Microchemical Journal 112 (2014) 104–108

Contents lists available at ScienceDirect

Microchemical Journal

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

Geographical origin classification of Argentinean honeys using a digital image-based flow-batch system

Marina Alejandra Domínguez [a], Paulo Henrique Gonçalves Dias Diniz [b], María Susana Di Nezio [a], Mário César Ugulino de Araújo [b], María Eugenia Centurión [a,*]

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

ARTICLE INFO

Article history:
Received 12 August 2013
Received in revised form 13 September 2013
Accepted 13 September 2013
Available online 1 October 2013

Keywords:
Honey
Classification
Geographical origin
Flow-batch
Digital image

ABSTRACT

In this paper a geographic origin classification of honey samples from Argentina was proposed. For this reason, a conventional flow-batch system with a simple webcam to capture digital images was employed. In this methodology, analytical information is generated from color histograms obtained from the digital images employing different color models (RGB (red–green–blue), HSB (hue–saturation–brightness) and Grayscale). Three chemometric tools were employed for geographic origin classification (SPA-LDA (successive projections algorithm-linear discriminant analysis), SIMCA (soft independent modeling for class analogy), and PLS-DA (partial least squares-discriminant analysis)). The proposed method is a good option to be used in quality control laboratories for the classification of honey samples according to their geographical origin.

1. Introduction

During the last decade, Argentina was positioned as one of the first honey producers in the world, being simultaneously the principal exporter of bulk honey. Due to the diversity of climates, Argentina has a wide variety of floral resources that allows these activities to be carried out in most of the territory. However, the province of Buenos Aires concentrates more than 50% of national production. The southwest of the province of Buenos Aires is located in the south of the Pampean region and presents different natural regions: mountain range, valleys, plains and forest. These environments are related to the climatic conditions and vegetal species valuable for beekeeping. The organoleptic characteristics of honey are strongly associated with the botanical origin, depending on geographical area in which bees collect pollen. Then, the honey produced in Argentina is internationally appreciated by physicochemical, microbiological and sensory (flavor and color) attributes and it is positioned at the top of the global preferences [1–3].

Currently, the honey market tends to classify them according to their geographical or botanical origin. Baroni et al. [4] evaluate the floral origin of honey determining its volatile organic compounds by using headspace solid-phase microextraction and gas chromatography coupled to mass spectrometry and chemometric techniques. On the other hand, the classification of floral origin of honeys was carried out by chemical and physical characteristics combined with chemometrics [5]. The determinations of 14 trace elements applying Instrumental Neutronic Activation Analysis (INAA) allowed the differentiation of multifloral honeys from Argentine Pampas in combination with chemometric techniques [6]. Escriche et al. show a method to differentiate two types of citrus honey that comes from two different botanical origins by determining the content of flavonoids, phenolic acids and volatile compounds and statistical data evaluation techniques [7]. Classification of honeys according to their botanical and geographical origin by using only seven elements and chemometric techniques was carried out [8]. Cometto et al. determined free amino acid composition of honey samples by reversed-phase high-performance liquid chromatography and statistical analyses [9]. Honeys produced in the regions of the province of Córdoba (Argentina), were classificated by geographical origin using chemometrics and chemical properties and mineral profile [10].

Some of these methods have been developed based on palynological studies, sensory characterization, mineral analysis, flavonoid contents, etc. The palynological studies have the disadvantages of being slow, tedious and require trained analysts [11]. Within sensory characterization, color determination is relevant. The international reference techniques for defining honey color are based on Pfund method which is based on optical comparison. These color units are obtained using subjective visual assessment that can lead to differences in the reported values between users and does not distinguish between small variations of color. On the other hand, the determination of this parameter is laborious, time consuming and requires large amounts of sample [12]. Moreover, some determinations involve time-consuming and laborious procedures, for example to evaluate mineral honey composition a muffle furnace at a temperature of 550 °C over night should be used [13].

* Corresponding author.
E-mail address: mecentur@criba.edu.ar (M.E. Centurión).

0026-265X/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.microc.2013.09.008

In some cases, are necessary purification steps, for example to determine flavonoids as reported by Bogdanov et al. [14].

In order to overcome these drawbacks, digital images as a potential tool for qualitative and quantitative analysis, have been used in the recent years [15–18]. Color digital image processing involves the use of color models in order to provide specification of colors by using some standard. In essence, a color model is a specification of a three-dimensional coordinate system and a subspace within the system where each color component is represented by a single point.

There are many models used to measure and describe color, including RGB (red–green–blue), HSB (hue–saturation–brightness) and Grayscale. The RGB color model is based on the theory that all visible colors can be created using the primary additive colors red, green and blue. These colors are known as primary additives because when combined other colors are produced. The HSB model defines a color space in terms of three constituent components: Hue (color type), Saturation (intensity of color) and Brightness (brightness of color). Finally, grayscale is a range of shades of gray without apparent color [19–21].

In order to obtain the digital images, it would be appropriate to implement an automatic system that includes a web cam. Among the automatic methods, Flow–Batch methodology (FB) is a good option since it combines the intrinsic favorable features of the flow and batch techniques. Therefore, they can be considered as a multipurpose analytical accessory. These systems are characterized by the use of a chamber and threeway solenoid valves fully computer-controlled. One of the advantages of the FB is that it is considered as a universal purpose accessory tool easily attached to any conventional equipment for instrumental analysis. Flow–Batch systems allow high sampling frequencies, low cost per analysis, less consumption of reagent and sample, and less chemical waste than classical methods, principles considered in green analytical chemistry [22–25].

The aim of this work was to carry out a digital image-based flow-batch system for geographic origin classification of honey samples from southwest of the province of Buenos Aires, Argentina. For this purpose, a Flow–Batch system which includes a simple webcam to capture digital images was employed. Data treatment involved the use of a suitable variable selection technique, the successive projections algorithm (SPA) [26], associated to the linear discriminant analysis (LDA) to improve the classification results. Alongside of SPA-LDA, SIMCA (soft independent modeling for class analogy), and PLS-DA (partial least squares-discriminant analysis) were used for comparison.

2. Material and methods

2.1. Samples preparation

A set of 210 representative honey samples from southwest of the province of Buenos Aires, Argentina, were collected from supermarkets and local producers. Samples from three differing geographical origins were selected: 75 from the mountain region called the Sierra de la Ventana, 70 from the north and 65 from the south of this mountain region. In Fig. 1, the samples that were collected can be seen in the three regions. Samples were kept in a cool dry place until analysis. For image acquisition, a 50% (w/v) solution of each honey sample was prepared in ultra pure water (18 M Ω Barnstead).

2.2. Apparatus and software

In order to obtain digital images, a Philips Webcam SPC900NC VGA with a CCD sensor was used. LabView 7.1 (National Instruments) software used to control the Flow–Batch system was used. The digital images obtained were processed with the ImageJ program (a free internet download). Chemometric data treatment was implemented with The Unscrambler_ 9.7 (CAMO S/A), and Matlab_ 2009b (Mathworks Inc.) software.

2.3. Flow–Batch system

A schematic diagram of the proposed Flow–Batch system is shown in Fig. 2. This system was composed of a lab-made detection cell (DC) built in PTFE which has an inner volume of 5 mL, a quartz circular window and the corresponding holes which allow the entry and exit of the honey solutions. Three three-way solenoid valves (model 137 161T031, Nresearch) were used as follows: VH valve allows the admission of honey sample solution to the DC; VW valve allows the access of

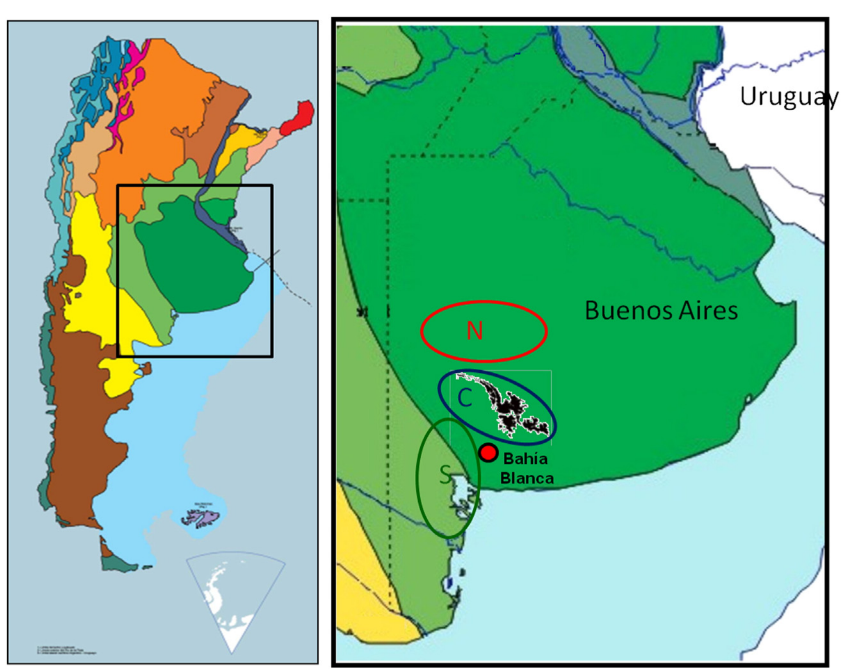

Fig. 1. Geographical location of the studied honey samples. N: north, S: south, and C: Sierra de la Ventana.

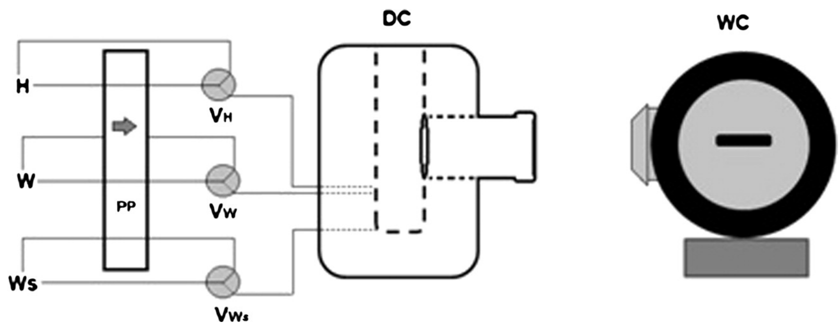

Fig. 2. Schematic diagram of the proposed digital image-based flow-batch system for honey classification.

water in order to clean the DC before the new honey solution enters to the camera and VWs valve removes the liquids of the cell. A four-channel Gilson Minipuls-3M312 peristaltic pump was used to propel the solutions through Tygon® pumping tubes of different internal diameters. For image acquisition a webcam was placed on a support located in front of the quartz window of the detection cell. The DC and the webcam were placed in a sealed black box to prevent the passage of light. A white-light LED was used to generate a uniform illumination. The interior walls of the box were covered with white paper in order to provide to avoid light scattering. A PC microcomputer supplied with a laboratory-made parallel interface was employed to an automatic handling of the valves and the webcam. A computer program developed in LabVIEW 7.1 (National Instruments) was used to control all the proposed digital image-based flow-batch system. Fig. 3 presents a photograph of the proposed Flow–Batch system.

2.4. Flow–Batch procedure

Before starting the acquisition of the images, all solutions in their respective channels were pumped and recycled towards their reservoirs. Then the valves VH and VW were switched on during 5.0 s and the honey solution and water were pumped towards the DC in order to fill in the channels between the valves and the cell. Then, VWs was immediately switched on during 10 s and the excess of the solutions into DC was aspirated to waste. A cleaning step of the DC was carried out by switching on VW during 16 s and afterward VWs to assure the total emptying of the cell. This procedure should be repeated every time a new sample is analyzed.

For image acquisition, DC was filled with the honey solution by switching on VH during 58 s. Before taking the photo, it is necessary to wait up to 30 s to stabilize the solution in order to avoid the formation of concentration gradients. Finally, five digital images of each honey sample during 1 s were captured.

2.5. Histograms and data analysis

In this work the digital images were processed using free downloadable software ImageJ 1.44p, which creates a histogram for each red, green and blue (RGB), hue, saturation and brightness (HSB) and grayscale intensity. The color histograms describe the statistical distribution of pixels as a function of the recorded color components. To perform the analysis of the photos a circular region of 60×60 pixels was selected at the center of each one. Using the selected regions, histograms were constructed.

The 256 tonnes that compose each color component were used as analytical information. To evaluate the influence of each color, three different models were employed: (a) RGB, (b) grayscale and (c) HSB. The models were composed of (a) 3×256, (b) 256 and (c) 3×256 variables, respectively.

The extracted analytical information from the histograms was employed to construct the chemometric classification models using Soft Independent Modeling of Class Analogy (SIMCA), Linear Discriminant Analysis (LDA) with variable selection by the Successive Projections Algorithm (SPA) and Partial Least Squares-Discriminant Analysis (PLS-DA). The data obtained from each histogram were separated into: training (50%), validation (25%) and prediction (25%) sets using the Kennard–Stone algorithm [27] Chemometric data treatment was implemented with The Unscrambler_9.7 (CAMO S/A) and Matlab_2009b (Mathworks Inc.) softwares.

3. Results and discussion

3.1. Exploratory analysis

Initially, an exploratory analysis was carried out using principal component analysis (PCA). For this purpose, the three color models were tested. PCA was performed for dimensionality reduction of the

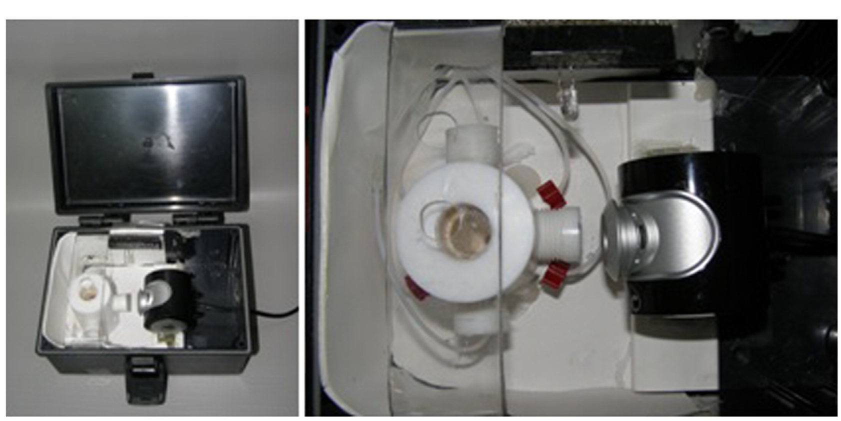

Fig. 3. Photograph of the proposed digital image-based Flow–Batch system for honey classification.

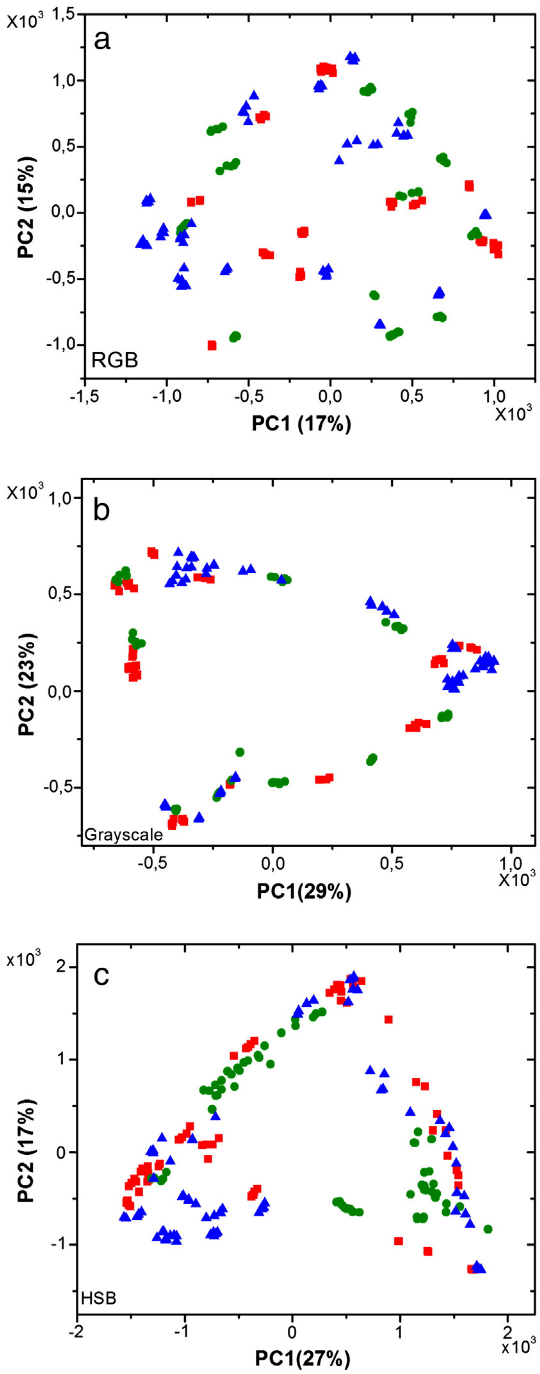

Fig. 4. PCA scores obtained from (a) RGB, (b) grayscale and (c) HSB histograms for all honey samples: north (■), south (●) and Sierra de la Ventana (▲).

data into a smaller number of principal components and to examine the grouping of the 210 studied honey samples according to their geographical origin.

Fig. 4 shows a two dimensional scatter plot (or map) of scores of the first two PCs using the three color models. The plot gives the information about the patterns in the samples. The score plot for (PC1, PC2) is especially useful, since these two components summarize more variation in the data than any other pair of components. In these study the first two principal components accounted for 32, 52 and 44% of the total variance for RGB (PC1 = 17%, PC2 = 15%), Grayscale (PC1 = 29%, PC2 = 23%) and HSB (PC1 = 27%, PC2 = 17%) model respectively. As can be seen, there is a strong overlapping between the studied groups. This was expected because despite having different geographical origins, the flora of the three studied regions is similar. It should be noted that the percentage of variance accounted by PC1 and PC2 for the three models was quite limited. These results obtained with PCA were limited and unsatisfactory. Therefore, further supervised classification techniques were used to classify honey samples according their geographical origin.

3.2. Classification

The SIMCA, SPA-LDA and PLS-DA results for the classification of honey samples in the test set using the RGB, HSB and grayscale color models can be seen in Table 1. By using SIMCA, all samples from the *Sierra de la Ventana* using RGB and from the south using HSB were correctly classified. Summarily, SIMCA modeling obtained 92.1, 90.3 and 66.2% of correct classification using the RGB, HSB and grayscale models, respectively. In addition, when PLS-DA was used, misclassified samples were always obtained for all classes and color models, reaching a mean correct classification of 88.5, 94.3 and 60.4% using the RGB, HSB and grayscale models, respectively. For the SPA-LDA modeling, no classification errors were found for the south samples using both RGB and HSB color models. A mean correct classification of 97.3, 88 and 81.5% was achieved for SPA-LDA using the RGB, HSB and grayscale models, respectively.

The best results were obtained employing SPA-LDA and the RGB color model, where only three samples were incorrectly classified, reaching a mean correct classification of 97.3%. One sample of the North region was classified as belonging to the mountain range, and two other samples belonging to the *Sierra de la Ventana* were misclassified as from the Southern region. This behavior may be due to the fact that these three samples were acquired from apiaries located close to the boundaries of these three regions. Fig. 5 shows the score plot of the two first Fisher's discriminant functions (DF) for SPA-LDA classification using the RGB color model. A good discrimination between the North, South and *Sierra de la Ventana* honey samples was obtained, demonstrating the feasibility of variable selection.

4. Conclusions

A simple digital image-based flow-batch system that includes a webcam to allow geographic origin classification of honey samples was designed. For this purpose, three chemometric tools with different color models were employed. The obtained results show that SPA-LDA and the RGB color model are analytical tools for discriminating honey samples to their geographical origin, reaching a mean correct classification of 97.3%.

Table 1
Summary of the classification results of the honey samples in the test set for SIMCA, PLS-DA, and SPA-LDA at a 95% confidence level.

	SIMCA			PLS-DA				SPA-LDA			
	Errors (%)				Errors (%)				Errors (%)		
Histograms used in the classification models	N	S	C	Optimal no of factors	N	S	C	Selected variables/cost function value	N	S	C
RGB	17.6	6.2	0	12	8.8	12.5	13.2	38/0.71513	2.9	0	5.3
HSB	26.5	0	2.6	4	8.8	3.1	5.3	14/0.74326	17.6	0	18.4
Grayscale	35.3	34.4	31.6	15	44.1	40.6	34.2	19/0.88140	14.7	25	15.8

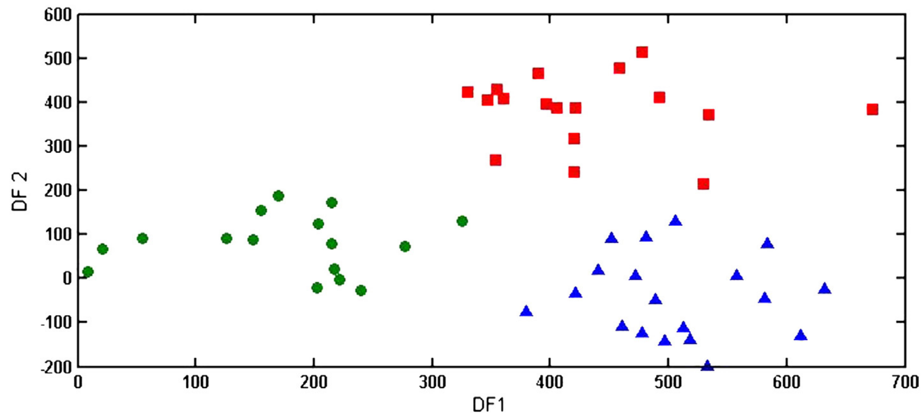

Fig. 5. Discriminant functions plot employing the SPA-LDA and RGB histogram for north (■), south (●) and Sierra de la Ventana (▲) honey samples.

The proposed method has the following advantages: simple, low cost, low consumption of reagents and less time consuming. Furthermore, it requires a minimal human intervention and the sample can be directly analyzed without any special pre-treatment. The Flow–Batch system allows to keep all devices unchanged in order to obtain reproducible images. Also, it enables a rapid admission and removal of the sample solution. Therefore, this system is a good alternative to be used in quality control laboratories for the classification of honey samples according to their geographical origin.

Acknowledgments

The authors acknowledge the financial support from Universidad Nacional del Sur (Dpto. de Química, INQUISUR (UNS-CONICET)) and Proyecto Grupo de Investigación (PGI) granted for Secretaría General de Ciencia y Tecnología (Argentina). M.A. Domínguez wishes to thank Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina). The authors gratefully acknowledge the Brazilian agencies CNPq and CAPES for the scholarships.

References

[1] M.S. Finola, M.C. Lasagno, J.M. Marioli, Microbiological and chemical characterization of honeys from central Argentina, Food Chem. 100 (2007) 1649–1653.
[2] M.I. Isla, A. Craig, R. Ordoñez, C. Zampini, J. Sayago, E. Bedascarrasbure, A. Alvarez, V. Salomón, L. Maldonado, Physico chemical and bioactive properties of honeys from Northwestern Argentina, LWT Food Sci. Technol. 44 (2011) 1922–1930.
[3] N.H. Malacalza, M.A. Caccavari, G. Fagúndez, C.E. Lupano, Unifloral honeys of the province of Buenos Aires, Argentine, J. Sci. Food Agric. 85 (2005) 1389–1396.
[4] M.V. Baroni, M.L. Nores, M.P. Díaz, G.A. Chiabrando, J.P. Fassano, C. Costa, D.A. Wunderlin, Determination of volatile organic compound patterns characteristic of five unifloral honey by solid-phase microextraction–gas chromatography–mass spectrometry coupled to chemometrics, J. Agric. Food Chem. 54 (2006) 7235–7241.
[5] E. Corbella, D. Cozzolino, Classification of the floral origin of Uruguayan honeys by chemical and physical characteristics combined with chemometrics, LWT Food Sci. Technol. 39 (2006) 534–539.
[6] R.G. Pellerano, M.A. Uñates, M.A. Cantarelli, J.M. Camiña, E.J. Marchevsky, Analysis of trace elements in multifloral Argentine honeys and their classification according to provenance, Food Chem. 134 (2012) 578–582.
[7] I. Escriche, M. Kadar, M. Juan-Borrás, E. Domenech, Using flavonoids, phenolic compounds and headspace volatile profile for botanical authentication of lemon and orange honeys, Food Res. Int. 44 (2011) 1504–1513.
[8] C. de Alda-Garcilope, A. Gallego-Picó, J.C. Bravo-Yagüe, R.M. Garcinuño-Martínez, P. Fernández-Hernando, Characterization of Spanish honeys with protected designation of origin "Miel de Granada" according to their mineral content, Food Chem. 135 (2012) 1785–1788.
[9] P.M. Cometto, P.F. Faye, R.D. Di Paola Naranjo, M.A. Rubio, M.A.J. Aldao, Comparison of free amino acids profile in honey from three Argentinian regions, J. Agric. Food Chem. 51 (2003) 5079–5087.
[10] M.V. Baroni, C. Arrua, M.L. Nores, P. Fayé, M.P. Díaz, G.A. Chiabrando, D.A. Wunderlin, Composition of honey from Córdoba (Argentina): assessment of North/South provenance by chemometrics, Food Chem. 114 (2009) 727–733.
[11] A.C. Soria, M. González, C. de Lorenzo, I. Martínez-Castro, J. Sanz, Characterization of artisanal honeys from Madrid (Central Spain) on the basis of their melissopalynological, physicochemical and volatile composition data, Food Chem. 85 (2004) 121–130.
[12] IRAM, standar 15941, Pfund Color Determination, 1997.
[13] IRAM, standar 15932, Ash Determination, 1994.
[14] S. Bogdanov, K. Ruoff, L. Persano Oddo, Physico-chemical methods for the characterisation of unifloral honeys: a review, Apidologie 35 (2004) S4–S17.
[15] S.I.E. Andrade, M.B. Lima, I.S. Barreto, W.S. Lyra, L.F. Almeida, M.C.U. Araújo, E.C. Silva, A digital image-based flow-batch analyzer for determining Al(III) and Cr(VI) in water, Microchem. J. 109 (2013) 106–111.
[16] M.B. Lima, S.I.E. Andrade, I.S. Barreto, L.F. Almeida, M.C.U. Araújo, A digital image-based micro-flow-batch analyzer, Microchem. J. 106 (2013) 238–243.
[17] E. da Nobrega Gaiao, V.L. Martins, W.S. Lyra, L.F. Almeida, E.C. Silva, M.C.U. Araújo, Digital image-based titrations, Anal. Chim. Acta. 570 (2006) 283–290.
[18] R.E. Larraín, D.M. Schaefer, J.D. Reed, Use of digital images to estimate CIE color coordinates of beef, Food Res. Int. 41 (2008) 380–385.
[19] G. Foca, F. Masino, A. Antonelli, A. Ulrici, Prediction of compositional and sensory characteristics using RGB digital images and multivariate calibration techniques, Anal. Chim. Acta. 706 (2011) 238–245.
[20] P.H.G.D. Diniz, H.V. Dantas, K.D.T. Melo, M.F. Barbosa, D.P. Harding, E.C.L. Nascimento, M.F. Pistonesi, B.S.F. Band, M.C.U. Araújo, Using a simple digital camera and SPA-LDA modeling to screen teas, Anal. Methods 4 (2012) 2648–2652.
[21] A. Ulrici, G. Foca, M.C. Ielo, L.A. Volpelli, D.P. Lo Fiego, Automated identification and visualization of food defects using RGB imaging: Application to the detection of red skin defect of raw hams, Innovative Food Sci. Emerg. Technol. 16 (2012) 417–426.
[22] P.H.G.D. Diniz, L.F. Almeida, D.P. Harding, M.C.U. Araújo, Flow–batch analysis, Trends Anal. Chem. 35 (2012) 39–49.
[23] M.F. Razuc, M. Grünhut, E. Saidman, M. Garrido, B.S.F. Band, Green method based on a flow-batch analyzer system for the simultaneous determination of ciprofloxacin and dexamethasone in pharmaceuticals using a chemometric approach, Talanta 115 (2013) 314–322.
[24] M.A. Domínguez, M. Grünhut, M.F. Pistonesi, M.S. Di Nezio, M.E. Centurión, Automatic flow-batch system for cold vapor atomic absorption spectroscopy determination of mercury in honey from Argentina using online sample treatment, J. Agric. Food Chem. 60 (2012) 4812–4817.
[25] J.A. Belloni, M.S. Di Nezio, M.F. Pistonesi, M.E. Centurión, Automatic flow–batch system for the sample treatment and determination of hydroxyproline in sausages, Talanta 89 (2012) 526–530.
[26] S.F.C. Soares, A.A. Gomes, A.R. Galvão Filho, M.C.U. Araújo, R.K.H. Galvão, The successive projections algorithm, Trends Anal. Chem. 42 (2013) 84–98.
[27] R.W. Kennard, L.A. Stone, Computer aided design of experiments, Technometrics 11 (1969) 137–148.