Food Chemistry 192 (2016) 374–379

Contents lists available at ScienceDirect

Food Chemistry

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

Analytical Methods

Using UV–Vis spectroscopy for simultaneous geographical and varietal classification of tea infusions simulating a home-made tea cup

Paulo Henrique Gonçalves Dias Diniz [a,*], Mayara Ferreira Barbosa [a], Karla Danielle Tavares de Melo Milanez [a], Marcelo Fabián Pistonesi [b], Mário César Ugulino de Araújo [a,*]

[a] Universidade Federal da Paraíba, CCEN, Departamento de Química, P.O. Box 5093, 58051-970 João Pessoa, Brazil
[b] Universidad Nacional del Sur, INQUISUR, Departamento de Química, 8000 Bahía Blanca, Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 30 May 2014
Received in revised form 20 November 2014
Accepted 7 July 2015
Available online 8 July 2015

Keywords:
Tea
Classification
Variety
Geographical origin
UV–Vis spectroscopy
Successive Projections Algorithm

ABSTRACT

In this work we proposed a method to verify the differentiating characteristics of simple tea infusions prepared in boiling water alone (simulating a home-made tea cup), which represents the final product as ingested by the consumers. For this purpose we used UV–Vis spectroscopy and variable selection through the Successive Projections Algorithm associated with Linear Discriminant Analysis (SPA-LDA) for simultaneous classification of the teas according to their variety and geographic origin. For comparison, KNN, CART, SIMCA, PLS-DA and PCA-LDA were also used. SPA-LDA and PCA-LDA provided significantly better results for tea classification of the five studied classes (Argentinean green tea; Brazilian green tea; Argentinean black tea; Brazilian black tea; and Sri Lankan black tea). The proposed methodology provides a simpler, faster and more affordable classification of simple tea infusions, and can be used as an alternative approach to traditional tea quality evaluation as made by skilful tasters, which is evidently partial and cannot assess geographic origins.

1. Introduction

Since ancient times, tea has been used by the Asiatic cultures not only as an herbal medicine, but also for its characteristic flavour and aroma. The great present-day popularity of teas as beverages is mainly due to the presence of polyphenols and caffeine, which respectively determine up to 30% and up to 4% of the dry weight (Kumar, Murugesan, Kottur, & Gyamfl, 2012). Polyphenols have been shown to present various benefits for human health, nutrition, and physiology (Khan & Mukhtar, 2007; Sharma, 2014), while caffeine is principally attractive due to its stimulatory effects, which are frequently used by the pharmacological industry (Spiller, 1997; Wang, Wan, Hu, & Pan, 2008). The additional health benefits of teas are also frequently described in the literature (Sharangi, 2009; Chow & Hakim, 2011; Pinto, 2013).

Tea is the second most consumed non-alcoholic beverage in the world (after water), and is prepared by brewing the dried leaves of *Camellia sinensis* in water. The types of tea (white, yellow, green, oolong, black, and Pu-ehr) basically differ with regards to the extent of fermentation. Green (unfermented) and black (fully fermented) teas are the two most popular categories, together accounting for around 98% of both world production and consumption (Diniz et al., 2012; Pinto, 2013). The plant is highly cultivated in Asia, Africa, and South America. In South America, Argentina and Brazil are the main tea producers, respectively harvesting 90.7 and 7.7 thousand tons in 2012 (Food & intergovernmental group on tea. Current situation, 2012). In the 1920's Japanese immigrants to Brazil initiated cultivation using tea seeds from Sri Lanka and India. Despite a relatively small tea industry in Brazil, the tea producers here have achieved some increase in market share due to their efforts to improve the quality of Brazilian teas. In Argentina, tea was first introduced in 1920 (with Russian seeds) and it is currently the world 9th largest tea producer.

In worldwide tea trading, consumer interest and a country's reputation (clearly indicating geographic origin), has increasingly become synonymous with higher than average prices (Ye, 2012). As an example the famous "Lion" logo of "Ceylon" or "Sri Lankan" teas (administered by the Sri Lankan Tea Board) is still regarded worldwide as a sign of quality and taste.

Various analytical methods to verify the geographic origins of teas have been proposed in the literature with the purpose of providing some sort of security to both tea traders and consumers, and to prevent fraudulent labelling (Ye, 2012). However, most of these techniques require laborious sample preparation, and induce

* Corresponding authors.
E-mail addresses: paulodiniz.quimica@gmail.com (P.H.G.D. Diniz), laqa@quimica.ufpb.br (M.C.U. de Araújo).

http://dx.doi.org/10.1016/j.foodchem.2015.07.022
0308-8146/

significant operational expenditures. We therefore propose a tea classification strategy that provides precise and reliable results, and can be implemented in a routine laboratory. Ultraviolet–Visible (UV–Vis) spectroscopy is one of the most common techniques used in routine analysis, and has already been used to differentiate between black, green, and Pu-erh tea varieties (Pallacios-Morillo, Alcázar, de Pablos, and Jurado (2013)). However in this method, methanol was used as the extractor solvent creating very broad spectra with highly correlated variables. This requires more sophisticated non-linear classifiers such as Artificial Neural Networks (ANNs), and Support Vector Machines (SVMs). Methanol also presents toxicity to humans as well as to the environment (Clary, 2013). Geographical classification of teas using UV–Vis spectroscopy has not been reported in the literature. Simultaneous classification of both geographic origin and variety for teas has been proposed using digital images (Diniz et al., 2012), and near-infrared spectroscopy (NIRS) (Diniz, Gomes, Pistonesi, Band, & Araújo, 2014). However, these methodologies were carried out directly on the tea as contained in commercialized bags, whereas the infusion represents the final product as ingested by the consumer. Tea quality moreover is traditionally evaluated by skilful tasters based on the infusion's appearance, taste, and aroma, which is evidently partial and cannot assess a tea's geographic origin (Diniz et al., 2014).

In this work we propose a method to verify the differentiating characteristics of simple tea infusions prepared in boiling water alone (simulating a home-made tea cup). This form represents the final product as ingested by the consumer. For this purpose, UV–Vis spectroscopy, and variable selection using the Successive Projections Algorithm associated with Linear Discriminant Analysis (SPA-LDA) (Soares, Gomes, Galvão Filho, Araújo, & Galvão, 2013) was used for simultaneous classification of teas according to their variety (black or green tea), and their geographic origin (Argentina, Brazil, or Sri Lanka). For comparison, other supervised pattern recognition techniques such as K-nearest neighbours (KNN), Classification, Regression Tree (CART), Soft Independent Modelling by Class Analogy (SIMCA), Partial Least Squares Discriminant Analysis (PLS-DA), and Principal Component Analysis-Linear Discriminant Analysis (PCA-LDA) were used. It is worth noting that SPA-LDA has been successfully applied to classify other foods such as edible vegetable oils using square wave voltammetry (Gambarra-Neto et al., 2009), coffees using UV–Vis spectroscopy (Souto et al., 2010), beers using NIR spectroscopy (Ghasemi-Varnamkhasti et al., 2012), and honeys using digital images (Domínguez, Diniz, Di Nezio, Araújo, & Centurión, 2014).

2. Materials and methods

2.1. Samples

One hundred tea samples were purchased from local supermarkets in the cities of João Pessoa (Brazil), and Bahía Blanca (Argentina): 20 Brazilian black teas, 20 Brazilian green teas, 20 Argentinean black teas, 20 Sri Lankan black teas, and 20 Argentinean green teas. A sample quartering step was performed as described by Diniz et al. (2014). The contents of the 100 tea bags from each batch were quartered, and then reduced to a final sample containing 25 g, they were subsequently stored in sealed plastic bags to prevent contamination and/or adulteration.

2.2. Apparatus and procedure

The infusions were prepared using 1 g of each tea sample in 100 mL of double-distilled water at 90 °C, and let to stand for 5 min. The infusions were filtered with medium speed (8 μm) retention filter paper, and completed to 100 mL with double-distilled water. They were kept in Nalgene plastic bottles and left to cool to room temperature. Of each infusion 10 mL were transferred to a 50 mL volumetric flask and diluted with double-distilled water. The spectrum for each sample was immediately acquired using a Hewlett Packard 8453 spectrophotometer equipped with a quartz cell, with an optical path of 1 cm, and with a photodiode array in the range 190–800 nm with a resolution of 1 nm. A blank spectrum was also recorded using double-distilled water alone.

2.3. Data analysis

The tea samples were divided into training (75%), and test (25%) sets by applying the Kennard-Stone (KS) uniform sampling algorithm (Kennard & Stone, 1969). Differences between unsupervised (PCA), and supervised (KNN, CART, SIMCA, PLS-DA, PCA-LDA, and SPA-LDA) pattern recognitions were evaluated. The validation step for each of the algorithms was performed using full cross-validation. The test samples were used for the final data evaluation, and for comparison of the classification models only (Soares et al., 2013).

The KS and SPA-LDA algorithms were performed with Matlab® 2009b (Mathworks Inc.) software. The other chemometric approaches were performed by using the Classification toolbox for Matlab® (version 2.0) released by Milano Chemometrics, and QSAR Research Group (Ballabio & Consonni, 2013), and may be found on the following site: http://michem.disat.unimib.it/chm/.

3. Results and discussion

3.1. Exploratory analysis of the data

Fig. 1a shows the absorbance spectra of the simple tea infusions in the range of 190–800 nm. The spectra present a profile similar to that of data published by Pallacios-Morillo et al. (2013), although their spectra were much broader and highly correlated. This is because the solvent (in this case, methanol) affects the position of the spectral band, and the maximum absorbance, i.e. the values of λ_{max}, molar absorptivity ε, and the shape of the spectrum. As can be seen in Fig. 1a, the most informative portion of the spectra is found in the UV region (190–500 nm), where the useful transitions are $\pi \rightarrow \pi^*$ for compounds with conjugated double bonds, some $n \rightarrow \sigma^*$, and some $n \rightarrow \pi^*$ transitions. Two absorption bands are easily seen in the ranges from 190 to 250 nm, and from 250 to 300 nm, and another broad absorption band appears around 300–400 nm. These band absorptions are related to phenolic compounds presented in the tea infusions (Fernández, Martín, González, & Pablos, 2000; Obanda, Owuor & Mang'oka, 2001; Fernández, Pablos, Martín, & González, 2002; Yao et al., 2006; Kilinc, 2009).

Fig. 1b shows the average spectra of the five studied tea classes, in the range from 240 to 340 nm. As can be seen, the shape of the mean exhibits a clear separation tendency into two major groups: the black teas; Brazilian black teas (BrB, magenta line), Argentinean black teas (ArB, blue line), and Sri Lankan black teas (SkB, green line), and the green teas; Brazilian green teas (BrG, red line), and Argentinean green teas (ArG, black line). In order to verify the above, PCA was performed using the entire spectral range (Fig. 2a) and five selected intervals: 190–250 nm (Fig. 2b), 251–310 nm (Fig. 2c), 311–370 nm (Fig. 2d), 371–430 nm (Fig. 2e), and 431–490 nm (Fig. 2f). The region above 490 nm presents no analytical information. We noted that the ranges 251–310 nm, 371–430 nm, and 431–490 nm contribute significantly towards differentiating between the black and green teas.

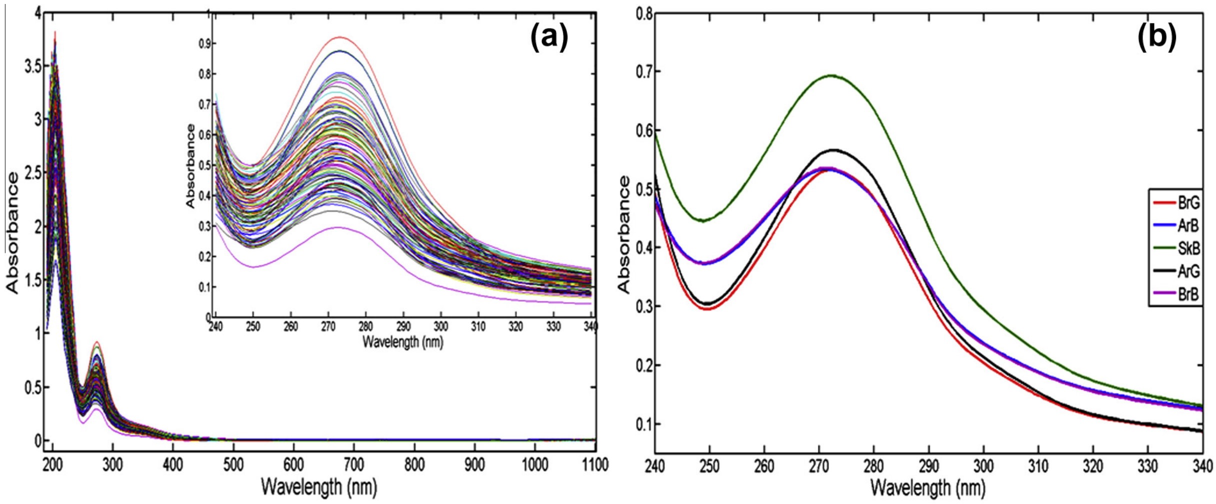

Fig. 1. (a) Raw UV–Vis spectra of all studied tea samples. (b) Mean spectra of the five studied tea classes. Argentinean green (ArG, ▬), Brazilian green (BrG, ▬), Argentinean black (ArB, ▬), Brazilian black (BrB, ▬), and Sri Lankan black (SkB, ▬).

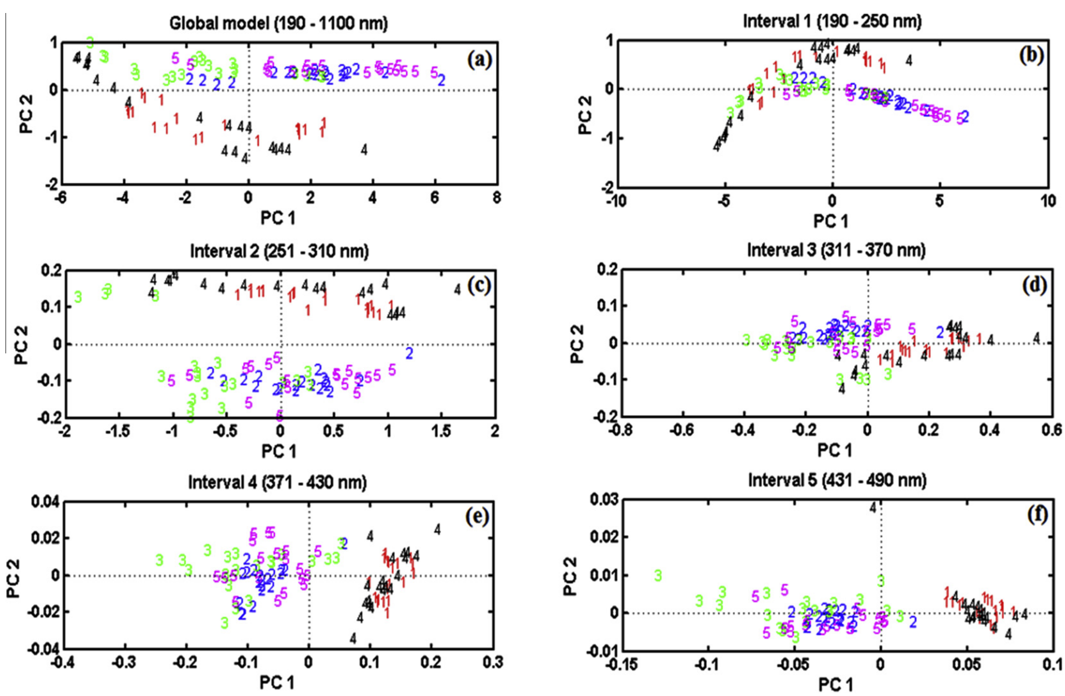

Fig. 2. PCA score plots (PC1 × PC2) using the (a) entire spectra and five selected intervals: (b) 190–250 nm, (c) 251–310 nm, (d) 311–370 nm, (e) 371–430 nm, and (f) 431–490 nm. Brazilian green teas (**1**), Argentinean black teas (**2**), Sri Lankan black teas (**3**), Argentinean green teas (**4**), and Brazilian black teas (**5**).

Based on this result obtained by PCA, we choose to study two differing spectral ranges: (a) the entire UV–Vis, and (b) the range of 251–490 nm. Moreover, it is worth to note that geographical discrimination of the samples could not be achieved, which justifies the use of supervised pattern recognition techniques. In the work of Pallacios-Morillo et al. (2013), PCA exhibited a noteworthy discrimination between black, Pu-her, and green teas, which could dispense with the use of more sophisticated non-linear classifiers (machine learning approaches), such as ANNs and SVMs.

3.2. Classification

The construction of the multivariate classification models was performed using a training set (75% of the studied samples). Each model was validated using the leave-one-out cross-validation technique. A test set (25% of the studied samples) was then used for final data evaluation and comparison to the classification models. The performance of the models was evaluated by accuracy, which is defined using the ratio of samples in the test set correctly assigned into their respective classes. Table 1 presents the assignation of the test set samples into the five studied tea classes using KNN, CART, SIMCA, PLS-DA, PCA-LDA and SPA-LDA for both the entire UV–Vis spectral range as well as the range of 251–490 nm. The summary of the classification accuracy for each model is shown in Table 2.

3.2.1. K-nearest neighbours

KNN is a classification method based on a distance matrix, in which an object is classified according to the classes of its K-nearest neighbors in the data space, i.e. it classifies unlabelled objects based on their similarity with samples in the training set. The optimal KNN models were obtained using 4 and 1

Table 1
Assignation of the test set samples into the five studied tea classes using KNN, CART, SIMCA, PLS-DA, PCA-LDA, and SPA-LDA.

Class	UV–Vis range					251–490 nm				
	BrG	ArB	SkB	ArG	BrB	BrG	ArB	SkB	ArG	BrB
KNN										
BrG	5	0	0	0	0	3	0	0	2	0
ArB	0	3	0	0	2	0	5	0	0	0
SkB	0	1	4	0	0	0	0	5	0	0
ArG	0	0	0	5	0	0	0	0	5	0
BrB	0	0	0	0	5	0	2	0	0	3
CART										
BrG	3	0	0	2	0	3	0	0	2	0
ArB	0	4	1	0	0	0	4	1	0	0
SkB	0	0	4	0	1	0	0	4	0	1
ArG	0	0	0	5	0	1	0	0	4	0
BrB	0	0	0	0	5	0	0	0	0	5
SIMCA										
BrG	1	0	0	1	0	5	0	0	0	0
ArB	0	0	1	0	1	0	5	0	0	0
SkB	0	0	2	0	0	0	1	4	0	0
ArG	0	0	0	2	0	0	0	0	5	0
BrB	0	0	0	0	5	0	2	0	0	3
PLS-DA										
BrG	3	1	0	0	0	2	0	0	0	0
ArB	0	5	0	0	0	0	4	0	0	0
SkB	0	0	4	0	0	0	0	3	0	0
ArG	0	0	0	3	0	0	0	0	3	0
BrB	0	0	0	0	5	0	0	0	0	5
PCA-LDA										
BrG	5	0	0	0	0	5	0	0	0	0
ArB	0	5	0	0	0	0	5	0	0	0
SkB	0	1	3	1	0	0	0	5	0	0
ArG	0	0	0	5	0	0	0	0	5	0
BrB	0	0	0	0	5	0	0	0	0	5
SPA-LDA										
BrG	5	0	0	0	0	5	0	0	0	0
ArB	0	5	0	0	0	0	5	0	0	0
SkB	0	1	4	0	0	0	0	5	0	0
ArG	0	0	0	5	0	0	0	0	5	0
BrB	0	0	0	0	5	0	0	0	0	5

BrG: Brazilian green teas; ArB: Argentinean black teas; SkB: Sri Lankan black teas; ArG: Argentinean green teas; BrB: Brazilian black teas.
Gray shaded values represents the values included correctly into their own classes.

(respectively) for nearest neighbours, for the entire spectral range and for the range of 251–490 nm, thus reaching a classification accuracy of 88% and 84%. In the first case, two ArB samples were misclassified as BrB, and one SkB sample was misclassified as ArB tea, meanwhile in the last case two BrG samples were misclassified as ArG, and two BrB samples as ArB (Table 1).

3.2.2. Classification and Regression Tree

CART is a tree-building technique based on rule induction, in which the data space is successively partitioned into different class subsets based on associated variables being significantly related to the response variable. In this case, the response variable is categorical. The final classification model consists of a collection of nodes (tree) that define the classification rule. The partitioning in each node is obtained by maximizing the purity of the new subsets. The optimal CART models were obtained with a classification accuracy of 84% and 80%, respectively for the entire UV–Vis spectral range and the range of 251–490 nm. In the first case, two BrG samples were misclassified as ArG, one ArB sample misclassified as SkB, and one SkB sample misclassified as BrB, in the last case one ArG sample was misclassified as BrG, besides the same samples being misclassified for the entire UV–Vis spectral range (Table 1).

Table 2
Summary of the classification accuracy for KNN, CART, SIMCA, PLS-DA, PCA-LDA, and SPA-LDA models using both the entire UV–Vis range and the range of 251–490 nm.

	Classification accuracy (%)	
	UV–Vis range	251–490 nm
KNN	88	84
CART	84	80
SIMCA	45	88
PLS-DA	80	68
PCA-LDA	92	100
SPA-LDA	96	100

3.2.3. Soft Independent Modelling by Cluster Analysis

SIMCA is a class modelling technique, in which the final classification model consists of a collection of PCA models, one for each class. A new object is then assigned by comparing the distances of the class models to the object. The optimal SIMCA models for both the entire UV–Vis spectral range and the range of 251–490 nm were obtained with a respective classification accuracy of 45% and 88%, respectively. As can be seen in Table 1, when the entire UV–Vis spectral range was used, one BrG sample was misclassified as ArG, and two SkB sample misclassified as ArB and BrB. Moreover, twelve samples (three BrG, three ArB, three SkB, and three ArG) were not assigned. For the range of 251–490 nm, one SkB and two BrB were misclassified as ArB.

3.2.4. Partial Least Squares Discriminant Analysis

PLS-DA is a modification of Partial Least Squares Algorithm for classification purposes. It is based on the PLS2 algorithm that searches for latent variables with a maximum covariance for the categorical variables (Y). The new object is then assigned to the class with the maximum value in the Y vector or, alternatively, a threshold between zero and one is determined for each class. In our work, the optimal PLS-DA model using the entire UV–Vis spectral range with 19 latent variables obtained a classification accuracy of 80%, being one BrG sample misclassified as ArB, and four other samples (one BrG, two ArG, and one SkB) that were not assigned (Table 1). On the other hand, three BrG, one ArB, two SkB, and two ArG samples were not assigned when using the range of 251–490 nm with 9 latent variables, totalling a classification accuracy of 68%.

For the differing organic molecules presented in the tea infusions, vibrational and rotational energy levels are overlapped on the electronic energy levels in the UV–Vis region, i.e. many transitions with different energies can occur, making the absorption bands broadened and highly correlated. Since PLS-DA is a full-spectrum-based technique, the analytical information comes overlapped, which interferes in the performance of the classification models in terms of its accuracy. This problem is circumvented using a suitable reduction variable technique that selects wavelengths whose information content is uncorrelated and/or minimally redundant. This fact is corroborated by the classification results obtained in Sections Principal Component Analysis-Linear Discriminant Analysis and Successive Projections Algorithm-Linear Discriminant Analysis.

3.2.5. Principal Component Analysis-Linear Discriminant Analysis

LDA classification methods employ linear decision boundaries (hyperplanes), which are defined in order to maximize between-class separability while minimizing within-class variability. For this purpose, the number of objects in the training set must be larger than the number of variables included in the LDA model, requiring a reduction in variables. This can be circumvented by using the PCA scores as input data; since linear combinations of the original variables called principal components (PCs) are uncorrelated. The optimal PCA-LDA models were obtained using 14 and

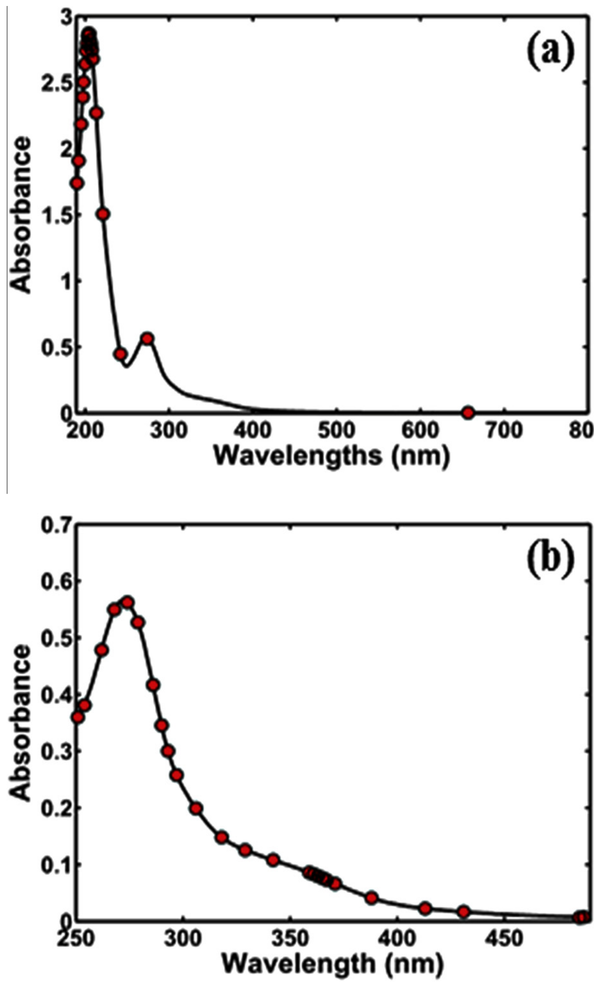

Fig. 3. Mean spectrum for all the tea samples studied with selected wavelengths using SPA for both (a) the entire spectral range and (b) the selected interval of 251–490 nm.

19 PCs respectively, for both the entire UV–Vis spectral range and the range of 251–490 nm, and reached a classification accuracy of 92% and 100%. When using the entire UV–Vis spectral range, one SkB sample was misclassified as ArB, and one ArG sample was misclassified as BrG (Table 1).

3.2.6. Successive Projections Algorithm-Linear Discriminant Analysis

Another way to circumvent LDA limitations is to use an adequate variable selection technique such as the Successive Projections Algorithm. SPA is an iterative forward selection method that selects such wavelengths as whose information content is minimally redundant, which solves collinearity problems. These chains of variables are then sequentially evaluated as based on the *G* cost function (Eq. (1)), which is calculated in the validation set as the average risk *G* of misclassification by LDA when the subset of variables under study is used (Soares et al., 2013):

$$G = \frac{1}{K\upsilon} + \sum_{k=1}^{K\upsilon} \frac{r^2(\mathbf{x}_k, \boldsymbol{\mu}_{Ik})}{\min_{Ij \neq Ik} \mathbf{r}^2(\mathbf{x}_k, \boldsymbol{\mu}_{Ij})} \tag{1}$$

where the numerator $r^2(\mathbf{x}_k,\boldsymbol{\mu}_{Ik})$ in the summation term is the squared Mahalanobis distance between object $\mathbf{x}_k$ and the sample mean $\boldsymbol{\mu}_{Ik}$ of its true class; and the denominator corresponds to the squared Mahalanobis distance between object $\mathbf{x}_k$ and the center of the closest wrong class. Ideally, object $\mathbf{x}_k$ should be close to the center of its true class and distant from the centers of all other classes.

In our work, the optimum number of variables selected using SPA was determined from the *G cost* function minimum of 0.7256 for the entire spectral range, and with 0.4426 for the range of 251–490 nm. When using the entire UV–Vis spectral range, the optimal SPA-LDA model selected twenty variables (190, 192, 195, 197, 198, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 213, 221, 242, 274, and 657 nm) (Fig. 3a) with a classification accuracy of 96%, being one SkB sample alone misclassified as ArB (Table 1). For the range of 251–490 nm, all of the samples in the test set were correctly classified using the twenty-five wavelengths (251, 254, 262, 268, 274, 279, 286, 290, 293, 297, 306, 318, 329, 342, 359, 361, 363, 365, 367, 371, 388, 413, 431, 485, and 487) selected by SPA-LDA, as indicated in Fig. 3b. A chemical or biochemical attribution to the selected wavelengths cannot be performed, because, for the differing organic molecules presented in the infusions, vibrational and rotational energy levels overlap the electronic energy levels in the UV–Vis region, i.e. many transitions with different energies can occur, broadening the bands. The broadening is even greater in solutions, owing to solvent–solute interactions. However, SPA selected the minimally correlated variables, and the application of the resulting LDA models for both the entire UV–Vis spectral range and the range of 251–490 nm classified correctly 96% and 100% of their respective samples. For illustration, Fig. 4 shows the discrimination of the test set samples for both cases in a three-dimensional graph, corresponding to the first three

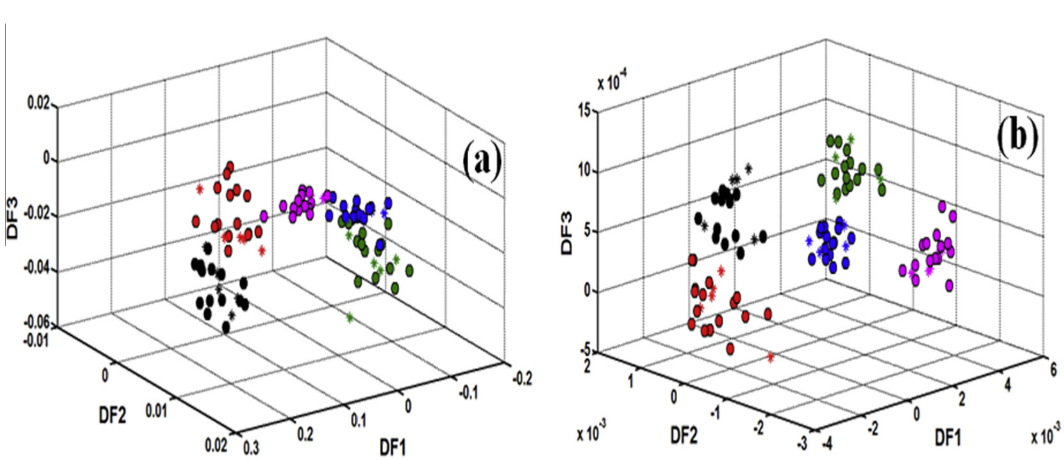

Fig. 4. Fisher's discriminant functions obtained by SPA-LDA for classification of the training (circle), and test (asterisk) samples for both (a) the entire spectral range and (b) the selected interval of 251–490 nm. (●) Argentinean green tea, (●) Brazilian green tea, (●) Argentinean black tea, (●) Brazilian black tea, and (●) Sri Lankan black tea.

Fisher's discriminant functions. SPA-LDA was therefore the most appropriate approach for simultaneous classification of the tea samples into the five differing (Argentinean green; Brazilian green; Argentinean black; Brazilian black; and Sri Lankan black) tea classes.

4. Conclusions

A simultaneous classification of both geographic origin and variety of teas using UV–Vis spectroscopy and pattern recognition techniques was proposed. In order to verify their differentiating characteristics, simple tea infusions prepared in boiling water alone (simulating a home-made tea cup) were analysed, this instead of extraction with methanol, as done by Pallacios-Morillo et al. (2013). Apart from its toxicity, the use of methanol as a solvent extractor makes the spectra much broader with highly correlated variables, requiring more sophisticated non-linear classifiers such as ANNs and SVMs. In our case, the use of water extraction, besides avoiding laborious sample preparation and additional operational costs, highlighted the analytical information contained in the spectra, allowing for the visualized discrimination tendencies between the black and green teas. Through this, the different pattern recognition methods applied successfully identified the differentiating characteristics of the tea samples for simultaneous geographic and varietal classification. SPA-LDA and PCA-LDA provided significantly better results for tea classification into the five classes (Argentinean green; Brazilian green; Argentinean black; Brazilian black; and Sri Lankan black). The proposed method therefore provides simpler, faster and more affordable classification of simple tea infusions, and can be used as an alternative approach to traditional tea quality evaluations as made by skilful tasters, which is evidently partial and cannot assess geographic origin. However, to guarantee any generalization of the proposed methodology, a larger and more varied testing of tea samples, using more varieties and geographic origins must be implemented.

Acknowledgements

The authors gratefully acknowledge the Universidad Nacional del Sur, CIC (Comisión de Investigaciones Científicas de la Provincia de Buenos Aires), and CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas) for their financial support. The authors also gratefully acknowledge the Capes and CNPq Brazilian scholarships and research fellowships.

References

Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: Linear models. PLS-DA. *Analytical Methods, 5*, 3790–3798.
Chow, H.-H. S., & Hakim, I. A. (2011). Pharmacokinetic and chemoprevention studies on tea in humans. *Pharmacological Research, 64*, 105–112.
Clary, J. J. (2013). *The toxicology of methanol* (1st ed.). Hoboken: Wiley.
Diniz, P. H. G. D., Dantas, H. V., Melo, K. D. T., Barbosa, M. F., Harding, D. P., Nascimento, E. C. L., et al. (2012). Using a simple digital camera and SPA-LDA modeling to screen teas. *Analytical Methods, 4*, 2648–2652.
Diniz, P. H. G. D., Gomes, A. A., Pistonesi, M. F., Band, B. S. F., & Araújo, M. C. U. (2014). Simultaneous classification of teas according their varieties and geographic origins by using NIR spectroscopy and SPA-LDA. *Food Analytical Methods, 7*, 1712–1718.
Domínguez, M. A., Diniz, P. H. G. D., Di Nezio, M. S., Araújo, M. C. U., & Centurión, M. E. (2014). Geographical origin classification of Argentinean honeys using a digital image-based flow-batch system. *Microchemical Journal, 112*, 104–108.
Fernández, P. L., Martín, M. J., González, A. G., & Pablos, F. (2000). HPLC determination of catechins and caffeine in tea. Differentiation of green, black and instant teas. *Analyst, 125*, 421–425.
Fernández, P. L., Pablos, F., Martín, M. J., & González, A. G. (2002). Study of catechin and xanthine tea profiles as geographical tracers. *Journal of Agricultural and Food Chemistry, 50*, 1833–1839.
Food and Agriculture Organization of the United Nations (FAOUN). Committee on commodity problems, intergovernmental group on tea. Current situation and medium term outlook for tea. (2012). URL <http://www.fao.org/fileadmin/templates/est/COMM_MARKETS_MONITORING/Tea/Documents/IGG_20/12-CRS7-CurrentSit_01.pdf> Acessed 23.04.14.
Gambarra-Neto, F. F., Marino, G., Araújo, M. C. U., Galvão, R. K. H., Pontes, M. J. C., Medeiros, E. P., et al. (2009). Classification of edible vegetable oils using square wave voltammetry with multivariate data analysis. *Talanta, 77*, 1660–1666.
Ghasemi-Varnamkhasti, M., Mohtasebi, S. S., Rodriguez-Mendez, M. L., Gomes, A. A., Araújo, M. C. U., & Galvão, R. K. H. (2012). Screening analysis of beer ageing using near infrared spectroscopy and the successive projections algorithm for variable selection. *Talanta, 89*, 286–291.
Kennard, R. W., & Stone, L. A. (1969). Computer aided design of experiments. *Technometrics, 11*, 137–148.
Khan, N., & Mukhtar, H. (2007). Tea polyphenols for health promotion. *Life Sciences, 81*, 519–533.
Kilinc, E. (2009). Significance of chromatographic and voltammetric data for the classification of green teas in Turkiye: A principle component analysis approach. *Journal of Liquid Chromatography and Related Technologies, 32*, 221–241.
Kumar, R. S. S., Murugesan, S., Kottur, G., & Gyamfl, D. (2012). Black tea: The plants, processing/manufacturing and production. In V. Preedy (Ed.), *Tea in health and disease prevention* (pp. 41–57). San Diego: Academic Press.
Obanda, M., Owuor, P. O., & Mang'oka, R. (2001). Changes in the chemical and sensory quality parameters of black tea due to variations of fermentation time and temperature. *Food Chemistry, 75*, 395–404.
Pallacios-Morillo, A., Alcázar, A., de Pablos, F., & Jurado, J. M. (2013). Differentiation of tea varieties using UV–Vis spectra and pattern recognition techniques. *Spectrochimica Acta A, 103*, 79–83.
Pinto, M. S. (2013). Tea: A new perspective on health benefits. *Food Research International, 53*, 558–567.
Sharangi, A. B. (2009). Medicinal and therapeutic potentialities of tea (*Camellia sinensis* L.) – A review. *Food Research International, 42*, 529–535.
Sharma, R. (2014). Polyphenols in health and disease: Practice and mechanisms of benefits. In R. R. Watson, V. R. Preedy, & S. Zibadi (Eds.), *Polyphenols in human health and disease* (pp. 757–778). San Diego: Academic Press.
Soares, S. F. C., Gomes, A. A., Galvão Filho, A. R., Araújo, M. C. U., & Galvão, R. K. H. (2013). The successive projections algorithm. *Trends in Analytical Chemistry, 42*, 84–98.
Souto, U. T. C. P., Pontes, M. J. C., Silva, E. C., Galvão, R. K. H., Araújo, M. C. U., Sanches, F. A. C., et al. (2010). UV–Vis spectrometric classification of coffees by SPA–LDA. *Food Chemistry, 119*, 368–371.
Spiller, G. A. (1997). *Caffeine* (1st ed.). Boca Raton: CRC Press.
Wang, X., Wan, X., Hu, S., & Pan, C. (2008). Study on the increase mechanism of the caffeine content during the fermentation of tea with microorganisms. *Food Chemistry, 107*, 1086–1091.
Yao, L. H., Jiang, Y. M., Caffin, N., D'Arcy, B., Datta, N., Liu, X., et al. (2006). Phenolic compounds in tea from Australian supermarkets. *Food Chemistry, 96*, 614–620.
Ye, N. S. (2012). A minireview of analytical methods for the geographical origin analysis of teas (*Camellia sinensis*). *Critical Reviews in Food Science and Nutrition, 52*, 775–780.