



Screening analysis of beer ageing using near infrared spectroscopy and the Successive Projections Algorithm for variable selection

Mahdi Ghasemi-Varnamkhasti^{a,b,*}, Seyed Saied Mohtasebi^c, Maria Luz Rodriguez-Mendez^a, Adriano A. Gomes^d, Mario Cesar Ugulino Araújo^d, Roberto K.H. Galvão^e

^a *Inorganic Chemistry Department, E.T.S Ingenieros Industriales, University of Valladolid, Valladolid, Spain*

^b *Department of Agricultural Machinery Engineering, Faculty of Agriculture, Shahrekord University, Shahrekord, Iran*

^c *Agricultural Machinery Engineering Department, Faculty of Agricultural Engineering & Technology, University of Tehran, Karaj, Iran*

^d *Universidade Federal da Paraíba, CCEN, Departamento de Química, Caixa Postal 5093, CEP 58051-970 João Pessoa, PB, Brazil*

^e *Instituto Tecnológico de Aeronáutica, Divisão de Engenharia Eletrônica, 12228-900, São José dos Campos, SP, Brazil*

ARTICLE INFO

Article history:

Received 30 August 2011

Received in revised form

30 November 2011

Accepted 6 December 2011

Available online 19 December 2011

Keywords:

Beer

Ageing

Near infrared spectroscopy

Wavelength selection

Classification

Successive Projections Algorithm

Linear Discriminant Analysis

ABSTRACT

This work proposes a method for monitoring the ageing of beer using near-infrared (NIR) spectroscopy and chemometrics classification tools. For this purpose, the Successive Projections Algorithm (SPA) is used to select spectral variables for construction of Linear Discriminant Analysis (LDA) classification models. A total of 83 alcoholic and non-alcoholic beer samples packaged in bottles and cans were examined. To simulate a long storage period, some of the samples were stored in an oven at 40 °C, in the dark, during intervals of 10 and 20 days. The NIR spectrum of these samples in the range 12,500–5405 cm⁻¹ was then compared against those of the fresh samples. The results of a Principal Component Analysis (PCA) indicated that the alcoholic beer samples could be clearly discriminated with respect to ageing stage (fresh, 10-day or 20-day forced ageing). However, such discrimination was not apparent for the non-alcoholic samples. These findings were corroborated by a classification study using Soft Independent Modelling of Class Analogy (SIMCA). In contrast, the use of SPA-LDA provided good results for both types of beer (only one misclassified sample) by using a single wavenumber in each case, namely 5550 cm⁻¹ for non-alcoholic samples and 7228 cm⁻¹ for alcoholic samples.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

An important challenge in the brewing industry consists of minimizing changes in the quality of beer from production to consumption stages. In this context, beer ageing is considered to be a major quality problem since the ageing flavours are mostly experienced as unpleasant [1]. Therefore, flavour stability over the ageing process has become one of the most important topics in brewing research over the past few years [2–4].

Beer flavour is conventionally assessed through the combination of common analytical tools (e.g. gas chromatography) and organoleptic profiling panels known as human sensory panels. However, these expensive and time-consuming methods may not be appropriate for on-line monitoring in breweries. In addition, organoleptic panels are prone to assessor fatigue and subjectivity,

which may compromise the accuracy and reproducibility of the results.

Within this scope, one of the most promising directions for the development of innovative solutions is the use of spectroscopic methods. In fact, the speed and on-line capabilities of such methods meet the trends of automation and continuous processing in the brewing industry. Nuclear magnetic resonance (NMR) spectroscopy, for example, can be used to monitor the change in chemical composition of beers during the period of storage [5]. However, the cost associated to this technique may be prohibitive for routine use in the production line. Alternatively, the application of vibrational spectroscopy, such as near infrared (NIR), may constitute a less costly approach to characterize changes in the organic compounds involved in beer ageing. The joint use of NIR spectroscopy and chemometrics techniques has been reported in the literature in several applications of quality control and analysis of fuel [6], vegetable oil [7], cigarettes [8], and biodiesel [9], for example.

The NIR range extends from 13,300 to 4000 cm⁻¹ and comprises overtones and combinations of fundamental vibrational transitions that occur in the middle infrared region [10]. As a result, a NIR

* Corresponding author at: Department of Agricultural Machinery Engineering, Faculty of Agriculture, Shahrekord University, Shahrekord, Iran.
Tel.: +98 3814424403; fax: +98 3814424428.

E-mail address: ghasemymahdi@gmail.com (M. Ghasemi-Varnamkhasti).

spectrum is formed by several overlapping bands, and thus the use of multivariate analysis tools is usually required. For classification purposes, methods based on Principal Component Analysis (PCA), such as Soft Independent Modelling of Class Analogy (SIMCA), can be employed to exploit the spectrum in the entire working range [11–13]. As an alternative, variable selection methods can be used to identify specific spectral variables that convey useful information for the analytical problem at hand.

Variable selection may have several advantages, such as removal of noise and nonlinearity, as compared to using the full spectrum [14]. Algorithms described in the literature for selection of variables in chemical data include the Genetic Algorithm [15], Simulated Annealing [16], Tabu Search [17] and Colony of Ants [18] among others. In this context, Araújo et al. proposed the Successive Projections Algorithm (SPA) for selection of variables in multiple linear regression (MLR) [19]. In a subsequent work, Pontes et al. proposed a modification to SPA so that it could also be applied to classification problems in conjunction with Linear Discriminant Analysis (LDA) models [20].

Although studies on spectroscopic analysis of beer have been reported in the literature [21–27], this approach has not yet been pursued for the characterization of ageing. The objective of the present investigation consists of using NIR spectroscopy together with SPA-LDA for screening analysis of ageing in alcoholic and non-alcoholic beers. For comparison, full-spectrum SIMCA models are also employed.

2. The Successive Projections Algorithm for Linear Discriminant Analysis

The SPA-LDA algorithm is aimed at selecting a subset of variables with small collinearity and suitable discriminating power for use in classification problems involving $C \geq 2$ different classes. For this purpose, it is assumed that a training set of N objects with known class labels is available to guide the variable selection process. In the case of spectroscopic data, for example, each object consists of a spectrum recorded over K wavenumbers (or wavelengths).

The SPA-LDA algorithm can be divided into two main phases. In Phase 1, the N training objects are centered on the mean of their respective classes and stacked in the form of a matrix \mathbf{X} ($N \times K$). Each column of \mathbf{X} is associated to a variable (i.e. a wavenumber in the case of spectroscopic data). Projection operations involving the columns of \mathbf{X} are then carried out to form K chains of L variables, where $L = \min(K, N - C)$. Each chain is initialized with one of the K available variables. Subsequent variables in the chain are selected in order to display the smallest collinearity (as assessed by the projection operations) with the preceding ones.

In Phase 2, different subsets of variables are extracted from the chains and then evaluated in terms of a cost function G defined as

$$G = \frac{1}{N} \sum_{n=1}^N g_n \quad (1)$$

where

$$g_n = \frac{r^2[\mathbf{x}_n, \bar{\mathbf{x}}(I_n)]}{\min_{I_j \neq I_n} r^2[\mathbf{x}_n, \bar{\mathbf{x}}(I_j)]} \quad (2)$$

In Eq. (2), the numerator $r^2[\mathbf{x}_n, \bar{\mathbf{x}}(I_n)]$ is the squared Mahalanobis distance [28] between object \mathbf{x}_n (of class index I_n) and the sample mean $\bar{\mathbf{x}}(I_n)$ of its true class (both row vectors). This distance is calculated as

$$r^2[\mathbf{x}_n, \bar{\mathbf{x}}(I_n)] = [\mathbf{x}_n - \bar{\mathbf{x}}(I_n)]\mathbf{S}^{-1}[\mathbf{x}_n - \bar{\mathbf{x}}(I_n)]^T \quad (3)$$

Table 1
Sample set partitioning for SIMCA and SPA-LDA.

Class	Alcoholic samples		Non-alcoholic samples	
	Training	Test	Training	Test
N	7	7	7	7
F1	7	7	7	7
F2	7	7	7	6
Total	21	21	21	20

where \mathbf{S} is a pooled covariance matrix, which is calculated in accordance with the standard LDA procedure [29]. The denominator in Eq. (2) corresponds to the squared Mahalanobis distance between object \mathbf{x}_n and the center of the nearest wrong class. A small value of g_n indicates that \mathbf{x}_n is close to the center of its true class and distant from the centers of the remaining classes. Therefore, minimizing the cost G results in a better separation of the objects according to their true classes.

SPA-LDA has been employed in several applications involving spectroscopic data, such as the classification of vegetable oils [20] and coffee [30] by UV–vis spectroscopy, as well as soil samples by LIBS (laser-induced plasma breakdown spectroscopy) [31]. In the case of NIR spectroscopy, applications have included the classification of diesel samples according to sulphur content [20] and cigarettes with respect to brand [8].

3. Materials and methods

3.1. Samples

A total of 83 samples of regular beer (42 alcoholic and 41 non-alcoholic) packed in bottles and cans were used in this study. To simulate a long storage period, 55 samples were stored in an oven at 40 °C, in the dark, in the original closed containers (bottles and cans). This procedure is known as forced ageing in the literature [32–34]. After 10 days, 28 samples (termed class F1) were removed from the oven. The remaining 27 samples (termed class F2) were removed after an additional period of 10 days. The 28 fresh beer samples (i.e. those not subjected to forced ageing) were used to form the class of non-aged samples, termed N.

3.2. NIR measurements

The beer samples were prepared for NIR measurements by following standard procedures described in the literature [35]. The measurements were carried out with a Nicolet Magna-IR 760 NIR spectrometer equipped with a glass cell with optical path of 8 mm. The spectrum of each sample was acquired in the range 12,500–5405 cm^{-1} with a resolution of 4 cm^{-1} by coaveraging 500 scans. The same procedure was employed to acquire the spectrum of the background (water), which was subtracted from the sample spectra. The room temperature was kept constant at 25 °C throughout the process. The spectra were acquired after the beer samples reached thermal equilibrium with the room.

3.3. Software

In order to remove baseline features, the spectra were processed by using the first derivative Savitzky-Golay filter with window of 121 points and second-order polynomial. The resulting derivative spectra comprised 1800 spectral variables in the range 5521–8991 cm^{-1} . Principal Component Analysis was used for a preliminary evaluation of the discriminating power of the NIR spectrum in relation to the ageing of the beer samples. Subsequently, the data were partitioned into training and test sets

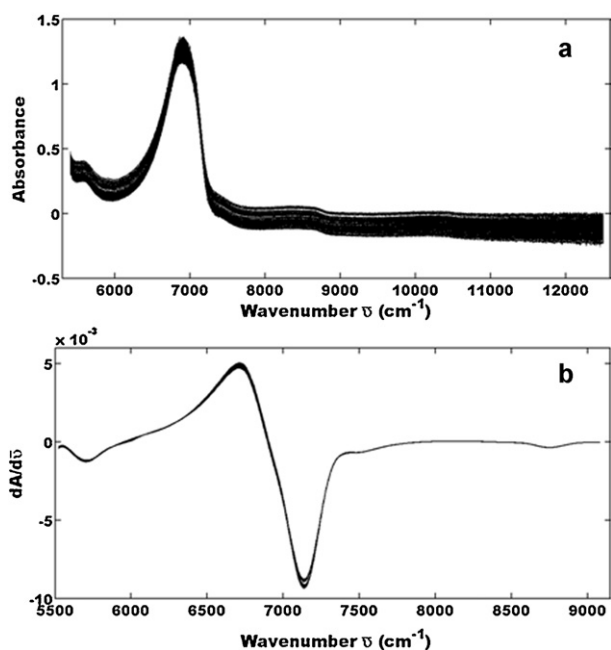


Fig. 1. Raw (a) and derivative (b) spectra of the 83 beer samples.

by using the classic Kennard–Stone algorithm [36] as indicated in Table 1. SPA-LDA and SIMCA models were built to classify the beer samples with respect to degradation time (classes N, F1, F2). Kennard–Stone and SPA-LDA calculations were carried out in Matlab 2010a, while Savitzky–Golay, PCA and SIMCA calculations were carried out in Unscrambler 9.7.

4. Results and discussion

4.1. NIR spectra

As seen in Fig. 1a, the raw spectra of the 83 beer samples with water background subtracted display high noise levels, as well as systematic variations of baseline. Such problems were corrected by using the first derivative Savitzky–Golay filter, which resulted in the derivative spectra shown in Fig. 1b. These spectra were employed throughout the study.

4.2. Principal Component Analysis

Fig. 2a presents the PCA score plot for the overall data set, with the percentage explained variance indicated at each axis. As can be seen, the beer samples with and without alcohol can be roughly discriminated along PC1, which indicates that the main source of variability within the spectral data is related to the OH ethanol band. In fact, according to the PC1 loadings plot in Fig. 2b, the largest loading corresponds to wavenumber 7135 cm^{-1} , which is associated to the first OH overtone [37]. It is worth noting that some alcoholic samples are close to the non-alcoholic ones in Fig. 2a. It may be argued that the ethanol in these alcoholic samples has been partly degraded during the forced ageing process [1,3], which would explain the proximity to the non-alcoholic samples. However, an analysis of the chemical composition of the samples (by using NMR, as in [5], for example), would be required to corroborate this explanation.

Due to the large influence of ethanol on the NIR spectra, the alcoholic and non-alcoholic groups were analyzed separately for the purpose of evaluating the ageing effects. Fig. 2c presents the new PCA score plot obtained for the group of alcoholic samples.

Table 2

Type-II SIMCA errors (training/test) obtained in the classification of the samples with respect to ageing.

Number of samples	Significance level		
	0.05	0.10	0.25
Alcoholic (21/21)	6/10	5/5	0/2
Non-alcoholic (21/20)	21/20	21/20	12/11

As can be seen, the samples can be separated along PC1 according to the ageing stage (N, F1, F2). According to Fig. 2d, the largest PC1 loading is found at 7155 cm^{-1} , which is also located in a spectral region associated to the first OH overtone [37]. It may thus be argued that ageing effects may be related to chemical reactions resulting in alterations of ethanol content, as discussed above.

The new PCA score plot for the non-alcoholic samples is presented in Fig. 2e. In this case, a separation according to ageing state is not observed. According to the PC1 loading plot in Fig. 2f, the main variability among the non-alcoholic sample spectra seems to be related to the first OH overtone. Such a finding suggests that these samples may have a residual ethanol content. However, the ethanol concentration is too small to cause changes in the NIR spectrum over the ageing process.

It is also worth noting that PC1 accounts for 81% of the explained variance in Fig. 2c (alcoholic samples) and only 38% of the explained variance in Fig. 2e (non-alcoholic samples). Such a finding suggests that changes in ethanol content are indeed the major source of spectral variability over the ageing process.

4.3. SIMCA classification

Full-spectrum SIMCA models were built for the three classes under consideration (N, F1 and F2). Full cross-validation using the training set was applied to determine the optimal number of principal components for each class. In the classification process, three significance levels were tested (0.05, 0.10, 0.25). It is worth noting that SIMCA classification errors can be of two types. A type-I error consists of a sample not included in its own class and a type-II error consists of a sample included in an incorrect class. All samples were correctly included in their respective classes, regardless of the significance level employed in the classification, i.e. no type-I errors occurred. However several samples were also included in an incorrect class. Table 2 presents the number of type-II errors in the training and test sets. Even in the best case, which corresponded to the significance level of 0.25, a considerable number of errors were obtained, especially for the non-alcoholic samples. This result is in agreement with the strong overlapping among the N, F1, F2 classes observed for the non-alcoholic samples in Fig. 2e.

4.4. SPA-LDA classification

Fig. 3 presents the cost function values obtained for the training set in the SPA-LDA variable selection process. As discussed in [20], the optimal number of variables corresponds to the minimum value of the cost, which can be interpreted as the risk of misclassification by the LDA model. Therefore, a single wavenumber was selected for the alcoholic samples and also for the non-alcoholic samples. The selected wavenumbers are shown in Fig. 4.

The SPA-LDA model correctly classified all alcoholic samples, in both the training and test sets, with respect to the ageing stage (N, F1, F2). As regards the non-alcoholic samples, only one test sample was incorrectly classified. These results can be interpreted by inspecting the univariate plots corresponding to the selected wavenumber in Fig. 5a (7228 cm^{-1} , alcoholic samples) and Fig. 5b (5550 cm^{-1} , non-alcoholic samples). The horizontal lines indicate the boundaries between classes, while the vertical line separates

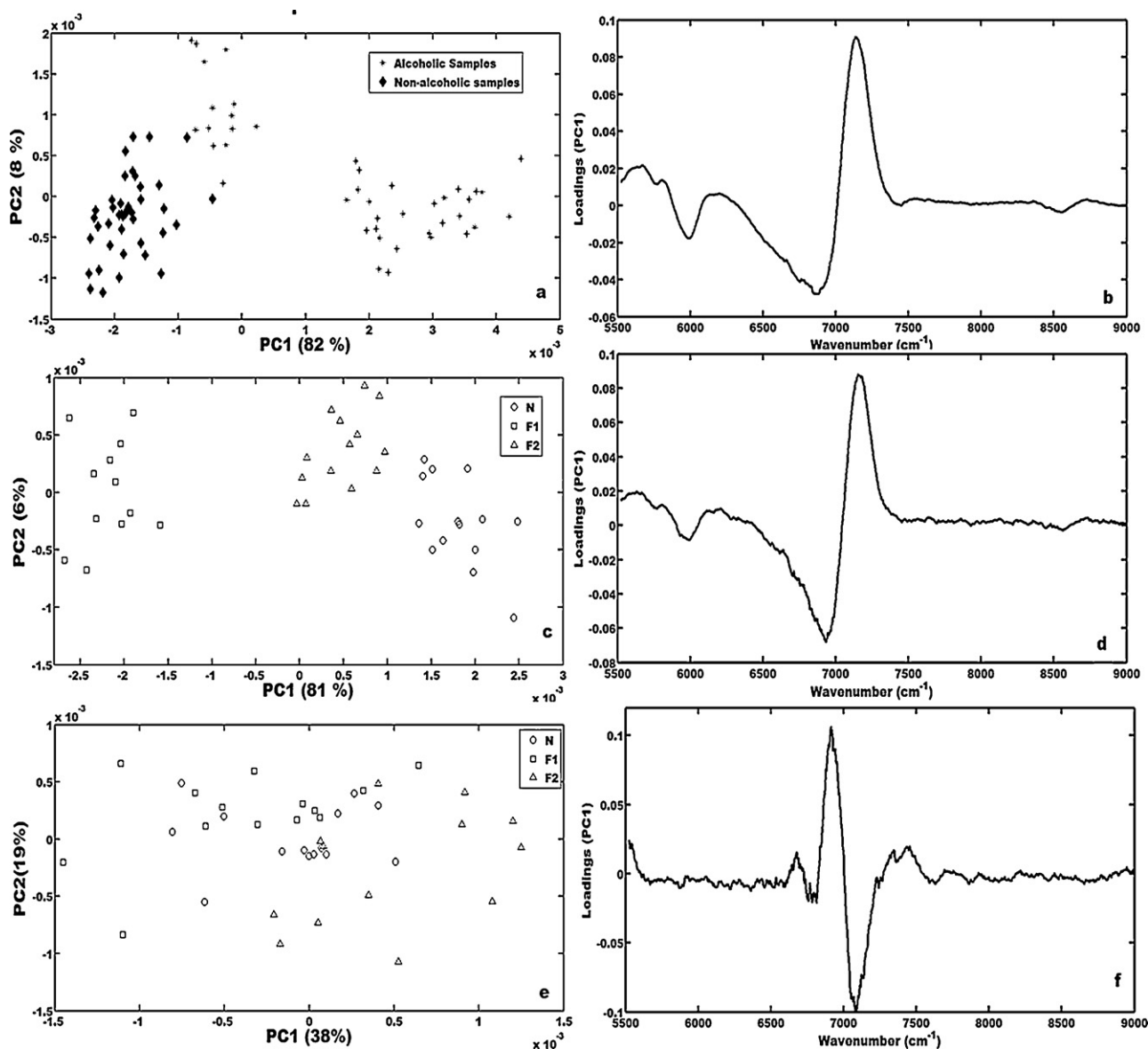


Fig. 2. Principal Component Analysis: score and loading plots for all 83 beer samples (a and b) for alcoholic samples (c and d) and for non-alcoholic samples (e and f).

the training and test sets. As can be seen, there is no overlap between the N, F1, F2 classes. The only sample that was incorrectly classified is indicated with a circle in Fig. 5b. Although the classification was incorrect in this case, the sample was close to the boundary of its true class.

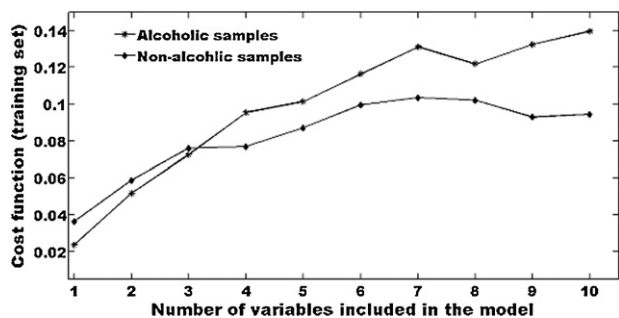


Fig. 3. Determination of the optimal number of variables in SPA-LDA.

The separation among the N, F1, F2 classes can also be visualized in Fig. 6a and b for the alcoholic and non-alcoholic samples, respectively. The insets in these figures present an expanded view of the derivative spectra around the wavenumbers selected by SPA-LDA.

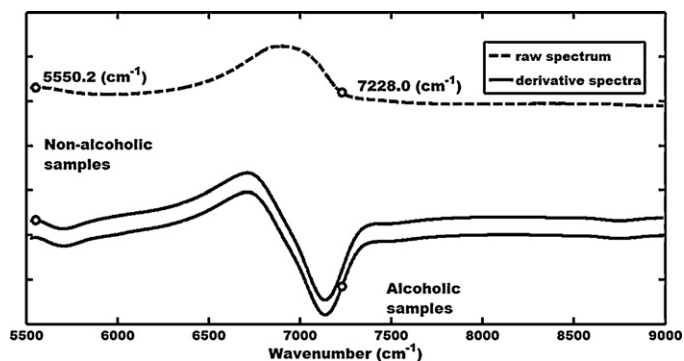


Fig. 4. Average raw and derivative spectrum of the overall data set with indication of the wavenumbers selected by SPA-LDA.

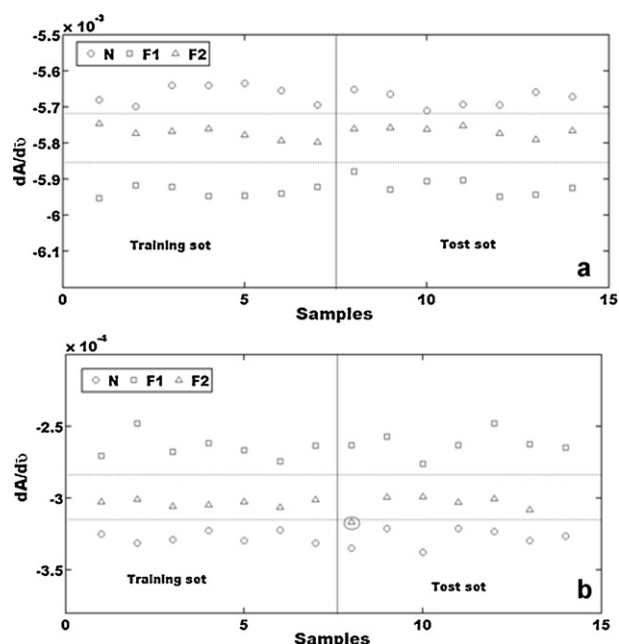


Fig. 5. Value corresponding to the derivative spectrum at the wavenumber selected by SPA-LDA for alcoholic (a) and non-alcoholic (b) samples.

For the alcoholic samples, the selected wavenumber (7228 cm^{-1}) is associated with the first OH overtone, which again suggests that the ageing process may be related to changes in ethanol content. In contrast, the wavenumber selected for the non-alcoholic samples (5550 cm^{-1}) is associated with the first SH overtone [37]. This finding may be ascribed to deterioration processes resulting in the formation of minor amounts of sulphur compounds. It is worth noting that such a separation among N, F1 and F2 classes was not apparent in the PCA score plot presented in Fig. 2e. In fact, when the overall spectrum is employed in PCA, the small alterations around 5550 cm^{-1} are masked by measurement noise and other sources of data variability that are unrelated to the ageing effects.

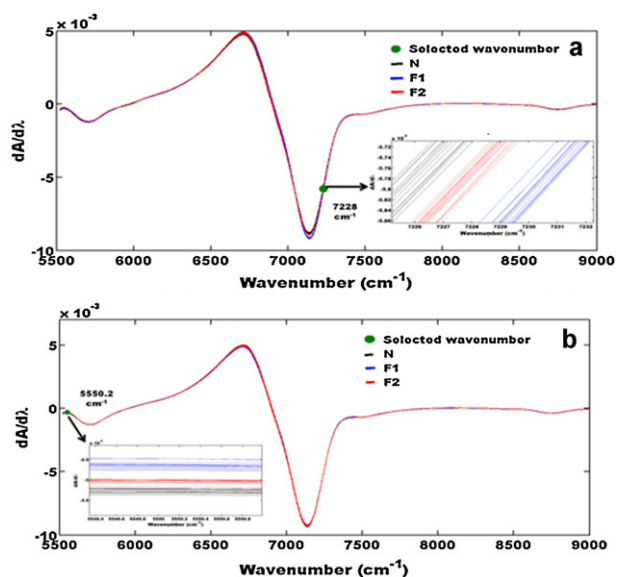


Fig. 6. A detailed picture of the NIR derivative spectra around the wavenumbers selected by SPA-LDA for alcoholic (a) and non-alcoholic (b) samples.

5. Conclusions

This paper proposed a methodology for screening analysis of beer ageing employing NIR spectrometry and Linear Discriminant Analysis coupled with the Successive Projections Algorithm for wavenumber selection. PCA results indicated that the NIR spectrum in the adopted working range ($12,500\text{--}5405\text{ cm}^{-1}$) provides a clear discrimination of alcoholic beer samples with respect to the ageing stage (fresh, 10-day or 20-day forced ageing). However, such a discrimination was not apparent for the non-alcoholic samples. The PCA findings were corroborated by a classification study using SIMCA models, which provided good results for the alcoholic samples, but were unable to discriminate the non-alcoholic samples. In contrast, the use of SPA-LDA provided appropriate results for both alcoholic and non-alcoholic beers by using only one wavenumber in each case.

It can be concluded that the use of NIR spectroscopy with chemometrics methods may provide valuable information about the ageing stage of beer.

Acknowledgments

The support of the Iran National Science Foundation (INSF) is gratefully appreciated regarding the finance of this research. Financial support by the Spanish Ministry of Science (grant AGL2009-12660/ALI) is gratefully acknowledged as well. Also, M. C. U. Araujo, R. H. K. Galvão and A. A. Gomes acknowledge the support of CNPq (research fellowships) and CAPES (scholarships).

References

- [1] B. Vanderhaegen, F. Delvaux, L. Daenen, H. Verachtert, F.R. Delvaux, *Food Chem.* 103 (2007) 404.
- [2] D.J. Evans, D.J.M. Schmedding, A. Bruijnje, T. Heideman, B.M. King, N.M. Groesbeek, *J. Inst. Brew.* 105 (1999) 301.
- [3] B. Vanderhaegen, H. Neven, H. Verachtert, G. Derdelinckx, *Food Chem.* 95 (2006) 357.
- [4] M. Ghasemi-Varnamkhasti, S.S. Mohtasebi, M.L. Rodriguez-Mendez, J. Lozano, S.H. Razavi, H. Ahmadi, *Trends Food Sci. Technol.* 22 (2011) 165.
- [5] J.A. Rodrigues, A.S. Barros, B. Carvalho, T. Brandão, A.M. Gil, *Anal. Chim. Acta* 702 (2011) 178.
- [6] M.C. Breikreitz, I.M. Raimundo Jr., J.J.R. Rohwedder, C. Pasquini, H.A.D. Filho, G.E. José, M.C.U. Araújo, *Analyst* 128 (2003) 1204.
- [7] A.F.C. Pereira, M.J.C. Pontes, F.F.G. Neto, S.R.B. Santos, R.K.H. Galvão, M.C.U. Araújo, *Food Res. Int.* 41 (2008) 341.
- [8] E.D.T. Moreira, M.J.C. Pontes, R.K.H. Galvão, M.C.U. Araújo, *Talanta* 79 (2009) 1260.
- [9] G. Veras, A.A. Gomes, A.C. Silva, A.L.B. Brito, P.B.A. Almeida, E.P. Medeiros, *Talanta* 83 (2010) 565.
- [10] C. Pasquini, *J. Braz. Chem. Soc.* 14 (2003) 198.
- [11] S. Wold, *Pattern Recogn.* 8 (1976) 127.
- [12] M. Daszykowski, J. Orzel, M.S. Wrobel, H. Czarnik-Matusewicz, B. Walczak, *Chemom. Intell. Lab. Syst.* 109 (2011) 86.
- [13] K.R. Beebe, R.J. Pell, B. Seasholtz, *Chemometrics – A Practical Guide*, Wiley, New York, 1998.
- [14] C.M. Andersen, R. Bro, *J. Chemom.* 24 (2010) 728.
- [15] H.C. Goicoechea, A.C. Olivieri, *J. Chemom.* 17 (2003) 338.
- [16] U. Horchner, J.H. Kalivas, *J. Chemom.* 9 (1995) 283.
- [17] J.A. Hageman, M. Streppel, R. Wehrens, L.M.C. Buydens, *J. Chemom.* 17 (2003) 427.
- [18] M. Shamsipura, V. Zare-Shahabadi, B. Hemmateenejad, M. Akhond, *QSAR Comb. Sci.* 28 (2009) 1263.
- [19] M.C.U. Araújo, T.C.B. Saldanha, R.K.H. Galvão, T. Yoneyama, H.C. Chame, V. Visani, *Chemom. Intell. Lab. Syst.* 57 (2001) 65.
- [20] M.J.C. Pontes, R.K.H. Galvão, M.C.U. Araújo, P.N.T. Moreira, O.D. Pessoa Neto, G.E. José, T.C.B. Saldanha, *Chemom. Intell. Lab. Syst.* 78 (2005) 11.
- [21] F. Inon, R. Liario, S. Garrigues, M. de la Guardia, *Anal. Bioanal. Chem.* 382 (2005) 1549.
- [22] A. Nordon, A. Mills, R.T. Burn, F.M. Cusick, D. Littlejohn, *Anal. Chim. Acta* 548 (2005) 148.
- [23] F. Inon, S. Garrigues, M. de la Guardia, *Anal. Chim. Acta* 571 (2006) 167.
- [24] B.M. Nicolai, K. Beullens, E. Bobelyn, A. Peirs, W. Saeys, K.I. Theron, K.I. Lamertyn, *Postharvest Biol. Technol.* 46 (2007) 99.
- [25] G. McLeod, K. Clelland, H. Tapp, E.K. Kemsley, R.H. Wilson, G. Poulter, D. Coombs, C.J. Hewitt, *J. Food Eng.* 90 (2009) 300.

- [26] H. Li, Y. Takahashi, M. Kumagai, K. Fujiwara, R. Kikuchi, N. Yoshimura, T. Amano, J. Lin, N. Ogawa, J. Near Infrared Spectrosc. 17 (2009) 238.
- [27] F. Liu, Y. Jjiang, Y. He, Anal. Chim. Acta 635 (2009) 45.
- [28] R. de Maesschalck, D. Jouan-Rimbaud, D.L. Massart, Chemom. Intell. Lab. Syst. 50 (2000) 1.
- [29] W. Wu, Y. Mallet, B. Walczak, W. Penninckx, D.L. Massart, S. Heuerding, F. Erni, Chemom. Intell. Lab. Syst. 329 (1996) 257.
- [30] U.T.C.P. Souto, M.J.C. Pontes, E.C. Silva, R.K.H. Galvão, M.C.U. Araújo, F.A.C. Sanches, F.A.S. Cunha, M.S.R. Oliveira, Food Chem. 119 (2010) 368.
- [31] M.J.C. Pontes, J. Cortez, R.K.H. Galvão, C. Pasquini, M.C.U. Araújo, R.M. Coelho, M.K. Chiba, M.F. Abreu, B.E. Madari, Anal. Chim. Acta 642 (2009) 12.
- [32] M.T. Walters, A.P. Heasman, P.S. Hughes, J. Am. Soc. Brew. Chem. 55 (1997) 83.
- [33] L.F. Guido, N.A. Fortunato, J.A. Rodrigues, A. Barros, J. Agric. Food Chem. 51 (2003) 3911.
- [34] J.A. Rodrigues, A.S. Barros, B. Carvalho, T. Brandao, A.M. Gil, A.C.S. Ferreira, J. Chromatogr. A 1218 (2011) 990.
- [35] K. Siebert, P.Y. Lynn, J. Am. Soc. Brew. Chem. 65 (2007) 229.
- [36] R.K.H. Galvão, M.C.U. Araújo, G.E. José, M.J.C. Pontes, E.C. Silva, T.C.B. Saldanha, Talanta 67 (2005) 736.
- [37] Z. Xiaobo, Z. Jiewen, M.J.W. Povey, M. Holmes, M. Hanpin, Anal. Chim. Acta 667 (2010) 14.