Multivariate Classification of Cigarettes According to Their Elemental Content Determined by Inductively Coupled Plasma Optical Emission Spectrometry

Carla C. Crispino,[*1,*2] **Kelly G. Fernandes,**[*1,*2] **Marcos Y. Kamogawa,**[*4] **Joaquim A. Nóbrega,**[*1] **Ana Rita A. Nogueira,**[*2†] **and Márcia M. C. Ferreira**[*3]

*1 *Grupo de Análise Instrumental Aplicada, Departamento de Química, Universidade Federal de São Carlos, São Carlos SP, Brazil*
*2 *Embrapa Pecuária Sudeste, P. O. Box 339, 13560-970, São Carlos SP, Brazil*
*3 *Instituto de Química, Universidade Estadual de Campinas, Campinas SP, Brazil*
*4 *Universidade de São Paulo, Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba SP, Brazil*

In this paper we describe our study on the characterization of cigarette samples according to their mineral content. Acid digestion assisted by microwaves was employed, and inductively coupled plasma optical emission spectrometry was the analytical technique used for the determination of Al, Ba, Ca, Cu, Fe, K, Mg, Mn, Na, P and Sr in conventional, light, and flavorized cigarettes. Multivariate techniques, such as hierarchical clusters analysis (HCA) and principal-component analysis (PCA), were applied to discriminate among different types of cigarettes. Cluster analysis and principal-component analysis showed differences in cigarettes according to the type and mineral composition. The cigarette samples were divided within the 3 groups according to their mineral composition. Ca, Sr, Cu, K and Na were the most important elements for cigarette classification, and only these 5 variables were sufficient for the classification and discrimination of the evaluated types of cigarettes.

Introduction

Tobacco has been extensively studied in order to assess the effects of nicotine on the human body.[1] The smoking of tobacco products has been in the etiology of respiratory diseases, cancer and cardiovascular diseases related to arteriosclerosis.[2] The cigarette is a form better known concerning the destination and consumption of tobacco. It is the product more consumed and disposed in the world.

A cigarette contains numerous cytotoxic substances, such as polycyclic aromatics hydrocarbons (PAHs), aromatics amines, nitrosamines, heavy metals, poisonous gases and pesticide residues.[3]

Little attention has been given to the presence of heavy metals and other toxic and trace elements in tobacco smoke and their possible effects on biochemical processes in the human body.[1]

Therefore, there are no specifications for metal contents in cigarettes. The specifications normal, light, and flavor in cigarettes are related to the content of nicotine and tar present in the smoke.

In cigarettes, element concentrations vary among brands, and even within the same brand.[1] The content in elements of the cigarette can be influenced by such factors as the concentration of these elements in soil, fertilizing practices, and processing conditions.[2]

Multi-element analysis techniques have been proposed for the characterization of mineral contents in water,[4] juices and soft drinks,[5] wines,[6] beers,[7] and medicinal plants.[8] Inductively coupled plasma optical emission spectrometry (ICP OES) has been used for determining the mineral content. ICP OES is a multi-element technique with very good detection power, rapidly determining the analytes with high precision.[4,5,9,10]

Multivariate methods of data analysis, such as principal-component analysis (PCA) and hierarchical cluster analysis (HCA), have been very useful to characterize sample types.

In the present work, the contents of Al, Ba, Ca, Cu, Fe, K, Mn, Na, P, and Sr were determined in samples of different types of cigarettes (conventional, light, and flavor) by ICP OES after previous acid digestion, and used as chemical descriptors. PCA and HCA were applied to explore the relationship between the mineral content and discrimination among different cigarette types.

Experimental

Apparatus

A cavity microwave oven (Anton Paar, multiwave sample

† To whom correspondence should be addressed.
E-mail: anarita@cppse.embrapa.br

Table 1   Closed-vessel microwave program for sample digestion

Step	Power/W	Time/min	Step	Power/W	Time/min
1	250	1	4	400	5
2	0	1	5	750	10
3	250	5	6	0	10

Table 2 Operating parameters for ICP OES

Element[a]	Wavelength/nm	LD/µg g[-1]
Al I	396.152	0.28
Ba II	493.398	0.73
Ca II	317.937	2.60
Cu I	327.395	1.17
Fe II	259.935	0.71
K I	766.506	1.25
Mg I	280.264	1.16
Mn II	259.369	0.75
Na I	588.995	1.18
P I	177.434	0.35
Sr II	407.771	1.25

Power/kW	1.2
Plasma gas flow rate/L min[-1]	15
Auxiliary gas flow rate/L min[-1]	1.5
Nebulizer gas flow rate/L min[-1]	0.8
Sample flow rate/L min[-1]	0.8

a. I, atomic line; II, ionic line.

Table 3 Determination of elemental content in a certified reference material (NIST 1515 apple leaves) by ICP OES (µg g[-1] ± SD, $n = 3$)

	Certified value	Found
Al	286.00 ± 9	278.66 ± 5
Ba	49.00 ± 2	48.03 ± 3
Ca	15260.00 ± 150	15020.74 ± 290
Cu	5.64 ± 0.24	7.35 ± 1.67
Fe	83.00 ± 5	79.83 ± 3
K	16100 ± 200	16015 ± 354
Mg	2710 ± 80	2690 ± 190
Mn	54.00 ± 3	52.57 ± 4
Na	24.4 ± 1.2	23.1 ± 2.5
P	1590.00 ± 110	1601.74 ± 30
Sr	25.00 ± 2	26.68 ± 2

preparation system) was used for sample decomposition. The microwave-heating program is listed in Table 1. Elemental analysis was carried out in an ICP OES with an axial view configuration (Varian, Australia) and a nebulizer V-Groove. Table 2 gives the analytical lines and the limit of detection (3 × SD of 10 measurements of the blank divided by the slope of calibration curve) used for each element, as well as the instrumental conditions.

Reagents and samples

Monoelemental containing stock solutions of Al, Ba, Ca, Cu, Fe, K, Mg, Mn, Na, P and Sr 1000 µg mL[-1] (Spex Sample Preparation, Metuchen, NJ) were used for preparing reference solutions for the calibration curve and optimization of the analytical conditions. Nitric acid 65% (v/v) (Carlo Erba, Milan, Italy) and hydrogen peroxide 30% (m/v) (Merck, Darmstadt, Germany) were used for acid digestion of the samples, both of which were of analytical grade. All aqueous solutions and dilutions were prepared with ultrapure water (18 MΩ cm[-1]), obtained by using a Milli-Q system (Millipore, Bedford, MA).

Cigarette samples

Three different types of cigarettes were obtained from local stores, representing the common types readily available to consumers. Besides conventional brands (designated as N), the following brands were selected for analysis: light (labeled as L) and flavored (menthol) (identified by code F).

Analytical procedure

Microwave sample preparation. Three 0.2-g portions of each individual cigarette sample were homogenized, dried at 60°C for 12 h and directly weighed in microwave vessels. To each vessel, 3.0 mL of concentrated nitric acid and 1.0 mL of hydrogenous peroxide 30% (m/v) were added. Microwave-assisted acid-decomposition was performed at high pressure (30 bar) and temperature (190°C). In each decomposition cycle, 6 vessels (1 blank and 5 samples) were used simultaneously. After completing the digestion program, the digested products were transferred to 25.0 mL volumetric flasks, and the volumes were made up with deionized water. The samples were digested in triplicate.

Analysis of the elemental content. Two emission lines for each element were tested before selection. The choice of the analytes

spectral lines was based on both their sensitivity and spectral interference. Linear concentration intervals for each element range between the detection limit (in concentration units) and a value are less than or equal to the maximal concentration, recommended in the Varian guidelines. The samples were measured in triplicate. Certificate reference material samples, NIST SRM 1515 (apple leaves) were used to validate the sample-preparation procedure and analytical measurements in ICP OES (Table 3).

Multivariate analysis

Data analysis. In this work, data analysis was performed using the software Pirouette 2.2 (Infometrix, Seatle, WA). A data matrix (24 × 11) shown in Table 4, was constructed with 24 rows representing cigarette samples and 11 columns corresponding to metals concentrations (Al, Ba, Ca, Cu, Fe, K, Mg, Mn, Na, P and Sr). A data pretreatment was applied in order to avoid any differences in the measurement scale. Autoscaling is the most indicated scaling technique in this case (data are mean centered and scaled to unit variance).

Hierarchical clustering analysis (HCA). Hierarchical cluster analysis (HCA)'s primary goal is to display the data in such a way as to emphasize their natural clusters and patterns in a two-dimensional space. The results, qualitative in nature, are usually presented in the form of a dendrogram, allowing the visualization of clusters and correlations among samples or variables.[11,12] In HCA, the Euclidean distances between samples or variables are calculated and transformed into similarity indices ranging from 0 to 1; a smaller distance means a larger index, and therefore a larger similarity.

Principal-component analysis (PCA). PCA is a data-compression method based on the correlation among variables.[13] Its aim is to group those correlated variables, while replacing the original ones by a new set, called the principal components, PCs, onto which the data are projected. These PCs are completely uncorrelated and are built as a simple linear combination of the original variables. It is important to point out here that the PCs contain most of the variability in the data set, albeit in a much lower dimensional space. The first principal component, PC_1, is defined in the direction of maximum variance of the whole data set. PC_2 is the direction that describes the maximum variance in the orthogonal subspace to PC_1. The subsequent components are taken orthogonal to those previously chosen, and describe the maximum of the remaining variance. Once the redundancy is removed, only the first few principal components are required to describe most of the information contained in the original data set. The data

Table 4   Values for the metals in cigarette samples according to their mineral composition and type of cigarette ($\mu g\ g^{-1} \pm SD$, $n = 3$)

	Al	Ba	Ca	Cu	Fe	K	Mg	Mn	Na	P	Sr
N1	480.2 ± 30.5	52.8 ± 6.5	21130.0 ± 34.7	3.5 ± 0.4	424.2 ± 41.5	19593.6 ± 46.2	4331.0 ± 42.5	152.0 ± 10.6	441.3 ± 21.2	2599.9 ± 54.1	119.2 ± 7.3
N2	423.2 ± 23.0	44.5 ± 6.9	22025.8 ± 82.5	3.4 ± 0.4	391.3 ± 46.4	19256.1 ± 108.0	4715.4 ± 28.4	103.0 ± 16.4	305.9 ± 18.1	2323.1 ± 48.0	116.5 ± 2.5
N3	500.7 ± 28.3	64.9 ± 8.9	26481.0 ± 36.6	4.6 ± 0.1	466.2 ± 34.9	18802.2 ± 92.4	5176.7 ± 35.9	105.9 ± 11.5	429.9 ± 29.4	2373.4 ± 50.1	165.4 ± 8.2
N4	386.6 ± 32.2	120.7 ± 3.9	21961.4 ± 30.5	20.5 ± 0.8	296.0 ± 40.6	21729.5 ± 67.7	4465.8 ± 55.6	116.4 ± 10.2	204.5 ± 25.1	2290.0 ± 49.9	146.5 ± 7.0
N5	380.1 ± 29.0	132.9 ± 6.8	21271.6 ± 29.7	31.3 ± 0.8	303.4 ± 45.0	22374.1 ± 23.9	4547.3 ± 52.8	108.9 ± 11.6	291.9 ± 19.7	2447.4 ± 49.6	146.2 ± 4.9
N6	408.3 ± 25.7	128.5 ± 1.9	23133.7 ± 5.7	3.0 ± 0.5	322.5 ± 33.7	22026.3 ± 92.2	4418.1 ± 42.2	134.3 ± 10.0	397.2 ± 26.6	2369.5 ± 51.8	165.3 ± 7.9
N7	588.3 ± 28.3	73.6 ± 8.3	23784.4 ± 48.1	9.9 ± 0.4	537.6 ± 33.1	22886.6 ± 130.1	4507.5 ± 27.8	65.2 ± 12.9	387.0 ± 20.2	2585.6 ± 46.1	109.0 ± 1.1
N8	540.2 ± 25.1	77.8 ± 2.6	23879.0 ± 4.5	9.9 ± 0.2	486.6 ± 44.4	22838.8 ± 50.3	4721.4 ± 32.7	61.7 ± 10.3	323.0 ± 20.2	2339.5 ± 53.8	120.8 ± 7.6
N9	540.0 ± 29.6	64.4 ± 6.5	24329.5 ± 86.8	9.6 ± 1.3	467.2 ± 36.1	22759.8 ± 124.6	4695.2 ± 23.8	55.8 ± 17.2	338.6 ± 20.9	2325.3 ± 45.2	112.4 ± 9.6
L1	426.3 ± 32.7	165.0 ± 3.3	18652.7 ± 8.1	0	409.3 ± 34.1	28381.9 ± 81.2	4394.2 ± 49.1	134.9 ± 9.5	1607.6 ± 23.8	2193.0 ± 53.8	93.8 ± 3.7
L2	375.6 ± 30.9	149.7 ± 6.9	17165.6 ± 5.3	0	403.7 ± 39.2	27323.5 ± 74.1	4283.8 ± 31.6	131.5 ± 12.1	1578.1 ± 17.6	2295.2 ± 53.1	72.7 ± 5.0
L3	310.7 ± 29.6	179.3 ± 4.7	16987.4 ± 36.2	0	317.0 ± 40.5	27067.7 ± 86.7	4067.0 ± 41.3	172.8 ± 10.5	1714.3 ± 25.4	2296.9 ± 52.6	85.1 ± 5.5
L4	353.3 ± 31.3	80.9 ± 6.2	16935.1 ± 33.0	0	293.2 ± 43.0	26813.0 ± 88.1	5009.4 ± 36.0	92.4 ± 12.3	987.5 ± 17.1	2506.4 ± 50.9	86.9 ± 7.7
L5	390.5 ± 28.3	112.9 ± 9.9	17963.1 ± 4.4	0	302.5 ± 48.4	27203.9 ± 80.6	5133.8 ± 40.1	127.9 ± 9.3	1025.6 ± 17.0	2563.0 ± 51.3	95.2 ± 9.0
L6	402.3 ± 31.4	81.4 ± 4.5	17583.0 ± 12.3	0	310.5 ± 38.7	27084.9 ± 56.1	4860.9 ± 32.7	127.7 ± 11.2	1013.0 ± 16.0	2546.7 ± 54.1	94.5 ± 9.3
F1	673.4 ± 29.9	81.5 ± 7.7	19112.7 ± 10.5	6.5 ± 0.1	674.5 ± 25.6	19030.4 ± 66.6	4719.7 ± 28.9	202.0 ± 13.3	468.0 ± 30.0	2534.8 ± 50.5	75.6 ± 10.8
F2	702.3 ± 33.5	77.3 ± 3.9	18819.5 ± 76.1	8.0 ± 0.5	692.2 ± 28.6	18666.8 ± 102.5	4723.9 ± 25.1	136.1 ± 3.7	426.1 ± 19.4	2535.6 ± 51.4	74.7 ± 2.5
F3	655.8 ± 25.2	86.3 ± 2.5	18968.5 ± 64.4	12.2 ± 1.8	640.6 ± 32.9	18703.4 ± 120.0	4507.9 ± 32.7	173.1 ± 11.3	342.7 ± 22.4	2721.7 ± 43.6	81.0 ± 9.9
F4	439.0 ± 30.5	93.0 ± 5.1	17096.9 ± 17.6	7.9 ± 0.6	446.1 ± 46.5	21096.9 ± 67.2	4756.0 ± 42.0	115.6 ± 10.7	754.0 ± 19.6	2447.2 ± 54.4	72.2 ± 4.5
F5	409.1 ± 29.8	75.8 ± 7.4	16974.3 ± 28.5	6.3 ± 0.9	401.9 ± 57.4	19289.7 ± 79.5	4976.2 ± 32.3	99.6 ± 12.2	686.7 ± 18.1	2253.5 ± 56.0	57.2 ± 7.1
F6	484.3 ± 31.1	88.5 ± 6.4	18635.7 ± 31.9	6.7 ± 0.7	449.1 ± 42.1	20822.3 ± 47.8	5089.0 ± 33.4	116.9 ± 11.0	589.9 ± 19.4	2508.4 ± 43.7	84.3 ± 8.5
F7	391.1 ± 30.2	23.0 ± 7.8	19139.6 ± 8.9	3.3 ± 0.8	349.0 ± 43.6	17784.3 ± 81.6	4139.0 ± 32.7	100.8 ± 12.3	179.0 ± 21.6	2545.3 ± 49.5	20.1 ± 1.1
F8	331.4 ± 32.3	22.0 ± 5.0	18902.7 ± 1.1	3.9 ± 0.2	289.2 ± 47.1	19560.1 ± 23.3	4399.1 ± 52.8	94.2 ± 9.3	187.1 ± 24.2	2356.7 ± 50.6	18.7 ± 4.9
F9	372.0 ± 33.4	22.2 ± 3.5	20345.0 ± 0.1	3.0 ± 0.3	318.0 ± 41.1	18550.7 ± 24.1	4429.4 ± 36.8	120.0 ± 11.3	172.7 ± 20.4	2223.8 ± 49.5	22.3 ± 4.0

matrix $\mathbf{X}$ (24 × 11) is decomposed into two matrices, $\mathbf{T}$ and $\mathbf{L}$, such that $\mathbf{X} = \mathbf{T}\mathbf{L}^T$.

The matrix $\mathbf{T}$, known as the "score" matrix, represents the positions of the sample in the new coordinate system (PCs coordinate system). $\mathbf{L}$ is the "loading" matrix, whose columns describe how the new axes, *i.e.* the PC's, are built from the old axes.

Results

Mineral content in cigarettes

The above-mentioned elements were analyzed in 24 cigarette samples. The results, expressed in $\mu g\ g^{-1}$, were obtained from triplicate measurements by ICP OES (shown in Table 4). Looking at these values, Ca, K, Mg, and P are the major elements in all samples. Al, Fe, and Na are elements with an intermediate content. Most of the other analyzed metals appear with values lower than 200 $\mu g\ g^{-1}$. These concentrations of metals obtained in the cigarette samples, are dependent on a number of factors, including the soil in which the tobacco is grown, the variety of tobacco, and how the tobacco leaves are processed.[2]

Data analysis

Initially, in a HCA and PCA study, 11 variables (Al, Ba, Ca, Cu, Fe, K, Mg, Mn, Na, P and Sr) were considered. Once the types of cigarettes were known "*a priori*", this information was used to identify which variables were useful for discrimination. Upon a visual inspection of the samples scatter plots, 5 variables (Sr, Ca, Cu, K and Na) were selected for discrimination among the evaluated types of cigarettes. Figure 1 shows scatter plots as an illustration.

Hierarchical clustering analysis (HCA)

HCA results are shown in Fig. 1. This dendrogram was obtained by calculating the Euclidean distance among samples and grouping them by the complete linkage method.[11] There are 3 main groups for a similarity index of approximately 0.55. As can be seen, the classes F, N, and L are clearly discriminated and distinguished. All samples from each group are clustered, while showing no mixing in different types of cigarettes.

Principal-component analysis (PCA)

PCA was applied to a matrix of 5 analytical parameters for 24 samples. The result of PCA projection of the data from the original five-dimensional space into the plane of the first two principal components, carrying 79.7% of the total variance (Table 5), is shown in Fig. 2.

Table 4 contains the loadings for three first-principal components with their variances. PC_1 describes 50.2% of the variance in the data set, and has high positive loadings for the Na and K and negative loadings for Ca and Cu. On the other hand, PC_2, accounting for 29.5% of the original information has a significant contribution from Sr. PC_3 (13.6% of total variance) describes the information from Cu. Together, these three PCs account for 93.3% of the total variance in the data.

In the scatter plot of PC_1 *vs.* PC_2 (Fig. 2) it can be seen that class L is clearly separated from the other two classes, N and F, in PC_1. The N and F samples tend to have negative scores, while in L samples they are positive.

The first-principal component was responsible for the separation of light cigarette samples, while in the PC_2 it can be seen that the flavored cigarettes have negative scores, being discriminated from the others. All samples are grouped correctly according to their common source, confirming PCA to be a powerful method for uncovering information contained in cigarettes elemental content analyses.

Through Fig. 3 it can be seen that the group of N samples (PC_1) tends to have higher contents of Ca, Cu and Sr, and simultaneously lower contents of K and Na. The position of group-L cigarettes in PC_1 indicates that are highly dependent on K and Na present in high concentrations and low dependencies of Ca, Cu and Sr. The group of F samples, besides having fairly high concentration of Ca, tends to have a higher concentration of Cu, but lower on Sr.

Therefore, with the obtained separation among the groups, this data set could be used as a training set to classify future unknown samples of cigarettes and to check for their authenticity. This was not done in this work, due to the sample size.

Conclusions

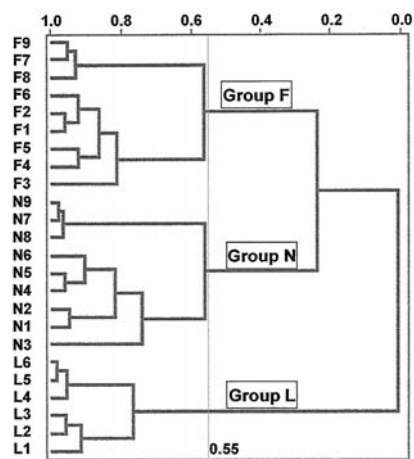ICP OES allows for the multi-element determination of many

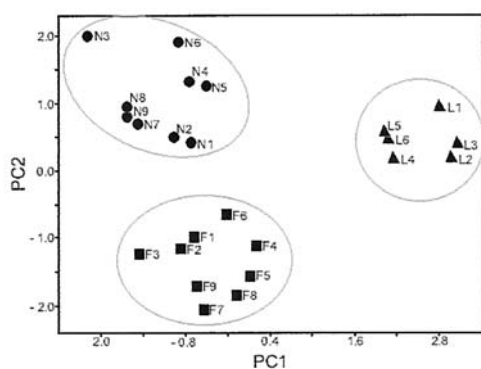Fig. 1   Dendogram presenting the results of hierarchical clustering for samples.

Fig. 2   Projection of a sample into the PC_1 and PC_2 planes. Codes for cigarette samples: (●) conventional, (▲) light, and (■) flavored.
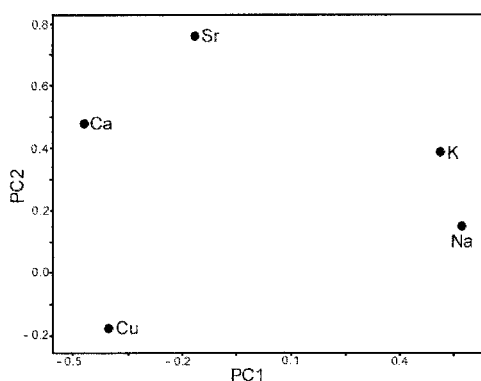

Fig. 3   Projection of the variables into the PC_1 and PC_2 planes.

Table 5   Loadings for three PCs and total variance (%) in a matrix of 24×11

Component	PC_1	PC_2	PC_3
Ca	−0.4713	0.4771	−0.0817
Cu	−0.4042	−0.1767	0.8941
K	0.5103	0.3832	0.2613
Na	0.5711	0.1473	0.3494
Sr	−0.1668	0.7567	0.0600
Variance explained	57.7	33.9	15.6
Total variance (%)	50.1	29.5	13.5
Cumulative variance	50.1	79.7	93.3

brands (N), light (L) and flavored (F). Ca, Cu, K, Na and Sr were the most important variables for discriminate cigarette types and to observe their authenticity.

References

1. M. Chiba and R. Masironi, *Mark Bulletin of the World Health Organization*, **1992**, *70(2)*, 269.
2. S. J. Stohs, D. Bagchi, and M. Bagch, *Inhalation Toxicol.*, **1997**, *9*, 867.
3. Agência Nacional de Vigilância Sanitária (ANVISA), www.anvisa.gov.br/legis/resol/105-01rdc.htm 25-02-2005.
4. F. V. Silva, M. Y. Kamogawa, M. M. C. Ferreira, J. A. Nóbrega, and A. R. A. Nogueira, *Eclet. Quím.*, **2002**, *27*, 91.
5. E. C. Ferreira, S. H. B. G. Rodrigues, M. M. C. Ferreira, J. A. N. Nóbrega, and A. R. A. Nogueira, *Eclet. Quím.*, **2002**, *27*, 77.
6. S. Frías, J. E. Conde, J. J. Rodríguez-Bencomo, F. García-Montelongo, and J. P. Pérez-Trujillo, *Talanta*, **2003**, *59*, 335.
7. A. Alcázar, F. Pablo, M. J. Martín, and A. G. González, *Talanta*, **2002**, *57*, 45.
8. S. Razic, A. Onjia, and B. Potkonjak, *J. Pharm. Biomed. Anal.*, **2003**, *33*, 845.
9. M. F. Giné-Rosias, "*Espectrometria de emissão atômica com plasma acoplado indutivamente*", **1998**, Vol. 3, Piracicaba, CENA, 143.
10. G. Thiel and K. Danzer, *Fresenius J. Anal. Chem.*, **1997**, *357*, 553.
11. M. A. Sharaf, D. L. Illman, and B. R. Kowalski, "*Chemometrics*", **1986**, John Wiley & Sons, New York.
12. K. R. Beebe, R. J. Pell, and M. B. Seasholtz, "*Chemometrics: A Pratical Guide*", **1998**, Wiley, New York.
13. M. M. C. Ferreira, A. M. Antunes, M. S. Melgo, and P. L. O. Volpe, *Quím. Nova*, **1999**, *22(5)*, 724.

samples in a simple and fast way. The generated data can be interpreted easily by using chemometric tools. Good results were obtained using HCA and PCA to differentiate cigarette samples.

The results of the elemental content present in cigarette samples can be used in the discrimination of conventional