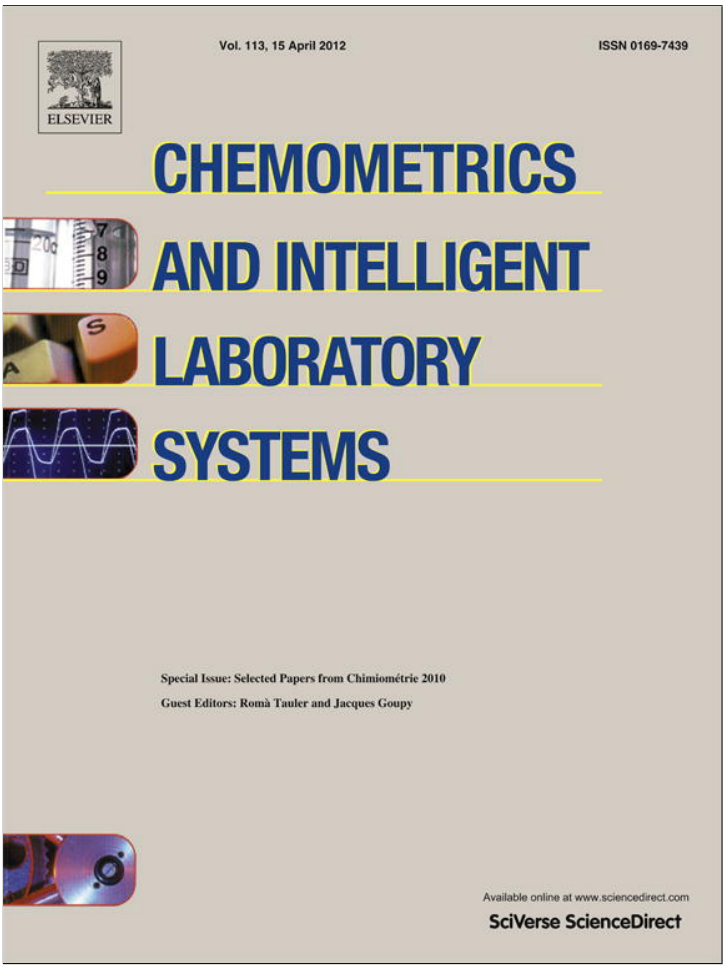

Chemometrics and Intelligent Laboratory Systems 113 (2012) 2–9

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

# Trappist beer identification by vibrational spectroscopy: A chemometric challenge posed at the 'Chimiométrie 2010' congress

Juan Antonio Fernández Pierna [a,*], Ludovic Duponchel [b], Cyril Ruckebusch [b], Dominique Bertrand [c], Vincent Baeten [a], Pierre Dardenne [a]

[a] Walloon Agricultural Research Centre, Valorisation of Agricultural Products Department, 24 Chaussée de Namur, 5030 Gembloux, Belgium
[b] Université Nord de France Lille1, LASIR CNRS F-59655 Villeneuve d'Ascq, France
[c] Institut National de la Recherche Agronomique, Equipe Bioinformatique, Rue de la Géraudière-BP 71627, 44316 Nantes, France

ARTICLE INFO

*Article history:*
Received 25 February 2011
Received in revised form 7 April 2011
Accepted 12 April 2011
Available online 19 April 2011

*Keywords:*
Trappist
Beer
Chemometrics
Spectroscopy
NIR
MIR
Raman
Challenge

ABSTRACT

A chemometric challenge was posed at the annual 'Chimiométrie' congress organized by the French Chemometrics Society in 2010. The congress was held in Paris on 2–3 December and the data relating to the challenge are available on the congress website (http://www.chimiometrie.fr/chemom2010). The aim of the challenge was to test the ability of congress participants to discriminate Trappist Rochefort 8° beer (one of seven authentic Trappist beers in the world) from other Trappist beers, as well as other special beers. All the beers were measured using three vibrational spectroscopic techniques (NIR, MIR and Raman). Three participants took up the challenge and, as in previous congress challenges, the organizing committee asked them to describe their approaches to face the challenge. This paper summarizes the three approaches put forward by participants, as well as the approach put forward by the organizers.



## 1. Introduction

Trappist monks brew beer to provide for their needs and to earn the income necessary for abbey operations and charity work [1]. In order to better protect the name, the International Trappist Association specified certain criteria for defining an Authentic Trappist Product. According to these criteria, the beer should be brewed at a Trappist abbey, by or under the control of Trappist monks. The brewery, the brewing procedures and the commercial factors involved depend upon the monastic community, and the economic purpose of the brewery is oriented towards providing for the monks' needs and social assistance activities, rather than financial profit. Currently, there are seven breweries that are allowed to use the Authentic Trappist Product logo on their products, indicating their authentic Trappist origin. Six of these beers are from Belgium (Achel, Chimay, Orval, Rochefort, Westmalle and Westvleteren) and one is from The Netherlands (La Trappe).

Until now, only a few studies have focused on the authenticity of beer products [2–8]. The published work concerns mainly on the detection of fraud, the authentication of the beer ingredients and the authentication of beer brands, using such techniques as gas chromatography–isotope ratio mass spectrometry [2], head-space solid-phase microextraction (HS–SPME) [3], isotope analysis [4] and gas chromatography–time-of-flight mass spectrometry (GC–TOFMS) [5]. Recently, three papers have been published on the confirmation of brand identity and the authentication of Belgian Trappist beers by LC–MS [6], UV spectroscopy [7] and near infrared (NIR) transflectance spectroscopy [8], respectively.

As in previous years [9–13], a challenge was posed at the 'Chimiometrie' congress held in Paris in December 2010, concerning the applicability of spectroscopy and multivariate analysis for guaranteeing the authenticity of a particular brand of beer. In particular, it focused on the discrimination of Trappist Rochefort 8° beer (one of seven authentic Trappist beers) from other Trappist beers, as well as other selected special beers. It involved using three vibrational spectroscopic techniques to measure several beers from a study conducted within the framework of the EU TRACE project (www.trace.eu.org). Spectroscopy is an analytical technique based on the interaction between a species and electromagnetic radiation. The electromagnetic radiation that is absorbed, emitted or scattered by the molecule is then analyzed. In vibrational absorption spectroscopy, by varying the frequency of the radiation, a spectrum can be produced that indicates the intensity of the exiting radiation for each frequency.

* Corresponding author.
*E-mail address:* fernandez@cra.wallonie.be (J.A. Fernández Pierna).

0169-7439/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2011.04.005

This spectrum shows which frequencies of radiation have been absorbed by the molecule to raise it to higher vibrational energy states. Each molecule has its own characteristic spectrum, making spectroscopy indispensible in analytical chemistry.

The three vibrational techniques applied in the challenge were near infrared (NIR), mid-infrared (MIR) and Raman spectroscopy. From a chemical point of view, both IR and Raman spectroscopy are based on the vibrational transitions that occur in the ground electronic state of molecules. Raman spectroscopy concerns the scattering of radiation by the sample, rather than an absorption process. Raman scattering arises from the changes in the polarisability or shape of the electron distribution in the molecule as it vibrates; in contrast, IR absorption requires a change in the intrinsic dipole moment with the molecular vibration. Although the mechanism of Raman scattering differs from that of IR absorption, Raman and IR spectra provide complementary information about the vibrations of molecules. Typical applications of vibrational spectroscopy techniques are in structure determination, multicomponent qualitative and quantitative analysis.

The aim of the challenge at the 'Chimiometrie' congress was to guarantee the authenticity of Rochefort 8° beer through multivariate analysis; that is, to discriminate as accurately as possible between Rochefort 8° and the rest of the beers included in the datasets, and then to predict the blind spectra included in an available test set.

## 2. Material and methods

### 2.1. Instrumentation

Table 1 shows the characteristics of the instruments used in the study, including the number of spectra obtained by sample, the mode of measurement, the background used and the acquisition parameters (resolution, number of scans and laser power).

### 2.2. Datasets

All the samples used in the study came from the TRACE project in which a complete experimental design was used. For the congress challenge, a set of 130 bottles of Trappist and non-Trappist beers was used. All the bottles were measured randomly on the same day and each beer was measured twice with each instrument (NIR, MIR and Raman), in line with the parameters shown in Table 1. From this set, 100 beers were selected for the calibration set and given to the congress participants, with information about the samples; the remaining 30 beers were selected for the blind test set. For the calibration data set, three data files of 200 spectra (corresponding to the 100 samples measured in duplicate) with information about the samples were provided, corresponding to each of the spectroscopic techniques used (see Figs. 1–3). Table 2 shows the list of beers used for the calibration set, including the number of bottles for each beer. Because the aim was to discriminate between Rochefort 8° and the rest of the beers, there were more bottles of this beer. For the test set, three data files of 60 spectra were provided (corresponding to the 30 samples measured in duplicate), but without any information about the samples.

**Table 1**
Characteristics of the different spectral data obtained from the analysis of the samples.

| | NIRS | MRI | Raman |
|---|---|---|---|
| Instrument | XDS (Foss) | Vertex 70 (Bruker) | Vertex 70 (Bruker) |
| Number of spectra per sample | 2 spectra/sample | 3 spectra/sample | 2 spectra/sample |
| Background | Internal reference | Pure water | NA |
| Mode of measurement | Transflexion | Reflectance | 180° geometry |
| Sample holder | transflexion cell | Diamond ATR accesory | test tube |
| Acquisition parameter | | | |
| Resolution: | 2 nm | 4 $cm^{-1}$ | 4 $cm^{-1}$ |
| Number of scans: | 16 | 64 | 128 |
| Range: | 400–2498 nm | 4000–600 $cm^{-1}$ | 3500–100 $cm^{-1}$ |
| Laser power: | NA | NA | 600 mW |

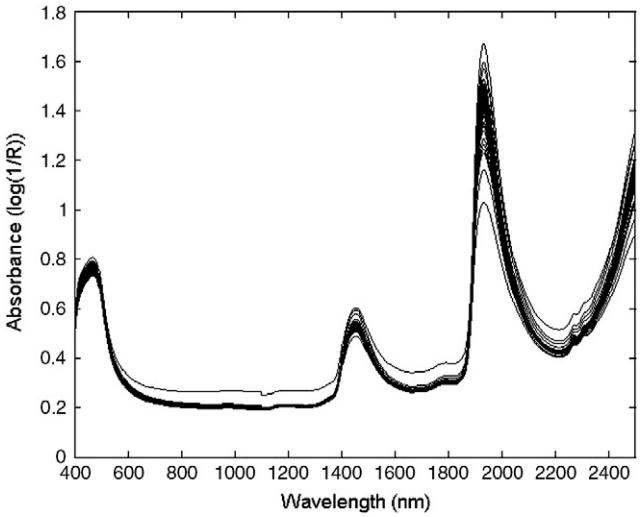


**Fig. 1.** NIR data: calibration set.

## 3. Results

### 3.1. Participant N° 1

From an initial analysis of the calibration dataset it was clear that the number of Rochefort 8° samples was very low in comparison with the total number of beers (only 9%). Therefore, to build a Rochefort 8°/non-Rochefort 8° classification model would be risky. This imbalance in numbers would lead to overtraining whatever classification and cross-validation methods were used. In addition, it was not possible to develop a good classification model without a validation step in order to check its accuracy. Even the calibration dataset would probably have to be split, thus further reducing again the number of Rochefort 8° cases.

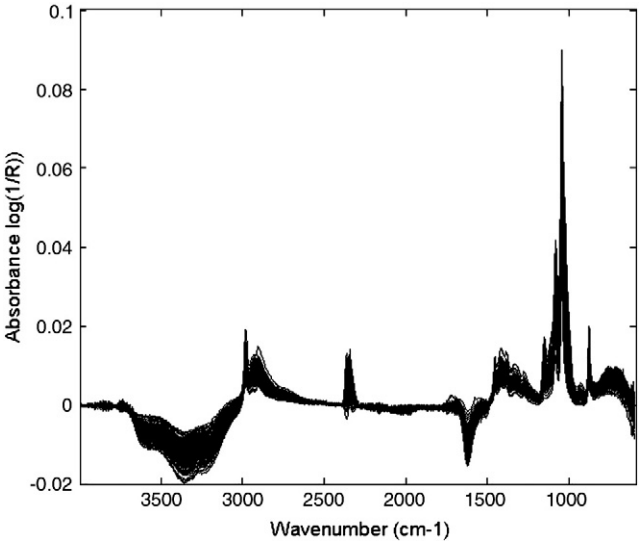


**Fig. 2.** MIR data: calibration set.

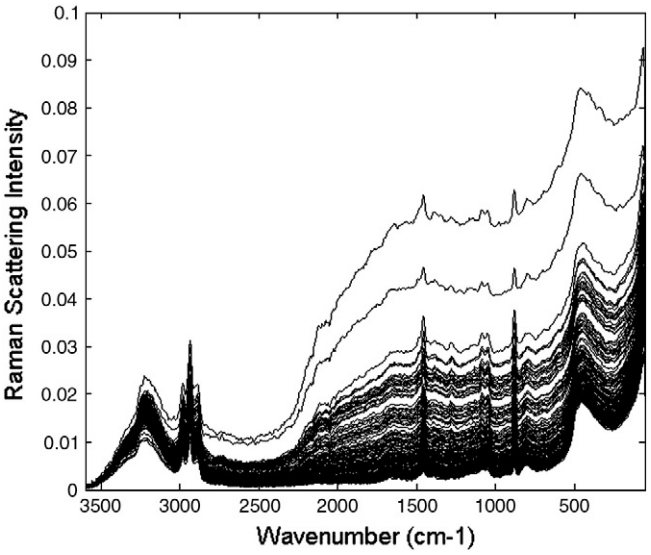


**Fig. 3.** Raman data: calibration set.

The aim was therefore to find some criteria from the spectroscopic data in order to classify a maximum number of samples as non-Rochefort 8°. This was an artificial way to increase the relative number of Rochefort 8° samples in the calibration dataset. It was possible that interlinked classifications would solve this problem, discriminating first the non-Trappist beers from the Trappist ones, then the non-Rochefort beers from the Rochefort ones, and finally the non-Rochefort 8° beers from the Rochefort 8° ones. The chemical composition of the beers in the study was 80–90% water, 5–11.3% alcohol, a low percentage of sugar and less than 1% protein. Given a specific alcohol spectral contribution, it might be possible to classify Rochefort 8° beers because of the high alcohol levels due to successive fermentations. The problem was, however, that Rochefort 8° beers do not have the highest alcohol content (only 9.2%). The classification would therefore have to use information about minor compounds. This might be close to the detection limits of the proposed spectroscopy techniques, and it would therefore be necessary to find a good pre-processing method in order to identify these small variances.

It was decided to use Principal Component Analysis (PCA) to prepare the classification models. It is a very good visualization tool for multivariate datasets. Therefore, based on the principle that the classification had to be based on minor compounds, it was necessary first to study the spectroscopy technique with the highest spectral specificity (i.e., the lowest bandwidth) as well as the highest sensitivity. Mid-Infrared and Raman spectroscopy were both good candidates. When looking at the data, it was easy to observe a lower signal to noise ratio in Raman spectroscopy than in Mid-Infrared. A high fluorescence background was also observed, which masked the Raman scattering. PCA was then applied to the MIR dataset. The aim was to find a factorial plane that roughly presented clusters for the non-Trappist/Trappist groups, the non-Rochefort/Rochefort groups and the non-Rochefort 8°/Rochefort 8° groups. Many pre-processing combinations were tested, but without any interesting results. Various selections of spectral domains were also investigated, such as the well-known fingerprint region of the MIR domain. Considering the bad signal to noise ratio of the Raman dataset, the focus was then on the NIR spectral domain. Fig. 1 shows the two well-known absorption bands of water (the major compounds) centered approximately on 1450 nm and 1930 nm. At this stage, both the visible range and the water absorption domains were removed in order to prevent overlap with other variables needed for the classification. Again, the aim was to find clusters with numerous factorial planes and various pre-processing combinations, but the results were the same as previously.

The Raman dataset was then analyzed. Derivative and baseline correction algorithms were explored in order to try to remove the fluorescence background contribution, but once again no interesting factorial planes were found. As Raman scattering is very sensitive to laser instability, a spectral normalization was only tested directly on the raw data (i.e. without background removal). Fig. 4 presents the first factorial plane (PC1 *vs.* PC3) with its Hotelling confidence ellipse where defined zones for Rochefort 8° beers were evident. When looking at the loadings, the first PCs showed that the fluorescence background provided important information for Rochefort 8° beer classification. The blind test samples were then projected onto the selected plane (Fig. 5).

Thus, it was possible to declare, with near certainty, that the test samples outside the rectangular zone were non-Rochefort 8° beers. A decision had to be taken about the 14 remaining test samples within the zone. From this subspace, there was a new distribution of calibration samples consisting of fourteen Rochefort 8° beers and eight non-Rochefort 8° beers (four Binchoise Brune and four Maredsous). It was assumed that the color of the beers could be used to classify them, even if all the samples of this sub-population were amber. PCA was then performed with samples belonging to the previous box zone on the visible spectral range. Fig. 6 presents a factorial plane (PC2 *vs.* PC3) for a first derivative used as spectral pre-processing. Well-defined clusters were observed for Rochefort 8° and non-Rochefort 8° beers. At this stage, the fourteen remaining

**Table 2**
List of the beers used for the calibration set.

| Trappist beers | Number of bottles | Non-trappist beers | Number of bottles | | Number of bottles |
|---|---|---|---|---|---|
| Achel Blond | 2 | Affligem Triple | 1 | Jupiler | 2 |
| Achel Brune | 2 | Binchoise Brune | 2 | Kwak | 2 |
| Chimay Bleu | 1 | Bon Secoures | 2 | Leffe Blonde | 2 |
| Chimay Rouge | 2 | Brugges Triple | 2 | Leffe Brune | 2 |
| Chimay Triple | 2 | Delirium Tremens | 2 | Maredsous 8 | 2 |
| La Trappe Blance | 2 | Duvel | 2 | Moinette Blonde | 2 |
| La Trappe Blonde | 2 | Gueuze Girardin | 2 | Moinette Brune | 2 |
| La Trappe Double | 2 | Gouden Carolus | 1 | Primator 22 | 1 |
| La Trappe Quadrupple | 1 | Grimbergem Double | 2 | Primator 24 | 1 |
| La Trappe Triple | 2 | Grimbergem Triple | 2 | Quintine | 2 |
| Orval | 2 | Hapkin | 1 | St Bernardus Prior | 2 |
| Rochefort 10 | 9 | Hercule | 1 | St Feuillien | 1 |
| Rochefort 6 | 5 | Het Kapitel Watou | 2 | Triple Karmeliet | 2 |
| Rochefort 8 | 9 | Hoegarden Grand cru | 2 | Val-Dieu Triple | 2 |
| Westmalle Double | 2 | Hotteuse Grand cru | 2 | | |
| Westvleteren | 3 | Judas | 1 | | |
| Total bottles | 48 | | | | 52 |

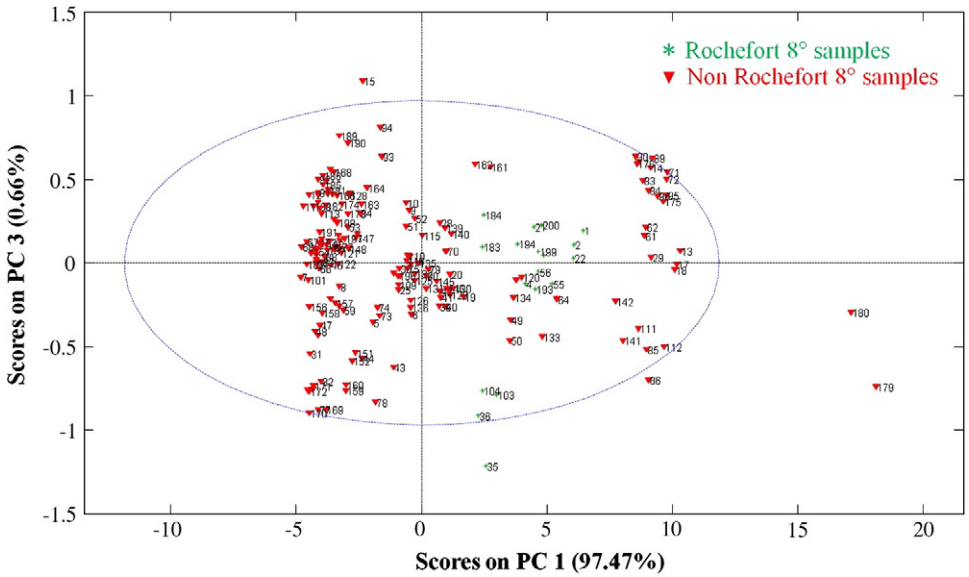


**Fig. 4.** PC1 vs. PC3 plot for the NIR data.

samples from the blind test set were unambiguously declared to be Rochefort 8° beers.

In conclusion, it would have been possible to go further in this classification challenge with the proposed hierarchical approach. Even if no real classification model was developed, the interpretation of sequential PCA applied to different spectroscopic techniques succeeded in providing solutions to meet the challenge.

### 3.2. Participant N° 2

Considering that only 9% of the samples are Rochefort 8°, the first objective was to try to better define the classification problem by reducing the calibration dataset, excluding the non-Rochefort 8° samples that were found to play no relevant role in the classification process (e.g., those that were far from the boundary of the class). After this initial step, it was expected that, ideally, there would be two classes of about the same size. In order to reduce the calibration dataset, the data were analyzed mainly by performing PCAs on individual Raman, NIR and MIR calibration data. The second step involved applying supervised classification methods to the reduced dataset, either Soft Independent Modeling of Class Analogies (SIMCA) [14] or High-Dimensional Discriminant Analysis (HDDA) [15]. The first method, which incorporates dimensionality reduction in a classification model by conducting disjoint PCA analyses for each group, is a standard chemometric classification tool. In contrast, the second method, HDDA, has only recently been introduced and applied in chemometrics [16]. It is a parametric method for clustering and classification, based on Gaussian mixture models for high-dimensional data. HDDA offers an alternative way of dealing with the specific problems related to high-dimensional data. In particular, it has two main advantages over other generative classification methods. First, the formulation of the inverse covariance matrix is explicit. Second, it is possible to build the classifier when the number of learning observations (samples) is smaller than the number of variables (dimension).

PCA was applied on MIR and Raman data and a rapid inspection of the results showed clear clustering of the Rochefort data in both cases.

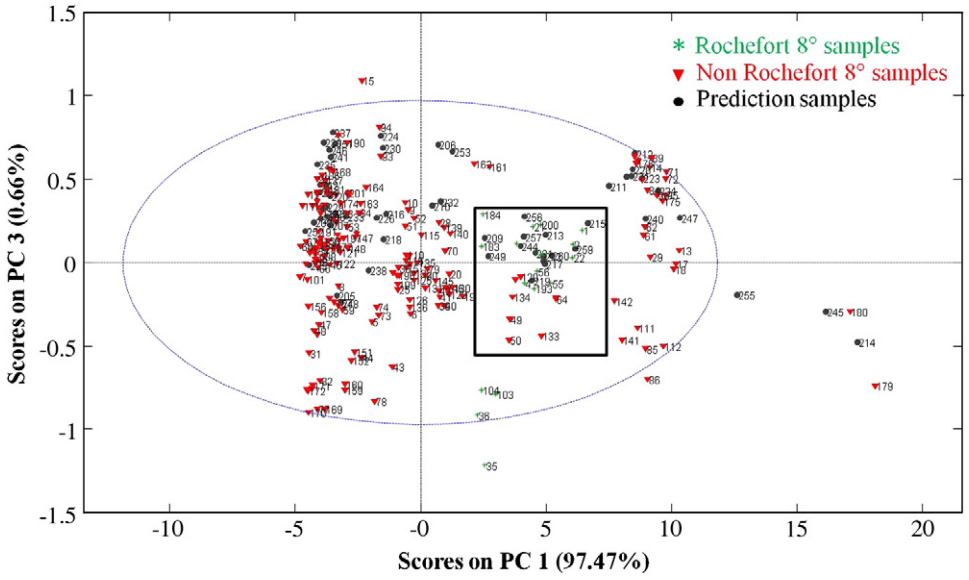


**Fig. 5.** Projection of the blind test samples onto the PC1 *vs.* PC3 plane for the NIR data.

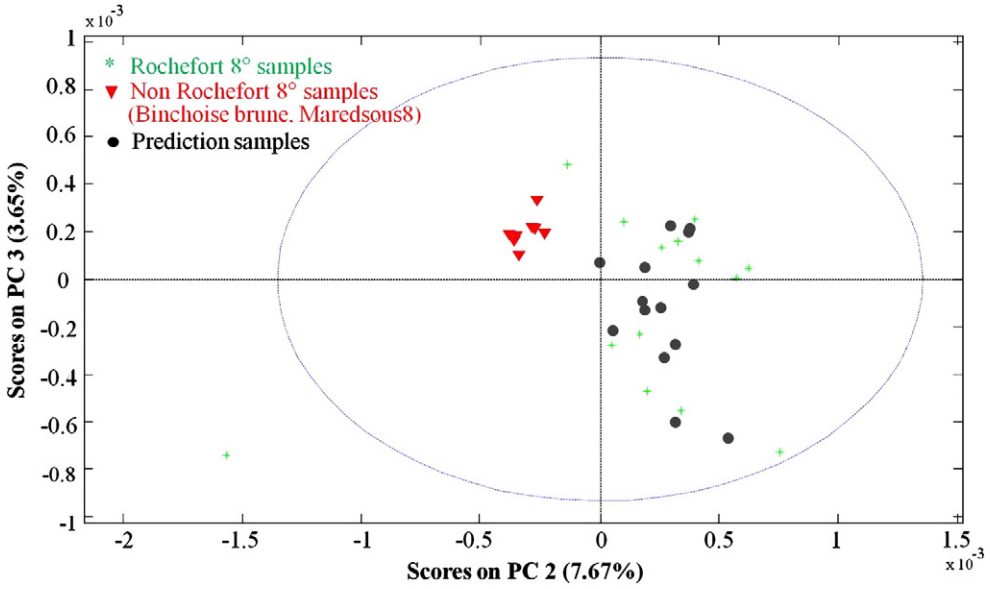


**Fig. 6.** PC2 *vs.* PC3 plot for the Raman data after the first derivative.

A deeper investigation led to the following conclusions: (i) all the Rochefort 8° calibration samples had positive PC3 scores with the MIR calibration data and (ii) all of them had positive PC1 scores with Raman calibration data.

The combination of findings (i) and (ii) meant that 161 samples could be excluded from the calibration (i.e., all the samples that did not fulfill the conditions with regard to the PC scores). The reduced calibration set consisted of 39 samples, of which 20 were Rochefort 8°. Fig. 7 was obtained by building a PCA model on this 'new' reduced calibration set (results shown for MIR data only).

The first objective had been reached in that there were now two groups of about the same size. The reliability of this approach should be treated with some caution, however, particularly because it was based on the assumption that the calibration and test data showed quite good overlap, without much extrapolation.

Once this clear group separation had been observed in the PCs subspace, the SIMCA and HDDA methods were applied to the reduced calibration MIR dataset. Validation was performed using cross-validation in both methods. For HDDA, the Bayesian informative criterion (BIC) was also used for model selection [16]. The main result was that a 100% accurate classification rate was obtained in cross-validation in both cases. With regard to the prediction ability of both models for the test samples, 12 samples were predicted as Rochefort 8° using the SIMCA model, whereas only 11 were predicted with HDDA (the same samples, except for one). It should be noted that with the SIMCA model, classification prediction was based on the projection of the samples into the models built around each class. Both methods gave reliable and consistent results, but this does not guarantee accurate group-membership predictions if the assumptions made are not true enough. Considering the information available for the challenge, no conclusion can be reached about the prediction of these particular samples and it was thus decided to put forward the prediction results obtained with HDDA. In addition, to roughly evaluate the risk taken by applying the proposed procedure, which actually consists in "sample pruning", it was found interesting to project the test set into the PCA subspaces obtained previously for the full calibration set. It was observed that quite strong extrapolation is required for the prediction and the classification of some samples, for which the prediction should be regarded more cautiously.

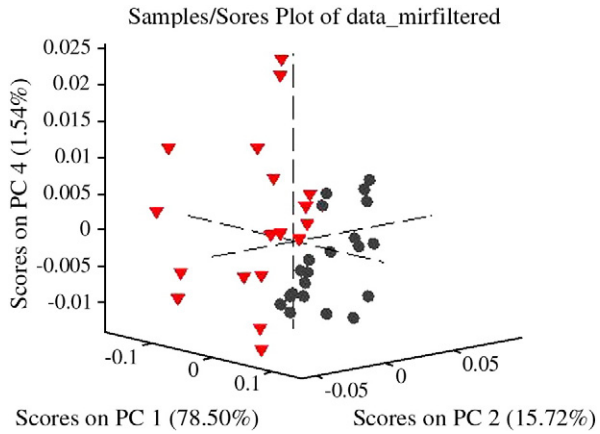


**Fig. 7.** PCA of MIR calibration data reduced by applying rules (i) and (ii), see text for details.

### 3.3. Participant N° 3

In this approach, the repetitions were considered as independent tables. An initial examination of the data tables led to various conclusions:

The vector correlation (or RV) coefficients between instrument data matrices were computed [17] (Table 3). These coefficients indicate consensus between pairs of tables. They take the value of 1 where the tables provide the same (linear) type of information. A value of 0 indicates that the tables cannot be rotated in order to have superimposed observations (as in the procrustean approaches). It therefore seemed inappropriate to use a multi-table approach such as those developed by Tenenhaus and Hanafi [18].

The spectral repetitions were not very stable.

The qualitative groups seemed to form complex 'clouds' in a multidimensional space.

**Table 3**
RV coefficients between the (pre-treated) data tables (the RV or vector correlation coefficients evaluate the 'similarity' between tables).

| | Raman | MIR | NIR |
|---|---|---|---|
| Raman | 1 | 0.04 | 0.13 |
| MIR | 0.04 | 1 | 0.04 |
| NIR | 0.13 | 0.04 | 1 |

As an initial step, the data were pre-treated. The choice of the spectral pre-treatments used for each technique was drawn from previous Analysis of Variance (ANOVA) studies. MIR and Raman spectra were corrected by using the standard normal variate (SNV) method, and NIR spectra were corrected using a second derivative. Then, in order to build the classification models, the k-nearest neighbor (kNN) approach was used. With this approach, a new sample was projected into the space created by the calibration dataset, and then a set of distance calculations was done; this sample was in the same class as most of the pre-defined number of neighbors in the dataset. The approach involved constructing independent models for each of the six tables (3 methods x 2 replicates). It was necessary to select the variables involved in the kNN discrimination. For this purpose, a simple forward introduction of the variables was followed. For a fixed number of neighbors, the variable giving the best discrimination in the calibration set was identified and selected. A second variable was then introduced. Once a variable had been introduced, there was no attempt to remove the previously introduced ones (as in a stepwise approach). The neighbors were computed using a type of Mahalanobis distance. This involved applying PCA to the matrices of the selected variables. A subset of the scores associated with the highest eigenvalues was then defined. By dividing each score by the corresponding standard deviation (reduction), it was possible to build a new data matrix. The Euclidian distances on this new matrix were equivalent to the Mahalanobis distances on the original dataset.

Several models for discriminating the samples have been tested with neighbors ranging from 1 to 6 and PCA-dimensions from 1 to 10. For each table among the 6 ones available, there were thus $10\times60$ discriminant models. These models were independently tested on the data of the blind test. As there were 60 observations in the blind test, the eventual results could be gathered in 6 matrices $T^{(i)}$ with $i=1\ldots6$. An element $t_{jk}^{(i)}$ of the matrix $T^{(i)}$ took the value 1 if the application of the discriminant model numbered $k$ (among 60 available) led to attribute the nature Rochefort 8° to the observation $j$, and 0 otherwise. Fig. 8 illustrates these predictions (with the observations in rows, and the discriminant models in columns). A black color indicates that the observation was not classified in the Rochefort 8° group. Gray and white colors indicate that the tested discriminant model led to a positive classification. White color gives the eventual decision based on the agreement between instruments, while gray one indicates that there is no consensus.

### 3.4. Challenge organizers' approach

The organizers' approach was based on using ensemble classifiers constructed by the supervised classification algorithm Support Vector Machines (SVM), where the important point is the decision rule, which determines whether a new sample belongs to a predicted group or not. Like other supervised classifiers, a training dataset is needed to define the decision boundaries within the feature space, and based upon this the classification decision rules are made. This technique works according to the principle of structural risk minimization, which is controlled by maximizing the margin between the training data and the decision boundary [19,20]. SVM projects, in a non-linear way, the training data from the original space to a feature space of higher dimension through the use of a kernel function. In this higher dimensional space, the decision boundary becomes linear in order to separate the groups of interest and is equivalent to placing a non-linear separator in the original space. The Gaussian radial basis is commonly used as a kernel function and SVM then has only two parameters to optimize: the width of the Gaussian and the C parameter in relation to the minimization of the classification errors. A good balance has to be found between both parameters in order to guarantee a maximization of the margin and a minimization of the misclassified points. In order to perform an accurate optimization of

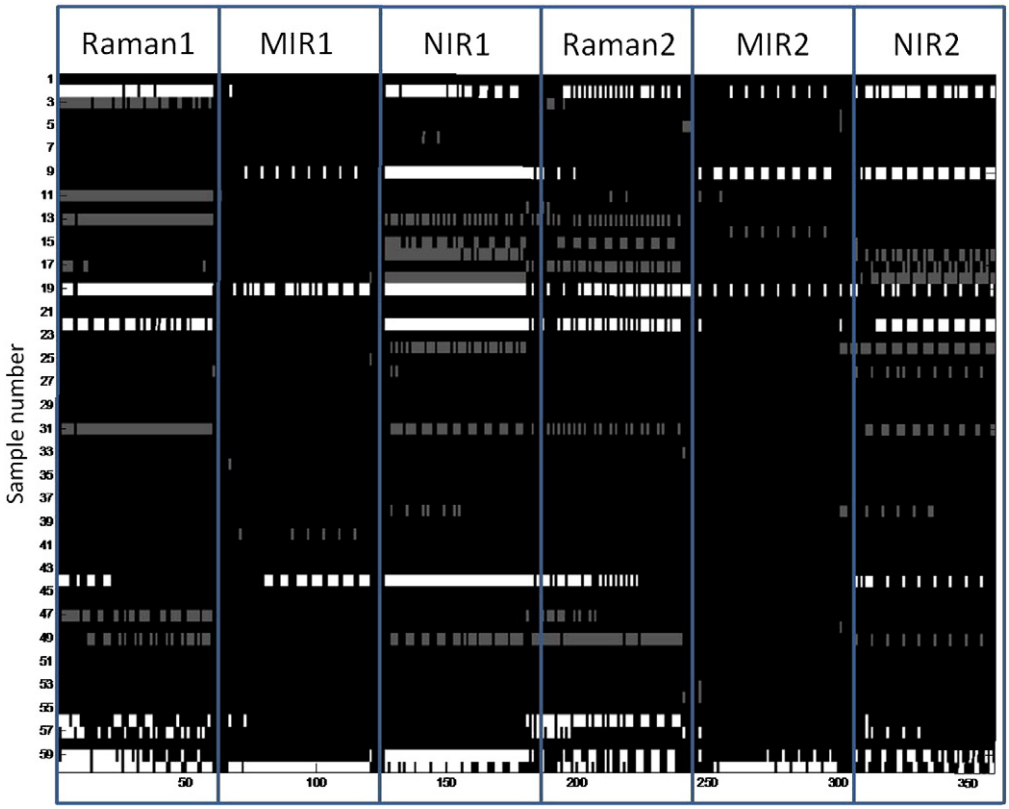


**Fig. 8.** Predictions for the blind test set. Rows: sample number. Columns: the discriminant models. Black: sample not classified in the Rochefort 8° group. Gray: Rochefort 8° detected, but no consensus between instruments. White: Rochefort 8° detected, with consensus.

**Table 4**
Results for all the participants expressed in terms of success rates and number of samples (between brackets), as well as the sensitivity and specificity values (in %).

| | Participant 1 | | Participant 2 | | Participant 3 | | Challenge organizer | |
|---|---|---|---|---|---|---|---|---|
| | Rochefort 8° | Other | Rochefort 8° | Other | Rochefort 8° | Other | Rochefort 8° | Other |
| Rochefort 8° | 100% (14) | 0% | 71% (10) | 29% (4) | 64% (9) | 36% (5) | 86% (12) | 14% (2) |
| Other | 0% | 100% (46) | 2% (1) | 98% (45) | 0% | 100% (46) | 0% | 100% (46) |
| Overall classification rate | 100% | | 84.5% | | 82.0% | | 93.0% | |
| Sensitivity | 100.0% | | 84.5% | | 82.0% | | 93.0 | |
| Specificity | 100.0% | | 97.8% | | 100.0% | | 100.0% | |

| | | Rochefort 8° | Other |
|---|---|---|---|
| Sensitivity = a/(a + c) | Rochefort 8° | a | c |
| Specificity = d/(b + d) | Other | b | d |

the parameters, all the datasets (NIR, MIR and Raman) were split into three subsets: a calibration dataset to build the model, a stop set to optimize the SVM parameters, and a test set to validate the models. Once the optimal parameters had been selected, the final model was constructed by combining the calibration and stop sets. The three sets of blind samples were then predicted using the ensemble classifier technique. The idea behind this technique is to produce a final classification result that is dependent upon a pool of classifiers instead of on a single equation. As explained by Parikh and Palikir, combining classifiers ensures diversity because each classifier will make a different error, and by combining them the total error can be reduced [21,22]. In this work, having samples measured in different instruments to train different classifiers ensured diversity. Each individual NIR, MIR or Raman blind spectrum then went through its equation and the corresponding prediction class was obtained. For each blind sample, three predictions corresponding to each spectroscopic technique were obtained. The final results for the new samples were based on a combination rule through the simple majority voting ensemble technique. This majority voting strategy constructs multiple decision terms to estimate the performance of the decision function.

In this case, an individual SVM model between Rochefort 8° and non-Rochefort 8° beers was constructed for each spectroscopic dataset (NIR, MIR and Raman) and the prediction of the blind samples was performed as previously explained. The NIR data were pre-processed by SNV and first derivative SG (9:2:1), the MIR data had no pre-processing and the peak of $CO_2$ (around 2400 $cm^{-1}$) was removed, and for the Raman data the best results were obtained without pre-processing.

## 4. Final results

All the results presented were obtained taking into account only individual spectroscopic datasets by proposing either the use of individual discrimination models or the use of sequential PCA models. But in all the cases, the evaluation of the approaches was conducted on the basis of the best assignment of the 60 beers in the blind test set. Table 4 shows the success rates expressed as the percentage of correctly classified samples, as well as the sensitivity and specificity (in %) obtained for the blind test set. The success rate was defined as:

$$success\ rate = \frac{\left(\sum_{i=1}^{K} \frac{correctly\ classified\ samples\ in\ class\ i}{total\ number\ of\ samples\ in\ class\ i}\right)}{K} \times 100$$

with $K$ being the number of classes. A success rate of, for example, 92 indicated that 92% of the objects are correctly classified.

In this study, the sensitivity was defined as the probability of Rochefort 8° beers correctly identified as such, while the specificity measured the probability of non-Rochefort 8° beers that were correctly identified as such [23].

In all cases, good sensitivity and specificity values were obtained. The results indicated that Participant 1 obtained a sensitivity and specificity of 100% (i.e. all the samples were correctly classified as Rochefort 8° or non-Rochefort 8°). Participant 2 had a specificity of 97.8%, and participant 3 and the challenge organizers had a specificity of 100%, but in all cases they had lower sensitivity than Participant 1, indicating the presence of a number of false negative results, which should be avoided. In terms of the overall classification rate, participant 1 has 100% of samples correctly classified, followed by the challenge organizer with 93% and participants 2 and 3 with 84.5% and 82% respectively. The perfect results obtained by participant 1 could be explained by the fact that they are based on the use of PCA exclusively, which is a non-parametric approach and which is valid only for this dataset.

## 5. Conclusion

The main feature of the challenge was the use of three vibrational spectroscopy techniques (NIR, MIR and Raman) to characterize the beers. This is a rather unusual situation when working on classification problems and some questions need to be asked. How should all this spectral information be used? Is it necessary to use the three spectral datasets at the same time, or even two of them? Is it necessary to develop one classification model per spectroscopic technique? Should only one of these models be chosen or is it possible to obtain predictions working with several models? All the participants in the challenge took into account the individual spectroscopic datasets by proposing the use of individual discrimination models or the use of sequential PCA models. In view of the results, and of the very few papers on this topic, it is difficult to answer these questions yet. Nevertheless, they will need to be addressed because such datasets will soon become increasingly common.

## Acknowledgments

We thank the European Commission, through the Sixth Framework Programme under the Food Quality and Safety Priority, part of the TRACE project (Integrated Project 006942) (www.trace.eu.org), for funding the samples included in this paper, Ouissam Abbas for her helpful work and discussions on spectroscopic aspects, and the CRA-W technical team for the analyses.

We would also like to thank all the participants who spent time analyzing the data and presenting their results.

The information contained in this article reflects the authors' views; the European Commission is not liable for any use of the information contained therein.

## References


[1] International Trappist Association, available at, http://www.trappist.com via the Internet ,accessed 01 February 2011.
[2] G. Calderone, C. Guillou, F. Reniero, N. Naulet, Helping to authenticate sparkling drinks with 13 C/12 C of $CO^2$ by gas chromatography–isotope ratio mass spectrometry, Food Research International 40 (2007) 324–331.
[3] G.A. da Silva, F. Augusto, R.J. Poppi, Exploratory analysis of the volatile profile of beers by HS–SPME–GC, Food Chemistry 111 (4) (2008) 1057–1063.
[4] W. Papesch, M. Horacek, Forensic applications of stable isotope analysis: case studies of the origins of water in mislabeled beer and contaminated diesel fuel, Science & Justice 49 (2009) 138–141.
[5] T. Cajka, K. Riddellova, M. Tomaniova, J. Hajslova, Recognition of beer brand based on multivariate analysis of volatile fingerprint, Journal of Chromatography A 1217 (25) (2010) 4195–4203.
[6] E. Mattarucchi, M. Stocchero, J.M. Moreno-Rojas, G. Giordano, F. Reniero, C. Guillou, Authentication of Trappist beers by LC–MS fingerprints and multivariate data analysis, Journal of Agricultural and Food Chemistry 58 (23) (2010) 12089–12095.
[7] J. Weeranantanaphan, G. Downey, Identity confirmation of a branded, fermented, cereal product by UV spectroscopy: a feasibility study involving a Trappist beer, Journal of the Institute Brewing 116 (1) (2010) 56–61.
[8] V. Di Egidio, P. Oliveri, T. Woodcock, G. Downey, Confirmation of brand identity in foods by near infrared transflectance spectroscopy using classification and class-modelling chemometric techniques—the example of a Belgian beer, Food Research International 44 (2011) 544–549.
[9] P. Dardenne, J.A. Fernández Pierna, A NIR data set is the object of a chemometric contest at 'Chimiométrie 2004', Chemometrics Intelligent Laboratory Systems 80 (2006) 236–242.
[10] J.A. Fernández Pierna, P. Dardenne, Chemometric contest at 'Chimiométrie 2005': a discrimination study, Chemometrics Intelligent Laboratory Systems 86 (2007) 219–223.
[11] J.A. Fernández Pierna, P. Dardenne, Soil parameter quantification by NIRS as a chemometric challenge at 'Chimiométrie 2006', Chemometrics Intelligent Laboratory Systems 91 (2008) 94–98.
[12] J.A. Fernández Pierna, F. Chauchard, S. Preys, J.M. Roger, O. Galtier, V. Baeten, P. Dardenne, How to build a robust model with only a few reference values: a chemometric challenge at 'Chimiométrie 2007', Chemometrics and Intelligent Laboratory Systems 106 (2011) 152–159, doi:10.1016/j.chemolab.2010.05.015.
[13] J.A. Fernández Pierna, H. Duval, P. Valderrama, D.N. Rutledge, V. Baeten, P. Dardenne, A case study of extrapolation in NIR modelling – A chemometric challenge at 'Chimiométrie 2009', Chemometrics and Intelligent Laboratory Systems 106 (2011) 205–209, doi:10.1016/j.chemolab.2010.04.001.
[14] S. Wold, Pattern recognition by means of disjoint principal component models, Pattern Recognition 8 (1976) 127–139.
[15] C. Bouveyron, S. Girard, C. Schmid, High dimensional discriminant analysis. Communications in statistics, Theory Methods 36 (14) (2007) 2607–2623.
[16] J. Jacques, C. Bouveyron, S. Girard, O. Devos, L. Duponchel, C. Ruckebusch, Journal of Chemometrics 24 (2010) 719–727.
[17] C. Lavit, Analyse conjointe de tableaux quantitatifs, Masson Ed, Paris, 1988.
[18] M. Tenenhaus, M. Hanafi, A bridge between PLS path modeling and multi-block data analysis, Handbook of Partial Least Squares. Springer Handbooks Comp. Statistics, Part 1, 2010, pp. 99–123.
[19] A. Gidudu, G. Hulley, T. Marwala, Image classification using SVMs: one-against-one vs. one-against-all, In Proceedings of the 28th Asian Conference of Remote Sensing, Kuala Lumpur, Malaysia, Nov. 12–16 2007.
[20] J.A. Fernández Pierna, V. Baeten, A. Michotte Renier, R.P. Cogdill, P. Dardenne, Combination of support vector machines (SVM) and near infrared (NIR) imaging spectroscopy for the detection of meat and bone meat (MBM) in compound feeds, Journal of Chemometrics 18 (2004) 341–349.
[21] D. Parikh, R. Polikar, An ensemble based incremental learning approach to data fusion, IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics 37 (2) (2007) 437–500.
[22] R. Polikar, Ensemble based systems in decision making, IEEE Circuits Systems Magazine 6 (3) (2006) 21–45.
[23] S. Spitalnic, Test properties I: sensitivity, specificity, and predictive values, Hospital Physician 40 (9) (2004) 27–31.