DISCRIMINATION OF BRANDS OF STRONG AROMA TYPE LIQUORS USING SYNCHRONOUS FLUORESCENCE SPECTROSCOPY AND CHEMOMETRICS METHODS

Z.-W. Zhu [1*], G.-Q. Chen [1], Y.-M. Wu [1], Y. Xu [2], T. Zhu [1]

[1] *School of Science, Jiangnan University, Wuxi, 214122, China; e-mail: zhuzhuowei2004@163.com*
[2] *School of Biotechnology, Jiangnan University, Wuxi, 214122, China*

The application of synchronous fluorescence spectroscopy combined with chemometrics using pretreated spectra was explored to develop a rapid, low-cost, and nondestructive method for discriminating between brands of different strong aroma type liquors. Principal component analysis, partial least square discriminant analysis, support vector machine, and back-propagation artificial neural network techniques were used to classify and predict the brands of liquor samples. Compared with the other models, the SVM model achieved the best results, with an identification rate of 100% for the calibration set, and 96.67% for the prediction set. The overall results showed that synchronous fluorescence spectroscopy with an efficient chemometrics method can be used successfully to identify different brands of liquor.

Keywords: chinese liquor, synchronous fluorescence spectroscopy, discrimination of liquors, chemometrics of discrimination.

ПРИМЕНЕНИЕ СИНХРОННОЙ ФЛУОРЕСЦЕНТНОЙ СПЕКТРОСКОПИИ И ХЕМОМЕТРИЧЕСКИХ МЕТОДОВ ДЛЯ РАСПОЗНАВАНИЯ АЛКОГОЛЬНЫХ НАПИТКОВ С СИЛЬНЫМ АРОМАТОМ

Z.-W. Zhu [1*], G.-Q. Chen [1], Y.-M. Wu [1], Y. Xu [2], T. Zhu [1]

УДК 535.372

[1] *Научная школа Университета Цзяннань, Уси, 214122, Китай; e-mail: zhuzhuowei2004@163.com*
[2] *Школа биотехнологии Университета Цзяннань, Уси, 214122, Китай*

Исследованы возможности использования синхронной флуоресцентной спектроскопии совместно с хемометрией применительно к предварительно обработанным спектрам для создания быстрого, экономичного и неразрушающего метода распознавания брендов различных типов ликеров с сильным ароматом. Для классификации и определения марок образцов использовались методы главных компонент, наименьших квадратов, опорных векторов и искусственных нейронных сетей обратного распространения. По сравнению с другими моделями модель опорных векторов позволила достичь наилучших результатов со степенью идентификации 100 % для калибровочного набора и 96.67 % для прогнозируемого набора. Показано, что синхронная флуоресцентная спектроскопия с эффективным методом хемометрии может быть успешно использована для идентификации марок ликеров.

Ключевые слова: китайский ликер, синхронная флуоресцентная спектроскопия, распознавание напитков, хемометрия распознавания.

Introduction. Chinese liquor is a distilled alcoholic beverage popular in China because of its unique flavor and excellent taste. On the basis of its aroma characteristics, the types of chinese liquor can be classified as: strong, light, sauce, miscellaneous aroma type, or rice [1]. Of these, the sales of the strong aroma type of liquor have been greatest [2, 3]. These strong aroma types of liquor vary according to their geographical origin, such as Sichuan, Jiangsu, or Anhui with liquor from Sichuan being the most famous and typical. Many famous brands of chinese liquor, such as "Wuliangye," "Jiannanchun," and "Luzhoulaojiao" are distilled and bottled in this region. To some extent, a good brand can guarantee the quality of the product

and so command a higher market price. Therefore, research has focused on identifying the different brands of Sichuan strong aroma type liquor to ensure their authenticity.

In recent years, the quality of chinese liquor has been assessed using chemical analysis tools such as gas chromatography–mass spectrometry (GC-MS) [4, 5], gas chromatography–olfactometry (GC-O) [6], and solid-phase microextraction (SPME) [7]. These techniques can accurately measure the chemical components of chinese liquor but require complicated pre-treatment procedures which are time-consuming. The colorimetric artificial nose has recently been developed and used to discriminate between the different aroma styles of chinese liquor [8, 9]. However, the equipment is costly, inconvenient, and artificially affected [10]. Compared with chromatographic and other techniques, spectroscopic techniques such as near-infrared spectroscopy (NIR) [11, 12], Fourier transformation near-infrared (FTIR) [13], and fluorescence spectroscopy [14, 15] are simple, rapid, and nondestructive. Fluorescence spectroscopy has the advantages of high sensitivity and selectivity with one of its variants, synchronous fluorescence (SF), being obtained by simultaneously scanning the excitation and emission monochromators at a constant offset. Synchronous fluorescence has been widely used because of its advantages of achieving narrow bandwidth and avoiding Rayleigh scattering [16, 17]. The technique has been widely used with chemometric methods to classify alcoholic beverages, such as brandy [18], juniper-flavored spirit [19], fruit spirit [20], plum spirit [21], and beer [22]. Chinese liquor has a higher ethanol content (normally 40–55 vol.%) than other alcoholic beverages [1]. Its aroma can be developed during the fermentation, distillation, and aging processes.

The present study aims to use synchronous fluorescence spectroscopy combined with chemometrics analysis to discriminate between six brands of Chinese strong aroma type liquor produced in Sichuan province. The total SF spectrum technique will be used to construct the fingerprints of these brands. Principal component analysis (PCA) will first be applied to investigate the characteristic parameters of the total synchronous fluorescence spectra of the liquors. Three supervised pattern recognition algorithms, partial least square discriminant analysis (PLS-DA), error back-propagation neural network (BP-NN), and support vector machines (SVM), will then be used to develop a robust identification model. PLS-DA is a linear algorithm, and BP-NN and SVM are both nonlinear algorithms. Some parameters of the algorithms will be optimized by cross-calibration and the model performance evaluated by comparing the discrimination rates using the calibration and prediction sets.

Experimental. For the present study, 192 samples of six different brands of Sichuan liquor with a strong aroma were used as the calibration and prediction sets. The calibration set consisted of 132 samples (22 from each brand), with the prediction set consisting of 60 samples (10 from each brand). The manufacturers produced and provided all the liquor samples in 2014. The brand name, ethanol content, raw materials, geographical origin, and code for the six brands are listed in Table 1. Before analysis, the samples were stored at 20°C.

All the SF spectra were recorded using an FLSP920 spectrofluorometer (Edinburgh Instruments Ltd., Edinburgh, UK) equipped with a Xenon lamp, and an R928 photomultiplier tube (Hamamatsu Photonics KK, Hamamatsu City, Japan). A 10×10×45 mm cuvette was used for SF spectra acquisition using right-angle geometry. The spectra were measured in the excitation range from 200 to 600 nm, at increments of 2 nm with a synchronous interval (the interval between the excitation and emission wavelengths) that varied from 10 to 160 nm in steps of 5 nm. The excitation and emission slits were both set at 5 nm. The data obtained were exported in ASCII format then transferred to an Intel Core 2 ™ PC for subsequent manipulation.

To obtain full and effective information from the spectral data and to eliminate as much spectral deviation from the offset, and drift as possible, the spectra need to be preprocessed using mathematical methods. In the present study, the raw SF spectra were first pretreated using the Savitzky–Golay algorithm (nine data points, three-order polynomial) to smooth the spectra. Then, the smoothed spectra by first derivative were used in the model. In this work, seven smoothing points and two orders of polynomial were used to obtain the first derivative of the spectra.

TABLE 1. Details of the Brands of Chinese Liquor Samples from Sichuan Studied

Liquor code	Brand	Geographical origin	Raw material	$C_{Ethanol}$, % v/v
WU	Wuliangye	Yibing, Sichuan	Sorghum, rice, wheat, corn, sticky rice	52
JI	Jiannanchun	Mianzhu, Sichuan	Sorghum, rice, wheat, corn, sticky rice	52
LU	Luzhoulaojiao	Luzhou, Sichuan	Sorghum, rice, wheat	53
TU	Tuopai	Suining, Sichuan	Sorghum, rice, wheat, corn, barley, sticky rice	53
XU	Xufu	Yibing, Sichuan	Sorghum, wheat, rice, corn, sticky rice	52
QU	Quanxing	Chendu Sichuan	Sorghum, barley, wheat, pea	52

PCA is an unsupervised method for describing the unique variances using linear combinations of the original variables with these PCs (principal components) being orthogonal [23]. Models established with data processed by PCA have been reported as being superior to those established using the original data [24]. In this work, to visualize the clustering trends of all the samples, a score plot was obtained using the two most important principal components from PCA based on data from the first derivative pre-processing.

Partial least square discriminant analysis (PLS-DA), a variant of partial least squares regression (PLS) [25], is a classification method where the response variable is categorical expressing the class membership of the statistical units [26]. The classification of the Chinese liquor samples according to brands was based on the threshold error of recognition. In the present study, for all models these were set as ±0.3. The prediction performance of the PLS model was evaluated using the coefficient of determination (R^2) and the root mean square error (RMSE).

The SVM method was first proposed by Vapnik based on the principle of structural risk minimization [27]. It functions by constructing hyperplanes in a multidimensional space that separate the cases of different class labels. The transformation into higher-dimensional space is implemented by a kernel function [28]. Generally, there are three classic kernel functions: polynomial, radial basis function (RBF), and sigmoid. Compared with other kernel functions, the RBF kernel is the best choice, without prior experienced knowledge. Not only can it handle the linear and nonlinear relationships between the class labels and the spectral data but it can also reduce the computational complexity of the training procedure to provide a good performance under general smoothness assumptions [29]. Therefore, only the RBF kernel function was used in this work.

To obtain a good performance when using RBF kernels in the SVM model, the penalty parameter, C, and kernel parameter, γ, must be optimized; C determines the trade-off between minimizing the training error and minimizing the model complexity; γ implicitly defines the bandwidth of the RBF kernel function. In this work, a "grid-search" on parameters C and γ was used by cross-validation.

In the present study, the important supervised pattern recognition method, BP-NN, was used [30]. Some parameters, such as learning rate factor, momentum factors, and initial weights, exert a certain influence on the performance of BP-NN models. The selection of parameter values was evaluated using the minimal RMSE value of the discrimination model. After several attempts, the optimized parameters for the BP-NN algorithm were as follows: three layers (an input layer, a hidden layer, and an output layer) were used to construct the discrimination model; the number of neurons in the hidden layer was set to 8; the learning rate factor and momentum factor were all set to 0.1; the initial weights were set to 0.3; the training function was based on the Levenberg-Marquard optimum algorithm [31]; the permitted training error was set to 0.001; and the maximal times of training were set to 10,000.

All the calculations were performed using MATLAB 2013a (The Math Works, Natick, MA, USA). The SVM model was completed using the LIBSVM toolbox (National Taiwan University, Taipei, Taiwan).

Results and discussion. *Total synchronous fluorescence spectra.* Figure 1 shows the average total SF spectra as contour maps for the six different brands of liquor. The contour maps reflect the specificity of intrinsic fluorophores and their microenvironments in the liquor. The spectral characteristics of the six brands were similar, with only slight difference. For example, some samples had several common fluorescence peaks: II ($\Delta\lambda/\lambda_{ex} \approx 35/310$ nm), IV ($\Delta\lambda/\lambda_{ex} \approx 75/468$ nm), and V ($\Delta\lambda/\lambda_{ex} \approx 80/350$ nm). Peak I ($\Delta\lambda/\lambda_{ex} \approx \approx 55/254$ nm) was found in the spectra of the WU, JI, LU, TU, and XU brands and can be ascribed to the ethanol-water mixture [32]. Because the fluorescence intensity was too weak in WU and JI, this fluorescence peak position was not reflected in the contour plots. Peak I and peak III ($\Delta\lambda/\lambda_{ex} \approx 40/366$ nm) were also hard to find in QU. Except for peak I, the other peaks could be attributed to the presence of flavor components in the liquor samples: esters, acids, aldehydes, and phenols. Each of these components plays a role in characterizing the different brands of liquor.

Because these liquor samples have the same flavor, they also had similar chemical components, so the position of their fluorescence peaks was also basically the same. Though these contour plots look very similar, the positions of the wavelength of the excitation maxima and intensities differed. These differences in spectral characteristics arise because of the different fluorophore concentrations and molecular environments in liquors of different botanical origin. This also implied that the production technology and raw materials used by the different manufacturers had certain differences. Thus, the total SF is sufficiently well correlated with the quality of liquor samples to be used for distinguishing samples of liquor from different manufacturers. However, it is difficult to determine the differences between brands using the naked eye, so chemometrics methods were introduced to help identify the liquor brands automatically and accurately.
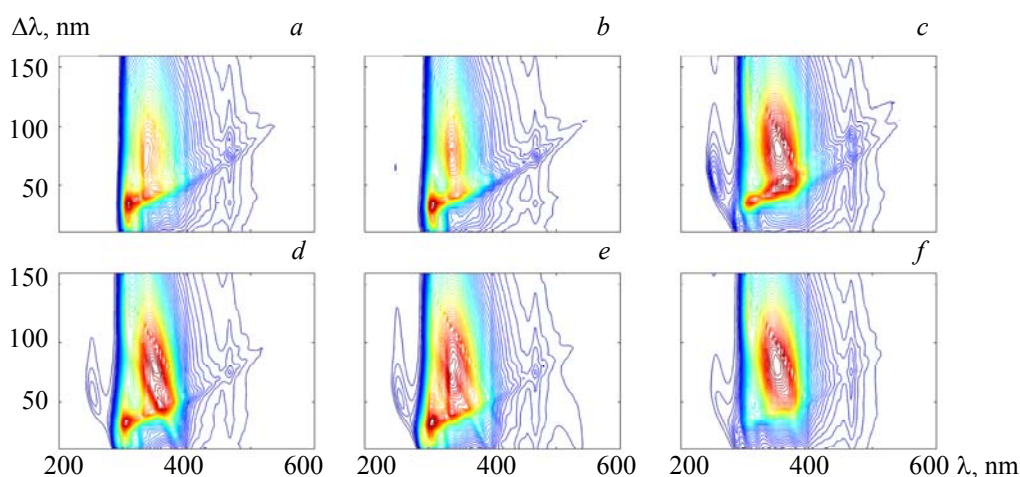
Fig. 1. Contour plots of total SF of liquor samples from six different brands. (a) Wuliangye; (b) Jiannanchun; (c) Luzhoulaojiao; (d) Tuopai; (e) Xufu; and (f) Quanxing.

To provide a visual representation of the dominant patterns in the SF spectra of the liquor samples, PCA was used to analyze the data matrix. The data were all auto-scaled before PCA analysis to avoid variables with high intensities being considered more important than those with low intensities [33]. Figure 2 shows the score plots (PC1 × PC2) resulting from applying PCA to the raw and derivative spectra. In Fig. 2a, the samples from different brands were separated but not very clearly. In particular, the brands, QU, XU, and LU, overlapped each other. In contrast, Fig. 2b, using the derivative spectra, shows that the samples from the six brands were clearly separated except for those of TU which clustered in the middle of those of the XU brand. This shows that data preprocessing can lead to a more appropriate classification of the brands of the liquor samples studied. Therefore, supervised classification techniques were applied to the preprocessed spectra.
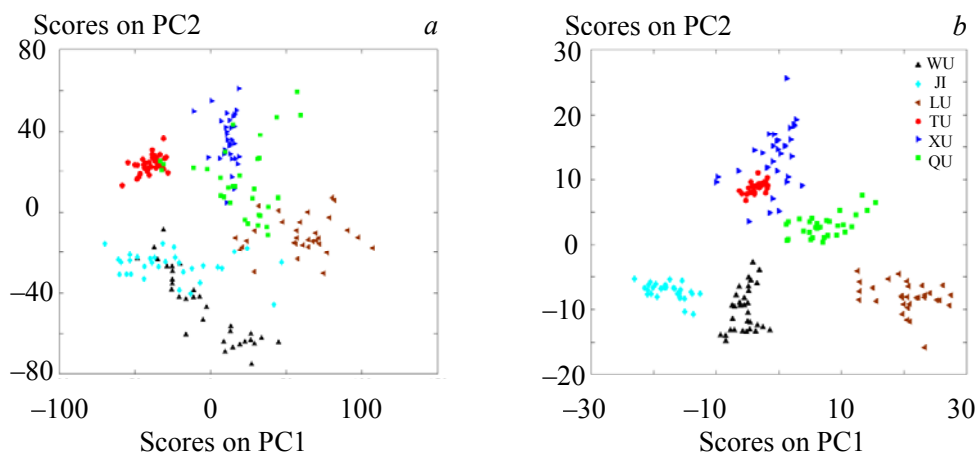

Fig. 2. Score cluster plot with two most important PCs of Chinese liquor samples from six different brands. (a): raw spectra; (b): derivative spectra. Brands: WU: Wuliangye; JI: Jiannanchun; LU: Luzhoulaojiao; TU: Tuopai; XU: Xufu; QU: Quanxing.

Spectra preprocessed by the first derivative method were used to build the PLS-DA model. It is important to decide on the number of latent variables (LVs) to be used to avoid overfitting. The optimal model was obtained using 15 LVs with full cross-validation, because there was only a small change in RMSE after 15 LVs. The prediction set was then used to test the PLS-DA model. Figure 3 shows the predictions for the different brands of liquor in the prediction set from the measured values using the PLS-DA model. From this figure, we can observe that most misjudgment samples occurred in adjacent groups. Overall, 55 samples

were correctly predicted out of 60 samples, a correct identification rate of 91.67%. The values of R^2 and RMSE were 0.984 and 0.176, respectively. In the calibration set, the correct identification rate was 93.94% with $R^2 = 0.986$ and RMSE = 0.148. A correct identification means that it is difficult to separate these brands using linear discriminant analysis; therefore nonlinear discriminant methods need to be developed.
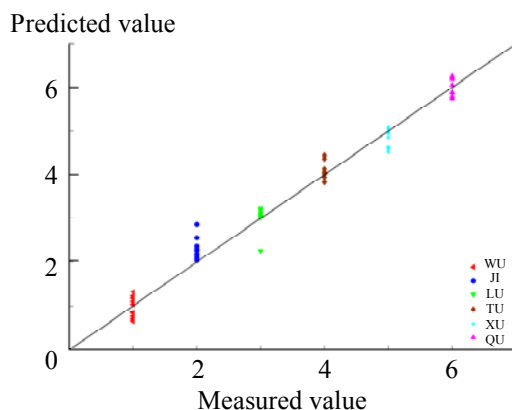

Fig. 3. A comparison of the measured values with predictions of the brands of Chinese liquor in the prediction set using the PLS-DA model. Brands: WU: Wuliangye; JI: Jiannanchun; LU: Luzhoulaojiao; TU: Tuopai; XU: Xufu; QU: Quanxing.

The optimal number of PCs was determined according to the highest discrimination rate observed in both the calibration and predication sets. The optimum SVM model used 4PCs with identification rates in the calibration and prediction sets of 100% and 96.67%, respectively. Figure 4a shows the contour plot of cross-validation accuracy using different colors, for different values of parameters C and γ, with a model input of 4 PCs. This showed that the highest cross-validation accuracy was obtained with $C = 4$ and $\gamma = 0.5$ when a minimal value of RMSECV (RMSE obtained with cross-validation) was achieved, indicating that the optimal SVM model had been obtained.
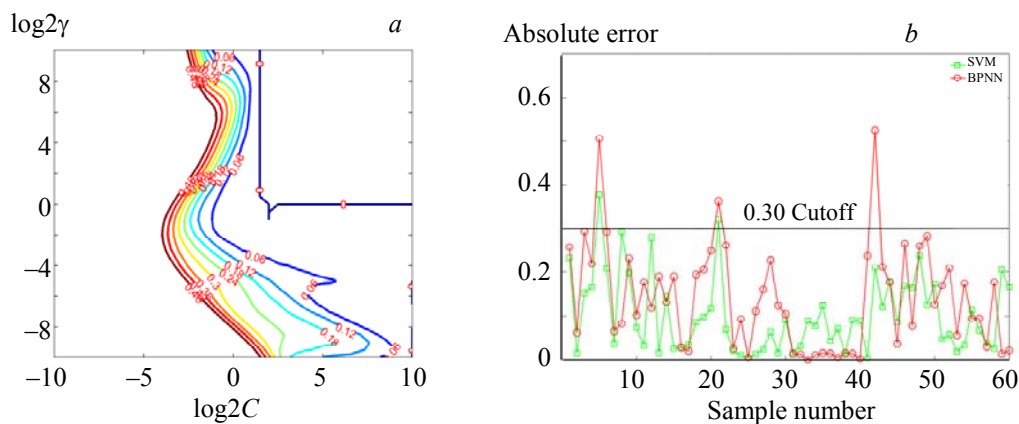

Fig. 4. (a) Results of grid search for the optimal C and γ values of the SVM model with the RBF kernel. (b) The absolute error of prediction results by BP-NN and SVM.

Another nonlinear approach, BP-NN was also used in the present study. The optimal BP-NN model was obtained when five PCs were used where its identification rates were 100% and 95% in the calibration and prediction sets, respectively. The absolute error of the prediction results of each sample is shown in Fig. 4b. Clearly, the prediction error of the SVM model was stable and lower than that of the BP-NN model. The recognition ratio of 96.67% was acceptable for discrimination analysis with a threshold value of ±0.3. If the cutoff value was set as normal ±0.4, the recognition value was 100%.

To achieve a good performance for identifying different brands of Chinese liquors using total SF, several methods of multivariate analysis and parameter optimization have been used systematically in the present study. Table 2, listing the discrimination results from the PLS-DA, BP-NN, and SVM models, shows that the identification rates using the BP-NN and SVM models were higher than those using the PLS-DA model in the calibration and prediction sets.

TABLE 2. Comparison of the Identification Results from the PLS-DA, BP-NN, and SVM Models

Methods	Calibration set			Prediction set		
	R^2	RMSE	Correct, %	R^2	RMSE	Correct, %
PLS-DA	0.986	0.148	93.94	0.984	0.176	91.67
SVM	0.998	0.084	100	0.998	0.120	96.67
BP-NN	0.994	0.093	100	0.996	0.160	95

This indicated that the nonlinear models (BP-NN and SVM) achieved better results than the linear model (PLS-DA). The nonlinear models, with their stronger capability for self-learning and self-adjustment, were able to handle the complex relationship among the different brands of liquor, and achieved better results than the linear models.

Furthermore, the performance of the SVM model was equal to that of the BP-NN model in the calibration set, but slightly better in the prediction set. The identification rate from the SVM model was 96.67% in the calibration set using 4 PCs, and that of the BP-NN model was 95% in the prediction set using 5 PCs. Therefore, the SVM model was considered more feasible to use than the BP-NN method for discriminating brands of Sichuan liquor.

Conclusion. This study has proposed a methodology for discriminating between six famous brands of Chinese liquor. SF spectroscopy has been used with three different supervised pattern recognition techniques (PLS-DA, SVM, and BP-NN). After comparing these techniques, chemometric analysis using SVM provided the best results with a mean correct prediction ability of 100 % in the calibration set and 96.67 % in the prediction set.

Overall the results demonstrate that SF spectroscopy can be enhanced by chemometric techniques to successfully classify brands of Chinese liquor samples. This methodology is a promising alternative to the time-consuming, expensive and laborious chemical analysis methods for assessing the authenticity of some types of beverage. Further studies are still needed to cover more brands of Chinese liquor and to provide a more robust discrimination model.

REFERENCES

1. W. Fan, M. C. Qian, *J. Agric. Food Chem.*, **54**, 2695–2704 (2006).
2. Y. Xu, D. Wang, W. Fan, X. Mu, J. Chen, *Adv. Biochem. Eng. Biotechnol.*, **122**, 189–233 (2010).
3. Y. Ma, D. Huo, H. Qin, C. Shen, P. Yang, C. Hou, *J. Appl. Spectrosc.*, **84**, 361–368 (2017).
4. P. Cheng, W. Fan, Y. Xu, *Food Res. Int.*, **54**, 1753–1760 (2013).
5. P. Cheng, W. Fan, Y. Xu, *Food Control*, **35**, 153–158 (2014).
6. W. Fan, M.C. Qian, *J. Agric. Food Chem.*, **53**, 7931–7938 (2005).
7. P. Wang, Z. Li, T. Qi, X. Li, S. Pan, *Food Chem.*, **169**, 230–240 (2015).
8. H. Qin, D. Huo, L. Zhang, L. Yang, S. Zhang, M. Yang, C. Shen, C. Hou, *Food Res. Int.*, **45**, 45–51 (2012).
9. M. Liu, X. Han, K. Tu, L. Pan, J. Tu, L. Tang, P. Liu, G. Zhan, Q. Zhong, Z. Xiong, *Food Control*, **26**, 564–570 (2012).
10. Q. Zhang, C. Xie, S. Zhang, A. Wang, B. Zhu, L. Wang, Z. Yang, *Sens. Actuators B: Chem.*, **110**, 370–376 (2005).
11. Z. Li, P. Wang, C. Huang, H. Shang, S. Pan, X. Li, *Food Anal. Methods*, **7**, 1337–1344 (2014).
12. V. Uríčková, J. Sádecká, *Spectrochim. Acta, A*, **148**, 131–137 (2015).
13. D. Dong, W. Zheng, W. Wang, X. Zhao, L. Jiao, C. Zhao, *Food Chem.*, **155**, 45–49 (2014).
14. H. Qiao, W. Zhang, W. Wang, *J. Biosci. Bioeng.*, **115**, 405–411 (2013).

15. Y. Wan, F. Pan, M. Shen, *Spectrochim. Acta, A*, **96**, 605–610 (2012).

16. D. Patra, A. K. Mishra, *Trends Anal. Chem.*, **21**, 787–798(2002).

17. F. Cui, Q. Zhang, X. Yao, H. Luo, Y. Yang, L. Qin, G. Qu, Y. Lu, *Pestic. Biochem. Phys.*, **90**, 126–134 (2008).

18. J. Sádecká, J. Tóthová, P. Májek, *Food Chem.*, **117**, 491–498 (2009).

19. J. Sádecká, V. Uríčková, K .Hroboňová, P. Májek, *Food Anal. Methods*, **8**, 58–69 (2015).

20. M. Tomková, J. Sádecká, K. Hroboňová, *Food Anal. Methods*, **8**, 1258–1267 (2015).

21. J. Sádecká, M. Jakubíková, P. Májek, A. Kleinová, *Food Chem.*, **196**, 783–790 (2016).

22. J. Tan, R. Li, Z. Jiang, *Food Chem.*, **184**, 30–36 (2015).

23. H. Hotelling, *J. Educ. Psychol.*, **24**, 417–520 (1933).

24. W. Li, L. Bagnol, M. Berman, R. A. Chiarella, M. Gerber, *Int. J. Pharm.*, **380**, 49–54 (2009).

25. P. Geladi, B. Kowalski, *Anal. Chim. Acta*, **185**, 1–17 (1986).

26. M. Barker, W. Rayens, *J. Chemometr.*, **17**, 166–173 (2003).

27. C. Cortes, V. Vapnik, *Mach. Learn.*, **20**, 273–297 (1995).

28. U. Thissen, M. Pepers, B. Üstün, W.J. Melssen, L. M. C. Buydens, *Chemometr. Intell. Lab. Syst.*, **73**, 169–179 (2004).

29. S. S. Keerthi, C. Lin, *Neural Comput.*, **15**, 1667–1689 (2003).

30. R. H. Nielsen, *Proc. Int. Joint Conf. Neural Networks*, IEEE Press, 593–605 (1989).

31. J. J. More, *Lect. Notes Math.*, **630**, 105–116 (1978).

32. Y. Liu, X. Luo, Z. Shen, J. Lu, X. Ni, *Opt. Rev.*, **13**, 303–307 (2006).

33. S. Wagner, K. Scholz, M. Donegan, L. Burton, J. Wingate, W. Volkel, *Anal. Chem.*, **78**, 1296–1305 (2006).