



Chemometric Methods for the Simultaneous Spectrophotometric Determination of Caffeine, Theobromine and Theophylline in Tea

A. HAKAN AKTAS* and HÜLYA PEKCAN

Department of Chemistry, Science and Art Faculty, Süleyman Demirel University, 32260 Isparta, Turkey

*Corresponding author: E-mail: hakanaktas@sdu.edu.tr

(Received: 23 October 2012;

Accepted: 21 August 2013)

AJC-13944

In this study, the simultaneous determination of caffeine, theobromine and theophylline in tea samples by chemometric approaches using UV spectrophotometry has been reported as a simple alternative to using separate models for each component. Spectra of caffeine, theobromine and theophylline were recorded at several concentrations within their ranges and were used to compute the calibration mixture between wavelengths 200 and 305 nm at an interval of 1 nm was used for data acquisition. Partial least squares regression and artificial neural network were used for chemometric analysis of data and the parameters of the chemometric procedures were optimized. The analytical performances of these chemometric methods were characterized by predicted residual error sum of squares (PRESS), standard error of prediction (SEP) and recoveries (%) and were compared with each other. A series of synthetic solution containing different concentrations of caffeine, theobromine and theophylline were used to check the prediction ability of the partial least squares and artificial neural network. These two methods were successfully applied to real samples, with no interference from excipients as indicated by recovery study results. The results obtained in this investigation strongly encourage us to apply these techniques for a routine analysis and quality control of the teas.

Key Words: Caffeine, Theobromine, Theophylline, Spectrometry, Multivariate calibration.

INTRODUCTION

The methylxanthines *e.g.*, caffeine (CAF), theobromine (TBR) and theophylline (TPH) (Fig. 1) are regularly consumed from a wide variety of foods and beverages with coffee and tea being most popular worldwide¹.

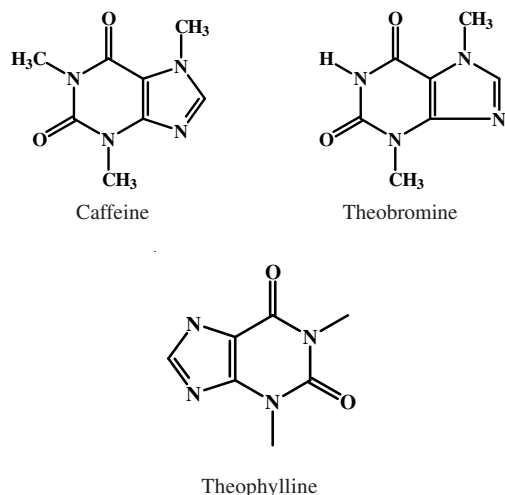


Fig. 1. Structures of the two studied compounds

These three alkaloids can produce extensive bioactivity. However, excessive intake of them will cause many undesirable side effects². Food is most important exposure factor to humans for caffeine and theobromine³. It is considered that food can play an important role in methylxanthine toxicity, not only in a direct way by providing large doses of caffeine and theobromine from beverages or chocolate, but also by accelerating the dissolution and absorption rates of theophylline from sustained release theophylline preparations⁴. Therefore, it is important to determine accurately the contents of these three alkaloids in various foods in order to estimate a person's daily intake.

Although a great variety of analytical techniques have been applied to the simultaneous determination of these three alkaloids in various matrices, liquid chromatography (LC) is the most frequently used method nowadays. Some approaches liquid chromatography methods have been proposed in combination with solid surface room temperature phosphorimetric detection⁵, UV detection⁶ and photodiode array detection⁷ and capillary electrophoresis⁸ were applied to determine the caffeine content in tea. However, UV-visible spectrophotometric determination is preferred because it is possible to obtain rapidly high accuracy and reproducibility from small samples

using a relatively simple and inexpensive procedure when compared a chromatographic techniques. Unfortunately, spectrophotometric methods simultaneously determine methylxanthines have been poorly developed given the high spectral overlap of these compounds (Fig. 2) and the matrix interference usually associated with food products. To avoid interferences⁹, developed an extractive-spectrophotometric method of pre-treated samples where they determine caffeine and theobromine in cocoa beans by extracting the former into chloroform at pH 12.5 with theobromine remaining in the aqueous phase. While this procedure presented good analytical results, it involved several preparation steps and methylxanthines determination was achieved in separated stages. Since the absorption spectra of these compounds overlap in the working wavelength range, it is not possible to determine simultaneously caffeine, theobromine and theophylline in their mixture by conventional methods. In order to solve this problem the two chemometric methods (partial least squares and artificial neural network) were applied to the quantitative analysis of the three component mixture system of the subject matter compounds. The chemometric methods are an effective way to analyze simultaneously several analytes¹⁰. The abilities of different chemometric methods to resolve mixtures of different compound with overlapped spectra have ben widely utilized. The main advantage of multicomponent analysis by chemometric methods is the rapid determination of the components in mixtures avoiding to the overlapped signals. Recently, the chemometric methods such as classical least squares (CLS), principal component analysis (PCA) and partial least squares (PLS) have found increasing applications for multicomponent determination^{11-14,15-17}. The literature, tea and coffee as caffeine, theobromine and theophylline article reached a large number of PLS, but the work could not be observed with artificial neural network (ANN). Therefore, in addition to the PLS in this study and achieved positive results by trying to ANN.

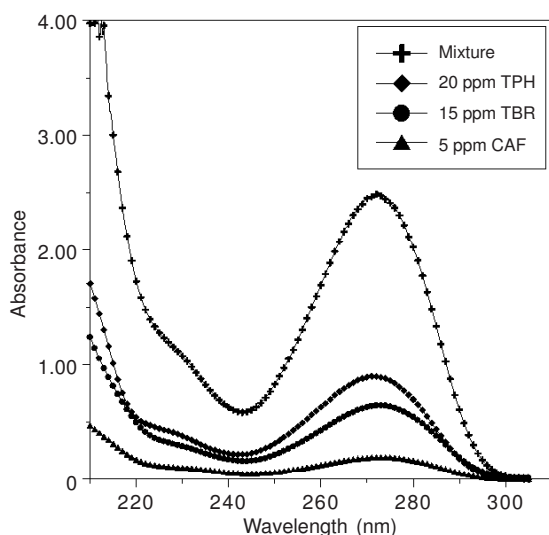


Fig. 2. Absorption spectra of 5 ppm caffeine, 15 ppm theobromine, 20 ppm theophylline and their mixture in 0.1 M HCl

EXPERIMENTAL

An absorbance measurement was carried out by using a Shimadzu (Model UV-1700) UV-visible spectrometer (Shimadzu,

Kyoto, Japan), equipped with 1 cm matched quartz cells and was used for spectrometric measurements.

Chemometric software: Application of PLS algorithms was supported by the software package "Minitab® 16". The software is dedicated to both multivariate analysis and experimental design and is equipped with several multivariate methods. It allows to optimize the calibration models and to develop validation procedures. The back-propagation neural network algorithm three layers were used in MATLAB (version 7.0, Math Work Inc.) using NN toolbox. All programs were run on a Pentium, personal computer, with windows XP home edition.

Caffeine, theobromine and theophylline were kindly donated by the pharmaceutical industries and were used without further purification. All solvents and reagents were of analytical reagent grade (Sigma and Fluka). Stock solutions of caffeine, theobromine and theophylline 10 mg/100 mL were prepared by dissolving the compounds in distilled water. A Britton-Robinson buffer solution pH 12 was prepared by mixture consists of 0.0286 M citric acid, 0.0286 M potassium dihydrogen phosphate, 0.0286 M boric acid, 0.0286 M veronal and 0.0286 M hydrochloric acid with a solution 0.2 M of sodium hydroxide.

Methods

Tea sample preparation: One gram of tea sample was boiled for 20 min in 50 mL of water. The obtained well-steeped tea was then filtered and the filtrate volume adjusted 50 mL. 10 % of 0.5 mL lead acetate was added to a 25 mL aliquot of sample and the mixture stirred 5 min at room temperature. Final solution was filtered and 0.1 mL of sodium carbonate 10 % was added to filtrate to remove lead acetate excess. Mixture was filtered and water added to raise volume to 50 mL. Samples prepared following the described procedure showed no variances in methylxanthine concentration up to 3 weeks.

Sample preparation for spectrophotometric analysis: In 10 mL volumetric flasks, an aliquot of the pre-treated sample of tea was introduced, followed by the appropriate addition of standards of caffeine, theobromine and theophylline, addition of 2.0 mL Britton -Robinson buffer solution (pH 12.0), diluted to the mark with the doubly distilled water and mixed well. The absorbance spectrum was measured with respect to a reagent blank between 200 and 305 nm at a 1 nm interval and all these data were recorded and uses for calculations.

RESULTS AND DISCUSSION

Multivariate analysis: Partial least square is factor analysis method, based on a two-stage procedure; a calibration step, in which a mathematical model is built by using component concentrations and spectral data from a set of references, followed by a prediction step in which the model is used to calculate the concentrations unknown sample from its spectrum. These methods are also called "factor methods" because they transform the original variables into a smaller number of orthogonal variables called factors or principal components (PCs), which are linear combinations of the original variables. When multivariate calibration approaches are applied in spectrophotometric multicomponent analysis, a relationship between spectral and concentration data from reference samples, representing the variables of the system, is established. A new matrix constituted by the new variables principal components

and scores is built. The calculation of this new matrix is planned by algorithm specific to the regression method adopted. The most used regression methods are PLS. The theory of such techniques has been fully described by several authors¹⁸⁻²¹.

Artificial neural network is a multivariate calibration method used mainly for modeling non-linear data, although, some applications use the neural network for modeling linear data. It is important to state that this method is computationally more complex than linear methods, they have limitation of being prone to over fitting and they heavily depend on amount and quality of data available. In many cases the principal disadvantage of the neural network is time required for its training.

Wavelength selection: Multivariate calibration methods have generally been considered as full-spectrum in the processes of calibration and prediction. But, most often, selecting the undesired regions of the spectrum evaluated is provided more noise in the analytical results^{22,23} showed that accuracy can be improved by careful wavelength selection. So far, various criteria have been developed to allow for wavelength selection. In the present work, the UV-spectra of caffeine, theobromine and theophylline solutions were recorded, in the wavelength range of 200-305 nm with 1 nm interval (Fig. 2). Spectrum as shown the simultaneous determination of the related compounds in samples is not possible by using classical spectrophotometric approaches. In this study, without any separation step using PLS and ANN chemometric approaches caffeine, theobromine and theophylline is focused 3 mixtures of quantitative resolution.

Multivariate methods: The first step in simultaneous determination of the ternary mixture of teas by multivariate calibration methods involves constructing the calibration matrix for ternary mixture of caffeine, theobromine and theophylline. Twenty-five ternary mixtures were selected by random design as the calibration set. The composition of the samples was randomly designed in order to obtain non-correlated concentration profiles (Table-1). In order to minimize the correlation between concentration vectors, the correlation coefficient matrix is considered as a criterion. The calibration model in each chemometric method was validated with 15 synthetic mixtures set containing the food under study in different proportions selected randomly. The predictive abilities of PLS and ANN were examined for simultaneous determination caffeine, theobromine and theophylline in sample mixtures. The common requirement for all mentioned methods is that unknown samples and standards be of the same nature.

Partial least squares calibration: In the UV-visible spectra, the absorbance data (A) and concentration data (C) are mean centered to give data matrix A_0 and vector C_0 . The orthogonalized PLS algorithm has the following steps. The loading weight vector W has the following expression:

$$W = \frac{A_0^T C_0}{C_0^T C_0}$$

The scores and loadings are given by:

$$t_1 = A_0 W; p_1 = \frac{A_0^T t_1}{t_1^T t_1}; q_1 = \frac{C_0^T t_1}{t_1^T t_1}$$

The matrix and vector of the residuals in A_0 and C_0 are:

$$A_1 = A_0 - t_1 p_1^T$$

No.	CAF	TBR	TPH	No.	CAF	TBR	TPH
1	5	4	3	14	30	16	18
2	10	8	6	15	5	8	3
3	15	12	9	16	10	12	6
4	20	15	12	17	15	16	9
5	25	18	15	18	20	20	12
6	10	4	6	19	25	24	15
7	15	8	9	20	30	28	18
8	20	12	12	21	10	2	3
9	25	16	15	22	15	4	6
10	30	19	18	23	20	6	9
11	15	4	9	24	25	8	12
12	20	8	12	25	30	10	15
13	25	12	15	-	-	-	-

$$C_1 = C_0 - t_1 q_1^T$$

From the general linear equation, the regression coefficients were calculated by;

$$b = W(P^T W)^{-1} q$$

$$a = C_{\text{mean}} - A_{\text{mean}}^T b$$

The builded calibration equation is used for the estimation of the compounds in the samples.

Artificial neural network calibration: A feed-forward ANN model with three layers of nodes was constructed as in Fig. 3. The artificial neuron is the building component of ANN designed to simulate the function of biological neuron. The arriving signals, called inputs, multiplied by the connection weighted (adjusted) are first summed (combined) and then passed through a transfer function to output that neuron. The activation function is the weighed sum of the neuron's inputs and the most commonly used transfer function is sigmoid function (Fig. 3).

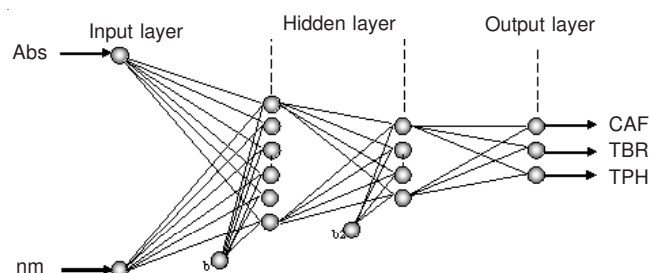


Fig. 3. Network architecture used in the spectrophotometric determination

The logistic function was used as the activation function in a neural network. The training and testing data sets must be normalized into a range 0.1-0.9. The input and the output data sets were normalized by using following equation:

$$X_N = 0.1 + \frac{0.8(X - X_{\min})}{(X_{\max} - X_{\min})} \quad (1)$$

where X_N is normalized value of a variable (the network input or the network output), X is a original value of a variable and X_{\max} and X_{\min} are maximum and minimum original values of the variables, respectively. In order to produce sufficient data for training and testing of the model shown in Fig. 3. Fifteen different standard solutions were prepared using different

caffeine, theobromine and TBH concentrations and each standard solution was subjected to spectrophotometric determination. Randomly chosen 200 data pairs from these 1050 data pairs were used in the training of the neural network and the rest of the data were used in the testing. The root mean square error values were calculated from following equation to prove quantitatively the accuracy of the testing results of neural network models:

$$\text{RMS} = \sqrt{0.5N^{-1} \sum_{i=1}^N (X'_i - X_i)^2} \quad (2)$$

where N is the number of testing data and X'_i is target value.

MATLAB 7.0 software was used to construct ANN models which have sigmoidal logistic function with back propagation of error algorithm. For this neural network modeling an input layer, one or two hidden layers and output layer were used.

To obtain the best network performance, the optimal network architecture and parameters must be chosen. Studies of the network structure include the selection of the number of layers and number of nodes in each layer. The number of layers used for this neural network modeling was four, *i.e.*, an input layer, one or two hidden layers and an output layer. As can be seen from Fig. 1, two neurons were used in the input layer, which were the absorbance and wavelength (nm). The absorbance and wavelength of the solution were considered as independent variables of the spectrophotometric method. Therefore, these variables were used as input variables in the network architecture. Table-1 showed the concentration of standard solutions (or output data of the network).

Various neural network models, which have the logistic function, were trained and tested. In this step, the number of the hidden layer units of the network was determined by performance evaluating of the network models defined in Table-2. According to RMS errors given in Table-2, the NN3 2-3-8-3 model, which performs best on a testing data set, were selected as neural network model to predict the teas concentrations.

Chemometric parameters: The application competence of a calibration model can be explained in several ways. These results can be examined numerically. One of the best ways to do this, by examining the predicted residual error sum of squares (PRESS). To calculate PRESS computed the errors between the expected and predicted values for all the samples, square them and sum them together.

$$\text{PRESS} = \sum_{i=1}^n (C_i^{\text{added}} - C_i^{\text{found}})^2$$

Strikingly speaking, this is not a correct way to normalize the PRESS values when not all of the data sets contain the same number of samples. If want correctly compare PRESS values for data sets that contain differing numbers of samples, should convert to standard error of prediction (SEP), which is given by following formula.

$$\text{SEP} = \sqrt{\frac{\sum_{i=1}^n (C_i^{\text{added}} - C_i^{\text{found}})^2}{n-1}}$$

where C_i^{added} the added concentration of drug is, C_i^{found} is the found concentration of drug and n is the total number of the synthetic mixtures. The SEP can provide a good measure of how well, on average, the calibration model performs. Often, however, the performance of the calibration model varies depending on the analyte level.

In the application of two chemometric techniques to the synthetic mixtures containing teas in variable compositions, the mean recoveries and relative standard deviations for PLS and ANN were found to be 100.26 and 3.05 %, 100.36 and 1.12 %, respectively for caffeine, 97.97 and 5.71 %; 100.08 and 1.51 %, respectively for theobromine, 98.96 and 4.39 %; 100.02 and 1.48 %, respectively for theophylline (Table-3). These experimental results indicate that these two methods are suitable for simultaneous determination of the above mentioned compounds in samples.

According to the added concentration and the concentration found in samples, the SEP and PRESS values of PLS and ANN techniques were calculated 0.4152 and 0.1851; 0.1724 and 0.0342, respectively for caffeine, 0.5114 and 0.1539; 0.2616 and 0.0237, respectively for theobromine and 0.3821 and 0.1466; 0.1460 and 0.0215, respectively for theophylline (Table-4).

The linear regression analysis of the added concentration and the concentration found in the synthetic mixtures were realized for each drug and for each calibration technique. In this regression analysis, the correlation coefficient (r), intercept, slope and relative standard deviation values were found satisfactory for the proposed chemometric techniques in Table-4. In order to test the proposed calibrations, an independent set of the validation set in Table-4 was analyzed and used for the calculations of the standard error of prediction (SEP) in the validation step. The standard error of calibration (SEC) and the errors of prediction (SEP) for n = 15 calibration-prediction samples were calculated PLS and ANN method and their values were summarized in Table-3. In same table, the results of linear regression analysis were applied to relationships

TABLE-2
COMPARISON OF THE PERFORMANCES OF THE NEURAL NETWORK MODELS

Model	RMS error					
	CAF		TBR		TPH	
	Training	Testing	Training	Testing	Training	Testing
NN1 2-2-7-3	0.19272	0.24185	0.22387	0.02674	0.27568	0.21883
NN2 2-2-8-3	0.00517	0.00894	0.00302	0.00474	0.02382	0.02754
NN3 2-3-8-3	0.00026	8.06×10^{-5}	0.00039	6.02×10^{-5}	0.00055	0.00032
NN4 2-3-9-3	0.04218	0.10280	0.02516	0.03081	0.04109	0.19625
NN5 2-4-5-3	0.03812	0.02802	0.17960	8.06×10^{-5}	0.00787	0.09889

TABLE-3
COMPOSITION OF SYNTHETIC SAMPLES, THEIR RECOVERIES BY PLS AND ANN MODELS

Mixtures added ($\mu\text{g/mL}$)			Recovery (%)					
			PLS			ANN		
CAF	TBR	TPH	CAF	TBR	TPH	CAF	TBR	TPH
5	14	3	99.21	89.55	91.57	99.84	99.68	96.92
10	14	3	105.42	97.61	90.77	102.02	98.24	102.40
15	14	3	98.24	105.46	102.28	99.64	102.28	100.21
20	14	3	95.61	103.98	106.77	100.02	100.20	99.98
25	14	3	102.30	97.86	95.75	101.20	99.81	98.88
4	3	19	102.16	90.04	96.47	100.08	98.76	100.02
4	6	19	100.74	101.46	106.05	100.28	100.84	100.08
4	9	19	104.29	98.94	98.39	98.12	102.26	99.68
4	12	19	97.19	94.57	98.46	102.01	97.85	98.80
4	15	19	102.83	98.25	99.37	101.26	98.98	100.06
24	2	4	98.04	104.09	98.72	98.94	101.29	100.24
24	2	8	99.60	96.75	99.26	102.01	102.64	100.24
24	2	12	97.43	106.74	101.92	99.98	98.68	99.82
24	2	16	101.66	89.01	99.33	99.88	100.02	100.08
24	2	20	102.21	95.18	99.36	100.26	99.66	98.66
Mean	–	–	100.26	97.97	98.96	100.36	100.08	100.02
RSD ^a	–	–	3.05	5.71	4.39	1.12	1.51	1.48

^aRelative standard deviation.

TABLE-4
STATISTICAL PARAMETERS IN THE
CALIBRATION-PREDICTION

Parameter	Method	CAF	TBR	TPH
PRESS	PLS	0.1724	0.2616	0.1460
	ANN	0.0342	0.0237	0.0215
SEP	PLS	0.4152	0.5114	0.3821
	ANN	0.1851	0.1539	0.1466
r	PLS	0.9979	0.9921	0.9923
	ANN	0.9997	0.9992	0.9996
Intercept	PLS	0.0755	0.0309	0.0111
	ANN	0.0183	0.0194	0.0187
Slope	PLS	0.9940	0.9864	0.9981
	ANN	1.0059	0.9980	0.9976
RSD	PLS	3.05	5.71	4.39
	ANN	1.12	1.51	1.48

between actual and predicted concentrations in the calibration and validation sets. Their statistical results were presented in Table-4. As can be seen, all the statistic values indicated that all techniques are convenient for the determination of caffeine, theobromine and theophylline in synthetic mixtures.

Applications: In order to assess the applicability of the proposed method to the analysis of real samples, they were applied to the determination of these methylxanthines in tea sample. Six replicate measurements were made. The results are shown in Table-5. The good agreement between the results and the label claims indicates the successful applicability of the proposed procure for simultaneous determination of caffeine, theobromine and theophylline in real sample. To check the validity of the proposed method, after the addition of the known amounts of caffeine, theobromine and theophylline to the real samples, we found that the amount of these samples did not change. Moreover we compare the spectra obtaining from the mixture caffeine, theobromine and theophylline in standard and real sample formulation solutions that showed similar patterns in their spectra. These findings indicate that excipients placed in real sample preparation did not interfere in the measurement of caffeine, theobromine and theophylline in real sample formulation. Tea samples were prepared to section "Tea

TABLE-5
COMPARISON OF RESULTS OF CAF, TBR AND TPH
ANALYSES OBTAINED TWO DIFFERENT METHODS

Teas	PLS ^a	ANN ^a	HPLC ^b
CAF/mg g ⁻¹ mean \pm	60.894 \pm 0.01	63.856 \pm 0.42	19.60
TBR mg g ⁻¹ mean \pm	1.338 \pm 0.31	1.430 \pm 0.25	0.41
TPH mg g ⁻¹ mean \pm	1.427 \pm 0.62	1.472 \pm 0.12	nd ^c

^aResults obtained are average of six experiments for each technique.

^bKhanchi *et al.*, method. ^cNot detected.

sample preparation" samples were analyzed using UV-VIS spectrophotometry. The PLS and ANN methods proposed in this study was compared with the high performance liquid chromatographic method of Khanchi *et al.*¹⁵ methods. The results obtained PLS and ANN solution and the other method are shown in Table-5.

Conclusion

Two chemometric technique in spectrometric analysis, PLS and ANN were proposed for the simultaneous determination of caffeine, theobromine and theophylline in their ternary mixtures. These techniques were applied with great success to commercial teas. The resolution of highly overlapping drug mixtures was achieved by the use of PLS and ANN techniques. A selection of working wavelength having high correlation values with concentration due to interference coming from matrix sample or additional analytes outside the working range. Comparison of the results of the ANN method with those of the revealed the lower prediction errors and higher correlation coefficients for ANN. Meanwhile, the data acquisition with ANN was easier than that with the PLS method. The proposed chemometric techniques can be applied for the routine analysis of food without any a priori chemical separation and without time consuming.

ACKNOWLEDGEMENTS

The authors thank to Research Council of Süleyman Demirel University of Turkey for financial support of this investigation.

REFERENCES

1. L.L. Martinez, P.L.L. de-Alba, R.G. Campos and L.M.D.L. Rodriguez, *Anal. Chim. Acta*, **493**, 83 (2003).
2. Q.C. Chen, S.F. Mou, X.P. Hou and Z.M. Ni, *Anal. Chim. Acta*, **371**, 287 (1998).
3. C.S. Yang, P. Maliakal and X. Meng, *Ann. Rev. Pharm. Toxicol.*, **42**, 25 (2002).
4. B. Stavaric, *Food Chem. Toxicol.*, **26**, 645 (1998).
5. Y.G. Campiglia, J.J. Laserna, A. Berthod and J.D. Winefordner, *Anal. Chim. Acta*, **244**, 215 (1991).
6. S.X. Candeias, E. Gallardo and A.C. Matos, *Food Anal. Methods*, **251**, 256 (2007).
7. Y.G. Zuo, H. Chen and Y.W. Deng, *Talanta*, **57**, 307 (2002).
8. H. Hideki, M. Toshihira and K. Katsunori, *J. Chromatogr. A*, **758**, 332 (1997).
9. L. Shufen, J. Berger and S. Hartland, *Anal. Chim. Acta*, **409**, 412 (1990).
10. D. Galeano, A. Guiberteau, O. Burguillas and F. Salinas, *Analyst*, **122**, 513 (1997).
11. Y.N. Ni and X.F. Gong, *Anal. Chim. Acta*, **354**, 163 (1997).
12. D.I. Meras, M. Pena, E. Mansilla and F. Salinas, *Analyst*, **118**, 807 (1993).
13. O. Aksu, A. Bozdogan and G.K. Kunt, *Anal. Lett.*, **31**, 859 (1998).
14. Y. Ni, G. Zhang and S. Kokot, *Food Chem.*, **89**, 465 (2005).
15. A.L. Khanchi, M.K. Mahani, M. Hajihosseini, M.G. Maragheh, M. Chaloosi and F. Bani, *Food Chem.*, **1062**, 1068 (2007).
16. T.S.T. Mansor, Y.B. Che Man and A. Rohman, *Food Anal. Methods*, **365**, 372 (2011).
17. A. Rohman and Y.B. Che Man, *Food Anal. Methods*, **155**, 162 (2011).
18. J. Zupan and J. Gastreiger, *Neural Networks in Chemistry and Drug Design*, VCH (1993).
19. H. Wold, *Research Papers in Statistics*. Wiley, New York (1996).
20. E. Dinç, D. Balenau, G. Ioele, M. De Luca and G. Ragno, *J. Pharm. Biol. Anal.*, **48**, 1471 (2008).
21. M. De Luca, F. Oliveria, G. Ioele and G. Ragno, *Chemometrics Intelligent Lab. Syst.*, **96**, 14 (2009).
22. P.J. Brown, *J. Chemometrics*, **151**, 161 (1992).
23. D.T. Rossi and H.L. Pardue, *Anal. Chim. Acta*, **153**, 161 (1985).