominator in Eq. 6 will be zero. To get a measure of how well the consensus product configuration represents the whole group of individuals one can compute the mean across all individuals’ similarity indices. This is done as follows:

 (7)

The lower the value of , the better are the individuals described by the consensus. Below we will also consider the standard deviation of the values.

## 2.9 Data Analysis Software

Monte Carlo simulations of random data in subsection 3.1 were carried out in a Python programming language environment using the numerical package Numpy (Oliphant 2007). From the Python environment, GPA and MFA functions were called to do the computations on the random projective mapping data. The GPA and MFA functions are part of the FactoMineR package (Le et al. 2008) coded in R and were accessed through PypeR (Xia et al. 2010), which is an interface between the Python and R programming languages.

For the constructed and real world data of subsection 3.2 and 3.3 the commercial XLSTAT software was used for computation of results. Both GPA and MFA are part of the XLSTAT-MX add-on package for market research and sensory analysis. In particular, the Gower implementation of GPA in XLSTAT (version XLSTAT 2013.4.08) was used for analysis.

## 3. Projective Mapping data used in study

Three types of data were used: (I) random data generated with different settings of the number of consumers and the number of samples; (II) constructed structures data that simulate certain simple situations; (III) real data from nine Napping experiments. The focus for the first is an investigation of the overfitting tendency of the methods, i.e. their ability to find a consensus in cases where there is no real underlying structure. The focus of the second is to assess how the two methods react to changes in translation, rotation, scaling and changing of the distances between the products. Focus here will be on consensus plots and individual differences. In the real world data section focus will be on similarity of consensus plots along two components only, but the potential of extracting more components will also be discussed briefly. Focus will be on how RV results correspond to a visual appearance of the configurations and how well the individuals fit to the consensus.

## 3.1 Monte Carlo simulations with random data

The main objective in this part of the study was to investigate the similarity of consensus product configurations and over a large number of simulations. For each fictive individual taking part in the projective mapping trial, random data were generated that fell within a standard projective mapping sheet of size 60 x 40 cm. The random data were generated using a uniform distribution. The amount of random data used in each Monte Carlo simulation depended on the number of individuals taking part in the trial, and the number of products simulated for all blocks . The upper limit of 16 products was chosen based on one of the real world data sets described in section 3.3 that compared 16 products (data set 4 in Table 1). The upper limit of 100 consumers was chosen based on the fact that such a high number is within realistic limits in practical situations (see data set 7 in Table 1 that uses 97 consumers to evaluate the products).

Using all possible pair-wise combinations of number of individuals (k = 20, 60, 100) and number of products (i = 4 … 16) a total of 39 Monte Carlo simulations (3 levels of individuals x 13 levels of products) were carried out. For each of the 39 Monte Carlo simulations 1000 runs were used, i.e. 1000 consensus product configurations and were computed and for each run their similarity was measured using the RV coefficient. The average across the 1000 RV coefficient was then computed and applied as an indicator for general similarity between and . For this calculation the data were standardised which is the default values in the program used (see below). This essentially corresponds to making the Napping sheet square and has for this particular simulation no influence on the conclusions.

### 3.2 Constructed data

The intention behind this part of the study is to show the importance of using simple simulations based on known structures for illustrating properties of methods. The main aim here was to investigate in a controlled setting how GPA and MFA handle individual product configurations from a number of fictive individuals. The individual product configurations were initially identical (as shown in Figure 1) before they were subject to one or more targeted manipulations. These targeted manipulations included off-sets (translation) from the projective mapping sheet centre, different degrees of rotations, reflections and scaling, as well as variation of relative product distances. Except for the last, these are exactly the types of situations GPA was designed for to handle with its Procrustes transformations.

The questions were: (a) how would MFA handle translation, rotation, reflection, scaling and change of relative product distances?; (b) would the MFA consensus configuration be different from that of GPA?; (c) which of the two consensus configurations would provide a better representation of the individuals?

For this purpose, different scenarios were created with manipulated data sets for 8 fictive individuals. To answer question (a), the data in each scenario were manipulated by applying at least one or a combination of the manipulations mentioned above. In this paper, we will put main emphasis on two scenarios described below (constructed data 1 and 2) where a combination of manipulations was applied to the data, but results for individual manipulations will also be mentioned briefly. To answer question (b), the first and second principal components of , of the resulting consensus product configurations were compared with one another. This was done by using the RV coefficient and scatter plots to make a statement regarding their similarity. To answer question (c), the similarity ratios (see description in section 2.6.) across all individuals were computed for GPA and MFA and compared with each other.

In order to avoid numerical computation problems with MFA in XLSTAT (encountered in constructed data set 1 below), 1 % random noise was added to each of the individual configurations after they were manipulated and prior to analysis with GPA and MFA. Our interpretation of this problem in XLSTAT is that convergence problems may occur when axes have the same variance which was the case for one of the standardised examples.

**Constructed data 1: translation, rotation and reflection**

Figure 2 visualises a data set where three types of manipulations were applied to eight individual configurations (starting with the initial product configuration as shown in Figure 1): different degrees of translation, rotation and reflection. Scaling and change of relative distances between products were not applied to the individual configurations.

**Constructed data 2: scaling and changing relative distances between products**

Figure 3 visualises a data set where relative distances between products were changed by stretching and shrinking the individual product configurations in different ways and combinations. Stretching and shrinkage were applied along either axis spanned by product 1, 2 and 3 or the axis spanned by product 2, 4 and 5 or both. Some of the individual configurations were scaled up or down from the initial product configuration (Figure 1). Changes applied to distance between product 1 and 2 were always identical to those of distance between product 2 and 3. Furthermore, changes of the distance between product 2 and 4 were always identical to those of distance between products 4 and 5. This resulted in 8 differently shaped triangles as visualised in Figure 3. Before and after stretching and shrinking all individual product configurations were centred in the middle of the projective map and pointing “south”. Translation, rotation or reflection, were not applied to the individual configurations.

### 3.3 Real World data

In this part of the study, nine Napping data sets from real experiments were analysed. Table 1 provides a short summary of the products tested in each experiment and the number of individuals that participated. It is also indicated if there is a connection to any of the other data sets. Note that data set 1 and 2 were acquired through experiments carried out at Nofima and that the remaining data sets were kindly provided by F. Husson (see Husson, 2013) on his own web site. The tested products are of varying sensory complexity ranging from relative low complexity products like apple and orange juices to relatively high complexity products such as wine. From experience, it is known that complex products generate more variation across consumers and experts than products of rather low sensory complexity. One of the data sets will be considered in detail, while the rest will only be considered for simpler numerical comparisons.

# 4. Results and Discussion

## 4.1 Monte Carlo simulations with random data

The average RV coefficient between consensus configurations and for the random data is shown in Figure 4. Each data point represents the average RV coefficient from 1000 simulations for a specific combination of number of individuals and number of products. As can be seen, the highest average similarity between the scores and are present with a low number of products. For all tested numbers of individuals, i.e. 20, 60 and 100 individuals, RV coefficients are the highest for 4 products and in general decreasing with an increasing number of products. As can be seen the decrease of the average RV coefficient by adding another product to simulations is larger when the number of products is low. Moreover the changes are getting continuously smaller when the number of products increases. Furthermore, it seems that GPA and MFA consensus product configurations are less similar when the number of consumers increases, which is to be expected. Overall, we can conclude that the similarity of the consensus configurations is remarkably high even when there is no such thing as an underlying consensus, in particular for small data sets. These results may indicate an overfitting tendency which means that one should always be cautious when interpreting consensus results, and if possible test the validity of the consensus as was discussed above

## 4.2 Constructed data

For constructed data set 1, which contains translation, reflections and rotations (no scaling, nor change of relative distances between products), GPA and MFA returned almost identical consensus results (Figure 5). This is a case which is ideal for GPA, but as can be seen, MFA practically gives identical results. This indicates that in such a simple setting, the two methods simultaneously handle translation, rotation and reflection very easily and in a quite similar way which of course was to be expected. The same was true when only one of the transforms above was used on the individual configurations (results not shown here).

At this point, it is of interest to take into account individual product configurations from the eight fictive consumers. In Fig. 5 these are plotted on top of the consensus plots in the way described above. The closer an individual product configuration is to the consensus configuration, the better it is represented by the consensus configuration. As can be seen, in both cases, individuals lie very close to the consensus and any real deviations are not present.

The consensus and individual differences results for constructed data set 2 are presented in Figure 6. The consensus product configurations are quite similar. The individual differences are comparable for the two plots, which may be a bit surprising since MFA is not limited to orthogonal transformations. In the figure we also present the results based on the standardised MFA and as can be seen the individual differences now more or less vanish completely. This is natural since individual differences in Figure 3 are essentially corrected for by the standardisation. Still we see that the consensus is comparable to the other two. On the other hand, as also indicated above, the practice of standardisation may be questionable.

For all the situations considered here, only two componets are used for visualisation. This is natural for this type for idealised differences and for visualising the differences and similarities of interest here. For the real data sets to be considered next, the situation is different and some emphasis is also put on components beyond component 2.

All the results based on the graphical displays are supported by the calculations in Table 2. Looking at the table more in detail, one can see that the individual differences are smaller for GPA than for MFA which is natural since GPA by definition tries to minimise individual differences.

## 4.3 Real World Data

It is important to mention that the nine data sets (see Table 1) discussed in this section are of varying degree of complexity and size. The degree of complexity is mainly attributed to the sensory dimensionality of the tested products.

**Validation of the consensus.**

In Table 3 are presented the values (explained variances) from the GPA permutation tests (see above and Wakeling et al., 1992) for the significance of the consensus configuration. While for the “simpler” products, as for example the apple juices in data set 1 and 2, the is relatively high (0.713 and 0.577 respectively), the value for the wine data set (data set 9) is the lowest with =0.304. This is a clear indication that there is far less clear consensus among individuals for the most complex situation considered (wines). The percentiles next to the in Table 3 are computed from permutation tests with 10 000 permutations and indicate at which level the real is compared to the distribution of 10 000 values from the permutation test. Note that for the first eight data sets the consensus configuration is considered highly significant with each of their being at the 100th percentile, which means that none of the from permutations is larger than from the found consensus configurations. The from the wine data (data set 9), on the other hand, indicates that the found consensus configuration might have been a product of chance if level of significance is set to 5%.

**The RV coefficients between the GPA and MFA consensus configurations**

Table 4 shows the RV coefficients between the first two PC's of and for consensus configurations from GPA and MFA for the nine data sets. The RV coefficients in general are relatively high indicating that very often GPA and MFA provide similar consensus configurations. Many of the RV coefficients across the nine data sets are well above 0.9, some of them close to 1. The lowest single RV coefficient is given for data set 9 for the wine products with RV = 0.875. Referring back to the simulation example of random data (average RV for 20 individuals and 10 products is approximately 0.84 vs 0.85 for 18 individuals and 10 products in data set 9) and the fact that the GPA is not significant, it is natural to conclude that the consensus configuration in this case may possibly be a result of chance.

**Visualization of the RV coefficient**

In order to visualise some problems regarding direct interpretation of the RV coefficient, we made a thorough comparison of the GPA and MFA for one of the data sets (data set 5 in Table 1) In this particular case the MFA is based on standardised variables because this shows the danger of over-interpreting the RV even more clearly than for non-standardised variables. The RV coefficients between the two consensus product configurations indicate high similarity (0.874), but there is no doubt when comparing the plots in Figure 7 that conclusions regarding the products may be quite different depending on which statistical method is used for analysis of the projective mapping data. Both consensus configurations separate the products in a very similar manner along the first component. Products 4, 7, 8, 9 and 12 are on one side of the plot, while products 1, 2, 3, 6, 10 and 11 are found on the opposite side along component 1. Product 5 is placed about in the middle in each of the consensus maps. Problems however arise when the placement of the products are compared along component 2. One can see substantial differences, as for example the placing of product 1 and 11. In the GPA consensus, the two products may be considered quite different regarding the second component whereas in the MFA consensus they may be interpreted to be very similar overall. The position of product 9 is another example of where interpretation is obviously very dependent on the choice of statistical method. Moreover, products 4 and 12 have positive scores in GPA and MFA, but products 3, 10 and 11 have positive and negative scores in GPA and MFA, respectively. If a user should decide to compare the consensuses from GPA and MFA, he or she may face a dilemma of how to properly interpret the findings. In other words, direct interpretation of the RV coefficient is not always obvious. This is illustrated further in Figure 8 where the first two components for both plots are plotted against each other. As can be seen, the correlation along the first axis is very large while it is almost equal to 0 for the second component.

The reason for these results is that the RV coefficient puts more emphasis on the first component than on the second. This indicates clearly, that although the RV coefficient may be a useful method for matrix comparisons, one should not trust it blindly if the differences in explained variance for the interpreted axes are large. One should, whenever possible, support the RV calculations by visual inspection. Note, however, that a high RV coefficient is a clear indication about similarity, but not necessarily of all the components considered. In the present case, for instance, all the similarity that led to a large RV value was found in the first component.

**Differences between individuals.**

The next important step is to investigate how well individuals are represented by the consensus product configurations from GPA and MFA. As in section 4.2 similarity ratios will be used to investigate which one represents individuals in the best manner. The right part of Table 4 shows and as well as standard deviation of for each of the nine data sets. It can be seen that for all data sets the values are lower for the GPA than it is for MFA. It seems that the GPA consensus product configuration in general provides the most similar individual configurations according to the projection methods discussed in the methods section. This is in good correspondence with the focus of GPA, namely to minimise individual differences..

Figure 9 shows an example of what the looks like for both methods for one of the data set, data set 5. The plot clearly illustrates that GPA finds a consensus product configuration that represents all individuals well. The of GPA varies very little compared to which means that for GPA the differences or distances between the consensus product configuration and individual product configurations are relatively small. For one can observe that the individual product configurations of assessors 4 and 6 are relatively different from the MFA consensus product configuration. In general we also recommend that one looks at the individual differences graphically as done in Figure 6.

**More than two components in the MFA solution**

As was stated above, the MFA consensus solution may contain several components. This is illustrated in Figure 10 for all the real data sets. As can be seen, the explained variance using two components varies from 40% to about 80%. For data set 5 the first two components of the MFA solution describe about 60% of the total variance with the third component describing almost as much as the second component. Looking further at how the different assessors relate to the different axes should be done using for instance cluster analysis and external sensory and assessor data. This is an important topic which has been given relatively little attention in the literature. More research is needed involving also how GPA could be modified for providing information about more than two underlying componentss.

# 5. Conclusion

Change this one. The results of this paper have shown that GPA and MFA can give very similar results as measured by RV coefficient even for random data. In other words, the methods can find quite similar structures even when there is no real structure underlying the data. An implication of this is that the methods are potentially sensitive to overfitting and a proper validation of the consensus should therefore be done either by using a permutation test or cross-validation.

Constructed data with clear structure have revealed that both methods are able to find the underlying consensus structure in the data in the presence of translation, rotation and reflection of the different assessors. Scaling and moderate changes of relative distance between products in individual configurations also gave similar consensus configurations.

A comparison of the RV coefficient with graphical illustration of the MFA and GAP solutions clearly showed that even though the RV coefficient is quite high (even higher than 0.85), the differences between samples can be quite large along the second component. The reason for this is that the RV coefficient gives most emphasis to the direction with the largest eigenvalue. Our recommendation here is that one computes both MFA and GPA and looks at the plots before putting too much emphasis on the interpretation of the components.

The relative size of the individual differences between individuals is generally smaller for GPA than for MFA. This may look a bit surprising since MFA is based on a non-restricted linear transform, but may indicate towards an advantage of the GPA, possibly accompanied with the idea presented above for how to extend GPA to more than 2 components. It should be noted, however, that the MFA results reported are based on the calculation and plotting procedure recommended in Abdi et al. (2013). Other approaches could be envisioned. The results from the MFA clearly show that a large portion of the variation is left in component 3 and further. For a complete analysis of the data, these components should also be considered. This may point towards the use of cluster analysis for improved interpretation (see e.g. Dahl and Næs, 2004), but this aspect is beyond the scope of the present paper.

# Acknowledgements

The author would like to thank Francois Husson for kindly providing Napping data on his web page at Agrocampus Ouest. We would also like to thank FFL and Research Council of Norway for financial support through the project RapidCheck. We would like to thank the referees for useful comments.

# References

Abdi, H. and Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews:Computational Statistics,* 2, 433–349.

Abdi, H., Williams, L. J., and Valentin, D. (2013). Multiple factor analysis: principal component analysis for multitable and multiblock data sets. *Wiley Interdisciplinary Reviews: Computational* *Statistics*, 5, 149–179.

Ares, G., Antúnez, L., Giménez, A., Roigard, C.M., Pineau, B., Hunter, D.C., and Jaeger, S.R. (2014). Further investigations into the reproducibility of check-all-that-apply (CATA) questions for sensory product characterization elicited by consumers. *Food Quality and Preference*, 36, 111-121.

Carroll, J.D. and Chang, J.J. (1970). Analysis of individual differences in multi-dimensional scaling via n-way generalization of Eckhart-Yong decomposition. Psychometrika, 35, 283-319.

Dahl, T. and Næs, T. (2004). Outlier and group detection in sensory analysis using hierarchical clustering and the Procrustes distance. Food Quality and Preference. 15, 3, 195-208.

Dehlholm, C., Brockhoff, P. B., Meinert, L., Aaslyng, M. D., and Bredie, W. L. (2012). Rapid descriptive sensory methods - comparison of free multiple sorting, partial napping, napping,

flash profiling and conventional profiling. *Food Quality and Preference*, 26, 267–277.

Dijksterhuis, G.B. and Heiser, W.J (1995). The role of permutation tests in exploratory multivariate data analysis. Food Quality and preference, 6, 263-270.

Dijksterhuis, G. B. (1996). Procrustes Analysis in Sensory Research*.* . In (Næs, T. and Risvis, E.ets.). Multivariate analysis of data from sensory science. Elsevier, Amsterdam.

Escofier, B. and Pagès, J. (1994). Multiple factor analysis. *Computational Statistics & Data Analysis*, 18, 121–140.

Garbez, M., Galopin, G., Sigogne, M., Favre, P., Demontes-Mainard, S., Symoneaux, R. (2014). Assessing the visual aspect of rotating virtual rose bushes by a labelled sorting task. *Food Quality and Preference*, DOI: 10.1016/j.foodqual.2014.06.008.

Gower, J. C. (75). Generalized procrustes analysis. *Psychometrika*, 40, 33–51.

Gower, J.C. and Dijksterhuis, G.B. (2009). Procrustes problems. Oxford University Press, UK.

Husson, F. (2013). Data sets. http://math.agrocampus-ouest.fr/infoglueDeliverLive/

membres/Francois.Husson/data.sets. Accessed: 2013-12-07.

Kennedy, J. (2010). Evaluation of replicated projective mapping of granola bars. *Journal of Sensory Studies*. 25, 672-684

King, B. and Arents, P. (1991). A statistical test of consensus obtained from Generalized Procrustes Analysis of sensory data. *Journal of Sensory Studies*, 6, 37–48.

Lawless, H.T., and Glatter, S. (1990). Consistency of multidimensional scaling models derived from odor sorting. *Journal of Sensory Studies*, 5, 217-230.

Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software*, 25, 1–18.

Martens, H. and Næs, T. (1989). *Multivariate Calibration*. John Wiley & Sons Ltd, Chichester.

Morand, E. and Pagès, J. (2006). Procrustes multiple factor analysis to analyse the overall perception of food products. *Food Quality and Preference*, 17, 36–42.

Næs, T., Brockhoff, P. B., and Tomic, O. (2010). *Statistics for Sensory and Consumer Science*. Wiley, Chichester.

Nestrud, M. and Lawless, H. (2011). Recovery of subsampled dimensions and configurations

derived from napping data by MFA and MDS. *Attention, Perception and Psychophysics*, 73, 1266–1278.

Nestrud, M. A. and Lawless, H. T. (2008). Perceptual mapping of citrus juices using projective mapping and profiling data from culinary professionals and consumers. *Food Quality and* *Preference*, 19, 431–438.

Oliphant, T. E. (2007). Python for scientific computing. *Computing in Science and Engineering*, 9, 10–20.

Pagès, J. (2005). Collection and analysis of perceived product inter-distances using multiple

factor analysis: Application to the study of 10 white wines from the Loire valley. *Food Quality*

*and Preference*, 16, 642–649.

Perrin, L., Symoneaux, R., Maitre, I., Asselin, C., Jourjon, F., and Pagès, J. (2008). Comparison of three sensory methods for use with the napping procedure: Case of ten wines from loire valley. *Food Quality and Preference*, 19, 1–11.

Ramsey, J.O. ten Berge, J. Styan, G. P. H. (1984). Matrix correlation. Psychometrika. 49, 403-423.

Risvik, E., McEwan, J. A., Colwill, J. S., Rogers, R., and Lyon, D. H. (1994). Projective mapping: A tool for sensory analysis and consumer research. *Food Quality and Preference*, 5, 263–269.

Risvik, E., McEwan, J. A., and Rødbotten, M. (1997). Evaluation of sensory profiling and projective mapping data. *Food Quality and Preference*, 8, 63–71.

Robert, P. and Escoufier, Y. (1976). A unifying tool for linearmultivariate statistical methods: The RV-coefficient. *Applied Statstics*, 25, 257–265.

Schlich. P. (1996). Defining and validating assessor compromises about product distances and attribute correlations. In (Næs, T. and Risvis, E.ets.). Multivariate analysis of data from sensory science. Elsevier, Amsterdam.

Smilde, A.K., Westerhuis, J.A. and de Jong, S. (2003). A framework for sequemtial multiblock components methods. J. Chemometrics, 17, 323-337.

Smilde, A. K., Kiers, H. A. L., Bijlsma, S., Rubingh, C. N., and van Erk, M. J. (2009). Matrix correlations for high-dimensional data: the modified RV-coefficient. *Bioinformatics*, 25, 401–405.

Tomic, O., Forde, C., Delahunty, C., and Næs, T. (2013). Performance indices in descriptive sensory analysis - a complimentary screening tool for assessor and panel performance. *Food* *Quality and Preference*, 28, 122–133.

Tucker, L.R.(1964). The extension of factor analysis in three-dimensional matrices, In Frederiksen, N. and Gulliksen, H. eds.). Contributions to mathematical psychology, New York, Holt, Rinehart and Winston. Pages

Valentin, D., Chollet, S., Lelièvre,M., and Abdi, H. (2012). Quick and dirty but still pretty good: a review of new descriptive methods in food science. *International Journal of Food Science and* *Technology*, 47, 1563–1578.

Varela, P. and Ares, G. (2012). Sensory profiling, the blurred line between sensory and consumer science. A review of novel methods for product characterization. *Food Research International*, 48, 893–908.

Vidal, L., Cadena, R. S., Correa, S., Ábalos, R. A., Gómez, G., Giménez, A., et al. (2014). Assessment of global and individual reproducibility of projective mapping with consumers. *Journal of Sensory Studies*, 29, 74–87.

Wakeling, I., Raats, M., and MacFie, H. M. (1992). A new significance test for consensus in generalized Procrustes analysis. *Journal of Sensory Studies*, 7, 91–96.

Williams, A.A. and Arnold, G.M. (1985). A comparison of the aromas of six coffees characterised by coventional profiling, free-choice profiling and similarity scaling methods. J. Food Science and Agriculture. 36, 204-214.

Xia, X.-Q.,McClelland, M., andWang, Y. (2010). PypeR, a Python package for using R in Python. *Journal of Statistical Software*, 35, 1–8.

# Tables

**Table 1:** Overview over the real world data sets used in this study.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data set number | Product type | Number of products | Type of individuals | Number of individuals | Remarks |
| 1 | Apple juices | 8 | Students | 16 | Same products as in data set 2 |
| 2 | Apple juices | 8 | Trained sensory panel | 11 | Same products as in data set 1 |
| 3 | Biscuits | 8 | Consumers | 18 |  |
| 4 | Cocktails | 16 | Consumers | 10 |  |
| 5 | Orange juices | 12 | Consumers | 20 | tested in 2005/06; same brands as in data set 6 |
| 6 | Orange juices | 12 | Consumers | 28 | tested in 2006/07; same brands as in data set 5 |
| 7 | Perfumes | 12 | Consumers | 97 | same products as in data set 8 |
| 8 | Perfumes | 12 | Students | 23 | same products as in data set 7 |
| 9 | Wines | 10 | Consumers | 18 |  |

**Table 2:** Numerical results for each of the two constructed data scenarios as described in section 3.2. The RV coefficients refer to the first two PC’s of consensus configurations from GPA and MFA in each scenario. The similarity ratios *SRk*,*c* represent differences between individuals and the consensus.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | *GPA* | *MFA* |
|  | RV | *SRmean,c* | *std(SRk,c)* | *SRmean,c* | *std(SRk,c)* |
| Constructed data 1: translation, rotation and reflection |  1.000 | 5.78E-05 | 1.26E-05 | 1.91E-04 | 1.41E-04 |
| Constructed data 2:changing relative distances between products |  1.000 | 0.051 | 0.047 | 0.160 | 0.095 |

**Table 3:** Overview over the *Rc* values (i.e. the explained variances of the consensus) for GPA for each of the nine real world data sets and their percentiles from 10000 permutations. As can be seen, for all data set except one (data set 9) the consensus is highly significant at 5% level.

|  |  |  |
| --- | --- | --- |
| **Data set** | **Rc** | **Quantile** |
| 1 | 0.713 | 100.000 |
| 2 | 0.577 | 100.000 |
| 3 | 0.527 | 100.000 |
| 4 | 0.524 | 100.000 |
| 5 | 0.492 | 100.000 |
| 6 | 0.427 | 100.000 |
| 7 | 0.335 | 100.000 |
| 8 | 0.326 | 100.000 |
| 9 | 0.304 | 91.210 |

**Table 4:** RV coefficients for consensus configurations for the first two PC’s from GPA and MFA and summaries of computations of the similarity ratios *SRk*,*c.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | GPA |  | MFA |  |
| **Data set** | **RV** | *SRmean,c* | *std(SRk,c)* | *SRmean,c* | *std(SRk,c)* |
| **Data set 1: Apple juices** | 0.995 | 0.359 | 0.128 | 0.604 | 0.387 |
| **Data set 2: apple juices** | 0.985 | 0.592 | 0.109 | 0.936 | 0.387 |
| **Data set 3: biscuits** | 0.986 | 0.712 | 0.092 | 0.897 | 0.395 |
| **Data set 4: cocktails** | 0.990 | 0.751 | 0.046 | 1.049 | 0.720 |
| **Data set 5: orange juices** | 0.961 | 0.888 | 0.052 | 1.286 | 0.840 |
| **Data set 6: orange juices** | 0.975 | 1.165 | 0.071 | 1.356 | 0.569 |
| **Data set 7: perfumes** | 0.970 | 1.676 | 0.357 | 2.092 | 0.964 |
| **Data set 8: perfumes** | 0.939 | 1.931 | 0.440 | 2.068 | 0.463 |
| **Data set 9: wines** | 0.896 | 2.111 | 0.576 | 2.317 | 0.889 |



Figure 1: The plot shows a projective mapping sheet and a constructed individual product configuration with five products numbered 1 to 5. This individual product configuration was used as a starting point for all fictive individuals prior to targeted manipulation. This initial configuration represents a triangle shape and is centred in the middle of the projective map (600 mm x 400 mm) pointing “north”. Product 4 is located exactly in the middle of the projective map. The axis formed by products 1, 2 and 3 is orthogonal to the axis formed by products 2, 4 and 5.



Figure 2: The plot shows schematically how the individual product configurations from the 8 fictive individuals (in plot abbreviated with FI) are placed in relation to one another. The real configurations had different centres, but the centres used in the plot are different and used for giving a better visualisation.. If the real placement coordinates were used the individual product configurations would overlap to a great extent.



**Figure 3:** Constructed data set 2: product configurations of 8 fictive individuals (abbreviated with FI) are shown. No translation, rotation or reflection applied was applied to the data, only change of relative distances between the products. For each individual configuration product 4 was located on the centre of its mapping sheet.



Figure 4: Each data point displays average RV values across 1000 simulations of specific consumer-product combinations. The x-axis represents different number of products and the three lines represent different number of individuals.





Figure 5: Constructed data set 1. PCA scores , of the consensus product configurations acquired with GPA and MFA and the respective individual “projected” scores and , of the projected individual product configuration. The raw data for these results are shown in Figure 2.







Figure 6: Constructed data set 2. Consensus and individual projections for GPA, MFA and MFA based on standardised variables. The raw data of these results are shown in Figure 3.





**Figure 7:** PCA scores and (MFA standardised) of consensus product configurations from GPA and MFA for real world data set 5. RV=0.874.



Figure 8. Plot of component 1 for GPA and MFA (standardised) and component 2 for GPA and MFA (standardised). The first component is indicated by blue crosses and the second axis by green circles.



Figure 9: Individual *SRk*,*c* for for GPA and MFA for real world data set 5 (orange juice).



**Figure 10:** Explained variances for the first five dimensions in MFA models for the real world data sets as shown in Table. 1