

Use of FTIR-ATR Spectroscopy Combined with Multivariate Analysis as a Screening Tool to Identify Adulterants in Raw Milk

Daniele G. Conceição,^a Ben-Hur R. F. Gonçalves,^a Fernanda F. da Hora,^b
Amanda S. Faleiro,^b Leandro S. Santos^a and Sibelli P. B. Ferrão^{✉*,a}

^aPrograma em Engenharia e Ciência de Alimentos, Universidade Estadual do Sudoeste da Bahia (UESB),
Rodovia BR 415, km 03, s/n, 45700-000 Itapetinga-BA, Brazil

^bLaboratório de Processamento de Leite e Derivados, Universidade Estadual do Sudoeste da Bahia (UESB),
Rodovia BR 415, km 03, s/n, 45700-000 Itapetinga-BA, Brazil

The objective of this study was to use Fourier transform infrared (FTIR) spectroscopy combined with multivariate analysis to identify adulterations in raw milk and in samples from producers. Five levels of concentration of sodium bicarbonate, sodium hydroxide, hydrogen peroxide, starch, sucrose and urea were used. A total of 620 samples previously adulterated, frozen and lyophilized were analyzed in FTIR-attenuated total reflection (ATR) equipment and 15 peaks of the spectra were obtained. With the multiple linear regression method for samples adulterated with sodium bicarbonate, sucrose and urea, a coefficient greater than 75% was obtained, and with artificial neural networks all adulterated samples obtained a percentage of correctness greater than 76.6%, making it possible to identify adulterants from 0.1%. Of the 249 samples of producers analyzed, 2.4% were adulterated. With the use of FTIR allied to the multivariate analysis as a screening method, it was possible to obtain a satisfactory classification for the adulterated samples in this study.

Keywords: dairy products, adulteration, chemometrics

Introduction

Milk consists of an aqueous phase, composed of lactose, water-soluble vitamins and minerals; a phase in suspension state, composed of caseins bound to salts; and a phase in the state of emulsion, composed of fat and fat-soluble vitamins.¹ Milk is a source of protein of high biological value and calcium, besides being composed of essential fatty acids like stearic and linoleic.^{2,3} Some physicochemical properties are important to determine their quality, which can be altered when milk is adulterated, such as total acidity, relative density and cryoscopy.

There are two possible types of adulteration in milk: (i) adulteration by substitution, which occurs when there is complete or partial removal of some component; and (ii) by addition, when substances are added in order to mask the lower quality. All these practices are considered adulterations when there is no consumer knowledge.⁴

Common adulterants added to milk are water and cheese whey to increase volume, generally followed by the

addition of density replenishers such as starch, sucrose and urea; compounds to mask acidity, such as sodium hydroxide and sodium bicarbonate; substances for the purpose of preserving milk, such as hydrogen peroxide; and urea, to simulate the increase in protein content.⁵⁻⁷

The adulterations can be identified by routine analysis (fat content, titratable acidity, dried extract, degreasing, cryoscopic index and density) and analyses such as electrophoresis, chromatography, enzyme linked immunoenzymatic assay and Fourier transform infrared (FTIR) spectroscopy.⁸ FTIR is a technique that has been used because it is fast and does not require chemical reagents.^{9,10}

Spectroscopy studies the interaction of electromagnetic radiation with matter in order to evaluate the chemical bonds, which acquire vibrational and rotational motions, and the difference between the emitted radiation and the radiation absorbed by the sample gives rise to the spectrum,¹¹ which is considered the “fingerprint” of the sample. This technique has been used to evaluate the quality of dairy products and to identify adulterations, being possible by analyzing the peaks obtained by the spectrum.^{5,12}

*e-mail: sibpass@yahoo.com.br

The multivariate analysis applied to the spectroscopy allows the assessment of the data obtained by the FTIR associated with more than one frequency, because two or more variables are analyzed at the same time, in order to predict quality parameters.^{8,9} It is a tool used to analyze data and identify adulterants in dairy products such as maltodextrin,¹² soy extract,¹ melamine¹³ and cheese whey.¹⁴ Multivariate analysis techniques such as multiple linear regression analysis and artificial neural networks can be used to detect adulterants in milk.

The objective of this study was to use a screening tool to identify neutralizers, preservatives and replenishers in raw milk samples using FTIR and multivariate analysis.

Experimental

Collection and adulteration of raw milk samples

Milk of crossbred Holstein × Zebu cows was obtained under satisfactory sanitary hygiene conditions and mechanical milking for 20 days between March and May 2017. For each day of collection, 32 treatments were obtained, one being a control sample and 30 adulterations, totaling 620 experimental units. The milk samples were adulterated as follows: milk with 0.03, 0.05, 0.1, 0.3 and 0.5% sodium hydroxide (Vetec; ≥ 97% purity); milk with 0.03, 0.05, 0.1, 0.3 and 0.5% sodium bicarbonate¹⁵ (Vetec; ≥ 99% purity); milk with 0.05, 0.1, 0.2, 0.3 and 0.4% of 3.0% hydrogen peroxide (Cap-Lab); milk with 10% water and 2.0, 4.0, 6.0, 8.0 and 10.0% urea (Vetec; ≥ 99% purity);⁸ milk with 10% water and 2.0, 2.5, 3.0, 3.5 and 4.0% sucrose (Dynamics; ≥ 99% purity); and milk with 10% water and 0.1, 0.3, 0.5, 1.0 and 1.5% starch (Dynamics; ≥ 99.5% purity).⁶

From August 2016 to June 2017, 249 samples of producers were obtained at the time of receipt in a dairy. For each adulteration and samples of producers, 1 mL of the sample was collected in an Eppendorf tube and frozen at -80 °C during 48 h for later lyophilization for 72 h (Terroni Fauvel L2000 lyophilizer).

Spectroscopic analysis

The freeze-dried samples (0.5 g) were analyzed in a Cary 630 FTIR-attenuated total reflection (ATR) spectrometer, and the spectra obtained in the region of mid-infrared of 4000 to 600 cm⁻¹. Data were processed using Microlab and Resolution Pro softwares (Agilent). The spectral regions and the main functional groups of milk and adulterants analyzed in the FTIR-ATR were identified in the literature.¹⁶⁻¹⁹

Multivariate analysis

In order to discriminate the samples in milk and adulterated milk, the data of each adulterant were organized in an $m \times n$ (row × column) matrix where m represents the milk samples and n represents the absorbance values obtained by the spectrum in the FTIR-ATR (variables). The outlines of the matrices were identified using Cook's D method, in which the samples that showed deviation greater than 2 were removed from the analysis. For the analysis of multiple linear regression and artificial neural network, the data were randomized using the Kennard-Stone algorithm and divided into 70% for training and 30% for validation of each concentration (Table 1).²⁰

Table 1. Set of samples, training set and validation set of adulterants

Group	Total set	Training set	Validation set
Sodium bicarbonate	113	79	34
Sodium hydroxide	115	81	34
Hydrogen peroxide	114	79	35
Starch	114	79	35
Sucrose	113	78	35
Urea	114	79	35

The data for multiple linear regression were standardized for mean value = 0 and standard deviation = 1 and the multicollinearity analysis was performed between the explanatory variables using the diagnostic method. The multiple linear regression analysis was performed by the stepwise method, and the backward, forward and stepwise forms were tested. The ones that obtained the least number of observations, the lowest mean square error (MSE) and the highest determination coefficient to obtain the equation were chosen.

For artificial neural network classification, data were normalized from 0 to 1, expressed by the equation 1, and inserted into the Java Neural Network Simulator (JavaNNS) 1.1,²¹ based on the Stuttgart Neural Network Simulator (SNNS).

$$x_{\text{stand}} = \frac{x_i - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \quad (1)$$

where x_{stand} is the standardized data, x_i is the experimental data, x_{min} is the minimum data value and x_{max} is the maximum data value.

The multi-layer perceptron type network is able to solve problems of high degrees of non-linearity and the resilient propagation algorithm makes the convergence process more efficient. The chosen artificial neural networks had the following architectures: supervised learning rate,

logistic activation function, 15 input variables, 2 to 3 intermediate layers, 2 output variables, 20 to 100 neurons in the intermediate layers and 200 to 600 cycles (Figure 1).

The performances of the different formations of artificial neural networks were compared according to the correlation coefficient between output data and predicted by root MSE. Six neural networks were used to classify the data (pure milk or milk adulterated with sodium bicarbonate; pure milk or milk adulterated with sodium hydroxide; pure milk or milk adulterated with hydrogen peroxide; pure milk or milk adulterated with starch; pure milk or milk adulterated with sucrose; pure milk or milk adulterated with urea). The input of the network was constructed by the standardized data and the output was represented by a vector formed by an identity matrix, where the dimension corresponds to the number of groups that compose the data used, (1.0) for pure milk and (0.1) for milk adulterated.

The samples collected from the producers were tested in the 6 networks chosen in the classification stage, in order to observe if there was any sample adulterated with any of the adulterants and their concentrations tested in this study. For this, the Neural Works program²² was used with the training data of each adulterant and the samples of producers as validation data.

Results and Discussion

Multiple linear regression analysis

With the multiple linear regression analysis, it was possible to obtain a mathematical function capable of relating the response variables to the explanatory variables, a function that was used to predict the results. For all the adulterants tested in the FTIR-ATR, multicollinearity was less than 100, that is, there was no correlation between 2 or more explanatory variables.²³ In the training stage, the mathematical model was obtained for the calibration and later the model was validated. The obtained models were evaluated based on the MSE and the correlation coefficient. The MSE defines how much the model has adapted to the data, and a value of up to 3% is considered acceptable.²⁴

The backward model was the best to detect adulteration of milk with starch, sodium bicarbonate, urea and sucrose, while the forward model was the best for milk adulterated with sodium hydroxide and hydrogen peroxide. Six equations were obtained for the estimation of milk and adulterated samples. The number of variables tested, MSE and correlation coefficient of adulterated samples are shown in Table 2.

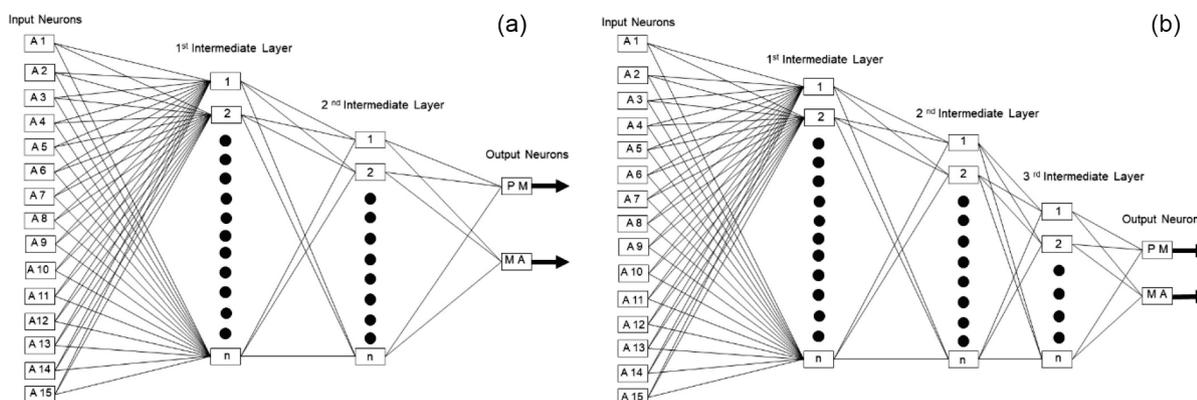


Figure 1. Configuration of the neural networks chosen for (a) neural network with 2 intermediate layers and (b) neural network with 3 intermediate layers. A1-A15: absorbances; P M: pure milk; M A: adulterated milk.

Table 2. Variables tested, MSE, correlation coefficient of adulterated samples

Group	Correlation	Training			Validation		
		VT	MSE	R	VT	MSE	R
Sodium bicarbonate	0.91	8	0.18	0.73	8	0.16	0.84
Sodium hydroxide	0.74	13	0.14	0.85	13	0.29	0.64
Hydrogen peroxide	0.60	12	0.15	0.80	8	0.29	0.48
Starch	0.73	10	0.22	0.62	8	0.27	0.53
Sucrose	0.76	4	0.21	0.62	6	0.25	0.58
Urea	0.85	9	0.16	0.80	9	0.21	0.72

VT: variable tested; MSE: mean square error; R: coefficient of multiple determination.

The spectra of milk and milk adulterated with sodium bicarbonate, sodium hydroxide, hydrogen peroxide, starch, sucrose and urea are presented in Figure 2. For the samples adulterated with sodium bicarbonate, the important variables in the validation step were: (i) lactose-related peaks, which are associated with the wavenumbers 882 (functional group with a vibration ring) and 1021 cm^{-1} (CO/CC/COO/CH functional group); (ii) protein-related peaks that are associated with wavenumbers 1232 (amide functional group III C–H/N–H); 11534 (amide functional group II N–H) and 1306 cm^{-1} (functional group CH); (iii) lipid-related peaks that are associated with wavenumbers 2305, 2845 and 2914 cm^{-1} , which have CH_2 functional group.^{8,16,25,26} The samples adulterated with sodium bicarbonate showed a correlation of 0.91; this high correlation may be associated with the variables of the equation that are associated with the absorption of this substance at 1100–1250 cm^{-1} (Table 2).

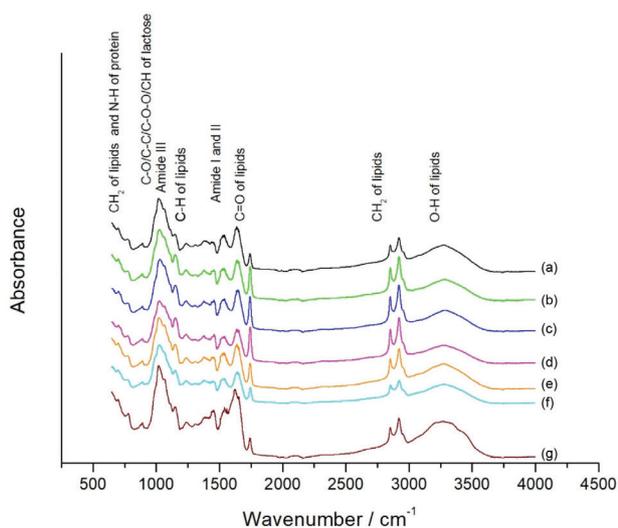


Figure 2. Spectra obtained by FTIR-ATR spectrometer from samples of (a) milk; milk adulterated with (b) 0.5% sodium bicarbonate, (c) 0.5% sodium hydroxide, (d) 0.4% hydrogen peroxide; and milk samples with 10% water and (e) 2.5% starch, (f) 4.0% sucrose, and (g) 10.0% urea.

For the samples with sodium hydroxide, 13 variables were necessary to predict the adulteration. The wavenumbers associated with lipids are 772, 2321, 2374, 2851 and 2915 (which have a CH_2 functional group); 1380 (CH functional group); 1740 (C=O functional group); and 3274 cm^{-1} (O–H functional group). Lactose-associated wavenumbers 883 (CH functional group); 1010 and 1148 cm^{-1} (C–O/C–C/O–O functional groups). Wavenumbers associated with protein are 1634 (amide functional group IC=O) and 1534 cm^{-1} (amide functional group IIN–H). The important variables to obtain the prediction function have absorption at the same wavenumbers of sodium hydroxide (1100–1200 cm^{-1}).^{8,16,25-27}

For samples adulterated with hydrogen peroxide, parameters associated with the wavenumbers 777, 1364, 1454, 1729, 2321, 2851, 2914 and 3274 (lipids); 883, 1026 and 1153 (lactose); and 1655 cm^{-1} (protein) were used to obtain the equation. Even with variables associated with the absorption of hydrogen peroxide in 2000–2500 cm^{-1} , the correlation for this adulterant was 0.60.^{8,16,18,25-27} The prediction difficulty of the function may be associated with the freeze-drying of the samples and due to the high volatility of the hydrogen peroxide, therefore, the presence of small amount of peroxide in the milk was not capable of satisfactory prediction.

For samples adulterated with starch, the following parameters were associated with the lactose (882, 1021 and 1148 cm^{-1}), lipids (767, 1734, 2851, 2914 and 3274 cm^{-1}) and protein (1232 and 1650 cm^{-1}). Only 4 parameters were tested for the sucrose-adulterated samples associated with the 1158 (lactose), 1369, 1465 and 3269 cm^{-1} (lipids) wavenumbers. The correlation of the equation for the prediction of adulteration with starch was 0.73 and for samples adulterated with sucrose, 0.76. These values may be related to the fact that starch and sucrose have many hydroxyls as well as lactose, and, therefore, absorption occurs at the same wavenumber. The absorption of the starch is in the region of 900–1200 cm^{-1} , which is associated with the glycosidic vibrations α 1–4 C–H (fold), C–O–H (fold), C–O and C–C (stretching), and sucrose absorption is in the region of 1000–1200 cm^{-1} , which is associated with the functional groups C–H, C–O (stretching) and C=O (stretching and flexing).^{8,16,17,27}

Samples adulterated with urea in the training and validation stages were tested for 9 parameters associated with the compounds and respective lactose (882, 1009, 1147 cm^{-1}), lipids (1437, 2851, 2915 cm^{-1}) and protein (1236, 1300, 1617 cm^{-1}) wavenumbers (Figure 1). The 1617 cm^{-1} variable has absorption in the same region of the urea (1600–1680 cm^{-1}), which is related to the functional groups CO, CN and NH_2 , being an important variable for the prediction of milk adulterated with urea.^{8,16,19,27}

The explanation of the dependent variable regarding the independent variable is explained by the value of the determination coefficient, therefore, for the adulterated samples in this study, sodium bicarbonate, sodium hydroxide, sucrose and urea obtained values greater than 60%, and referring to the MSE, all samples presented values lower than 3%. The models obtained were able to predict the adulteration of the samples, but for the samples adulterated with hydrogen peroxide the correlation was 60%, showing that the model has low damping capacity for this adulterant.

Artificial neural network

Artificial neural networks are techniques that use mathematical models inspired by the nervous system that are able to classify, predict and optimize data, and have the capacity to learn real data and provide this knowledge for later use. They are formed primarily of the input layer, the intermediate layer and the output layer, which are connected to each other by means of mathematical connections. Depending on the complexity of the data, the network may consist of more than one intermediate layer.²⁸

For milk adulterated with sodium bicarbonate, the neural network with 2 intermediate layers and 50 neurons in each layer obtained a classification of 80% for milk samples and 93.2% for samples of adulterated milk, with MSE of 0.27. Samples with sodium hydroxide presented an error of 0.42, in a network constructed with 2 intermediate layers and 100 neurons in each layer, where 40 and 96.6% of the milk samples were classified as milk and adulterated milk, respectively. The neural network developed for samples adulterated with sodium bicarbonate and sodium hydroxide presented high classification capacity, while the neural network did not have a high classification rate to differentiate milk from milk adulterated with sodium hydroxide, which may be related to quantity of milk samples at the validation stage, which contributed to the increase of classification error.

Samples adulterated with hydrogen peroxide and pure milk obtained 80% of classification, with an average square error of 0.46 and a network constructed with 2 intermediate layers, the first composed of 80 neurons and the second of 50 neurons. Samples adulterated with starch obtained a classification of 86.6%, while pure milk obtained 100%, with an average square error of 0.55. The developed neural networks had high classification rate for samples adulterated with hydrogen peroxide, with starch and pure milk. This fact did not occur with the analysis of multiple linear regression for samples adulterated with hydrogen peroxide, where the prediction capacity was low, which can be related to the greater capacity of generalization of the neural network.

A network constructed with 3 intermediate layers composed of 100, 50 and 20 neurons each was used for the samples adulterated with sucrose, with an average square error of 0.41. Classification obtained was 40% for adulterated samples and 100% for pure milk. For samples adulterated with urea, the net was constructed with two intermediate layers and 20 neurons in each layer; an error of 0.58 was obtained, with 100% classification for samples without adulteration and 76.6% for adulterated milk samples. The developed neural network presented

low classification rate for the samples adulterated with sucrose.

With the artificial neural network, it was possible to identify samples adulterated with sodium bicarbonate, sodium hydroxide and hydrogen peroxide with concentrations from 0.1%, from 0.5% for samples adulterated with starch, 2.0% for samples adulterated with sucrose and 6.0% for samples adulterated with urea.

The artificial neural network was used to identify possible adulterations in the samples obtained from the producers. Of the 249 samples obtained, the artificial neural networks were able to identify 3 samples adulterated with sodium hydroxide, 2 samples adulterated with starch and 1 sample adulterated with urea, corresponding to 2.4% of the samples studied.

Therefore, the chosen artificial neural networks obtained good classification rates, but also identified the presence of adulterants when used in low concentrations. The FTIR-ATR was able to generate data for prediction and classification of samples adulterated with sodium bicarbonate, sodium hydroxide, hydrogen peroxide, starch, sucrose and urea, being a fast and valuable analytical tool in the investigation of adulterants in raw milk, and can be used as an efficient screening method.

Conclusions

With the FTIR-ATR allied to the multivariate analysis it was possible to identify milk adulterated with sodium bicarbonate, sodium hydroxide, hydrogen peroxide, starch, sucrose and urea. It was possible to obtain functions to predict milk adulterated with sodium bicarbonate and urea with high correlation rates using multiple linear regression analysis. Artificial neural networks obtained classification rate higher than 76.6% for all adulterants and were able to identify sodium bicarbonate with concentration from 0.1%. When the multivariate analysis was applied to samples of producers, it was possible to identify adulterated samples. The use of FTIR-ATR allied with multivariate analysis proved to be a screening tool for the identification of adulteration in raw milk.

References

1. Jaiswal, P.; Jha, S. N.; Borah, A.; Gautam, A.; Grewal, M. K.; Jindal, G.; *Food Chem.* **2015**, *168*, 41.
2. Kolberg, D. I. S.; Presta, M. A.; Wickert, C.; Adaime, M. B.; Zanella, R.; *J. Braz. Chem. Soc.* **2009**, *20*, 1220.
3. Simionato, J. I.; Garcia, J. C.; Santos, G. T.; Oliveira, C. C.; Visentainer, J. V.; Souza, N. E.; *J. Braz. Chem. Soc.* **2010**, *21*, 520.

4. Moore, J. C.; Spink, J.; Lipp, M.; *J. Food Sci.* **2012**, *77*, 118.
5. Domingo, E. C.; Tireli, A. A.; Nunes, C. A.; Batista, A. V.; Guerreira, M. C.; Pinto, S. A.; *Talanta* **2015**, *132*, 535.
6. Gondim, C. S.; Junqueira, R. G.; Souza, S. V. C.; Ruisánchez, I.; Callao, M. P.; *Food Chem.* **2017**, *230*, 68.
7. Nascimento, C. F.; Santos, P. M.; Pereira Filho, E. R.; Rocha, F. R. P.; *Food Chem.* **2017**, *221*, 1232.
8. Santos, P. M.; Pereira Filho, E. R.; Saona, L. E. R.; *J. Agric. Food Chem.* **2013**, *61*, 1205.
9. Souza, A. M.; Poppi, R. J.; *Quim. Nova* **2012**, *35*, 223.
10. Lohumi, S. H.; Lee, S.; Lee, H.; Cho, B. K.; *Trends Food Sci. Technol.* **2015**, *46*, 85.
11. Carvalho, B. M. A.; Carvalho, L. M.; Coimbra, J. S. R.; Minim, L. A.; Barcellos, E. S.; Silva Júnior, W. F.; Detmann, E.; Carvalho, G. G. P.; *Food Chem.* **2015**, *174*, 1.
12. Liu, J.; Ren, J.; Liu, Z. M.; Guo, B. H.; *Food Chem.* **2015**, *172*, 251.
13. Chen, H.; Tan, C.; Lin, Z.; Wu, T.; *Spectrochim. Acta, Part A* **2017**, *173*, 832.
14. Rocha, R. A.; Paiva, I. M.; Anjos, V.; Furtado, M. A. M.; Bell, M. J. V.; *J. Dairy Sci.* **2015**, *9*, 3559.
15. Cassoli, L. D.; Sartori, B.; Machado, P. F.; *Braz. J. Vet. Res. Anim. Sci.* **2011**, *40*, 2591.
16. Stuart, B.; *Infrared Spectroscopy: Fundamentals and Applications*; John Wiley & Sons: Chichester, 2004.
17. Botelho, B. G.; Reis, N.; Oliveira, L. S.; Sena, M. M.; *Food Chem.* **2015**, *181*, 31.
18. Santos, P. M.; Pereira Filho, E. R.; Soana, L. E. R.; *Food Chem.* **2013**, *138*, 19.
19. Jha, S. N.; Jaiswal, P.; Borah, A.; Gautam, A. K.; Srivastava, N.; *Food Bioprocess Technol.* **2015**, *8*, 926.
20. Wicklin, R.; *Statistical Programming with SAS/IML® Software*; SAS Institute Inc.: Cary, 2010.
21. Fischer, I.; Henecke, F.; Bannes, C.; Zell, A.; *JavaNNS, Java Neural Network Simulator Version 1.1*; University of Tübingen, Germany, 2001.
22. NeuralWare; *Neural Computing: A Technology Handbook for Neural Works Professional II/Plus*; Neural Ware Inc.: Carnegie, 2000.
23. Montgomery, D. C.; Peck, E. A.; *Introduction to Linear Regression Analysis*; John Wiley & Sons: New York, 1982.
24. Rebecchi, S. R.; Vélez, M. A.; Vaira, S.; Perotti, M. C.; *Food Chem.* **2016**, *192*, 1025.
25. Lei, Y.; Qun, Z.; Zhang, Y. L.; Chen, J. B.; Sun, S. Q.; Noda, I.; *J. Mol. Struct.* **2010**, *974*, 88.
26. Pappas, C. S.; Tarantilis, P. A.; Moschopoulou, E.; Moatsou, G.; Kandarakis, I.; Polissiou, M. G.; *Food Chem.* **2008**, *106*, 1271.
27. Nicolau, N.; Xu, Y.; Goodacre, R.; *J. Dairy Sci.* **2010**, *93*, 5651.
28. Silva, I. N.; Spatti, D. H.; Flauzino, R. A.; *Redes Neurais Artificiais para Engenharia e Ciências Aplicadas*; Artliber: São Paulo, 2010.

Submitted: August 7, 2018

Published online: October 16, 2018