Talanta 79 (2009) 1260–1264

Contents lists available at ScienceDirect

Talanta

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

Near infrared reflectance spectrometry classification of cigarettes using the successive projections algorithm for variable selection

Edilene Dantas Teles Moreira[a], Márcio José Coelho Pontes[a], Roberto Kawakami Harrop Galvão[b], Mário César Ugulino Araújo[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 21 March 2009
Received in revised form 15 May 2009
Accepted 18 May 2009
Available online 27 May 2009

Keywords:
Cigarettes
Near infrared reflectance spectroscopy
Classification
Successive projections algorithm
Linear discriminant analysis

ABSTRACT

This paper proposes a methodology for cigarette classification employing Near Infrared Reflectance spectrometry and variable selection. For this purpose, the Successive Projections Algorithm (SPA) is employed to choose an appropriate subset of wavenumbers for a Linear Discriminant Analysis (LDA) model. The proposed methodology is applied to a set of 210 cigarettes of four different brands. For comparison, Soft Independent Modelling of Class Analogy (SIMCA) is also employed for full-spectrum classification. The resulting SPA–LDA model successfully classified all test samples with respect to their brands using only two wavenumbers (5058 and 4903 cm^{-1}). In contrast, the SIMCA models were not able to achieve 100% of classification accuracy, regardless of the significance level adopted for the *F*-test. The results obtained in this investigation suggest that the proposed methodology is a promising alternative for assessment of cigarette authenticity.

1. Introduction

Cigarette authenticity is an important matter, which involves economic aspects and consumer health issues. In fact, cigarette brands may differ in retail price, as well as in the levels of potentially hazardous substances such as nicotine and tar [1,2]. Therefore, the assessment of compliance with the cigarette label and the identification of counterfeit products are analytical problems that merit investigation.

The discrimination of cigarette types is usually carried out on the basis of visual aspect, flavour and aroma. However, such an inspection is subjective and may lead to unreliable results. As an alternative, instrumental techniques have been employed to obtain a more objective and accurate assessment of cigarette samples. Examples include gas chromatography (GC) and liquid chromatography (LC) [3–5], inductively coupled plasma mass spectrometry (ICP-MS) [6], inductively coupled plasma optical emission spectrometry (ICP-OES) [7,8], nuclear magnetic resonance (NMR) [9] and pyrolysis single-photon ionisation time-of-flight mass spectrometry (Py-SPI-TOFMS) [10]. However, these techniques are laborious and time-consuming, require harmful reagents and involve expensive equipment with high operation and/or maintenance costs. An interesting alternative to overcome such drawbacks would be the use of near infrared (NIR) spectroscopy, a technique that enables practical, fast and less dispendious analyses.

NIR spectroscopy has been successfully applied to discrimination and/or classification of various materials, including alcoholic beverages [11,12], food products [12–15], fuel samples [16–18], polymers [19] and agricultural goods [20], among others [21,22]. However, only a single paper [23] has been published on the use of NIR spectroscopy for cigarette discrimination. In that work, 142 cigarettes of two different brands were distinguished by using the Adaboost algorithm and Linear Discriminant Analysis (LDA) applied to near infrared reflectance (NIRR) measurements. Feature extraction was performed by principal component analysis (PCA) or Kernel Principal Component Analysis (KPCA).

The present paper proposes an analytical methodology for cigarette classification based on the use of NIRR spectroscopy and variable selection. For this purpose, the Successive Projections Algorithm (SPA) [17] is employed to choose an appropriate subset of wavenumbers for a Linear Discriminant Analysis (LDA) model. Recently, SPA–LDA has been successfully applied to the classification of edible vegetable oils and soil samples by using square wave voltammetry (SWV) [24] and laser-induced breakdown spectroscopy (LIBS) [25], respectively. In comparison with the approach adopted in [23], SPA–LDA provides a simpler model in the sense that the classification variables correspond to actual reflectance measurements, rather than PCA/KPCA scores.

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

0039-9140/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2009.05.031

The proposed methodology is applied to a set of 210 cigarettes comprising four brands of different chemical composition and retail price. For comparison, Soft Independent Modelling of Class Analogy (SIMCA) [26] is also employed. SIMCA is a well-known method for full-spectrum classification, which has been widely employed in applications involving NIR data [27–30].

2. Background

2.1. Notation

Matrices will be represented by bold capital letters, column vectors by bold lowercase letters, and scalars by italic characters. The matrix of instrumental responses will be denoted by **X**. The *n*th object in matrix **X** will be denoted by $\mathbf{x}_n$ (that is, $\boldsymbol{x}_n^T$ will correspond to the *n*th row of matrix **X**). The *k*th column of matrix **X** will be denoted by $\mathbf{x}^k$.

2.2. Linear Discriminant Analysis

The LDA classification method employs the Mahalanobis distance [31,32], which can be defined as follows. Let $x = [x_1, x_2, \ldots x_d]^T$ be an object that must be assigned to one out of *c* possible classes. In the case of NIRR data, the classification variables $x_1, x_2, \ldots x_d$ correspond to reflectance measurements acquired at *d* wavenumbers. The squared Mahalanobis distance $r^2(\mathbf{x}, \boldsymbol{\mu}_j)$ between **x** and the center of the *j*th class ($j = 1, 2, \ldots, c$) is defined as

$$r^2(\boldsymbol{x}, \boldsymbol{\mu}_j) = (\boldsymbol{x} - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_j) \quad (1)$$

where $\boldsymbol{\mu}_j$ ($d \times 1$) and $\boldsymbol{\Sigma}_j$ ($d \times d$) are the mean vector and covariance matrix for the class under consideration [32]. If the true mean and covariance values for the population are unknown (which is usually the case), maximum likelihood estimates $\mathbf{m}_j$ and $\mathbf{S}_j$ may be employed in place of $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j$, respectively. These estimates can be obtained from a finite set of training objects of known classification [31]. It is worth noting that LDA estimates a single pooled covariance matrix **S**, instead of using a separate estimate for each class. This regularization procedure simplifies the classification model and results in linear decision surfaces (hyperplanes) in R^d [31,33,34]. With this modification, the squared Mahalanobis distance between **x** and the center of the *j*th class is calculated as

$$r^2(\boldsymbol{x}, \boldsymbol{m}_j) = (\boldsymbol{x} - \boldsymbol{m}_j)^T \boldsymbol{S}^{-1} (\boldsymbol{x} - \boldsymbol{m}_j) \quad (2)$$

Object **x** is then assigned to the class *j* for which $r^2(\mathbf{x}, \mathbf{m}_j)$ has the smallest value.

In order to have a well-posed problem, the number of training objects must be larger than the number *d* of variables to be included in the LDA model. Otherwise, the estimated covariance matrix **S** will be singular, which prevents the calculation of the matrix inverse in Eq. (2). Therefore, the use of LDA for classification of spectral data usually requires appropriate variable selection procedures [17,33,35]. In the present work, the Successive Projections Algorithm (SPA) is adopted for this purpose.

2.3. Successive Projections Algorithm

The Successive Projections Algorithm [36,37] was originally proposed by Araújo et al. [38] in the context of multivariate calibration. In SPA, variable selection is formulated as a constrained combinatorial optimization problem, in which subsets of variables are tested and compared with respect to the performance of the resulting model. The optimization is said to be constrained because the search for an optimum is restricted to certain subsets of variables. Such subsets are formed according to a sequence of projection operations involving the matrix **X** of instrumental responses, as follows.

Suppose that the available *x*-data are disposed in a matrix **X** of dimensions ($N \times K$) such that the *k*th variable x_k is associated to the *k*th column vector $\mathbf{x}^k \in \Re^N$. The column vectors are assumed to be mean-centered. Starting from each variable x_k, $k = 1, \ldots, K$, the following sequence of projection operations is carried out [39].

Step 1 (initialization). Let

$$\boldsymbol{z}^1 = \boldsymbol{x}^k$$
$$i = 1$$
$$\boldsymbol{x}^{j,i} = \boldsymbol{x}^j, \quad j = 1, \ldots, K$$
$$SEL(1, k) = k$$

Let *M* be the largest number of variables to be included in a subset, as specified by the analyst.

Step 2. Calculate the matrix $\mathbf{P}^i$ of projection onto the subspace orthogonal to $\mathbf{z}^i$ as

$$\boldsymbol{P}^i = \boldsymbol{I} - \frac{\boldsymbol{z}^i (\boldsymbol{z}^i)^T}{(\boldsymbol{z}^i)^T \boldsymbol{z}^i} \quad (3)$$

where **I** is an identity matrix of appropriate dimensions.

Step 3. Calculate the projected vectors $\boldsymbol{x}^{j,i+1}$ as

$$\boldsymbol{x}^{j,i+1} = \boldsymbol{P}^i \boldsymbol{x}^{j,i} \quad (4)$$

for all $j = 1, \ldots, K$.

Step 4. Determine the index j^* of the largest projected vector and store this index in matrix **SEL**:

$$j^* = arg \max_{j=1,\ldots,K} ||x^{j,i+1}|| \quad (5)$$

$$SEL(i + 1, k) = j^* \quad (6)$$

Step 5. Let $\mathbf{z}^{i+1} = \mathbf{x}^{j^*,\, i+1}$

Step 6. Let $i = i + 1$. If $i < M$ return to Step 2.

After these operations are completed, a total of $K \times M$ subsets of variables will be considered in the search for the optimum solution. For each value of *k* (ranging from 1 to *K*), and for each value of *i* (ranging from 1 to *M*), a subset of *i* variables is defined by the indexes $SEL(1, k), SEL(2, k), \ldots, SEL(i, k)$.

In a subsequent paper [17], SPA was adapted for use in classification problems. As in the original formulation [38], candidate subsets of variables are formed as the result of projection operations carried out on the matrix of instrumental responses for the training data. However, prior to these operations, the objects belonging to the same class are centered in the mean of the class. The resulting subsets of variables are then compared in terms of a cost function *G* calculated for a given validation data set as

$$G = \frac{1}{N_v} \sum_{n=1}^{N_v} g_n, \quad (7)$$

where g_n is defined as

$$g_n = \frac{r^2(\boldsymbol{x}_n, \boldsymbol{m}_{I(n)})}{\min_{I(m) \neq I(n)} r^2(\boldsymbol{x}_n, \boldsymbol{m}_{I(m)})}. \quad (8)$$

where $I(n)$ is the index of the true class for the *n*th validation object $\mathbf{x}_n$. In Eq. (8), the numerator $r^2(\mathbf{x}_n, \mathbf{m}_{I(n)})$ is the squared Mahalanobis distance between $\mathbf{x}_n$ and the center of its true class, whereas the denominator corresponds to the squared Mahalanobis distance between $\mathbf{x}_n$ and the center of the closest wrong class. The cost function *G* can be interpreted as an average risk of misclassification of the validation data.

3. Experimental

3.1. Samples

A total of 210 cigarette samples of different lots and four brands (A, 45; B, 57; C, 57 and D, 51) were acquired in the city of João Pessoa, Paraíba, Brazil. Before NIRR spectral recording, these samples were dried in an oven at 60 °C for 24 h, ground, sieved to a particle size smaller than 300 μm and stored in desiccators.

3.2. NIRR spectra measurements

The spectra were recorded in triplicate by using a Spectrum GX FTIR spectrophotometer (PerkinElmer), with spectral resolution of 1 cm^{-1} and 32 scans in the near infrared range of 15,000–2700 cm^{-1}. After a preliminary inspection of the spectra, those regions in which the detector was saturated or the signal-to-noise ratio was poor were discarded. As a result, the 5420–4252 cm^{-1} interval was selected for the study.

A mean spectrum was then calculated for each sample by averaging the triplicate spectra. The spectrum of a KBr sample was used as blank.

3.3. Software

The samples were divided into training, validation and test sets by applying the classic Kennard-Stone (KS) uniform sampling algorithm [40] to the NIRR spectra. Each class was treated separately, as described in Ref. [17]. The number of samples in each set is presented in Table 1.

As in Ref. [17], the training and validation samples were used in the modelling procedures (including SPA variable selection for LDA and determination of principal components in SIMCA) whereas the test samples were only used in the final evaluation and comparison of the classification models.

Spectrum differentiation, Savitzky–Golay smoothing [41], principal component analysis (PCA) and SIMCA were carried out in Unscrambler® 9.6 (CAMO S.A.). PCA and SIMCA were performed with the default settings of the software. Four different significance levels (1%, 5%, 10%, 25%) of the *F*-test for SIMCA classification were tested. The KS and SPA–LDA algorithms were coded in Matlab® 6.5.

4. Results and discussions

4.1. NIRR spectra

Fig. 1a presents the raw NIRR spectra of the 210 cigarette samples in the range of 5420–4252 cm^{-1}. As can be seen, the spectra are noisy and display systematic variations in the spectral baseline. These problems were circumvented by applying the Savitzky–Golay first derivative procedure with a second-order polynomial and a 121-point window, as shown in Fig. 1b. Each resulting spectrum had 1049 points.

Table 1
Number of training, validation and test samples in each class.

Class	Set		
	Training	Validation	Test
A	25	10	10
B	27	15	15
C	27	15	15
D	27	12	12
Total	106	52	52

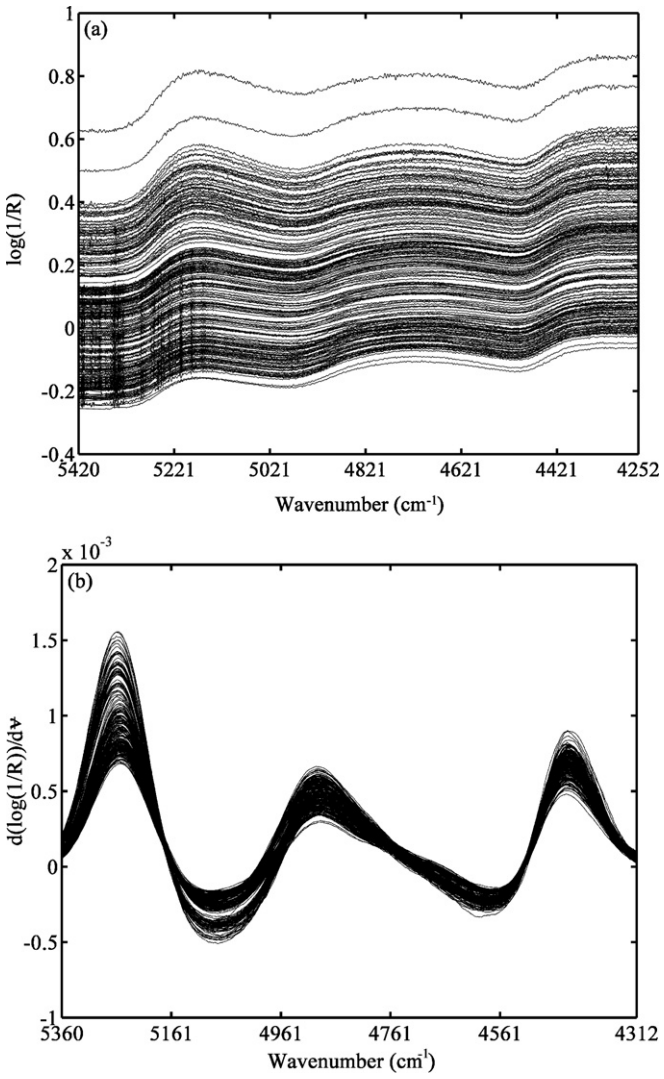

Fig. 1. (a) Original and (b) pre-processed NIRR spectra of the cigarette samples.

4.2. Principal Component Analysis

Fig. 2 presents the PC2 × PC1 score plot resulting from the application of PCA to the derivative spectra. As can be seen, there is no overlapping between the four cigarette brands, which indicates that the NIRR spectrum conveys appropriate information for the classification task.

4.3. SIMCA classification

A SIMCA model was built for each of the four cigarette brands. Table 2 presents the classification results obtained by applying the SIMCA models to the test set. It is worth noting that SIMCA errors can be of two types. A type-I error consists of a sample not included in its own class, such as the A sample that was not included in the A model at the 25% significance level. A type-II error consists of a sample included in an incorrect class, such as the four A samples included in the B model at the 1% significance level.

4.4. SPA–LDA classification

The optimum number of variables for SPA–LDA was determined from the minimum of the cost function *G* displayed in Fig. 3. As

Table 2
Number of SIMCA errors in the test set for four different significance levels of the *F*-test (1%, 5%, 10%, and 25%).

	Model															
	A (7 PCs)				B (4 PCs)				C (7 PCs)				D (5 PCs)			
Level (%)	1	5	10	25	1	5	10	25	1	5	10	25	1	5	10	25
A	–	–	–	1	4	–	–	–	–	–	–	–	1	–	–	–
B	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
C	–	–	–	–	–	–	–	–	–	–	–	2	1	–	–	–
D	3	3	2	–	–	–	–	–	1	–	–	–	–	–	–	1

The number of principal components in each SIMCA class is also indicated.

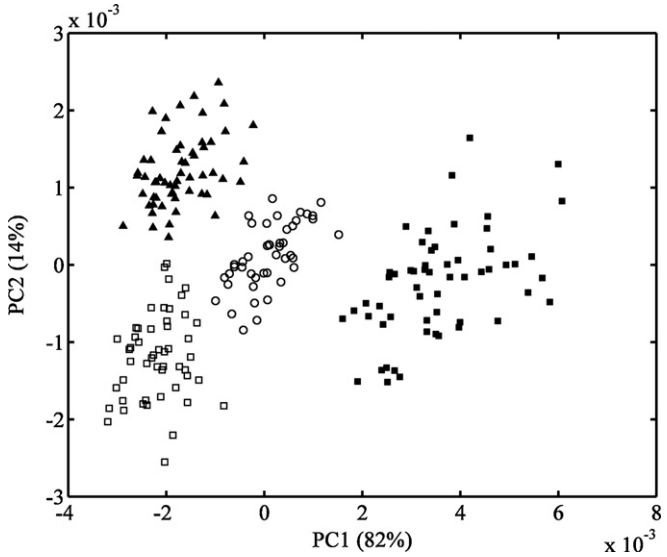

Fig. 2. PC2 × PC1 score plot for the overall set of 210 cigarette samples (○: A; ■: B; ▲: C and □: D). The variance explained by each principal component is indicated in parenthesis.

can be seen, a well-localized minimum is obtained for two variables. These variables correspond to the wavenumbers 4903 and 5058 cm^{-1}, as indicated in Fig. 4.

The resulting LDA model correctly classified all test samples. As shown in Table 3, this classification performance is not achieved by SIMCA regardless of the significance level adopted for the *F*-test.

For illustration, Fig. 5 presents the overall set of 210 cigarette samples in a bivariate plot for the two wavenumbers selected

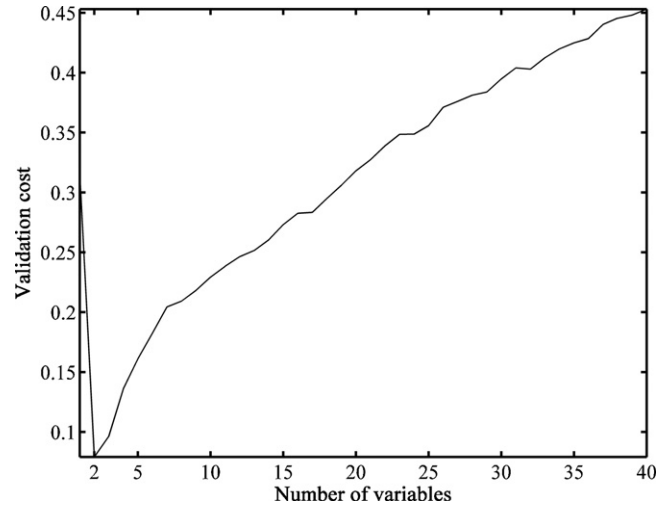

Fig. 3. Determination of the optimum number of variables in SPA–LDA.

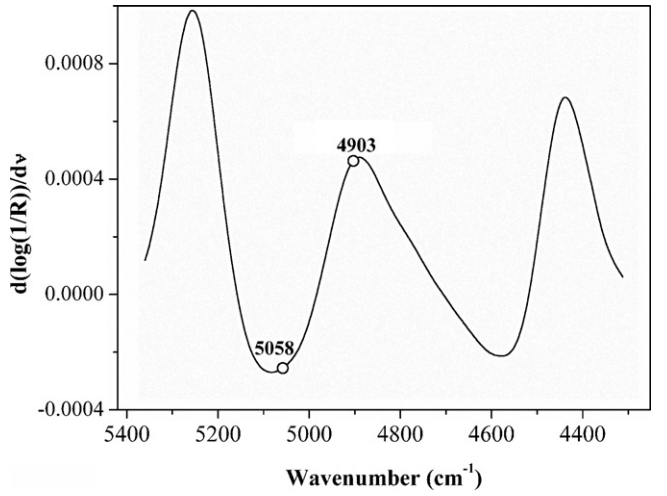

Fig. 4. Mean derivative spectrum of the data set with indication of wavenumbers selected by SPA–LDA.

Table 3
Summary of results (classification errors in the test set) for SPA–LDA and SIMCA (four different significance levels of the *F*-test).

	SPA–LDA	SIMCA (1%)	SIMCA (5%)	SIMCA (10%)	SIMCA (25%)
Type I	0	0	0	0	6
Type II	0	10	3	2	0
Total	0	10	3	2	6

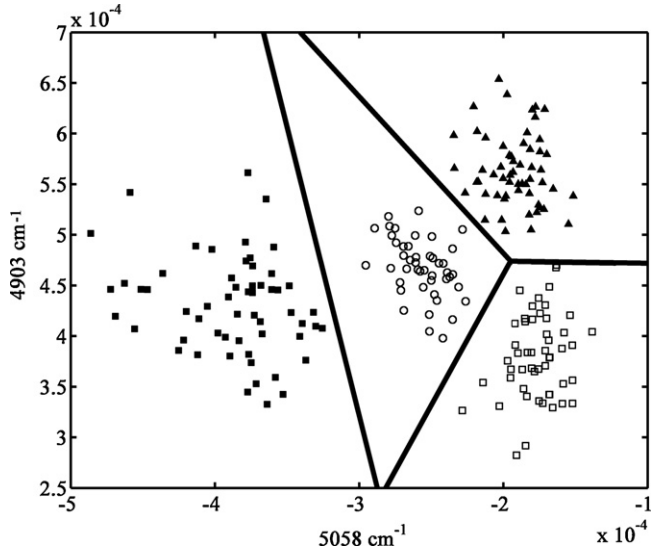

Fig. 5. Bivariate plot of the 210 cigarette samples (○: A; ■: B; ▲: C and □: D) for the two wavenumbers selected by SPA (4903 and 5058 cm^{-1}). The thick lines are the decision boundaries established by LDA.

by SPA. As can be seen, the four brands are properly separated by the linear boundaries resulting from LDA, which indicates that the two spectral variables are appropriate for discrimination purposes.

5. Conclusions

This paper proposed a methodology for cigarette classification employing NIRR spectrometry and Linear Discriminant Analysis coupled with the Successive Projections Algorithm for wavenumber selection. In a case study involving four different cigarette brands, the resulting SPA–LDA model successfully classified all test samples using only two wavenumbers (5058 and 4903 cm^{-1}). In contrast, traditional full-spectrum SIMCA models were not able to achieve 100% of classification accuracy, regardless of the significance level adopted for the *F*-test.

The results obtained in this investigation suggest that the proposed methodology is a promising alternative for assessment of cigarette authenticity. It is worth noting that the methodology is based solely on spectroscopic measurements and chemometrics techniques. Therefore, laborious procedures for chemical characterization of the cigarettes are not required. On the other hand, representative sets of training and validation samples must be available for variable selection and model-building purposes. Moreover, since the classification procedure is data-driven, the results cannot be easily generalized to the analysis of cigarette brands that were not included in the study. In particular, the number of required spectral variables (wavenumbers) would likely depend on the chemical similarity between the brands, as well as the signal-to-noise ratio of the measurements.

Acknowledgments

This work was supported by CAPES (PROCAD Grant 0081/05-1, MSc and DSc studentship) and CNPq (research fellowships).

References

[1] S.D. Bolboacă, L. Jäntschi, Int. J. Environ. Res. Public Health 4 (2007) 233.
[2] J.F. Pankow, J.E. Henningfield, B.E. Garrett, Nicotin Tob. Res. 6 (2004) 199.
[3] G. Pieraccini, S. Fulanetto, S. Orlandini, G. Bartolucci, I. Giannini, S. Pinzauti, G. Moneti, J. Chromatogr. A 1180 (2008) 138.
[4] S.L. Choua, Y.C. Ling, M.H. Yang, C.Y. Pai, Anal. Chim. Acta 598 (2007) 103.
[5] L.F. Huanga, K.J. Zhong, X.J. Sun, M.J. Wu, K.L. Huang, Y.Z. Liang, F.Q. Guo, Y.W. Li, Anal. Chim. Acta 575 (2006) 236.
[6] M.J. Chang, J.D. Naworal, K. Walker, C.T. Connell, Spectrochim. Acta: B 58 (2003) 1979.
[7] R.S. Pappas, G.M. Polzin, L. Zhang, C.H. Watson, D.C. Paschal, D.L. Ashley, Food Chem. Toxicol. 44 (2006) 714.
[8] C.C. Crispino, K.G. Fernandes, M.Y. Kamogawa, J.A. Nóbrega, A.R.A. Nogueira, M.M.C. Ferreira, Anal. Sci. 23 (2007) 435.
[9] D.E. Axelson, J.B. Wooten, J. Anal. Appl. Pyrolysis 78 (2007) 214.
[10] T. Adam, E.T. Ferge, E.S. Mitschke, E.T. Streibel, R.R. Baker, E.R. Zimmermann, Anal. Bioanal. Chem. 381 (2005) 487.
[11] M.J.C. Pontes, S.R.B. Santos, M.C.U. Araújo, L.F. Almeida, R.A.C. Lima, E.N. Gaião, U.T.C.P. Souto, Food Res. Int. 39 (2006) 182.
[12] T. Woodcock, G. Gowney, C.P.O. Donnell, J. Near Infrared Spectrosc. 16 (2008) 1.
[13] D. Toher, G. Downey, T.B. Murphy, Chemom. Intell. Lab. Syst. 89 (2007) 102.
[14] L.A. Berrueta, R.M. Alonso-Salces, K. Héberger, J. Chromatogr. A 1158 (2007) 196.
[15] A.M.C. Davies, J.G. Franklin, A. Grant, N.M. Griffiths, R. Shepherd, G.R. Fenwick, Vib. Spectrosc. 2 (1991) 161.
[16] R.M. Balabin, R.Z. Safieva, Fuel 87 (2008) 1096.
[17] 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.
[18] M. Kim, Y.H. Lee, C. Han, Comput. Chem. Eng. 24 (2000) 513.
[19] R. Leitner, H. Mairer, A. Kercek, Real-Time Imaging 9 (2003) 245.
[20] K.D. Shepherd, M.G. Walsh, J. Near Infrared Spectrosc. 15 (2007) 1.
[21] C. Pasquini, J. Braz. Chem. Soc. 14 (2003) 198.
[22] V.R. Kondepati, M. Keese, R. Mueller, B.C. Manegold, J. Backhaus, Vib. Spectrosc. 44 (2007) 236.
[23] C. Tan, M. Li, X. Qin, Anal. Bioanal. Chem. 389 (2007) 667.
[24] F.F. Gambarra-Neto, G. Marino, M.C.U. Araújo, R.K.H. Galvão, M.J.C. Pontes, E.P. Medeiros, R.S. Lima, Talanta 77 (2009) 1660.
[25] 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) 12–18.
[26] K.R. Beebe, R.J. Pell, B. Seasholtz, Chemometrics—A Practical Guide, Wiley, New York, 1998.
[27] G. Downey, P. Mcintyre, A.N. Davies, J. Agric. Food. Chem. 50 (2002) 5520.
[28] J. Luypaert, D.L. Massart, Y.V. Heyden, Talanta 72 (2007) 865.
[29] E. Smidt, K. Meissl, M. Schwanninger, P. Lechner, Waste Manage. 28 (2008) 1699.
[30] P. Sirisomboon, Y. Hashimoto, M. Tanaka, J. Food Eng. 93 (2009) 502.
[31] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, 2nd ed., John Wiley, New York, 2001.
[32] R.D. Maesschalck, D. Jouan-Rimbaud, D.L. Massart, Chemom. Intell. Lab. Syst. 50 (2000) 1.
[33] A.R. Caneca, M.F. Pimentel, R.K.H. Galvão, C.E. Matta, F.R. Carvalho, I.M. Raimundo Jr., C. Pasquini, J.J.R. Rohwedder, Talanta 70 (2006) 344.
[34] W. Wu, Y. Mallet, B. Walczak, W. Penninckx, D.L. Massart, S. Heuerding, F. Erni, Anal. Chim. Acta 329 (1996) 257.
[35] Y. Mallet, D. Coomans, O.D. Vel, Chemom. Intell. Lab. Syst. 35 (1996) 157.
[36] S. Ye, D. Wang, S. Min, Chemom. Intell. Lab. Syst. 91 (2008) 194.
[37] R.K.H. Galvão, M.C.U. Araújo, W.D. Fragoso, E.C. Silva, G.E. José, S.F.C. Soares, H.M. Paiva, Chemom. Intell. Lab. Syst. 92 (2008) 83.
[38] 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.
[39] R.K.H. Galvão, M.C.U. Araújo, in: S.D. Brown, R. Tauler, B. Walczak (Eds.), Comprehensive Chemometrics: Chemical and Biochemical Data Analysis, Elsevier, Oxford, 2009, p. 233.
[40] R.W. Kennard, L.A. Stone, Technometrics 11 (1969) 137.
[41] A. Savitzky, M.J.E. Golay, Anal. Chem. 36 (1964) 1627.