Talanta 83 (2011) 1352–1358

Contents lists available at ScienceDirect

Talanta

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

Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy

J.S. Ribeiro [a,b], M.M.C. Ferreira [a], T.J.G. Salva [b,*]

[a] Theoretical and Applied Chemometrics Laboratory, Chemistry Institute, University of Campinas, P.O. Box 6154, 13083-970 Campinas, SP, Brazil
[b] Agronomic Institute of Campinas-Coffee Centre, P.O. Box 28, 13012-970 Campinas, SP, Brazil

ARTICLE INFO

Article history:
Received 22 June 2010
Received in revised form 28 October 2010
Accepted 2 November 2010
Available online 10 November 2010

Keywords:
Chemometrics
Near infrared diffuse reflectance
Partial least squares
Arabica coffee
Sensory analysis

ABSTRACT

Mathematical models based on chemometric analyses of the coffee beverage sensory data and NIR spectra of 51 Arabica roasted coffee samples were generated aiming to predict the scores of acidity, bitterness, flavour, cleanliness, body and overall quality of coffee beverage. Partial least squares (PLS) were used to construct the models. The ordered predictor selection (OPS) algorithm was applied to select the wavelengths for the regression model of each sensory attribute in order to take only significant regions into account. The regions of the spectrum defined as important for sensory quality were closely related to the NIR spectra of pure caffeine, trigonelline, 5-caffeoylquinic acid, cellulose, coffee lipids, sucrose and casein. The NIR analyses sustained that the relationship between the sensory characteristics of the beverage and the chemical composition of the roasted grain were as listed below: 1 – the lipids and proteins were closely related to the attribute body; 2 – the caffeine and chlorogenic acids were related to bitterness; 3 – the chlorogenic acids were related to acidity and flavour; 4 – the cleanliness and overall quality were related to caffeine, trigonelline, chlorogenic acid, polysaccharides, sucrose and protein.

1. Introduction

Nowadays, the term "quality" is one of the most widespread keywords used in commercial trade. Nevertheless, product quality can assume different meanings for consumers, producers and regulating organizations. In the case of coffee, quality may result from factors like the production system; the aspect and chemical composition of the green or roasted beans and to the final beverage characteristics.

The quality of coffee as a beverage is strictly related to the chemical constituents of the roasted beans, whose composition depends on the composition of green beans (*i.e.*, un-roasted). Unroasted coffee beans contain a wide range of different chemical compounds, which react and interact amongst themselves at all stages of coffee roasting, resulting in even more diverse final products.

Despite the reliability of the cupping method for the evaluation of beverage quality for commercial purposes, the research for more objective, simpler and faster analytical methods is of scientific interest [1–3], since that could be used easily and reproductively in the routine coffee beverage analyses. The impact of chemical components of coffee, *e.g.*, chlorogenic acids, carbohydrates, proteins, trigonelline and caffeine on the final quality of the beverage has already been established in the literature [1–3].

Near infrared spectroscopic (NIRS) is a widespread methodology for qualitative and quantitative analyses in the chemical [4–6], pharmaceutical [7,8] and food industries [9–14].

Several studies have also been carried out on the application of NIRS in coffee analyses. NIRS has been applied to discriminate between the species *Coffea arabica* and *Coffea canephora*, either in pure or blended samples [15–17], to quantify caffeine [13], trigonelline, chlorogenic acid [18,19], total sugar [20] and minerals [21] and to define the roasting degree of coffee beans [22]. Moreover, NIR spectroscopy has also been used to evaluate the quality of espresso beverages. To this end, Esteban-Diez et al. [23] constructed PLS models for the prediction of acidity, body, aftertaste and bitterness from spectra of Arabica and Robusta species. Some important wavelengths related to sensory attributes were highlighted in this study [23]. However, an external validation to evaluate the proposed models was not performed.

The goal of this study was to develop mathematical predictive models for an objective and reproducible sensory quality analyses of coffee. To this end, chemometric tools for the exploitation of the near infrared spectra of roasted Arabica coffee were applied. Moreover, using the NIR spectra analyses, it was tried to establish a relationship between the sensory attributes of the beverage and the chemical components of the coffee beans.

* Corresponding author. Tel.: +55 19 3202 1717; fax: +55 19 3202 1721.
E-mail address: tsalva@iac.sp.gov.br (T.J.G. Salva).

0039-9140/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2010.11.001

Table 1
Scale used for the sensory evaluation of Arabica coffee samples.

Score	Acidity	Bitterness	Flavour	Cleanliness	Body	Overall quality
1	Very low desirability or very highly undesirable	Very strong	Very low desirability or very highly undesirable	Rio	Weak	Very bad
1.5	–	–		Rio/Rioysh	–	Very bad/bad
2	Low desirability or less undesirable	Strong/regular	Low desirability or less undesirable	Rioysh	Weak/regular	Bad
2.5	–	–	Low/regular	Hard−	–	Bad/regular
3	Desirable	Regular	Regular	Hard	Regular	Regular
3.5	–	–	Regular/good	Hard+	–	Regular/good
4	Highly desirable	Regular/normal	Good	Softish	Regular/strong	Good
4.5	–	–	Good/excellent	Soft	–	Good/excellent
5	Very highly desirable	Normal	Excellent	Strictly soft	Strong	Excellent

2. Materials and methods

2.1. Reagents

In order to compare and identify wavelengths possibly related to important compounds normally presents in coffee beans, spectra of pure caffeine (minimum 99% purity), trigonelline (99%), sucrose (99.5%), 5-CGA (5-caffeoylquinic acid) (95%), protein (casein 90%) and carbohydrate (cellulose high purity) were performed. These reagents were supplied by Sigma–Aldrich (Munich). The lipid fraction was extracted from crude coffee beans by mechanical pressing [24].

2.2. Coffee samples

Fifty-one green Arabica coffee samples with different beverage characteristics were roasted separately in a gas fired drum roaster (Pinhalense S/A Máquinas Agrícolas, Brazil) to the medium roast point (Agtron # 55 – according to SCAA's roast color classification system). The roasted coffee samples were packed in special plastic films (polystyrene and polyethylene) and aluminium foil, to avoid a loss of aroma and a contamination from external substances. The samples were stored at −5 °C for a maximum of 3 h before being ground in a mortar and analyzed.

2.3. Sensory analysis

All the 51 samples were submitted to a sensory evaluation by five experts using 10 g of roasted and ground coffee in 100 mL of hot water [25]. The cup quality was assessed according to acidity, bitterness, flavour, cleanliness (clean cup), body (mouthfeel) and overall quality. The quality and intensity of each attribute were evaluated simultaneously by using a scale varying from 1 to 5 (Table 1). For each attribute an assignment of one point was considered very low quality (in terms of a more intense perception in the case of an undesirable sensation or less intense perception if it was a desirable sensation). When analyzing acidity, an assignment of one point may refer to an either low, or too high and undesirable acidity. On the other hand, when evaluating bitterness, a score of one point refers to a beverage bitterer than one with five points. In this case, five points would be assigned to a coffee with normal and desirable bitterness.

2.4. NIR spectra acquisition

Diffuse reflectance spectra of roasted and ground coffee and of trigonelline, caffeine, lipids, cellulose, casein, 5-CGA and sucrose, were obtained using a near-infrared NIRSystems 6500 spectrophotometer (Foss NIRSystems, Raamsdonksveer, The Netherlands) equipped with a reflectance detector and sample transport module. Each spectrum was profiled with 256 scans in the 1100–2500 nm range and a resolution of 4 nm. In this work, three different aliquots of the sample were used and the spectrum of each aliquot was recorded.

2.5. Chemometric data treatment

The original spectroscopic profiles were organized into a matrix format $\mathbf{X}$ ($I \times J$), where each replicate was considered as one sample. Data analysis was carried out using Matlab 6.5 software (The MathWorks, Co., Natick, MA, USA) with the PLS_Toolbox computational package (Eigenvector Research, Inc. –PLS_Toolbox version 3.02.) [26].

In the present study, two pre-treatments were applied to the original data matrix: Savitzky–Golay smoothing with a window size of 5 points and first derivative [27]. The algorithm ordered predictor selection (OPS) was used for variable selection [28], according to the following sequence:

- Step 1 – Selection of an informative vector that contains information about the location of the best independent variables for prediction;
- Step 2 – Differentiation of the original variables according to the corresponding values of the informative vector selected in step 1;
- Step 3 – Sorting variables in decreasing order of magnitude;
- Step 4 – Building and evaluating the multivariate regression models through a cross validation strategy.

The partial least square regression (PLSR) was the method used for modelling. More information on the regression method can be found in Ferreira et al. [29] and Ribeiro et al. [30].

3. Results and discussion

The original spectra of the coffee samples were organized into a format × matrix (153 × 700). The original (A) and pre-treated (B) spectra ($\mathbf{X}_p$) are depicted in Fig. 1.

The average values of sensory scores assigned by the experts for each attribute were used as the dependent variables ($\mathbf{y}$) and the pre-treated spectra (matrix $\mathbf{X}_p$) of the coffee samples were used as the independent variables to develop the regression model.

The data set was split as follows: 41 samples (123 spectra) were randomly selected to be the calibration set, and the remaining 10 samples, corresponding to 30 spectra, were used for external validation. Leave five out cross-validation was performed to select the number of components in the models. In this case, the three replicates of five samples were left out at a time.

From the initial 700 variables (NIR spectrum) the OPS algorithm selected: 76 (18 regions) to build the acidity model (A), 116 (23 regions) for bitterness (B), 118 (29 regions) for flavour (C), 99 (20 regions) for cleanliness (D), 143 (29 regions) for body (E) and 85 (22 regions) for overall quality (F).

Table 2
Regions selected by OPS for construction of the regression models.

Region selected	General region ranges (nm)	Wavelength ranges selected for each model (nm)	Vibrational modes	Charts
1	1208–1236[a,b,d,e,f]	1208–1222,[a] 1214–1220,[b] 1218–1232,[d] 1218–1236,[e] 1218–1236[f]	2nd *overtone* of C–H	CH CH_2
2	1230–1238[b]	–	2nd *overtone* of C–H	CH
3	1340–1344[a]	–	1st *overtone* of C–H combination bands	CH_3
4	1352–1358[d]	–	1st *overtone* of C–H combination bands	CH_3
5	1362–1366[b]	–	1st *overtone* of C–H combination bands	CH_3
6	1396–1398[f]	–	1st *overtone* of O–H	ArOH CH_2 CH3
7	1412–1444[b,c,d,e,f]	1416–1418,[b] 1412–1418,[c] 1426–1438,[c] 1412–1444,[d] 1420–1422,[e] 1432–1444,[e] 1412–1418,[f] 1430–1412,[f] 1438–1442[f]	1st *overtone* of O–H and N–H	ArOH CH CH_2 H_2O ROH $CONH_2$
8	1472–1476[d,e,f]	1472–1476,[d] 1474–1476,[e] 1472–1476[f]	1st *overtone* of N–H	ROH $CONH_2$ CONHR
9	1506–1508[a]	–	1st *overtone* of N–H	RNH_2
10	1520–1528[e]	–	1st *overtone* of N–H	RNH_2
11	1528–1530[c]	–	1st *overtone* of N–H	RNH_2
12	1532–1536[a]	–	1st *overtone* of N–H	RNH_2
13	1552–1556[a,b]	1554–1556,[a] 1552–1556[b]	1st *overtone* region	–
14	1590–1594[a,e]	1590–1592,[a] 1592–1594[e]	1st *overtone* region	–
15	1614–1618[b]	–	1st *overtone* of C–H	ArCH
16	1634–1646[b,c]	1640–1644,[b] 1634–1646[c]	1st *overtone* of C–H	ArCH CH_3
17	1658–1662[b]	–	1st *overtone* of C–H	CH_3
18	1678–1686[c,d,e,f]	1682–1684,[c] 1678–1686,[d] 1670–1686,[e] 1680–1686[f]	1st *overtone* of C–H	CH_2 CH_3
19	1700–1704[c]	–	1st *overtone* of C–H	CH CH_2 CH_3
20	1704–1720[c,e,f]	1710–1720,[c] 1704–1708,[e] 1706–1708[f]	1st *overtone* of C–H	CH CH_2 CH_3
21	1728–1732[c,e]	1728–1732,[c] 1728–1730[e]	1st *overtone* of C–H	CH CH_2
22	1738–1740[c]	–	1st *overtone* of C–H and S–H	CH CH_2 SH
23	1780–1790[a]	–	1st *overtone* of C–H	CH
24	1908–1912[b]	–	2nd *overtone* of C=O stretching	H_2O POH RCO_2H
25	1934–1636[b]	–	2nd *overtone* of C=O stretching	H_2O POH $CONH_2$ RCO_2R'
26	1938–1940[d]	–	2nd *overtone* of C=O stretching	H_2O RCO_2R' $CONH_2$
27	1944–1982[c,d,e]	1956–1958,[c] 1968–1970,[c] 1978–1982,[c] 1948–1950,[d] 1944–1982[e]	2nd *overtone* of C=O stretching	H_2O RCO_2R' $CONH_2$
28	1988–1992[a]	–	1st *overtone* of C=O and O–H combination bands	–
29	1998–2002[a]	–	1st *overtone* of C=O and O–H combination bands	–
30	2020–2022[a]	–	1st *overtone* of C=O and O–H combination bands	–
31	2028–2034[c]	–	1st *overtone* of C=O and O–H combination bands	–
32	2040–2082[a,c,d,e,f]	2066–2074,[a] 2040–2046,[c] 2054–2056,[c] 2040–2056,[d] 2062–2064,[d] 2040–2056,[e] 2042–2082[f]	1st *overtone* of C=O and O–H combination bands	CO
33	2088–2118[b,e]	2088–2118,[b] 2092–2094[e]	1st *overtone* of C=O and O–H combination bands	ROH $CONH_2$(R) CO
34	2126–2132[c,f]	2126–2132,[c] 2128–2132[f]	N–H combination bands	$CONH_2$(R)
35	2138–2140[e]	–	N–H combination bands	$CONH_2$(R)
36	2150–2154[c,e]	2150–2152,[c] 2150–2154[e]	N–H combination bands	RNH_2
37	2180–2182[c]	–	N–H combination bands	RNH_2 CC
38	2190–2192[b,d,f]	–	1st *overtone* of O–H	RNH_2 CC CHO
39	2196–2198[e]	–	N–H and O–H combination bands	ArOH CH_2 CH_3
40	2200–2238[a,b,c,d,e,f]	2226–2228,[a] 2200–2214,[b] 2222–2238,[b] 2210–2234,[c] 2202–2216,[d] 2224–2232,[d] 2206–2216,[e] 2202–2216,[f] 2224–2232[f]	N–H and O–H combination bands	RNH_2 CC CHO
41	2238–2244[e]	–	N–H and O–H combination bands	RNH_2 CHO CH_3
42	2246–2270[a,b,c,d,e,f]	2248–2254,[a] 2248–2272,[b] 2248–2250,[c] 2258–2266,[c] 2246–2270,[d] 2250–2258,[e] 2266–2268,[e] 2248–2254,[f] 2260–2264[f]	N–H and O–H combination bands	H_2O CH_3
43	2274–2298[a,b,c,d,e,f]	2276–2284,[a] 2290–2296,[a] 2280–2286,[b] 2292–2298,[b] 2274–2276,[c] 2288–2294,[c] 2290–2296,[d] 2278–2294,[e] 2280–2282,[f] 2290–2296[f]	C–H + C–H combination bands	H_2O CH_2 CH_3
44	2300–2316[b,d,e,f]	2310–2312,[b] 2306–2312,[d] 2300–2312,[e] 2310–2312[f]	C–H + CC combination bands	CH CH_2 CH_3
45	2324–2334[b,c,d,e,f]	2324–2334,[b] 2324–2334,[c] 2326–2334,[d] 2324–2326,[e] 2326–2332[f]	C–H + C–H combination bands	CH CH_2 CH_3
46	2342–2352[e]	–	C–H + CC combination bands	CH CH_2 CH_3
47	2358–2394[a,e]	2362–2394,[a] 2358–2368,[b] 2374,[e] 2388[e]	C–H + C–H combination bands	CH CH_2 CH_3
48	2400–2402[c,e]	2400–2402,[c] 2400[e]	C–H + C–H combination bands	CH CH_2 CH_3
49	2410–2416[c]	–	C–H + C–H combination bands	CH CH_2
50	2424–2428[b,c,e]	2424–2428,[b] 2426–2428,[c] 2424–2426[e]	C–H + C–H combination bands	CH CH_2
51	2440–2456[a,b,c,d,e,f]	2448–2452,[a] 2440–2452,[b] 2452–2456,[c] 2442–2444,[d] 2440–2454,[e] 2442–2444[f]	C–H + C–H combination bands	CH
52	2464–2490[a,c,d,e,f]	2474–2478,[a] 2460–2490,[c] 2480–2488,[d] 2464–2476,[e] 2480–2488[f]	C–H + CC combination bands	–

[a] Acidity.
[b] Bitterness.
[c] Flavour.
[d] Cleanliness.
[e] Body.
[f] Overall quality.

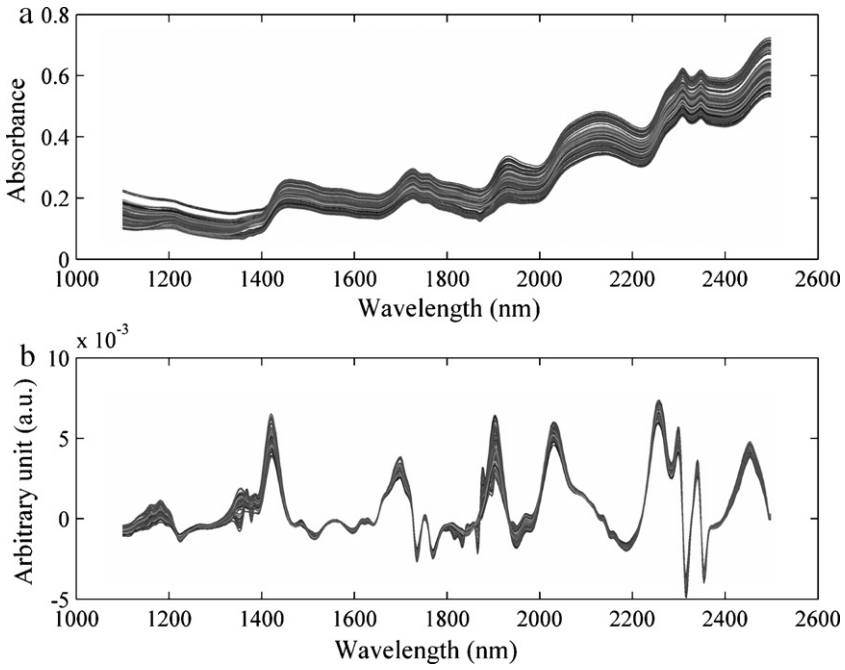

Fig. 1. Original (A) and pre-treated (B) diffuse reflectance spectra of the roasted coffee samples.

Table 2 shows the regions defined by the variables selected for the six calibration models as well as the corresponding vibrational modes and charts. Taking into account the first row of this table, in the region named one, the variables ranging from 1208 to 1222 were selected for the calibration model of the beverage acidity (A) and they are related to the 2nd overtone of the CH or CH_2 charts.

The number of latent variables used in the PLS models was determined from the root mean square error of cross validation (RMSECV) values. The number of latent variables selected for each sensory attribute and the respective statistical parameters RMSECV and r_{cv} is shown in Table 3.

Using the number of latent variables (Table 3) for all the calibration models, in general it was possible to describe 99% and 86%

Table 3
Latent variable numbers, RMSECV and r_{cv} for the PLS models.

Model	N° LV[a]	RMSECV[b]	r_{cv}[c]
Acidity	8	0.28 ± 0.02	0.84 ± 0.01
Bitterness	8	0.35 ± 0.01	0.87 ± 0.01
Flavour	7	0.31 ± 0.01	0.93 ± 0.00
Cleanliness	8	0.38 ± 0.01	0.91 ± 0.01
Body	9	0.27 ± 0.01	0.88 ± 0.01
Overall quality	8	0.39 ± 0.01	0.91 ± 0.00

[a] Number of latent variables.
[b] Root mean square error of cross validation.
[c] Cross validation correlation coefficient.

Table 4
Measured values by the experts and predicted values from the regression models.

	Acidity		Bitterness		Flavour	
Sample	Measured	Predicted	Measured	Predicted	Measured	Predicted
1	3.00 ± 0.06	2.56 ± 0.08	4.13 ± 0.63	3.65 ± 0.22	2.75 ± 0.29	2.70 ± 0.09
2	3.70 ± 0.02	3.35 ± 0.05	4.38 ± 0.75	4.40 ± 0.06	4.19 ± 0.52	4.16 ± 0.06
3	3.69 ± 0.06	3.54 ± 0.11	4.13 ± 0.63	4.12 ± 0.11	3.63 ± 0.48	3.39 ± 0.05
4	3.38 ± 0.03	3.10 ± 0.02	4.13 ± 0.63	4.15 ± 0.08	3.75 ± 0.29	3.84 ± 0.12
5	2.56 ± 0.05	3.01 ± 0.05	2.69 ± 0.63	3.36 ± 0.08	2.25 ± 0.00	2.78 ± 0.01
6	2.19 ± 0.03	2.38 ± 0.05	3.25 ± 0.50	3.31 ± 0.03	2.63 ± 0.48	2.75 ± 0.07
7	2.91 ± 0.06	2.85 ± 0.02	3.50 ± 0.50	3.46 ± 0.20	4.50 ± 0.48	4.24 ± 0.10
8	3.00 ± 0.04	2.53 ± 0.05	3.50 ± 0.63	3.42 ± 0.15	4.00 ± 0.21	4.08 ± 0.02
9	3.7 ± 0.02	3.68 ± 0,04	4.00 ± 0.75	3.71 ± 0.10	4.50 ± 0.52	4.17 ± 0.10
10	3.00 ± 0.04	2.99 ± 0.08	4.00 ± 0.63	3.28 ± 0.18	4.50 ± 0.65	4.34 ± 0.13
	Cleanliness		Body		Overall quality	
Sample	Measured	Predicted	Measured	Predicted	Measured	Predicted
1	3.83 ± 0.29	3.51 ± 0.07	3.33 ± 0.58	3.15 ± 0.19	3.38 ± 0.48	2.93 ± 0.08
2	4.31 ± 0.38	3.82 ± 0.07	3.44 ± 0.38	3.29 ± 0.15	4.38 ± 0.25	3.96 ± 0.04
3	3.63 ± 0.48	3.52 ± 0.07	3.66 ± 0.58	3.33 ± 0.16	4.13 ± 0.25	3.99 ± 0.05
4	3.88 ± 0.14	3.47 ± 0.05	2.69 ± 0.31	2.71 ± 0.10	4.00 ± 0.00	3.54 ± 0.12
5	1.50 ± 0.20	1.83 ± 0.06	1.94 ± 0.24	2.52 ± 0.01	1.50 ± 0.20	2.31 ± 0.09
6	2.38 ± 0.43	2.54 ± 0.02	2.38 ± 0.43	2.50 ± 0.02	2.19 ± 0.13	2.45 ± 0.04
7	3.75 ± 0.29	3.88 ± 0.21	3.50 ± 0.48	3.50 ± 0.33	3.50 ± 0.41	3.94 ± 0.04
8	4.00 ± 0.31	3.77 ± 0.05	3.50 ± 0.25	3.35 ± 0.06	4.00 ± 0.25	3.80 ± 0.07
9	4.50 ± 0.24	4.25 ± 0.12	4.00 ± 0.50	3.61 ± 0.07	4.50 ± 0.25	4.17 ± 0.08
10	4.50 ± 0.41	4.35 ± 0.07	4.00 ± 0.41	3.82 ± 0.18	4.50 ± 0.50	4.35 ± 0.18

of the variance used in blocks **Y** and **X**, respectively. The models were validated by the external data set (10 samples). Fig. 2 shows the experimental values for each sensory attribute vs. the respective values estimated from cross validation. The predicted values for the external validation samples were also included in this figure to show that they are in the same range as the other samples. The values supplied by the experts compared to the predicted by the PLS models for the 10 samples used in the external validation step are shown in Table 4. The RMSEP (root mean square error of prediction) values were 0.30 acidity, 0.37 for bitterness, 0.25 for flavour, 0.37 for cleanliness, 0.30 for body and 0.42 for overall quality.

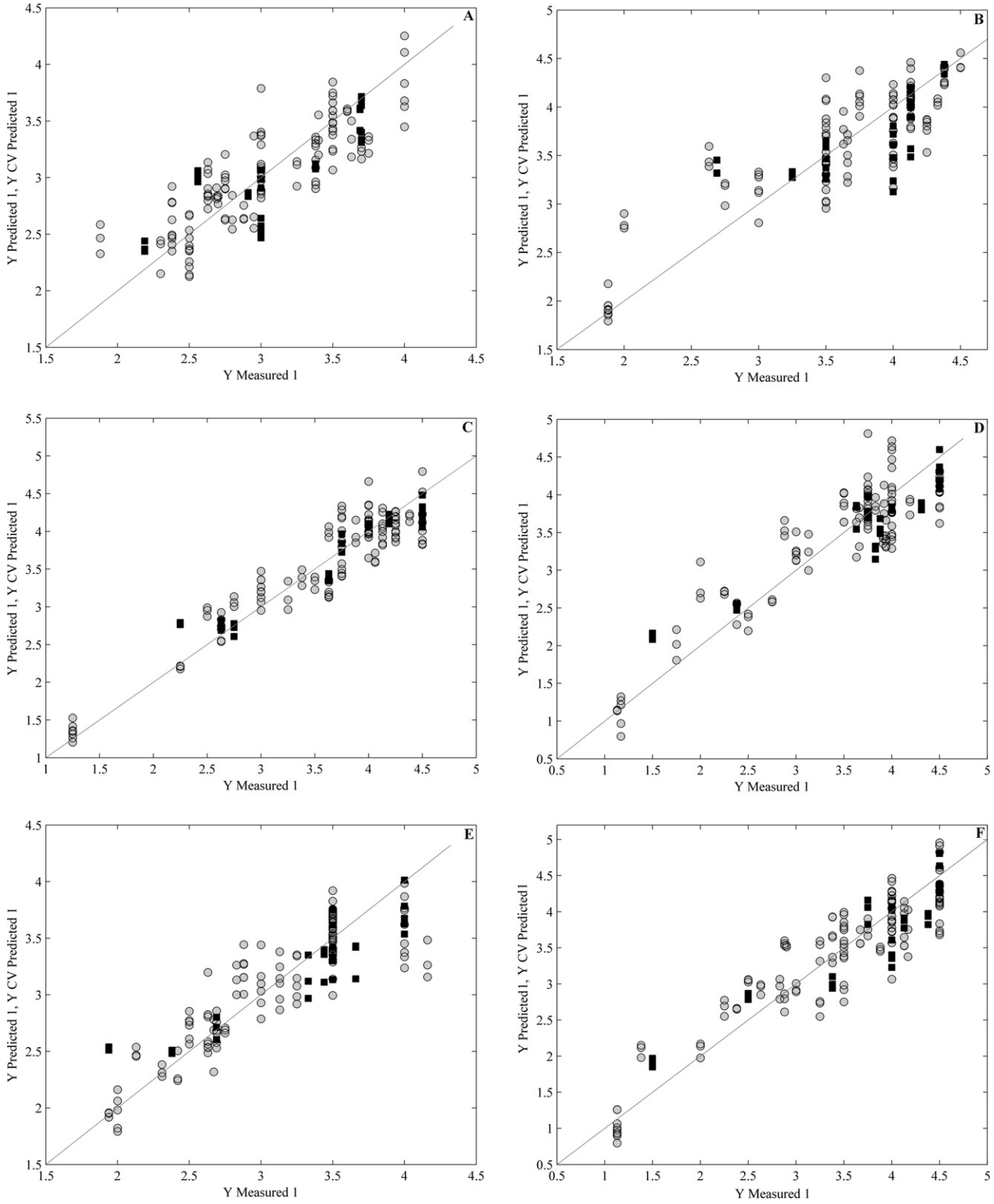

Fig. 2. Plots of measured vs. predicted values for the calibration (○) and prediction (■) sets. Acidity (A), bitterness (B), flavour (C), cleanliness (D), body (E) and overall quality (F).

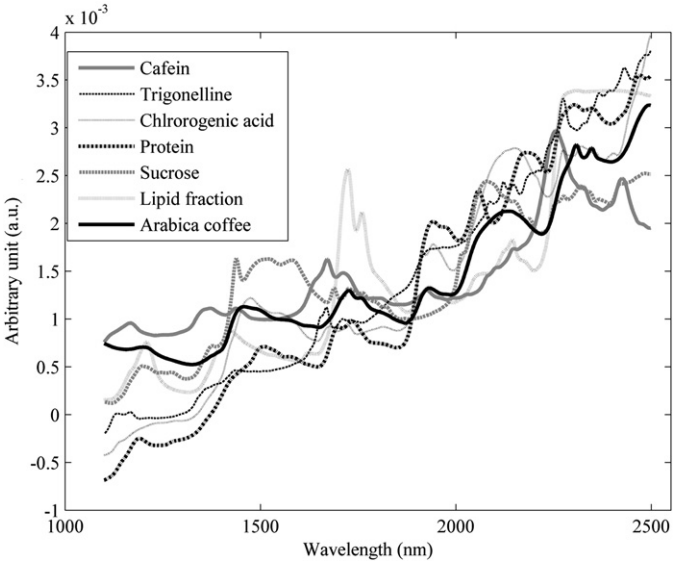

Fig. 3. Spectra of pure compounds present in coffee beans.

A total of 135 wavelength ranges used to construct the six PLS prediction models for the sensory attributes studied are listed at Table 2. Some of these ranges either were selected for more than one sensory attribute or was inside the wavelength range important of other attribute. Thus, these 135 wavelength ranges were grouped in 52 regions (second column of Table 2). Moreover, these regions listed in Table 2 were compared to the pure spectra of the most relevant compounds present in coffee (Fig. 3).

Using the information of the NIR spectra indicated in Fig. 3, the wavelength ranges, in which occurred the greatest absorbance of pure compounds, were shown in Fig. 4.

Some researchers have indicated that the acidity of coffee is due to the phosphoric acid, chlorogenic acids, quinic acid and aliphatic organic acids present in roasted coffee [31,32]. Looking at the 18 regions selected by OPS for the attribute acidity, it can be seen that 8 of them are related to aliphatic and chlorogenic acid structures (Table 2). Thus regions **28**, **29**, **30** and **32** (related to the absorbance of the 1st overtone of C=O and O–H combination bands) and region **13** (1st overtone of C–H) could have originated from organic acids. Since pure aliphatic acids were not analyzed by NIR in this study, only region **32** showed the relationship between acidity and chlorogenic acid, as shown in Fig. 4. The relationship between beverage acidity and chlorogenic acids is sustained by regions **42**, **43** and **47**, where the values for absorbance are relative to the O–H and CH + CH combination bands, according to the analyses of pure compounds (Fig. 4).

According to Esteban-Díez et al. [23], the influence of chlorogenic acids could be justified by taking into account their possible decomposition during roasting, and the influence of the resulting decomposition products on the relative amounts of many groups of compounds, with sensory implications on acidity. One of the wavelengths used for the acidity model by these authors was also selected by OPS (1212 nm). Coincident wavelengths were also identified in the models built for bitterness (2096 nm) and body (1968, 1972, 1978 and 2142 nm).

Beverage bitterness is highly related to the roasting degree and arises from caffeine, some heterocyclic and peptide compounds, chlorogenic acids and sugar degradation products [31,32].

Fifteen of the 23 regions selected by the OPS algorithm for the bitterness model (Fig. 4), could be related to chlorogenic acids and caffeine, as follows: region **7** (1st overtone of O–H and N–H), regions **13**, **15**–**17** (1st overtone of C–H), **24**, **25** (1st overtone of C=O stretch), **33** (1st overtone of C=O and O–H combination bands), **38** (1st overtone of O–H), **42** (N–H and O–H combination bands), **43**–**45**, **50** and **51** (C–H + C–H combination bands). Thus, the variable selection performed by the OPS algorithm also denoted the importance of caffeine and the chlorogenic acids in the composition of beverage bitterness.

Of the 29 regions selected by the OPS algorithm for the model of the attribute body (Table 2), fifteen (**7**, **20**, **21**, **32**, **35**, **36**, **39**, **44**–**48** and **50**–**52**) could be attributed to lipid absorbance, as shown in Fig. 4. However, 10 were coincided with chlorogenic acids and 8 with the absorbance of protein molecules. Thus, the evidence of the relationship between the lipid content and beverage body was shown. However, these data clearly indicated that the beverage body could also be related to the protein and chlorogenic acid content. According to Illy and Viani [33] the body of espresso coffee is closely related to emulsified lipids and proteins, justifying the contribution of some bands assigned to these classes of compounds. Esteban-Díez et al. [23] suggested that the protein content was related to the viscosity, which, in turn, could be related to the body of the coffee.

Flavour is the olphatic perception caused by free gases from the roasted and warm coffee, after preparation of the infusion. Cleanliness is related to the "transparence" of the cup. In a "dirty" cup, all the sensory attributes, except the chemical taste, are masked. In the overall quality evaluation, flavour, taste, bitterness, balance, body and harmony are simultaneously considered [34].

Thus the overall quality is highly dependent on all the sensory attributes that were studied here, especially flavour, cleanliness (high correlation) and body (slightly less correlation). Therefore,

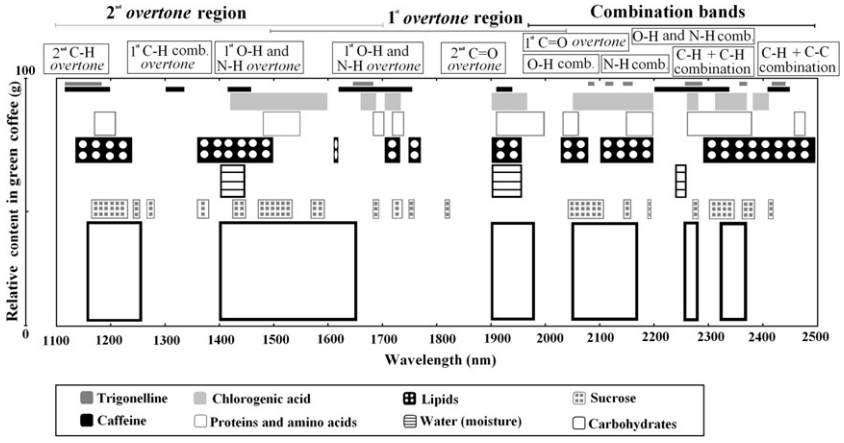

Fig. 4. Schematic representation of the absorption regions of the main components found in coffee.

it is expected that coffee with good flavour and high cleanliness will also have high overall quality.

A total of nine spectral regions selected by the OPS algorithm (**7**, **18**, **32**, **40**, **42**, **43**, **45**, **51** and **52**) were shown to be important for the attributes of flavour, cleanliness and overall quality. Another six were important for overall quality and flavour or for cleanliness (**1**, **8**, **20**, **34**, **38** and **44**).

The regions **7** (1412–1444 nm), **32** (2040–2082 nm), **42** (2246–2270 nm), **43** (2274–2298 nm) and **45** (2324–2334 nm), for example, were important for almost all the compounds, as indicated in Fig. 4. Several other regions were more specific for some classes of compounds: regions **1** (1218–1242 nm) and **44** (2306–2312 nm) for sucrose and other carbohydrates; region **8** (1472–1478 nm) for chlorogenic acids and phenols; regions **34** (2128–2132 nm) and **38** (2190–2192 nm) for proteins and/or chlorogenic acids; and regions **20** (1706–1714 nm), **51** (2436–2475 nm) and **52** (2480–2488 nm) for lipids. Since the overall quality is a synthesis of the attributes involved in the coffee beverage evaluation, mainly flavour and cleanliness, it is expected that the spectral regions of all the compounds would be relevant to the construction of the three regression models.

4. Conclusions

The present study indicates that it is possible to estimate the quality of coffee using PLS regression models obtained by using NIR spectra of roasted Arabica coffees. Parameters predicted were sensory scores for flavour, acidity, bitterness, body, cleanliness and overall quality. Variable selection by using OPS algorithm was a primary step to determine the best spectral regions describing each sensory attribute studied.

The values for RMSECV, r_{cv} and RMSEP computed by the models established using 7–9 latent variables, were 0.28 ± 0.02, 0.84 ± 0.01 and 0.30, 0.35 ± 0.01, 0.87 ± 0.01 and 0.37, 0.31 ± 0.01, 0.93 ± 0.00 and 0.25, 0.38 ± 0.01, 0.91 ± 0.01 and 0.37, 0.27 ± 0.01, 0.88 ± 0.01 and 0.30, 0.39 ± 0.01, 0.91 ± 0.00 and 0.37 for acidity, bitterness, flavour, cleanliness, body and overall quality, respectively.

The considerable stability of the models allowed for the establishment of a correlation between the spectral regions selected and the spectra of the pure compounds. Thus the lipids and proteins in the roasted bean were closely related to the attribute of body in the coffee beverage, caffeine and chlorogenic acids to bitterness and chlorogenic acid to acidity, and the flavour, cleanliness and overall quality were related to the caffeine, trigonelline, chlorogenic acid, polysaccharides, sucrose and protein present in the roasted coffee beans.

Acknowledgements

This work was supported by grants from CAPES and FAPESP.

References

[1] A.S. Franca, J.C.F. Mendonça, S.D. Oliveira, LWT 38 (2005) 709.
[2] T.S. Alves, R.C.E. Dias, M.T. Benassi, Quím. Nova 29 (2006) 1164.
[3] A. Farah, M.C. Monteiro, V. Calado, Food Chem. 98 (2006) 373.
[4] W. Li, D. Huang, L. Liu, J. Mater. Sci. 41 (2006) 7183.
[5] D. Wang, F.E. Dowell, R.E. Lacey, Cereal Chem. 76 (1999) 30.
[6] G. Siebielec, G.W. McCarty, T.I. Stuczynski, J.B. Reeves, J. Environ. Qual. 33 (2004) 2056.
[7] G. Reich, Adv. Drug Deliv. Rev. 57 (2005) 1109.
[8] Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, N. Jent, J. Pharm. Med. Anal. 44 (2007) 683.
[9] Q. Chen, J. Zhao, H. Zhang, X. Wang, Anal. Chim. Acta 572 (2006) 77.
[10] D. Cozzolino, H.E. Smyth, K.A. Lattey, W. Cynkar, L. Janik, R.G. Dambergs, I.L. Francis, M. Gishen, Anal. Chim. Acta 563 (2006) 319.
[11] R. Karoui, L. Pillonel, E. Schaller, J.O. Bosset, J. De Baerdemaeker, Food Chem. 101 (2006) 1121.
[12] H. Cen, Y. He, Trends Food Sci. Technol. 18 (2007) 72.
[13] C. Pizarro, I. Esteban-Díez, J.M. González-Sáiz, M. Forina, J. Agric. Food Chem. 55 (2007) 7477.
[14] A.M.K. Pedro, M.M.C. Ferreira, Appl. Spectrosc. 63 (2009) 1308.
[15] E.K. Kemsley, S. Ruault, R.H. Wilson, Anal. Nutr. Clin. Meth. Sec. 54 (1995) 321.
[16] C. Pizarro, I. Esteban-Díez, J.M. González-Sáiz, Anal. Chim. Acta 585 (2007) 266.
[17] I. Esteban-Díez, J.M. González-Sáiz, C. Sáenz-González, C. Pizarro, Talanta 71 (2007) 221.
[18] M.A. Morgano, C. Camargo, A.P. Pagel, M.F. Ferrão, M.M.C. Ferreira, II Simpósio de Pesquisa dos Cafés do Brasil, 2001 (Vitória, ES, Brazil, p. 1502).
[19] F. Davrieux, J.C. Manez, Durand, B. Guyot, Near infrared spectroscopy, in: Proceedings of the 11th International Conference, 2004, p. 441.
[20] M.A. Morgano, C.G. Faria, M.F. Ferrão, M.M.C. Ferreira, Quím. Nova 30 (2007) 346.
[21] M.A. Morgano, L.F. Pauluci, B.M.V. Mantovani, E.E.M. Mory, Ciênc. Tecnol. Aliment. 22 (2002) 19.
[22] L. Alessandrini, S. Romani, G. Pinnavaia, M.D. Rosa, Anal. Chim. Acta 95 (2008) 95.
[23] I. Esteban-Díez, J.M. González-Sáiz, C. Pizarro, Anal. Chim. Acta 525 (2004) 171.
[24] E. Uquiche, M. Jerez, J. Ortiz, Innov. Food Sci. Emer. Technol. 9 (2008) 495.
[25] http://www.agricultura.gov.br (accessed in 2009).
[26] 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.
[27] A. Savitzky, M.J.E. Golay, Anal. Chem. 36 (1964) 1627.
[28] R.F. Teófilo, J.P.A. Martins, M.M.C. Ferreira, J. Chemom. 23 (2009) 32.
[29] M.M.C. Ferreira, A.M. Antunes, M.S. Melgo, Quím. Nova 22 (1999) 724.
[30] J.S. Ribeiro, F. Augusto, T.J.G. Salva, R.A. Thomaziello, M.M.C. Ferreira, Anal. Chim. Acta 634 (2009) 172.
[31] M.N. Clifford, in: M.N. Clifford, K. Willson (Eds.), Coffee: Botany, Biochemistry and Production and Beans and Beverage, Croom Helm, London, 1985.
[32] R.J. Clarke, in: I.D. Morton, A. Macleod (Eds.), Food Flavour: Part B. The Flavours of Beverages, Elsevier Science Publ., Amsterdam, 1986.
[33] A. Illy, R. Viani, Espresso Coffee – The Chemistry of Quality, Academic Press, London, 2005.
[34] http://www.scaa.org (accessed in 2010).