Analytica Chimica Acta 634 (2009) 172–179

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Prediction of sensory properties of Brazilian Arabica roasted coffees by headspace solid phase microextraction-gas chromatography and partial least squares

J.S. Ribeiro[a,c], F. Augusto[b], T.J.G. Salva[c], R.A. Thomaziello[c], M.M.C. Ferreira[a,*]

[a] Theoretical and Applied Chemometrics Laboratory, Chemistry Institute, University of Campinas, P.O. Box 6154, 13083-970 Campinas, SP, Brazil
[b] Gas Chromatography Laboratory, Chemistry Institute, University of Campinas, P.O. Box 6154, 13083-970 Campinas, SP, Brazil
[c] Centro de Análise e Pesquisa Tecnológica do Agronegócio do Café "Alcides Carvalho", Instituto Agronômico de Campinas, P. O. Box 28, 13001-970 Campinas, SP, Brazil

ARTICLE INFO

Article history:
Received 23 September 2008
Received in revised form 11 December 2008
Accepted 12 December 2008
Available online 25 December 2008

Keywords:
Processed coffee
Solid phase microextraction
Correlation optimized warping
Principal component analysis and partial least squares

ABSTRACT

Volatile compounds in fifty-eight Arabica roasted coffee samples from Brazil were analyzed by SPME-GC-FID and SPME-GC-MS, and the results were compared with those from sensory evaluation. The main purpose was to investigate the relationships between the volatile compounds from roasted coffees and certain sensory attributes, including body, flavor, cleanliness and overall quality. Calibration models for each sensory attribute based on chromatographic profiles were developed by using partial least squares (PLS) regression. Discrimination of samples with different overall qualities was done by using partial least squares-discriminant analysis (PLS-DA). The alignment of chromatograms was performed by the correlation optimized warping (COW) algorithm. Selection of peaks for each regression model was performed by applying the ordered predictors selection (OPS) algorithm in order to take into account only significant compounds. The results provided by the calibration models are promising and demonstrate the feasibility of using this methodology in on-line or routine applications to predict the sensory quality of unknown Brazilian Arabica coffee samples.

According to the PLS-DA on chromatographic profiles of different quality samples, compounds 3-methypropanal, 2-methylfuran, furfural, furfuryl formate, 5-methyl-2-furancarboxyaldehyde, 4-ethylguaiacol, 3-methylthiophene, 2-furanmethanol acetate, 2-ethyl-3,6-dimethylpyrazine, 1-(2-furanyl)-2-butanone and three others not identified compounds can be considered as possible markers for the coffee beverage overall quality.

1. Introduction

The establishment of mechanisms that allow the evaluation, assurance and certification of the quality of food products is an indispensable strategy for maintaining commercial competitiveness. The great number of norms created by international organizations, like the International Organization for Standardization (ISO), indicates the importance of the quality guarantee determined by a set of parameters, frequently used in commercial transactions [1]. These measurable parameters must be faced as an essential element to improve the aggregate value of the agro-industrial production worldwide. To attain an objective guarantee, research has been carried out for better evaluation of the coffee beverage in order to correlate its quality with physicochemical characteristics and the chemical composition of green or roasted beans [2–7].

Commonly, the quality of coffee is evaluated according to criteria such as bean size, color, shape, cupping and number of defects [2,6,8,9]. However, cupping, also known as cup tasting, is still the most widespread technique employed to evaluate the final quality of this product. Arabica coffee, generally regarded as superior to Robusta coffee in terms of sensory attributes, accounts for approximately 70% of world production of this commodity [10].

Flavor plays an important role in sensory analyses and could be considered a "fingerprint" of products, but despite its importance, there are few studies that correlate this characteristic with the final quality of coffee beverage [9]. This correlation, using multivariate analysis [11], is an excellent tool in the quality control of foods and agricultural products and is being applied successfully in analyses of hazelnut, vinegar, juices and wine [12–16].

The chemistry of coffee flavor is highly complex and is still not completely understood. The main families of chemical compounds

* Corresponding author. Tel.: +55 19 3521 3102; fax: +55 19 3521 3023.
E-mail address: marcia@iqm.unicamp.br (M.M.C. Ferreira).

0003-2670/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2008.12.028

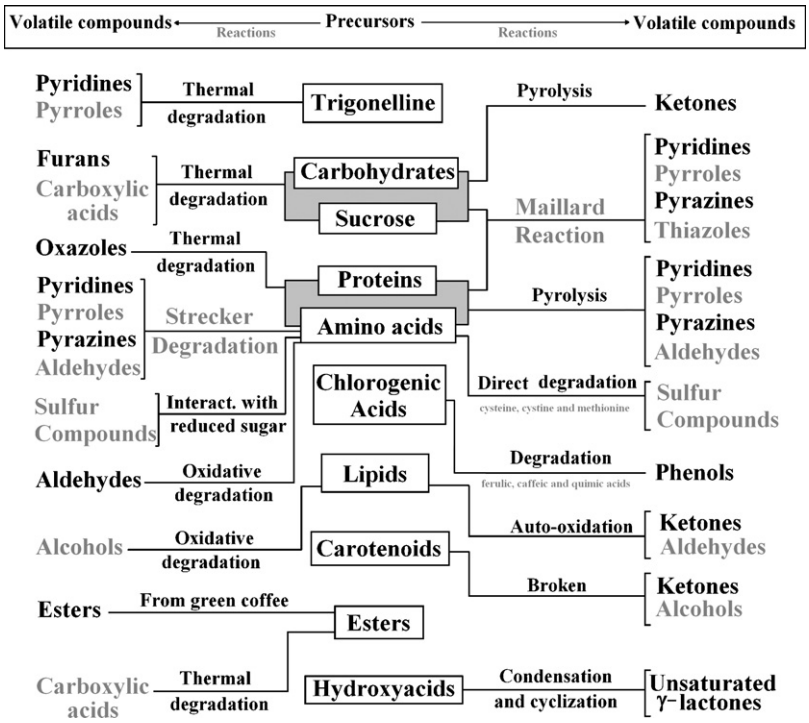

Fig. 1. Schematic representation of the main volatile formation reactions during the coffee roasting process.

found in green coffee, and responsible for the volatiles in roasted coffee, are alkaloids like trigonelline, chlorogenic acids, carbohydrates, free sugars like sucrose, lipids and proteins. During the roasting process, the composition of coffee beans is drastically changed and several hundreds of substances associated with coffee flavor and taste are formed [17]. A general description of the volatile compounds, their precursors and the main volatile formation reactions are shown in Fig. 1.

Several efforts have been made to identify the main volatile compounds responsible for the real flavor of roasted coffee [18–25]. However, the question of which volatiles are the most relevant contributors for the quality of coffee has not yet been elucidated.

Among the analytical techniques used to analyze and separate volatile fractions of different products, gas chromatography has been established as one of the most important. Coupled with gas chromatography, solid phase microextraction (SPME) has been shown to be an excellent sampling method, allowing simultaneous extraction and concentration of analytes from sample matrices [26].

With the aim to identify characteristic volatile compounds that could be responsible for prediction of certain sensory attributes of Brazilian Arabica coffee, chromatographic profiles and sensory profiling were compared in this work using chemometric data treatment.

2. Materials and methods

2.1. Coffee samples

Fifty-eight Arabica green (not roasted) coffee samples from different origins were supplied by Instituto Agronômico de Campinas. Flat coffee beans were visually inspected, and those with defects like black, insect-damaged, immature and broken were excluded. The roasting process was carried out in a gas faired drum roaster (Pinhalense S/A Máquinas Agrícolas) to the medium roast point. Roasted coffee samples were packed in films consisting of plastic (polystyrene and polyethylene) and aluminum, to avoid aroma losses and contamination by external substances, and stored at −5 °C for a maximum period of 48 h before chromatographic analyses.

2.2. Sensory analysis of the coffees

All 58 Arabica coffee samples were evaluated by two cuppers. The cup quality was assessed by flavor, body, cleanliness and overall quality using sample preparation according to Brazilian legislation (Normative instruction no. 8, from 11 June 2003) obtained from www.pr.gov.br/claspar/pdf/cafebenef008_03.pdf.

Thus, for the four sensory attributes selected for evaluation a fivepoint scale was adopted, in such a way that each of the attributes, according to the degree of sensory magnitude perceived were given corresponding scale points, e.g., the cleanliness classifications 'rio' (1) and 'strictly soft' (5) defined the extreme scores on the rating scale.

2.3. SPME devices and GC-FID parameters

SPME fibers coated with 65-μm thick polydimethylsiloxane/divinylbenzene (PDMS/DBV) and the manual holder were purchased from Supelco (Bellefonte, PA). The fibers were conditioned according to the SPME data Sheet (T7941231) from Supelco in the GC injector port. The analyses were performed on a G-6850 GC-FID system (Agilent, Wilmington, DE) fitted with a HP-5 capillary column (30 m × 0.25 mm × 0.25 μm). Helium (1 mL min^{-1}) was the carrier gas. The oven temperature was programmed as follows: 40 °C → 5 °C min^{-1} → 150 °C → 30 °C min^{-1} → 260 °C. The injection port was equipped with a 0.75 mm i.d. liner and the injector was maintained at 220 °C in the splitless mode. Under these conditions, no sample carry-over was observed on blank runs conducted between extractions.

Identification of the extracted analytes was performed on a HP-5890 gas chromatographer (Hewlett-Packard, Wilmington, DE, USA) equipped with a HP-5973 mass-selective detector fitted with the same column and operated under the same conditions as

Table 1
Main compounds identified from the mass analyses by comparison of their MS spectra with those of the NIST MS data base and literature.

Peaks	Retention time (min)	Compounds	Main *m*/*z* ions observed in MS spectra[a,b]	Math
1	1.93	Methanethiol	47 (B)	923
2	2.02	Acetonitrile	41(B)	759
3	2.22	2-Methylpropanal	43(B), 72	879
4	2.32	2,3-Butadione	43(B), 86	973
5	2.44	2-Methylfuran	82(B), 53, 39	909
6	2.68	3-Methylbutanal	44(B), 58, 39	866
7	2.68	2-Methylbutanal	41(B), 57, 39	916
8	2.74	Thiophene	84(B), 58	877
9	3.1	2,3-Pentadione	43(B), 29, 57	928
10	3.2	Acetic acid	43(B), 60	949
11	3.25	2,3-Pentanone	43(B), 57, 100	928
12	3.45	2,5-Dimethylfuran	96(B), 43, 53	928
13	3.55	3-Methylpyridazine	94(B), 39, 65	865
14	3.68	Methyl acetate	43 (B)	915
15	3.9	Pyrazine	80(B), 53	892
16	4.17	1-Methylpyrrole	81(B)	951
17	4.4	Pyridine	79(B), 52	978
18	4.4	1H-pyrrole	67(B), 39	933
19	4.6	4,5-Dimethyloxazole	97(B), 43, 55	850
20	4.7	Toluene	91(B)	958
21	4.93	3-Methylthiophene	97(B)	890
22	5.1	2,3-Hexanedione	43(B), 71	892
23	5.36	3,4-Hexanedione	57(B)	907
24	5.52	Dihydro-2-methyl-3(H)furanone	43(B), 72, 100	904
25	5.77	Methylpyrazine	94(B), 67	954
26	5.83	2-Fururyl methyl ether	81(B), 53, 112	918
27	5.88	3-Methylphenol (*m*-cresol)	108(B)	827
28	5.95	Furfural (furancarboxaldehyde)	96(B), 39	966
29	6.0	2,*N*-Dimethylpyrrole	94(B)	845
30	6.4	Trimethyloxazole	111 (B), 43, 55	802
31	6.42	2-Propenyl-2-furan	108(B), 79	830
32	6.65	2-Furanmethanol	98(B), 41, 53, 81, 69	950
33	6.9	3-Methylbutanoic acid	60(B), 87	789
34	6.9	2-Methylbutanoic acid	74(B), 41, 57	834
35	7.72	Furfuryl formate	81(B), 126, 53	880
36	7.84	2-Furanmethanethiol	81(B), 114	864
37	7.94	2,*N*-Dimethylpyrazine	108(B), 42	716
38	8.2	Ethylpyrazine	107(B)	900
39	8.5	Butyrolactone	42(B), 67, 86	919
40	8.6	2,*N*-Dimethylpyrrole	94 (B)	979
41	8.65	Ethenyl pyrazine	106(B), 52, 79	820
42	9.1	2-*n*-butylfuran	81(B), 43, 53,124	778
43	9.24	Benzaldeyde	106(B), 77, 51	893
44	9.38	5-Methyl-2-furancarboxaldeyde	110(B), 53	963
45	10.02	3-Methyl-2(5H)furanone	41(B), 98, 69	941
46	10.29	2-Furanmethanol acetate	81(B), 98, 43,140	954
47	10.36	2-Ethyl-*n*-methyl pyrazine	121(B)	894
48	10.6	1-Methyl-1H-pyrrole-2-carboxaldeyde	109(B), 53, 80	894
49	10.65	2-Proponylfuran	95(B), 124	872
50	11.17	2-Ethenyl-*n*-methylpyrazine	120(B), 52	829
51	11.43	2-Acetyl-5-methylfuran	109(B), 124	936
52	11.48	Benzeneacetaldeyde	91(B), 120	932
53	12.14	1-(2-Furanyl)-2-butanone	57 (B), 81, 53, 138	829
54	12.3	Acetylfuran	95(B), 110	725
55	12.5	2-Acetylpyrrole	94(B), 109, 66	934
56	12.68	2-Acetyl-*N*-methylpyrrole	108(B), 123	905
57	12.77	2-Ethyl-3,6-dimethylpyrazine	135(B)	931
58	12.85	*p*-Guaiacol	121(B), 135	915
59	13	3-Ethyl-2-hydroxycyclopenten-1-one	126(B)	752
60	13.2	Maltol	121(B), 135	815
61	14.68	N/I	–	–
62	15.52	2-Furfuryl-5-methylfuran	162(B), 91	742
63	15.6	1-Furfurylpyrrole	81(B), 147	936
64	15.8	N/I	–	–
65	16.9	N/I	–	–
66	18.18	3,4-Dihydroxyacetophenone	137(B), 152	834
67	18.28	4-Ethylguaiacol	137(B), 152	855
68	18.67	Furfuryl methyl disulfide	81(B)	823
69	18.67	Difurfuryl ether	81 (B)	929
70	18.82	4-Vinylguiaicol	135(B), 150, 77, 107	894

N/I: not identified.
[a] *m*/*z* of each compound from the NIST data bank.
[b] Base peak c.

the GC-FID. GC–MS data treatment was carried out using the Automated Mass Spectral Deconvolution and Identification System (AMDIS) v. 2.61 software and the NIST Mass Spectral Search Program v. 1.6d (NIST, Washington, DC, USA), as well as making comparisons with earlier reports on the volatile compounds of roasted coffee [27–29].

2.4. General SPME procedure for sampling and injection

Ground coffee (250 mg) and 2 mL of saturated aqueous sodium chloride solution were transferred to a septum-sealed glass sample vial (5 mL). After 10 min of sample/headspace equilibration under agitation of 900 rpm at 42.5 °C, the fibers were exposed to the sample headspace for 22 min. After sampling, the fiber was immediately placed in the injection port of the GC and the analytes were thermally desorbed at 220 °C. All analyses were carried out in triplicate.

2.5. Chemometric data treatment

The original chromatographic profiles were organized into a matrix format $\mathbf{X}$ ($I \times J$), where each replicate represented one sample. Data analysis was carried out in Matlab 6.5 software (The MathWorks Co., Natick, MA, USA) using the computational package PLS_Toolbox (Eigenvector Research, Inc., PLS_Toolbox version 3.02.) [30]. The chromatograms' alignments were performed using the correlation optimized warping (COW) algorithm [31] obtained from www.models.kvl.dk/source/. The chromatogram with the best peak resolution was used as the reference vector. The chromatograms were divided into 10 regions and, for each region, the segment length and the slack–parameter used were 10 and 1, respectively.

Each aligned profile was normalized to unit length, smoothed by the Savitzky–Golay algorithm, with a window size of 10 points followed by taking the absolute values of first derivative [32] and, lastly, the matrix $\mathbf{X}$ was column-wise autoscaled and the vector $\mathbf{y}$ of sensory notes was mean-centered. Variable selection was carried out by the ordered predictors selection (OPS) method [33].

The regression methods used for data treatment were partial least squares (PLS) [11] and partial least squares-discriminant analysis (PLS-DA) [34].

In PLS regression, a dependent variable, $\mathbf{y}$, is modeled using latent variables (LV), maximizing the covariance between $\mathbf{X}$ and $\mathbf{y}$. The PLS model can be presented as follows [35]:

$$\mathbf{X} = \mathbf{T}_k\mathbf{P}_k^{\mathbf{T}} + \mathbf{E} \quad (1)$$

$$\mathbf{y} = \mathbf{T}_k\mathbf{q} + \mathbf{f} \quad (2)$$

where $\mathbf{X}(I, J)$ represents the data matrix (chromatograms), vector $\mathbf{y}(I, 1)$ is a dependent variable (sensory analysis notes), $\mathbf{T}(I, k)$ is the score matrix, $\mathbf{P}^{\mathbf{T}}(k, J)$ denotes the transposed loadings matrix, $\mathbf{q}(I, 1)$ is a loading vector and $\mathbf{E}(I, J)$ and $\mathbf{f}(I, 1)$ are the residuals (k is the number of latent variables).

In order to predict y_i for a new autoscaled chromatogram $\boldsymbol{x}_{i(as)}$ $(1, J)$, the following equation can be used:

$$\hat{y}_i = \bar{y} + \mathbf{x}_{i(as)}\mathbf{b} \quad (3)$$

where $\hat{y}_i$ is the predicted dependent value for the ith new sample, $\bar{y}$ denotes the mean of the dependent values for the calibration samples, and $\mathbf{b}(J, 1)$ is the computed vector of PLS regression coefficients [36]:

$$\mathbf{b} = \mathbf{W}(\mathbf{P}^{\mathbf{T}}\mathbf{W})^{-1}\mathbf{q} \quad (4)$$

where $\mathbf{W}$ is the matrix of loading weights.

While PLS is used as a calibration method, PLS-DA is a discrimination method where the model is built between the matrix $\mathbf{X}$ and the matrix of known classes $\mathbf{Y}$. In PLS-DA each class is described by a column in $\mathbf{Y}$. To each class variable is assigned a value 1 or 0 depending on to which class an object belongs.

The optimal model complexity, i.e., the number of latent factors (k) in the PLS or PLS-DA models, can be determined by a cross-validation procedure. Leave-one-out cross-validation is performed by excluding one chromatogram at a time, the model is built and the estimated class ($\hat{y}_i$) for each sample is used to calculate the root mean square error of cross-validation (RMSECV). The performance of the final PLS or PLS-DA model is evaluated in terms of RMSECV (Eq. (5)), computed for different numbers of latent variables, and the correlation coefficient of cross-validation (r_{cv}) (Eq. (6)).

$$\mathrm{RMSECV}_k = \sqrt{\frac{\sum_{i=1}^{I}(y_i - \hat{y}_i)^2}{I}} \quad (5)$$

In Eq. (5), y_i is the measured response of the ith sample, $\hat{y}_i$ is a predicted response from a calibration equation obtained for the data without the ith sample and I is the number of samples in the calibration set. The optimal PLS model corresponds to the number of latent factors resulting in the lowest RMSECV.

The correlation coefficient between the estimated values in cross-validation and the experimental values of the reference method is

$$r_{cv} = \frac{\sum_{i=1}^{I}(\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\left[\sum_{i=1}^{I}(\hat{y}_i - \bar{\hat{y}})\right]^{1/2}\left[\sum_{i=1}^{I}(y_i - \bar{y})\right]^{1/2}} \quad (6)$$

where $\bar{\hat{y}}$ is the mean estimated response.

Once the model has been internally validated and tested by an external data set, it can be used for the prediction of new samples. For the external validation set, the root mean square error of prediction (RMSEP) is used:

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{i=1}^{I}(y_i^p - \hat{y}_i^p)^2}{I_p}} \quad (7)$$

where I_p is the number of samples in the test set and $\hat{y}_i^p$ and y_i^p are the predicted and measured response values for the test set samples.

3. Results and discussion

3.1. Mass detection of volatile compounds

Two of the fifty-eight samples supplied by the Agronomic Institute of Campinas were analyzed by mass spectrometry. These two samples represented products with the lowest (1.5) and the highest (4.5) notes in the overall quality attribute. More than 250 volatile compounds were detected in these mass analyses. Table 1 shows the 70 principal compounds and their respective retention times, nomenclature, fragmentation and number of match. These compounds are usually encountered in mass spectrometry analyses of roasted coffee [27–29].

3.2. Partial least squares-discriminant analysis

Fig. 2A reports the original overlapped chromatograms of volatile compounds obtained from a PDMS/DVB fiber. It is visible that a pretreatment is necessary to correct peak shift. Fig. 2B shows the profiles pretreated by the COW method, as indicated in Section 2.

In order to have insight into which peaks could be involved in the discrimination of the Arabica coffee samples according to their overall quality, a subset of 20 samples (60 replicates) were selected. Eleven of them had the best notes of overall quality, between 4 and 4.5, and were designated as class one, while the other nine samples

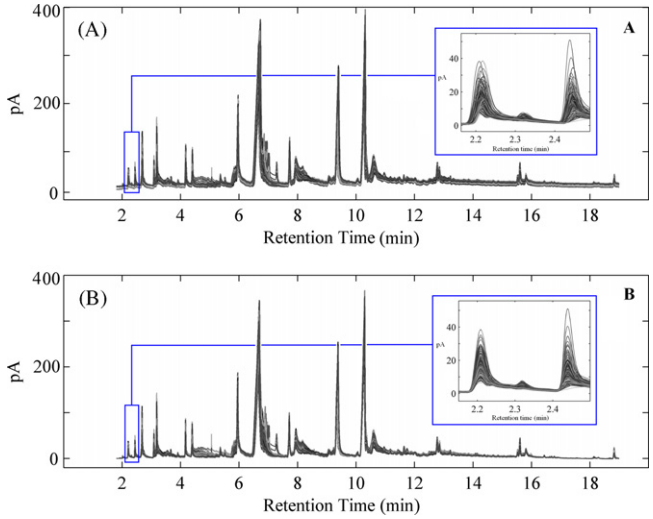

Fig. 2. Original chromatograms (A) and pre-treated chromatograms (B). The indicated regions in the two figures were expanded to show a detail of the alignment.

with notes below 2.5 composed class two. The OPS method was applied to this pretreated subset in order to identify the main peaks involved in discriminating the two classes. The 13 selected peaks are indicated as vertical lines in Fig. 3.

PLS-DA was applied to the selected peaks and the LV1 × LV2 × LV3 scores plot are indicated in Fig. 4A. In this plot, low overall quality coffee samples are distinguished from those of high quality. Samples of low overall quality, with negative scores in LV1 (51.28% of the selected information), are located on the left, well separated from samples with high overall quality, with positive scores, on the right side. Fig. 4B shows the LV1 versus LV2 loadings and scores biplot. The numbers indicate the peaks defined in Fig. 3.

Considering the LV1 loadings in Fig. 4B, it can be seen that when compounds 3-methypropanal (**3**), 2-methylfuran (**5**), furfural (**28**), furfuryl formate (**35**), 5-methyl-2-furancarboxaldehyde (**44**) and 4-ethylguaiacol (**67**) appear in higher amounts, the overall quality of the Arabica coffee is increased. On the other hand, when compounds such as 3-methylthiophen (**21**), 2-furanmethanol acetate (**46**), 2-ethyl-3,6-dimethylpyrazine (**57**) and 1-(2-furanyl)-2-butanone (**53**) are more abundant, the overall quality of the product drops. The compounds not identified (N/I) (**61**, **64** and **65**), also important to discriminate the quality of the Arabica coffee, are now being identified by new mass spectrometric analyses.

According to Arctander [37], 3-methylpropanal has a sweet and fruity flavor and is considered one of the key odorant compounds of roasted Arabica ground coffee [38]. Vernin [39] described 2-methylfuran as having a burnt material aroma with a sweet odor very similar to that of coffee. The flavor of furfural is similar to that of bread and caramel at certain concentrations, still possessing a bitter taste character [40]. The flavor of furfuryl formate is associated to malt, fruits and cereal aromas, while 5-methyl-2-furancarboxyaldehyde has a spicy, candy and slightly caramel odor [37]. 4-Ethylguaiacol has a smoky and burnt material flavor [41] and 3-methylthiophene is responsible for a slightly sweet off-flavor similar to tinned meat [42], with sulfuric odor like baked onions [43]. The flavor of 2-ethyl-3,6-dimethylpyrazine was described as burnt material and earthy by Wagner et al. [44] and as fermented stuff by Maeztu et al. [45]. However, except for the last three compounds, the others are not classified with high odoriferous activity, i.e., they are not considered as the main potent constituents of the coffee flavor [46]. From the above analysis, it is confirmed that the quality of coffee is rather complex and that the quality information can be effectively enhanced by the presence of less odoriferous compounds.

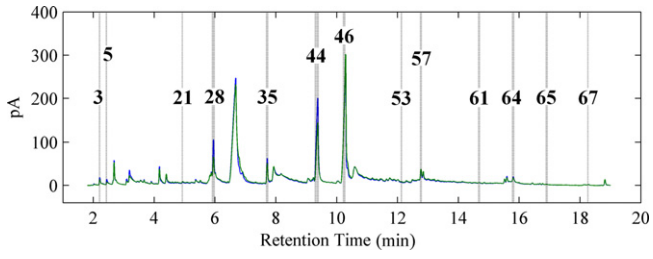

Fig. 3. Peaks selected by the OPS method and used in PLS-DA from chromatograms of good overall quality ( ) and bad overall quality ( ) coffees obtained by using a PDMS/DVB fiber and GC-FID.

More complete information can be given when taking into account the precursor compounds of these selected volatiles. Thus, the great number of furan derivatives indicates the important role that carbohydrates and free sugars, like sucrose, play in the final quality of the beverage. It is well known that furans are formed by thermal degradation of sugars and carbohydrates [47]. In wines, for example, fructose is an important compound that increases its quality [48].

According to Franca et al. [2], the highest quality coffee samples have higher protein levels in comparison to "rio" (low quality) samples. The oxidative degradation of proteins and sugars is the main route for aldehyde formation [49], while sulfur-containing amino acids (cysteine, cystine and methionine) are the precursor of sulfur compounds like 3-methylthiophene [50]. So, lower protein levels in bad quality samples indicate higher degradation ratios of sulfur-containing amino acids and, consequently, higher concentrations of 3-methylthiophene in the headspace.

Pyrazine derivatives are formed by Maillard reactions, Strecker degradation and pyrolysis of hydroxyl amino acids [51], and are considered as natural perfuming of foods [52]. The phenol derivatives are formed by degradation of free phenolic acids during the roasting process [50]. Methoxyphenols, for example, 4-ethylguaiacol, 4-vinylguaiacol and vanillin are among the 22 main compounds responsible for the flavor of roasted coffee [18].

3.3. Regression models

To build the regression models for the four descriptive quantitative sensory analyses (flavor, body, cleanliness and overall quality), the mean values of the notes indicated by the two cuppers were used as the dependent variables (**y**) and 174 chromatograms referring to 58 Arabica coffee samples as independent variables (matrix **X**) were used. Through a *t*-paired test [53], using a confidence limit of 95%, the notes supplied by the cuppers for three attributes (body, cleanliness and overall quality) did not present significant differences. A higher reliable limit was necessary only for the attribute flavor (99%).

To form the calibration sets of each model, 48 samples (144 chromatograms) were randomly selected. Leave five out cross-validation was the method used to select the number of

Table 2
Latent variable numbers, RMSECV and r_{cv} for PLS models.

Model	No. LV[a]	RMSECV[b]	r_{cv}[c]
Flavor	7	0.39 ± 0.05	0.89 ± 0.03
Body	7	0.18 ± 0.02	0.88 ± 0.03
Cleanliness	8	0.32 ± 0.04	0.91 ± 0.02
Overall quality	6	0.38 ± 0.05	0.91 ± 0.03

[a] Latent variable number.
[b] Root mean square error of cross-validation.
[c] Cross-validation correlation coefficient.

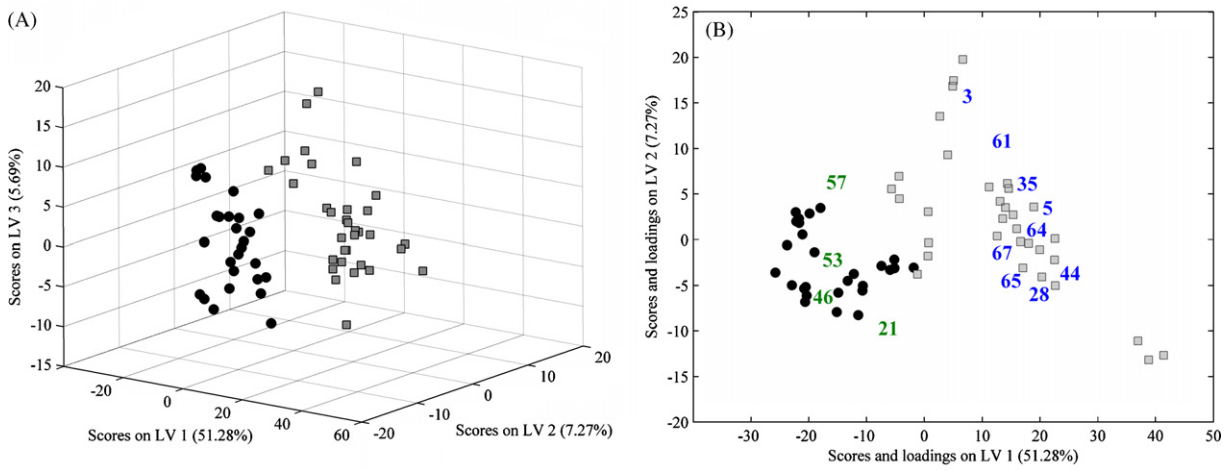

Fig. 4. LV1 × LV2 × LV3 scores plot for the PDMS/DVB fiber (A) and LV1 × LV2 scores and loadings plot (B): (□) Good overall quality coffee samples (class one) and (●) samples with low overall quality (class two): Green numbers (21, 46, 53, 57) represent important peaks for low quality and blue numbers (other numbers) indicate the important peaks for high overall quality. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

components in the models (15 replicates of five samples were left out at a time). The 10 remaining samples, corresponding to 30 chromatograms were used to form the test set.

From the pretreated data matrix (174 × 20,640), the baseline regions without chemical information were removed, with only the chromatographic peaks remaining. The variable selection for the construction of the models was carried out by the OPS method. In this way, from an initial set of 20640 variables, 1350 were selected for the construction of the flavor model (A), 1350 for body (B), 1550 for cleanliness (C) and 1350 for overall quality (D). These variables are indicated as vertical lines in Fig. 5.

The number of latent variables used in the PLS models was determined from the RMSECV values. Table 2 shows the number of latent variables selected for each sensory attribute and the respective statistical parameters RMSECV and r_{cv}.

Using the number of latent variables indicated in Table 2 for all calibration models, it was, in general, possible to describe 95% and 52% of the variance used in blocks **Y** and **X**, respectively. The models were validated by the external data set, composed of 10 samples (30 replicates). Fig. 6 contains the prediction samples of each model, distributed in the cross-validation sets and Table 3 shows the measured and predicted values of the prediction samples for each model. The RMSEP values were 0.32 for flavor, 0.22 for body, 0.28 for cleanliness and 0.34 for overall quality.

Due to the scale used to describe the sensory analysis notes (1–5 points), the relative errors tend to be higher for samples with low notes and decrease when the values became higher. Because of this, when calculating the cross-validation and prediction errors for each replicate, the following criterion was used: if the difference between the cuppers mean notes and the notes predicted

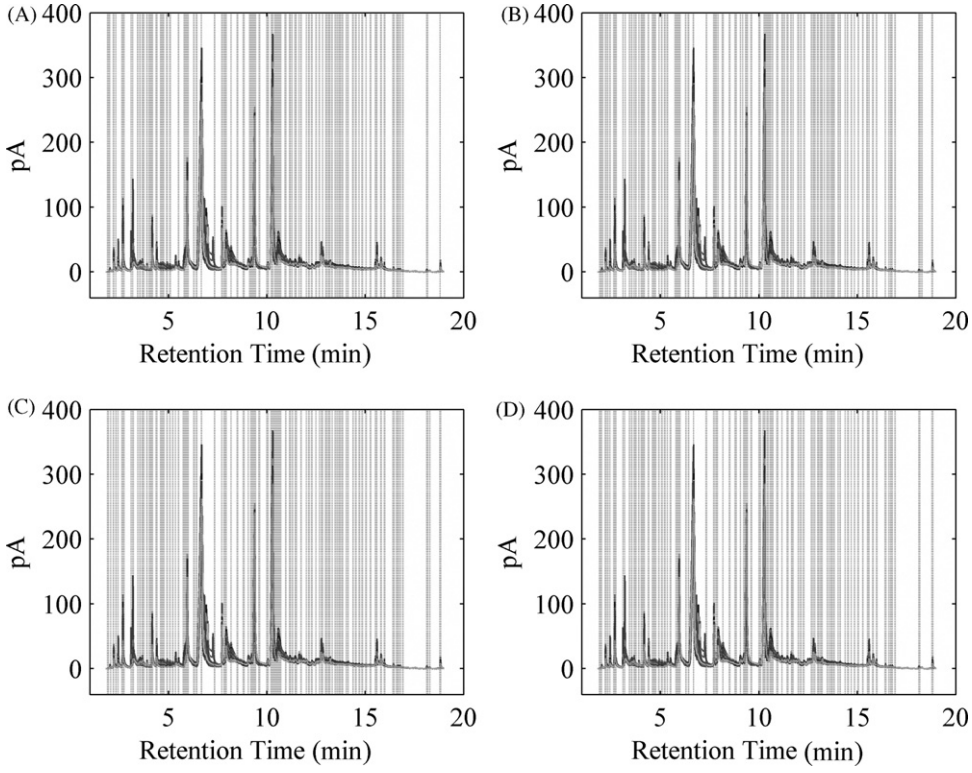

Fig. 5. Regions of the chromatograms selected by the OPS method for the regression models. Flavor (A), body (B), cleanliness (C) and overall quality (D).

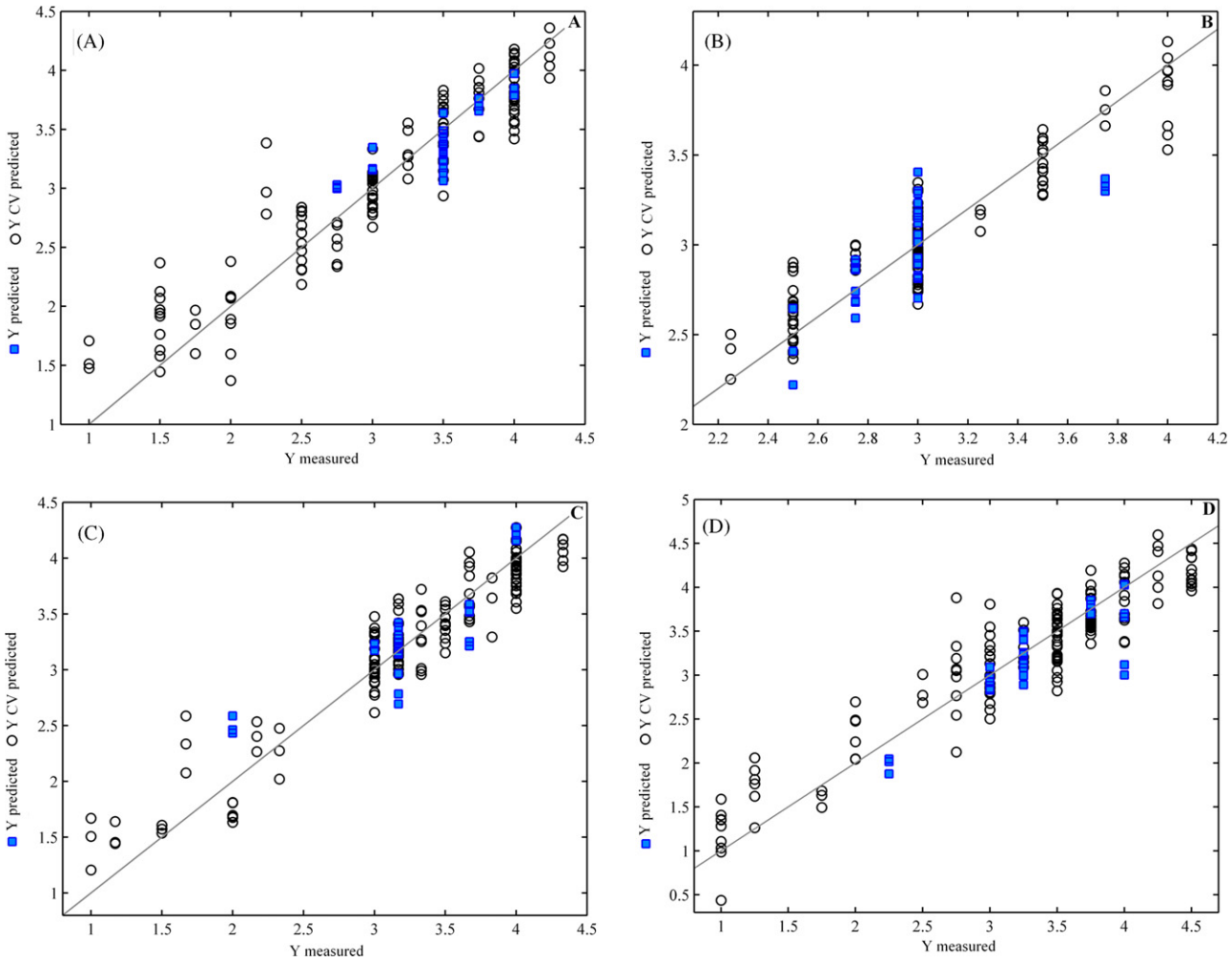

Fig. 6. Plots of measured vs. predicted samples in calibration (○) and prediction (■) sets. Flavor (A), body (B), cleanliness (C) and overall quality (D).

by the regression model was higher than two standard deviations obtained by the difference between the cuppers notes, the cross-validation and predicted notes of the replicates were disregarded. The calibration and prediction errors for each model are indicated in Table 4.

The flavor of coffee is composed of an extremely complex mixture of volatile compounds with different intensities, concentrations and odorific sensations. According to De Maria et al. [54], the same compound could present a positive as well a negative flavor to the beverage depending on its concentration and synergic

Table 4
Number of desregarded replicates predicted in the PLS models for cross-validation and prediction.

	Calibration sets	Prediction sets
Flavor	15	0
Body	0	0
Cleanliness	1	0
Overall quality	2	1

Table 3
Measured values given by the experts and predicted values from the regression models.

Flavor			Body			Cleanliness			Overall quality		
Sam.[a]	Meas.[b]	Pred.[c]	Sam.	Meas.	Pred.	Sam.	Meas.	Pred.	Sam.	Meas.	Pred.
1	2.75	3.01 ± 0.02	11	2.75	2.88 ± 0.03	1	3.17	3.20 ± .0.5	5	4	3.92 ± 0.19
4	4	3.87 ± 0.09	12	3.75	3.33 ± 0.03	21	3.17	2.81 ± 0.14	9	3.75	3.78 ± 0.08
13	3.5	3.27 ± 0.08	15	3	3.33 ± 0.07	22	3.67	3.35 ± 0.21	13	3.25	2.99 ± 0.10
24	3.5	3.34 ± 0.10	16	3	2.82 ± 0.11	24	3.17	3.39 ± .0.2	15	3.25	3.41 ± 0.15
37	3.75	3.70 ± 0.05	21	2.75	2.67 ± 0.07	31	3.17	3.22 ± .0.5	22	4	3.26 ± 0.35
45	3.5	3.41 ± 0.09	29	3	3.19 ± 0.04	37	3.67	3.52 ± 0.01	31	3.25	3.26 ± 0.14
50	3.5	3.52 ± 0.11	36	2.75	2.77 ± 0.11	39	2	2.49 ± 0.08	34	2.25	1.98 ± 0.09
51	3.5	3.33 ± 0.06	41	3	2.84 ± 0.04	42	4	4.21 ± 0.06	35	3	2.91 ± 0.06
57	3	3.22 ± 0.11	45	2.5	2.42 ± 0.21	44	3	3.21 ± 0.03	47	3.25	3.30 ± 0.09
58	3.5	3.18 ± 0.15	46	3	3.06 ± 0.05	56	3.17	3.20 ± 0.10	56	3	2.95 ± 0.13

[a] Prediction samples.
[b] Measured.
[c] Predicted.

effects, when combining with other compounds. Due to all these effects the majority of compounds detected with the SPME-GC-FID were selected by the OPS algorithm for the construction of the calibration models (Fig. 5).

Since the overall quality depends on other attributes, such as flavor, body and cleanliness, it is expected that a coffee well evaluated in terms of overall quality will also have a high evaluation in one or more of the other attributes. However, the same consideration cannot be taken as a rule when comparing flavor, body and cleanliness among themselves, although, in some cases, these attributes could have a certain degree of correlation. Due to the high correlation observed by the cuppers between the attributes of the analyzed samples, most of the selected peaks used for the construction of the regression models were important for the prediction of all sensory attributes.

4. Conclusions

The discriminant analysis (PLS-DA) carried out on the chromatographic profiles of sound beans indicated that the compounds 3-methypropanal, 2-methylfuran, furfural, furfuryl formate, 5-methyl-2-furancarboxyaldehyde, 4-ethylguaiacol, 3-methylthiophene, 2-furanmethanol acetate, 2-ethyl-3,6-dimethylpyrazine, 1-(2-furanyl)-2-butanone and three other not identified compounds (**61**, **64** and **65** from Table 1) could be considered as possible markers for the overall differentiation of coffee beverages.

The regression models (PLS) using chromatographic profiles predicted very well the notes conferred by the cuppers for flavor, body, cleanliness and overall quality of Brazilian Arabica coffees. From the results obtained in this study, the methodology proposed is a promising alternative tool for monitoring coffee beverage evaluation.

Acknowledgements

This work was supported by grants from CAPES and FAPESP.

References

[1] W.J. Keegan, M.C. Green, Princípios de marketing global, Ed. Saraiva, São Paulo, 2000.
[2] A.S. Franca, J.C.F. Mendonça, S.D. Oliveira, LWT 38 (2005) 709.
[3] S. Andueza, M.A. Vila, M.P. De Pena, J. Sci. Food Agric. 4 (2007) 586.
[4] C.F. Ross, K. Pecka, K. Weller, J. Food Qual. 29 (2006) 596.
[5] A. Farah, M.C. Monteiro, V. Calado, Food Chem. 98 (2006) 373.
[6] D. Selmar, G. Bytof, S.E. Knopp, Plant Biol. 8 (2006) 260.
[7] I. Esteban-Diez, J.M. Gonzalez-Saiz, C. Pizarro, J. Near Infrared Spectrosc. 12 (2004) 287.
[8] M. Banks, C. McFadden, C. Atkinson, The World Encyclopedia of Coffee, Anness Publishing Limited, London, 1999.
[9] P.D.C.M. Agresti, A.S. Franca, L.S. Oliveira, Food Chem. 106 (2008) 787.
[10] E. Nebesny, G. Budryn, Eur. Food Res. Technol. 224 (2006) 159.
[11] M.M.C. Ferreira, A.M. Antunes, M.S. Melgo, Quim. Nova 22 (1999) 724.
[12] C. Alasalvar, F. Shahidi, K.R. Cadwallader, J. Agric. Food Chem. 51 (2003) 5067.
[13] C. Durante, M. Cocchi, M. Grande, Chemom. Intell. Lab. Syst. 83 (2006) 54.
[14] M.G. Moshonas, P.E. Shaw, J. Agric. Food Chem. 45 (1997) 3968.
[15] A. Rapp, Nahrung-Food 42 (1998) 351.
[16] M. Aznar, R.L. Pez, J. Cacho, V. Ferreira, J. Agric. Food Chem. 51 (2003) 2700.
[17] L.M. Nijssen, C.A. Visscher, H. Maarse, L.C. Willemsens, M.M. Boelens, Volatile Compounds in Food: Qualitative and Quantitative Data, 7th ed., TNO Nutrition and Food Research Institute, Zeist, The Netherlands, 1996.
[18] P. Semmelroch, W. Grosch, J. Agric. Food Chem. 44 (1996) 537.
[19] P. Semmelroch, W. Grosch, Lebensm. Wiss. Technol. 28 (1995) 310.
[20] M. Czerny, F. Mayer, W. Grosch, J. Agric. Food. Chem. 47 (1999) 695.
[21] M. Czerny, W. Grosch, J. Agric. Food Chem. 48 (2000) 868.
[22] W. Grosch, F. Mayer, Abs. Pap. Am. Chem. Soc. 218 (1999) U53.
[23] F. Mayers, M. Czerny, W. Grosch, Eur. Food Res. Technol. 211 (2000) 272.
[24] F. Mayers, W. Grosch, Flav. Frag. J. 16 (2001) 180.
[25] F. Mayers, M. Czerny, W. Grosch, Eur. Food Res. Technol. 209 (1999) 242.
[26] J. Pawliszyn, Solid Phase Microextraction: Theory and Practice, Wiley–VCH, New York, 1999.
[27] C.G. Zambonin, L. Balest, G.E. De Benedetto, F. Palmisano, Talanta 66 (2005) 261.
[28] D. Ryan, R. Shellie, P. Tranchida, J. Chromatogr. A 1054 (2004) 57.
[29] J.S. Ribeiro, R.F. Teófilo, T.J. G. Salva, F. Augusto, M.M.C. Ferreira, J. Food Compos. Anal. (submitted for publication).
[30] B.M. Wise, N.B. Gallagher, R. Bro, J.M. Shaver, W. Windig, R.S. Koch, PLS_Toolbox 3.5, for Use with Matlab™, Eigenvector Research, 2004.
[31] N.P.V. Nielsen, J.M. Carstensen, J. Smedsgaard, J. Chromatogr. A 805 (1998) 17.
[32] A. Savitzky, M.J.E. Golay, Anal. Chem. 36 (1964) 1627.
[33] R.F. Teófilo, J.P.A. Martins, M.M.C. Ferreira, J. Chemometr. 22 (2008) 1.
[34] M. Barker, W. Rayens, J. Chemometr. 17 (2003) 166.
[35] T. Naes, T. Isaksson, T. Fearn, T. Davies, A User-friendly Guide to Multivariate Calibration and Classification, NIR Publications, Chichester, UK, 2002.
[36] L.M.V. Malmquist, R.R. Olsen, A.B. Hansen, O. Andersen, J.H. Christensen, J. Chromatogr. A 1164 (2007) 262.
[37] S. Arctander, Perfume and Flavor Chemicals (Aroma Chemical), Allured Publishing Corporation, Illinois, USA, 1969.
[38] P. Pollien, Y. Krebs, A. Chaintreau, Comparison of brew and instant coffee using a new GC-olfactometric method, in: Proceedings of the 17th International Colloquium of Chemistry of Coffee, Nairobi, July, 20–25, 1997, pp. 191–196.
[39] G. Vernin, in: M.N. Clifford, K.C. Wilson (Eds.), Coffee Botany, Biochemistry and Production of Beans and Beverage, Croom Helm, London, 1985.
[40] S. Fors, ACS Symp. Ser. 215 (1983) 185.
[41] M. Winter, F. Gautschi, I. Flament, M. Stoll, I. M. Goldman (Firmenich & Cie) Use of heterocyclic compounds as flavoring ingredients: Furfuryl ethers, US Patent 3,940,502 (1976).
[42] J. Maga, CRC Crit. Rev. Food Technol. 6 (1975) 241.
[43] M. Winter, F. Gautschi, I. Flament, M. Stoll, I.M. Goldman (Firmenich & Cie) Use of heterocyclic compounds as flavoring ingredients: Thiophenes, US Patent 3,900,581 (1975).
[44] R. Wagner, M. Czerny, J. Bielohradsky, W. Grosch, Eur. Food Res. Technol. 208 (1999) 308.
[45] L. Maeztu, C. Sanz, S. Andueza, M.P. De Penã, J. Bello, C. Cid, J. Agric. Food Chem. 49 (2001) 5437.
[46] R. Buffo, C. Cardelli-Freire, Flav. Frag. J. 19 (2004) 99.
[47] M.A. Gianturco, A.S. Giammarino, P. Friefel, V. Flanagan, Tetrahedron 20 (1964) 2951.
[48] R.C. Runnebaum, Am. J. Enol. Vitic. 58 (2007) 417A.
[49] S. Motoda, J. Ferment. Technol. 57 (1979) 395.
[50] S.K. Dart, H.E. Nursten, in: R.J. Clarke, R. Macrae (Eds.), Coffee Chemistry, vol. 1, Elsevier Applied Science Publishers, London, 1985.
[51] W. Baltes, G. Bochmann, J. Agric. Food Chem. 35 (1987) 340.
[52] J. Maga, CRC Crit. Rev. Food Sci. Nutr. 16 (1982) 1.
[53] P. Gemperline, Practical Guide to Chemometrics, 2nd ed., CRC Press, 2006.
[54] C.A.B. De Maria, R.F.A. Moreira, L.C. Trugo, Quím. Nova 22 (1999) 209.