

IDENTIFICATION OF PALE LAGER BEERS VIA IMAGE ANALYSIS

T. C. O. SILVA, M. S. GODINHO and A. E. de OLIVEIRA

*Instituto de Química, UFG, PO Box 131, 74001-970, Goiânia, GO, Brazil
elcana@quimica.ufg.br*

Abstract— This article presents an alternative color-based classification methodology which is used for pattern recognition of pale lager beers. Beer sample images are digitalized on a common desk scanner resulting in color histograms in the RGB scale. The frequency distribution of color indexes according to each color channel were obtained for each image, then decomposed into vector lines R, G and B, each vector having 256 components (indexes/color tones). PCA is used to represent the data in two-dimensional plots, and also to represent color changes of beer samples exposed to light and air. After one hour, significant changes in the yellow color of the beer can be distinguished in the score plot. In short, differences between brands of beer are directly related to differences in their color.

Keywords— PCA, chemical imaging, pale lager beers.

I. INTRODUCTION

Beer is an alcoholic beverage of intense consumption and can be found almost everywhere. It has been produced and consumed in many countries since antiquity, especially in Europe and Asia. Brazil is the fourth largest beer producer in the world (Kirin, 2009), having an annual beer consumption of 50 liters *per capita*. This beverage is produced by the brewing and fermentation of starches and it is a blend of water, malt, hops, and yeast.

Commercially, the most common beer in Brazil is pale lager, having an alcohol content of approximately 4.0% by volume, is light yellow in color, and has low contents of fermentable carbohydrates. This type of beer is 100% malt barley and it is produced with salt reduced water.

As a way of improving beer production techniques, new methods and rapid access to quality control is of great importance in the industry. Flame atomic spectroscopy and linear discriminant analysis were used to classify 25 samples of beer into stout, ale, lager or wheat types (Bellido-Milla *et al.*, 2000). This method was 100% efficient for both lager and wheat. Nuclear magnetic resonance and multivariate analysis were also used to optimize beer analysis (Duarte *et al.*, 2004; Lachenmeier *et al.*, 2005) which is currently restricted because of the extremely high cost of instruments. Fourier transform infrared spectroscopy analysis is also a useful tool in the quality control of beers, since it is fast and requires simple preparation (Lachenmeier, 2007). However, FTIR-partial least squares method of 461 samples

of beer showed a low correlation and accuracy with EBC color.

Among the many attributes of beer, one easily characterized is its color. To date, the classification of beer by color is done during the production process, through a European-wide standard known as the EBC (European Brewery Convention). According to this scale, pale lager beer must contain less than 20 EBC units. For darker beers, 20 units is the lowest allowable EBC number. The amber color of beer is due to pigments known as melanoidins and also from the caramel added to the malt. The Maillard reaction should be taken into account, since it is a non-enzymatic oxidation process during caramelization.

Our working group recently proposed a simple color-based classification methodology for soft drinks using color characteristics of investigated samples whose images were scanned digitally and manipulated in the RGB scale (Godinho *et al.*, 2008). This color scale has 256 color tones (color indexes), varying from 0 to 255 for each basic color channels: red (R), green (G) and blue (B). These 768 color tones ($256 + 256 + 256$), when combined, result in 16,777,216 color tones per pixel (256^3). In this format, a color tone corresponds to a point in a three-dimensional space formed by R, G, and B axes, Fig. 1.

Nowadays, imaging is widespread and has been associated with many development sectors such as industry, communication and research (Pratt, 1991). In chemistry, however, its use is still relatively little, but its application has grown in recent years. Some examples are hyperspectral imaging (Gowen *et al.*, 2008), ultrasound

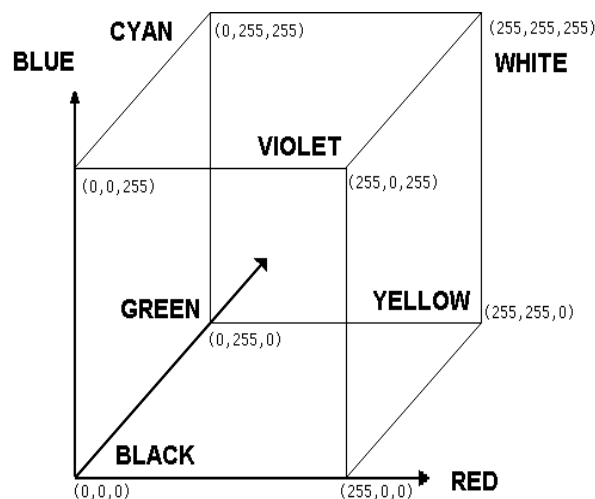


Figure 1. RGB color cube, with axes varying from 0 to 255.

Characteristic colors are on the edges of the cube. imaging (Jin *et al.*, 2004), in photometric analysis in paper matrices (Budantsev, 2004; Schimidt, 1997) in thin-layer chromatography (Hayakawa and Hirai, 2003), in determining the saponin content in quinoa (Souza *et al.*, 2004), in the study of L-glutamate flow in the brains of rats (Hirano *et al.*, 2003), two-dimensional gas chromatography (Reichenbach *et al.*, 2004), the food industry (Antonelli *et al.*, 2004; Yu and MacGregor, 2003), etc.

This paper presents a simple, accessible alternative, and uses a low cost color-based classification methodology for pattern recognition of Brazilian pale lager beers. Beer sample images are digitalized on a common desk scanner resulting in color histograms in the RGB (red, green and blue) scale. At the end, principal component analysis is applied as a pattern recognition tool.

II. METHODS

The beer sample collective was comprised of ten brands of pale lager beer from the retail trade in the state of Goiás, Brazil. For each brand, five 350 mL cans were acquired. All samples for each brand were taken from the same production lot and expiry date. They were randomly collected from bars, supermarkets and beer distributors, taking into account the available variety. All samples were collected and analyzed, and were not subjected to any form of stress.

A. Image Analysis

Approximately 200 mL of each sample at room temperature were transferred to a glass beaker and degassed in an ultrasonic bath for 30 min., then 50.00 mL of the degassed sample was transferred to a Petri dish. Digital sample images of the dish were recorded in JPEG format with 300 dpi (dots per inch) resolution. Three aliquots of each sample were analyzed. For each aliquot, three images were digitized, resulting in 45 images for each brand of beer. Images were then recorded on a Genius desk scanner (ColorPage Vivid, 1200XE). Scilab software (Scilab, 2009) and SIP (Scilab Image Processing) toolbox (Fabri, 2002; SIP, 2009) were used for image manipulation and multivariate analysis calculations.

Histograms of the frequency distribution of color indexes according to each color channel (R, G, and B), were obtained for each digital image, and decomposed into vector lines R, G and B, each vector having 256 components (indexes or color tones). A new vector was then obtained by the juxtaposition of these three vectors, generating a profile for each RGB sample image, having 768 quantified variables per image. For each brand of beer, one average histogram was obtained from each 45 frequency histograms, as seen in Fig. 2.

The data matrix, \mathbf{X} , has the size of the number of analyzed brands, times the number of image replicates, by 768 columns (R, G, and B color channel indexes). This procedure resulted in a 450 x 768 data matrix containing the frequency histograms of all digitalized beer images. The scheme for obtaining the image data matrix can be seen in Fig. 3. Both 300 and 600 dpi JPEG files were analyzed, but only the results for 300 dpi files are

shown, due to the fact that increasing image resolution did not result in a better classification for the analyzed beers.

B. PCA

To conveniently deal with a multidimensional problem such as the simultaneous study of the digitalized image of pale lager beers, the statistical method of principal component analysis (PCA) may be used to represent data in two-dimensional plots, where their axes are the two principal components that explain most data variance. Principal components are the eigenvectors of the matrix product $\mathbf{X}^t\mathbf{X}$ where \mathbf{X}^t is the transpose of the \mathbf{X} matrix. The \mathbf{X} matrix has 450 rows, corresponding to the frequency histograms, and 768 columns, one for each of the color indexes. The original data matrices were pre-processed before the eigenvector calculations. The frequency values were autoscaled (Sharaf *et al.*, 1986). The resulting matrices can be represented in a multidimensional space, as shown in Fig. 4. Each coordinate axis represents one of the color indexes, and the results of the frequency histograms correspond to points in this space.

The first eigenvalue of the matrix product $\mathbf{X}^t\mathbf{X}$ is equal to the amount of statistical variance explained by the first eigenvector (Mardia *et al.*, 1979). This eigenvector, which defines the first principal component axis, points in the direction of maximum statistical variance, as indicated in Fig. 4. The second eigenvector, the second principal component, is perpendicular to the first one and explains a maximum amount of the residual variance in the data – that is, variance not explained by the first eigenvector. If the first two eigenvectors explain a significant amount of the total variance, a principal component score plot in which they are the coordinate axes provides a faithful two-dimensional projection of the 768-dimensional color index space. In such situations, two-dimensional plots can be used to assess quality image data for pattern recognition of pale lager beers.

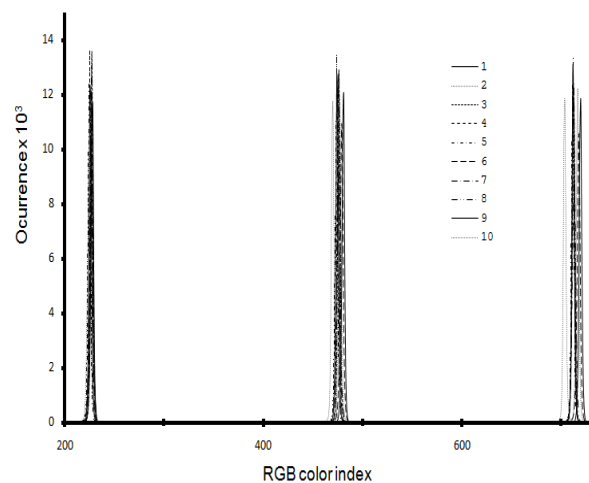


Figure 2. Average histograms *versus* RGB color indexes of ten brands of beer. Digitalized images were 300 dpi JPEG files. Red color indexes ranging from 0 to 255, Green color indexes ranging from 256 to 511, and Blue color indexes ranging from 512 to 767.

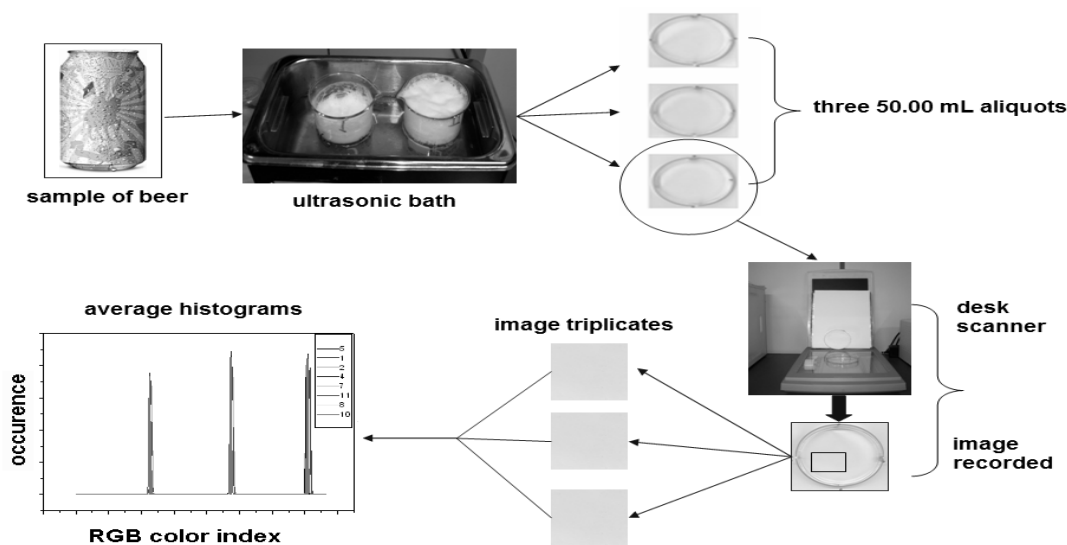


Figure 3. Scheme for obtaining the image data matrix.

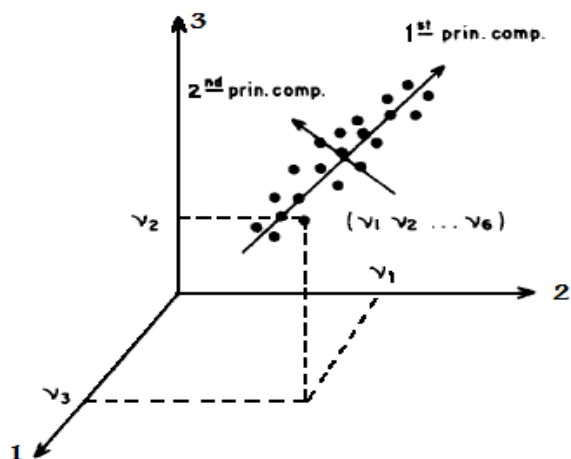


Figure 4. A three-dimensional representation of the 768-dimensional color index space.

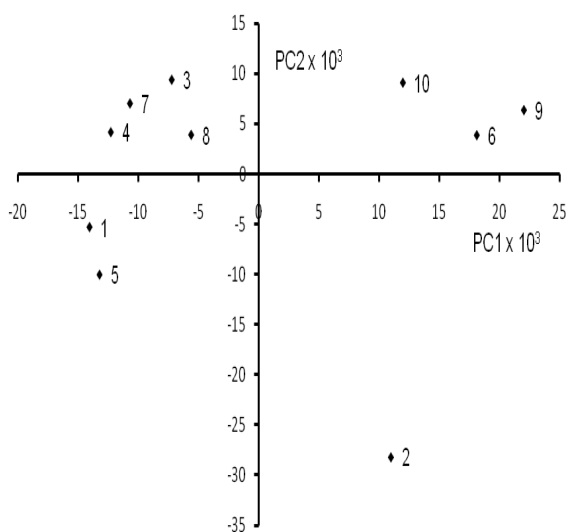


Figure 5. Principal component score graph of the average histograms of ten brands of beer. This graph accounts for 73% of the total data variance.

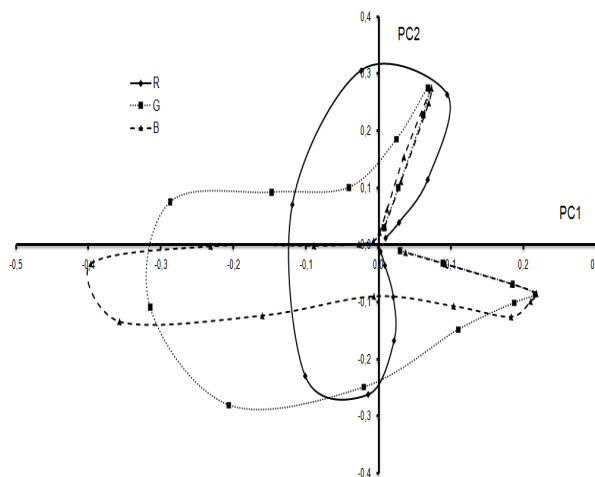


Figure 6. Loading graph for the first two principal components of the average histograms of ten brands of beer. The lowest and the highest color indexes for each color channel are represented, and these indexes increase counterclockwise for each channel.

Image analysis and principal component calculations were carried out using algorithms developed in our laboratory. The calculations were made with the Scilab computer package and SIP toolbox.

III. RESULTS AND DISCUSSION

The constituents of beer are very similar chemically and therefore the collective of beer images is appropriate for chemometric techniques. The data matrix containing average histograms for each brand of beer was mean centered and approximately 75% of the total variance was explained by the first two principal components. PC1 corresponds to 44.16%, and PC2 corresponds to 29.17%. Figs. 5 and 6 contain the graphs of scores and loadings for the first two principal components.

Beer brands 2, 6, 9 and 10, are separated in PC1 from brands 1, 3, 4, 5, 7 and 8, while PC2 separates brands 1, 2 and 5 from the group brands 3, 4, 6, 7, 8, 9 and 10. Four groups, represented in each of the four qu-

adrants, can be distinguished: group 1, brand 2; group 2, brands, 6, 9, and 10; group 3, brands 3, 4, 7, and 8; and group 4, brands 1 and 5.

The point distribution in Fig. 5 is related to the first two principal component loadings according to Fig. 6. Each R, G, and B color channel range is represented by the lowest and highest color index having loadings larger than 0.01. In this figure, only the lowest and the highest color indexes for each color channel are represented. These indexes increase clockwise for each channel. PC1 positive loadings (right) represent less clear tones (see Fig. 1) and also clearer tones (higher color indexes). Intermediary color indexes have negative loadings (left). PC2 separates clear tones (below), while positive values are characteristic of even clearer tones (above).

As a result, data points for each brand of beer are distributed in the score graph, Fig. 5, from the least clear brand, number 2, to the clearest, numbers 6, 9, and 10, represented clockwise according to the beer numbers 2-5-1-4-7-3-8-10-6-9. Based on the loadings in Fig. 5, beer brands are classified as follows: the clearer are numbers 6, 9 and 10, and the least clear is number 2. Brands number 1, 3, 4, 5, 7, and 8 have intermediate tones. Of these, 1 and 5 are not as clear as 3, 4, 7, and 8. This distinction can also be seen when analyzing the average histograms for B channel color indexes. In Fig. 7, for example, the further right, the clearer the brand.

PCA can also be used to represent color changes of samples of beer when exposed to light and air, as in Fig. 8. After one hour, significant changes in the yellow color of the beer can be distinguished in the score plot.

In short, differences shown between the brands are directly related to differences in their colors. What makes a particular brand of beer different from another in the same category is the fact that, although visually imperceptible, on the average, its colors are different. Naturally, it is a direct result from its chemical composition. Therefore, color pattern classification is a cheap and simple alternative compared to the usual physical-chemistry sample classification. This method could facilitate the identification of falsified beer samples and/or counterfeits, with direct involvement in the trade of this product.

IV. CONCLUSIONS

Unlike conventional analytical methods, digital beer images can be used to classify different brands of the beverages of the same type or category. Similarity patterns in the principal component score plot, resulting from the average histograms of R, G, and B color channels, can be obtained from the images generated on a desk scanner. Image color changes in the principal component graphs are accompanied by the weight of the color index in the principal component axes as a result of the average tone for each brand. In general, different brands of beer can be classified within the same lot based on their digitalized images.

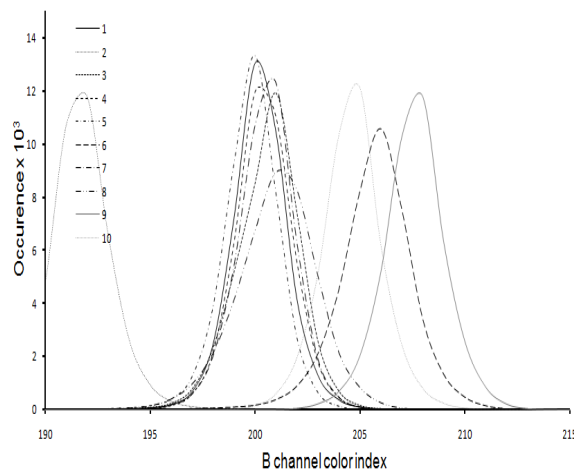


Figure 7. Average histograms versus Blue color indexes of ten brands of beer. Digitalized images were 300 dpi JPEG files.

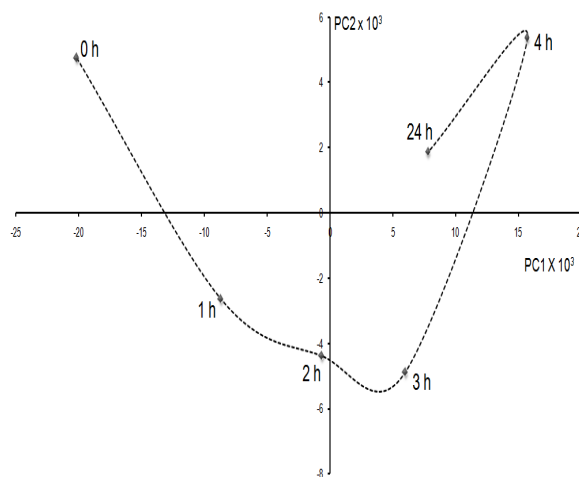


Figure 8. Principal component score graph of the average histograms of six samples of a brand of beer, when exposed to light and room air. Exposure times ranging from 0 h to 24 h are indicated. This graph accounts for almost all the total data variance.

V. ACKNOWLEDGEMENTS

Financial support from the Fundação de Apoio à Pesquisa da UFG (Funape) and the Conselho Nacional de Pesquisa (CNPq) is gratefully acknowledged.

REFERENCES

- Antonelli, A., M. Cocchi, P. Fava, G. Foca, G.C. Franchini, D. Manzini and A. Ulrichi, "Automated evaluation of food colour by means of multivariate image analysis coupled to a wavelet-based classification algorithm," *Anal. Chim. Acta*, **515**, 3-13 (2004).
- Bellido-Milla, D., J.M. Moreno-Perez and M.P. Hernández-Artiga, "Differentiation and classification of beers with flame atomic spectrometry and molecular absorption spectrometry and sample preparation assisted by microwaves," *Spectrochim. Acta B*, **55**, 855-864 (2000).
- Budantsev, A.Y., "Photometric Determination of Compounds in Paper Matrices Using Digital Imaging

- Technique and Transmitted Light Images,” *J. Anal. Chem.*, **59**, 703-707 (2004).
- Duarte, I.F., A. Barros, C. Almeida, M. Spraul and A.M. Gil, “Multivariate analysis of NMR and FTIR data as a potential tool for the quality control of beer,” *Journal of Agricultural and Food Chemistry*, **52**, 1031–1038 (2004).
- Fabri, R., *Desenvolvimento de um pacote de processamento de imagens para o Scilab*, Monograph, Universidade de São Paulo, Brazil (2002).
- Godinho, M.S., R.O. Pereira, K.O. Ribeiro, F. Schimidt, S.B. de Oliveira and A.E. de Oliveira, “Classificação de refrigerantes através de análise de imagens e PCA,” *Quim. Nova*, **31**, 1727-1732 (2008).
- Gowen, A.A., C.P. O’Donell, M. Taghizadeh, P.J. Cullen, J.M. Frias and G. Downey, “Hyperspectral imaging combined with principal component analysis for bruise damage detection on white mushrooms (*Agaricus bisporus*),” *J. Chemom.*, **22**, 259-267 (2008).
- Hayakawa, T. and M. Hirai, “An Assay of Ganglioside Using Fluorescence Image Analysis on a Thin-Layer Chromatography Plate,” *Anal. Chem.*, **75**, 6728-6731 (2003).
- Hirano, A., N. Moridera, M. Akashi, M. Saito and M. Sugawara, “Imaging of L-Glutamate Fluxes in Mouse Brain Slices Based on an Enzyme-Based Membrane Combined with a Difference-Image Analysis,” *Anal. Chem.*, **75**, 3775-3783 (2003).
- Jin, J.Y., G.T. Silva and A.C. Frery, “Sar despeckling filters in ultrasound imaging,” *Lat. Am. Appl. Res.*, **34**, 49-53 (2004).
- Kirin Holdings Company Website, http://www.kirinholdings.co.jp/english/news/2009/0811_01.html (2009).
- Lachenmeier, D.W., W. Frank, E. Humpfer, H. Schäfer, S. Keller and M. Mörtter, “Quality control of beer using high-resolution nuclear magnetic resonance spectroscopy and multivariate analysis,” *European Food Research and Technology*, **220**, 215–221 (2005).
- Lachenmeier, D.W., “Rapid quality control of spirit drinks and beer using multivariate data analysis of fourier transform infrared spectra,” *Food Chemistry*, **101**, 825-832 (2007).
- Mardia, K.V., J.T. Kent and J.M. Bibby, *Multivariate Analysis*, Academic Press, London (1979).
- Pratt, W.K., *Digital Image Processing* (2nd edn), John Wiley & Sons (1991).
- Reichenbach, S.E., M. Ni, V. Kottapalli and A. Visvanathan, “Information technologies for comprehensive two-dimensional gas chromatography,” *Chemom. Intell. Lab. Syst.*, **71**, 107-120 (2004).
- Schimidt, F., “Desenvolvimento de um programa computacional para tratamento de dados de imagens digitalizadas e sua aplicação em química analítica,” *Master Thesis*, Unicamp, Brazil (1997).
- Scilab, Software, <http://www.scilab.org.br/> (2009).
- Sharaf, M.A., D.L. Illman and B.R. Kowalski, *Chemometrics*, Wiley (1986).
- SIP Toolbox, <http://siptoolbox.sourceforge.net> (2009).
- Souza, L.A.C., C.R. Spehar and R.L.B. Santos, “Análise de imagem para determinação do teor de saponina em quinoa,” *Pesq. Agro. Bras.*, **39**, 397-401 (2004).
- Yu, H. and J.F. MacGregor, “Multivariate image analysis and regression for prediction of coating content and distribution in the production of snack foods.” *Chemom. Intell. Lab. Syst.*, **67**, 125-144 (2003)

Received: December 7, 2009.

Accepted: April 27, 2010.

Recommended by Subject Editor Jorge Solsona.

