Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 153 (2016) 586–590

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

Discrimination of genetically modified sugar beets based on terahertz spectroscopy

Tao Chen *, Zhi Li, Xianhua Yin, Fangrong Hu, Cong Hu

Guangxi Key Laboratory of Automatic Detecting Technology and Instruments, School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China

ARTICLE INFO

Article history:
Received 7 April 2015
Received in revised form 25 August 2015
Accepted 27 September 2015
Available online 30 September 2015

Keywords:
Terahertz spectroscopy
Genetically modified sugar beets
Discriminant partial least squares

ABSTRACT

The objective of this paper was to apply terahertz (THz) spectroscopy combined with chemometrics techniques for discrimination of genetically modified (GM) and non-GM sugar beets. In this paper, the THz spectra of 84 sugar beet samples (36 GM sugar beets and 48 non-GM ones) were obtained by using terahertz time-domain spectroscopy (THz–TDS) system in the frequency range from 0.2 to 1.2 THz. Three chemometrics methods, principal component analysis (PCA), discriminant analysis (DA) and discriminant partial least squares (DPLS), were employed to classify sugar beet samples into two groups: genetically modified organisms (GMOs) and non-GMOs. The DPLS method yielded the best classification result, and the percentages of successful classification for GM and non-GM sugar beets were both 100%. Results of the present study demonstrate the usefulness of THz spectroscopy together with chemometrics methods as a powerful tool to distinguish GM and non-GM sugar beets.

1. Introduction

Sugar beet (*Beta vulgaris*) is a biennial crop which grows a sugar-rich tap root in the first year (the vegetative stage) and a flowering seed stalk in the second year (the reproductive stage). Nowadays people meet the major needs of sugar by producing sugar from sugar beet and sugarcane. About 30% of the sugar supply comes from sugar beets in the world. The contribution of the sugar beet to the national agricultural economy is considerable, so how to produce the crop efficiently have been largely investigated in many countries [1]. The recent developments in biotechnology and genetic engineering technology has enabled the introduction of a gene sequence from a donor organism, which confer new characteristics, such as resistance to disease, herbicide tolerance and other new features [2,3]. Genetic modification has been realized in sugar beet by using several different target genes (e.g. rhizomania virus coat protein gene) to produce different GM sugar beet cultivars and lines. Nevertheless, some potential risks of GM organisms (GMOs) for ethics, environment and human health are unclear [4]. To detect the existence and the amount of GMOs, reliable and sensitive methods for qualitative and quantitative determination of GMOs are greatly needed [5,6].

A variety of methods which have been employed to detect GMOs have been developed, such as western blot [7], enzyme-linked immunosorbent analysis (ELISA) [8], lateral flow strip [9], Southern blot [10], polymerase chain reaction (PCR) [11] and so on. These protein- and DNA-based methods are precise, versatile and sensitive, but there are also some disadvantages, such as long duration, special need, high cost, etc. Terahertz (THz) refers to electromagnetic wave with the frequency range from 0.1 to 10 THz (wavelength 3 mm–30 μm) between microwave and infrared. Studies indicate that rotational and vibrational energy levels of many biomolecules (e.g. protein and DNA) fall in the THz frequency range. Therefore, it is possible to detect and identify GMOs using THz spectroscopy [12]. Moreover, THz spectroscopic analysis is simple, fast, non-destructive, and requires no sample pretreatment, which makes this technique a very good complement of the previous techniques, such as protein- and DNA-based testing methods, to detect the existence and amount of GMOs [13–15]. Because of the advantages over other analytical techniques, the aim of this paper was to apply THz spectroscopy combined with chemometrics methods for discrimination of GM and non-GM sugar beets.

2. Experimental section

2.1. Experimental system

The THz–TDS system used in the experiment has been depicted in [16,17]. The schematic of the experimental system is shown in Fig. 1. The apparatus consists of two parts: the Z-3 THz time-domain spectrometer (Zomega Terahertz Corp., USA), and the femtosecond laser, FemtoFiber pro NIR (TOPTICA Photonics Inc., Germany). The femtosecond laser is used as a radiation source. It has 780 nm central wavelength,

* Corresponding author.
E-mail address: ct63307@163.com (T. Chen).

http://dx.doi.org/10.1016/j.saa.2015.09.028

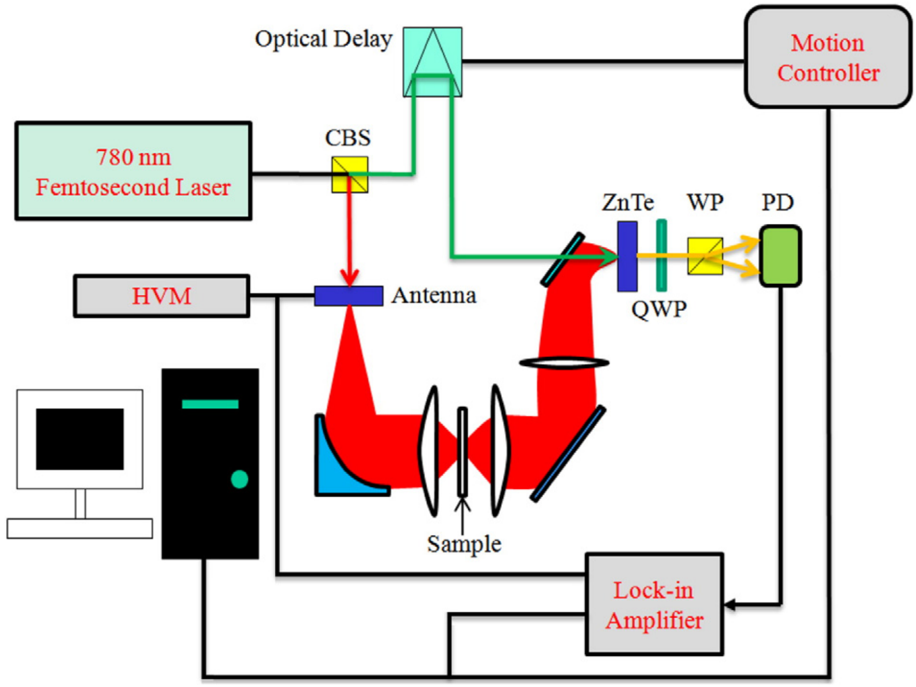

Fig. 1. Experimental setup of the THz–TDS system.

100 fs pulse width, 80 MHz repetition rate and nearly 140 mW average power. The laser beam is divided into the pump light (nearly 96 mW) and the probe light (about 16 mW) by a cubic beam splitter (CBS) for THz generation and detection, respectively. The whole experimental system has a spectral resolution of less than 5 GHz and a dynamic range of better than 70 dB. To avoid the interference of water vapor in the ambient air, the apparatus was placed in a closed box and dry air was injected until the indoor relative humidity (RH) is less than 2%.

2.2. Sample preparation

We bought the GM and non-GM sugar beet samples from the Sigma-Aldrich Shanghai Trading Co., Ltd., Shanghai City, China, with the purity of above 95%. To remove water, all the powder-form samples were dried at 50 °C for at least 1 h before sample preparation by using a vacuum drying oven. All the samples were carefully ground using a mortar and pestle, and were then pressed into about 1–2 mm circular slice with a hydraulic press under a pressure of about 3 tons. The diameter of a sample slice was approximately 13 mm and the nominal target weight was 200 mg. The slice surfaces were kept smooth and parallel to reduce the influences of multiple reflections. A total of 84 sugar beet samples (36 GM sugar beets and 48 non-GM ones) of similar size were prepared for THz transmission measurements. Of these, 60 sugar beet samples (24 GM and 36 non-GM ones) were used for the calibration set, while the remaining 24 samples (12 GM and 12 non-GM ones) were separated for the validation set. The samples of calibration set and validation set were selected randomly.

2.3. Chemometrics methods

2.3.1. Principal component analysis (PCA)

PCA is usually used to decrease the dimensionality of variables by projecting the *n*-dimensional data into a smaller number of linearly uncorrelated variables, namely, principal components (PCs). PCs are orthogonal and ordered such that the *k*th PC has the *k*th maximum variation between all PCs. That is to say, the first principal component (PC1) contains the maximum variations and is orthogonal the second principal component (PC2). By plotting the scores of PCA, we can view interrelationships among samples, and detect and explain sample patterns, classifications, differences or similarities [18,19].

2.3.2. Discriminant analysis (DA)

DA was used to discriminate the objects into categorical-dependent values, generally a dichotomy [20,21]. In DA, multiple quantitative attributes are applied to determine the variables, which distinguish between two or more naturally occurring clusters. DA provides statistical classification of samples and the samples can be grouped into several categories by computing the Mahalanobis distance between a sample and the mean value of a set of standards (THz spectra used here). In this study, the samples were divided into two classes (i.e. GMOs and non-GMOs), and the Mahalanobis distance is given as follows [22]:

$$MD_{ij} = \left(t_i - \overline{t_j}\right)^T S_{kj}^{-1} \left(t_i - \overline{t_j}\right) \tag{1}$$

where MD_{ij} is the Mahalanobis distance between the *i*th sample and the *j*th sample set, S_{kj} is the scores covariance matrix of the *j*th sample set, t_i and $\overline{t_j}$ are the score vector of *i*th sample and the average score vector of the *j*th sample set, respectively.

2.3.3. Discriminant partial least squares (DPLS)

DPLS is a classification method considered for a variety of applications such as medical diagnosis [23] and bioinformatics [24]. DPLS is based on a classic PLS-regression (PLSR) [25] in which the dependent variable is a categorical variable that designates the class of the sample. Each sample is assigned an integer as a reference value, which indicates whether the sample belongs to a particular class or not. For instance, a numeric value of 1 is used to indicate that the training sample belongs to class 1, and a numeric value of 2 indicates that the sample belongs to class 2. Classification of an unknown sample is derived from the value predicted by the PLS model. This predicted value is a real number, not an integer, which should be ideally close to the reference value (here either 1 or 2). A cutoff value between 1 and 2 (e.g. 1.5) is established so that a sample is assigned to class 1 if the predicted value is between 0.5 and 1.5, or assigned to class 2 if the value is between 1.5 and 2.5.

3. Results and discussion

3.1. THz spectroscopy of sugar beets

36 GM and 48 non-GM sugar beet samples were scanned by using THz–TDS system. Both the time-domain reference and sample THz spectra were obtained from free path and sample measurement. By performing fast Fourier transformation (FFT), the corresponding frequency-domain THz spectra of both the reference and the sample will be got, which were denoted as $E_r(\omega)$ and $E_s(\omega)$, respectively. Thus, we can calculate the THz absorbance spectra of the samples using the following formula [26]:

$$Absorbance(\omega) = -\log_{10}\left|\frac{E_s(\omega)}{E_r(\omega)}\right|^2. \quad (2)$$

Fig. 2 shows the THz absorbance spectra of 84 sugar beet samples (36 GM and 48 non-GM ones) in the frequency range of 0.2–1.2 THz obtained with THz–TDS system. As shown in Fig. 2, the THz absorbance spectra of GM and non-GM sugar-beet samples are very similar, and no obvious characteristic absorption peaks are observed in the THz band. Fig. 3 shows the average THz spectra for 36 GM and 48 non-GM sugar beet samples, and it indicates that the THz absorbance spectrum of GM sugar beets in the 0.2–1.2 THz region is less absorbent than that of non-GM sugar beets.

Because of the similarity of THz spectra of GM and non-GM sugar beet samples, it is difficult for us to discriminate the types of the samples by their spectral features. So, we can apply chemometrics methods to distinguish GM and non-GM sugar beet samples [27]. Here, three chemometrics methods, including PCA, DA and DPLS, were employed to classify sugar beet samples into transgenic and non-transgenic.

3.2. PCA

PCA was used to reduce the dimensionality of THz spectroscopic data and investigate qualitative differences between GM and non-GM sugar beet samples. A dataset of 84 GM and non-GM sugar beet samples was used to perform PCA. By performing PCA, the eigenvalues of the first two eigenvectors (PCs) extracted from the THz absorbance spectra of these 84 samples were 89.72% and 6.52%, respectively. The initial two PCs, which account for more than 95% variance in the data set, describe the most spectral variations related to origin and were used to make differentiation clearer.

Fig. 4 gives the two dimensional (2D) principal component scores scatter plot of GM and non-GM sugar beets using the first two PCs (PC1, PC2) and demonstrates how the two PCs are able to divide the samples. It can be seen from Fig. 4 that all 84 sugar beet samples can be separated clearly into two categories without overlapping each other by using the first two PCs as features. PCA offers significant information about fundamental data structure and probability of separation of objects, and it is very capable of clustering THz absorbance spectra with similar features. The results indicate that PCA can effectively extract THz spectral attributes of samples, and discrimination of GM and non-GM sugar beets is possible.

3.3. DA

The DA method was used to classify the GM and non-GM sugar beet samples. Fig. 5 shows a plot of the Mahalanobis distance of each sample from the two categories by using the DA method. As shown in Fig. 5, all samples are zonal distribution in the figure and have the apparent dividing line between GM and non-GM samples. It can be seen that all GM sugar beet samples were classified correctly and two non-GM ones were misclassified in the DA model. That is to say, using the DA method, the clusters of GM and non-GM sugar beets can be divided to a large extent and a good classification result was obtained.

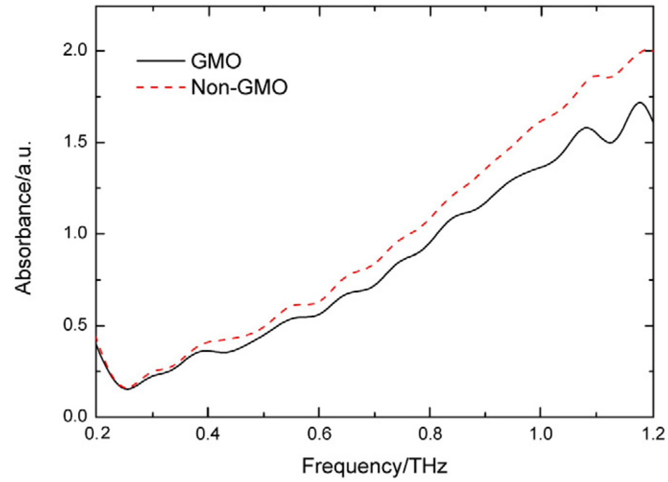

Fig. 3. Average THz spectra of GM (solid line) and non-GM (dashed line) sugar beets.

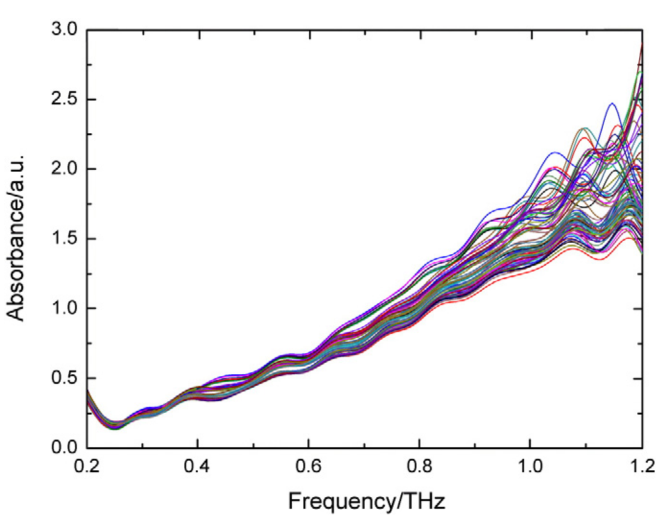

Fig. 2. THz absorbance spectra of all 84 GM and non-GM sugar beet samples.

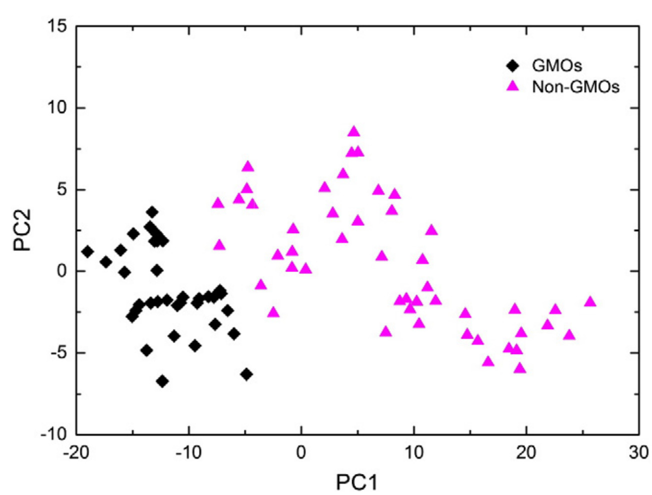

Fig. 4. Score plots of PC1 (89.72% variance) and PC2 (6.52% variance) for GM (◆) and non-GM (▲) sugar beets.

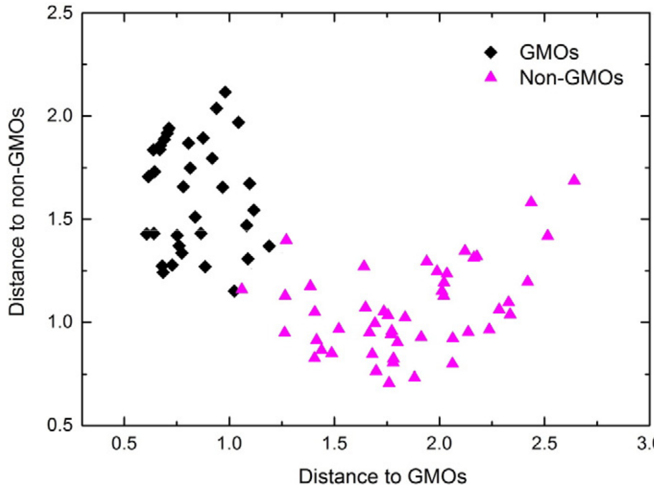

Fig. 5. Classification of GM (◆) and non-GM (▲) sugar beets by using the discriminant analysis method.

The Mahalanobis distance of the two misclassified non-GM samples to the non-GM group are 1.18 and 1.41, respectively, which is a little farther than to the GM group (1.05 and 1.27, respectively). Therefore, they are misclassified as the GM samples. Overall the percentage of correctly classified non-GM samples was 95.8%, and no GM samples were misclassified, as there was 100% correctness. The results of this study indicate that the DA method lead to an excellent classification for GM and non-GM sugar beet samples.

3.4. DPLS

The DPLS model was developed by regression of the THz spectrum data against the assigned reference value (set to 1 = GMOs and 2 = non-GMOs). Leave-one-out cross-validation (LOOCV) technique was used in development of the DPLS models. In order to estimate the accuracy of the models and the ability to classify samples, we calculated correlation coefficient (R) and root mean square error of cross-validation (RMSECV). Table 1 shows the statistics and classification results for GM and non-GM sugar beet samples using DPLS models. The results indicate that the DPLS models accounted for 97.8% and 96.1% of the variability in calibration and validation, respectively. The correct classifications for GM and non-GM samples were both 100% using DPLS. Fig. 6 shows the THz predictions of GM and non-GM sugar beet samples in the calibration set using the DPLS model. GM sugar beets with predicted values ranged from 0.61 to 1.38 and non-GM samples from 1.57 to 2.41 were defined as correctly classified by the DPLS model.

4. Conclusion

In this paper, an excellent classification was obtained by DPLS using THz spectroscopy, with classification accuracy up to 100%. The results of this study show that THz spectroscopy combined with chemometrics methods is a very powerful tool for discrimination of GMOs and non-GMOs, which avoids time-consuming, laborious and expensive sensory and chemical analyses. Although this work has achieved an excellent classification, only a limited number of samples and varieties involved in the experiment suggest caution in extending the applicability of such a technique to discriminate other GM products. In view of this, further studies should be focused on THz spectra collection and the development of robust and valuable models to discriminate other GMOs or blends.

Table 1
The statistics and classification results for GM and non-GM sugar beet samples using DPLS models.

	R	RMSECV
Calibration	0.978	0.103
Validation	0.961	0.146
	Correct classification, %	
	GMOs	non-GMOs
Classification	100	100

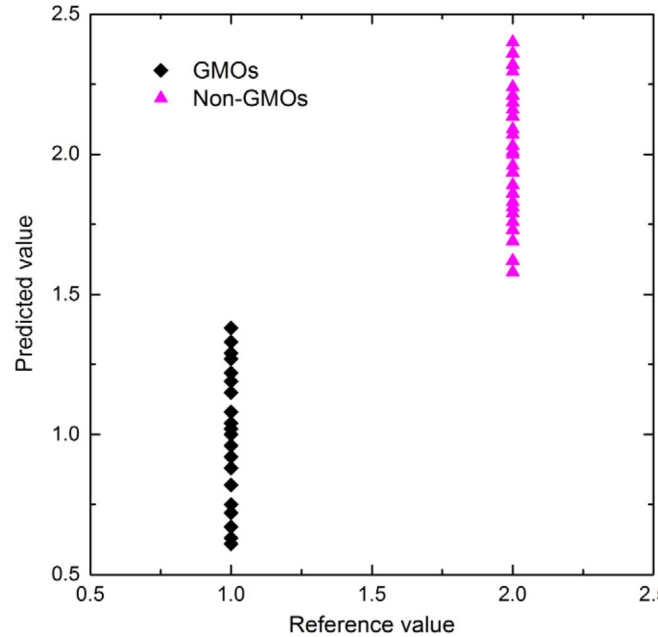

Fig. 6. Prediction for GM (◆) and non-GM (▲) sugar beet samples in the calibration set using DPLS regression.

Acknowledgments

This work was partly supported by the National Natural Science Foundation of China (No. 11574059); partly supported by the Guangxi Natural Science Foundation (No. 2015GXNSFBA139252); partly supported by the Foundation from Guilin University of Electronic Technology (No. UF14025Y); and partly supported by the program (No. YQ15104) from Guangxi Key Laboratory of Automatic Detecting Technology and Instruments.

References

[1] A. Philip Draycott, Sugar Beet, Blackwell, Oxford, 2006.
[2] N. Srivastava, V.K. Gupta, R. Pati, et al., Genetically modified crops: an overview, Biotechnology 10 (2011) 136–148.
[3] A. Karthikeyan, R. Valarmathi, S. Nandini, et al., Genetically modified crops: insect resistance, Biotechnology 11 (2012) 119–126.
[4] A.S. Bawa, K.R. Anilakumar, Genetically modified foods: safety, risks and public concerns—a review, J. Food Sci. Technol. 50 (2013) 1035–1046.
[5] D. James, A.M. Schmidt, E. Wall, et al., Reliable detection and identification of genetically modified maize, soybean, and canola by multiplex PCR analysis, J. Agric. Food Chem. 51 (2003) 5829–5834.
[6] A. Alishahi, H. Farahmand, N. Prietob, et al., Identification of transgenic foods using NIR spectroscopy: a review, Spectrochim. Acta A 75 (2010) 1–7.
[7] C.R. Lipton, J.X. Dautlick, C.D. Grothaus, et al., Guidelines for the validation and use of immunoassays for determination of introduced proteins in biotechnology enhanced crops and derived food ingredients, Food Agric. Immunol. 12 (2000) 153–164.
[8] G. Liu, W. Su, Q. Xu, et al., Liquid-phase hybridization based PCR-ELISA for detection of genetically modified organisms in food, Food Control 15 (2004) 303–306.
[9] J. Fagan, B. Schoel, A. Haegert, et al., Performance assessment under field conditions of a rapid immunological test for transgenic soybeans, Int. J. Food Sci. Technol. 36 (2001) 357–367.
[10] R. Ross, X.L. Ross, B. Rueger, et al., Nonradioactive detection of differentially expressed genes using complex RNA or DNA hybridization probes, BioTechniques 26 (1999) 150–155.
[11] K. Nakamura, H. Akiyama, N. Kawano, et al., Evaluation of real-time PCR detection methods for detecting rice products contaminated by rice genetically modified with a CpTI-KDEL-T-nos transgenic construct, Food Chem. 141 (2013) 2618–2624.

[12] B. Ferguson, X.C. Zhang, Materials for terahertz science and technology, Nat. Mater. 1 (2002) 26–33.
[13] J.V. Rudd, J. Johnson, D.M. Mittleman, Quadruple radiation from terahertz dipole antennas, Opt. Lett. 25 (2000) 1556–1558.
[14] E.P.J. Parrotta, Y. Sun, E. Pickwell-MacPherson, Terahertz spectroscopy: its future role in medical diagnoses, J. Mol. Struct. 1006 (2011) 66–76.
[15] T. Chen, Z. Li, W. Mo, Identification of biomolecules by terahertz spectroscopy and fuzzy pattern recognition, Spectrochim. Acta A 106 (2013) 48–53.
[16] T. Chen, Z. Li, W. Mo, et al., Simultaneous quantitative determination of multicomponents in tablets based on terahertz time-domain spectroscopy, Spectrosc. Spectr. Anal. 33 (2013) 1220–1225.
[17] T. Chen, Z. Li, W. Mo, et al., Application of interval selection methods in quantitative analysis of multicomponent mixtures by terahertz time-domain spectroscopy, Spectrosc. Spectr. Anal. 34 (2014) 3241–3245.
[18] Y.C. Shen, P.F. Taday, M.C. Kemp, et al., Terahertz spectroscopy of explosive materials, Proc. SPIE 5619 (2004) 82–89.
[19] A.D. Burnett, W.H. Fan, P.C. Upadhya, et al., Broadband terahertz time-domain spectroscopy of drugs-of-abuse and the use of principal component analysis, Analyst 134 (2009) 1658–1668.
[20] S. Shrestha, F. Kazama, T. Nakamura, Use of principal component analysis, factor analysis and discriminant analysis to evaluate spatial and temporal variations in water quality of the Mekong River, J. Hydroinf. 10 (2008) 43–56.
[21] C. Kandemir-Cavas, E. Nasibov, Classification of apoptosis proteins by discriminant analysis, Turk. J. Biol. 37 (2012) 54–61.
[22] H.Y. Yu, B. Ying, T. Sun, et al., Vintage year determination of bottled Chinese rice wine by VIS-NIR spectroscopy, J. Food Sci. 72 (2007) E125–E129.
[23] H.Y. Fu, S.Y. Huan, L. Xu, et al., Moving window partial least-squares discriminant analysis for identification of different kinds of bezoar samples by near infrared spectroscopy and comparison of different pattern recognition methods, J. Near Infrared Spectrosc. 15 (2007) 291–297.
[24] Z.Y. Mao, W.S. Cai, X.G. Shao, Selecting significant genes by randomization test for cancer classification using gene expression data, J. Biomed. Inform. 46 (2013) 594–601.
[25] S. Wold, M. Sjostrom, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2001) 109–130.
[26] Y.F. Hua, H.J. Zhang, H.L. Zhou, Quantitative determination of cyfluthrin in n-hexane by terahertz time-domain spectroscopy with chemometrics methods, IEEE Trans. Instrum. Meas. 59 (2010) 1414–1423.
[27] X.X. Yin, B.W.H. Ng, B.M. Fischer, et al., Support vector machine applications in terahertz pulsed signals feature sets, IEEE Sensors J. 7 (2007) 1597–1608.