Food Chemistry 119 (2010) 368–371

Contents lists available at ScienceDirect

Food Chemistry

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

Analytical Methods

UV–Vis spectrometric classification of coffees by SPA–LDA

Urijatan Teixeira Carvalho Polari Souto [a], Márcio José Coelho Pontes [a], Edvan Cirino Silva [a,*], Roberto Kawakami Harrop Galvão [b], Mário César Ugulino Araújo [a], Fátima Aparecida Castriani Sanches [a], Francisco Antônio Silva Cunha [a], Maria Socorro Ribeiro Oliveira [a]

[a] Universidade Federal da Paraíba, Departamento de Química – Laboratório de Automação e Instrumentação em Química Analítica/Quimiometria (LAQA), Caixa Postal 5093, CEP 58051-970 João Pessoa, PB, Brazil
[b] Instituto Tecnológico de Aeronáutica, Divisão de Engenharia Eletrônica, São José dos Campos, SP, Brazil

ARTICLE INFO

Article history:
Received 25 April 2008
Received in revised form 18 May 2009
Accepted 27 May 2009

Keywords:
Coffee classification
UV–Vis spectrometry
Linear discriminant analysis
Successive projections algorithm
Soft independent modelling of class analogy

ABSTRACT

UV–Vis spectrometry and chemometric techniques were used to classify aqueous extracts of Brazilian ground roast coffee with respect to type (caffeinated/decaffeinated) and conservation state (expired and non-expired shelf-life). Two classification methods were compared: soft independent modelling of class analogy (SIMCA) and linear discriminant analysis (LDA) with wavelength selection by the successive projections algorithm (SPA). The best results were obtained by SPA–LDA, which correctly classified all test samples. The classification accuracy of this model remained high (96%) even after the introduction of artificial spectral noise. These results suggest that UV–Vis spectrometry and SPA–LDA modelling provide a promising alternative for assessment of conservation state and decaffeination condition of coffee samples.

1. Introduction

Potential health hazards associated to the consumption of caffeine have been the subject of much research. It has been reported that large doses of caffeine may produce toxic effects including arrhythmias, hypokalemia, hyperglycaemia, vomiting, convulsions, among others (Forman, Aizer, & Young, 1997; Kerrigan & Lindsey, 2005; Riesselmann, Rosenbaum, Roscher, & Schneider, 1999). Even small amounts of caffeine can cause undesirable symptoms in sensitive people. According to Lafuente-Lafuente et al. (2008), a single cup of regular coffee taken in the evening may reduce the quality of sleep in caffeine-sensitive subjects. In some cases, caffeine may increase blood pressure (Myers, 1988) and even aggravate pre-existing predisposition to arrhythmogenesis (Dobmeyer, Stine, Leier, Greenberg, & Schaal, 1983). Fetal growth problems have also been associated to the consumption of caffeine by pregnant women (Vlajinac, Petrovic, Marinkovic, & Sipetic, 1997). In view of these undesirable health effects, sensitive individuals may opt to consume decaffeinated coffee. Therefore, the assessment of compliance with the decaffeination condition stated in the product label is an important issue. According to the Brazilian National Health Agency (ANVISA, 1999), for example, the caffeine content in commercial decaffeinated coffee must be smaller than 0.1% (w/w), which is 10 times less than the typical levels found in arabica and robusta coffee (Feldman, Ryder, & Kung, 1969).

Another aspect that must be taken into account for acceptability of coffee is its conservation state. In this context, the shelf-life can be defined as the time until coffee becomes unacceptable to consumers under a given storage condition. Such information must be stated in the product label. However, examination of conservation state is an important procedure because variations in temperature and/or partial pressure of oxygen may cause a substantial reduction in shelf-life (Cardelli & Labuza, 2001). In addition, regular inspections of conservation state may be of value to discourage the commercialisation of counterfeit products.

Several instrumental techniques have been developed for quality control, classification and authentication of coffee samples. Examples include high performance liquid chromatography (HPLC) (González, Pablos, Martin, León-Camacho, & Valdenebro, 2001), gas chromatography–mass spectrometry (GC–MS) (Agresti, Franca, Oliveira, & Augusti, 2008; Costa Freitas, Parreira, & Vilas-Boas, 2001), inductively coupled plasma optical emission spectrometry (ICP-OES) (Fernandes et al., 2005), nuclear magnetic resonance (NMR) (Charlton, Farrington, & Brereton, 2002), Raman (Rubayiza & Meurens, 2005) and Fourier transform infrared (FTIR) spectrometry (Briandet, Kemsley, & Wilson, 1996).

Most of these instrumental techniques require harmful reagents and/or expensive equipments with large operational/maintenance costs. In this context, UV–Vis spectrometry would be a simpler and

* Corresponding author. Tel.: +55 83 3216 7438; fax: +55 83 3216 7437.
E-mail address: edvan@quimica.ufpb.br (E.C. Silva).

0308-8146/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodchem.2009.05.078

less costly alternative. This technique has been successively applied for determination of caffeine (Belay, Ture, Redi, & Asfaw, 2008; López-Martínez, López-de-Alba, García-Campos, & León-Rodríguez, 2003; Moores, Dorothy, Mcdermott, & Wood, 1948), chlorogenic acid (Moores et al., 1948) and theobromine (López-Martínez et al., 2003) in coffee.

The objective of the present study is to evaluate the potential of UV–Vis spectrometry and chemometric techniques for classification of coffee samples with respect to type (caffeinated/decaffeinated) and shelf-life condition (expired/non-expired).

2. Experimental

2.1. Samples

A total of 175 samples of decaffeinated and caffeinated ground coffee (medium roast degree) from different lots and producers were acquired in supermarkets and roasting industries in the city of João Pessoa, Paraíba, Brazil. These samples consisted of dry-processed blends from arabica and robusta varieties, which are the two most employed species in Brazilian industry.

Ninety of the samples had been stored in the original commercial container without strict environmental control for 30–50 months past the expiration date. Such a large storage time was adopted in the present work to guarantee that the coffees were indeed unacceptable to consumers. Henceforth, these 90 samples will be termed "expired". For the classification study, the following classes were considered: non-expired decaffeinated, non-expired caffeinated, expired decaffeinated, and expired caffeinated with 31, 54, 22 and 68 samples, respectively.

2.2. Extraction procedure

The following aqueous extraction procedure, adapted from Vitorino, França, Oliveira, and Borges (2001), was employed: (1) 1.0 g of each coffee sample was weighted and transferred to a paper filter in a glass funnel; (2) three 50 ml aliquots of distilled water at 90–98 °C were sequentially poured on the sample; (3) after cooling to room temperature (about 25 °C), the extract was diluted in the proportion of 1:20 (v/v) with distilled water.

2.3. UV–Vis spectral measurements

The spectrum of each diluted extract was acquired immediately after the procedures described above. The measurements were carried out with a Hewlett Packard model HP 8453 UV–Vis spectrophotometer equipped with a quartz cell with optical path of 10 mm, and spectral resolution of 1 nm in the range 225–353 nm. The blank spectrum was recorded with recently distilled water.

2.4. Data analysis

The samples were divided into training (89), validation (43) and test (43) sets by applying the classic Kennard–Stone (KS) uniform sampling algorithm (Kennard & Stone, 1969). Each class was treated separately, as described by Pontes et al. (2005). The training sets comprised 17 non-expired decaffeinated, 26 non-expired caffeinated, 12 expired decaffeinated and 34 expired caffeinated samples. The validation set comprised 7 non-expired decaffeinated, 14 non-expired caffeinated, 5 expired decaffeinated and 17 expired caffeinated samples. The remaining samples formed the test set.

The training and validation samples were used in the modelling procedures (including SPA variable selection for LDA and determination of principal components in SIMCA). In particular, the validation samples were employed in SPA to select the best subset of spectral variables by minimising a cost function defined as an average risk of misclassification by LDA (Pontes et al., 2005). It could be argued that this optimisation procedure might lead to an LDA model that is over-fitted to the validation set. Therefore, the third set of test samples was used to check the generalisation ability of the resulting classifier, as recommended elsewhere (Marini et al., 2004). These samples can be deemed independent in that they were only used in the final evaluation and comparison of the classification models.

Principal component analysis (PCA) and SIMCA were carried out in Unscrambler® 9.6 (CAMO SA). PCA and SIMCA were performed with the default settings of the software. The KS and SPA–LDA algorithms were coded in Matlab® 6.5.

3. Results and discussion

3.1. UV–Vis absorbance spectra

Fig. 1 presents the UV–Vis absorption spectra of the aqueous extracts of the 175 coffee samples in the range 225–353 nm. This region can be associated with $n \rightarrow \pi^*$ electronic transition of caffeine, chlorogenic acids and trigonelline molecules. In particular, the band around 275 nm is related to the C=O chromophore absorption of caffeine (López-Martínez et al., 2003). It is worth noting that chlorogenic acids and trigonelline are decomposed with roasting (Illy & Viani, 1995). However, the presence of the bands around 290 and 320 nm suggests that these compounds were not entirely decomposed in the medium roasting process (Vitorino et al., 2001).

In order to investigate the repeatability of the spectral measurements, four samples (one from each class) were employed. Ten extractions were carried out for each sample and three spectra were acquired from each extract. By using these measurements, the pooled standard deviation estimated for each spectral variable ranged from 5.5×10^{-4} to 2.4×10^{-3} in absorbance. Such values, which correspond to the variability of the instrumental response for a given sample, are much smaller than the inter-sample variability, as illustrated in Fig. 2.

3.2. Principal component analysis

Fig. 3 presents the PC2 × PC1 score plot resulting from the application of PCA to the UV–Vis spectra. As can be seen, there is

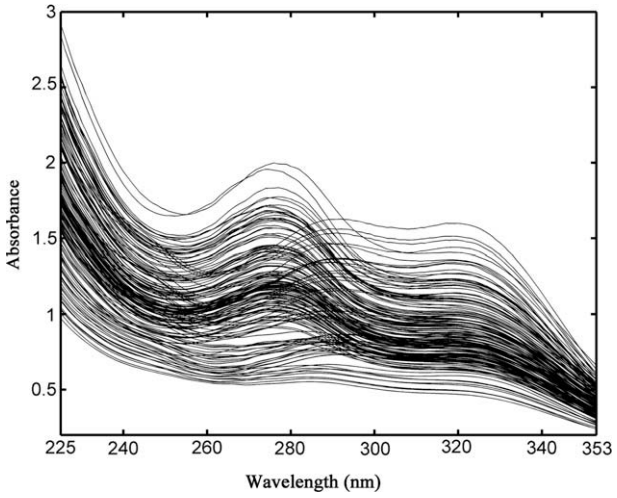

Fig. 1. UV–Vis spectra of the 175 coffee samples.

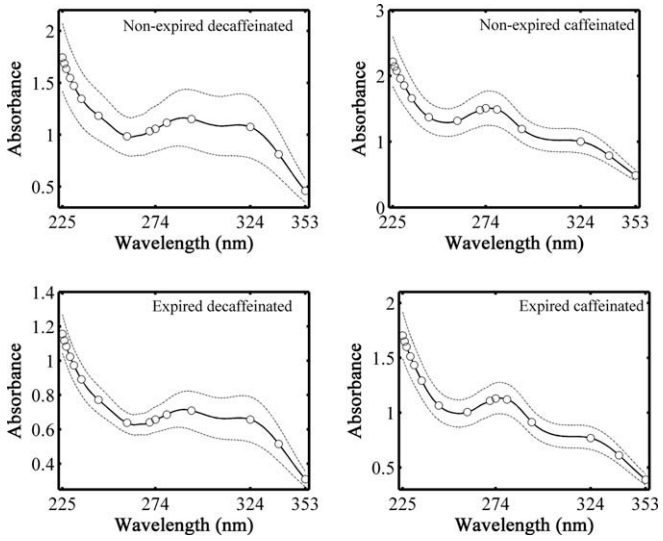

Fig. 2. Mean spectra of the four coffee classes (solid lines) with +/− one standard deviation limits (dash-dot lines). The 15 wavelengths selected by the successive projections algorithm are indicated.

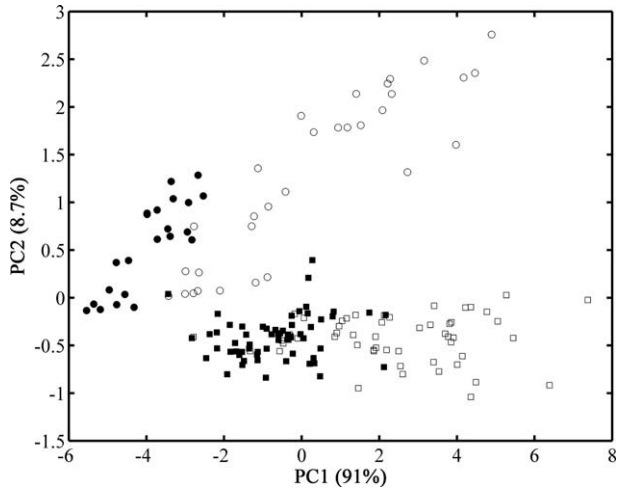

Fig. 3. Principal components score plot for all 175 coffee samples (○, non-expired decaffeinated; ●, expired decaffeinated; □, non-expired caffeinated and ■, expired caffeinated). The variance explained by each principal component is indicated in parenthesis.

substantial overlapping among the classes, especially regarding the non-expired (□) and expired (■) caffeinated samples.

3.3. SIMCA classification

A full-spectrum SIMCA model was built for each of the four coffee types. Table 1 presents the classification results obtained in the test set at a significance level of 5% for the *F*-test. It is worth noting that SIMCA errors can be of two types (I or II). Type-I errors correspond to samples not included in their own class (main diagonal of Table 1). Type-II errors correspond to samples included in an incorrect class (off-diagonal elements of Table 1). As can be seen, there is a large number of Type-II errors related to caffeinated samples (non-expired classified as expired and vice versa). This finding indicates that SIMCA was not able to provide adequate models for assessing the conservation state of caffeinated coffee.

Table 2
Summary of results (classification errors in the test set) for SIMCA and SPA–LDA in the coffee data set.

	SIMCA	SPA–LDA	SIMCA[a]
Type-I	1	0	2
Type-II	26	0	24
Total	27	0	26

[a] SIMCA models built with wavelength subset selected by SPA–LDA.

3.4. SPA–LDA classification

The 15 wavelengths selected by SPA–LDA, namely 225, 226, 227, 229, 231, 235, 244, 259, 271, 274, 280, 293, 324, 339 and 353 nm, are indicated in Fig. 2. The LDA model obtained with the 15 variables selected by SPA was applied to the classification of the test set. As a result, all samples were correctly classified.

In order to further assess the discriminatory power of the selected variables, SIMCA modelling was repeated by using only the 15 spectral variables selected by SPA–LDA instead of the full spectrum. By using this procedure, the results remained largely unaltered. One additional Type-I error was observed (a non-expired caffeinated sample was not included in its own class). However, as regards Type-II errors for caffeinated coffee, the number of expired samples classified as non-expired decreased from 17 to 16 and the number of non-expired samples classified as expired decreased from 8 to 7. These findings suggest that the discriminatory information conveyed by the full spectrum was preserved in the selected variables.

Table 2 summarises the SPA–LDA and SIMCA classification performance obtained with the test data set. As can be seen, SPA–LDA clearly outperformed SIMCA.

The good classification performance of SPA–LDA is further demonstrated by Fig. 4, which presents the scores of the first two discriminant functions (DF2 × DF1) for the overall data set. As can be seen, the separation of the four classes is much better as compared to the PC2 × PC1 score plot in Fig. 3.

3.5. Robustness to measurement noise

In order to evaluate the robustness of SPA–LDA and SIMCA with respect to measurement noise, a Monte Carlo simulation was carried out (Martins et al., 2003). For this purpose, artificial noise was added to the spectra of the test samples. The noise level was set

Table 1
Number of classification errors in the test set obtained by the full-spectrum SIMCA model. The number of samples in each class and the number of principal components in each SIMCA model are indicated in parenthesis.

Samples	Model			
	Non-expired decaffeinated (2PCs)	Non-expired caffeinated (2PCs)	Expired decaffeinated (2PCs)	Expired caffeinated (2PCs)
Non-expired decaffeinated (7)	–	–	–	–
Non-expired caffeinated (14)	–	–	–	8
Expired decaffeinated (5)	1	–	1	–
Expired caffeinated (17)	–	17	–	–

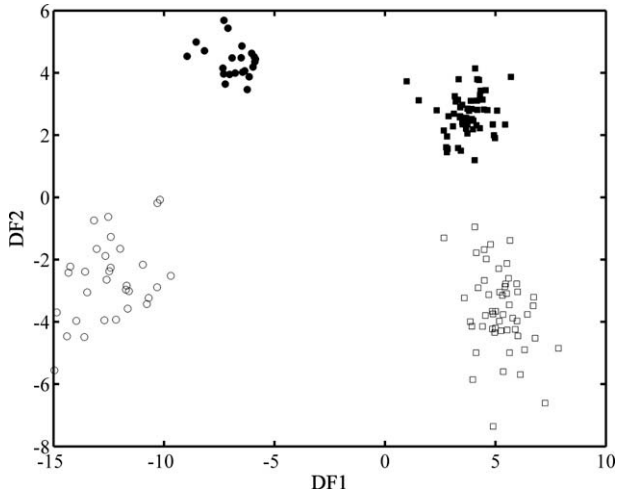

Fig. 4. Score plot of the two first discriminant functions (DF1 and DF2) for the overall data set of 175 coffee samples.

according to the pooled standard deviation calculated in Section 3.1. Ten different noise permutations were added to the spectra of each of the 43 test samples, thus generating 430 noisy spectra. By using the SPA–LDA model, a classification accuracy of 96% was achieved (19 errors out of 430). The full-spectrum SIMCA models yielded 10 Type-I and 260 Type-II errors. By using the variables selected by SPA–LDA, SIMCA resulted in 20 Type-I and 224 Type-II errors.

4. Conclusions

In this study, SPA–LDA displayed better classification performance compared with SIMCA, particularly in discriminating between conservation states of caffeinated samples. In fact, the SPA–LDA model correctly classified all 43 test samples. Moreover, this model was found to be robust with respect to measurement noise, as the classification accuracy remained high (96%) even after the introduction of artificial noise in the spectra. These results suggest that the proposed methodology is a promising alternative for assessment of conservation state and decaffeination condition of coffee samples.

Future works could investigate the possibility of using this methodology to monitor the ageing process of coffee samples over time. Moreover, the influence of roasting degree in the classification results could also be studied.

Acknowledgements

This work was supported by CAPES (PROCAD Grant 0081/05-1, MSc and DSc studentships) and CNPq (research fellowships). The authors are also grateful to São Braz and Santa Clara roasting industries for providing the coffee samples employed in this study.

References

Agresti, P. D. C. M., Franca, A. S., Oliveira, L. S., & Augusti, R. (2008). Discrimination between defective and non-defective Brazilian coffee beans by their volatile profile. *Food Chemistry, 106*(2), 787–796.

ANVISA (National Agency of Sanitary Vigilance) (1999). DOU, April 26, N. 377. <http://www.anvisa.gov.br/legis/portarias/377_99.htm> Accessed 1.10.2008.

Belay, A., Ture, K., Redi, M., & Asfaw, A. (2008). Measurement of caffeine in coffee beans with UV/vis spectrometer. *Food Chemistry, 108*(1), 310–315.

Briandet, R., Kemsley, K., & Wilson, R. H. (1996). Discrimination of arabica and robusta in instant coffee by Fourier transform infrared spectroscopy and chemometrics. *Journal of Agricultural and Food Chemistry, 44*(1), 170–174.

Cardelli, C., & Labuza, T. P. (2001). Application of weibull hazard analysis to the determination of the shelf life of roasted and ground coffee. *Lebensmittel-Wissenschaft und-Technologie, 34*(5), 273–278.

Charlton, A. J., Farrington, W. H. H., & Brereton, P. (2002). Application of ^{1}H NMR and multivariate statistics for screening complex mixtures: Quality control and authenticity of instant coffee. *Journal of Agricultural and Food Chemistry, 50*(11), 3098–3103.

Costa Freitas, A. M., Parreira, C., & Vilas-Boas, L. (2001). The use of an electronic aroma-sensing device to assess coffee differentiation–comparison with SPME gas chromatography–mass spectrometry aroma patterns. *Journal of Food Composition and Analysis, 14*(5), 513–522.

Dobmeyer, D. J., Stine, R. A., Leier, C. V., Greenberg, R., & Schaal, S. F. (1983). The arrhythmogenic effects of caffeine in human beings. *The New England Journal of Medicine, 308*(14), 814–816.

Feldman, J. R., Ryder, W. S., & Kung, J. T. (1969). Importance of nonvolatile compounds to the flavour of coffee. *Journal of Agricultural and Food Chemistry, 17*(4), 733–739.

Fernandes, A. P., Santos, M. C., Lemos, S. G., Ferreira, M. M. C., Nogueira, A. R. A., & Nóbrega, J. A. (2005). Pattern recognition applied to mineral characterization of Brazilian coffees and sugar-cane spirits. *Spectrochimica Acta Part B, 60*(5), 717–724.

Forman, J., Aizer, A., & Young, C. R. (1997). Myocardial infarction resulting from caffeine overdose in an anorectic woman. *Annals of Emergency Medicine, 29*(1), 178–180.

González, A. G., Pablos, F., Martin, M. J., León-Camacho, M., & Valdenebro, M. S. (2001). HPLC analysis of tocopherols and triglycerides in coffee and their use as authentication parameters. *Food Chemistry, 73*(1), 93–101.

Illy, A., & Viani, R. (1995). *Expresso coffee: The chemistry of quality*. San Diego: Academic.

Kennard, R. W., & Stone, L. A. (1969). Computer-aided design of experiments. *Technometrics, 11*(1), 137–148.

Kerrigan, S., & Lindsey, T. (2005). Fatal caffeine overdose: Two case reports. *Forensic Science International, 153*(1), 67–69.

Lafuente-Lafuente, C., Mouly, S., Delcey, V., Mahe, I., Diemer, M., Chassany, O., et al. (2008). Effects of an evening single cup of coffee on the quality of sleep in caffeine-sensitive patients: A randomised cross-over trial. *Fundamental and Clinical Pharmacology, 22*(s1), 60.

López-Martínez, L., López-de-Alba, P. L., García-Campos, R., & León-Rodríguez, L. M. D. (2003). Simultaneous determination of methylxanthines in coffees and teas by UV–Vis spectrophotometry and partial least squares. *Analytica Chimica Acta, 493*(1), 83–94.

Marini, F., Balestrieri, F., Bucci, R., Magri, A. D., Magri, A. L., & Marini, D. (2004). Supervised pattern recognition to authenticate Italian extra virgin olive oil varieties. *Chemometrics and Intelligent Laboratory Systems, 73*(1), 85–93.

Martins, V. L., Almeida, L. F., Castro, S. L., Galvão, R. K. H., Araújo, M. C. U., & Silva, E. C. (2003). A multiscale wavelet data treatment for reliable localization of inflection points for analytical purposes. *Journal of Chemical Information and Computer Science, 43*(6), 1725–1732.

Moores, R. G., Dorothy, L., Mcdermott, L., & Wood, T. R. (1948). Determination of chlorogenic acid in coffee. *Analytical Chemistry, 20*(7), 620–624.

Myers, M. G. (1988). Effects of caffeine on blood pressure. *Archives of Internal Medicine, 148*(5), 1189–1193.

Pontes, M. J. C., Galvão, R. K. H., Araújo, M. C. U., Moreira, P. N. T., Pessoa Neto, O. D., José, G. E., et al. (2005). The successive projections algorithm for spectral variable selection in classification problems. *Chemometrics and Intelligent Laboratory Systems, 78*(1–2), 11–18.

Riesselmann, B., Rosenbaum, F., Roscher, S., & Schneider, V. (1999). Fatal caffeine intoxication. *Forensic Science International, 103*(1), S49–S52.

Rubayiza, A. B., & Meurens, M. (2005). Chemical discrimination of arabica and robusta coffees by Fourier transform Raman spectroscopy. *Journal of Agricultural and Food Chemistry, 53*(12), 4654–4659.

Vitorino, M. D., França, A. S., Oliveira, L. S., & Borges, M. L. A. (2001). Metodologias de obtenção de extrato de café visando a dosagem de compostos não voláteis. *Revista Brasileira de Armazenamento, 26*(3), 17–24.

Vlajinac, H. D., Petrovic, R. R., Marinkovic, J. M., & Sipetic, S. B. (1997). Effects of caffeine intake during pregnancy on birth weight. *American Journal of Epidemiology, 145*(4), 335–338.