



Chemometric analysis of the volatile fraction evolution of Portuguese beer under shelf storage conditions



Ricardo Rendall^a, Marco S. Reis^{a,*}, Ana Cristina Pereira^a, Cristina Pestana^b, Vanda Pereira^{c,d}, José Carlos Marques^{c,d}

^a CIEPQPF, Department of Chemical Engineering, University of Coimbra, Rua Sílvio Lima, 3030-790, Coimbra, Portugal

^b Empresa de Cervejas da Madeira, Parque Empresarial Zona Oeste, 9304-003 Câmara de Lobos, Portugal

^c Centre of Exact Sciences and Engineering, University of Madeira, Campus da Penteada, 9000-390 Funchal, Portugal

^d Institute of Nanostructures, Nanomodelling and Nanofabrication (I3N), University of Aveiro, Aveiro, Portugal

ARTICLE INFO

Article history:

Received 17 September 2014

Received in revised form 24 November 2014

Accepted 24 January 2015

Available online 3 February 2015

Keywords:

Lager beer

Volatile profile

Chemometrics

Principal component analysis

Clustering

Variables' contributions

ABSTRACT

In this work we present a multivariate statistical analysis of the evolution of the volatile fraction of Portuguese beer over an extended period of 1 year under standard shelf storage conditions, using gas chromatography coupled with mass spectrometry (GC–MS). A systematic methodology is proposed for detecting the onset of meaningful changes in chemical composition during shelf storage and to monitor its evolution along time. We also put forward and discuss chemometric procedures for analyzing the contributions of different chemical components in the definition of dynamic ageing trends. In summary, the chemometric analysis reveals that the chemical composition of beer presents a statistically meaningful deviation from the reference scenario after a period of 7 months, although the deviation trend has its onset during the 6th month. The analysis performed also underlines the limitations of current variable contribution methods, and an alternative procedure was proposed based on the analysis in the original domain which finally led to a consistent and interpretable clustering structure of the volatile fraction compounds. Esters and higher alcohol compounds stand up on a cluster arrangement suggesting that their strict control can effectively point out meaningful changes on beer aroma. Organic acids, namely caprylic, capric and acetic acids can also be very helpful in that sense.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Brewery industry has been well-succeeded in dealing with a wide variety of relevant quality issues in its products. Quality aspects related with beer appearance and safety, namely colloidal stability, haze formation or microbiological spoilage, are now considered to be completely under control. Currently, the most important quality parameter and the shelf-life-limiting quality aspect is beer flavor [1–4] and efforts have been conducted towards maximizing and predicting beer shelf life taking into account flavor information.

Several compounds have been pointed out as key agents of beer sensory changes and suggested as deterioration markers during ageing, in particular diacetyl [5], acetaldehyde and ethyl acetate [6], (E)-2-nonenal [7], Iso- α -acids [8], riboflavin and amino acids [9] and 5-hydroxymethylfurfural [10]. However, the evaluation of beer shelf life should be more comprehensive and take into account the heterogeneity of chemical volatile groups presented in beer, which can either individually or simultaneously, influence the beer stale flavor in a

synergistic or antagonistic sense [11]. Furthermore, beer aroma is also an important aspect in customer preferences and loyalty, since the typical beer consumer, once he/she has approved and selected a given brand, expects the same “degree of excellence” (i.e., quality) from the product on future experiences [6,12–15].

Aware of the importance of aroma evolution during ageing, brewers and the scientific community have been developing approaches to evaluate the overall beer aroma quality as well as strategies to verify the accomplishment of expected attributes in the final product. Electronic sensors have deserved special attention in that sense. A number of papers have proposed the use of electronic tongue devices for quality control, in particular regarding the prediction of beer sensory attributes frequently evaluated according to the brewers' panel list [16–19]. Results achieved in these works suggest that e-tongue can be used as a fast screening tool to check quality control parameters in a routine way. Electronic nose systems have also been tested to evaluate aroma fingerprint changes in beer during the ageing process. M. Ghasemi-Varnamkhasti et al. reviewed the last applications of e-nose in brewery applied to quality assessment [20], reporting applications for hop's quality assessment [21], on-line process control [22], monitoring the odor and taste of active compounds [23] and as a brand authenticity checking tool [24,25]. All works reported in this review take advantage

* Corresponding author. Tel.: +351 239 798 700; fax: +351 239 798 703.
E-mail address: marco@eq.uc.pt (M.S. Reis).

of the use of multivariate techniques for data analysis, being reported a plethora of techniques used in exploratory data analysis, calibration and pattern recognition tasks. The information of spectroscopy techniques (FTIR, NIR, RMN) combined with chemometric modeling has also generated good results on monitoring the fermentation process [26,27] and on assessing quality control parameters [28]. Other works explored the information of beer aroma obtained from chromatography techniques to discriminate products by brands, categories and production areas [17,29–31]. Chemical composition underlying beer deterioration has also been analyzed. Most of the work has been conducted under ageing acceleration conditions, namely by submitting beer samples to high temperatures [32–35]. In these cases, the classification models typically achieve good discrimination performances concerning the differentiation of fresh and matured beer. Multivariate tools applied to classification and calibration models also elucidate the flavor characteristics responsible for the differences identified during ageing. However, and notwithstanding the value of the information reported in these studies, to the best of our knowledge, no strategy or methodology is yet available for specifically addressing the key issue of beer shelf life. This is the focus of the presented work, where a methodology is proposed for determining the period of time after which beer fresh features start to change in a statistically significant way under standard shelf storage conditions and, accordingly, all degradation phenomena begins to accelerate. In this study, bottles were maintained at normal ageing conditions and the entire analysis carried out during one year. This is another differentiating feature of the present study, with the purpose to make the whole assessment as realistic as possible, mitigating the risks in the transfer of results to the industrial practice.

The remaining parts of this article are organized as follows: Section 2 describes the experimental protocol and the analytic and chemometric methods. Section 3 presents the analysis workflow and the main results obtained. In Section 4, we discuss relevant aspects of the analysis and some of its implications. Finally, in Section 5, we summarize the main findings in this work and address future work worth undertaking in the sequence of the results reported in this article.

2. Materials and methods

2.1. Beer samples

A total of 39 lager beers were analyzed in this study. Samples were kept at room temperature, within 12 months on amber glass bottles of 330 mL. The analysis of volatile compounds was done monthly, for three different bottles randomly selected. All samples were kindly supplied by the *Empresa de Cervejas da Madeira, Lda*, and collected from the same production batch to minimize inter-batch variability.

Throughout the text, samples will be denoted by the month in which each bottle is open and analysis carried out and, when necessary, a letter (A, B and C) will be used to distinguish among the replicates analyzed for the same ageing time.

2.2. Analytical procedures

The analysis of volatile compounds was carried out on a TRACE GC Ultra gas chromatograph equipped with the ISQ single quadrupole (electron impact mode) and the TriPlus autosampler (liquid mode) from Thermo Scientific (Hudson, NH). The extract injected was obtained through solid phase micro-extraction as described elsewhere [36]. A total of 70 volatile compounds were identified and quantified in terms of the relative area with respect to the internal standard (3-octanol). The identification of compounds was made by comparison of the mass spectra obtained with those present in NIST08 and Wiley 6.0 MS library databases, and comparing the obtained Kovats indices with those stated in the NIST Chemistry WebBook [37].

2.3. Chemometric methods

Chemometrics encompass a wide range of methodologies and approaches for extracting useful information from complex data sets originated in analytical instrumentation and industrial processes. There is no standard and universal protocol to conduct a chemometric analysis, and the specific path to follow is always dependent on the problem in hand, goals to achieve and the data structure originated by the analytical procedure adopted. In fact, chemometrics can be seen as an integral part of the scientific method when it comes to analyzing chemical information with modern instrumentation.

For the current problem, the collected data is structured in a two-way table, where one mode stands for the different components quantified with resource to gas chromatography coupled with mass spectrometry, GC–MS, (columns' mode) and the other mode regards time information relative to the moments where bottles from a homogeneous group were randomly opened and analyzed (rows' mode). The existence of such two meaningful modes in the data structure implies that one should select and adopt data analysis methodologies with the potential to extract and explore the fundamental pieces of information relative to both the composition and time modes, and eventually their interaction, if relevant. In this section we present the main frameworks employed in the chemometric analysis, which will integrate the workflow presented in Section 3.

2.3.1. Principal component analysis (PCA)

Principal component analysis is a well-known methodology to extract the linear correlation structure shared by multiple variables [38–40]. It essentially aims at explaining most of data variability using as few linear combinations of all variables (called variates) as possible. The linear combinations of the variables constitute new transformed variables, the principal components, PC_a , whose values are the scores, stored in $(n \times 1)$ column vectors t_a , (where n stands for the number of observations in the data set – in this work it is the total number of beer samples collected after storage in shelf conditions and analyzed in the laboratory). The coefficients of the linear combination are called loadings, which for each PC can be stored in the $(m \times 1)$ column vector l_j (where m is the number of variables – in the present work it corresponds to chemical compounds). As a scale-dependent technique, PCA is usually applied after proper pre-processing of the original $(n \times m)$ raw data matrix, \mathbf{X} . The resulting pre-processed matrix with the same size, \mathbf{Z} , presents better properties to achieve the purposes of analysis. In this work we applied autoscaling (centering to zero mean followed by scaling to unit variance) in order to provide all components with the same a priori weight in the PCA procedure.

In mathematical terms, the scores and loading vectors are gathered in matrices, called the score matrix, \mathbf{T} , and the loading matrix, \mathbf{L} , leading to the following decomposition of the pre-processed data matrix:

$$\mathbf{Z} = \mathbf{T} \cdot \mathbf{L}^T + \mathbf{E} \quad (1)$$

where \mathbf{T} and \mathbf{L} are $(n \times a)$ and $(m \times a)$ matrices, respectively, a is the number of retained components, also known as the pseudorank ($a < m$) and \mathbf{E} is a $(n \times m)$ residual matrix, with information that is orthogonal to that described by the retained principal components. In practice, an analyst looks for information regarding the variable mode in PC loadings (such as groups of correlated variables) and concerning the observation mode in the scores (such as trends and clusters of observations).

By construction, PCA is entirely focused on the variable mode and has no built-in capacity to describe or incorporate any part of the observation structure in its implementation. It tacitly assumes samples to be generated from an independent and identically distributed multivariate distribution (the usual *iid* assumption). Even its evolution to cope with the time-related mode, dynamic principal component analysis (DPCA), consists of using the trick of adding artificial variables corresponding to

lagged or shifted versions of the original variables, forming an expanded raw data table and implementing the usual PCA methodology over it [41]. However, it can be shown that such an approach is incapable of removing entirely the autocorrelation in the scores, meaning that time structure is not yet being fully taken into account [42]. This means that one has to find a complementary way to extract and analyze the relevant information contained in the time mode. This will be achieved through clustering analysis, as explained further ahead in the text (Section 2.3.3).

2.3.2. Change point detection in high dimensional systems

The main goal of the present work is to develop a data-driven approach to monitor the evolution of the chemical composition of Portuguese commercial beer and to propose a systematic way to detect the point after which its properties begin to change beyond what could be interpreted as a normal variation corresponding to the early stages of storage. In this context, the aim is not to develop a dynamical model for the evolution of the chemical composition, but a procedure for detecting the instant after which, with high probability, a change may have occurred in the beer composition, indicating a potential seed of deterioration.

In order to do so, the framework of multivariate statistical monitoring will be employed after adaptation to the specific conditions of this work. Multivariate statistical monitoring encompasses a class of different methodologies for monitoring multivariate systems, including the use of Hotelling's T^2 statistic [43] for low dimensional systems and of latent variable frameworks, such as PCA and PLS (partial least squares) especially suited for handling a large number of correlated *iid* variables [44–47]. As the data set under analysis consists of 70 correlated variables, a latent variable multivariate process monitoring approach based on PCA will be employed.

For such, let us consider again the PCA decomposition presented in Eq. (1). This numerical decomposition of the matrix \mathbf{Z} can be reinterpreted as a statistical description of the variability exhibited by the variables in its columns. According to the latent variable interpretation, such model suggests that the variability presented by the variables is originated at the level of several (a , the pseudorank of the system) latent variables or scores, \mathbf{T} , which are also affected by the additional interferences from unstructured sources of variation, \mathbf{E} . Therefore, by assigning suitable statistical descriptions to specify the behavior of the terms \mathbf{T} and \mathbf{E} , the statistical behavior of the variables in the \mathbf{Z} matrix is obtained. With such a reference model structure, one can finally implement a confirmatory-like analysis in order to evaluate if a given sample is consistent with the statistical model considered to properly describe the system, or alternatively decide that it defers considerably (at a given significance level, specified a priori). This amounts to a sequence of statistical hypothesis tests, which is the foundation of the statistical monitoring approach adopted here.

The monitoring procedure based on PCA consists of: (i) first estimating a PCA model representing the normal behavior of the process under supervision (in this case, regarding the initial stages of storage, before degradation); (ii) specifying the statistical limits of detection (for a given level of significance); and (iii) implementing the process monitoring procedure by computing the pair of monitoring statistics (see below) for each new incoming sample and evaluate if the process is normal or instead some significant change has occurred (in which case a subsequent diagnostic stage is usually conducted).

For each sampling time, the following pair of monitoring statistics is computed:

- (i) Hotelling's T^2 statistic of the PCA scores to monitor variability within the PCA subspace, verifying whether the projections of new samples onto this a -dimensional subspace fall inside or outside the region of normal behavior:

$$T_{PCA}^2(k) = \sum_{i=1}^a \frac{t_i^2(k)}{\lambda_i} = \mathbf{x}(k) \mathbf{L} \mathbf{A}_a^{-1} \mathbf{L}^T \mathbf{x}(k)^T \quad (5)$$

where $t_i(k)$ is the k th entry of the i th score (k stands for the current time index, where the statistic is to be computed), \mathbf{A}_a^{-1} is the inverse of the diagonal matrix with the a largest eigenvalues in descendent order of magnitude along the main diagonal, and $\mathbf{x}(k)$ is the $(1 \times m)$ row vector with \mathbf{X} data for the current multivariate observation.

- (ii) Q statistic (also referred as SPE – square prediction error) measures the deviation (squared Euclidean distance) between each multivariate observation, $\mathbf{x}(k)$, and the PCA subspace.

$$Q = \mathbf{e}^T \mathbf{e} = (\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{x} - \hat{\mathbf{x}}) = \mathbf{x}^T (\mathbf{I} - \mathbf{L} \mathbf{L}^T) \mathbf{x} \quad (2)$$

where $\hat{\mathbf{x}}(k) = \mathbf{T}(k) \mathbf{L}^T$ is the $(1 \times m)$ row vector corresponding to the projection of \mathbf{x} onto the PCA subspace. Therefore, contrary to the T^2 statistic for the PCA scores that monitors the variability of the projected samples, i.e., within the PCA subspace, the Q statistic strictly addresses the variability around the PCA subspace. With this pair of monitoring statistics it is possible to follow the behavior of an arbitrarily large number of variables without incurring in collinearity problems, as it happens with the Hotelling's T^2 statistic. In fact, with this statistic, the inversion of the covariance matrix may become ill-conditioned or even unfeasible, due to the presence of highly collinear variables. This will not happen with the current monitoring scheme, as the scores are uncorrelated by design and only contemplate the main contributors to the \mathbf{X} variability (large eigenvalues). Thus, the computation of the Hotelling's T^2 for the scores offers no problems in such conditions, nor will the computation of the residual statistic, Q , which is not prone to any numerical computational issues.

The control limits for the T^2 and Q statistics, that define the upper limits for the normality region, are usually computed as follows [45–49]. The upper control limit (UCL) of the T^2 statistic is given by:

$$UCL_{T^2} = \frac{a(n-1)(n+1)}{n^2-na} F_{\alpha,a,n-a} \quad (3)$$

where $F_{\alpha,a,n-a}$ is the upper α percentile of the F distribution with a and $n-a$ degrees of freedom. On the other hand, The UCL for the Q statistic (UCL_Q) is computed through the corresponding percentile of a weighted χ^2 distribution, $g \cdot \chi_{\nu}^2$, whose parameters are estimated from the first ($\hat{\mu}_E$) and second ($\hat{\sigma}_E^2$) empirical moments of the values of Q obtained from reference data [50]:

$$\begin{aligned} \hat{g} &= \frac{\hat{\sigma}_E^2}{2\hat{\mu}_E} \\ \hat{\nu} &= \frac{2\hat{\mu}_E}{\hat{\sigma}_E^2} \end{aligned} \quad (4)$$

There is a mild assumption of Gaussianity in the derivation of these statistics, which is quite reasonable in most circumstances as the scores result from the contribution of many variables which makes them more Gaussian than each of the isolated variables, due to the Central Limit Theorem (CLT). Furthermore, the behavior of the measurements collected from the beer samples undergoing shelf storage during the initial period is quite stable and thus amenable to be described by a static model such as the one implicitly considered in the above described multivariate statistical approach for detecting the early onset of meaningful changes in the chemical composition of beer.

However such control limits tend to be somewhat optimistic in the assessment of multivariate variability, leading to some underestimation of the actual multivariate variability. In order to protect against this feature, the control limits were computed using the same formulas presented above but using data obtained through a leave-one-out procedure.

2.3.3. Clustering

Another class of methodologies employed in the chemometric analysis of the data generated in the experimental procedure conducted in this work is clustering. Like PCA, clustering is another unsupervised machine learning methodology, but now focused on identifying natural groups of objects. The objects can be either variables or samples. When addressing the clustering of samples, one seeks to find groups of observations that present similar behavior or identical patterns, from the standpoint of the variables characterizing them. On the other hand, clustering variables usually address the identification of groups of variables presenting strong mutual associations and that, conjecturally, could be closely related or even involved in the same phenomenon. Both perspectives will be employed in this work.

There are several classes (and sub-classes) of clustering algorithms [51–53]. In this work we will use agglomerative hierarchical clustering. In this technique, there are initially as many clusters as objects. Then, the procedure successively groups similar objects into clusters, and similar clusters into new larger clusters, until only one big cluster exists that contains all the objects. As the groups are merged into larger clusters, the similarity of the objects in the same cluster tends to decrease. Therefore, there is usually an intermediate level of agglomeration that should be used in order to identify the natural clusters present in the data. The identification of this level is facilitated with the analysis of the so called dendrogram, which is a two dimensional diagram where the history of agglomeration of all the variables and clusters is represented and the hierarchical structure in which the successive clusters were formed is made explicit (in hierarchical clustering, once a given object is assigned to a cluster containing a set of objects, they will all be maintained together until the end). Also presented in this plot is the distance at which two groups merge into a new cluster. It represents a measure of the difference between the objects in the two clusters: clusters that merge at short distances are more similar while those that only do so at larger agglomeration distances are composed with more diverse objects.

The fundamental aspects to define before implementing an agglomerative hierarchical clustering (AHC) technique are: (i) the measure of distance and (ii) the agglomeration criterion. The first aspect defines how we want similarity to be assessed. Considering two objects, $x_1, x_2 \in \mathcal{R}^m$ (m -dimensional column vectors), one may want to call “similar” those objects that are characterized by identical values of their m properties, in which case the Euclidean distance (d_E) or the city-block distance (d_{CB}) may be suitable choices.

$$d_E(x_1, x_2) = \sqrt{\sum_{i=1}^m [x_1(i) - x_2(i)]^2} \quad (5)$$

$$d_{CB}(x_1, x_2) = \sum_{i=1}^m |x_1(i) - x_2(i)| \quad (6)$$

But we may also be interested in identifying objects whose properties are strongly correlated, in which case a measure of similarity based on the scale-free Pearson correlation coefficient, (d_p), is more appropriate:

$$d_p(x_1, x_2) = 1 - |\rho(x_1, x_2)|. \quad (7)$$

The second aspect establishes the way clusters are combined into larger clusters, by defining the rule to compute the distance between two clusters of objects. Common options include: single linkage (minimum distance, as specified in i. between objects in the two clusters; also known as nearest-neighbor); complete linkage (maximum distance between objects in the two clusters; also known as farthest neighbor) and average linkage (or average distance between all pairs of points in the two clusters).

As referred above, natural clusters are identified by analyzing the dendrogram and establishing a threshold on the agglomeration distance: the clusters defined at such a distance will be the ones considered for analysis (see for instance Fig. 3a). The consistency of the clusters assigned to each object, should also be evaluated. This can be carried out by computing the so called “silhouette” score for each object x_i , $S(x_i)$, defined by:

$$S(x_i) = \frac{(\min(AVER_BETWEEN(x_i, k)) - AVER_WITHIN(x_i))}{\max(AVER_WITHIN(x_i), \min(AVER_BETWEEN(x_i, k)))} \quad (8)$$

where, $AVER_WITHIN(x_i)$ stands for the average distance from object x_i to the other objects laying in the same cluster, and $AVER_BETWEEN(x_i, k)$ is the average distance between object x_i and all other objects from cluster k . The silhouette score for object x_i provides a relative measure of the closeness of one object belonging to one group from the objects grouped in other clusters. This measure ranges from +1 (for objects that are distant from those belonging to neighbor clusters, indicating a well-defined clustering structure), to –1 (for objects whose attribution to a given cluster is not so clear, or that were eventually assigned to the wrong cluster). An example of a silhouette plot can be found in Fig. 3b.

3. Results

This work addresses the development of an effective procedure for monitoring and detecting meaningful changes in the chemical composition of Portuguese beer, during the period it undergoes shelf storage. In this section we present the main results obtained in this study. Aligned with the main goal, the chemometric analysis conducted was centered on detecting the early onset of meaningful changes in chemical composition (detection task) and then on the analysis and characterization of the evolution of groups of compounds (diagnosis task). The data analysis workflow followed is schematically presented in Fig. 1.

The results of the detection task will be presented in the next subsection, after briefly addressing the exploratory data analysis (EDA) task. Then, in the following subsection we report on the results of the diagnosis task.

3.1. Large-scale monitoring of the evolution of chemical composition (detection task)

We begin our study with a global analysis of the experimental data set consisting of compositions collected for a wide variety of chemical components (70 analytes overall) quantified through GC–MS, corresponding to beer samples taken over time for a period of 12 months in shelf storage conditions (including a sample taken at the beginning of

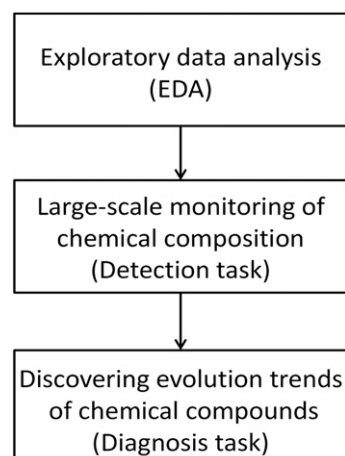


Fig. 1. Data analysis workflow.

the period, amounting to a total of 13 samples with 3 replicates each). Performing a PCA analysis of the entire data set one obtains the results presented in Fig. 2 (data was preliminarily autoscaled). From the analysis of the scree plot (Fig. 2a), the presence of a strong correlation structure is evident, with the first component gathering the equivalent variation of over 35 original autoscaled variables (50% of the compounds quantified by the analytical methods). The 2D scatter plot (Fig. 2b) illustrates a “smile”-like shape for the dynamic behavior of chemical composition when projected onto the first two PCA scores, starting from the top right region of the plane, moving to the left, downwards and then upwards. There are however some areas where several samples collected at different times overlap, indicating that some variation may exist along higher dimensions that is not being properly captured by the low dimensional 2D projection. With the 3D scatter plot one is able to discern some of these trends (Fig. 2c). From the analysis of the scores behavior it is apparent that, after an initial period where the evolution is rather weak or almost absent (with the exception of the first sample), the system undergoes a change, indicating a modification of the original chemical composition.

This remark can be further investigated with resort to an agglomerative hierarchical clustering (AHC) analysis performed on the observations. In this analysis, the time sequence is not considered at all, and the grouping is made strictly on the basis of the similarity between the observations (replicates of the 12 samples collected over time). Even so, the results obtained clearly indicate that the first 6 samples (or 18 observations) form a natural cluster (Fig. 3a), which confirms the similarity between such samples and substantiates the conjecture of an essentially static behavior in the initial period of shelf storage. The consistency of this grouping is high, as can be appreciated from the silhouette plot presented in Fig. 3b, where the scores of the observations are essentially located above 0.5 and none is less than 0.

In this context, the first 6 samples were considered to represent the variability one expects to obtain under normal storage conditions without any significant evolution of chemical composition, constituting the reference set. All the reminding observations are gathered in the so-called, test set. The reference subset of normal condition observations was used to develop a new PCA model which faithfully describes the multivariate variability in such conditions. This model was then used to implement the change point detection procedure for high dimensional systems presented in Section 2.3.2, as described next.

First the number of latent variables to retain in such PCA model was selected, using a stratified 3-fold cross-validation approach (in order to assure that in each subgroup left aside during the cross-validation procedure, always a replicate from each sample will be found; this is necessary to guarantee the representativeness of the data set in each cross-validation trial). The cross-validation plot is presented in Fig. 4, based on which 10 principal components were selected.

Then, with the normal conditions PCA model estimated, the same reference set of observations was used to estimate the control limits for the T^2 and Q detection statistics, using the formulas presented before and a leave-one-out procedure. Finally, each observation in the test set was projected onto the PCA models after proper pre-processing (using the sample processing parameters as the reference samples), and the corresponding values of the T^2 and Q detection statistics computed and plotted in the respective charts (called control charts). The results obtained are presented in Fig. 5. In this figure, one can verify that the first detection, i.e., the first statistically significant deviation from normal conditions, occurs at observation 25 (9th sample, corresponding to 8 months of shelf storage) for the T^2 statistic and at observation 22 (8th sample) for the Q statistic.

In summary, the chemometric analysis reveals that the chemical composition of the beer begins presenting a statistically meaningful

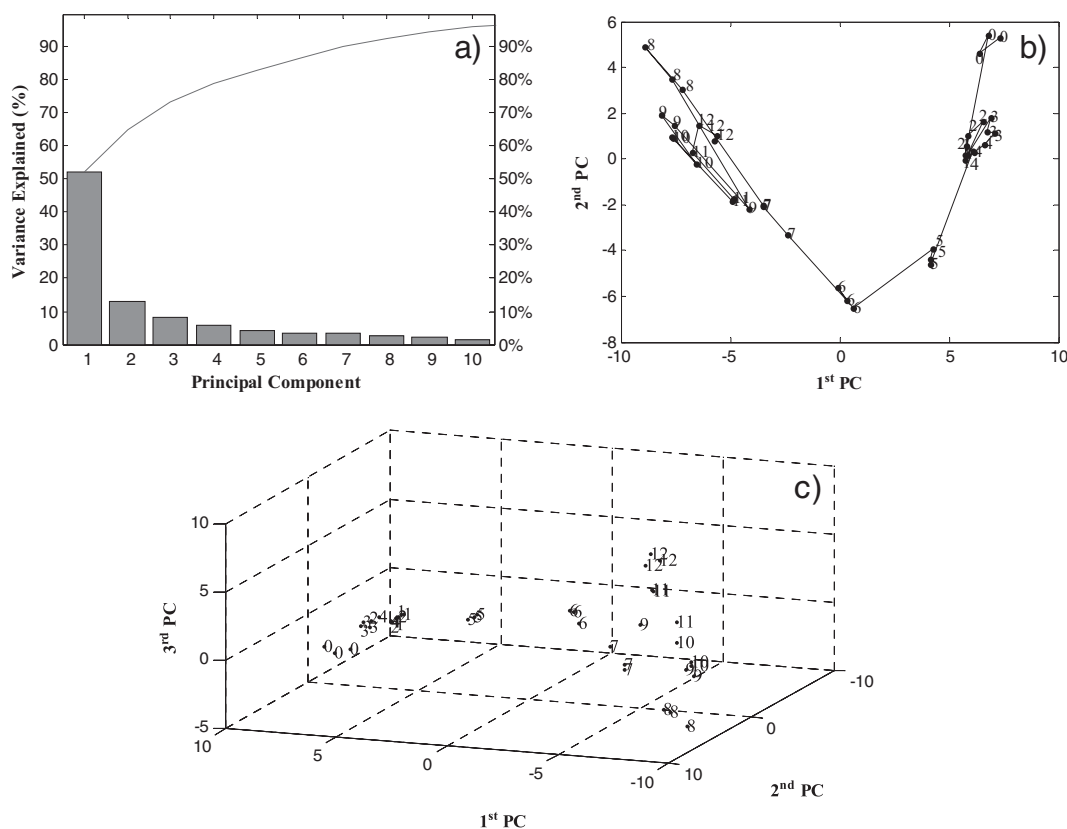


Fig. 2. Results of the PCA analysis performed on the entire experimental data set: a) Scree plot and b) 2D scatter plot of the scores of PC1 and PC2; 3D scatter plot of the scores of PC1, PC2 and PC3.

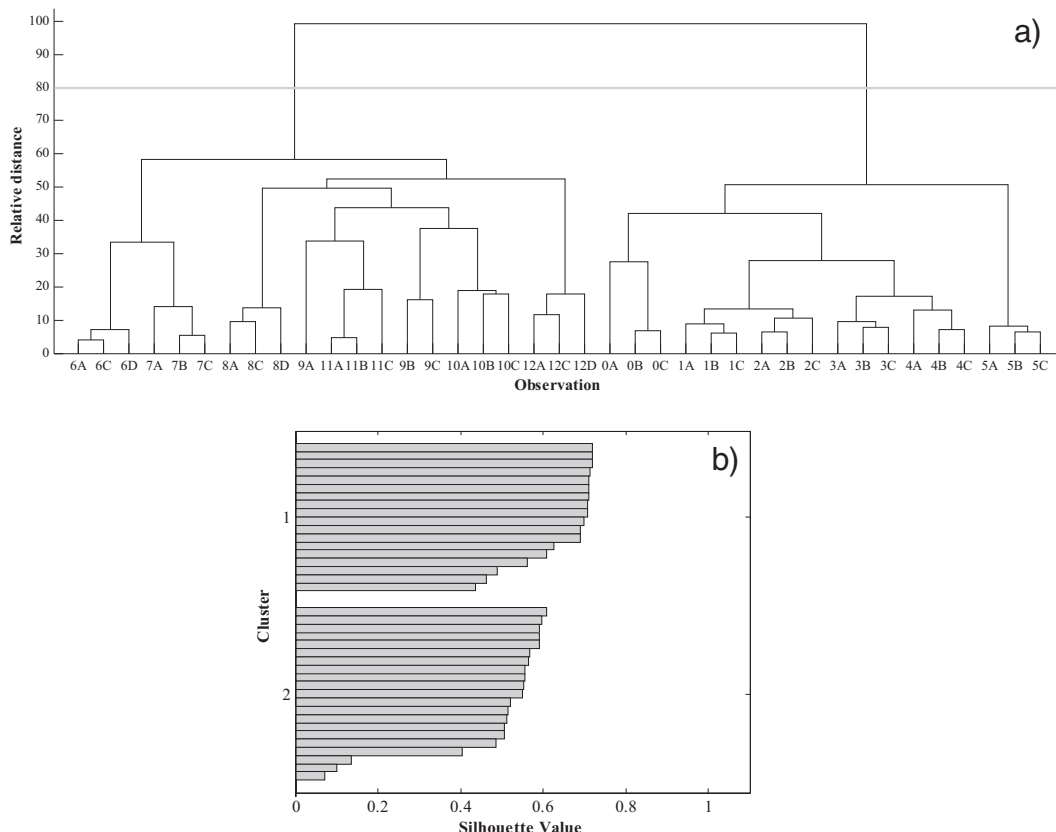


Fig. 3. Clustering observation in the beer data set: a) dendrogram and b) silhouette plot. The distance measure used is the city-block distance and the agglomeration criterion was average linking.

deviation from the reference scenario after a period of 7 months, although the deviation trend has its onset during the 6th month.

3.2. Discovering evolution trends of chemical compounds (diagnosis task)

Having established the procedure for detecting the onset of the dynamical trend in chemical composition, the chemometric analysis turns the focus now to the characterization of such trend and the discovery of correlated groups of compounds that might share some phenomenological affinity (diagnosis task). Thus, contrary to the analysis in the previous section which was mainly focused on the samples mode, now it will be essentially concerned with the variable or compound mode of the two-way data table.

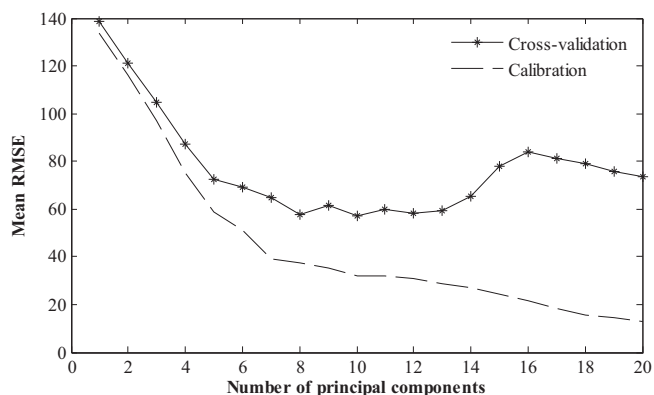


Fig. 4. Cross-validation plot for the selection of the number of principal components to retain in the normal conditions PCA model built with the reference data set.

In the scope of principal component analysis, contribution plots are well-known tools to explore this type of research question [54,55]. Namely, they are instrumental in finding out the variables that may be driving a given change observed at the scores or residuals level, such as those observed in the T^2 and Q charts. The contributions of each variable, j , to the square of the normalized score i at the k th replicate is computed as follows:

$$c_{i,j}^T = \frac{t_i}{s_{t_i}^2} l_{i,j}^T (x_j - \bar{x}_{REF,j}), \quad (9)$$

$$c_j^T(k) = \sum_{i=1}^a c_{i,j}^T(k) \quad (10)$$

where $\bar{x}_{REF,j}$ represents the average value of the j th variable in the reference set and $s_{t_i}^2$ is the estimated variance of the latent variable t_i in the reference set (also given by the corresponding eigenvalue). The total contribution of each variable is obtained by summing the contributions arising from all a retained principal components, as given by Eq. (10). Similarly, the contribution of the individual variable, j , for a change in the Q statistic is given by:

$$c_j^Q(k) = [\mathbf{x}_j(k) - \hat{\mathbf{x}}_j(k)]^2 \quad (11)$$

where $\mathbf{x}_j(k)$ is a vector containing a new observation and $\hat{\mathbf{x}}_j(k)$ is an estimate of the new observation obtained by multiplying its scores and the transpose of the loading matrix (\mathbf{L} , with a latent variables).

Even though quite widespread, the application of contribution plot analysis is rather limited to the present situation, namely because: i. The information provided by contribution plots is essentially univariate, but the underlying phenomena most likely encompass the collective behavior of several components varying together; ii. The overall dynamic

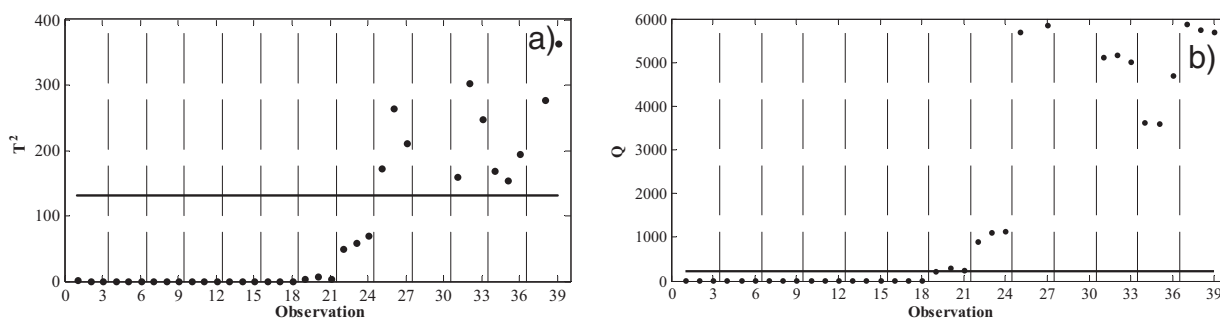


Fig. 5. Control charts for the T^2 and Q statistics (vertical lines delimitate sequences of three replicates). The first clear out-of-control point for the Q statistics occurs at observation 22. Some future observations are not presented because the deviation is so high that it becomes out of scale in this plot.

profile is not considered, as the analysis is based solely on the change verified from a reference scenario to another condition; what happens in between or afterwards is not incorporated in the analysis; and iii. Contribution plots suffer from the so-called smearing effect [55], which consists of a propagation of the estimated effect due to a change in one variable, to several other variables highly correlated with it.

Given the above limitations of contribution plots, two alternative procedures were developed with the ability to overcome the problems described above: i) an improved dynamic variable contribution analysis and ii) a variable clustering approach conducted in the original domain.

3.2.1. Dynamic variable contribution analysis

Analyzing variable contributions to displacements in the score space or changes in the Q statistic using the current available procedure presents the drawbacks of being univariate and neglecting information regarding the specific trend in analysis. Therefore, a new framework is proposed here that addresses such limitations, called dynamic variable contribution analysis.

The proposed approach consists in considering the set of samples representing normal conditions (in the present case all the replicates from samples 1–6, i.e., replicates 1–18) and use the centroids in the score space and residual space as the reference points kept fixed when computing the respective contributions. Then, for the first observation of the test set (i.e., not belonging to the reference set), one computes the contribution of each variable to the change observed in the score space on one hand, and in the complementary residual space, on the other, and register them in one row of two separate tables: one for the variable contribution to the scores, table CS, and another for the variable contribution to the residuals, table CR. This procedure is repeated for the next observation in the test set, and the variable contributions saved in the second row of tables CS and CR. Proceeding in this way to all the n_{test} test observations (in our case, all the replicates from samples 7 to 13 (i.e., replicates 22–39; see also Fig. 6), one obtains finally the complete ($n_{test} \times m$) CS and CR tables of variable contributions.

Tables CR and CS have in each column, all the contributions of a given variable to the successive deviations obtained from sample to sample relatively to the same reference point. Fig. 7 illustrates the profiles obtained for the contributions of each variable in the two complementary spaces. Analyzing this figure, it is possible to appreciate the existence of some common trends in the contributions profiles. In fact the proposed procedure and the way the two tables were constructed, potentiates the identification of clusters of variables with similar dynamic patterns. This feature will be further explored.

This analysis also enables a richer assessment of the relative importance of each variable in the establishment of the entire trajectory in the test set. In other words, it provides information on the importance of each chemical component in the changes observed in the score and residual spaces. For such, we computed the ranking of the contribution of each variable in explaining the change observed in each observation in

the test set regarding the reference condition. Again, one variable will have a ranking score for each test set observation, and in the end it will have as many rankings as test observations. Table 1 present the average ranking for each variable in decreasing order of importance, regarding the analysis of the contributions in the score space, and Table 2 presents the corresponding information regarding the contributions in the residual space.

As referred before, analyzing the isolated importance of the contributions of each variable may bring some useful information, but still of a univariate nature. Most often chemical phenomena originates correlated changes in the components, for which reason it is particularly interesting to develop a procedure that is able to identify such occurrences. This can be accomplished within the framework of the proposed analysis, by performing clustering of variables on the CS and CT tables. Variables (i.e., chemical compounds) whose contributions to the change in each space share similar patterns would be naturally gathered in the same group. In order to do so, an appropriate distance metric should be adopted that reflects the similarity concept one aims to implement: similar variables present strongly correlated profiles. One appropriate distance metric for this scenario is the correlation distance $d_p(x_1, x_2)$ (Eq. (7)). The dendrograms obtained from clustering of variables in the CS and CR tables are presented in Figs. 8 and 9.

Analyzing the dendrograms presented above, one can observe that there is a larger cluster of highly correlated variables or chemical components, to which other variables join in at larger agglomeration distances. However, there is no clear separation between the highly correlated variables and others and even looking at the original variables, some of the associations are not clearly justifiable. Moreover, variables highly correlated in terms of their contribution to changes in the scores are not necessarily identified as such when analyzing the

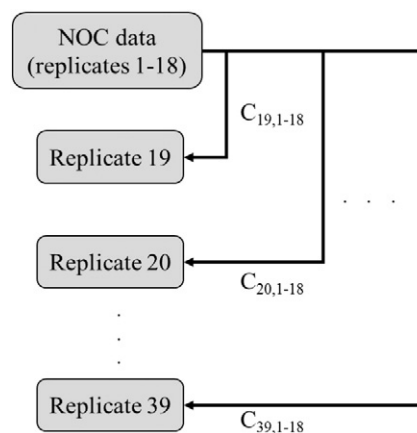


Fig. 6. Schematic representation of the procedure followed to build the 2-way table used in dynamic variable contribution analysis.

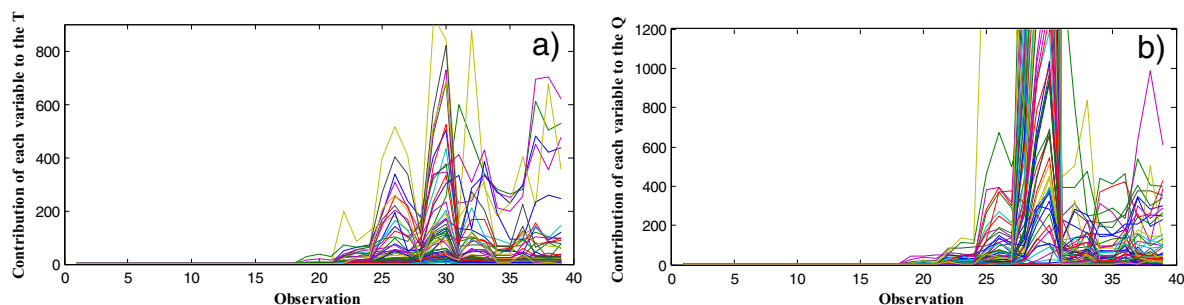


Fig. 7. Contribution profiles for all variables regarding changes in the (a) score space and (b) residual space (in this case, tables CS and CR also contain information of the contributions regarding the initial reference samples to compare with those after the beginning of the change).

dendrogram for the contributions in the residual space. This brings ambiguity to the analysis and conclusions, which may also be promoted by the smearing effect referred before.

Therefore, we conclude that the analysis of contribution plots on correlated data with dynamical patterns is rather involved and the decomposition of such an analysis into two complementary spaces introduces one more difficulty in the process. In fact, the final aim is to access the importance of each chemical component in the changes observed and the identification of groups of correlated components, an analysis that can be properly conducted in the original space, as described in the next subsection.

3.2.2. Clustering variables in the original domain

Instead of adopting the chemometric framework of contribution plots to analyze the influence of a given chemical component in

explaining an observed change in the score and residual spaces, separately, an alternative procedure explored in this work was to analyze the change in chemical composition in the original domain. In fact, the aim of analyzing the contributions in both the scores and residuals after a PCA decomposition is precisely to take into account all sources of variation. The natural alternative is therefore to analyze directly the structure of the original space, namely the similarity of the component profiles using a similarity metrics defined by the correlations distance.

Fig. 10 presents the dendrogram that results from conducting a clustering analysis in the original uncompressed domain, where it is possible to observe a strong clustering structure with two clearly identifiable clusters plus a series of other smaller clusters or individual instances at larger agglomeration distances representing components with rather different variation profiles (see description in Table 3). Rather interesting is the similarity in the trends found in each cluster,

Table 1

Mean ranking for each variable (in decreasing order, i.e., from the most important in average to the less important), regarding their contributions to changes in the score space relatively to the reference condition.

Component ID	Mean rank	Component ID	Mean rank
Isopentanol	6.8	1-Decanol	36.3
Nerolidol	7.3	Phenylethyl butyrate	36.6
Myrcenol	9.4	Furfural	36.9
Acetic acid	9.5	Pelargonic acid	37.1
Isoamyl acetate	11.3	Difurfuryl ether	37.8
Linalool	11.7	Hexyl acetate	38.8
Ethyl caproate	12.1	2-Methylcaproic acid	39.4
Caprylic acid	12.4	Ethyl caprylate	40.7
Isobutanol	14.3	Unknown 4	43.8
1-Heptanol	14.7	Ethyl valerate	43.8
Nonanal	16.5	3-Octanone	44.4
Isobutyl acetate	17.4	Dimethyloctanol	44.4
Phenethyl alcohol	17.4	Unknown 1	44.8
Citronellol	17.5	1-Hexanol	45.2
Caproic acid	17.7	Octyl acetate	46.3
Ethyl acetate	19.4	Ethyl hydrocinnamate	46.4
Phenyl ethyl acetate	22.6	3-Methyl-1-hexanol	46.7
Ethyl butyrate	23	Linalool tetrahydride	47.1
Ethyl phenylacetate	24.6	p-Cymene	47.3
Vinylguaiaicol	24.6	Ethyl 3-hexenoate	47.3
Propyl laurate	24.7	Heptyl acetate	49.3
Agidol	25.8	Ethyl capronate	49.4
2-Propyl-1-pentanol	26.6	Decanal	50
Unknown 3	28.5	Myrcene	50.4
Geranyl acetone	28.9	Citronellol acetate	51.3
1-Octanol	29	Furfuryl alcohol	53
Unknown 5	29.1	β-damascenone	54.1
2-Decanol	29.2	Lauryl alcohol	58.3
Capric acid	32.3	6-Methyl-2-heptanone	62.0
Ethyl 5-methylhexanoate	33	Methyl geraniate	62.6
Ethyl 2-nonenoate	34	Ethyl caproate	63
Unknown 6	34.2	Isoamyl butyrate	63.6
Caproic acid	34.2	Ethyl 9-decenoate	64
Ethyl pelargonate	34.9	Furfuryl acetate	64.8
Unknown 2	35	Carbitol	68.4

Table 2

Mean ranking for each variable (in decreasing order, i.e., from the most important in average to the less important), regarding their contributions to changes in the residual space relatively to the reference condition.

Component ID	Mean rank	Component ID	Mean rank
Unknown 5	9.8	Ethyl 3-hexenoate	37.1
Ethyl caproate	11.5	Heptyl acetate	38.3
3-Methyl-1-hexanol	12.4	Ethyl capronate	38.7
Nerolidol	13.2	Caproic acid	38.7
Linalool	13.8	Ethyl phenylacetate	39.2
1-Hexanol	14	Isobutanol	40
Ethyl 5-methylhexanoate	14.1	β-damascenone	40.3
Furfuryl alcohol	15.6	Agidol	40.3
6-Methyl-2-heptanone	17.7	2-Propyl-1-pentanol	42.8
Myrcenol	19	2-Decanol	42.9
Unknown 6	20	Phenylethyl butyrate	43
Lauryl alcohol	20.3	Furfural	44.1
Nonanal	22.6	Unknown 2	44.2
Ethyl 2-nonenoate	22.8	1-Decanol	44.2
Unknown 1	22.9	Hexyl acetate	44.4
Isoamyl acetate	23.9	Citronellol acetate	45.1
1-Heptanol	24.5	2-Methylcaproic acid	45.4
Isobutyl acetate	24.8	3-Octanone	45.7
Caproic acid	25.1	Methyl geraniate	46.3
Phenethyl alcohol	25.3	Linalool tetrahydride	47.2
Carbitol	25.3	Unknown 3	47.5
Citronellol	25.5	Ethyl caproate	48.1
Pelargonic acid	25.5	1-Octanol	48.2
Ethyl caprylate	28	Octyl acetate	48.7
Isopentanol	28.3	Ethyl 9-decenoate	49
Ethyl butyrate	29.6	Ethyl valerate	49.3
Geranyl acetone	30.1	Isoamyl butyrate	51
Acetic acid	30.8	p-Cymene	52.1
Difurfuryl ether	31	Ethyl acetate	52.4
Vinylguaiaicol	31	Propyl laurate	53.7
Capric acid	34.8	Ethyl pelargonate	56
Ethyl hydrocinnamate	34.8	Unknown 4	59.6
Caprylic acid	36.4	Myrcene	60
Dimethyloctanol	36.5	Furfuryl acetate	61.5
Phenyl ethyl acetate	36.5	Decanal	62.7

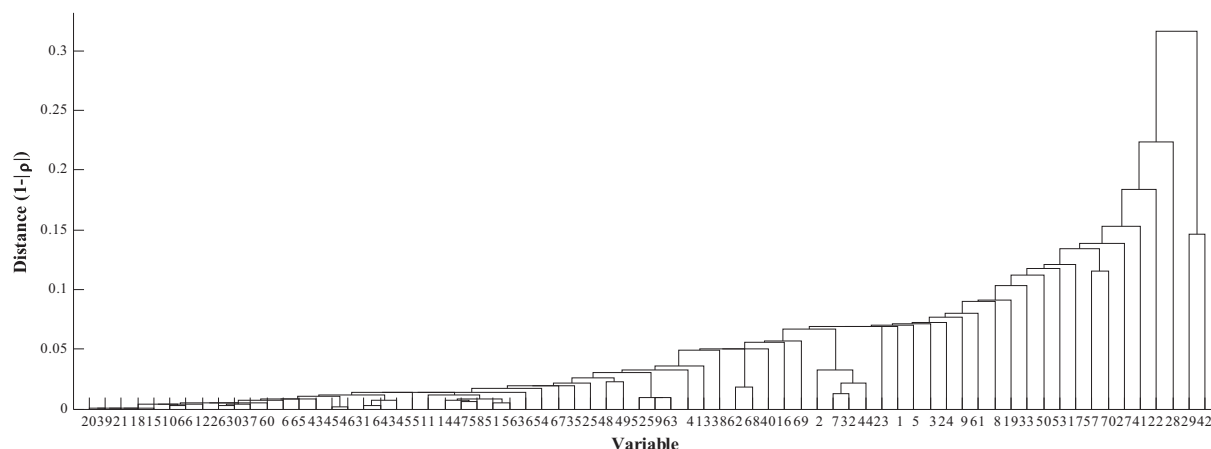


Fig. 8. Dendrogram for the clustering of variables made on the variable contribution table for the scores (CS).

as depicted also in Fig. 10, which are all of the same type despite in Eq. (7) one is using the absolute value of the correlation coefficient in the definition of the distance metric.

The clustering of components obtained with this approach is therefore much more consistent than that obtained using the PCA analysis based on the residuals, even considering the improvements implemented in dynamic variable contribution analysis.

4. Discussion

The main focus of this paper is the analysis of the evolution of the volatile profile of lager beer during 12 months at regular conditions of storage. The GC–MS analyses of thirty nine samples lead to the identification of 70 volatile compounds, from several chemical families, namely 25 esters, 15 alcohols, 7 acids, 6 aldehydes and ketones, 7 terpenes, 4 furans and other 6 compounds. One relevant question is to find out, among this array of compounds, those with an important role regarding the evolution of product quality.

The proposed procedure detects statistically meaningful differences in the chemical composition of future observations, regarding those of the initial reference set. However, a statistically meaningful deviation does not necessarily mean that the product is not conforming to the consumer requirements. Statistical significance is evaluated on the basis of natural variation present in the process under monitoring, whereas consumer requirements and perception of quality depend on factors external to the production process. Therefore, the statistical limits of variation for the detection statistics are fundamentally different from the specification limits for product acceptance. One important

problem, however, is the absence of consensus or guidelines on specification and requirements of the chemical composition of commercial beer. In this context, the data-driven approach presented here has the potential to bring more information to process owners and R&D&I personal on the dynamic evolution of beer composition along time, increasing their understanding and awareness on the chemical phenomena and their time scale (evolution kinetics), which ultimately will allow them to correlate such features with consumer requirements.

Regarding the analysis based on contribution plots, we would like to point out that they are the chemometric tool commonly used to investigate the influence of variables in the appearance of step perturbations or other simple changes in the scores or residuals. Under certain conditions they perform reasonable well, especially in stationary systems. However, this work brings out their limitations when addressing intrinsically dynamic trends. In fact, they were unable to provide a reasonable clustering of the components with a well-defined structure and meaning. Not even the proposed improved methodology of dynamical contribution plots was enough to achieve a reasonably good clustering performance in this case. Given their failure, we have turned the analysis to the original domain, which finally led to a clear and interpretable clustering structure. There is not, in our opinion, any formal inconsistency in adopting this procedure. In fact, the monitoring procedure based on PCA, follows the entire variability of data through the two complementary statistics, T^2 and Q . In other words, it performs a monitoring of the full domain, after properly decomposing it into two complementary subspaces which are monitored simultaneously. The same applies to the contribution plot analysis, where, once again, the influences of variables in both the scores and residuals of the PCA decomposition

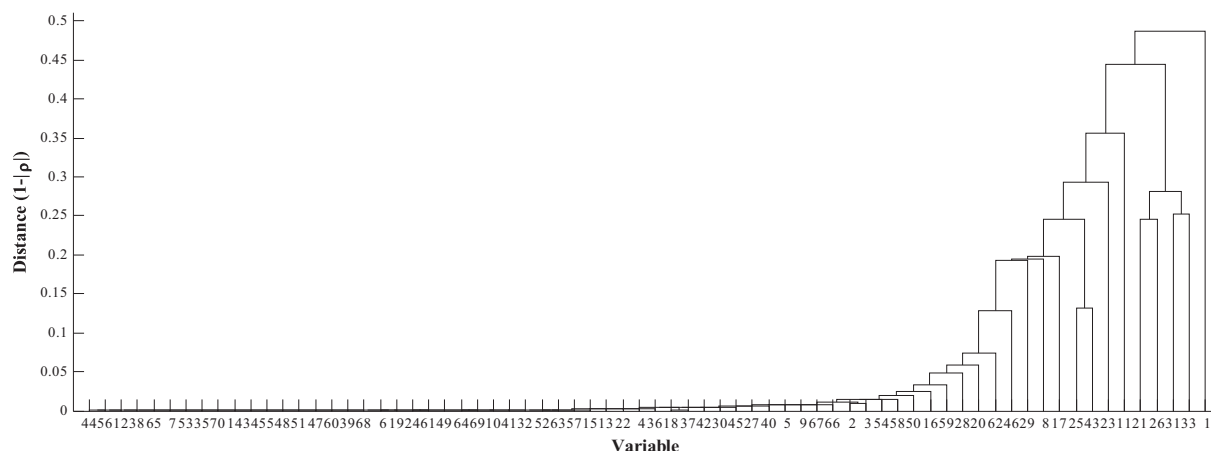


Fig. 9. Dendrogram for the clustering of variables made on the variable contribution table for the residuals (CR).

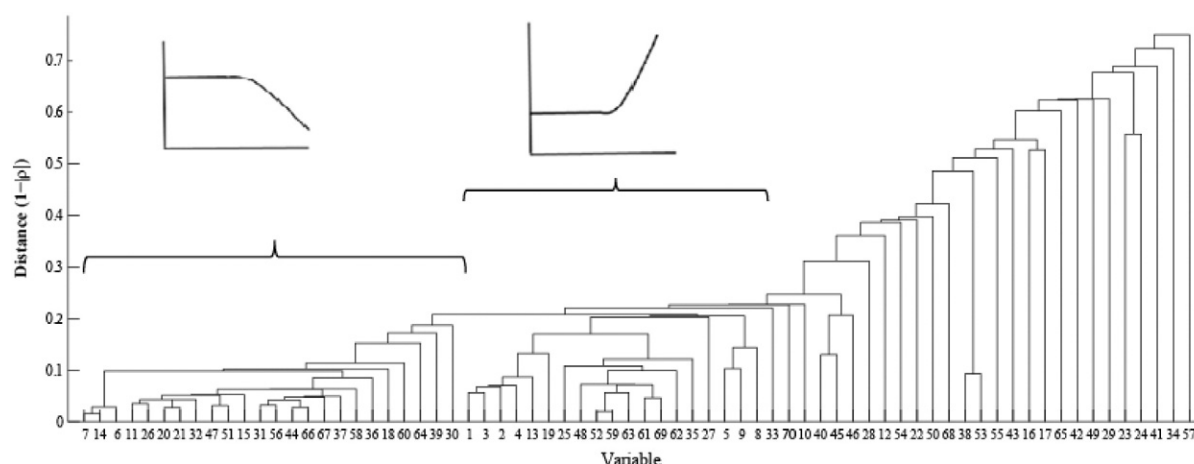


Fig. 10. Identifying groups of correlated compounds through clustering of variables in the original domain.

was conducted simultaneously, which altogether constitute an analysis of the original domain. In this context, our clustering analysis of the components according to the correlation of their dynamic trends in the original domain is entirely consistent with the endeavor of the analysis carried out with the chemometric tools previously applied, and furthermore leads to good results. We believe that this is an interesting outcome of this work, worthwhile to be retained and further explored in future applications.

Focusing on our analysis in Table 3, it is possible to summarize the chemical compounds evolution for lager beer under study. Cluster 2 is formed essentially by esters compounds (about 40% of the esters quantified are in this clusters). For lager beer, high esters concentration are generally undesirable, hence production process parameters like wort concentration and aeration, temperature and oxygen levels should be controlled to supervise and monitor their levels. Nevertheless, when present at lower concentrations and below their odor thresholds they contribute positively to the beer flavor [56,57]. For example, ethyl acetate flavor varies from a light pear-like character to solvent-like in high concentrations [58]. Cluster 2 gathers ester compounds with a similar evolution trend, namely after an initial period during which their concentration is low and follows a period with an increase in concentration levels. The isoamyl acetate, the ethyl caproate and the ethyl caprylate reached concentrations higher than their odor threshold (a rough estimate was obtained from an internal standard concentration, 3-octanol) and therefore during ageing they can contribute negatively to the flavor and beer deterioration. Most of the remaining ester compounds do not have a clear pattern of evolution, and therefore their expression on the beer's global aroma depends on synergies with other compounds. Cluster 2 also includes the most abundant organic acids found in this beer, the caprylic acid and capric acid. Both compounds are responsible for caprylic flavor [59] that resembles the odor of a goat. Despite being abundant, none of these acids exceeded their flavor thresholds, 15 mg/L to caprylic acid and 10 mg/L to capric acid, but their trends after the 6th month deserve special attention. About 80% of

higher alcohols are located at clusters 1 and 2. In the beer analyzed it is estimated that none of these compounds exceeded their odor thresholds. In general terms, for a balanced flavor, it is expected that the ratio of the amount of higher alcohol to esters is about 1:2.5–3 [56]. At the 7th month, the concentration seems to depart from this scenario, which can be in line with the onset detected for meaningful changes. Furfural, a typical marker of beer flavor deterioration [60] does not present a clear trend during the storing period analyzed, been quantified at about five times less of its odor threshold. The remaining furanic compounds also do not present clear trends. However, their combined effect is supposed to contribute to beer staling during the later months.

From our analysis, there is no single resulting compound that can be identified as a unique ageing marker, but rather two sets of compounds acting in a synergistic or antagonistic way to produce meaningful changes on fresh beer flavor. This result reinforces the idea that aroma deterioration requires a comprehensive analysis, not centered in a specific marker but analyzing the combined effect of several volatile substances.

5. Conclusions

In this article we presented a data-driven approach to monitor the evolution of the chemical composition of Portuguese commercial beer and proposed a systematic way to detect the point after which its properties begin to change significantly (i.e., beyond what could be interpreted as the normal variation corresponding to the early stages of storage). Furthermore, we proposed and discussed chemometric procedures for analyzing the contributions of different beer components in the definition of the dynamic ageing trends: dynamic contribution plots and analysis in the original domain.

The study is based on experimental data on the volatile fraction of beer obtained from gas chromatography coupled with mass spectrometry, GC–MS. Samples were collected over an extended period of 1 year under standard shelf storage conditions.

Table 3

Clusters of variables identified in the analysis in the original chemical compounds' domain (component codes).

Cluster 1 (Downward trend)	Cluster 2 (Rising trend)	Other clusters (No clear pattern)
Ethyl valerate; myrcene; 3-octanone; <i>p</i> -cymene; ethyl 5-methylhexanoate; 3-methyl-1-hexanol; unknown 1; 1-hexanol; 1-heptanol; 2-propyl-1-pentanol; decanal; ethyl pelargonate; ethyl 2-nonenolate; myrcenol; carbitol; 2-methylcaproic acid; unknown 2; propyl laurate; unknown 3; agidol; lauryl alcohol; unknown 4; unknown 5; unknown 6	Ethyl acetate; isobutyl acetate; ethyl butyrate; isobutanol; isoamyl acetate; isopentanol; ethyl caproate; hexyl acetate; ethyl capronate; ethyl caprylate; acetic acid; 1-octanol; 1-decanol; phenyl ethyl acetate; phenethyl alcohol; phenylethyl butyrate; nerolidol; caprylic acid; capric acid	6-Methyl-2-heptanone; isoamyl butyrate; difurfuryl ether; ethyl 3-hexenoate; heptyl acetate; nonanal; linalool tetrahydride; octyl acetate; furfural; linalool; furfuryl acetate; 2-decanol; ethyl caproate; dimethyloctanol; citronellol acetate; furfuryl alcohol; ethyl 9-decenoate; methyl geranate; citronellol; ethyl phenylacetate; β -damascenone; caproic acid; geranyl acetone; ethyl hydrocinnamate; pelargonic acid; vinylguaicol; caproic acid

In summary, the chemometric analysis reveals that the chemical composition of the beer presents a statistically meaningful deviation from the reference scenario after a period of 7 months, although the deviation trend has its onset during the 6th month. The analysis performed underlines some limitations of contribution plots methods and an alternative procedure was put forward based on the analysis of data in the original domain, which finally led to a consistent and interpretable clustering structure of the volatile fraction.

As to future work, we will extend the use of the proposed methodologies to the analysis of ageing behavior in other economically relevant beverages, namely the potential of using dynamic contribution plots and the analysis in the original time domain. Furthermore, the following interesting aspect may also be explored. In this work, we have not adopted a sub-group statistical approach and therefore incorporated some inter-sample variability in the normal condition model. However, an alternative procedure, widely used in statistical process monitoring that retains the classical consideration of inter-sample variation as a faithful representation of process variation may also be considered.

Conflict of interest

There is no conflict of interests in this work.

References

- [1] B. Vanderhaegen, F. Delvaux, L. Daenen, H. Verachtert, F.R. Delvaux, Aging characteristics of different beer types, *Food Chem.* 103 (2007) 404.
- [2] S. Malfliet, F.V. Opstaele, J.D. Clippeler, E. Stryn, K. Gooris, L.D. Cooman, G. Aerts, Flavour instability of pale lager beers: determination of analytical markers in relation to sensory ageing, *J. Inst. Brew.* 114 (2008) 180.
- [3] J.-J. Dong, Q.-L. Li, H. Yin, C. Zhong, J.-G. Hao, P.-F. Yang, Y.-H. Tian, S.-R. Jia, Predictive analysis of beer quality by correlating sensory evaluation with higher alcohol and ester production using multivariate statistics methods, *Food Chem.* 161 (2014) 376.
- [4] M. Riu-Aumatell, P. Miró, A. Serra-Cayuela, S. Buxaderas, E. López-Tamames, Assessment of the aroma profiles of low-alcohol beers using HS-SPME-GC-MS, *Food Res. Int.* 57 (2014) 196.
- [5] G.C. da Silva, Abner A.S. da Silva, L.S.N.d. Silva, R.L.d.O. Godoy, L.C. Nogueira, S.L. Quiterio, R.S.L. Raices, Method development by GC-ECD and HS-SPME-GC-MS for beer volatile analysis, *Food Chem.* 167 (2015) 71.
- [6] B.M. King, C.A.A. Duineveld, Changes in bitterness as beer ages naturally, *Food Qual. Prefer.* 10 (1999) 315.
- [7] L.F. Guido, J.R. Carneiro, J.R. Santos, P.J. Almeida, J.A. Rodrigues, A.A. Barros, Simultaneous determination of E-2-nonenal and β -damascenone in beer by reversed-phase liquid chromatography with UV detection, *J. Chromatogr. A* 1032 (2004) 17.
- [8] I. Caballero, C.A. Blanco, M. Porras, Iso- α -acids, bitterness and loss of beer quality during storage, *Trends Food Sci. Technol.* 26 (2012) 21.
- [9] E. Sikorska, A. Gliszczynska-Swiglo, M. Insińska-Rak, I. Khmelinskii, D.D. Keukeleire, M. Sikorski, Simultaneous analysis of riboflavin and aromatic amino acids in beer using fluorescence and multivariate calibration methods, *Anal. Chim. Acta* 613 (2008) 207.
- [10] J.A. Rodrigues, A.S. Barros, B. Carvalho, T. Brandão, A.M. Gil, A.C.S. Ferreira, Evaluation of beer deterioration by gas chromatography-mass spectrometry/multivariate analysis: a rapid tool for assessing beer composition, XIII Weurman Flavour Research Symposium, Academic Press, Zaragoza, 2014, p. 435.
- [11] O. Pinho, I.M.P.L.V.O. Ferreira, L.H.M.L.M. Santos, Method optimization by solid-phase microextraction in combination with gas chromatography with mass spectrometry for analysis of beer volatile fraction, *J. Chromatogr. A* 1121 (2006) 145.
- [12] B. Vanderhaegen, H. Neven, H. Verachtert, G. Derdelinckx, The chemistry of beer ageing – a critical review, *Food Chem.* 95 (2006) 357.
- [13] M.J. Lewis, C.W. Bamforth, Sanitation and quality, in: Springer (Ed.) *Essays in Brewing Science*, 2006, p. 161.
- [14] D. Giacalone, M.B. Frøst, W.L.P. Bredie, D. Bénédicte Pineau b, C. Hunter, A.G. Paisley, M.K. Beresford, S.R. Jaeger, Situational appropriateness of beer is influenced by product familiarity, *Food Qual. Prefer.* 39 (2014) 16.
- [15] D.E. Briggs, C.A. Boulton, P.A. Brookes, R. Stevens, *Malts, adjuncts and supplementary enzymes in: brewing science and practice*, Woodhead Publishing Limited, Cambridge, 2004.
- [16] Á.A. Arrieta, M.L. Rodríguez-Méndez, J.A.d. Saja, C.A. Blanco, D. Nimubona, Prediction of bitterness and alcoholic strength in beer using an electronic tongue, *Food Chem.* 123 (2010) 642.
- [17] A. Rudnitskaya, E. Polshin, D. Kirsanov, J. Lammertyn, B. Nicolai, D. Saison, F.R. Delvaux, F. Delvaux, A. Legin, Instrumental measurement of beer taste attributes using an electronic tongue, *Anal. Chim. Acta* 646 (2009) 111.
- [18] M. Ghasemi-Varnamkhasti, S.S. Mohtasebi, M.L. Rodríguez-Méndez, M. Siadat, H. Ahmadi, S.H. Razavi, Electronic and bioelectronic tongues, two promising analytical tools for the quality evaluation of non alcoholic beer, *Trends Food Sci. Technol.* 22 (2011) 245.
- [19] E. Polshin, A. Rudnitskaya, D. Kirsanov, A. Legin, D. Saison, F. Delvaux, F.R. Delvaux, B.M. Nicolai, J. Lammertyn, Electronic tongue as a screening tool for rapid analysis of beer, *Talanta* (2010) 88.
- [20] M. Ghasemi-Varnamkhasti, S.S. Mohtasebi, M.L. Rodríguez-Méndez, J. Lozano, S.H. Razavi, H. Ahmadi, Potential application of electronic nose technology in brewery, *Trends Food Sci. Technol.* 22 (2011) 164.
- [21] A. Lamagna, S. Reich, D. Rodríguez, N.N. Scoccola, Performance of an e-nose in hops classification, *Sensors Actuators B Chem.* 102 (2004) 278.
- [22] J.A. Ragazzo-Sanchez, P. Chaliier, D. Chevalier-Lucia, M. Calderon-Santoyo, C. Ghommidh, Off-flavours detection in alcoholic beverages by electronic nose coupled to GC, *Sensors Actuators B Chem.* 140 (2009) 29.
- [23] H. Kojima, S. Araki, H. Kaneda, M. Takashio, Application of a new electronic nose with fingerprint mass spectrometry to brewing, *J. Am. Soc. Brew. Chem.* 63 (2005) 151.
- [24] Z. Tao, W. Lei, J. Teng, Pattern recognition of the universal electronic nose, Second International Symposium on Intelligent Information Technology Application, 2008, p. 249.
- [25] E. Sikorska, J. Chmielewski, T. Gorecki, I. Khmelinskii, M. Sikorski, D.D. Keukeleire, Discrimination of beer flavours by analysis of volatiles using the mass spectrometer as an electronic nose, *J. Inst. Brew.* 113 (2007) 110.
- [26] S. Grassi, J.M. Amigo, C.B. Lyndgaard, R. Foschino, E. Casiraghi, Beer fermentation: monitoring of process parameters by FT-NIR and multivariate data analysis, *Food Chem.* 155 (2014) 279.
- [27] G. McLeod, K. Clelland, H. Tapp, E.K. Kemsley, R.H. Wilson, G. Poulter, et al., A comparison of variate pre-selection methods for use in partial least squares regression: a case study on NIR spectroscopy applied to monitoring beer fermentation, *J. Food Eng.* 90 (2009) 300.
- [28] I.F. Duarte, A. Barros, C. Almeida, M. Spraul, A.M. Gil, Multivariate analysis of NMR and FTIR Data as a potential tool for the quality control of beer, *J. Agric. Food Chem.* 52 (2004) 1031.
- [29] X. Cetó, M. Gutiérrez-Capitán, D. Calvo, M.d. Valle, Beer classification by means of a potentiometric electronic tongue, *Food Chem.* 141 (2013) 2533.
- [30] P. Ciosek, W. Wróblewski, The recognition of beer with flow-through sensor array based on miniaturized solid-state electrodes, *Talanta* 69 (2006) 1156.
- [31] A.G. Mignania, L. Ciaccheria, A.A. Mencaglia, H. Ottevaere, E.E.S. Bábac, H. Thienpont, Optical measurements and pattern-recognition techniques for identifying the characteristics of beer and distinguishing Belgian beers, *Sensors Actuators B Chem.* 179 (2013) 140.
- [32] J.A. Rodrigues, A.S. Barros, B. Carvalho, T. Brandão, A.M. Gil, A.C.S. Ferreira, Evaluation of beer deterioration by gas chromatography-mass spectrometry/multivariate analysis: a rapid tool for assessing beer composition, *J. Chromatogr. A* 1218 (2011) 990.
- [33] M. Ghasemi-Varnamkhasti, S.S. Mohtasebi, M.L. Rodríguez-Méndez, A.A. Gomes, M.C.U. Araújo, R.K.H. Galvão, Screening analysis of beer ageing using near infrared spectroscopy and the Successive Projections Algorithm for variable selection, *Talanta* 30 (2012) 286.
- [34] M. Ghasemi-Varnamkhasti, M.L. Rodríguez-Méndez, S.S. Mohtasebi, C. Apetrei, J. Lozano, H. Ahmadi, S.H. Razavi, J.A.d. Saja, Monitoring the aging of beers using a bio-electronic tongue, *Food Control* 25 (2012) 216.
- [35] M. Ghasemi-Varnamkhasti, S.S. Mohtasebi, M. Siadat, J. Lozano, H. Ahmadi, S.H. Razavi, A. Dicko, Aging fingerprint characterization of beer using electronic nose, *Sensors Actuators B Chem.* 159 (2011) 51.
- [36] D. Saison, D.P.D. Schutter, F. Delvaux, F.R. Delvaux, Optimisation of a complete method for the analysis of volatiles involved in the flavour stability of beer by solid-phase microextraction in combination with gas chromatography and mass spectrometry, *J. Chromatogr. A* 1190 (2008) 342.
- [37] S.E. Stein, NIST standard reference database, in: P.J.La.W.G. Mallard (Ed.), *National Institute of Standards and Technology*, Gaithersburg, 2008.
- [38] H. Martens, T. Naes, *Multivariate calibration*, Wiley, Chichester, 1989.
- [39] J.E. Jackson, *A user's guide to principal components*, Wiley, New York, 1991.
- [40] R.A. Johnson, D.W. Wichern, *Applied multivariate statistical analysis*, Prentice Hall, Englewood Cliffs, 1992.
- [41] W. Ku, R.H. Storer, C. Georgakis, Disturbance detection and isolation by dynamic principal component analysis, *Chemom. Intell. Lab. Syst.* 30 (1995) 179.
- [42] T.J. Rato, M.S. Reis, Advantage of using decorrelated residuals in dynamic principal component analysis for monitoring large-scale systems, *Ind. Eng. Chem. Res.* 52 (2013) 13685.
- [43] H. Hotelling, Selected techniques of statistical analysis, in: C. Eisenhart, M.W. Hastay, W.A. Wallis (Eds.), *Multivariate quality control*, illustrated by the air testing of sample bombsights, McGraw-Hill, New-York, 1947.
- [44] J.E. Jackson, Quality control methods for several related variables, *Technometrics* 1 (1959) 359.
- [45] J.E. Jackson, G.S. Mudholkar, Control procedures for residuals associated with principal component analysis, *Technometrics* 21 (1979) 341.
- [46] J.F. MacGregor, T. Kourti, Statistical process control of multivariate processes, *Control. Eng. Pract.* 3 (1995) 403.
- [47] T.T. Kourti, J.F. MacGregor, Process analysis, monitoring and diagnosis, using multivariate projection methods, *Chemom. Intell. Lab. Syst.* 28 (1995) 3.
- [48] J.V. Kresta, J.F. MacGregor, T.E. Marlin, Multivariate statistical monitoring of process operating performance, *Can. J. Chem. Eng.* 69 (1991) 35.
- [49] B.W. Wise, N.B. Gallagher, The process chemometrics approach to process monitoring and fault detection, *J. Process Control* 6 (1996) 329.
- [50] T. Kourti, Application of latent variable methods to process control and multivariate statistical process control in industry, *Int. J. Adapt. Control* 213 (2005).
- [51] W.R. Dillon, M. Goldstein, *Multivariate analysis – methods and applications*, Wiley, New York, 1984.
- [52] J. Han, M. Kamber, *Data mining – concepts and techniques*, 2001.

- [53] R.A. Johnson, D.W. Wichern, Applied multivariate statistical analysis, 5th ed. Prentice Hall, Upper Saddle River, 2002.
- [54] T. Kourti, J.F. MacGregor, Multivariate SPC methods for process and product monitoring, *J. Qual. Technol.* 28 (1996) 409.
- [55] J.A. Westerhuis, S.P. Gurden, A.K. Smilde, Generalized contribution plots in multivariate statistical process monitoring, *Chemom. Intell. Lab. Syst.* 51 (2000) 95.
- [56] W. Kunze, Beer production – fermentation, maturation and filtration, *Technology Brewing and Malting*, VLB Berlin, Verlagsabteilung, Berlin, 1999, p. 323.
- [57] H.H. Jeleń, K. Wlazły, E. Wąsowicz, E. Kamiński, Solid-phase microextraction for the analysis of some alcohols and esters in beer: comparison with static headspace method, *J. Agric. Food Chem.* 46 (1998) 1469.
- [58] D. Saison, D.P.D. Schutter, B. Uyttenhove, F. Delvaux, F.R. Delvaux, Contribution of staling compounds to the aged flavour of lager beer by studying their flavour thresholds, *Food Chem.* 114 (2009) 1206–1215.
- [59] J.F. Clapperton, D.G.W. Brown, Caprylic flavour as a feature of beer flavour, *J. Inst. Brew.* 84 (1978) 90.
- [60] M. Li, Z. Yang, M. Yang, L. Shan, J. Dong, Determination of furfural in beer by high-performance liquid chromatography with solid-phase extraction, *J. Inst. Brew.* 115 (2009) 226.