Talanta 83 (2011) 1239–1246

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

# Relating gas chromatographic profiles to sensory measurements describing the end products of the Maillard reaction

I. Stanimirova[a], C. Boucon[b], B. Walczak[a,*]

[a] *Department of Chemometrics, Institute of Chemistry, The University of Silesia, 9 Szkolna Street, 40-006 Katowice, Poland*
[b] *Unilever R&D Vlaardingen, Olivier van Noortlaan 120, 3130 AC Vlaardingen, The Netherlands*

ARTICLE INFO

*Article history:*
Available online 13 October 2010

*Keywords:*
Chemometrics
Multiblock methods
Data fusion
Sensory analysis
Product optimisation

ABSTRACT

Often in analytical practice, a set of samples is described by different types of measurements in the hope that a comprehensive characterisation of samples will provide a more complete picture and will help in determining the similarities among samples. The main focus is then on how to combine the information described by different measurement variables and how to analyse it simultaneously. In other words, the main goal is to find a common representation of samples that emphasises the individual and common properties of the different blocks of variables. Several methods can be adopted for the simultaneous analysis of multiblock data with a common object mode. These are: consensus principal component analysis (CPCA), SUM-PCA, multiple factor analysis (MFA) and structuration des tableaux à trois indices de la statistique (STATIS).In this article we present a comparison of the performances of these methods for data describing the chemistry and sensory profiles of the Maillard reaction products. The aroma compounds formed during the reaction of thermal heating between one or two selected amino acids and one or two reducing sugars have been analysed by head space gas chromatography and the intensity and nature of the odour of the resulting products has been evaluated according to selected descriptors by a panel of sensory experts.The results showed that using the information of the chromatographic and sensory data in conjunction enhanced the interpretability of the data. SUM-PCA and more specifically multiple factor analysis, MFA, allowed for a detailed study of the similarities of mixtures in terms of reaction products and sensory profiles.

## 1. Introduction

Nowadays, a primary issue for every food company in meeting consumers' requirements is product optimisation. The traditional approach to quality control used by chemists is to analyse the chemical composition of food samples. Employing separation methods such as liquid chromatography (LC) or gas chromatography (GC), which are suitable for the analysis of complex mixtures, the components of the analysed product are identified and quantified. To understand the importance of the chemical components analysed for the sensory preferences of the consumers, the final product is also evaluated according to a number of relevant sensory descriptors of smell and/or taste. The relationship between the physico-chemical composition of a product and its sensory properties can be very complex. Often substances present in very low concentrations in a food mixture can give rise to very intense aromas and strong flavours.

Perception depends on the individual tasting the product, some people being sensitive to certain components and others very insensitive. In addition, the context in which a product is consumed, previous experience with similar products and expectation will also influence the perception and ultimately the preference of consumers. Therefore, experts trained in judging a number of carefully selected descriptors of flavour and smell usually evaluate the final product. The use of a specially trained panel is quite expensive, but to date has been an irreplaceable approach for the sensory description of unknown samples. Electronic systems of chemical sensors like an electronic nose or tongue, which mimic human perception, are used for routine sensory control of products of a known quality [1–3].

The purpose of this work is to show that the use of information about the composition of a product and sensory measurements in conjunction can provide a more complete picture for the product or system being studied. Specifically, the goal is to reveal the (dis)similarities of a set of mixture samples characterised by at least two sets of variables, e.g. chromatographic profiles and sensory measurements and to explore the relationship among the groups of different variables. In other words, we seek a well-defined common representation of samples which emphasises the individual

---

* Corresponding author. Tel.: +48 32 359 2115; fax: +48 32 259 9978.
*E-mail address:* beata@us.edu.pl (B. Walczak).

0039-9140/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2010.09.018

and common properties of the blocks of variables. To find such a compromise of different blocks of variables measured on the same set of samples, several approaches can be used. Here, we will illustrate the performance of consensus principal component analysis, CPCA [4], SUM-PCA [4], multiple factor analysis, MFA [5–7] and structuration des tableaux à trois indices de la statistique, STATIS [8–10] for a dataset describing the products of the Maillard reaction. The products of the chemical reaction between selected amino acids and selected reducing sugars during thermal heating have been analysed by head space gas chromatography and the intensity and nature of the odour of the resulting products have been evaluated by experts in sensory analysis according to carefully selected descriptors.

Several studies [11,12] were carried out on the dataset chosen to be used in this work. Clustering methods such as the Neural-Gas network, Generative Topographic Mapping, Bottleneck Neural Networks, the Sammon's mapping and the Kohonen self-organising maps were all performed using the chromatographic data only. In this study, we offer a concept for the simultaneous analysis of chromatographic profiles and sensory measurements using multiblock methods.

Depending on the way in which the common components are obtained from all of the blocks, one can distinguish between sequential [4,13] and simultaneous methods. A general overview of simultaneous methods for multiblock analysis has recently been described in the literature [14]. Even though CPCA and SUM-PCA initially belong to the first category of methods, the same results can be obtained in a simultaneous way. These particular methods were selected for use in this study because of the different ways in which they handle the variation and size differences among blocks of variables. When combined with an appropriate preprocessing procedure to correct for scale differences among variables, these methods may enhance the interpretability of the data being studied and it is our intention to illustrate this using the Maillard reaction data [11,12].

In the theoretical Section 2, the properties of the chemometric methods for multiblock analysis are described. Some relevant technical details concerning the Maillard reaction, experimental design and analytical method used are discussed in Section 3 of this article. In Section 4, a detailed description of the strategy for data analysis and the results obtained from different methods are presented and discussed. Finally, the concluding remarks follow in Section 5.

## 2. Theory

As was already mentioned, we focus mainly on the methods that are used for the analysis of multiblock data where a set of samples is described by several sets of variables. The $K$ blocks of variables can be merged in one matrix $\mathbf{X} = [\mathbf{X}_1\ \mathbf{X}_2\ \ldots\ \mathbf{X}_K]$ so that each row is the $i$th analysed sample ($i = 1, 2, \ldots, I$) characterised by $J_1 + J_2 + \cdots + J_K$ variables. Each set of variables measured for $I$ samples has the dimensions $I \times J_k$. The index $k = 1, 2, \ldots, K$ shows the number of blocks. To visualise the distribution of samples in a low dimensional space spanned by a few uncorrelated factors, principal component analysis, PCA, can be used on the concatenated $\mathbf{X}$. The model can be presented in the following way:

$$\mathbf{X} = [\,\mathbf{X}_1\quad \mathbf{X}_2\quad \ldots\quad \mathbf{X}_K\,] = \mathbf{T}[\,\mathbf{P}_1{}^T\quad \mathbf{P}_2{}^T\quad \ldots\quad \mathbf{P}_K{}^T\,] + [\,\mathbf{E}_1\quad \mathbf{E}_2\quad \ldots\quad \mathbf{E}_K\,] \tag{1}$$

In this model, $\mathbf{T}$ is a scores matrix of the dimension $I \times F$, the $F$ columns of which are the principal components obtained as the weighted sum of the original variables by maximising the data variance. The scores matrix $\mathbf{T}$ is related to the common objects mode. The elements of matrices $\mathbf{P}_1$ ($J_1 \times F$), $\mathbf{P}_2$ ($J_2 \times F$) and $\mathbf{P}_K$ ($J_K \times F$) are called loadings and give information about the impact of each block variable on the common principal components constructed. The concatenated matrix $\mathbf{E} = [\mathbf{E}_1\ \mathbf{E}_2\ \ldots\ \mathbf{E}_K]$ of the dimensions $I \times J_1 + J_2 + \cdots + J_K$ is called a residuals matrix and contains the part of the data variance not explained by the model with a definite number of components.

In a more general way, the model presented in Eq. (1) can be expressed by

$$[\,q_1\mathbf{X}_1\quad q_2\mathbf{X}_2\quad \ldots\quad q_K\mathbf{X}_K\,] = \mathbf{T}[\,\mathbf{P}_1{}^T\quad \mathbf{P}_2{}^T\quad \ldots\quad \mathbf{P}_K{}^T\,] + [\,\mathbf{E}_1\quad \mathbf{E}_2\quad \ldots\quad \mathbf{E}_K\,] \tag{2}$$

The difference is that each variable within a block is now weighted by a specific weight, $q_1, q_2, \ldots, q_K$ accounting for variation differences among the blocks. All variables of a given block are weighted with the same weight. In general, the weighting scheme is different depending on the particular method used. Accordingly, each type of weighting scheme puts different weights on the blocks of variables indicating the differences in the importance of the block variables in the common structure obtained from various methods. In the next paragraphs, a brief description of the weighting schemes used by different methods will be given along with their specific properties.

### 2.1. Consensus-principal component analysis, CPCA

Consensus-PCA [4] is in fact principal component analysis applied on a concatenated matrix $\mathbf{X}$. In the original NIPALS-like algorithm of the method, the so-called super score vectors $\mathbf{T}$ are obtained sequentially, one by one in an iterative way. The block loadings are then found by projecting the original block variables onto the score vectors. The decomposition follows the model described by equation 1 which is equivalent to the model presented by Eq. (2) with weights $q_k = 1$. It can also be performed using the singular value decomposition, SVD, algorithm on the concatenated $\mathbf{X}$. Because the method does not account for variation and size differences among the blocks of variables, the group with the largest number of variables will have a predominant influence in the analysis. This method is preferred when the variables of different blocks are expected to give different information.

### 2.2. SUM-PCA

SUM-PCA [4] can also be seen as PCA applied to data which must be preprocessed so that the sum of the squares of each block, $SS_k$, of variables equals one. This gives a comparable weight $(1/\sqrt{SS_k})$ to each bock emphasising the variables of the smaller blocks. The scaling procedure is equivalent to assigning weights to the data and as a consequence the variables of the larger block will be down-weighted more in comparison with the variables belonging to the smaller blocks. To eliminate the differences in the units of variables, autoscaling can be performed beforehand. It has been shown in Ref. [4] that another multiblock method, CPCA-W, gives exactly the same results as SUM-PCA.

### 2.3. Multiple factor analysis, MFA

MFA [5–7] is a factor analysis in which the variables of each group are weighted so as to correct for individual matrix size, while ensuring that none of the variables have too profound an effect on the first factor in the global solution of MFA analysis. The weights $q_k$ are defined as the inverse of the first singular value or the inverse of the square root of the first eigenvalue of $\mathbf{X}_k{}^T\mathbf{X}_k$ block. Thanks to this specific weighting scheme, MFA highlights the similarity of the information content of different groups of variables.

### 2.4. STATIS

Like the other methods described earlier, STATIS [8–10] can be viewed as PCA of the concatenated data where the variance of each data block is weighted. The weights reflect the similarity between each individual block of variables and a compromise defined as the weighted sum of block cross-product matrices, $\sum_{k=1}^{K} q_k \boldsymbol{X}_k \boldsymbol{X}_k^T$. The block that is the least similar to the compromise has a weight close to zero. The weights are determined as normalized elements (their sum is equal to one) of the first eigenvector of the so-called RV [15] matrix which holds the RV coefficients. The RV coefficient is a measure of closeness between any two cross-product matrices. It is positive and its value varies in the range of [0,1]. Two cross-product matrices are more similar if their respective RV coefficient is closer to one. As a consequence of the weighting scheme in STATIS, the block of variables which shares more information with the other blocks of variables is highlighted. This holds true for more than two blocks of variables ($K > 2$). Without correction for block sizes more weight will be put on the matrix with the larger sum of squares.

## 3. Experimental

### 3.1. Maillard reaction

The term Maillard reaction, or nonenzymatic browning, was coined in honour of its inventor, Louis-Camille Maillard (1913). It is one of the major reactions responsible for the formation of aromas and colours in food products where thermal processing has been applied. As is shown in Fig. 1, it is a very complex series of reactions which involves the presence of the carbonyl group of a reducing sugar and the free amino group of an amino acid or protein. Numerous factors such as temperature, time, composition of the system, water activity, pH and sulphur dioxide are known to affect the rate and products of the reaction. Due to the large number of reactants that can combine in many different ways, it generates a very large variety of compounds with desirable as well as undesirable sensory characteristics. It may also result in a reduction in nutritional value, the development of antioxidant properties, and the formation of potentially toxic compounds. It also has some physiological ramifications [16]. Therefore, the Maillard reaction has been studied widely by scientists from different fields. The food industry is particularly interested in it as it can provide a way to control and improve food quality by "natural" means and thus fulfil consumers demands for tasty and natural products [17,18].

### 3.2. Material and methods

*Chemicals*. The following compounds were obtained commercially: D-glucose, D-fructose, α-lactose, L-rhamnose, D-xylose, maltose, (DL)-alanine, glycine, L-threonine, (DL)-asparagine, L-glutamine, L-lysine.HCl, L-arginine.HCl, L-cysteine, L-glutamic acid, L-proline, L-methionine (Acros Organics, Geel, Belgium), glycerol, dichloromethane, 3,5-trimethoxybenzene and p-xylene (Merck, Darmstadt, Germany).

*Sample preparation*. Samples were prepared by mixing sugars and amino acids in a 50 mL glycerol buffer solution at pH 3. The reaction mixture was heated in a metal heating block at 120 °C for 60 min. After cooling at −40 °C for 2 minutes, 10 mL of water was added and samples were vortexed briefly

*Isolation of the volatiles*. On the one hand, 10 g of the samples were collected for sensory evaluation. On the other hand, 4 mL of dichloromethane containing two internal standards (1,3,5-trimethoxybenzene and p-xylene) were added to the reminder of the reaction mixture for GC analysis. The samples were then centrifuged again for 90 sec. After separation, the dichloromethane layered was pipetted, filtered and transferred into a GC vial

*Gas chromatography (GC)*. All analyses were performed on a Hewlett-Packard 6890 gas chromatograph coupled with a Flame Ionisation Detector (FID) using a HP ChemStation software. An aliquot of 0.1 μL was injected in the splitless mode onto a CP-WAX 52-CB capillary column (10 m × 100 μm i.d. × 0.2 μm film thickness). The oven temperature was set to 50 °C for 1 min, then raised to 240 °C with a heating rate of 540 °C/min, and held for 1 min. Finally the oven temperature was set to 160 °C. A total of 159 volatile compounds were identified and quantified.

### 3.3. Data description

The data set presented in this article consists of 576 chemical mixtures prepared so that they contain one or two out of six reducing sugars (fructose, glucose, lactose, maltose, rhamnose and xylose) and one or two out of 11 amino acids (alanine, asparagine,

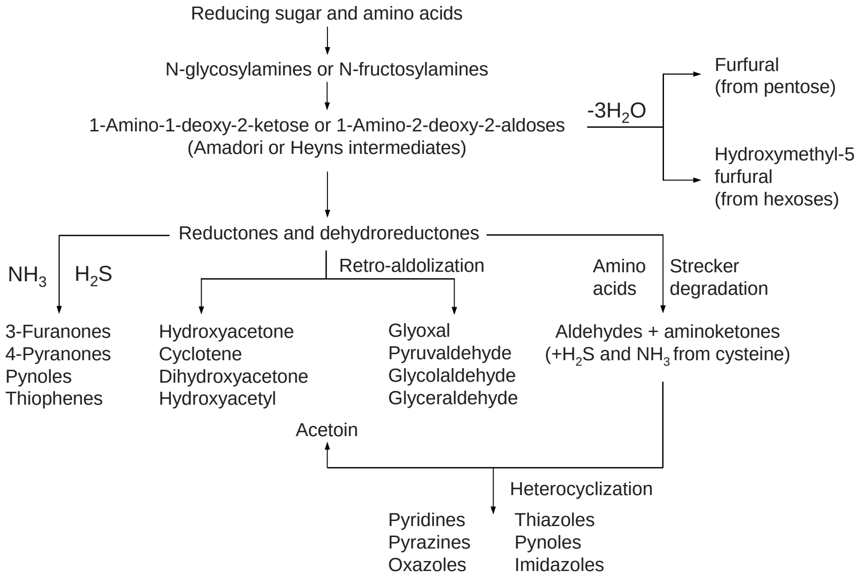

**Fig. 1.** Formation of the Maillard compounds in food products.

threonine, arginine, cysteine, glutamine, glutamate, glycine, lysine, methionine and proline).

Each of the 576 mixtures was also evaluated using 10 descriptors of smell (overall intensity, sulphur, meaty, caramel, burn, nutty, popcorn, jammy, potato and aldehyde) by a small expert panel. For each sample and each descriptor, each expert gave a score in the range [0,4]. A score of 0 indicated that the smell is not present in the samples being considered, while a score of 4 indicated the presence of a highly intense smell. The average scores of the panel were then calculated and are reported in the sensory block. The final augmented data $\mathbf{X}$ is of the dimension**s** $576 \times 169$ containing two blocks of variables ($K = 2$), gas chromatographic data, $\mathbf{X}_1$, of the dimensions $576 \times 159$ and sensory data, $\mathbf{X}_2$, of the dimensions $576 \times 10$.

All calculations using in-house implemented routines were performed with MATLAB 7.0 (R14) on a personal computer (Intel(R) Pentium(R) M, 1.60 GHz with 2 Gb RAM) using Microsoft Windows XP (service pack 2) operating system.

## 4. Results and discussion

Before describing the multiblock analysis of the data, it is useful to look at the analysis of gas chromatographic and sensory data separately. The variables of both sets were subjected to an autoscaling procedure. Autoscaling is a combination of column centering and scaling to a unit standard deviation which gives the variables the same importance in the analysis. This procedure will be used with all methods of multiblock analysis to enable a better comparison.

In order to investigate what amount of information explained by chromatographic data overlaps with the information in the sensory measurements, the RV [15] coefficient can be calculated. The RV coefficient is a rotation invariant measure of the similarity between two cross-product matrices that has values in the range of [0,1]. As was already mentioned, the closer the RV coefficient is to one, the more similar the two cross-product matrices are. In this case, the RV coefficient is used to measure the degree of correlation of the two autoscaled data sets, sensory and chromatographic data, respectively. Recently, a modified version of the RV correlation coefficient has been introduced [19]. The modified RV coefficient eliminates the possibility of obtaining a relatively high RV coefficient when a real underlying relationship between the two sets of variables is not expected. Compared to the original RV, the modified RV has values in the range of −1 and 1 and can be interpreted as the Pearson's correlation coefficient. The modified RV for the two autoscaled sets is equal to 0.30 indicating a relatively low degree of overlap. This RV value also shows that both datasets contain some common information, but that they also have unique parts. From an experimental point of view, this observation indicates that probably not all of the compounds analysed by GC are associated with the development of the specific smells the intensity of which was evaluated by the experts in sensory analysis. Accordingly, one may also expect that somewhat different results might be obtained from the various methods for simultaneous multiblock analysis.

### 4.1. Analysis of the individual GC and sensory data

Another way to look at the data sets separately is to investigate the essential correlation between the blocks after reducing the level of noise in each of them using PCA, and comparing the variance explained by the number of principal components. Again the RV coefficient can be used to see the similarity of each two score spaces defined for a given number of PCs. Table 1 summarises the results.

**Table 1**
Variance explained by the number of PCs obtained from individual principal component analyses and the RV and $\text{RV}_{\text{modified}}$ coefficients describing the degree of overlap between the two score matrices of a definite number of PCs.

| PC | GC data [%] | Sensory data [%] | RV | $\text{RV}_{\text{modified}}$ |
|---|---|---|---|---|
| 1 | 11.2 | 30.3 | 0.0013 | $5.10^{-4}$ |
| 2 | 16.0 | 46.7 | 0.11 | 0.10 |
| 3 | 20.2 | 60.9 | 0.20 | 0.19 |
| 4 | 23.5 | 71.6 | 0.28 | 0.27 |
| 5 | 26.2 | 79.3 | 0.27 | 0.27 |
| 6 | 28.8 | 85.9 | 0.30 | 0.29 |
| 7 | 30.9 | 91.9 | 0.30 | 0.29 |
| 8 | 32.9 | 95.9 | 0.30 | 0.30 |
| 9 | 34.8 | 99.3 | 0.30 | 0.30 |

There is not a substantial change in the RV coefficient values after four principal components. Slightly lower values are observed for the modified RV in comparison with the original RV indicating that in this case, both coefficients reveal the real underling relationship between the two sets of variables. The results confirm that even though there is an overlap between the GC and sensory data, there are also substantial parts that are unique for the individual sets. From the values of the percentages of variance for the number of PCs presented in Table 1, it can be seen that four PCs explain 23.5% of the chromatographic data variance, while the same number of PCs explains 71.6% of the variance in the sensory data. For illustrative purposes, the projections of mixtures and variables in the spaces spanned by the first two PCs for both analyses are presented in Fig. 2.

An interesting general observation is that the variance explained by the first two PCs in the chromatographic data (see Fig. 2a and b) is associated with the distinction between the mixtures containing rhamnose and xylose, while in the sensory dataset the difference of cysteine and methionine mixtures from all of the other mixtures is highlighted (see Fig. 2c and d). The lack of influence of amino acids on the distinction of the mixtures using their chromatographic profiles has also been observed with other clustering methods such as the Neural-Gas network, Generative Topographic Mapping, Bottleneck Neural Networks, the Sammon's mapping and the Kohonen self-organising maps and was reported in Ref. [11].

The loadings plot in Fig. 2b shows that the variables responsible for the clustering trend observed along PC 2 in Fig. 2a are nos. 45, 137, 128, 71, 51 and 70 with negative loading values and no. 64 with a positive loading value. The high intensity of peak 64 and low intensities of peaks 45, 137, 128, 71, 51 and 70 are characteristic for rhamnose-containing mixtures, whereas the opposite tendency holds true for xylose-containing mixtures. Cysteine-containing mixtures seem to have the most intense meat-like and sulphur smell. Cysteine actually is a sulphur-containing amino acid. The other amino acid-containing sulphur is methionine, but generates a highly intense smell of potatoes. In the food industry, methionine is usually added to some products to enhance the soft flavour of potatoes. Mixtures containing amino acids other than cysteine or methionine in a combination with one of the reducing sugars are characterised mostly by a caramel and jammy smell.

The projections of mixtures in the spaces spanned by the other pairs of principal components were also carefully inspected. Since they did not provide any new information about the clustering tendency in both chromatographic and sensory data analyses, they are not presented here.

As was mentioned before, when the aim of the analysis is to relate the blocks of variables, we look for a compromise representation of samples in a component space of the various block variables. In the next sub-section, a comparison of the compromise representations obtained from different methods for multiblock analysis is presented. Again each variable was subjected to autoscaling. With STATIS autoscaling per variable and correction of blocks size were applied.

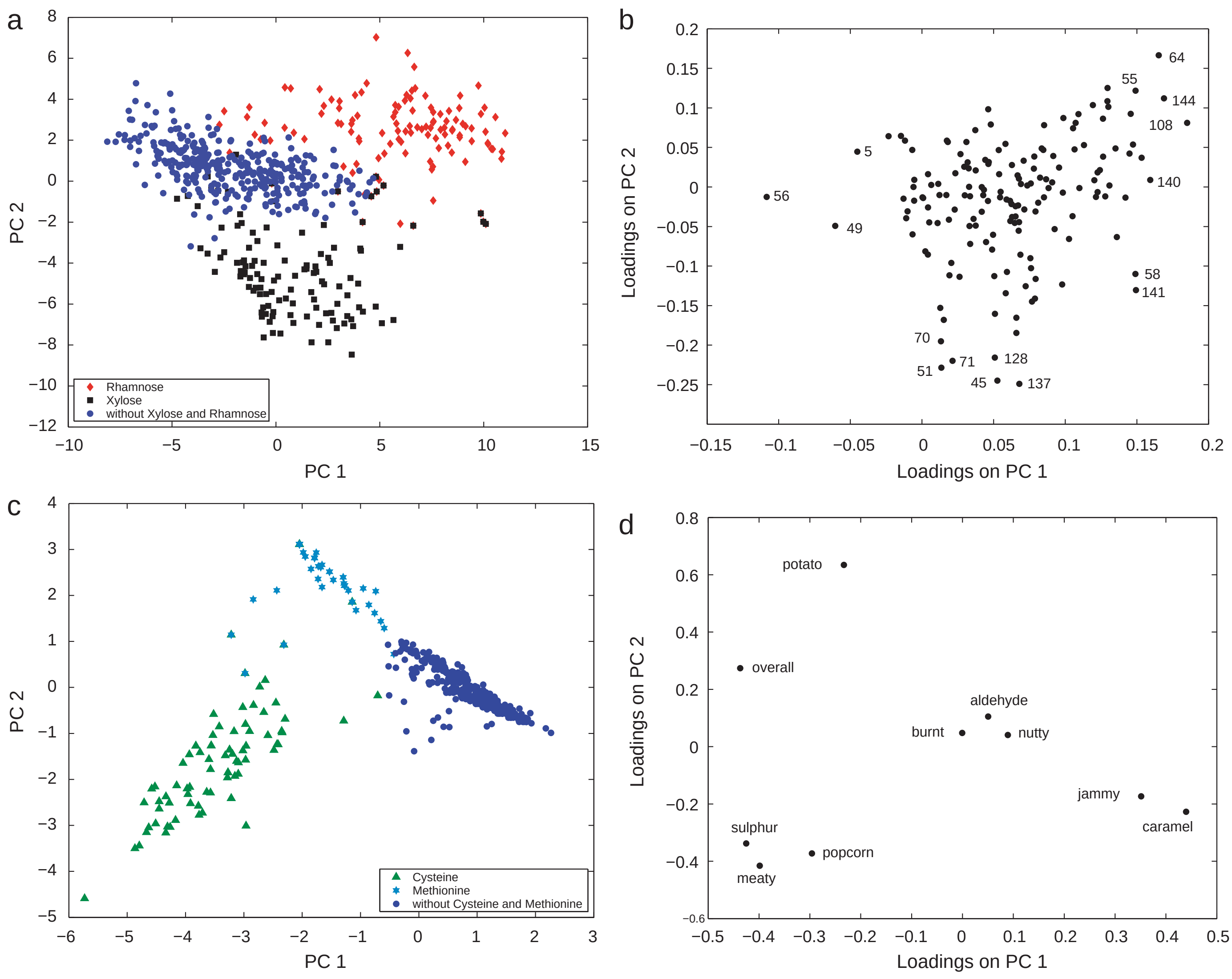

**Fig. 2.** Principal component analysis applied to 576 × 159 gas chromatographic and 576 × 10 sensory data separately: (a) projections of mixtures into the space spanned by the first two principal components after PCA of GC data, (b) projections of GC variables into the space spanned by the first two principal components, (c) projections of mixtures into the space spanned by the first two principal components after PCA of sensory data and (d) projections of sensory descriptors of smell into the space spanned by the first two principal components.

### *4.2. Analysis of the similarity between GC and sensory data using methods for simultaneous multiblock analysis*

In order to investigate the similarity of the compromise representations, i.e. score spaces, **T**, obtained using various methods that can handle multiblock data, the Tucker's coefficient of congruence can be used. For two sets of score vectors, $\mathbf{T}_1$ and $\mathbf{T}_2$ obtained from methods 1 and 2, the coefficient of congruence, $\varphi$, is calculated using the following equation:

$$\varphi = \frac{\mathrm{tr}(\mathbf{T}_1{\mathbf{T}_2}^T)}{\sqrt{\mathrm{tr}(\mathbf{T}_1{\mathbf{T}_1}^T)\mathrm{tr}(\mathbf{T}_2{\mathbf{T}_2}^T)}} \qquad (3)$$

In this equation 'tr' denotes the sum of the diagonal elements of the resulting matrices obtained from the matrix multiplication in the brackets. The definition of $\varphi$ is the same as the definition of the correlation coefficient, in which the centering procedure is omitted. It takes values between −1 and 1 for perfectly congruent distributions of samples. Because of the rotational freedom of PCA, the estimation of $\varphi$ was performed after the Procrustes rotation and isotropic scaling of the two score matrices investigated. The values of $\varphi$ estimated for each pair of four-component score matrices obtained using the different methods and from the individual principal component analyses of gas chromatographic and sensory data are presented in Table 2.

The lowest congruence was observed for the individual analysis of GC and sensory data, which confirms the results obtained earlier that the PC analysis of GC data mainly highlights information about the sugar in the mixtures, while the sensory data emphasise the mixtures containing some specific amino acids like cysteine and methionine. For autoscaled data, the methods that put a higher weight on GC or sensory data show higher coefficients

**Table 2**
The Tucker's coefficient of congruence between the score matrices with four components ($f = 4$) obtained from simultaneous methods of multiblock analysis.

| | $\mathrm{PCA_{sensory}}$ | $\mathrm{PCA_{GC}}$ | CPCA | MFA | SUM-PCA | STATIS |
|---|---|---|---|---|---|---|
| $\mathrm{PCA_{sensory}}$ | 1 | | | | | |
| $\mathrm{PCA_{GC}}$ | 0.50 | 1 | | | | |
| CPCA | 0.62 | 0.98 | 1 | | | |
| MFA | 0.92 | 0.74 | 0.82 | 1 | | |
| SUM-PCA | 0.98 | 0.62 | 0.72 | 0.98 | 1 | |
| STATIS | 0.98 | 0.62 | 0.72 | 0.98 | 1 | 1 |

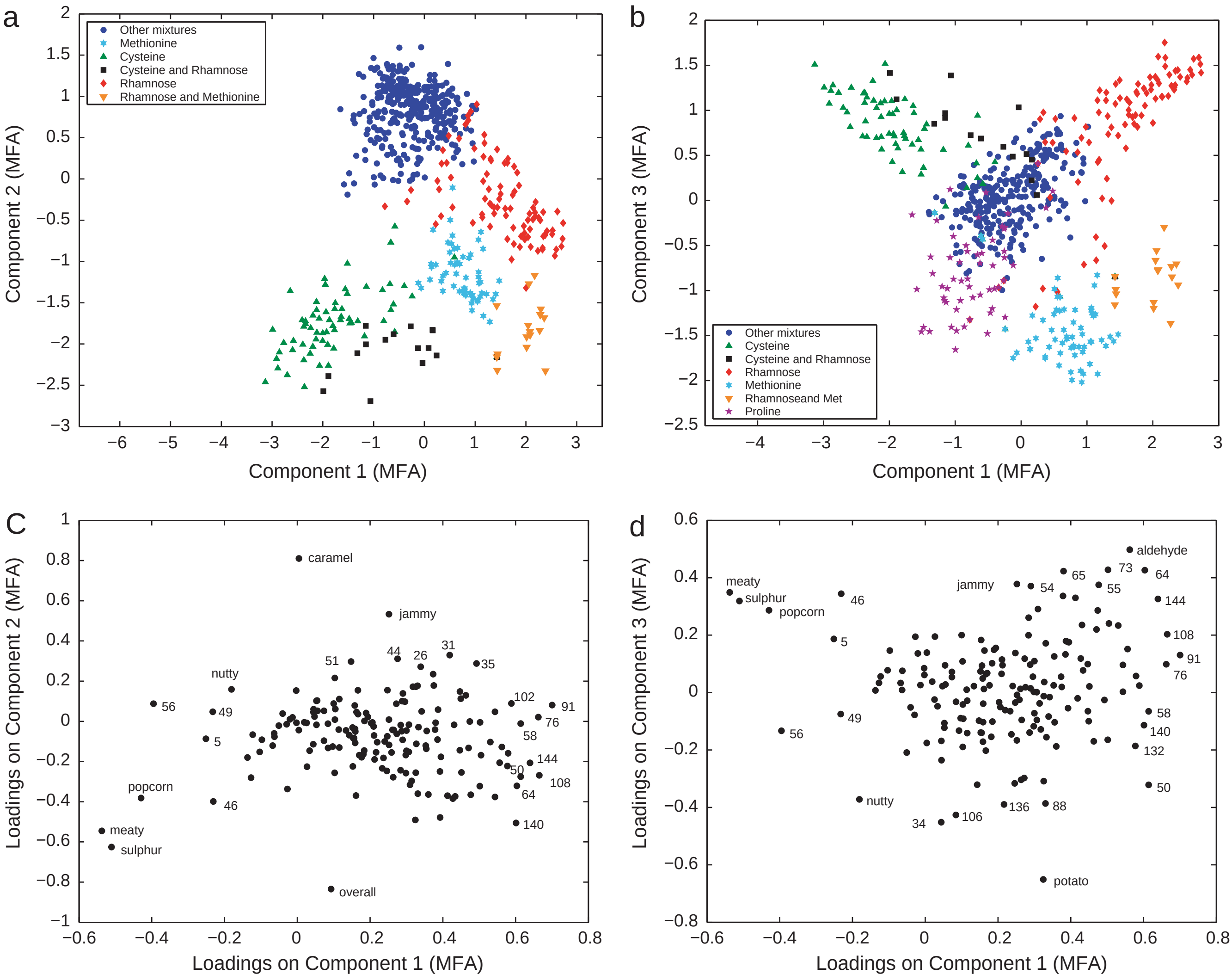

**Fig. 3.** Multiple factor analysis, MFA, of the concatenated data **X** (576 × 169): (a) projections of mixtures into the space spanned by the first two components of MFA, (b) projections of mixtures into the space spanned by the first and third components of MFA, (c) projections of GC variables and sensory descriptors into the space spanned by the first two components of MFA and (d) projections of GC variables and sensory descriptors into the space spanned by the first and third components of MFA.

of congruencies with the compromises of individual PCA analyses of GC or sensory data, respectively. In this context, CPCA puts a higher weight on the block with more numerous variables, i.e. the GC data. This results in a better congruence between the compromise scores matrix of CPCA and the scores matrix from PCA of GC data ($\varphi = 0.98$) than the congruence between the compromise scores matrix of CPCA and the scores matrix from PCA of sensory data ($\varphi = 0.62$). Due to the lack of a correction for block size and variation differences, the compromise representation from CPCA has relatively lower values of $\varphi$ with the compromises from the other methods, i.e. 0.82 with MFA, 0.72 with SUM-PCA and STATIS. An opposite tendency was observed for SUM-PCA. Because each block of variables was scaled to the sum of squares one, the block with fewer variables was highlighted, while the block with more variables was downweighted. Consequently a better congruence between the compromise scores matrix of SUM-PCA and the scores matrix from PCA of sensory data ($\varphi = 0.98$) was observed in comparison with the congruence between the compromise scores matrix of SUM-PCA and the scores matrix from the PCA of GC data ($\varphi = 0.62$). It is worth noting that for two blocks of variables SUM-PCA and STATIS gave the same compromise representations as indicated from the values presented in Table 2. A perfect congruence of both representations was also indicated by $\varphi = 1$. Please note that with STATIS we used autoscaling per variable of a block and a correction for block size. Without a correction of block size, STATIS uses a higher relative weight to the GC data because of their larger sum of squares.

As can be seen from the values presented in Table 2, a more balanced representation of the two blocks of autoscaled variables is found with MFA. A higher weight, but lower than in SUM-PCA, was attributed to the sensory data and a lower weight, but higher compared to SUM-PCA, was put on the GC data. In general, if the aim of the analysis is to find a common structure, which agrees with the individual principal component analyses, both MFA and SUM-PCA would be suitable. Here, we will mainly focus on the interpretation of the common representation of samples obtained by MFA. The reason is that we are mainly interested in the possibility of comparing the information content of GC and sensory variables by eliminating the predominant influence of the GC variables in the simultaneous analysis.

The projections of mixtures (see Fig. 3a and b) and the corresponding projections of variables (see Fig. 3c and d) onto the selected compromise components of multiple factor analysis are shown in Fig. 3.

Although there is a slight overlap, some groups of a higher density can be observed in both plots representing the distribution of

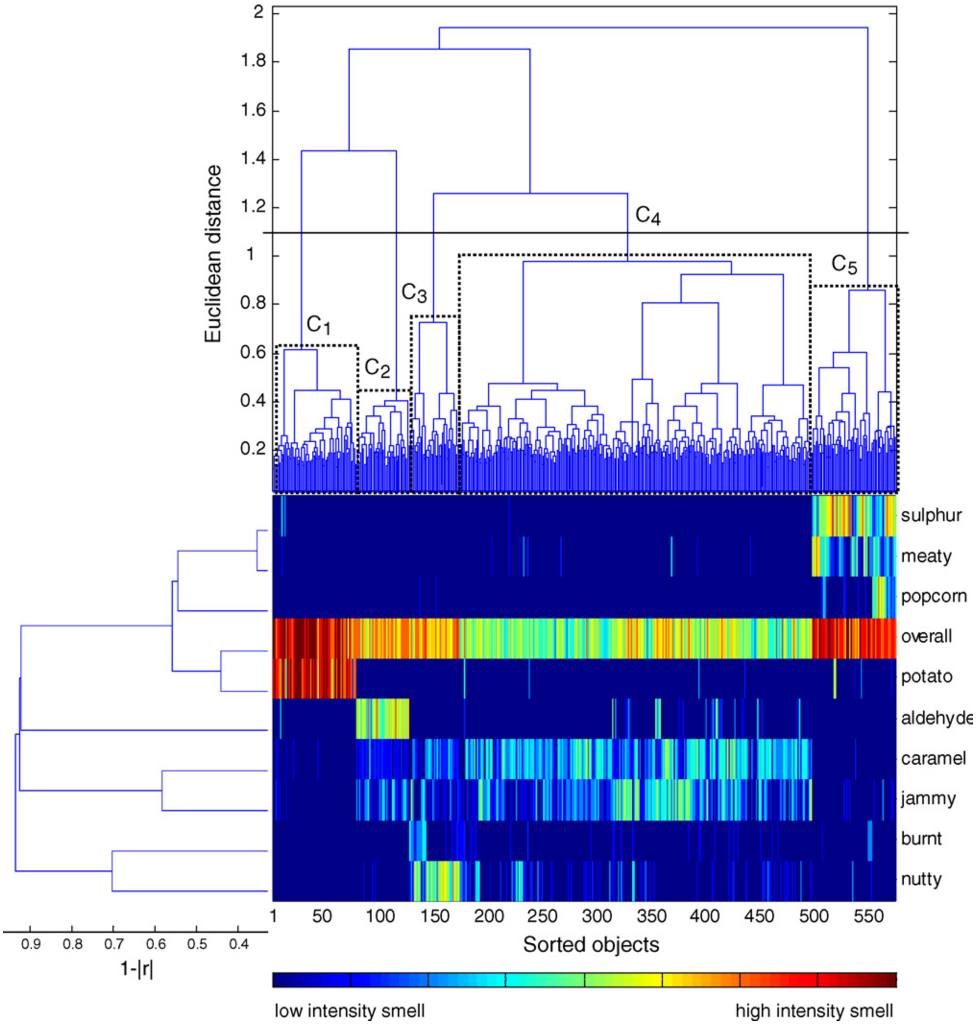

**Fig. 4.** Two-way hierarchical clustering applied to autoscaled and MFA-weighted data. The Ward's method for hierarchical clustering with the Euclidean distance as a similarity measure was used for the mixture samples, while the single linkage method with the $1 - |r|$ as a similarity measure was applied for the sensory variables. The corresponding heat map is related to the intensity of a smell.

mixtures. This clustering tendency is more clearly emphasised in the plot of the first versus the third component (compare Fig. 3a and b). Compared with the projections obtained from the analysis of individual blocks of data that are shown in Fig. 2, a more comprehensive picture can now be seen. An interesting observation is that in addition to methionine- and cysteine-containing mixtures distinguished in the individual analysis of sensory data, one can now differentiate all the rhamnose-containing mixtures from the rest of the mixtures. The combinations of rhamnose and cysteine as well as rhamnose and methionine are also clearly separated. Looking at the projections of variables (see Fig. 3c and d) indicating the impact of each variable on the construction of a given component, it can be observed that cysteine-, methionine- and methionine- and rhamnose-containing mixtures have a strong overall odour in comparison with the other mixtures. The changes in the intensity of the meat-like or sulphur smell can be indicated by peak 56 and all of the chromatographic variables with positive loading values along the first component (see Fig. 3c). The highly intense sulphur smell of cysteine mixtures is related to the high intensity of peak 46 and the low intensity of all peaks with positive loading values along the first component. Among the three groups with the meat-like and sulphur smell, the rhamnose- and methionine-containing mixtures have the least intense smell. In general, the groups containing methionine have a potato-like smell, which is indicated by the projections in Fig. 3b and d along with the third component of MFA. The potato smell is also related to peaks 34, 106, 136, 88 and 50. A less intense potato smell is characteristic for the methionine- and rhamnose-containing mixtures, which might be explained by the aldehyde smell associated with rhamnose that partially masks the potato smell. The high intensity of peaks 73, 64, 55, 65 and 144 is also characteristic for the rhamnose-containing mixtures. A nutty smell is attributed to the proline mixtures. Peaks 49 and 56 are also characteristic for these mixtures (see Fig. 3d).

In order to be able to observe the pattern of all mixtures in the whole space of variables, the autoscaled and MFA-weighted data can also be used as input into a two-way hierarchical clustering. Using the methods of hierarchical clustering, a unique order of mixtures is established by joining them together according to a similarity measure. Here, the Ward's method for hierarchical clustering with the Euclidean distance as a similarity measure was used [20] for the mixture samples, while the single linkage method with the correlation coefficient as a similarity measure was applied for the sensory variables. Specifically, $1 - |r|$ was adopted, which implies that highly correlated variables are linked at the lower levels of the dendrogram. The objects linked at the lower levels of the dendrogram are more similar to each other. The interpretation of the dendrograms can be extended by a so-called heat or colour map [21]. Each pixel of the heat map has a colour proportional to the values of variables in the original data. The two dendrograms for the autoscaled and MFA-weighted data and the corresponding heat map constructed for the sensory variables are shown in Fig. 4.

Five clusters denoted as $C_1$–$C_5$ in the figure were obtained by cutting the links of the dendrogram at the selected level (see Fig. 4 above). The first group, $C_1$, consists of methionine-containing mixtures. The heat map shows that these mixtures, together with the cysteine-containing mixtures, $C_5$, are characterised by the most intense overall smell. However, the first group has the smell of a potato, while mixtures in the fifth group have an intense meat-like and sulphur smell. The second group, $C_2$, of rhamnose mixtures and the third group, $C_3$, of proline-containing mixtures produce a less intense overall smell. An interesting finding is that the mixtures containing proline have a nutty-like smell, while the presence of rhamnose in the mixtures contributes to an intense aldehyde-like smell as indicated by the heat map in Fig. 4. The largest group, $C_3$, of mixtures contains all of the other sugars with the exception of rhamnose and these mixtures have the least intense overall smell. For them, a strong caramel/jammy smell is often recognised. The results from the two-way hierarchical clustering of weighted data complete the results obtained from the MFA analysis. A more detailed data interpretation can be obtained by considering the similarity among the mixtures at the lower levels on the dendrogram and the impact of the sensory variables ordered according to $1 - |r|$, can be observed with the help of a heat map.

## 5. Conclusions

A comparison of the results obtained from the individual principal component analyses and the results from the methods for simultaneous multiblock analysis showed that the use of the information of the chromatographic and sensory data in conjunction enhanced the interpretability of the data. When the aim of the study is to obtain a common structure of the blocks of variables which share similar information, but also have unique parts, both SUM-PCA and multiple factor analysis, MFA are suitable. In the case of the example data described in this paper, MFA was preferred because of the possibility of reducing the predominant influence of gas chromatographic variables in the compromise representation of mixtures. Thanks to the specific weighting scheme of MFA, it was possible to relate the content of the chemical mixtures to the sensory profiles of the products from the Maillard reaction. Moreover, a more detailed interpretation of the similarities among the mixtures was obtained using the Ward's method of hierarchical clustering on the autoscaled and MFA-weighted data.

## Acknowledgment

The authors would like to thank Guillaume Desclaux for his help with the description of the analytical method.

## References

[1] T. Naes, E. Risvik, Multivariate Analysis of Data in Sensory Science, Elsevier, Amsterdam, The Netherlands, 1996.
[2] S. Roussel, V. Bellon-Maurel, J.-M. Roger, P. Grenier, Chemometrics and Intelligent Laboratory System 65 (2003) 209–219.
[3] J.L.H.C. Busch, K. Hrncirik, E. Bulukin, C. Boucon, M. Mascini, Journal of Agricultural and Food Chemistry 54 (2006) 4371–4377.
[4] A.K. Smilde, J. Westerhuis, S. de Jong, J. Chemometrics 17 (2003) 323.
[5] B. Escofier, J. Pages, Analyses Factorielles Simples et Multiples: Objectifs, Methodes, Interpretations, Dunod, Paris, 1990.
[6] B. Escofier, J. Pages, Computational Statistics & Data Analysis 18 (1994) 121.
[7] I. Stanimirova, B. Walczak, D.L. Massart, Analytica Chimica Acta 545 (2005) 1.
[8] Ch. Lavit, Y. Escoufier, R. Sabatier, P. Traissac, Computational Statistics & Data Analysis 18 (1994) 97.
[9] P. Schlich, in: T. Naes, E. Risvik (Eds.), Multivariate Analysis of Data in Sensory Science, Elsevier, Amsterdam, The Netherlands, 1996, p. 259.
[10] I. Stanimirova, B. Walczak, D.L. Massart, V. Simeonov, C.A. Saby, E. Di Crescenzo, Chemometrics and Intelligent Laboratory Systems 73 (2004) 219.
[11] F. Questier, Q. Guo, B. Walczak, D.L. Massart, C. Boucon, S. de Jong, Chemometrics and Intelligent Laboratory Systems 61 (2002) 105.
[12] M. Daszykowski, B. Walczak, D.L. Massart, Chemometrics and Intelligent Laboratory Systems 65 (2003) 97.
[13] A. Smilde, M.J. van der Werf, S. Bijlsma, B.J.C. van der Werf-van der Vat, R.H. Jellema, Analytical Chemistry 77 (2005) 6729.
[14] K. Van Deun, A.K. Smilde, M.J. van der Werf, H.A.L. Kiers, I. Van Mechelen, BMC Bioinformatics 10 (2009) 246.
[15] P. Robert, Y. Escoufier, Applied Statistics 25 (1976) 257.
[16] C.-T. Ho, Thermal generation of Maillard aromas, in: R. Ikan (Ed.), The Maillard Reaction. Consequences for the Chemical and Life Sciences, J. Wiley & Sons, Chichester, UK, 1996, pp. 27–53.
[17] J.M. Ames, Trends in Food Science & Technology 1 (1990) 150.
[18] S.I.F.S. Martins, et al., Meat flavor generation in complex Maillard model systems, controlling Maillard pathways to generate flavors, ACS Symposium Series 1042 (2010) 71.
[19] A.K. Smilde, H. A.L. Kiers, S. Bijlsma, C.M. Rubingh, M.J. van Erk, Data and Text Mining 25 (2009) 401.
[20] D.L. Massart, L. Kaufman, The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis, Robert E. Krieger Publishing Co., Inc., Malabar, FL, USA, 1989.
[21] A. Smoliński, B. Walczak, J. Einax, Chemometrics and Intelligent Laboratory Systems 64 (2002) 45.