

Original Research Article

Simplified tea classification based on a reduced chemical composition profile via successive projections algorithm linear discriminant analysis (SPA-LDA)



Paulo Henrique Gonçalves Dias Diniz^a, Marcelo Fabián Pistonesi^b,
Mónica Beatriz Alvarez^b, Beatriz Susana Fernández Band^b,
Mário César Ugulino de Araújo^{c,*}

^a Universidade Estadual da Paraíba, Centro de Ciência e Tecnologia, Laboratório de Química Analítica e Quimiometria (LQAQ), Zip Code: 58429-500, Campina Grande, Paraíba, Brazil

^b Universidad Nacional del Sur, INQUISUR, Departamento de Química, Zip Code: 8000, Bahía Blanca, Buenos Aires, Argentina

^c Universidade Federal da Paraíba, Centro de Ciências Exatas e da Natureza, Departamento de Química, P.O. Box: 5093, Zip Code: 58051-970, João Pessoa, PB, Brazil

ARTICLE INFO

Article history:

Received 10 June 2013

Received in revised form 6 November 2014

Accepted 25 November 2014

Available online 4 January 2015

Chemical compounds studied in this article:

Caffeine (PubChem CID: 2519)

Gallic acid (PubChem CID: 370)

Fluoride (PubChem CID: 28179)

Potassium (PubChem CID: 5462222)

Aluminum (PubChem CID: 5359268)

Magnesium (PubChem CID: 5462224)

Keywords:

Camellia sinensis

Tea leaves

Tea infusions

Food composition

Food analysis

Elemental fingerprint

Classification

SPA-LDA feature selection

ABSTRACT

In this study, several possible approaches for simultaneous discrimination of teas based on a linear discriminant analysis with variables selected by the successive projections algorithm (SPA-LDA), featuring selection from the chemical composition according to variety (black or green tea) and geographical origin (Argentina or Sri Lanka), are explored. Chemical composition (moisture, ash, caffeine, fluoride, polyphenols, and 15 elements from both tea leaves and infusions) was used as input data for identification of the differentiating characteristics of tea samples. Thus, a strategy that allows tea discrimination using a reduced number of chemical parameters was developed. SIMCA (soft independent modeling of class analogy) and PLS-DA (partial least squares-discriminant analysis) were used along with SPA-LDA for comparison. The elemental fingerprint (chemical signature) can be used for identifying the variety and origin of the tea, and SPA-LDA provided the most successful result (100% correct classification), despite having selected just three chemical parameters (namely K, Al, and Mg). The result is extremely positive from the viewpoint of chemical analyses, because quantifications made using fewer elements naturally provide simpler, faster and less expensive methods.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Aside from water, tea is the most widely consumed beverage in the world. An infusion of the leaves of *Camellia sinensis*, it is usually divided into three basic types, or varieties; green tea (non-fermented), oolong tea (semi-fermented), and black tea (fully fermented). Its worldwide popularity is due to its unique aroma

and characteristic flavor. Tea also presents several bioactive compounds with important nutritional and therapeutic properties, and several other potential health benefits (Chaturvedula and Prakash, 2011; Chow and Hakim, 2011; Sharangi, 2009).

Due to the global importance of tea, the Food and Agricultural Organization of the United Nations has created an Intergovernmental Group for Teas. In a recent report, the 20th Session of this Committee, the current market situation and medium term outlook for teas (until 2021) were overviewed (FAO, 2012). Notably, Argentina has presented remarkable increases in tea production, placing it in the 9th position worldwide. Primary production

* Corresponding author. Tel.: +55 8332167438; fax: +55 8332167438.

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

(green shoot) has reached 383,000 tons. Industrial production is about 91,200 tons. Argentina exports about 95% of the total (86,000 tons), and imports 335 tons for blends (FAO, 2012).

In the worldwide tea trade, there is increasing consumer interest in geographical origin. The country's reputation or even that of a specific plantation may allow the producer to ask for significantly higher than average prices (Ye, 2012). For instance in Argentina, a Sri Lankan tea costs seven to ten times more than a national tea. Methods for analysis and verification of the origin of a tea provide security to both buyers and consumers, and prevent fraudulent labeling.

Several analytical techniques have been used for classification of varieties or geographical origin of teas: elemental analysis (Moreda-Piñeiro et al., 2003; Pilgrim et al., 2010), fluorescence (Seetohul et al., 2012; Ye, 2010), electronic tongue (He et al., 2009; Kovács et al., 2010), electronic nose (Chen et al., 2011; Chen et al., 2013), multi- or hyperspectral imaging (Zhao et al., 2009; Li et al., 2011), digital images (Diniz et al., 2012), laser-induced fluorescence (Mei et al., 2012), NIR spectroscopy (Zhou et al., 2009; Chen et al., 2009; Liu et al., 2010; Ren et al., 2013; Diniz et al., 2014), and UV–vis spectroscopy (Palacios-Morillo et al., 2013). In this context, only Diniz et al. (2012) proposed simultaneous discrimination of teas by variety (green or black), and by geographical origin (whether from Brazilian and Argentinean, or other). They used color histograms obtained from digital images of tea samples, and subsequent chemometric treatment using linear discriminant analysis (LDA), with variables selected by the successive projections algorithm (SPA) (Soares et al., 2013). However, this methodology cannot be applied to identify blended teas, because it uses simple digital imaging as input data, instead of chemical information. Using chemical composition as the input data for discrimination of tea samples is of great interest when identifying the variety (manufacturing process), and geographical origin, it allows the establishing the differentiating characteristics of tea products (Palacios-Morillo et al., 2013).

The Argentine Food Codex (AFC, 2013) recommends controlling the content of stems and petioles, moisture at 100–105 °C, total ash at 500–550 °C, HCl insoluble ash, ash (from the total ash) soluble in water, aqueous extraction, caffeine, tannins, and crude fiber. However, the procedures for determining these parameters are slow, laborious, and require several steps for sample treatment. Moreover, the determination results for these parameters do not furnish information directly related to tea origin. It would be interesting to propose a strategy that allows tea discrimination using a reduced number of chemical parameters, and develop in time to come a simple, fast, and inexpensive methodology. This could be done by algorithm, to find the most relevant analytical information for the problem under consideration, and then combining the variables that optimize performance of the analytical result. In this sense, SPA-LDA has also been successfully applied to classify other foods, such as coffees using UV–vis spectroscopy (Souto et al., 2010), vegetable oils using square wave voltammetry (Soares et al., 2010; Gambarra-Neto et al., 2009), beers using NIR spectroscopy (Ghasemi-Varnamkhasti et al., 2012), and honeys using digital images (Domínguez et al., 2014). In other words, SPA-LDA was used to select spectral variables in spectroscopy and potentials in voltammetry, which are continuous analytical signals.

In this study, we investigated for the first time the SPA-LDA algorithm for feature selection in discrete chemical data, in order to achieve a simplified tea classification based on a reduced chemical composition profile. Simultaneously with SPA-LDA, SIMCA (soft independent modeling for class analogy) and PLS-DA (partial least squares-discriminant analysis) were used to obtain a simultaneous discrimination of teas according to their variety (black or green tea), and geographical origin (Argentina or Sri Lanka). For this purpose, we selected and quantified five of the

main chemical parameter required by the Argentine regulation (AFC, 2013), namely moisture, total ash, caffeine, and tannins. However, since the determination of tannins identifies only a specific group of phenolic compounds, it is advisable to abandon the use of the term “tannin”, and replace it by the determination of total polyphenols due to their important contribution to flavor, aroma and color of tea infusions (Brignardello and Parodi, 2011).

Besides the determinations of the main chemical parameters described before, we also quantified the content of fluoride and several elements (Na, Mg, Al, P, K, Ca, Cr, Mn, Fe, Co, Ni, Cu, Zn, Cd, and Pb) in both leaves and infusions. These parameters proved to be important to estimate tea quality (Lu et al., 2004; Szymczycha-Madeja et al., 2012). Tea is a significant source of fluorine, because their leaves accumulate more fluoride due to the contamination of soil and air than other kinds of plants. Moreover, the fluoride absorbed from the soil forms F-Al complexes, transports Al to the leaves and, subsequently, to the catechins (the primary polyphenols in the tea) (Lu et al., 2004). The chemical signature (fingerprint) of the leaves and their infusions can be appropriate for a reliable differentiation because it is specific to a plant's geographical origin, and effectively presents no changes over relatively long periods of time (Szymczycha-Madeja et al., 2012).

2. Materials and methods

2.1. Samples and reagents

Sixty samples from 15 commercial brands of black and green teas were purchased at different supermarkets in Bahía Blanca, Argentina. Five brands of each type, whether green or black, were purchased to assure the natural variability of the samples.

All chemicals and reagents used in the determinations of contents of moisture, total ash, caffeine, fluoride, total polyphenols, and elements for both tea leaves and infusions were of analytical reagent grade. Double-distilled demineralized water was used throughout the work. All glass and plastic material was soaked in 10% HNO₃ (Carlo Erba Reagents, Milan, Italy) for 24 h. The washing was then completed with several rinses with double-distilled water.

2.2. Instrumentation

A Shimadzu ICPS-1000 (Shimadzu Corporation, Tokyo, Japan) sequential optical emission spectrometer with inductively coupled plasma (ICP-OES) was used for the determination of the elemental content in leaves and infusions of black and green tea samples. The operation of RF power was 1.2 kW. An observation height of 12 mm for the measurements was selected, with the following flow rates of argon: plasma 14 L/min; auxiliary 1.2 L/min; and nebulizer 1.0 L/min. The integration time was 5 s. Stock standards (1000 mg/L) for all the analytes were obtained from Cica-Merck Standards (Cica-Merck, Tokyo, Japan). Calibration solutions of 15 elements (Na, Mg, Al, P, K, Ca, Cr, Mn, Fe, Co, Ni, Cu, Zn, Cd, and Pb) were prepared in 0.1 mol/L HNO₃ (Carlo Erba Reagents, Milan, Italy) by successive dilutions from standards. Determination limits for the spectrochemical procedures were estimated from solutions containing the matrix components, *i.e.* simulating the composition of the solutions. Practical values were calculated as three times the IUPAC detection limit, defined as the ratio between three times the standard deviation of the blank signal (for $n = 10$), and the slope of the calibration curve. Table 1 presents the wavelengths selected for the determination of the elements and their respective determination limits.

Fluoride determinations in the tea infusions were made by using a fluoride selective electrode Cole-Parmer 27502-19 (Cole-Parmer, Vernon Hills, IL, USA) and a Potentiometer Orion Model 710A (Orion Research Inc, Boston, MA, USA).

Table 1

Wavelengths selected for the determination of elements by ICP-OES and their respective limits of determination.

Element	Wavelength (nm)	Limit of determination (µg/L)
Al	396.153	20
Ca	422.673	0.5
Cd	226.502	5
Co	228.616	5
Cr	267.716	5
Cu	327.396	5
Fe	259.940	5
K	766.491	100
Mg	279.553	0.5
Mn	257.610	1
Na	588.995	50
Ni	231.604	5
P	178.287	50
Pb	220.351	50
Zn	213.856	5

Determination of the total polyphenols content in the tea infusions was performed in a UV–vis spectrophotometer GBC Cintra 20 (GBC Scientific, Braeside, Australia).

The univariate calibration strategy was used for the determination of fluoride, total polyphenols, and elemental contents in both tea leaves digests and freshly prepared infusions.

2.3. Analytical procedures

2.3.1. Determinations in tea leaves

2.3.1.1. Moisture. For moisture determination, 5 g of each tea sample was weighed and placed in an oven at 103 ± 2 °C in order to achieve a constant weight. The percent loss in weight is reported as moisture content.

2.3.1.2. Total ash. Determination of mineral content by ashing consisted of calcination of 5 g of each sample in a muffle furnace, with a ramp up of temperature increasing 50 °C every 30 min until reaching 550 °C, maintaining this temperature for a period of 24 h. This approach guarantees the removal of all organic material.

2.3.1.3. Caffeine. The caffeine content was determined by the modified Cortés method (AFC, 2013). For this purpose, 1 g of tea sample was placed in a 100 mL Erlenmeyer, which was added 4 mL of concentrated H_2SO_4 (Mallinckrodt Pharmaceuticals, Dublin, Ireland) and heated in a water bath for 15 min in order to destroy all organic matter and form the caffeine sulfate. Then 50 mL of boiling distilled water were added and kept in a water bath for 20 min. The hot suspension was filtered and the solution containing caffeine was collected in a 150 mL vial. The solution was cooled at room temperature for 10 min and alkalized with NaOH 40% w/v (Mallinckrodt Pharmaceuticals, Dublin, Ireland) until a pH of approximately 8.0. After these steps, three extractions using 25 mL of chloroform (Mallinckrodt Pharmaceuticals, Dublin, Ireland) each one are performed. The organic phase of each extract containing caffeine was collected in a tared distillation flask. The distillation flask was then placed in a rotoevaporator to reduce the volume of organic solution and following placed in a drying oven at 103 ± 2 °C to eliminate the solvent residue to constant weight. Thus, white crystals of caffeine is obtained and quantified gravimetrically.

2.3.1.4. Total content of elements. The determination of total content of elements (Na, Mg, Al, P, K, Ca, Cr, Mn, Fe, Co, Ni, Cu, Zn, Cd, and Pb) was made by calcination of 5 g of each tea sample in a muffle furnace using a temperature ramp with increments of

50 °C every 30 min until reaching 450 °C, during 24 h. The obtained product was slowly dissolved in 10 mL of 6 mol/L hydrochloric acid (Carlo Erba Reagents, Milan, Italy), and heated on a hot-plate until almost dryness. The residue was then dissolved in 0.1 mol/L nitric acid, and completed to 25 mL. The solution was placed in Nalgene plastic bottles, and kept refrigerated at 4.0 °C until the determination of elements by ICP-OES. A filtration step through a Millipore 0.45 µm filter was carried out in order to prevent ICP nebulizer blocking. Sample blanks were prepared in an identical way, but omitting the sample. Dilution of samples was done by the addition of double-distilled water. The calibration procedure is described in Section 2.

2.3.2. Determination in tea infusion

2.3.2.1. Infusion preparation. The infusions were prepared using 2 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 and retention (8 µm) filter paper, and completed to 100 mL with double-distilled water. They were kept in Nalgene plastic bottles in refrigerator at 4.0 °C.

2.3.2.2. Elements in the infusion. Determination of elements (Na, Mg, Al, P, K, Ca, Cr, Mn, Fe, Co, Ni, Cu, Zn, Cd, and Pb) in the freshly prepared infusions was carried out by ICP-OES, using the calibration procedure described in Section 2.

2.3.2.3. Fluoride. Fluoride determinations in the tea infusions were made by using a fluoride selective electrode. A fluoride standard solution of 1000 mg/L was employed. A total ionic strength adjustment buffer (TISAB) solution was prepared by dissolving 22.05 g of dried sodium citrate (Mallinckrodt Pharmaceuticals, Dublin, Ireland) and 0.80 g of sodium hydroxide (Mallinckrodt Pharmaceuticals, Dublin, Ireland) with 100 mL double-distilled water. The pH was adjusted to 5.3 with 0.1 mol/L of HClO_4 (Mallinckrodt Pharmaceuticals, Dublin, Ireland). In order to keep the ionic strength and pH constant, a 1 mol/L sodium nitrate (Mallinckrodt Pharmaceuticals, Dublin, Ireland) solution was employed.

2.3.2.4. Total polyphenols. Total polyphenol content was determined by Folin-Ciocalteu method, which is based on the transfer of electrons in alkaline medium from phenolic compounds to phosphomolybdic/phosphotungstic acid complexes (with yellow coloring). The phenolic compounds contained in the sample are oxidized in alkaline medium, resulting in the formation of O^{2-} , which reacts with acids forming a complex with an intense blue color absorbance maximum at 760 nm. The measurements were performed in a UV–vis spectrophotometer GBC Cintra 20. Total polyphenols was expressed as gallic acid (Mallinckrodt Pharmaceuticals, Dublin, Ireland) equivalent, which was used as the calibration standard.

2.4. Chemometric procedure for data analysis

Tea samples were divided into training (50%), validation (25%), and test (25%) sets by applying the Kennard–Stone (KS) uniform sampling algorithm (Kennard and Stone, 1969). Differing methods of pattern recognition were evaluated: (a) SIMCA, (b) PLS-DA, and (c) SPA-LDA. The variables chosen by SPA were those corresponding to the smallest cost function G value. The modeling procedures were accomplished using the training and validation samples (including the variable selection for the SPA-LDA modeling, and the determination of the principal components for each SIMCA modeling). The test samples were only used for the final data evaluation and comparison of the classification models (Soares et al., 2013).

PCA, SIMCA and PLS-DA calculations were carried out using The Unscrambler® 9.7 (CAMO S/A) software. The KS algorithm and SPA-LDA models were performed with Matlab® 2009b (Mathworks Inc.) software. All calculations were performed at 95% confidence level.

3. Results and discussion

Table 2 shows the means and their respective standard deviations of the measurements of moisture, total ash, caffeine, fluoride, total polyphenols, and elemental contents for both tea leaves and infusions in the three studied tea classes (see Table S1 in Appendix A, in online Supplementary Material, for accessing the precision of the 15 studied commercial brands of black and green teas). The studied tea samples were spiked with a known amount of analytes (50.0 mg/kg) in both tea leaves digests and freshly prepared infusions in order to perform recovery tests, which were in the range from 87.1 to 112% for tea leaves and from 85.2 to 113% for tea infusions (see Table S2 in Appendix A, in online Supplementary Material). No significant matrix effects and/or inter-element interferences were observed at the used concentration levels.

Among the studied parameters, the Argentinean Food Codex (AFC, 2013) establishes limits only for moisture, total ash and

caffeine. This regulation establishes that both moisture and total ash contents must not exceed 7 and 8% w/w on dry basis of tea, respectively. The caffeine content must be at least 1.6% w/w on dry basis of tea. It is worth to highlight that the Argentine regulation sets up a minimum of 12% w/w tannins on dry basis of tea. However, since the determination of tannins identifies only a specific group of phenolic compounds, it was replaced by the determination of total polyphenols (Brignardello and Parodi, 2011).

In relation to the moisture content, we can observe that the Argentinean black tea samples presented values higher than both the Sri Lankan black teas and Argentinean green teas. It is known that the transfer of water into these systems is uncertain, since the material for packaging (cardboard, paper and aluminum) have their own absorption isotherms, which can be combined in a pseudo-equilibrium with the tea isotherm. Thus, this difference can be due to the fact that Argentinean black teas are commercialized without involucre, whereas the other studied teas are sold in paper or aluminum sachets. Moreover, bad storage conditions are a worsening factor to the moisture retention in teas. So it is important to note that virtually all samples had moisture content above 7% (w/w) on dry basis, recommended by Argentine regulation.

For the total ash determination it was verified that all analyzed samples presented values under the maximal established by the

Table 2

Results of the determinations of moisture, ash, caffeine, fluoride, polyphenols, and several elements in both tea leaves and infusions for the three studied tea classes: Argentine black tea; Sri Lankan black tea; and Argentine green tea.

	Argentine black tea			Sri Lankan black tea			Argentine green tea		
	Range (mg/kg)	Mean value (mg/kg)	SD (mg/kg)	Range (mg/kg)	Mean value (mg/kg)	SD (mg/kg)	Range (mg/kg)	Mean value (mg/kg)	SD (mg/kg)
Moisture	7.50–11.9	9.1	0.8	7.00–8.20	7.6	0.3	5.80–8.30	7.0	0.6
Ash	6.00–7.40	6.7	0.3	5.00–6.40	6.0	0.2	5.40–7.80	6.1	0.5
Caffeine	1.20–2.60	1.8	0.3	1.40–3.00	2.1	0.3	0.80–2.30	1.4	0.3
Fluoride	73.9–183	110	19	95.2–666	416	151	90–410	168	82
Polyphenols	6.56–9.60	7.8	0.7	7.78–15.1	11.0	1.9	7.94–14.9	10	1.8
Elements in tea leaves									
Na	21.9–46.2	34	6	32.3–75.8	55.9	12	35.7–71.9	53.6	12
Mg	1094–1633	1397	89	735–1492	1047	219	85.3–1397	856	299
Al	1111–3166	1642	344	291–1231	693	211	440–2169	1178	524
P	866–2355	1952	214	1679–3349	2383	357	1384–2964	2009	413
K	8279–17,202	13,282	1262	11,492–17,335	14,882	1299	5540–21,676	13,248	4189
Ca	2732–6387	4450	453	2576–4710	3590	513	2785–5179	4082	539
Cr	0.66–3.13	1.8	0.5	0.25–3.65	1.8	1.0	0.77–2.58	1.2	0.3
Mn	537–1192	927	141	64.8–1164	383	321	196–1346	1017	193
Fe	40.9–182	94	25	57.9–289	132	57	58.9–210	106	38
Co	0.05–0.24	0.16	0.03	0.03–0.24	0.12	0.07	0.09–0.24	0.19	0.04
Ni	1.65–11.0	4.0	0.9	2.71–5.57	4.0	0.6	2.13–8.43	3.8	0.8
Cu	0.58–26.7	13	2.5	11.3–23.3	17	3.8	6.15–27.6	13	4.4
Zn	4.57–26.6	19	3.1	14.4–34.9	24	5.8	12.3–30.2	19	4.9
Cd	0.04–0.39	0.22	0.06	0.10–0.29	0.18	0.04	0.10–0.40	0.23	0.08
Pb	0.86–6.48	2.8	0.7	1.11–3.28	1.7	0.5	0.17–8.16	2.4	1.8
Elements in tea infusions									
Na	14.2–20.3	18	1.5	15.0–21.0	18	1.4	2.58–23.4	15	3.8
Mg	34.3–177	118	24	12.3–182	127	21	45.2–153	84	33
Al	50.9–134	97	13	13.9–82	46	16	19.4–87.0	63	17
P	39.3–126	95	17	110–166	136	14	39.5–93.0	64	16
K	994–2532	1979	246	19.8–2440	1858	284	721–1578	1121	223
Ca	24.4–49.3	41		15.8–37.6	29	6.3	23.5–56.5	36	5.0
Cr	0.01–0.21	0.11		0.01–0.09	0.03	0.03	0.01–0.06	0.02	0.02
Mn	15.4–66.6	39	11	4.28–71.7	22	16	19.3–69.0	37	12
Fe	0.16–1.32	0.40	0.24	0.19–0.38	0.26	0.05	0.04–0.62	0.42	0.12
Co	0.01–0.20	0.05	0.03	0.01–0.10	0.06	0.02	0.01–0.09	0.04	0.03
Ni	0.25–0.54	0.43	0.06	0.21–0.65	0.41	0.13	0.18–0.58	0.38	0.06
Cu	0.35–0.60	0.52	0.04	0.36–4.48	0.79	0.48	0.09–2.84	0.37	0.26
Zn	0.31–0.82	0.61	0.09	0.69–1.99	1.3	0.30	0.45–1.07	0.73	0.16
Cd ^a	–	–	–	–	–	–	–	–	–
Pb	0.01–0.67	0.35	0.18	0.01–0.66	0.22	0.22	0.01–0.46	0.17	0.17

^a Below the limit of detection.

Argentinean regulation, which is 8% w/w on dry basis. The other studied samples were in accordance with the permitted values. Since the process of fermentation changes the chemical composition of teas, the caffeine content in green tea leaves (processed before they begin to ferment) is generally lower than that of fully fermented black teas.

Despite the mentioned above, the aim of this work is to achieve a simplified tea classification based on a reduced chemical composition profile, finding the best analytical approach for a simultaneous discrimination of teas according to their variety (black, or green tea), and geographical origin (Argentina or Asia). The classification results for four differing data sets were then evaluated. The first data set (“All parameters”) was constituted of the determination results for moisture, total ash, caffeine, fluoride, total polyphenols, and several elements in both tea leaves and infusions. The second data set (“AFC parameters”) was formed by the determination results of moisture, total ash, caffeine, total polyphenols, and fluoride. Since fluoride content is an important chemical parameter that can be used as a quality indicator for tea evaluation (Lu et al., 2004), we placed these results in the second data set. The third data set (“Total content of elements”) was constituted of the determination results of Na, Mg, Al, P, K, Ca, Cr, Mn, Fe, Co, Ni, Cu, Zn, Cd, and Pb content in the tea leaves alone. Finally, the last data set (“Elements in tea infusions”) was composed by the determination results of the elements mentioned before in tea infusions alone.

3.1. Principal component analysis

A screening analysis of all samples was developed by using PCA. Fig. 1 presents the score plots and their respective loadings for the two first principal components employing the determination results for (a) “All parameters”, (b) “AFC parameters”, (c) “Total content of elements”, and (d) “Elements in tea infusions” data sets.

As can be seen, Fig. 1a and c were very similar. In fact, analyzing the PCA loadings we see that the elemental content in the tea leaves (mainly K, Ca, Al, Mg, Mn, and P) was responsible for the variance described in the two first PCs of both data sets. Fig. 1b and d exhibit an apparent tendency of separation between the three studied classes. Loadings for “Elements in tea infusions” presented the same metals (Al, Mn, Mg, Ca, P, and K) as being the most important for the two first PCs, which corroborates with the fact that these metals are highly extractable from leaves to infusions. Loadings for the “AFC parameters” had total polyphenols and

fluoride contents as the most relevant for the two first PCs. However, some overlapping was observed in all cases, which motivated the use of supervised classification techniques, as reported below.

3.2. Classification

Table 3 summarizes the classification results in the test set by using SIMCA, PLS-DA, and SPA-LDA at 95% confidence level. The test set was composed of 6 Argentine black, 4 imported black, and 5 Argentine green tea samples, totaling 15 samples. The classification was carried out by evaluating the discrimination of green and black teas as either national or imported black teas.

3.2.1. SIMCA

Table 3 shows the optimal PC number employed for modeling of each class and the two types of classification errors: (a) a type-I error, which consists of a sample not included in its own class; and (b) a type-II error, which consists of a sample included in an incorrect class. We can observe that all samples were correctly classified into their respective class for all used data sets, except for the Argentinean green teas when using the “AFC parameters” that had a classification error of 20%. In relation to the type II errors, the best SIMCA result was achieved by using the “Elements in tea infusions” data set, which reached a mean correct classification of 88.9%. The other “All parameters”, “AFC parameters”, and “Total content of elements” data sets presented mean classification errors of 25, 61.1 and 71.7%, respectively.

3.2.2. PLS-DA

In all cases, the PLS-DA results presented fewer misclassified samples than the respective SIMCA results. As can be also observed in Tab. 3, no errors were found for both “All parameters” and “Elements in tea infusions” data sets, whereas the other two data sets achieved correct classification at 86.7%. Low R^2 values for calibration and cross-validation were obtained for all data sets, despite “Elements in tea infusions” presenting reasonable R^2 values of 0.9131 and 0.8325 for calibration and cross-validation, respectively.

3.2.3. SPA-LDA classification results

The study involving SPA-LDA was aimed at investigating whether good classification results could be achieved by using a minimal subset of selected chemical parameters, instead of the

Table 3

Summary of the classification results in the test set for SIMCA, PLS-DA, and SPA-LDA at a 95% confidence level.

	All parameters			AFC parameters			Total content of elements			Elements in tea infusions		
	ArB	SrB	ArG	ArB	SrB	ArG	ArB	SrB	ArG	ArB	SrB	ArG
SIMCA												
N_{PC}	6	5	2	1	1	1	5	4	2	1	2	2
Type-I errors (%)	0	0	0	0	0	20	0	0	0	0	0	0
Type-II errors (%)	50	25	0	83.3	0	100	100	75	40	33.3	0	0
PLS-DA												
R^2 (calibration)		0.6678			0.7033			0.7031			0.9131	
RMSEC		0.2823			0.2668			0.2669			0.1444	
R^2 (cross-validation)		0.8082			0.8168			0.6581			0.8325	
RMSECV		0.2145			0.2097			0.2865			0.2005	
N_{LV}		9			4			8			12	
Classification error (%)	0	0	0	0	0	40	0	0	40	0	0	0
SPA-LDA												
Cost function		0.3615			0.3547			0.3350			0.3114	
N_{var}		9			3			9			3	
Classification error (%)	0	0	0	16.7	0	20	0	0	0	0	0	0

ArB: Argentinean black tea; SrB: Sri Lankan black tea; ArG: Argentinean green tea. N_{PC} : optimal number of PC; N_{LV} : optimal number of latent variables; N_{var} : number of selected variables; R^2 : coefficient of determination; RMSEC: root-mean-square error of calibration; RMSECV: root-mean-square error of cross-validation.

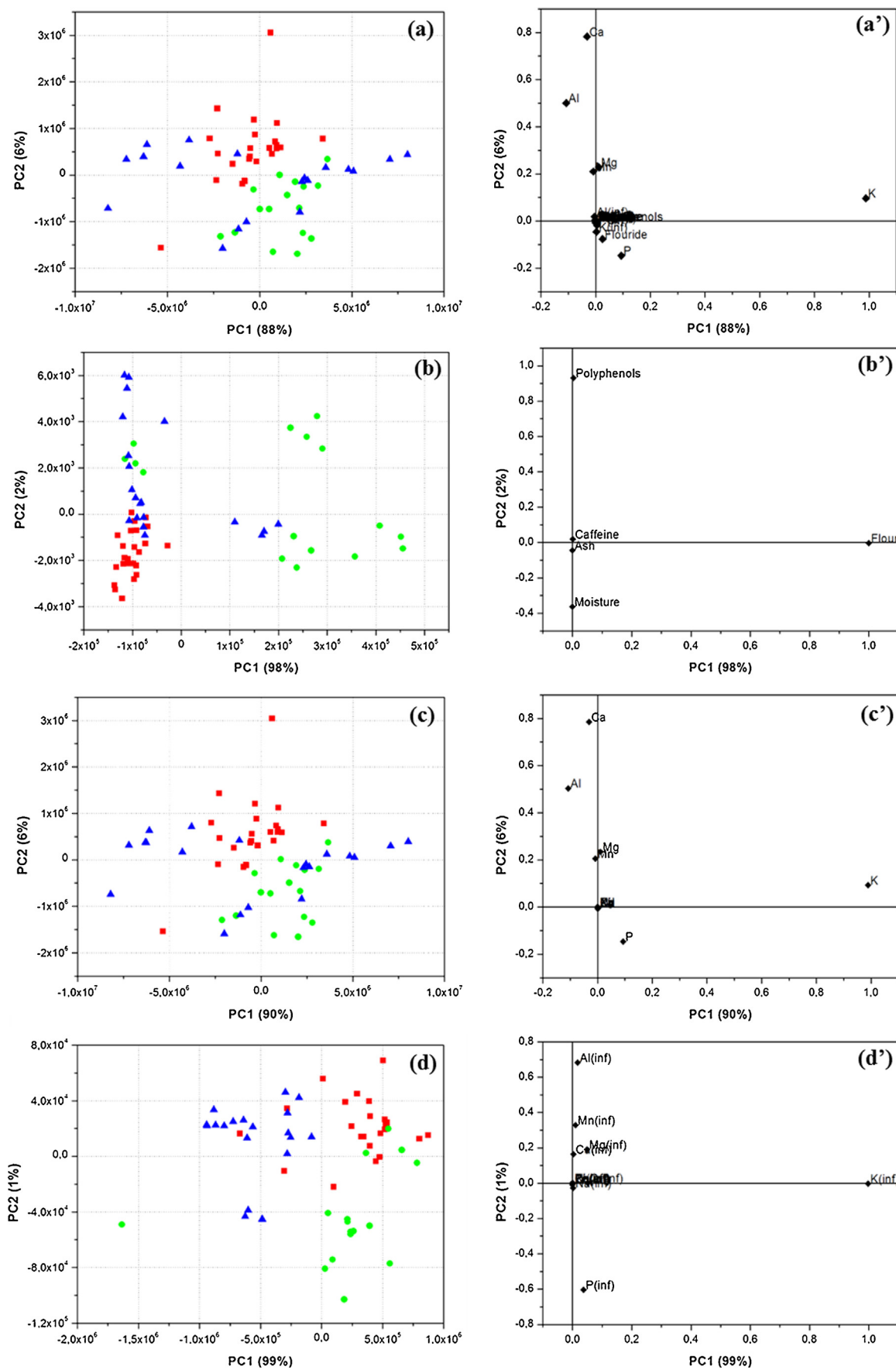


Fig. 1. PCA score (first column) plots and their respective loadings (second column) for the first PCs employing the determinations results for the following data sets: (a) all parameters, (b) ASACFF parameters, (c) total content of elements, and (d) elements in the infusion. Argentinean black tea (■), Sri Lankan black tea (●), and Argentinean green tea (▲).

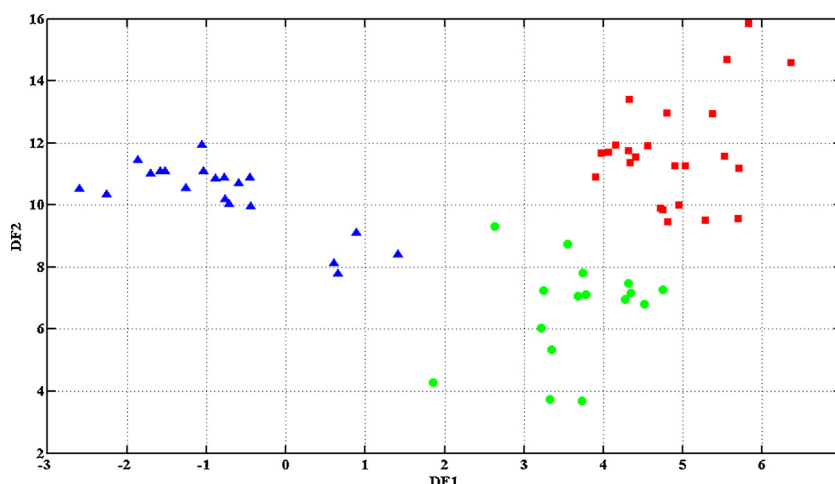


Fig. 2. Fisher's discriminant score plot obtained by using the parameters (K, Al and Mg) selected by SPA-LDA for elements in the infusion. Argentinean black tea (■), Sri Lankan black tea (●), and Argentinean green tea (▲).

entire variable set. The selected parameters by SPA for each model are presented. (a) "All parameters": moisture, and K, Ca, P, Mn, Al, Mg, Fe, and Na in the leaves. (b) "AFC parameters": caffeine, fluoride, and polyphenols. (c) "Total content of elements": K, Ca, P, Mn, Al, Mg, Fe, Na, and Co. (d) "Elements in tea infusions": K, Al, and Mg.

In relation to the classification errors, the SPA-LDA results were better than SIMCA, and similar to PLS-DA. The "AFC parameters" model presented a correct classification of 86.7% when using SPA-LDA. The other models correctly classified all the samples in the test set, which indicates that the parameters selected by SPA-LDA convey enough information for tea discrimination. Despite the good performance of all supervised classification techniques using the "Elements in tea infusions" data set, SPA-LDA provided the more significant result, a feature selection based on only three chemical parameters. Fig. 2 shows the Fisher's discriminant score plot obtained by using the parameters (K, Al and Mg) selected by SPA-LDA for "Elements in tea infusions". As can be seen, the separation between the classes is more apparent than those of the respective PCA score plot presented in Fig. 1d. Moreover, three of the most relevant parameters indicated in the loading plots in Fig. 1d were selected by SPA, which demonstrates its suitability for feature selection.

This is extremely positive from the viewpoint of chemical analyses, because quantifying only three major elements (namely K, Al, and Mg) in a simple tea infusion naturally results in a simpler, faster and less expensive methodology. Indeed, the chemical signature (fingerprint) of tea leaves, and tea infusions is appropriate for reliable differentiation, because it is not only specific for the given tea plants when grown in a specific geographical region, but also presents essentially no changes for relatively long periods of time (Szymczycha-Madeja et al., 2012). From a biological perspective, we note that potassium is an essential macronutrient for plant growth and magnesium is found in the plant chlorophyll porphyrins involved in photosynthesis. On other hand, *Camellia sinensis* is one of a few plants to accumulate aluminum, which suggests further study since ingestion of aluminum can precipitate health problems due to both neuro- and osteotoxicity.

4. Conclusions

In this study, chemical composition (moisture, ash, caffeine, fluoride, polyphenols, and 15 metals from both tea leaves and infusions) was used as input data for identification of the

differentiating characteristics of tea samples. Undoubtedly, the best classification results were obtained when the elemental profile in the tea infusions was used. Some approaches provided 100% of discrimination: (1) PLS-DA using "All parameters"; (2) PLS-DA using "Elements in tea infusions"; (3) SPA-LDA using "All parameters"; (4) SPA-LDA using "Total content of elements"; and (5) SPA-LDA using "Elements in tea infusions". Despite of that, SPA-LDA using "Elements in tea infusions" provided the most important result, selection of the three telling chemical parameters (namely potassium, aluminum, and magnesium), corroborating that the elemental fingerprint (chemical signature) can be used for simultaneous discrimination of both varieties and geographical origin of teas. The result is extremely positive from the view point of chemical analyses, because the quantification of only three major elements in simple tea infusions provides a simpler, faster and more affordable methodology for tea discrimination. However, to guarantee any generalization of the proposed methodology, a larger (more varied) testing of tea samples using more varieties and geographical 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 gratefully acknowledge Capes and CNPq Brazil scholarships and research fellowships. The authors also thank Eugenio Campaña for his collaboration in the experimental of this work, and also David Harding for reviewing the English in this article.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jfca.2014.11.012>.

References

- Argentinean Food Codex (AFC), 2013. *Productos estimulantes o frutivos*. Cap. XV Article 1185 (Res. 1542, 17.9.85).
- Brignardello, A.E., Parodi, N.B., 2011. Calidades químicas de té negro de Misiones-Argentina y de un producto derivado (bt) destinados a exportación. *Revista de Ciencia y Tecnología* 15, 30–37.

- Chaturvedula, V.S.P., Prakash, I., 2011. The aroma, taste, color and bioactive constituents of tea. *Journal of Medicinal Plants Research* 5, 2110–2124.
- Chen, Q., Liu, A., Zhao, J., Ouyang, Q., 2013. Classification of tea category using a portable electronic nose based on an odor imaging sensor array. *Journal of Pharmaceutical and Biomedical Analysis* 84, 77–83. <http://dx.doi.org/10.1016/j.jpba.2013.05.046>.
- Chen, Q., Zhao, J., Chen, Z., Lin, H., Zhao, D.-A., 2011. Discrimination of green tea quality using the electronic nose technique and the human panel test, comparison of linear and nonlinear classification tools. *Sensors and Actuators B: Chemical* 159, 294–300. <http://dx.doi.org/10.1016/j.snb.2011.07.009>.
- Chen, Q., Zhao, J., Lin, H., 2009. Study on discrimination of Roast green tea (*Camellia sinensis* L.) according to geographical origin by FT-NIR spectroscopy and supervised pattern recognition. *Spectrochimica Acta A* 72, 845–850.
- Chow, H.-H.S., Hakim, I.A., 2011. Pharmacokinetic and chemoprevention studies on tea in humans. *Pharmacological Research* 64, 105–112.
- Diniz, P.H.G.D., Dantas, H.V., Melo, K.D.T., Barbosa, M.F., Harding, D.P., Nascimento, E.C.L., Pistonesi, M.F., Band, B.S.F., Araújo, M.C.U., 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 geographical 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.
- FAO (Food and Agriculture Organization of the United Nations), 2012. Committee on Commodity Problems, Intergovernmental Group on Tea. Current Situation and Medium Term Outlook for Tea. Retrieved from: http://www.fao.org/fileadmin/templates/est/COMM_MARKETS_MONITORING/Tea/Documents/IGG_20/12-CRS7-CurrentSit_01.pdf (accessed 31.05.13).
- Gambarra-Neto, F.F., Marino, G., Araújo, M.C.U., Galvão, R.K.H., Pontes, M.J.C., Medeiros, E.P., Lima, R.S., 2009. Classification of edible vegetable oils using square wave voltammetry with multivariate data analysis. *Talanta* 77, 1660–1666.
- Ghasemi-Varnamkhasti, M., Mohtasebi, S.S., Rodríguez-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.
- He, W., Hu, X., Zhao, L., Liao, X., Zhang, Y., Zhang, M., Wu, J., 2009. Evaluation of Chinese tea by the electronic tongue: correlation with sensory properties and classification according to geographical origin and grade level. *Food Research International* 42, 1462–1467.
- Kennard, R.W., Stone, L.A., 1969. Computer aided design of experiments. *Technometrics* 11, 137–148.
- Kovács, Z., Dalmadi, I., Lukács, L., Sipos, L., Szántai-Kóhegyi, K., Kókai, Z., Fekete, A., 2010. Geographical origin identification of pure Sri Lanka tea infusions with electronic nose, electronic tongue and sensory profile analysis. *Journal of Chemometrics* 24, 121–130.
- Li, X., Nie, P., Qiu, Z.-J., He, Y., 2011. Using wavelet transform and multi-class least square support vector machine in multi-spectral imaging classification of Chinese famous tea. *Expert Systems with Applications* 38, 11149–11159.
- Liu, S.L., Tsai, Y.S., Ou, S.-M., 2010. Classifying the variety, production area and season of Taiwan partially fermented tea by near infrared spectroscopy. *Journal of Food and Drug Analysis* 18, 34–43.
- Lu, Y., Guo, W.-F., Yang, X.-Q., 2004. Fluoride content in tea and its relationship with tea quality. *Journal of Agricultural and Food Chemistry* 52, 4472–4476.
- Mei, L., Lundin, P., Brydegaard, M., Gong, S., Tang, D., Somesfalean, G., He, S., Svanberg, S., 2012. Tea classification and quality assessment using laser-induced fluorescence and chemometric evaluation. *Applied Optics* 51, 803–811.
- Moreda-Piñeiro, A., Fisher, A., Hill, S.J., 2003. The classification of tea according to region of origin using pattern recognition techniques and trace metal data. *Journal of Food Composition and Analysis* 16, 195–211.
- Palacios-Morillo, A., Alcázar, A., Pablos, F., Jurado, J.M., 2013. Differentiation of tea varieties using UV-vis spectra and pattern recognition techniques. *Spectrochimica Acta A* 103, 79–83.
- Pilgrim, T.S., Watling, R.J., Grice, K., 2010. Application of trace element and stable isotope signatures to determine the provenance of tea (*Camellia sinensis*) samples. *Food Chemistry* 118, 921–926.
- Ren, G., Wang, S., Ning, J., Xu, R., Wang, Y., Xing, Z., Wan, X., Zhang, Z., 2013. Quantitative analysis and geographical traceability of black tea using Fourier transform near-infrared spectroscopy (FT-NIRS). *Food Research International* 53, 822–826.
- Seetohul, L.N., Scott, S.M., O'Hare, W.T., Ali, Z., Islam, M., 2012. Discrimination of Sri Lankan black teas using fluorescence spectroscopy and linear discriminant analysis. *Journal of the Science of Food and Agriculture* 93, 2308–2314.
- Sharangi, A.B., 2009. Medicinal and therapeutic potentialities of tea (*Camellia sinensis* L.) – a review. *Food Research International* 42, 529–535.
- Soares, A.S., Galvão, R.K.H., Araújo, M.C.U., Soares, S.F.C., Pinto, L.A., 2010. Multi-core computation in chemometrics: case studies of voltammetric and NIR spectrometric analyses. *Journal of the Brazilian Chemical Society* 21, 1626–1634.
- 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., Cunha, F.A.S., Oliveira, M.S.R., 2010. UV-vis spectrometric classification of coffees by SPA-LDA. *Food Chemistry* 119, 368–371.
- Szymczycha-Madeja, A.S., Welna, M., Pohl, P., 2012. Elemental analysis of teas and their infusions by spectrometric methods. *Trends in Analytical Chemistry* 35, 165–181.
- Ye, N., 2010. Geographical classification of green teas based on MEKC with laser-induced fluorescence detection. *Chromatographia* 71, 529–532.
- 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.
- Zhao, J., Chen, Q., Cai, J., Ouyang, Q., 2009. Automated tea quality classification by hyperspectral imaging. *Applied Optics* 48, 3557–3564.
- Zhou, J., Cheng, H., He, W., Wang, L., Liu, X., Lu, W., 2009. Identification of geographical indication tea with Fisher's discriminant classification and principal components analysis. *Journal of Near Infrared Spectroscopy* 17, 159–164.