

Analytical Methods

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

Fast and accurate detection of organophosphorus pesticides using surface enhanced Raman scatter and chemometrics methods

Shizhuang Weng^{a, b}, Miao Li^{* a}, Cheng Chen^{a, b}, Xiang Gao^{a, b}, Shouguo Zheng^a, Xinhua Zeng^{* a}

In this paper, we develop an approach for the fast and accurate detection of organophosphorus pesticides combining surface enhanced Raman scatter (SERS) technology with chemometrics methods. Firstly, we measure the SERS spectra of three pesticides: methyl parathion, edifenphos and ethyl paraxon. And all the spectra are processed by the different preprocessing algorithms, namely baseline subtraction, Savitzky-Golay first derivative (FD), standard normal variate variant transformation (SNV) and multiplicative scatter correction (MSC). Subsequently, the principal components analysis (PCA) and nonnegative matrix factorization (NMF) are adopted to obtain the main feature of spectra, respectively. Lastly, the various data are utilized to develop the classification models by support vector machines (SVM) and random forest (RF), and the discrimination performance is evaluated through 5-fold cross-validation method. The experiments show that the baseline subtraction method can eliminate the fluorescence background and baseline drift perfectly. And PCA greatly shorten the training time on the premise of keeping classification accuracy, but NMF make the results worse. SNV can improve the discriminant accuracy up 4% except the usage of NMF, but MSC and the first derivative have a negative effect. The classification model of the highest accuracy (99.79%) is built by SVM with the spectra processed by SNV and PCA, and the training process spends 1115 s. The training time (81 s) of the RF model developed with the spectra processed by SNV is much shorter than the former though their accuracy is very approaching. And the classification accuracy of the model built with RF and the different data always keeps on the high level, which suggests that RF has the excellent robustness. These results demonstrate it is more suitable for the detection of organophosphorus pesticides via combining the SERS spectroscopy, RF with SNV.

1 Introduction

Organophosphorus pesticides, which are widely used to prevent and control a large variety of pests, play an important role to increase the agricultural production. Through inhibiting acetylcholinesterase^{1, 2}, the pesticides kill pests by the impairing nervous system. But the similar impairment can also be imposed to the human beings, and meanwhile the pesticide residues have widely distributed across the fruits, vegetables and other food now. Therefore, the detection of the organophosphorus pesticides residues is of much significance. Additionally, the complicated objects and arduous detection tasks are unavoidable in actual detections³, thus exploring and developing fast and accurate methods for pesticide detection are a preferred approach to solve these problems. At present, the fast detection methods mainly include two categories: biological and spectral methods. The biological methods, such as immunoassay^{4, 5}, enzyme inhibition⁶ and living bacterias⁷, have the advantages of short detection time, high sensitivity and lower testing cost. Nevertheless, these methods are sensitive to the measurement environment and require the synthesis of some intricate reagents, so they are not unfit for a wide variety of applications in the pesticide detection. The spectral methods as fluorescence^{8, 9}, infrared¹⁰ and Raman¹¹ spectroscopy

are universal, fast, and of simple sample operations for the various pesticides detection. And it is especially noted that SERS technology^{12, 13, 14, 15, 16, 17, 18}, which can enhance the Raman signal by 10^6 to 10^{15} times via using noble metal nanoparticles (Ag, Au) also possesses the advantages of fingerprint and high precision and is very suitable for the fast analysis of pesticides. The relevant works have been carried out by many researchers. Liu reports a shell thickness-dependent Raman enhancement of silver-coated gold nanoparticles (Au@Ag NPs) for the identification and detection of pesticide residues at various fruit peels¹³. And Lee develops a fast and ultra-sensitive trace analysis of methyl parathion pesticides using confocal SERS spectroscopy¹⁴. After we measure the SERS spectra of samples, the pesticides information of the samples can be subsequently obtained through the experts combining with the Raman spectral knowledge. The manual intervention, however, limits the development of SERS on the fast detection of organophosphorus pesticides. Moreover, the differences between the SERS spectra of analogues are minimal, and even the experts need a long time to differentiate them. Therefore, we aim at developing an approach independent of human for the intelligent analysis of pesticides using SERS spectroscopy.

A large number of experiments^{19, 20, 21} have demonstrated the classification methods in chemometrics can construct their own understanding (classification model) via learning the training data to realize the accurate identification of unknown samples. The classification methods, which are no longer dependent on the experts, are hence introduced into the analysis of SERS spectra in the paper.

^a Hefei Institute of Intelligent Machine, Chinese Academy of Sciences, Hefei, 230031, China

^b University of Science and Technology of China, Hefei, 230026, China

1
2 Taking the excellent classification ability and robustness into account,
3 SVM^{22,23} and RF^{24,25} will be used to develop classification models.
4 Besides, the variations, noise, fluorescence background and spectral
5 annihilation always arise in the spectra because the measurement can
6 be influenced by the SERS-active substrate, measurement
7 environment, and experimental operation and so on. To eliminate the
8 negative effects, consequently some pretreatment methods are
9 adopted in the paper. And the spectral data is of the high dimension,
10 so the feature extraction methods are also used to extract the main
11 information of SERS spectra.

12
13 In the paper, we measure the SERS spectra of methyl parathion,
14 edifenphos and ethyl paraxon solutions of different concentration,
15 and it should be noted that ethyl paraxon and methyl parathion are
16 the analogues. Firstly, the spectral fluorescence background and
17 baseline shift are eliminated by the baseline subtraction algorithm²⁶.
18 Subsequently, FD²⁷, MSC²⁸ and SNV²⁹ are adopted to diminish
19 some considerable noises, respectively. And PCA^{30,31} and NMF³²
20 are employed to extract the main information and reduce the
21 dimension of spectral data. Finally, various data are utilized to
22 develop the classification models by SVM and RF. In order to
23 determine the optimal chemometrics methods for the fast detection of
24 organophosphorus pesticides, the discrimination performance of all
25 models is evaluated via 5-fold cross validation method.

33 2 Experiments and methods

34 2.1 SERS spectroscopy measurement

35 Firstly, methyl parathion, edifenphos and ethyl paraxon (AR, 99%)
36 are dissolved in ethanol to get different concentration solutions,
37 which are 100, 50, 10, 5, 1, 0.5 and 0.1 mg/L. Up to date, the sol
38 solution of silver nanoparticle is reported as the most widely active
39 substrate in view of its synthetic simplicity and excellent enhanced
40 effect³³. The testing solutions are mixed with the Ag sol, which are
41 mixed by using ultrasonic dispersing method for 10 minutes³⁴.
42 Before measurement, the mixed solution is dropped on a silicon
43 wafer. After the evaporation of solvent sol at room temperature,
44 spectra are acquired from LabRAM HR800 Raman spectrometer
45 equipped with a 532 nm diode laser source. The excitation laser
46 power keeps constant at 20mW, and 50× microscope objective is
47 used for all measurements. For each original spectrum, we calculate
48 the averaged 10 scans with each integrating time of 1s. Every scan is
49 sampled with the 710 points over the range of 600-1800 cm⁻¹. For
50 every sample 50 spectra are measured, and a total of 1050 spectra are
51 obtained.

52 2.2 Chemometrics methods

53 2.2.1 Spectral pretreatment methods

54 The considerable noises caused by sample preparation,
55 inhomogeneities and measurement environments have inevitably
56 emerged in the SERS spectra of pesticides. Thus, all the spectra from
57 600 to 1800 cm⁻¹ are considered for preprocessing. Firstly, the
58 baseline subtraction method removes spectral fluorescence
59 background and baseline shift by the polynomial fit³⁵. And the
60 Savitzky-Golay FD reduces the background and baseline drift with a
window size of 7 points gradually²⁷. Then, SNV²⁸ centers each
spectrum separately by subtracting its mean and then scales it by its
own standard deviation. Finally, MSC²⁹ eliminates light scattering or
changes in path length effects for each sample relative to the average
of the calibration set by shifting and rotating each spectrum so that it
fits closely to the average spectrum of the dataset. Their effect on the
performance of the classification models is evaluated and discussed,
respectively.

61 2.2.2 Feature extraction methods

62 Allowing for the high-dimension of spectra, the analysis process is
63 time consuming. Besides, some information which is included in
64 spectra is not always additive effect on the analysis results. Therefore,
65 PCA and NMF are used to obtain the dominant features of SERS
66 spectroscopy, respectively.

67 Firstly, PCA obtains the orthogonal basis through the eigenvalue
68 decomposition of data covariance matrix and converts the data into
69 the scope matrix based on the basis. Then, the scopes which make
70 greater contribute to the variance of data are preserved to get the
71 principal components (PCs). Hence, PCA can reduce a
72 multidimensional data set to its most dominant features, and remove
73 random variation (noise)³⁶. The accumulating contribution rate is up
74 to 95% for PCA here. Then, NMF³² is a top-down generative
75 algorithm that optimizes its internal representation to minimize the
76 reconstruction error between input and reconstructed output.
77 Additionally, its weighted least-squared problem resolution can
78 prevent the occurrence of negative factors and avoid contradicting
79 physical reality, which is very significant since the negative factors
80 are unexplainable for the actual problems.

81 2.2.3 Classification methods

82 SVM, which is developed by Vapnik³⁷, is based on statistical
83 learning theory and aims at determining the location of decision
84 boundaries that produce the optimal separation of classes. In a
85 two-class pattern recognition problem where classes are linearly
86 separable, SVM can select the one linear decision boundary that
87 leaves the greatest margin between the two classes. The margin is
88 defined as the sum of distances to the hyperplane from the closest
89 points of the two classes here. If the two classes are not linearly
90 separable, SVM will try to find the hyperplane that maximizes the

margin and minimizes a quantity proportional to the number of misclassification errors. Besides, SVM can also be extended to handle the nonlinear problems with projecting the input data into a high-dimensional feature space through nonlinear mapping and formulating a linear classification problem in the feature space. Subsequently, the kernel functions are introduced to reduce the computational cost of mapping into high-dimensional feature space. In this paper, the radial basis function was adopted for its high effectiveness in training process. Through utilization of the techniques such as 'one against one' and the 'one against the rest', SVM can also be applied to the multi-class problems now. In case of SVM, the accuracy mainly depends on two parameters which are the regularization parameter C and the width of gaussian kernel g . The empirical value range of C runs from 0.001 to 1000, and the range of g is from 0.001 to 100³⁸. The value of C and g can be obtained by traversing the value range with the interval of 2^{0.5} times to guarantee the highest classification accuracy.

RF^{23,24} is developed based on the decision tree, which can be viewed as one ensemble method of trees to improve the algorithm performance further. In general, random forest classifier consists of a combination of tree classifiers where each classifier is generated by using a random vector sampled independently from the input vector, and each tree casts a unit vote for the most popular class to classify an input vector³⁹. Meanwhile, it also offers some unique features which are built-in estimation of prediction accuracy, measurement of descriptor importance, and measurement of similarity between molecules. During the training process, RF does not do any pruning at all. So it is noted that the classification model based on RF can be trained in less time even in cases where there are excessively a large number of descriptors (data dimension).

All computation and chemometric methods were implemented in MATLAB 2011b (The Mathworks Inc., Natick, MA, USA). The free SVM toolbox (Zhiren Lin, Taiwan) was applied in MATLAB to develop the classification models. The RF is carried out by using the RF toolbox.

3 Results and discussion

3.1 SRES spectra of methyl parathion, edifenphos and ethyl paraxon

The original SERS spectra obtained from ethyl paraxon, edifenphos and methyl parathion solutions in different concentrations (0.1, 0.5, 1, 5, 10, 50 and 100 mg/L) and the spectra processed by baseline are all shown in Fig. 1. From the figure (A, B, C), the strong fluorescence background and baseline drift presents in the original spectra of three pesticides. All the spectra are handled by the baseline subtraction method, and the results shown in fig. 1 (D, E, F) indicate the impact

is removed perfectly.

For the fingerprint characteristics SERS spectral peaks appear on some fixed positions and represent the vibration information of molecules, which is the basis for the analysis and detection of substances using SERS. According to the previous studies, the spectral assignation is discussed. The main peaks of methyl Parathion¹³ at 844, 1143, 1377, and 1581 cm⁻¹ should be attributed to the vibrations of phenyl stretch, N–O stretch, C–O stretch, and P–O stretch, respectively. For edifenphos, the peaks at 1362, 1297, 1207, 1110, 1071 cm⁻¹ are associated with CH₃ symmetry deformation, CH₂ twisting mode (in plane), C–C stretch and CH deformation in plane, respectively. And P=O stretch and P–O band occur at 1207, 927 cm⁻¹. Besides, the peaks at 1179, 1019, 997, 688 cm⁻¹ are due to aromatic ring breathing, and the ring stretching is at 1570, 1506 cm⁻¹. The main Raman peaks of ethyl paraoxon⁴⁰ included the aromatic ring (C=C) stretching at 1572 cm⁻¹, the symmetry stretching NO₂ at 1322 cm⁻¹, the aromatic ring (C–O) stretching at 1265 cm⁻¹, the C–H band (in plane) /NO₂ asymmetric stretching at 1104 cm⁻¹, the aromatic–NO₂ scissor at 860 cm⁻¹, and the C–C bending at 639 cm⁻¹.

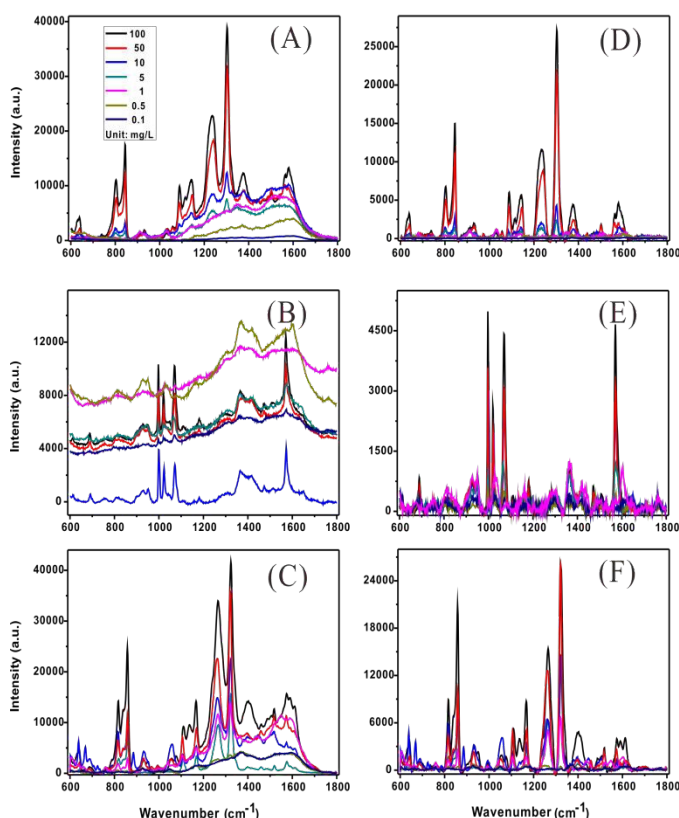


Fig. 1 The original SERS spectra of methyl parathion (A), edifenphos (B) and ethyl paraxon (C), and the spectra processed by baseline subtraction method of methyl parathion (D), edifenphos(E) and ethyl paraxon (F).

In summary, there is some different among the spectra of different

organophosphorus pesticides, which provides the feasibility for detection of pesticides using SERS. Nevertheless, we should also note that the spectroscopy of the structural analogs, such as ethyl paraxon and methyl parathion, is roughly similar. In this case, even some well-experienced experts also need to spend a lot of time on distinguishing with the two. Therefore, developing an intelligent discriminating approach of the organophosphorus pesticides is of great significant combining with SERS technology.

3.2 Classification results

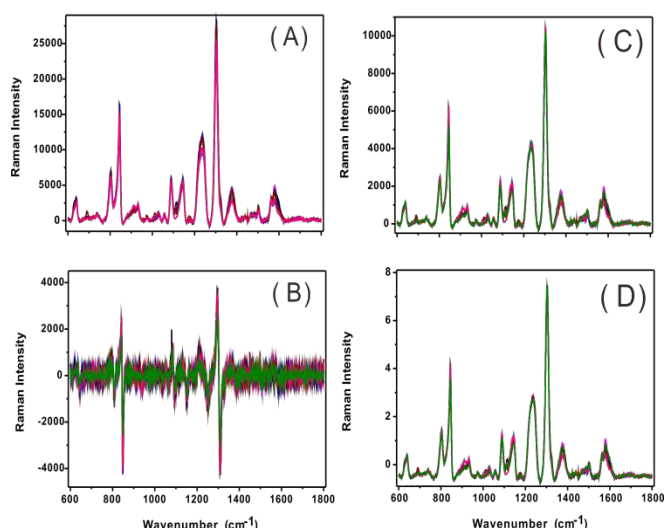


Fig. 2 The SERS spectra (A) of 100 mg/L methyl parathion and the spectra preprocessed by first derivative (B), SNV (C) and MSC (D).

Tab. 1 The classification results obtained with the different spectra and the model developed by SVM and RF. (Unit of Time: S)

Feature extraction	Pretreatment methods	SVM		RF	
		Accuracy	Time	Accuracy	Time
	*	95.17%	5215	95.09%	83
	FD	93.41%	5423	94.26%	99
	MSC	94.34%	3728	95.27%	82
	SNV	99.79%	4646	99.79%	81
	*	94.55%	172	94.09%	1.5
PCA	FD	91.50%	2538	94.61%	35
	MSC	94.03%	279	94.11%	3.5
	SNV	99.79%	1115	98.69%	17
	*	72.23%	598	91.09%	9.7
NMF	FD	70.23%	646	86.16%	11
	MSC	75.10%	605	88.85%	8.7
	SNV	77.03%	629	90.07%	8.3

* Without any methods

To suppress the noise and interference further, the spectra were preprocessed by FD, SNV and MSC before classification, respectively. And 10 original SERS spectra of 100 mg/L ethyl

paraxon and the spectra processed by different methods are shown in Fig. 2. Regarding the preprocessing effect on the spectra, from Fig. 3 it is seen that the first derivative preprocessing algorithm distorts the spectra (Fig. 2 B), while the MSC (Fig. 2 C) and SNV (Fig. 2 D) produce similar results, showing less pronounced differences to the original spectrum (Fig. 2 A), which are mostly related to the scale and position of the spectrum. The similarity between the has been pointed out by Helland⁴¹. Then the feature extraction methods as PCA and NMF are adopted to reduce the analysis time and extract the main information. Afterwards the original and processed spectra are utilized to build the classification model using RF and SVM. The discrimination performance of all the models is evaluated using 5-fold cross validation methods (Tab. 1).

Firstly, although NMF can ensure the transformed vectors from the original spectra nonnegative, the accuracy of the classification model developed with these vectors is obviously low than the other data. Therefore, it suggests that NMF is unfit for the qualitative analysis of SERS spectroscopy. But for PCA the classification accuracy of the corresponding models does not get promotion, but the training time is reduced dramatically which is mainly due to the reduction of data dimension. Then, for spectral pretreatment methods FD always worse the analysis results, MSC cannot promote the discrimination performance, and SNV can improve the classification accuracy up 4% except the usage of NMF. Finally, regarding the classification methods, the SVM model built with the spectra processed by SNV and PCA obtain the highest classification accuracy (99.79%) and spend the relative short training time (1115 s). However, the accuracy (99.75%) of the RF model developed with the spectra processed by SNV is a little lower than the former, and the training time (81 s) is much shorter. Additionally, the phenomenon that the classification accuracy of the model developed with RF and the different training data always keeps on the similar and high level should be noticed, which suggests that RF has the excellent robustness. Consequently, with SNV pretreatment, the RF is hereafter used to carry out the fast and accurate detection of organophosphorus pesticides using the SERS spectroscopy through taking the accuracy, robustness and training time into consideration simultaneously.

4 Conclusions

In the paper, a protocol for the fast and accurate detection of organophosphorus pesticides is developed on the basis of SERS technology and the chemometrics methods include the baseline subtraction, spectral pretreatment, feature extraction and classification methods. The spectra of pesticides handled by baseline subtraction methods can eliminate the fluorescence background and baseline drift perfectly. For the feature extraction methods PCA can

greatly shorten the training time on the premise of keeping classification accuracy, but NMF cause the sharp decrease of accuracy. Then, regarding the pretreatment methods, SNV can improve the discriminant accuracy up 4% except the using of NMF. However, MSC and the FD have a negative effect. Finally, the classification model of the highest accuracy (99.79%) is built by SVM with the spectra processed by SNV and PCA, and the training process spends 1115 s. Nevertheless, compared to SVM, RF performs better in the consideration of the accuracy, robustness and training time. Consequently, the results demonstrate that RF and SNV can be applied more potentially for the detection of organophosphorus pesticides using SERS spectroscopy.

Acknowledgement

This study is supported by National High Technology Research and Development Program of China (No. 2013AA102302).

References

1. F. Christa, K. J. Schmidt, *The chemistry of organophosphorus pesticides*, Springer-Verlag, 1982.
2. C. N. Pope, *J. Toxicol. Environ. Health. B: Crit. Rev.*, 1999, **2**, 161-181.
3. Y. Gilmo, I. M. White and X. Fan, *Sens. Actuator B: Chem.*, 2008, **133**, 105-112.
4. Y. J. Kim, Y. Cho, H. S. Lee, Y. T. Lee, S. J. Gee, and B. D. Hammock, *Anal. Chim. Acta*, 2003, **475**, 85-96.
5. Y. Liang, X. J. Liu, Y. Liu, X. Y. Yu, and M. T. Fan, *Anal. Chim. Acta*, 2008, **615**, 174-183.
6. A. Amine, H. Mohammadi, I. Bourais and G. Palleschi, *Biosens. Bioelectron.*, 2006, **21**, 1405-1423.
7. E. I. Rainina, E. N. Efremento, S. D. Varfolomeyev, A. L. Simonian, J. R. Wild, *Biosens. Bioelectron.*, 1996, **11**, 991-1000.
8. L. Viveros, S. Paliwal, D McCrae, J. Wild and A. Simonian, *Sens. Actuator B: Chem.*, 2006, **115**, 150-157.
9. A. L. Simonian, T. A. Good, S. S. Wang and J. R. Wild, *Anal. Chim. Acta*, 2005, **534**, 69-77.
10. M. Janotta, M. Karlowatz, F. Vogt and B. Mizaikoff, *Anal. Chim. Acta*, 2003, **496**, 339-348.
11. P. A. Tanner, K. H. Leung, *Appl. Spectrosc.*, 1996, **50**: 565-571.
12. Z. Liron, A. Zifman and V. Heleg-Shabtai, *Anal. Chim. Acta*, 2011, **703**, 234-238.
13. B. Liu, G Han, Z. Zhang, R. Liu, C. Jiang, S. Wang and M. Y. Han, *Anal. Chem.*, 2011, **84**: 255-261.
14. D. Lee, S. Lee, G. H. Seong, J. Choo, E. K. Lee, D. G. Gweon and S. Lee, *Appl. Spectrosc.*, 2006, **60**, 373-377.
15. Y. Wang, B. Yan, L. Chen, *Chem. Rev.*, 2012, **113**, 1391-1428.
16. J. Kneipp, H. Kneipp, K. Kneipp, *Chem. Soc. Rev.*, 2008, **37**, 1052-1060.
17. W. Zhang, B. Li, L. Chen, Y. Wang, D. Gao, X. Ma, A. Wu, *Anal. Methods*, 2014, **6**, 2066-2071.
18. B. Li, W. Zhang, L. Chen, B. Lin, *Electrophoresis*, 2013, **34**, 2162-2168.
19. L. Mariey, J. P. Signolle, C. Amiel and J. Travert, *Vib. Spectrosc.*, 2001, **26**, 151-159.
20. Eide, K. Zahlse, *Energ. Fuels*, 2007, **21**, 3702-3708.
21. D. Cozzolino, H. E. Smyth and M. Gishen, *J. Agric. Food Chem.*, 2003, **51**, 7703-7708.
22. Q. Chen, J. Zhao, C. H. Fang and D. Wang, *Spectrochimica Acta A*, 2007, **66**, 568-574.
23. A. Borin, M. F. Ferrão, C. Mello, D. A. Maretto and R. J. Poppi, *Anal. Chim. Acta*, 2006, **579**, 25-32.
24. Ham, Y. Chen, M. M. Crawford and J. Ghosh, *IEEE T. GEOSCI. REMOTE.*, 2005, **43**, 492-501.
25. V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan and B. P. Feuston, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 1947-1958.
26. M. Člupek, P. Matějka and K. Volka, *J. Raman Spectrosc.*, 2007, **38**, 1174-1179.
27. B. Jamshidi, S. Minaei, E. Mohajerani and H. Ghassemian, *Comput. Electron. Agr.*, 2012, **85**, 64-69.
28. Kachrimanis, D. E. Braun, U. J. Griesser, *J. Pharmaceut. Biomed.*, 2007, **43**, 407-412.
29. Y. Hu, A. Erxleben, A. G. Ryder and P. McArdle, *J. Pharmaceut. Biomed.*, 2010, **53**, 412-420.
30. C. Fan, Z. Hu, L. K. Riley, G. A. Purdy, A. Mustapha, and M. Lin, *J. Food Sci.*, 2010, **75**, M302-M307.
31. Kuroda, Y. Mori, M. Iizuka, and M. Sakakihara, *Comput. Stat. Data An.*, 2011, **55**, 143-153.
32. W. Li, D. Qu, M. Li, J. Liu, Y. Zhong, F. Zhang, B. Sun, D. Yu and J. Liu, *Anal. Methods*, 2012, **4**, 1704-1708.
33. G. Hu, L. Hu, J. Song, P. Li, X. Chen, H. Li, *Advances in Neural Networks-ISBN 2010. Springer Berlin Heidelberg*. 2010, 1-8.
34. A. Michota, J. Bukowska, *J. Raman Spectrosc.* 2003, **34**: 21-25.
35. H. Cen, Y. He, *Trends in Food Sci. Tech.*, 2007, **18**, 72-83.
36. H. M. Al-Qadiri, M. A. Al-Holy, M. Lin, N. I. Alami, A. G. Cavinato and B. A. Rasco, *J. Agr. Food Chem.*, 2006, **54**, 5749-5754.
37. J. A. Suykens and J. Vandewalle, *Neural Process. Lett.*, 1999, **9**, 293-300.
38. S. Liu, N. Jiang, *Mechatronics and Automation*, 2008. ICMA

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
-
2008. *IEEE International Conference on IEEE.*, 2008, 2014, **127**, 269-275.
509-513.
39. Breiman, *Mach. Learn.*, 2001, **45**, 5-32.
40. P. Li, R. Dong, Y. Wu, H. Liu, L. Kong and L. Yang, *Talanta.*,
41. T. Fearn, C. Riccioli, A. Garrido-Varo and J. E. Guerrero-Ginel, *Chemometr. Intell. Lab.*, 2009, **96**, 22-26.