Talanta 128 (2014) 393–400

Contents lists available at ScienceDirect

Talanta

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

Application of elastic net and infrared spectroscopy in the discrimination between defective and non-defective roasted coffees

Ana Paula Craig [a,b], Adriana S. Franca [a,c,*], Leandro S. Oliveira [a,c], Joseph Irudayaraj [b], Klein Ileleji [b]

[a] PPGCA, Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627, 31270-901 Belo Horizonte, MG, Brazil
[b] Department of Agricultural and Biological Engineering, Purdue University, 225 S. University Street, West Lafayette, IN 47907, USA
[c] DEMEC, Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627, 31270-901 Belo Horizonte, MG, Brazil

ARTICLE INFO

Article history:
Received 26 February 2014
Received in revised form
30 April 2014
Accepted 2 May 2014
Available online 10 May 2014

Keywords:
Defective coffee
Elastic net
FTIR
NIRS

ABSTRACT

The quality of the coffee beverage is negatively affected by the presence of defective coffee beans and its evaluation still relies on highly subjective sensory panels. To tackle the problem of subjectivity, sophisticated analytical techniques have been developed and have been shown capable of discriminating defective from non-defective coffees after roasting. However, these techniques are not adequate for routine analysis, for they are laborious (sample preparation) and time consuming, and reliable, simpler and faster techniques need to be developed for such purpose. Thus, it was the aim of this study to evaluate the performance of infrared spectroscopic methods, namely FTIR and NIR, for the discrimination of roasted defective and non-defective coffees, employing a novel statistical approach. The classification models based on Elastic Net exhibited high percentage of correct classification, and the discriminant infrared spectra variables extracted provided a good interpretation of the models. The discrimination of defective and non-defective beans was associated with main chemical descriptors of coffee, such as carbohydrates, proteins/amino acids, lipids, caffeine and chlorogenic acids.

1. Introduction

Food quality encompasses sensory properties, nutritive values, mechanical properties, functional properties and presence of defects. The majority of traditional techniques used for quality assurance are costly, labor intensive and time consuming, an example being sensory panel evaluation that, to this date, is the ultimate tool to assess coffee quality. Besides being costly and time consuming, sensory panels are inadequate for employment in routine analysis in food processing facilities. In this scenario, infrared spectroscopy is gaining attention since it has been demonstrated capable of solving some of the problems presented by traditional techniques. In particular, much attention has been given to FTIR and NIR spectroscopy. The FTIR spectrum detects fundamental vibrations in the mid-infrared region (4000–400 cm^{-1}) whereas the NIR spectrum (2500–800 nm) arises from the molecular absorptions of overtones and combinations of fundamental vibrational bands in the mid-infrared region. These techniques are rapid, nondestructive and require minimum sample preparation [1].

The feasibility of using infrared spectroscopy in combination with multivariate statistics has been examined with notable success in coffee quality evaluation, with applications including discrimination and quantification of arabica and robusta blends [2], detection of adulterants [3–5], prediction of sensory properties [6] and discrimination between high and low quality coffees [7–10]. Although successful classification and quantification models have been presented, the interpretation of those models can be challenging. Precise band assignments of complex sample matrices such as coffee are difficult due to the fact that a single band may be attributable to several classes of molecules. Particularly, the broadband nature of NIR spectra, consisting of overlapping combination and overtone bands, makes it harder to clearly attribute bands to specific chemical functionalities and molecules in comparison to mid-infrared, where fundamental peaks can be observed in isolated positions. At the very least, techniques for spectral manipulation are required to resolve

Abbreviations: ATR, Attenuated Total Reflectance; CGA, chlorogenic acids; D, particle diameter; DLATGS, Deuterated Triglycine Sulphate Doped with L-Alanine; DRIFTS, Diffuse Reflectance Fourier Transform Infrared Spectroscopy; FTIR, Fourier Transform Infrared Spectroscopy; LASSO, Least Absolute Shrinkage and Selection Operator; MSC, Multiplicative Scatter Correction; NIR, Near Infrared; NIRS, Near Infrared Spectroscopy.

* Corresponding author at: DEMEC, Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627, 31270-901 Belo Horizonte, MG, Brazil. Tel.: +55 31 34093512; fax: +55 31 34433783.
E-mail addresses: adriana@demec.ufmg.br, drisfranca@gmail.com (A.S. Franca).

http://dx.doi.org/10.1016/j.talanta.2014.05.001
0039-9140/

bands and compensate for background interferences [11], and novel statistical techniques are needed to select discriminant variables and interpret the developed models.

One of the major factors affecting coffee quality is the presence of defective coffee beans. The defects that contribute the most to the depreciation of beverage quality are black beans, associated with a heavy and ashy flavor; sour beans, related to sour and oniony tastes; and immature beans, that impart astringency and bitterness to the beverage [8]. Prior to roasting, physical, chemical and sensory parameters can be used to classify those beans, and, in particular, recent studies have shown that either DRIFTS (Diffuse Reflectance Fourier Transform Infrared Spectroscopy), ATR-FTIR (Attenuated Total Reflectance Fourier Transform Infrared) Spectroscopy [7,9] or NIR [10] can discriminate defective and non-defective crude coffees. On the other hand, sensory analysis still remains the ultimate tool employed to assess the quality of roasted coffees. However, the 'cup test' is not suitable for the classification and inspection of roasted coffees from the market, because the roasting conditions are usually unknown and low-quality (defective) coffees are included in the batch and generally roasted to a darker roasting degree to mask unpleasant flavors and/or aromas. We have shown, in a recent study [8], that DRIFTS provides satisfactory discrimination of non-defective/defective beans even after roasting. Nevertheless, the small amount of sample analyzed in the DRIFTS accessory can be viewed as a drawback if quantification of defects is of interest [7].

In view of the aforementioned, the objective of this work was to further investigate the potential of infrared spectroscopy (FTIR and NIR) for coffee quality evaluation. The major goals were to compare the performance of the referred techniques in the discrimination of high quality (non-defective) and low quality (defective) roasted coffees and to attempt correlation of discriminating bands to chemical compounds. Given the complexity of the infrared spectra of coffees, in this study, we have introduced a sparse learning dimensionally reduction algorithm named Elastic Net [12]. Elastic net is a version of penalized least squares that combines both Ridge and LASSO (Least Absolute Shrinkage and Selection Operator) regressions. Ridge regression, or Tikhonov regularization, shrinks (toward zero) the least square coefficients, while LASSO not only shrinks the coefficients but also provides sparsity and model selection. Sparsity makes the data more succinct and simpler, and can provide good interpretation of a model, revealing an explicit relationship between the objective of the model and the given variables. Unlike LASSO penalty, the Ridge penalty (L_2-penalty), drawn from a Gaussian distribution, is ideal if there are many predictors and all have nonzero coefficients. Therefore, in Elastic net the penalty is a compromise between the Ridge regression penalty ($\alpha=0$) and the LASSO regression penalty ($\alpha=1$) [13,14].

Besides the application of Elastic Net, other novelty aspects of the present work are associated to it being the first study employing NIRS for discrimination of roasted defective coffees, since up to this date application of this technique in association to the analysis of defective coffees has been restricted to crude beans [10]; and also to being the first study that evaluates both NIRS and ATR-FTIR for discrimination of roasted defective beans.

2. Material and methods

2.1. Preparation of samples of defective and non-defective roasted coffees

Arabica green coffee samples were acquired from a roasting company located in Minas Gerais State, Brazil. Samples consisted of coffee beans harvested by strip-picking that were rejected by color sorting machines. The beans were manually sorted (by a professional trained and certified for green coffee classification) into five lots or sample classes: non-defective, immature, black and sour (light and dark colored). Samples of 25 g were taken from each lot and roasted in a convection oven (Model 4201D Nova Ética, SP, Brazil) at 220, 235 and 250 °C. For each temperature, samples were roasted at three roasting times, resulting in nine different roasting conditions for each lot. Samples were ground in a coffee grinder (Arbel, Brazil) and subjected to color evaluation. Color measurements were performed using a tristimulus colorimeter (HunterLab Colorflex 45/0 Spectrophotometer, Hunter Laboratories, VA, USA) with standard illumination D_{65} and colorimetric normal observer angle of 10°. Roasting conditions were established for each specific lot, given that defective coffee beans have been reported to roast to a lesser degree than non-defective coffee beans when submitted to the same processing conditions [8]. Roasting degrees were then defined according to luminosity ($L*$) measurements similar to commercially available coffee samples ($19.0 < L* < 25.0$), corresponding to light ($23.5 < L* < 25.0$), medium ($21.0 < L* < 23.5$) and dark ($19.0 < L* < 21.0$) roasts. The corresponding roasting times ranged from 7 to 10 min (250 °C), 9 to 16 min (235 °C) and 12 to 33 min (220 °C), with the smaller and larger times for a given temperature corresponding to the light and dark roasts, respectively. A total of 45 samples were obtained. Samples were sieved and fractions with $0.25 > D > 0.15$ mm and $0.84 > D > 0.39$ mm were used in the ATR-FTIR and NIR experiments, respectively.

2.2. ATR-FTIR and NIR measurements and spectral collection

A Shimadzu IRAffinity-1 FTIR Spectrophotometer (Shimadzu, Japan) with a DLATGS detector was used in the ATR-FTIR measurements that were performed in dry atmosphere (20 ± 0.5 °C). A horizontal ATR sampling accessory (ATR-8200HA) equipped with ZnSe cell was employed. Approximately 2 g of the ground and roasted coffee samples were placed in the sampling accessory obtaining the best contact with the crystal. The empty accessory was used to obtain the background spectrum. The approximate total time required for spectral collection was 5 min. All spectra were recorded within a range of 3100–800 cm^{-1} with a 4 cm^{-1} resolution. Each spectrum was calculated as the average of 20 scans and subjected to background subtraction.

A SpectraStar 2400 Drawer NIR spectrophotometer (Unity Scientific) with an InGaAs detector was used in the measurements. Approximately 3 g of ground coffee samples were placed inside a glass cup, filling the entire empty space, and covered. Air was used to obtain the background spectra. The approximate time required for the spectral collection was 2 min. All spectra were recorded within a range of 1200–2400 nm with 1 nm resolution. Each spectrum was calculated as the average of 30 scans and subjected to background subtraction. In both FTIR and NIR experiments, each sample was read in triplicate, resulting on a total of 135 spectra for each technique. Intra-day precision was evaluated by analyzing three spectra replicates of a given sample every three hours, for a period of 12 h. Inter-day precision was determined by analyzing spectra triplicates in five consecutive days. Both were evaluated according to average relative standard deviations between spectra. Interday and intraday precisions for FTIR measurements were $4.2 \pm 0.9\%$ and $2.2 \pm 1.3\%$. Interday and intraday precisions for NIR measurements were $0.3 \pm 0.5\%$ and $0.7 \pm 0.6\%$.

2.3. Data analysis

Preprocessing techniques (baseline correction, area normalization, MSC and mean centering) were applied to raw data prior to statistical analysis to compensate for changes in experimental

conditions and enhance results. In addition, for the FTIR spectra, the regions above 2800 and below 800 cm^{-1} were excluded to avoid noise effects. Prior to classification analysis by Elastic net, Principal Component Analysis (PCA) was applied to the datasets to detect outliers. The combination of *Q* and *T*2 tests were used to detect abnormal observations. Given the significance level for the *Q* and *T*2 statistics, in this case, 99%, observations with *Q* and/or *T*2 values over the threshold were classified as outliers. After the elimination of the outliers from the models, the procedure was continually repeated until no outliers were identified. The softwares Matlab (The MathWorks, Co., Natick, MA) and PLS_Toolbox (Eigenvector Research, Inc.) were employed for the preprocessings calculation and PCA analysis.

Elastic net was used to develop classification models and select variables (wavenumbers/wavelengths) that presented explicit relationships with the different sample classes (non-defective, dark and light sour, black and immature). This algorithm was applied using the glmnet package for the R software that fits generalized linear models via penalized maximum likelihood. Samples were separated into training (75%) and validation (25%) data sets [15,16]. Approximately 7 spectra of each of the five sample classes were included in the validation set. The regularization parameter lambda causes coefficient shrinkage, minimizing the residual sum of squares. In order to obtain the lambda value that provides minimum cross-validated error, leave-one-out cross-validation was performed. In sequence, multinomial logistic models were fitted with the training data set at α values ranging from 0 to 1, in steps of 0.25. The developed models were used to predict the class of observations from the calibration and new observations from the validation data sets. The models with α-values that provided the highest predictability based on the lowest classification errors were selected as best fits. The discriminant variables from the best fit models and their respective coefficient estimates were then extracted.

3. Results and discussion

3.1. Observations on spectra

A comparative evaluation of the original spectra of defective and non-defective coffees obtained by FTIR, Fig. 1(a), indicates that they are quite similar. A considerable difference in their baseline was observed, with absorbance values being higher for light sour and non-defective coffees and lower for immature, dark sour and black beans. After the application of baseline correction (Fig. 1(b)) the spectra exhibited high similarity and could not be visually differentiated. Significant bands at 2920, 2850, 1747 and 1400–900 cm^{-1} have been previously identified in arabica and robusta roasted coffees [17] and arabica green coffees [7–9]. The first two regions are associated with symmetric and asymmetric stretching of CH bonds in CH_2 and CH_3 groups, respectively [18]. The region associated with CH_2 groups is highly related to the presence of lipids [3,19], while the CH_3 region has been used for caffeine quantification [20]. The sharp band at 1747 cm^{-1} is assigned to C=O stretch from aliphatic esters and is mostly related to the presence of lipids. The third region, from 1400–900 cm^{-1}, is commonly called *fingerprint* region because of the large amount of characteristic bands from single bonds or specific functional groups [18]. In particular, carbohydrates exhibit large bands in this region [17].

The original average spectra of defective and non-defective coffees obtained by NIR are shown in Fig. 1(c). The shape of the spectra was particularly dominated by broad water absorption bands at 1440 – 1480 nm (1st overtone of O–H stretching) and 1930–1950 nm (combination band of O–H stretching and O–H deformation). This feature was also observed after spectra were subjected to Multiplicative Scatter Correction (MSC), Fig. 1(d). Other few regions of the spectra that could be visually identified have been reported in the literature as characteristic absorption regions of specific compounds. For example, the two well defined bands between 1715–1760 and 2300–2350 nm are attributed to lipids and the region around 2100 nm is associated with carbohydrates and/or chlorogenic acids and proteins [6,21]. The spectra obtained in this study are similar to the ones presented by Santos et al. [10] for defective and non-defective crude coffees. Higher log (1/*R*) intensities were observed in the study by Santos et al. [10] except in the regions previously associated to lipids. Although the lipids loss is minimal during roasting, the loss of other compounds (e.g. water, volatiles) leads to a relative increase in their levels.

Different preprocessing methods, including area normalization, baseline correction, first and second derivatives, SNV and MSC, were applied to the raw data in order to compensate for variations in experimental conditions and to develop accurate calibration models. The preprocessing procedures that improved the discrimination of the samples based on PCA and were then considered for this study were: baseline correction followed by area normalization (FTIR) and MSC (NIR). PCA analysis allowed the detection of abnormal observations, at 99% of confidence level, based on the combination of *Q* and *T*2 tests (see Supplementary material, Figs. S1 and S2). In the first round of outlier removal, 0 and 4 outliers were detected from the original and preprocessed FTIR datasets, respectively. In the second round, no outliers were detected. Thus, the final PCA model and the subsequent statistical analysis performed in this study for the original and preprocessed FTIR data sets were developed with 135 and 131 spectra, respectively. In the first round of outlier removal for the NIR data sets, 4 and 5 outliers were detected and removed from the original and MSC spectral data, respectively. In the second round of outlier removal, 2 and 1 outliers were removed from the original and MSC PCA models, respectively. The optimized models did not exhibit additional outliers, so the remaining statistical analysis for both the original and MSC NIR spectra were developed with data sets containing 129 spectra. Furthermore, PCA analysis provided an explanation for the data variability.

A visual discrimination of the samples by PCA was not clearly evident. In the FTIR models, a better separation of the samples was observed in the scatter plots of samples versus PC1 (original data) and PC3 (preprocessed data). Non-defective and light sour tended to exhibit positive scores, while dark sour and black coffees tended to exhibit negative scores (Fig. 2a and b). Immature beans could not be visually discriminated based on the FTIR models. With regarding to the PCA analysis of the NIR data, the PC2 scores scatter plot of the original data (Fig. 2c) revealed that black coffee samples were fully discriminated from the remaining classes. On the PCA analysis of the NIR preprocessed data (Fig. 2d), a small percentage of variance explained by the PC3 separated, although not completely, (a) dark sour and black coffee (positive scores), from (b) non-defective, light sour and immature coffees (negative scores). In general, the results of the unsupervised PCA analysis were not conclusive in terms of sample discrimination and an analysis of the PC loadings would not provide insights on the discriminating spectral bands.

3.2. Classification by elastic net

Table 1 summarizes the results obtained by the multinomial logistic models constructed via Elastic net to classify defective and non-defective coffees, roasted at nine different conditions, based on their FTIR and NIR spectra. As expected, models constructed with preprocessed data exhibited higher accuracy. In general, models constructed with a lower number of nonzero variables

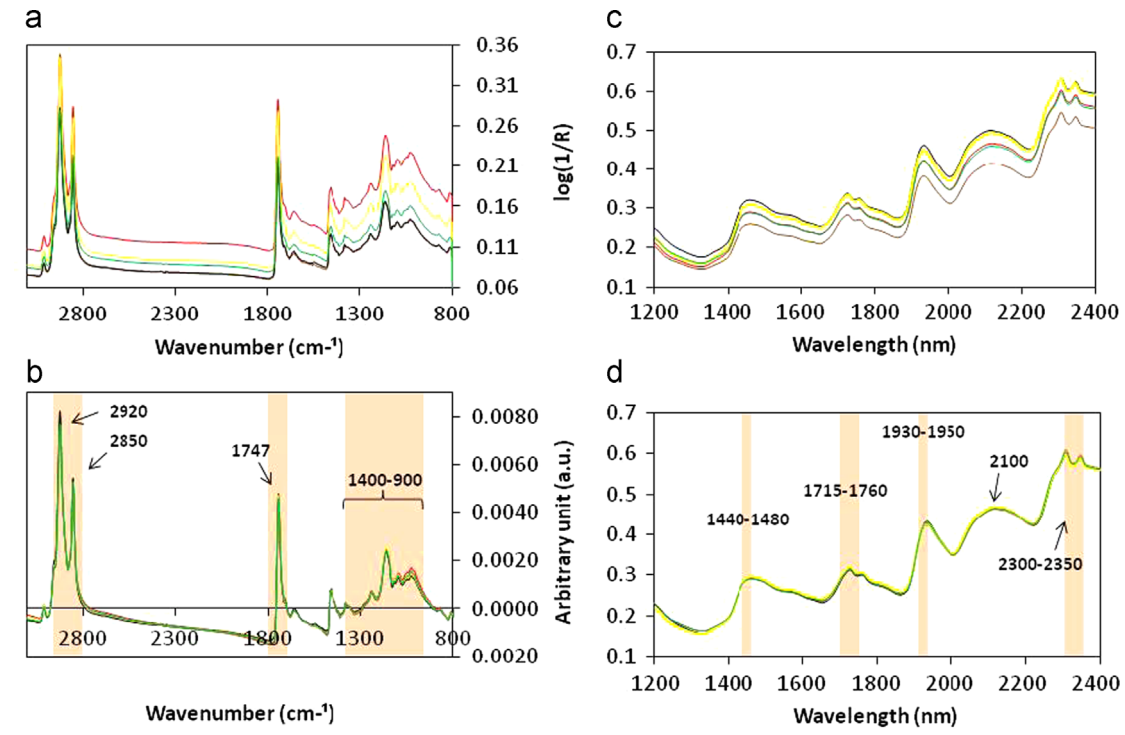

Fig. 1. ATR-FTIR spectra: (a) original; (b) submitted to baseline correction and area normalization; NIR spectra: (c) original; (d) submitted to MSC.—non-defective; — immature; — sour (light); — sour (dark); — black.

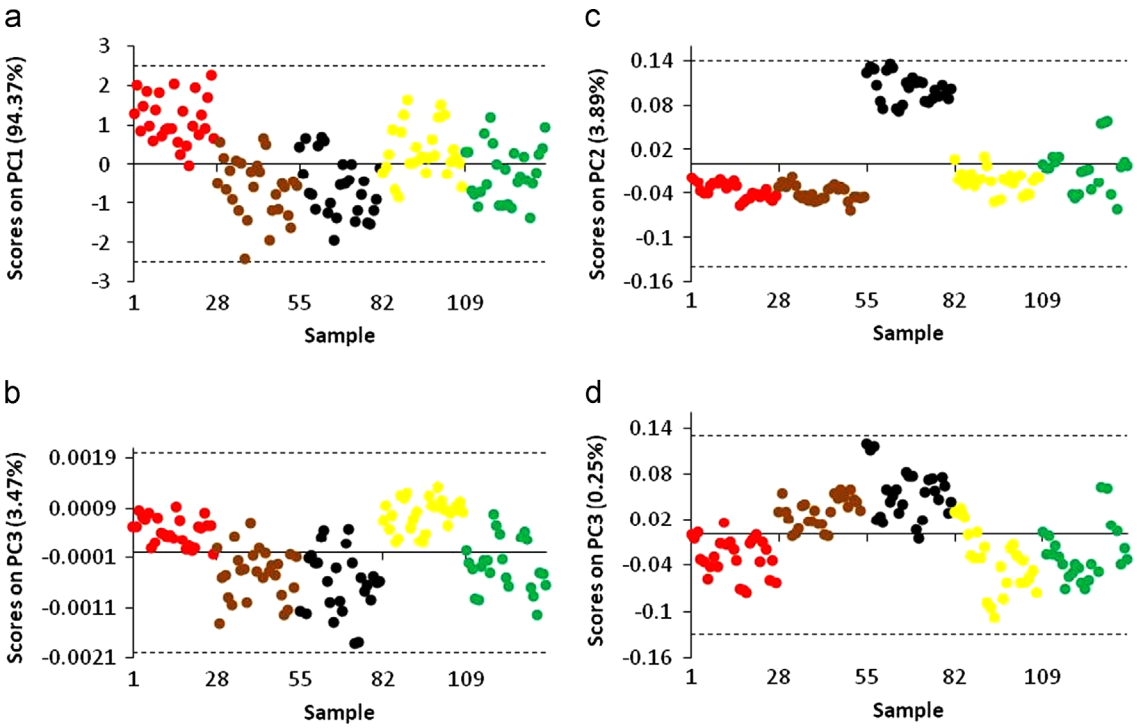

Fig. 2. PCA scores scatter plots of (a) original ATR-FTIR spectra and (b) spectra submitted to baseline correction and area normalization. PCA scores scatter plots of (c) original NIR spectra and (d) MSC spectra. ● non-defective, ● immature, ● black, ● light sour, ● dark sour.

($\alpha \neq 0$) provided better and comparable results. Excellent statistical classification of defective and non-defective coffees was achieved at α levels ranging from 0.25 to 1 for the models based on preprocessed data. These results indicated that accurate classification can be achieved from relatively small regions of the spectra, by means of imposing penalties in the models to reduce the number of explicit variables. The α parameter controls the mixing between Ridge and LASSO regression. Ridge regression ($\alpha=0$) imposes a L_2-penalty to the model inducing coefficient shrinkage, while LASSO regression ($\alpha=1$) imposes a L_1-penalty which expects many predictors to be close to zero and a small subset to be nonzero, providing automatic variable selection [22]. LASSO presents some limitations if there is a group of variables among which the pairwise correlations are very high, because it

Table 1
Percentage of correct classification obtained by Elastic net models based on ATR-FTIR and NIR spectra: comparing treatments and penalties.

Treatment	α	Nonzero variables	Correct Classification		
			Cal	CV	Val
FTIR					
Original spectra	0	676	0.63	0.54	0.71
	0.25	318	0.88	0.88	0.93
	0.5	239	0.89	0.89	1
	0.75	151	0.9	0.9	0.93
	1	39	0.96	0.96	0.99
Baseline correction	0	676	0.91	0.91	0.87
+area normalization	0.25	291	1	1	0.97
	0.5	175	1	1	0.97
	0.75	99	1	1	0.97
	1	35	1	1	0.97
NIR					
Original spectra	0	1200	0.84	0.84	0.7
	0.25	681	0.92	0.91	0.85
	0.5	519	0.95	0.92	0.85
	0.75	307	0.95	0.95	0.88
	1	27	0.95	0.95	0.88
MSC	0	1200	0.89	0.89	0.94
	0.25	446	1	1	0.94
	0.5	271	1	1	0.94
	0.75	174	1	1	0.94
	1	39	1	1	0.88

Cal=calibration; CV=cross-validation; Val=validation.

tends to select only one variable from the group with no regard for which one is selected. This problem can be overcome by using the regularization technique called Elastic net [12], a version of penalized least squares that combines both Ridge and LASSO regression, thus providing both shrinkage and variable selection. Considering the classification results in Table 1 and the previous discussion, models built with $\alpha=0.75$ were considered for the extraction of the coefficients estimates.

3.3. *Variable selection and characterization of coffee samples*

The main reason why Elastic net was used in this study is because it provides a good interpretation of a model, revealing an explicit relationship between the objective of the model and the given variables. Thus, it allows the visualization of discriminating variables and how they contribute to the correct classification of each coffee class. A positive or negative coefficient estimate indicates that a specific coffee class exhibited higher or lower absorption intensity at that range of the spectrum, respectively, and thus positive coefficients may be associated to higher concentrations of a specific compound in a class sample in comparison to other classes. The Elastic net coefficient estimates for models developed with the original and preprocessed spectra obtained by ATR-FTIR and NIR are shown in Fig. 3. The chemical attribution of variables was conducted by examining each of the nonzero coefficient estimates obtained by Elastic net and, based on the literature, conducting a tentative assignment of these coefficients to chemical compounds that may absorb in the selected region of the ATR-FTIR and NIR spectra. The detailed assignments of the discriminating peaks and bands selected by Elastic net are shown in Tables 2 and 3.

Evaluation of the plots presented in Fig. 3a and b shows non-zero coefficients associated with C–O stretching in carbohydrates including sucrose, oligo- and polysaccharides at the following regions of the mid-infrared: 1039, 1099, 1118–1132 and 1138–1165 cm^{-1} [18,23]. In those regions, non-defective/light sour coffees exhibited

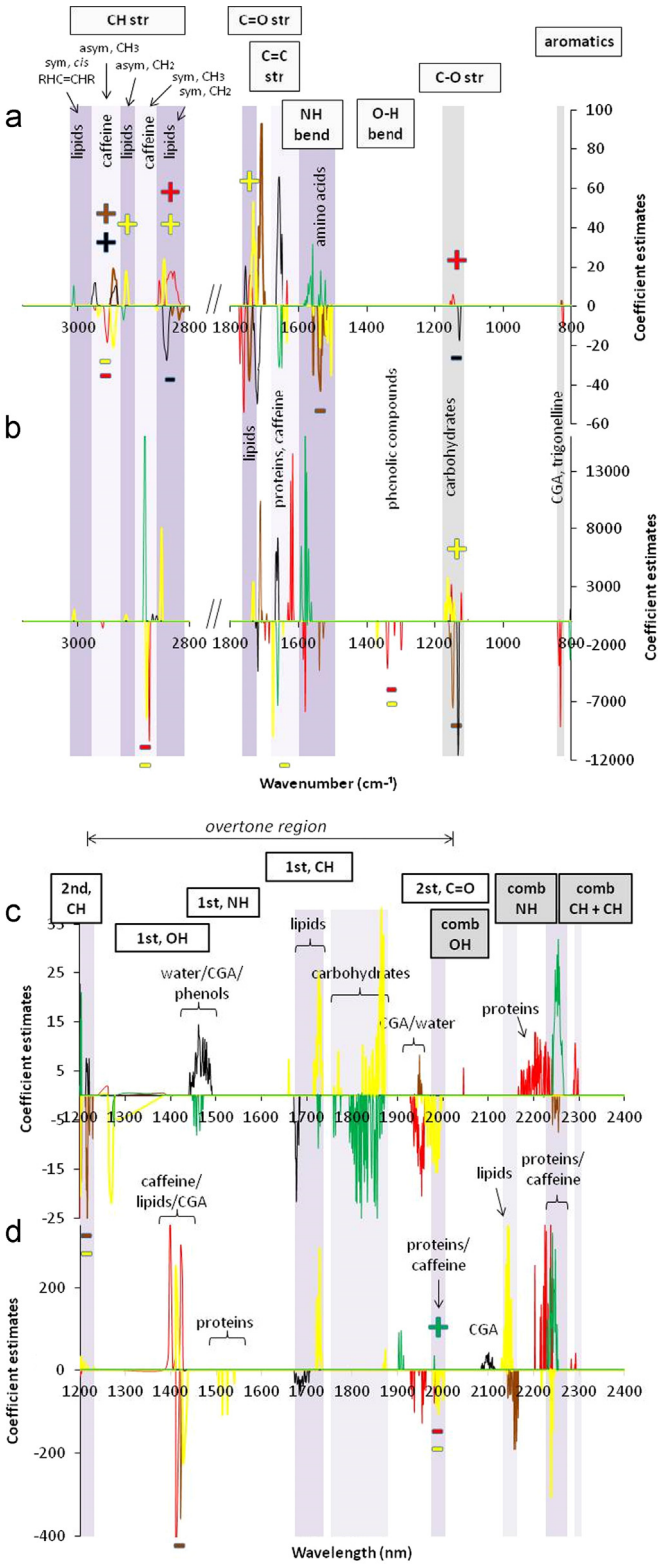

Fig. 3. Elastic net coefficient estimates at $\alpha=0.75$ for ATR-FTIR spectra: (a) original; (b) submitted to baseline correction and area normalization; NIR spectra: (c) original; (d) submitted to MSC. — non-defective, — light sour, — dark sour, — black and — immature. sym=symmetric; asym=asymmetric; str=stretching.

positive coefficients while dark sour/black coffees exhibited negative coefficients. In the NIR spectra (Fig. 2c and d), carbohydrates were mainly associated with region 1760–1871 nm, where the 1st overtone of C–H and a combination band of O–H and C–O stretching take place [11,21,24]. In this region of the spectrum non-defective

Table 2
Tentative chemical assignment of significant ATR-FTIR bands selected by Elastic net (α=0.75) for the classification of defective and non-defective coffees.

Gereral region ranges	ND	LS	DS	BL	IM	Vibration modes	Compounds	References
800–835		–	+	+	–	Out-of-plane CH bend, adjacent CH wag	Phenolic compounds, pyridine	[18]
1039				–		C–O str	Celulose	[23]
1099		+				C–O str	Carbohydrates	[23,33]
1118–1132		+		–		C–O str	Carbohydrates	[33,17]
1138–1165	+	+	–			C–O str	Polysaccharydes, celulose	[17]
1292–1294		–						
1334–1365	–	–				In-plane O–H bending	Phenolic compounds	[18]
1504–1589	–	–	–		+	sym NH_3 bend, amide II	Amino acids, protein	[8,19,23]
1610–1645	–	+				NH_2, NH bend amide II, lactam	Caffeine, protein	[18,19]
1649–1670				+	–	Amide I	Protein	[23]
1674–1701	–	–	+			C=O str in aryl conjugated acids	Aromatic acids, CGA	[28,34]
1705–1720			+	–		C=O str	Ketones, aliphatic acids	[18,28,35]
1722–1759	+	+	–	–		C=O str	Lipids, aldehyde	[19,23,28,35]
1760–1774		–				C=O str adjacent to C–O–group	Vynil esters, lactones	[28]
2810–2848	+	+	–	–		sym CH str in CH_2	Lipids	[18]
2850–2877	–	+/–		+	+	sym CH str in CH_3	Caffeine	[18]
2908–2920	+				–	asym CH_2 str	Lipids	[3,18,19]
2935–2970	–	–	+	+	+	asym CH_3 str	Caffeine	[18]

LS=light sour, DS=dark sour, BL=black, ND=non-defective, IM=immature.
sym=symmetric, asym=asymmetric, str=stretching, CGA=chlorogenic acids.

Table 3
Tentative chemical assignment of significant NIR bands selected by Elastic net (α=0.75) for the classification of defective and non-defective coffees.

Gereral region ranges	ND	LS	DS	BL	IM	Vibration modes	Compounds	References
1215–1224			–	+		2nd overtone CH in CH_2 and CH groups	Quinic acid, carbohydrates, amino acids, caffeine	[24,6,30]
1264–1276	–							
1398		+				2 × CH str+2 × CH def (=CH) comb	Caffeine	[30]
1411	+	–				1st overtone OH	ROH, oil	[21]
1420–1425	–	+	–			1st overtone OH in aromatic compounds	CGA	[6,21,11]
1450–1491			–	+	–	1st overtone OH str	Water, CGA and phenols	[6]
1504–1527	–					1st overtone NH str in proteins	Proteins	[24,11]
1680–1755	+	+		–	–	1st overtone CH (=CH)	Lipids	[24,31]
1760–1871	+				–	2nd overtone OH str+CO str	Carbohydrates	[24,11]
1937–1959		–	+			2nd overtone C=O in CO_2R OH str+OH def comb	CGA Water	[24]
1970–1993	–	–				NH asym str+NH in-plane ben comb in amides	Protein and nitrogenous compounds	[24]
2085–2114				+		2st overtone C=O OH comb bands	CGA	[6,11]
2132–2166	+		–			=CH str+C=C str (HC=CH) comb	Lipids	[24]
2170–2230		+				2 × amide I+amide III comb CH str+C=O str comb C=O str+amide III comb	Protein	[24,21]
2238–2266	–		–		+	CH str+CH def comb	Caffeine	[31,36]
2283–2293		+				CH str+CH def comb (CH_2 CH_3)	Caffeine, CGA	[6,31]

LS=light sour, DS=dark sour, BL=black, ND=non-defective, IM=immature.
comb=combination, def=deformation, str=stretching, ben=bending, CGA=chlorogenic acids, *less likely.

coffee exhibited positive coefficients. Reports regarding the total carbohydrate content in roasted defective and non-defective coffees are scarce in the literature. Vasconcelos et al. [25] found higher concentration of total carbohydrates in both crude and roasted non-defective coffee, which is in agreement with the results found in this study. However, the contents were estimated by difference. In terms of sucrose, previous studies have reported higher levels of sucrose in non-defective beans in comparison to defective ones, but after roasting, only traces were found in either defective or non-defective coffees. Low sucrose levels in immature crude beans are associated with bean maturation, whereas in the case of black and sour beans, are due to loss by fermentation [25].

A characteristic FTIR band associated with lipids occurs at 1740 cm^{-1} (C=O stretching in esters). Results presented in Fig. 3a and b, from 1722 to 1759 cm^{-1}, show positive coefficients for non-defective and light sour coffees and negative for dark sour coffee. Non-zero coefficients at regions 2810–2833 cm^{-1} and 2908–2920 cm^{-1} were associated with symmetric and asymmetric C–H stretching vibrations in CH_2 groups, respectively [18]. Again we observed positive coefficients for non-defective and light sour coffees and negative for dark sour and black coffees. Non-defective and immature coffees also displayed positive coefficients at 3007 cm^{-1}, where symmetric CH stretching of *cis*-olefinic groups (=C–H in *cis* RHC=CHR) absorb [1]. In the NIR spectra, regions characterized by lipids absorption that exhibited non-zero coefficients were the following: 1411, 1680–1755 and 2132–2166 nm, assigned to the 1st overtone of O–H, 1st overtone of C–H stretching in CH_2 groups and a combination of =CH stretching and C=C stretching, respectively [21,24]. Non-defective coffees exhibited positive coefficients at these regions, whereas dark sour and black coffees exhibited negative ones, corroborating the ATR-FTIR observations. Oliveira et al. [26] reported higher content of lipids for non-defective coffees in comparison to defective, crude or roasted. In addition, the region 1678–1686 nm was attributed to lipids in the study by Ribeiro et al. [6] and selected as an important region for the attributes of flavor, cleanliness and overall quality of the coffee beverage.

Chlorogenic acids (CGA) represent a family of esters formed between quinic acid and one to four residues of certain *trans*-cinnamic acids, most commonly caffeic, *p*-coumaric and ferulic [27]. In the mid-infrared region, a band from 675 to 900 cm^{-1} (out-of-plane C–H bending vibration) is a potential indicator of chlorogenic acids and trigoneline [18]. In this study, from 800 to 835 cm^{-1}, positive coefficients for dark sour and black coffees and negative coefficients for light sour and immature were observed. Light sour and non-defective coffees exhibited negative coefficients at 1334–1365 cm^{-1}, where in-plane O–H bending in phenolic compounds absorb [18]. An evaluation of Fig. 3a and b also show negative coefficients for non-defective coffee at 1674 cm^{-1} that could be related to the C=O stretching vibration in aryl conjugated acids [28].

In the NIR spectra, a number of regions were assigned to CGA. The ones selected by Elastic net, with positive coefficients for black and negative for immature coffee, were the regions 1450–1491 and 2085–2114 nm characterized by the 1st overtone of O–H and the 2nd overtone of C=O and O–H combination bands, respectively [6]. These results, together with the results obtained by ATR-FTIR, reinforce that, under the same roasting conditions, black and dark sour coffees undergo lighter roasting, resulting in less degradation and significant higher CGA levels after roasting. On the other hand, high levels of CGA are found in crude immature beans [29], but most of this content is expected to be degraded during roasting. The degradation and final level of CGA in coffees is highly dependent on the extent of roasting that coffee is subject to. Farah et al. [27] observed that light roasts would lead to a rise in the total CGA amount, which could be a result of the loss of other compounds that are more sensitive to heat, causing a relative increase in the levels of the remaining ones. Medium or dark roasts cause loss of total CGA that are isomerized, degraded and dehydrated giving rise to other compounds such as lactones.

Caffeine is generally observed at the range of 1600–1650 cm^{-1} of the mid-infrared spectrum, where cyclic amides absorb [19]. Non-defective and light sour coffees exhibited negative and positive coefficients at 1610–1645 cm^{-1}, respectively (Fig. 3a and b). In a study that aimed to quantify caffeine in soft drinks, Paradkar and Irudayaraj [20] attributed a peak at 2882 cm^{-1} to the symmetric stretching of C–H bonds of methyl groups in the caffeine molecule. The same peak was reported by Silverstein et al. [18] to occur at 2870 cm^{-1}, while the CH asymmetric stretching of methyl groups was reported to absorb at 2962 cm^{-1}. In our study, negative coefficients for non-defective and positive for dark sour, black and immature coffees were observed at 2850–2877 and 2935–2970 cm^{-1}, while light sour exhibited both positive and negative coefficients in these regions. Franca et al. [29] studied the physical and chemical differences between defective and non-defective coffees and reported lower levels of caffeine in non-defective coffees prior to roasting. While the caffeine content of non-defective and immature classes decreased, the roasting did not affect the caffeine levels of black or sour coffees, which remained relatively constant upon roasting. Thus, the authors suggested that under the same roasting conditions, black and sour beans were roasted to lesser extents than other beans. In the NIR models developed, light sour exhibited positive coefficients at 1398 and 2283–2293 nm, due to the combination of CH stretching and deformation modes [6,30]. The region 2238–2266 nm, that also attributed to CH combination bands in caffeine [31], was selected with positive coefficients for immature and negative coefficients for non-defective and dark sour coffees.

Protein bands are observed in the 1515–1670 cm^{-1} region of the mid-infrared spectrum. Although proteins may absorb in this wide range, strong absorption bands at 1550 – 1567 cm^{-1} and ~1653 cm^{-1} are expected, due to NH bending in amide II groups, and NH_2 vibrations in amide I groups, respectively [19,23]. The NH bending vibrations of amino groups from amino acids are observed at 1504–1550 cm^{-1} and 1575–1600 cm^{-1} [18]. In this study, non-zero coefficients were found from 1504 cm^{-1} to 1670 cm^{-1}. Overall, in the region 1504–1589 cm^{-1}, which can be attributable to proteins or free amino acids, immature coffee exhibited positive, and non-defective, light and dark sour coffees exhibited negative coefficients. In the region 1649–1670 cm^{-1}, attributable to proteins absorption, black coffee exhibited positive coefficients, while immature coffee exhibited negative coefficients. In the NIR models, non-defective coffee exhibited negative coefficients in the regions 1504–1527 and 1970–1993 nm, where the 1st overtone of NH and a combination of NH stretching and bending vibrations in proteins occur [11,24]. Light sour beans exhibited negative coefficients at 1970–1993 nm, but positive coefficients at 2170–2230 nm. The latter region is assigned to series of combination bands characteristic from proteins [21]. Mazzafera [32] determined the concentration of proteins and amino acids in non-defective, immature and immature-black crude coffee beans, finding higher levels of proteins in non-defective, and higher levels of amino acids in immature coffees. However, a direct comparison of the results obtained by Mazzafera [32] and the ones obtained in this study is not possible given the numerous chemical changes occurring during the roasting process. In particular, free amino acids are pyrolysed or react to form Maillard products, resulting in a considerable decrease in their levels, and proteins are denatured. On the other hand, Vasconcelos et al. [25] found lower levels of proteins in non-defective than defective coffees roasted to medium and dark degrees, which is in agreement with the NIR results obtained in this research. In the study by Santos et al. [10], the NIR region of 2170–2500 nm contributed the most for the discrimination between defective and non-defective crude coffees. Even though the authors did not attempt to correlate this discrimination with chemical compounds, in our study, the same region played an important role being assigned to several compounds such as proteins, lipids, caffeine and CGA.

Besides the major coffee compounds previously discussed, the carbonyl absorption region of 1680–1800 cm^{-1} in the mid-infrared exhibits low molecular weight compounds that are formed or degraded during roasting and affect the aroma and taste of the beverage [28]. Because of that, this region has been used to provide insights into the chemical changes that occur in coffees during roasting [19,28]. It is, however, challenging to clearly identify each of those compounds and evaluate their presence and influence in different coffees since their absorption peaks are near each other. According to Lyman et al. [28] and Wang and Lim [19], general band assignments of a number of known compounds are as follows: aromatic acid (1680–1700 cm^{-1}), aliphatic acid (1705–1714 cm^{-1}), ketone (1707–1714 cm^{-1}), aldehyde or ketone (~1726 cm^{-1}), aldehyde (1723–1729 and 1738–1741 cm^{-1}), aliphatic ester (1744–1754 cm^{-1}), and vinyl ester and/or lactone (1762–1780 cm^{-1}). A more detailed evaluation of the coefficient estimates at the region of 1680–1780 cm^{-1} indicates that non-defective coffee was associated with high content of aldehydes and low content of aromatic acids or CGA, while black coffee was associated with low levels of aldehydes, ketones and/or aliphatic acids. This result suggests that, under the same roasting conditions, non-defective coffee attained a high extent of roast, with a higher formation of volatile compounds, while the opposite was observed for black coffee. Light sour was associated with high levels of aldehydes or lipids and low levels of aromatic acids or CGA and unsaturated esters and lactones, and dark sour exhibited positive coefficients at regions characterized by the absorption of aromatic and aliphatic acids, although CGA may have also contributed to the latter region. The naturally higher acidity of sour beans in comparison to other

Table 4
Major chemical compounds assigned by both ATR-FTIR and NIR variable selection.

Compounds	ND	LS	DS	BL	IM
Carbohydrates	+				
Proteins and/or amino acids					
Lipids	+		−	−	−
Caffeine	−				+
Chlorogenic acids	−	−		+	−

LS=light sour, DS=dark sour, BL=black, ND=non-defective, IM=immature, +=higher level than other classes; −=lower level than other classes.

classes, due to bean fermentation, was previously demonstrated in the literature for green coffee [25,29]. Additionaly, a high content of CGA may indicate that dark sour attained a lighter extent of roast, confirming the previously discussed results.

A summarized compilation of the major chemical compounds assigned for the peaks from the spectra obtained by both techniques is shown in Table 4. Most of the assignments in the ATR-FTIR variable selection results were in agreement with the ones assigned in the NIR models. In general, the correct classification of non-defective coffee was associated with higher levels of carbohydrates and lipids, and lower levels of proteins, caffeine and chlorogenic acids than defective coffees. Although the chemical differences between defective and non-defective crude coffees was crucial for the correct classification of each sample class, the higher extent of roasting attained by non-defective and the lesser attained by defective beans, under the same roasting conditions, was also a key factor for the discrimination of the sample classes. The high content of free sugars available for reactions in non-defective beans resulted in a more efficient roasting, with extensive degradation of compounds such as amino acids and, consequently, large production of aroma compounds, including ketones and aldehydes. The same was not observed for defective beans. In particular, the low extent of roasting attained by dark sour and black beans caused a reduced degradation of CGA. It must be mentioned, however, that the establishment of the roasting degrees was based in luminosity measurements. Since black coffees naturally exhibit low luminosity before roasting, it is also possible that these beans attained an incomplete roasting.

4. Conclusion

Overall, both ATR-FTIR and NIR were found to be powerful techniques for the discrimination of roasted defective and non-defective coffees. A comparison between both techniques indicated that ATR-FTIR can provide more information and selectivity on the group frequencies present in the samples. It is well known that precise band assignments are difficult in the near-infrared region because of the fact that a single band may be attributable to several possible combinations of fundamental and overtone vibrations overlapped. However, the employment of Elastic net provided insights on the characterization of the samples and on the visualization of discrete spectral bands associated with the correct classification of defective and non-defective coffees. The main chemical descriptors that characterized the coffee samples were carbohydrates, proteins/amino acids, lipids, caffeine and chlorogenic acids. The positive classificatory results obtained in this study also indicate that FTIR and NIR could be used to predict the percentage of defective coffees in mixture with non-defective ones and is the subject of an ongoing investigation.

Acknowledgments

The authors acknowledge financial support from the Brazilian Government Agencies (BZG) CNPq (CNPq306139/2013-8; CNPq 475746/2013-9 CNPq505001/2013-6) and FAPEMIG (APQ-5112-6.01-07).

Appendix A. Supplementary Information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.talanta.2014.05.001.

References

[1] H. Yang, J. Irudayaraj, J. Am. Oil Chem. Soc. 78 (2001) 889–895.
[2] I. Esteban-Díez, J.M. González-Sáiz, C. Sáenz-González, C. Pizarro, Talanta 71 (2007) 221–229.
[3] N. Reis, A.S. Franca, L.S. Oliveira, Lebensm. Wiss. Technol. 50 (2013) 715–722.
[4] N. Reis, A.S. Franca, L.S. Oliveira, Talanta 115 (2013) 563–568.
[5] H. Ebrahimi-Najafabadi, R. Leardi, P. Oliveri, M. Chiara Casolino, M. Jalali-Heravi, S. Lanteri, Talanta 99 (2012) 175–179.
[6] J.S. Ribeiro, M.M.C. Ferreira, T.J.G. Salva, Talanta 83 (2011) 1352–1358.
[7] A.P. Craig, A.S. Franca, L.S. Oliveira, Food Chem. 132 (2012) 1368–1374.
[8] A.P. Craig, A.S. Franca, L.S. Oliveira, Lebensm. Wiss. Technol. 47 (2012) 505–511.
[9] A.P. Craig, A.S. Franca, L.S. Oliveira, J. Food Sci. 76 (2011) C1162–C1168.
[10] J.R. Santos, M.C. Sarraguça, A.O.S.S. Rangel, J.A. Lopes, Food Chem. 135 (2012) 1828–1835.
[11] J. Workman, L. Weyer, Practical Guide to Interpretive Near-infrared Spectroscopy, CRC Press, Boca Raton, FL, 2007.
[12] H. Zou, T. Hastie, J. Roy. Stat. Soc. B Met. 67 (2005) 301–320.
[13] J. Friedman, T. Hastie, R. Tibshirani, J. Stat. Softw. 33 (2010) 1–22.
[14] J. Zhu, T. Hastie, Biostatistics 5 (2004) 427–443.
[15] N. Damayanti, A.P. Craig, J. Irudayaraj, Analyst 138 (2013) 7127–7134.
[16] M.D. Dyar, M.L. Carmosino, E.A. Breves, M.V. Ozanne, S.M. Clegg, R.C. Wiens, Spectrochim. Acta B 70 (2012) 51–67.
[17] E.K. Kemsley, S. Ruault, R.H. Wilson, Food Chem. 54 (1995) 321–326.
[18] R.M. Silverstein, F.X. Webster, D.J. Kiemle, Spectrometric Identification of Organic Compounds, John Wiley & Sons, Hoboken, NJ, 2005.
[19] N. Wang, L.-T. Lim, J. Agr. Food Chem. 60 (2012) 5446–5453.
[20] M.M. Paradkar, J. Irudayaraj, Food Chem. 78 (2002) 261–266.
[21] J.S. Shenk, M.O. Westerhaus, J.W. Workman, Application of NIR spectroscopy to agricultural products, in: E.W. Ciurczak, D.A. Burns (Eds.), Handbook of Near-infrared Analysis, third ed.,CRC Press, Boca Raton, FL, 2007, pp. 347–386.
[22] R. Tibshirani, J. Roy. Stat. Soc. B Met. 73 (2011) 273–282.
[23] R. Karoui, G. Downey, C. Blecker, Chem. Rev. 110 (2010) 6144–6168.
[24] I. Esteban-Diez, J.M. Gonzalez-Saiz, C. Pizarro, Anal. Chim. Acta 525 (2004) 171–182.
[25] A.L.S. Vasconcelos, A.S. Franca, M.B.A. Gloria, J.C.F. Mendonca, Food Chem. 101 (2007) 26–32.
[26] L.S. Oliveira, A.S. Franca, J.C.F. Mendonca, M.C. Barros, Lebensm. Wiss. Technol. 39 (2006) 235–239.
[27] A. Farah, T. De Paulis, L.C. Trugo, P.R. Martin, J. Agr. Food Chem. 53 (2005) 1505–1513.
[28] D.J. Lyman, R. Benck, S. Dell, S. Merle, J. Murray-Wijelath, J. Agr. Food Chem. 51 (2003) 3268–3272.
[29] A.S. Franca, L.S. Oliveira, J.C.F. Mendonca, X.A. Silva, Food Chem. 90 (2005) 89–94.
[30] C. Pizarro, I. Esteban-Díez, J.-M. González-Sáiz, M. Forina, J. Agric. Food Chem. 55 (2007) 7477–7488.
[31] C.W. Huck, W. Guggenbichler, G.K. Bonn, Anal. Chim. Acta 538 (2005) 195–203.
[32] P. Mazzafera, Food Chem. 64 (1999) 547–554.
[33] R. Briandet, E.K. Kemsley, R.H. Wilson, J. Sci. Food Agric. 71 (1996) 359–366.
[34] Z. Fábián, V. Izvekov, A. Salgó, F. Örsi, Anal. Proc. 31 (1994) 261–263.
[35] N. Wang, Y. Fu, L.-T. Lim, J. Agric. Food Chem. 59 (2011) 3220–3226.
[36] G. Downey, J. Boussion, J. Sci. Food Agric. 71 (1996) 41–49.