

# Fast Determination of Biodiesel Content in Commercial Diesel/Biodiesel Blends by Using Digital Images and Multivariate Calibration

Mayara FERREIRA BARBOSA,\* Danielle SILVA DO NASCIMENTO,\*\* Marcos GRÜNHUT,\*\*  
Hebertty VIEIRA DANTAS,\* Beatriz S. FERNANDEZ BAND,\*\* Mario C. UGULINO DE ARAÚJO,\*† and  
Matías INSAUSTI\*\*†

\*Universidade Federal da Paraíba, Departamento de Química, Laboratório de Automação e Instrumentação em Química Analítica/Quimiometria (LAQA), Caixa Postal 5093, CEP 58051-970—João Pessoa, PB, Brazil

\*\*Laboratorio FIA, INQUISUR (CONICET)—Departamento de Química, Universidad Nacional del Sur, Av. Alem 1253, B8000CPB, Bahía Blanca, Argentina

A new simple, rapid and inexpensive analytical method was developed to determine the biodiesel percentage in biodiesel/diesel blends through simple digital images of samples obtained by scanning with a commercial scanner. Soybean biodiesel and petroleum diesel samples were acquired from refineries currently in operation. There were prepared several mixtures within the range 1.5 to 12.0% of biodiesel in diesel oil, using the same procedure as is done in industry. The analytical signals were images recorded with a scanner. This data was decomposed with different color systems: RGB, HSV, HLS, CMYK and Grayscale. Chemometrics models based on color signals obtained from different mixtures of biodiesel/diesel were built. The quantification by using partial least squares (PLS) resulted in a RMSEP value for biodiesel of 0.9% (w/w); this load approximately 10-times smaller than the corresponding calibration range, with a correlation of 0.96 between predicted and reference values.

**Keywords** Biodiesel content, digital images, chemometrics, diesel/biodiesel blends

(Received May 23, 2017; Accepted August 2, 2017; Published November 10, 2017)

## Introduction

Biodiesel is a biodegradable fuel derived from renewable sources, such as vegetable oils and animal fats. In recent years, biodiesel has gained attention because it is a biofuel with similar properties that can partly or completely replace petroleum-derived diesel.<sup>1</sup> Since biodiesel is completely miscible with diesel, the blending of both fuels in any proportion is recommended in order to improve its qualities.<sup>2</sup> Biodiesel has been proven to reduce the emissions of hydrocarbons, carbon monoxide and particulates when used alone or in blends that include petroleum diesel. Consequently, blends of biodiesel and diesel have been used worldwide because of their environmental, economic, and social advantages.<sup>3</sup>

Argentina ranks as the world's largest biodiesel exporter. Most of Argentina's pure biodiesel (known as B100) is already being exported to foreign markets; the production capacity is expected to continue increasing in the next years.<sup>4</sup> The Argentine and Brazilian biodiesel industries are mainly based on the use of soybean as the feedstock.<sup>5</sup> There is a regulated minimum amount of 10 and 7% (v/v), respectively, and the most commonly

used raw oil source is the soybean oil.<sup>6</sup> Therefore, the quality control of biodiesel is greatly important to the successful commercialization of this fuel and its blends.

Those requirements demand the development of methods for determining the biodiesel content in diesel blends. The standard methods, ASTM D7371-07 and EN 14078, use mid-infrared spectroscopy to determine the biodiesel content in diesel oil. For this kind of determination some studies indicate the possibility of using the near-infrared region,<sup>7,8</sup> as well as the visible one<sup>9,10</sup> associated to univariate or multivariate calibration methodologies, to perform a rapid analysis without any sample treatment.

The evolution of instrumental techniques has allowed the production of a large amount of information about the sample, enabling the development of instrumental analytical methods associated with multivariate regression. Chromatography and infrared spectroscopy techniques are most often used to analyze the quality and quantification of biodiesel in diesel; however, physical-chemical methods are also used as techniques for these analyses.<sup>11,12</sup>

The use of digital images can replace the human visual system, thus eliminating the subjective nature of the analysis, which is substantially influenced by environmental conditions and subject to inconsistencies. Using digital imaging data has proven to be a very effective alternative with univariate<sup>13</sup> or multivariate techniques.<sup>14</sup> Digital images can be obtained by various devices, including digital cameras, scanners, X-ray machines, electron

M. C. U. D. A. and M. I. contributed equally to this work

† To whom correspondence should be addressed.

E-mail: mariougulino@gmail.com (M. C. U. D. A.); matiasi@conicet.gov.ar (M. I.)

microscopes, ultrasound devices and radars, mobile phones, webcam, among others.<sup>15</sup> The application of these advantages involves coffee classification,<sup>16</sup> tea,<sup>17</sup> honey,<sup>18</sup> biodiesel,<sup>19</sup> and vegetable oil<sup>20</sup> among other matrices. Digital images can also be used for quantification in matrices such as air,<sup>21</sup> milk,<sup>22</sup> soda,<sup>23</sup> and water<sup>24</sup> among others.

In this sense, it would be interesting to develop a simple, low-cost and rapid methodology based on digital images and multivariate calibration techniques to quantify biodiesel in petroleum diesel that is rapid, inexpensive, and non-destructive. Despite the homogeneity of the samples, the use of color histograms and a statistical treatment in digital images amplify the proposal, since small variations of colors in the samples are more perceptible and provide possibilities of analytical applications. The aim of this study is to determine the biodiesel content in commercial diesel fuel by using digital images (scanner) and multivariate calibration. For this motion, different color histograms in the RGB (Red, Green, Blue), HSV (Hue, Saturation, Value), HLS (Hue, Luminosity, Saturation), CMYK (Cyan, Magenta, Yellow, Black) and Grayscale channels, and their combinations were extracted from digital images and used as analytical information, and then statistically evaluated with a PLS (partial least squares) algorithm.<sup>25</sup>

## Experimental

### Samples

Five commercial biodiesel samples of soybean were acquired from different producer companies located in diverse states of Argentina. One sample of pure diesel was supplied by a local refinery in the city of Bahía Blanca. With these raw materials, fourty diesel/biodiesel mixtures were prepared in order to obtain a percentage range of between of 1.5 to 12.0% biodiesel (w/w), in 1.5% (w/w) increments. Such diesel/biodiesel blends were used to calculate the multivariate calibration models. In addition, 5 commercial samples obtained in gas stations as real samples were collected. Images of the samples without any treatment were recorded in the scanner.

The choice to not study others biodiesel sources or samples contaminated with raw oil was because in Argentina and Brazil, the native countries of the authors, the biodiesel employed is solely from raw soybean oil, and the biofuel used in industry to blend the petroleum diesel is exhaustively analyzed before being employed. It therefore hardly has any substantial content of raw oil.

### Equipment and image acquisition

A multifunction (MFP) printer, HP Photosmart C3100 (Fig. 1(a)), was used to scan the samples. A black box (Fig. 1(c)) was used to eliminate potential interferences of light occurring during the scanning record. A sheet of white paper, A4 (Fig. 1(d)), was also used as a light diffuser in the cap of the box. A vial with dimensions of 15 × 45 mm, 4 mL of volume (Fig. 1(b)) was used as a sample container. A volume of 0.5 mL of each sample was used to record the image. Between each measurement, the vial was washed with water, detergent and ethanol to avoid any remnant portion of the sample measured before. Figure 1 shows the scheme of the system used. It should be noted that the proposal of this methodology could be easily extended to various vials in each scanning.

Image acquisition takes into account the overall visual characteristics of the sample surface. The images acquired were stored in the JPEG format, which is a compressed format. Whereas color histograms describe the statistical distribution of the pixels as a function of the recorded color component, the

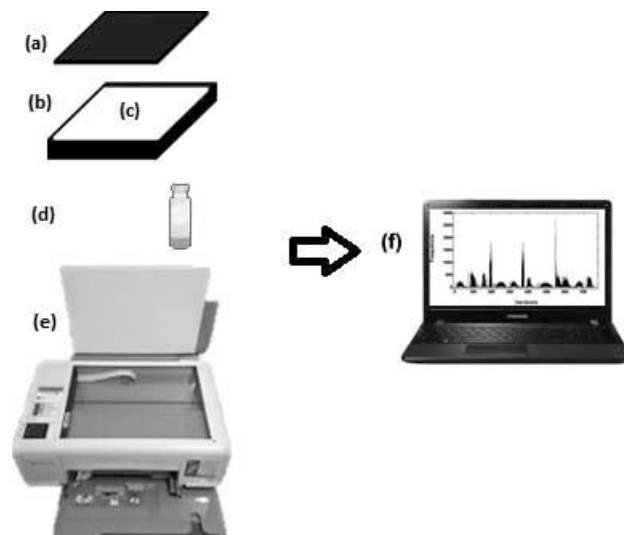


Fig. 1 Scheme of the equipment used for image acquisition. Box cover (a), black box (b), white paper sheet (c), with 0.5 mL of fuel sample (d), scanner (e), notebook (f).

selection of the region of interest provides sufficient information to obtain a suitable quantification.

The image acquisition and processing were performed by a notebook (Fig. 1(f)) connected to the MFP, and also the notebook was used in chemometric procedures. The acquisition of images from the MFP was carried out with a resolution of 300 dpi (dot per inch), the region of interest (ROI) with 51 × 53 pixels being selected in the center of the vial for a better characterization of the physical aspect of the sample. Each ROI was converted to RGB, HSV, HLS, CMYK, and Grayscale histograms using a publicly available, (<http://www.laqa.quimica.ufpb.br/index.php/downloads>) graphical interface.

The 5 commercial samples were tested to obtain reference values by using the fuel analyzer PetroSpec (PAC) (TD PPA), a near and mid-infrared spectroscopic analyzer. This instrument allows for the determination of biodiesel content, where the repeatability and reproducibility agree with those required by the standard ASTM methods.<sup>26</sup> It has an internal library with a diverse array of more than 600 fuels, obtaining a worldwide variability by the main producers of fuel. An automatic pressurized system purges and fills the new undiluted fuel sample (10 mL), and the total analysis takes 3 min to obtain the result. Outlier samples are detected by this instrument through the Mahalanobis distances from the calibration set.

### Chemometric calculation

Color histograms describing the statistical distribution of the pixels as a function of the recorded color component in the RGB, HSV, HLS, CMYK and Grayscale were obtained and used as analytical information. To assemble arrays, the data provided for each color system was grouped one next to each other, which together accounted for 3584 variables (256 levels for each of the 14 features under consideration: R, G, B (RGB), H, S, V (HSV), H, L, S (HLS), C, M, Y, K (CMYK) and Grayscale). A routine written in Matlab language was used to remove samples with above 50% of variables with a zero value, to simplify the dataset for calculations. Such a treatment leaves an array of 735 variables, which can be seen in Fig. 2. This limit was defined as above 50% because if half of the samples present zero value for a given variable it is because this variable

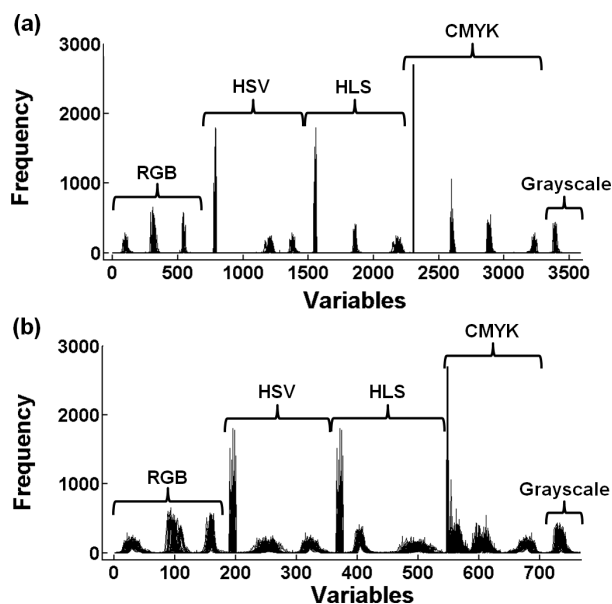


Fig. 2 Histograms of diesel/biodiesel blends. Histograms with all variables (a), histograms after the withdrawal of variables that contained above 50% of zeros (b).

can be removed without great losses. Despite the visual homogeneity of the biodiesel/diesel samples when the decomposition in color histograms is realized, it is possible to observe (Fig. 2) a great variation in the distribution of the pixels in the different color systems studied.

The 40 diesel/biodiesel mixtures were separated into a calibration (25 samples) and a prediction set (15 samples) by the modified Kennard-Stone algorithm (KS XY).<sup>27</sup> Such as real samples, 5 commercial blends were collected from local gas stations to be predicted with the calculated models. The PLS multivariate calibration models were built using a cross-validation procedure to select the best number of latent variables. This best model was used to predict a set of 15 samples and the real samples. The different models were evaluated in terms of the root mean square error of the prediction (RMSEP), correlation coefficient ( $r$ ) and bias (BIAS). It worth noticing that the prediction set was not used at any step of the model construction.

KS XY and PLS algorithms were calculated by a lab-made routine written in Matlab 2010.b (Mathworks, Inc.).

## Results and Discussion

A digital image acquired by cameras or scanners showed the physical aspects of samples. Since biodiesel samples have native absorption at between 400 and 500 nm, which almost correspond to green and blue. The histogram of RGB in Fig. 2 corroborates that green and blue were the predominant colors. Petroleum diesel fuel is colorless, so, the overall look of the fuel sample changes as a function of the added biodiesel. In this way, the compounds present in biodiesel will influence the final color of the diesel/biodiesel blend. This alteration makes possible a correlation between the biodiesel quantity and the components of the digital images.

Initially, the histograms of color systems were tested separately to find the system with the best correlation between the color and biodiesel content in diesel. Table 1 shows the results of

Table 1 PLS model results for the calibration and prediction set

Imaging system	Calibration set		Prediction set	
	$R^a$	RMSECV	$r$	RMSEP
RGB	0.98 (3)	0.9	0.86	1.8
HSV	0.94 (2)	1.3	0.91	1.4
HLS	0.98 (4)	0.8	0.91	1.4
CMYK	0.93 (1)	1.5	0.89	1.1
Grayscale	0.88 (1)	1.7	0.15	3.7
All	0.98 (3)	0.9	0.96	0.9

a. The number of latent variables employed in PLS models was indicated in parenthesis.

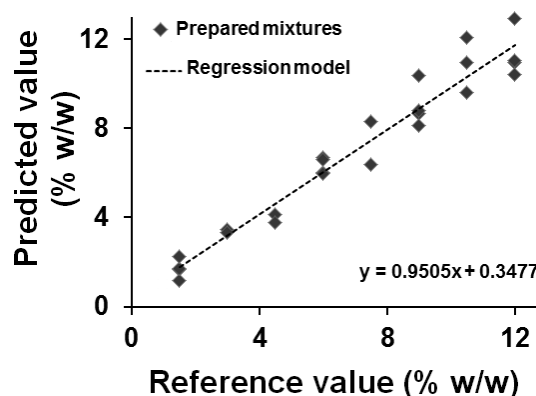


Fig. 3 Calibration set of samples. Prepared biodiesel blends in diesel oil vs. predicted values by the PLS model.

PLS models built by using each color system data, as well as all grouped as shown in Fig. 2.

Using the histograms of all color model available helps in characterizing the aspect of the digital image recorded. Thus, the use of chemometrics tools is remarkable to distinguish the variations in the histograms. The present paper used all color models together, whereas they were evaluated separately, and they did not provide satisfactory results. Better results were achieved by using a greater number of variables; namely, it provided a better mathematical description of the characteristics of the digital image.

The PLS model<sup>28</sup> was calculated with the calibration set of samples, and 3 latent variables were selected by cross-validation, obtaining RMSECV = 0.9% (w/w),  $r = 0.98$  e BIAS = 0.13% (w/w) (Fig. 3). The estimation of the concentrations of the prediction set gives values of RMSEP = 0.9% (w/w),  $r = 0.96$ , BIAS = 0.18% (w/w) and REP = 8.57% (Fig. 4). It is verified that there is no systematic error, since the results are distributed on both sides of the bisectrix lines, not indicating trends in the PLS models.

The elliptical joint confidence region (EJCR)<sup>29</sup> was obtained based on the linear regression between the reference values and the predicted concentrations of biodiesel in the built PLS model. This region can be visualized from Fig. 5, for a confidence level of 95%. The EJCR contains the ideal intersection point (slope = 1, and intercept = 0), indicating that the bias in the prediction of the PLS model is not significant.

Figure 5 shows that the Elliptical joint confidence regions (EJCR)<sup>29</sup> for the slope and intercept contain the ideal result of slope = 1, and intercept = 0; concluding that constant and

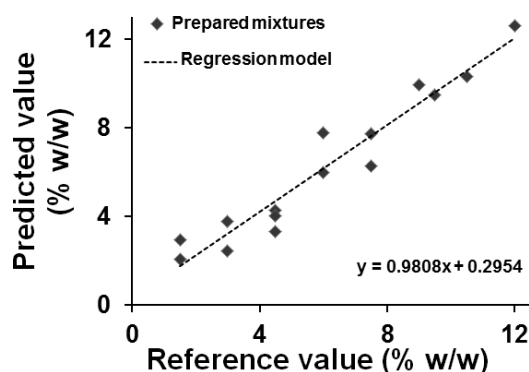


Fig. 4 Prediction set of samples.

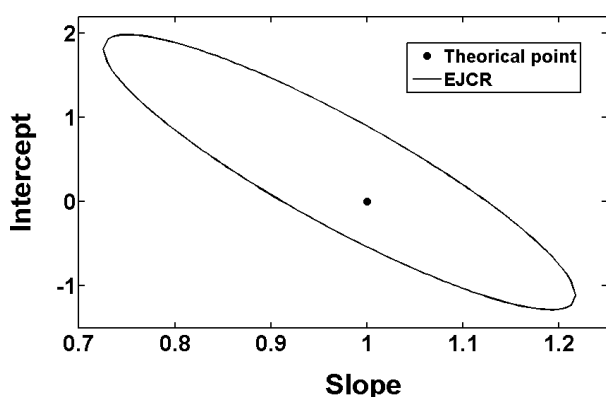


Fig. 5 Elliptical joint confidence region corresponding to the lineal regression analysis of predicted vs. reference values, as estimated from the PLS method. The point marks the theoretical (1, 0) point.

proportional biases are absent.

The results concerning the 5 commercial samples that were evaluated by the PLS model, are given in the Table 2.

## Conclusions

In this work a simple, rapid, inexpensive, and non-destructive methodology based on digital images and multivariate calibration techniques for biodiesel quantification in diesel oil blends was explored. Intrinsic advantages include fast analyses, no reagents, and the avoidance of chemical characterization of the samples, *i.e.* waste generation is remarkably reduced, contributing to the primary objective of green chemistry.

The multivariate calibration models provide excellent results by using partial least squares with the only data pretreatment of removing the variables with a maximum of zero values. The quantification by using PLS resulted in an RMSEP value for biodiesel of 0.9% w/w, approximately 10-times smaller than the corresponding calibration range, with a correlation of 0.96, between the predicted and reference value. The correct prediction of real samples acquired in different local gas stations confirms the immediate application of this methodology.

Taking into account the worldwide variation in the composition of diesel fuel, the present method could be re-calibrated with 5 mixtures of a pure diesel with any biodiesel and with only one digital image of the mixtures.

Table 2 Commercial samples results

Sample	Reference value (%, w/w)	Predicted value (%, w/w)	% Error
1	7.52	6.87	8.64
2	7.03	6.86	2.42
3	10.11	10.21	-1.00
4	8.21	8.76	-6.70
5	5.81	6.86	-18.1

The most valuable characteristics of the proposed method are the low cost and fast analysis of biodiesel blends; these advantages can be useful for an initial assessment of biodiesel in the diesel industry. In addition is the accessibility of the instrument needed (multifunctional printer), compared with the specific and expensive analyzer of the reference method. Whereas fuel diesel is a mixture of thousands of hydrocarbon compounds, it will not be necessary to use a huge number of samples to obtain a calibration model with the possibility to predict every sample in the world.

## Acknowledgements

The authors acknowledge the support of CAPES/SPU (PPCP 013/2011), CNPq, CONICET (PIP 11220120100625) and PGI-UNS (24Q054).

## References

1. R. K. Saluja, V. Kumar, and R. Sham, *Renewable Sustainable Energy Rev.*, **2016**, *62*, 5866.
2. J. Corach, P. A. Sorichetti, and S. D. Romano, *Fuel*, **2016**, *177*, 268.
3. M. Insausti and B. S. F. Band, *Energy Fuels*, **2016**, *30*, 5341.
4. "Argentina, *Energy Policy, Laws and Regulations Handbook*", **2015**, Vol. 1, International Business Publications, Washington, D.C.
5. O. M. Ali, R. Mamat, N. R. Abdullah, and A. A. Abdullah, *Renewable Energy*, **2016**, *86*, 59.
6. "Ministerio De Planificacion Federal, Inversion Publica Y Servicios. Secretaria De Energia. Resolución N° 450/2013 6.8.2013", <http://www.infoleg.gov.ar/infolegInternet/anexos/215000219999/218395/norma.htm>, accessed April 2017.
7. M. F. Pimentel, G. M. G. S. Ribeiro, R. S. Cruz, L. Stragevitch, J. G. A. Pacheco Filho, and L. S. G. Teixeira, *Microchem. J.*, **2016**, *82*, 201.
8. E. M. Paiva, J. J. R. Rohwedder, C. Pasquini, M. F. Pimentel, and C. F. Pereira, *Fuel*, **2015**, *160*, 57.
9. D. D. S. Fernandes, A. A. Gomes, G. B. Costa, G. W. B. Silva, and G. Vêras, *Microchem. J.*, **2011**, *87*, 30.
10. M. A. A. Silva, R. A. Correa, M. G. O. Tavares, and N. R. Antoniosi Filho, *Fuel*, **2015**, *143*, 16.
11. N. A. Portela, E. C. S. Oliveira, C. N. Alvaro, R. R. T. Rodrigues, S. R. C. Silva, E. V. R. Castro, and P. R. Filgueiras, *Fuel*, **2016**, *166*, 12.
12. G. Knothe, *Trans. ASAE*, **2001**, *44*, 193.
13. M. B. Lima, S. I. E. Andrade, I. S. Barreto, L. F. Almeida, and M. C. U. Araújo, *Microchem. J.*, **2013**, *106*, 238.
14. M. I. S. Gonçalves, W. T. S. Vilar, E. P. Medeiros, and M. J. C. Pontes, *Comput. Electron. Agric.*, **2016**, *123*, 89.

15. L. F. Capitán-Vallvey, N. López-Ruiz, A. Martínez-Olmos, M. M. Erenas, and A. J. Palma, *Anal. Chim. Acta*, **2016**, *899*, 23.
  16. U. T. C. P. Souto, M. F. Barbosa, H. V. Dantas, A. S. Pontes, W. S. Lyra, P. H. G. D. Diniz, M. C. U. Araújo, and E. C. Silva, *Food Anal. Methods*, **2014**, *8*, 1515.
  17. P. H. G. D. Diniz, H. V. Dantas, K. D. T. Melo, M. F. Barbosa, D. P. Harding, E. C. L. Nascimento, M. F. Pistoneli, B. S. F. Band, and M. C. U. Araújo, *Anal. Methods*, **2012**, *4*, 2648.
  18. M. A. Domínguez, P. H. G. D. Diniz, M. S. Di Nezio, M. C. U. Araújo, and M. E. Centurión, *Microchem. J.*, **2014**, *112*, 104.
  19. G. B. Costa, D. D. S. Fernandes, V. E. Almeida, T. S. P. Araújo, J. P. Melo, P. H. G. D. Diniz, and G. Vêras, *Talanta*, **2015**, *139*, 50.
  20. K. D. T. M. Milanez and M. J. C. Pontes, *Microchem. J.*, **2014**, *113*, 10.
  21. J. P. Filho, J. F. S. Petrucci, and S. S. Cardoso, *Talanta*, **2015**, *140*, 73.
  22. P. M. Santos and E. R. Pereira-Filho, *Anal. Methods*, **2013**, *5*, 3669.
  23. B. G. Botelho, L. P. Assis, and M. M. Sena, *Food Chem.*, **2015**, *159*, 175.
  24. A. L. Molinero, V. T. Cubero, R. D. Irigoyen, and D. S. Piazuolo, *Talanta*, **2013**, *103*, 236.
  25. S. Wold, M. Sjostrom, and L. Eriksson, *Chemom. Intell. Lab. Syst.*, **2001**, *28*, 109.
  26. ASTM D 1500-12 Standard Test Method for Color of Petroleum Products; ASTM D1500-07 Standard Test Method for ASTM Color of Petroleum Products (ASTM Color Scale), Book of Standards, Vol. 05.01.
  27. R. K. H. Galvão, M. C. U. Araújo, G. E. José, M. J. C. Pontes, E. C. Silva, and T. C. B. Saldanha, *Talanta*, **2005**, *67*, 736.
  28. K. R. Beebe, R. J. Pell, and M. B. Seasholtz, “*Chemometrics—A Practical Guide*”, **1998**, Wiley, New York.
  29. A. C. Olivieri, H. C. Goicochea, and F. A. Inon, *Chemom. Intell. Lab. Syst.*, **2004**, *73*, 189.
-