Analytical Methods

Dynamic Article Links

Cite this: *Anal. Methods*, 2012, **4**, 2648

www.rsc.org/methods

COMMUNICATION

Using a simple digital camera and SPA-LDA modeling to screen teas

Paulo H. G. D. Diniz,[a] Hebertty V. Dantas,[a] Karla D. T. Melo,[a] Mayara F. Barbosa,[a] David P. Harding,[a] Elaine C. L. Nascimento,[a] Marcelo F. Pistonesi,[b] Beatriz S. F. Band[b] and Mário C. U. Araújo†*[a]

Received 11th May 2012, Accepted 6th July 2012
DOI: 10.1039/c2ay25481f

Classification or screening analysis of natural unprocessed teas using simple digital images and a variable selection algorithm is described. The proposed methodology uses color histograms generated on free downloadable software *ImageJ 1.44p* as a source of analytical information. Two chemometric methods were compared for classification of the resulting images, namely Soft Independent Modeling of Class Analogy (SIMCA), and Linear Discriminant Analysis (LDA) with variable selection by the Successive Projections Algorithm (SPA). The results were evaluated in terms of errors found in a sample set separate from the modeling process. The choice of more informative photometric color attributes (red-green-blue (RGB), hue (H), saturation (S), brightness (B), and grayscale) for screening the tea samples was made during the color modeling because SIMCA failed to give good results. Therefore the data treatment used SPA-LDA, which correctly classified all samples according to their geographical regions, whether from Brazilian, Argentinian or foreign soils.

1. Introduction

After water, teas from the leaves of *Camellia Sinensis* are the most widely consumed beverages in the world.[1] Various factors such as altitude, climate, soil, *etc.* affect the plant which leads to differing leaf textures, physical appearance, and chemical properties. Depending on the manufacturing process, teas are classified into three major types: non-fermented green tea, semi-fermented oolong tea, and fermented black tea.[2,3] Tea has a wide variety of compounds such as flavonoids, phenolic acids, amino acids, chlorophyll, pigments, carbohydrates, organic acid, caffeine and other alkaloids, minerals, vitamins and enzymes.[1,4] Unlike black and oolong teas, green tea is not fermented, and more of the active constituents remain in the leaves. In terms of consumption, black teas are the most consumed around the world, and their quality is judged mainly on the basis of their components and color.

[a]*Universidade Federal da Paraíba, CCEN, Departamento de Química, João Pessoa, PB, Brazil. E-mail: laqa@quimica.ufpb.br; Fax: +55 83 3216 7437; Tel: +55 83 3216 7438*

[b]*Universidad Nacional del Sur, Departamento de Química, Bahía Blanca, Argentina*

† Present address: Department of Chemistry, CCEN, Federal University of Paraiba, Caixa Postal 5093, CEP 58051-970, Joao Pessoa, Brazil.

India and China are first in world tea production, being favoured by both climatic and topographic characteristics.[5,6] Argentinian production is around 70 000 tons per annum, 80% of which is sold in foreign markets. The province of *Misiones* in Argentina accounts for 90% of the country's output. Only 10% of the cultivated area is sown with seeds of national origin, the rest is planted with seeds from China and India[7] which is a source of considerable variation in the quality of the crops. Though they lack flavour, Brazilian teas are ideal for blending, and the majority of Brazilian tea is produced for this purpose. About 70% of the total tea production is sold to the United States. During the 1970s, the annual tea production was about 11 000 tons, yet there has been a steady decline ever since. In recent years, tea producers have been focusing their efforts on increasing the quality of Brazilian teas and have seen an increase in market share.

Quality control of tea depends mainly on its appearance, taste, and aroma. Traditionally, human sensory assessment by skillful tasters is employed for evaluating tea quality. However, the results of this type of assessment are not precise and can be markedly partial.[8] The "*Digital Image*", on the other hand, is an analytical tool with great potential for use in qualitative and quantitative analysis.[9–13] This is one of the emerging frontiers of advanced research, and deals with the process of capturing, conditioning, and measurement of the Digital Camera based *digital image* by using advanced soft-computing algorithms that extract important information and features from the acquired images.[14–16] The technique is non-invasive, facilitates the acquisition of data, and decreases costs. It also can reproduce a human evaluation in a standardized, impartial, and robust way.[17]

Many methodologies have been proposed to determine the physical/chemical properties of foods, using chemometrics, pattern recognition, and image analysis techniques. Applications for digital image processing techniques are expanding rapidly in the food processing industries.[18,19] Multi- and hyperspectral imaging, artificial olfactory devices, and spectrometric techniques that use wavelet transform, neural networks, and support vector machines have been developed but are expensive, and overly complex.[18,20–31]

Non-destructive classification is one of the most important applications for image analysis. Many studies have used Principal Component Analysis (PCA) to compress the information, and the derived latent variables are applied later with a classifier (*e.g.* Fisher's Linear Discriminant Analysis).[32–34] From a database made up of useful classification variables, and applying Soft Independent Modelling of Class Analogy (SIMCA), and Partial Least Squares-Discriminant Analysis (PLS-DA), the dimension of the data structure is reduced and the classification accuracy rate achieved by simpler classification techniques is improved upon.[35,36] Linear Discriminant Analysis (LDA) as a classifier, with variable selection by the Successive Projections Algorithm (SPA)[37] has been successfully used for this purpose in different classification problems, including the analysis of edible vegetable oils,[38,39] soils,[40] cigarettes,[41] coffees,[42] diesel/biodiesels,[43] alcohol fuels,[44] beers[45] and pen inks.[46]

SIMCA is a well-known multivariate pattern recognition method which discriminates between different classes of samples,[47–51] given various classes, classification rules are defined by a set of samples from known classes (the training set), together with distinct measurement values from each sample, and a sample validation set is included. The rules are then used to classify the new unknown samples (test set) on the basis of similar measurements.[52] LDA classification methods employ linear decision boundaries (hyperplanes), which are defined in order to maximize the ratio of inter-class to intra-class dispersion.[53] In order to have a well-posed problem, the number of calibration (training) objects must be larger than the number of variables included in the LDA model. The use of LDA for classification of spectral data usually requires an appropriate variable selection procedure.[37,41,54,55] The Successive Projections Algorithm (SPA) was adapted[37] for use in classification problems. In the original formulation,[56] the candidate variable subsets are formed as a result of projection operations intended to minimize multi-collinearity effects, which are known to cause poor performance in LDA.[39,57]

This study aimed to classify or screen non-processed green and black teas purchased in Brazil and Argentina using simple images captured on a digital camera. The methodology used chemo-metrically evaluated color histograms generated on free software *ImageJ 1.44p*.[58] The chemometric evaluation utilized SPA in combination with LDA. The study determined a relationship between the digital image and geographical origin of the tea, this being useful for investigating possible sample adulterations. It could provide an additional tool for primary tea producers to secure both quality and a differentiated product.

2. Experimental

2.1. Samples and image acquisition

One hundred samples of green and black teas were purchased from local supermarkets; 40 Brazilian samples (20 each green and black), and 40 Argentinian samples (20 each green and black), and 20 imported black samples. For all of the above, five brands of each type whether green or black were purchased to assure variability or uniform distribution of the model.

The method of imaging takes into consideration the overall visual feature of the sample surface. All samples were homogenized, and photographed without treatment using a digital camera, Olympus® model SP-510 UZ, with 7.1 Megapixels. To achieve standardization, manufacturers resolve questions of color balance red-green-blue (RGB), hue (H), saturation (S), and brightness (B), as well as the grayscale intensity in ways particular for each camera model sold. These manufacturer's considerations do not affect the overall method, but rather the individual scores for each value; RGB, HSB, grayscale. A total of 300 images (3 for each sample) were obtained. To ensure reproducibility, a mechanical digital camera support and a circular fluorescent lamp were mounted to maintain fixed positioning, luminosity, sample-to-camera distance, and focus. Fig. 1 shows the apparatus built for image capturing, and actual examples of the five groups (Argentinian green and black, Brazilian green and black, and imported black teas) in line to be photographed. The digital camera was set above the tea vertically. The distance between the lens and the tea was 12 cm, and between the illumination and the tea was 10 cm. The digital camera was placed in the center of a circular fluorescent lamp. The sample holder used in this study was composed of polytetrafluoroethylene, which minimizes light scattering and fluorescence effects, and their effects on image color histograms. Therewith, the proposed methodology dispensed with the need for further image manipulation.

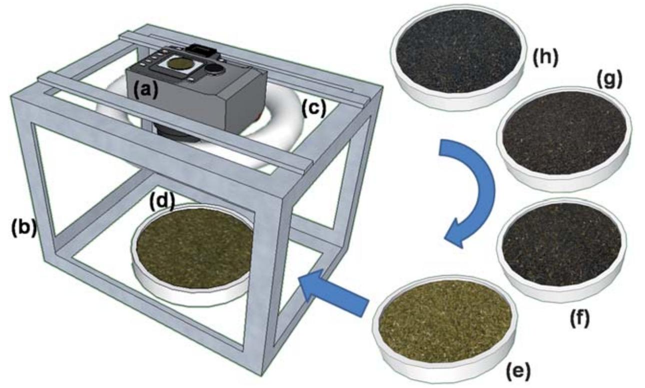

Fig. 1 Apparatus built for tea image capturing and examples of green and black tea samples. (a) Digital camera, (b) metallic support, (c) circular lamp, (d) Brazilian green, (e) Argentinian green, (f) Brazilian black, (g) Argentinian black, and (h) Imported black tea.

2.2. Histograms and data analysis

Color histograms describe the statistical distribution of the pixels as a function of the recorded color component, and not a physical–chemical behavior directly. The digital images were treated using free downloadable software *ImageJ 1.44p*, which created a histogram for each; red, green, blue, hue, saturation, brightness, and grayscale intensity. Fig. 2 shows an example of each histogram type (and its corresponding image), for a single green tea sample.

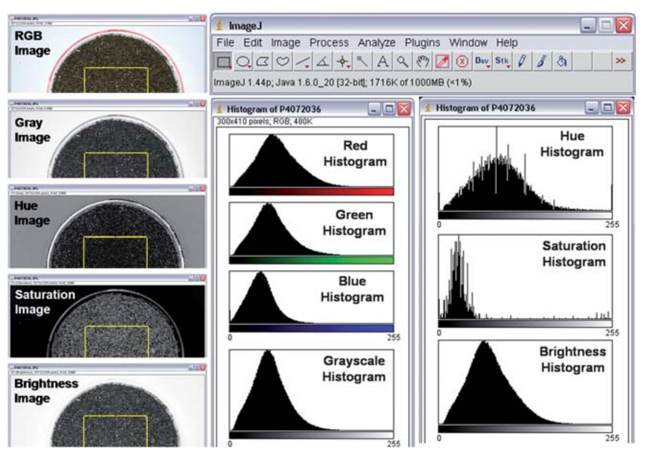

Fig. 2 The *ImageJ* software and the corresponding histograms and images for a green tea sample.

validation and prediction samples were initially selected in accordance with the Kennard–Stone results.[37,59–61] Chemometric data treatment was implemented with The Unscrambler® 9.7 (CAMO S/A), and Matlab® 2009b (Mathworks Inc.) software.

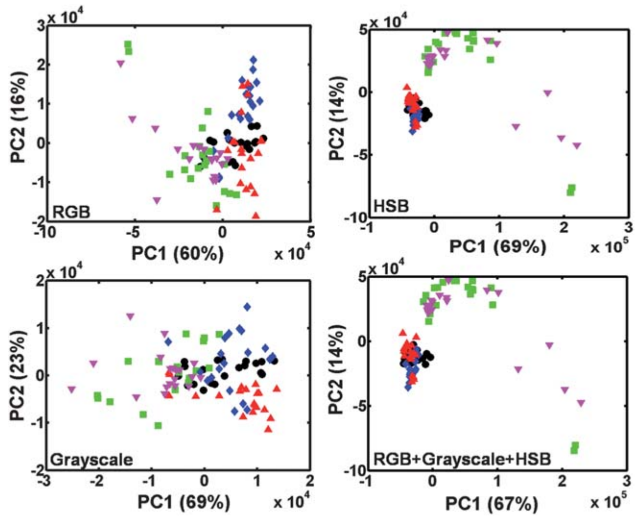

Fig. 3 PCA scores obtained from RGB, HSB, grayscale and RGB + grayscale + HSB histograms for all tea samples. Brazilian green tea (▼), Brazilian black tea (▲), Argentinian green tea (■), Argentinian black tea (◆) and Imported black tea (●). The variance for each principal component is indicated in parenthesis.

For data treatment we defined a 1000 × 1000 pixel square region at the center of each image, which represented about 14% of the total image area. Using only the selected region of the images, histograms employing red, green, blue, hue, saturation, brightness, and grayscale were constructed. Each color component of the models is composed of 256 tones, which are used as analytical information. In order to check for each color's relative influence, we selected four different color models employing (a) RGB, (b) HSB, (c) grayscale, (d) RGB + grayscale + HSB. The models were composed of (a) 3 × 256, (b) 3 × 256, (c) 1 × 256 and (d) 768 + 256 + 768 variables, respectively. From three photos for each sample a mean histogram was calculated.

The analytical information extracted from the histograms above was employed to construct the chemometric classification models using SIMCA and SPA-LDA. The data obtained from each histogram were separated into: training (60%), validation (20%) and prediction (20%) sets using the Kennard–Stone algorithm.[59] The Kennard–Stone (KS) algorithm was applied separately to each class. It is a classic method to extract a representative set of objects from a given dataset by maximizing the minimal Euclidean distance between already selected objects and the remaining objects. The calibration,

3. Results and discussion

3.1. Exploratory analysis

In Fig. 1, we can observe that it is difficult to distinguish between Fig. 1(d) and (e), in the case of green teas, and between Fig. 1(f)–(h) for black teas. Since there is variation in the colors of individual tea particles, the images are characterized by a natural stochastic image texture, *i.e.* they allow us to extract information about the spatial arrangement of color and intensities. Color histograms describe the statistical distribution of the pixels as a function of the recorded color component, and not physical–chemical properties.

An exploratory data analysis was made using principal component analysis employing the three color models RGB, HSB, and grayscale, and the one model using all the histograms together (RGB + grayscale + HSB), as described before. Fig. 3 shows the plotted scores of the two first principal components using these four colors models.

The (RGB) and (grayscale) plots present overlap between the samples, and very little separation. Plots (HSB) and (RGB + grayscale + HSB) present similar results which separate the samples into two major clusters of black and green tea, yet with some remaining overlap. This was expected since green and black teas are very different in color, surface texture, particle size, *etc.* The three black teas overlap more, while the two green teas are better resolved. This occurs because there is repetition of the image information, which requires variable selection to discover which tones of the histogram lead to a better classification model. A variable selection algorithm (SPA-LDA) was used to resolve this and consequentially to classify the tea samples. The SPA-LDA results were subsequently compared with the SIMCA classification results, as discussed in the next section.

3.2. Classification

Table 1 shows the SPA-LDA and SIMCA results with cost function values, for both green and black tea samples according to the geographical region for the four histogram classification models. Green teas were classified into Brazilian and Argentinian clusters; and black teas into Brazilian, Argentinian, and Imported clusters. Each class was composed of 20 tea histograms, which were separated into

Table 1 SIMCA and SPA-LDA classification results for green and black tea samples according to their geographical origin using color histograms in classification models[a]

	SPA-LDA						SIMCA[b]				
		No. of errors[c] (%)					No. of errors[c] (%)				
		Green		Black			Green		Black		
Histograms used in the classification models	Selected variables/cost function value	Br	Ar	Br	Ar	Im	Br	Ar	Br	Ar	Im
RGB	20/0.74728	0	37.5	50.0	37.5	12.5	100	87.5	87.5	100	87.5
HSB	16/0.54008	0	0	12.5	37.5	0	100	100	100	87.5	87.5
Grayscale	12/0.81383	50.0	50.0	12.5	12.5	12.5	100	75.0	100	87.5	87.5
RGB + grayscale + HSB	17/0.54907	0	0	0	0	0	100	75.0	100	87.5	87.5

[a] Br = Brazilian tea; Ar = Argentinean tea; Im = Imported tea. [b] 95% confidence level. [c] Validation + test errors.

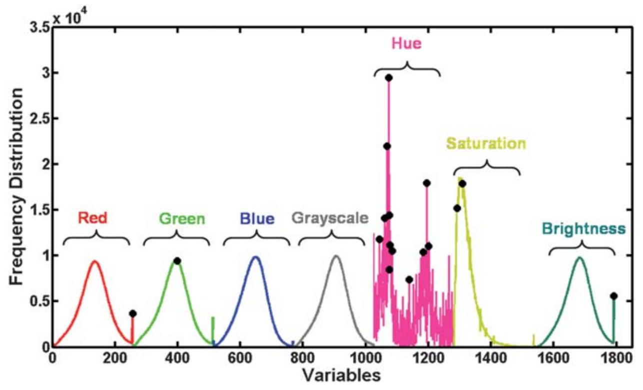

Fig. 4 The 17 variables selected by SPA-LDA (●) from the red, green, blue, grayscale, hue, saturation, and brightness histograms.

training (12), validation (4) and prediction (4) sets. The number of classification errors is presented as validation + prediction errors; their respective percentages are also shown. The SIMCA models were calculated with a 95% confidence level.

For the SIMCA models, all training, validation, and prediction samples separated by the Kennard–Stone algorithm were correctly classified into their respective class, but they are to a large degree classified into other classes as well. This can be ascribed to the high overlap between the classes, reaching errors of above 75%. SPA-LDA improved meaningfully the geographical discrimination ability of the classification models. Fig. 4 shows the 17 variables selected by SPA-LDA from the red, green, blue, grayscale, hue, saturation, and brightness histograms, which correspond to less than one percent of all the variables. These 17 variables (color tones, hue, saturation, and brightness) serve as the basis for a more intelligent discrimination between class characteristics. Although a similar separation pattern into two major classes (black and green teas) had been obtained using principal component analysis (see Fig. 3), the SIMCA model uses all of the information contained in the databases and its predictive ability becomes limited by superfluous information.

Fig. 5 shows plots of the four resulting SPA-LDA discriminant functions (DF), corresponding to the linear combination of the selected variables. We observed good discrimination between the green teas, into Brazilian and Argentinian clusters, and with respect to the black teas, the *Imported* tea class was separated from the Brazilian and Argentinian black tea clusters by the superior DF's, 3–4.

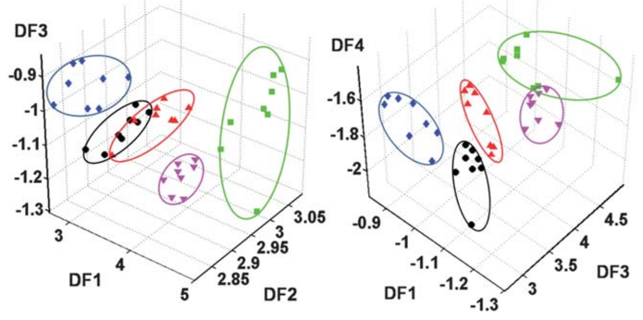

Fig. 5 Discriminant functions plot (DF1-4) obtained employing the SPA-LDA selected variables for Brazilian green tea (▼), Brazilian black tea (▲), Argentinian green tea (■), Argentinian black tea (◆), and Imported black tea (●).

4. Conclusions

This paper demonstrates the use of analytical information extracted from digital image generated color histograms. An appropriate variable selection algorithm (SPA), in combination with LDA, served as an analytical tool for discriminating natural green and black teas according to their geographical regions, whether from Brazilian and Argentinian, or foreign soils. The value of this method is that it approximates the expert/master's eye for discerning the origin of the tea being examined, yet remains impartial. The importance of sample positioning, lighting uniformity, and focal distance is worth mentioning, with these resolved, further image manipulation was unnecessary. The choice of an appropriate combination of histograms, and subsequent treatment using SPA-LDA reduced by about 99% the number of variables. The proposed method presents inherent advantages and requires no prior handling of the samples, no reagents, and generates no waste.

Acknowledgements

The authors gratefully acknowledge the Brazilian agencies CNPq and CAPES, and the Argentinian agencies CONICET and CIC for scholarships.

References

1 C. J. Dufresne and E. R. Farnworth, *J. Nutr. Biochem.*, 2001, **12**, 404–421.
2 M. Bancirova, *Food Res. Int.*, 2010, **43**, 1379–1382.
3 L. F. Wang, D. M. Kim and C. Y. Lee, *J. Agric. Food Chem.*, 2000, **48**, 4227–4232.
4 Y. Lu, W. F. Guo and Q. X. Yang, *J. Agric. Food Chem.*, 2004, **52**, 4472–4476.
5 M. A. Delavar, L. M. Sann, K. G. Lin, S. Tajuddin, B. S. Hassan and P. Hanachi, *Res. J. Int. Stud.*, 2008, **8**, 117–124.
6 D. L. McKay and J. B. Blumberg, *J. Am. Coll. Nutr.*, 2002, **21**, 1–13.
7 C. Anesini, G. E. Ferraro and R. Filip, *J. Agric. Food Chem.*, 2008, **56**, 9225–9229.
8 J. Zhao, Q. Chen, J. Cai and Q. Quyang, *Appl. Opt.*, 2009, **19**, 3557–3564.
9 A. Lopez-Molinero, D. Liñan, D. Sipiera and R. Falcon, *Microchem. J.*, 2010, **96**, 380–385.
10 W. S. Lyra, F. A. C. Sanches, F. A. S. Cunha, P. H. G. D. Diniz, S. G. Lemos, E. C. Silva and M. C. U. Araújo, *Anal. Methods*, 2011, **3**, 1975.
11 S. I. E. Andrade, M. B. Lima, I. S. Barreto, W. S. Lyra, L. F. Almeida, M. C. U. Araújo and E. C. Silva, *Microchem. J.*, 2012, DOI: 10.1016/j.microc.2012.03.029.
12 E. N. Gaiao, V. L. Martins, W. S. Lyra, L. F. Almeida, E. C. Silva and M. C. U. Araújo, *Anal. Chim. Acta*, 2006, **570**, 283–290.
13 A. C. Sousa, A. S. Soares, C. J. Coelho, E. N. Gaião, A. F. C. Pereira, R. K. H. Galvão and M. C. U. Araújo, *Anal. Methods*, 2012, DOI: 10.1039/C2AY055377F.
14 T. Oikawa, M. Tomizawa and S. Degawa, *Control Eng. Pract.*, 1997, **5**, 75–78.
15 R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, Boston, MA, 3rd edn, 1992.
16 R. Berstein, *IBM J. Res. Dev.*, 1976, **20**, 40–56.
17 G. S. Gill, A. Kumar and R. Agarwal, *J. Food Eng.*, 2011, **106**, 13–19.
18 M. S. Zenoozian and S. Devahastin, *J. Food Eng.*, 2009, **2**, 219–227.
19 A. R. Tôrres, W. S. Lyra, S. I. E. Andrade, R. A. N. Andrade, E. C. Silva, M. C. U. Araújo and E. N. Gaião, *Talanta*, 2011, **84**, 601–606.
20 N. Bhattacharyya, S. Seth, B. Tudu, P. Tamuly, A. Jana, D. Ggosh, R. Bandyopadhyay, M. Bhuyan and S. Sabhapandit, *Sens. Actuators, B*, 2007, **122**, 627–634.
21 S. Borah and M. Bhuyan, *Insight J.*, 2003, **35**, 55–58.
22 P. K. Mahanta and M. Hazarika, *J. Sci. Food Agric.*, 1985, **36**, 1133–1139.

23 M. Pietikäinen, T. Ojala and Z. Xu, *Pattern Recognit.*, 2000, **33**, 43–52.
24 X. Li, P. Nie, Z. J. Qiu and Y. He, *Expert Syst. Appl.*, 2011, **38**, 11149–11159.
25 Q. S. Chen, J. W. Zhao and J. R. Cai, *Trans. Am. Soc. Agric. Biol. Eng.*, 2008, **51**, 615–621.
26 Q. S. Chen, J. W. Zhao, C. H. Fang and D. M. Wang, *Spectrochim. Acta, Part A*, 2007, **66**, 568–574.
27 S. Borah, E. L. Hines and M. Bhuyan, *J. Food Eng.*, 2007, **79**, 629–639.
28 Q. S. Chen, J. W. Zhao, H. D. Zhang, M. H. Liu and M. Fang, *J. Near Infrared Spectrosc.*, 2005, **13**, 327–332.
29 D. Wu, X. J. Chen and Y. He, *Spectrosc. Spectral Anal.*, 2009, **29**, 1382–1385.
30 X. J. Chen, D. Wu, Y. He, X. L. Li and S. Liu, *Spectrosc. Spectral Anal.*, 2008, **28**, 2527–2530.
31 D. Wu, H. Q. Yang, X. J. Chen, Y. He and X. L. Li, *J. Food Eng.*, 2008, **88**, 474–483.
32 H. Yu and J. Yang, *Pattern Recognit.*, 2001, **34**, 2067–2070.
33 A. Moreda-Piñeiro, A. Fisher and S. J. Hill, *J. Food Compos. Anal.*, 2003, **16**, 195–211.
34 J. E. Lee, B. J. Lee, J. A. Hwang, K. S. Ko, J. O. Chung, E. H. Kim, S. J. Lee and Y. S. Hong, *J. Agric. Food Chem.*, 2011, **59**, 10579–10585.
35 J. M. Prats-Montalbán, A. Juan and A. Ferrer, *Chemom. Intell. Lab. Syst.*, 2011, **107**, 1–23.
36 F. López, J. M. Valiente, J. M. Prats and A. Ferrer, *Pattern Recognit.*, 2008, **41**, 1161–1172.
37 M. C. U. Araújo and R. K. H. Galvão, Linear regression modeling: variable selection, in *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis (Four-Volume Set)*, ed. B. Walczak, R. T. Ferré and S. Brown, Elsevier, Amsterdan, The Netherlands, 1st edn, 2009, vol. 03, pp. 233–283.
38 M. J. C. Pontes, R. K. H. Galvão, M. C. U. Araújo, P. N. T. Moreira, O. D. Pessoa Neto, G. E. José and T. C. B. Saldanha, *Chemom. Intell. Lab. Syst.*, 2005, **78**, 11–18.
39 F. F. Gambarra-Neto, G. Marino, M. C. U. Araújo, R. K. H. Galvão, M. J. C. Pontes, E. P. Medeiros and R. S. Lima, *Talanta*, 2009, **77**, 1660–1666.
40 M. J. C. Pontes, J. Cortez, R. K. H. Galvão, C. Pasquini, M. C. U. Araújo, R. M. Coelho, M. K. Chiba, M. F. Abreu and B. E. Madari, *Anal. Chim. Acta*, 2009, **642**, 12–18.
41 E. D. Moreira, M. J. C. Pontes, R. K. H. Galvão and M. C. U. Araújo, *Talanta*, 2009, **79**, 1260–1264.
42 U. T. C. P. Souto, M. J. C. Pontes, E. C. Silva, R. K. H. Galvão, M. C. U. Araújo, F. A. C. Sanches, F. A. S. Cunha and M. S. R. Oliveira, *Food Chem.*, 2010, **119**, 368–371.
43 M. J. C. Pontes, C. F. Pereira, M. F. Pimentel, F. V. C. Vasconcelos and A. G. B. Silva, *Talanta*, 2011, **85**, 2159–2165.
44 A. C. Silva, L. F. B. L. Pontes, M. F. Pimentel and M. J. C. Pontes, *Talanta*, 2012, **93**, 129–134.
45 M. Ghasemi-Varnamkhastia, S. S. Mohtasebic, M. L. Rodriguez-Mendeza, A. A. Gomes, M. C. U. Araújo and R. K. H. Galvão, *Talanta*, 2012, **89**, 286–291.
46 C. S. Silva, F. S. L. Borba, M. F. Pimentel, M. J. C. Pontes, R. S. Honorato and C. Pasquini, *Microchem. J.*, 2012, DOI: 10.1016/j.microc.2012.03.025.
47 S. Wold and M. Sjostrom, *Chemometrics: Theory and Application*, ed. B. R. Kowaslki, American Chemical Society, Washington, 1973, p. 243.
48 M. D. Derde and D. L. Massart, *Anal. Chim. Acta*, 1986, **191**, 181–192.
49 M. Sjostrom and B. R. Kowalski, *Anal. Chim. Acta*, 1979, **112**, 11–30.
50 R. De Maesschalck, A. Candolfi, D. L. Massart and S. Heuerding, *Chemom. Intell. Lab. Syst.*, 1999, **47**, 65–77.
51 S. Wold, K. Esbensen and P. Geladi, *Chemom. Intell. Lab. Syst.*, 1987, **2**, 37–52.
52 F. S. Oliveira, L. S. G. Teixeira, M. C. U. Araujo and M. Korn, *Fuel*, 2004, **83**, 917–923.
53 R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*, John Wiley, New York, 2nd edn, 2001.
54 A. R. Caneca, M. F. Pimentel, R. K. H. Galvão, C. E. Matta, F. R. Carvalho, I. M. Raimundo Jr, C. Pasquini and J. J. R. Rohwedder, *Talanta*, 2006, **70**, 344–352.
55 Y. Mallet, D. Coomans and O. de Vel, *Chemom. Intell. Lab. Syst.*, 1996, **35**, 157–173.
56 M. C. U. Araújo, T. C. B. Saldanha, R. K. H. Galvão, T. Yoneyama, H. C. Chame and V. Visani, *Chemom. Intell. Lab. Syst.*, 2001, **57**, 65–73.
57 T. Naes and B. H. Mevik, *J. Chemom.*, 2001, **15**, 413–426.
58 M. D. Abramoff, P. J. Magalhaes and S. J. Ram, *Biophotonics Int.*, 2004, **11**, 36–42.
59 R. W. Kennard and L. A. Stone, *Technometrics*, 1969, **11**, 137–148.
60 E. Bouveresse, C. Hartmann, D. L. Massart, I. R. Last and K. A. Prebble, *Anal. Chem.*, 1996, **68**, 982–990.
61 K. R. Kanduc, J. Zupan and N. Majcen, *Chemom. Intell. Lab. Syst.*, 2003, **65**, 221–229.