Optik 156 (2018) 834–840

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Original research article

Terahertz spectroscopy applied to quantitative determination of harmful additives in medicinal herbs

Huo Zhang [a], Zhi Li [a,b,*]

[a] School of Mechano-Electronic Engineering, Xidian University, Xi'an 710071, China
[b] Guilin University of Aerospace Technology, Guilin 541004, China

ARTICLE INFO

Article history:
Received 27 March 2017
Accepted 14 December 2017

Keywords:
Terahertz spectroscopy
Rapid measurement
Harmful additives
Variable importance in the projection

ABSTRACT

In this study, terahertz (THz) spectroscopy was used in combination with an improved partial least squares (PLS) method to quantitatively determine the harmful additives Auramine O in medicinal herb Pollen Typhae. A terahertz time-domain spectroscopy (THz-TDS) transmission system was built to collect the absorbance spectra, and stacked partial least squares based on variable contribution sorting (VIP-SPLS) was used to establish the correlation between the absorbance and the content of Auramine O. Compared with original PLS and SPLS, VIP-SPLS obtained a better performance. The results in this study indicated that THz spectroscopy combined with VIP-PLSR has a potential for the rapid and non-destructive prediction of harmful additives residue.

1. Introduction

For both modern and traditional medicines, herb is an important source. The quality of the raw herbs is related to the quality and safety of medicines. However, industrial coloring dye is abused severely to improve the appearance of raw herbs and arouse purchasing intention of customers. These industrial dyes pose a potential threat to people's health and safety. Auramine O ($C_{17}H_{21}N_3 \cdot HCl$) is a kind of yellow industrial dye which had been proved to cause cancer [1]. This dye was repeatedly found to be added in various kinds of raw herbs on the medicinal herb market in recent years [2–4]. Previous reports about the detection of harmful additives are mainly focused on liquid chromatography tandem mass spectrometry, chromatography, high performance liquid chromatography and capillary electrophoresis [5–8]. Although these methods are objective and accurate, plenty works of sample pretreatment need to be prepared, a lot of time would be spent on complex operations, and many chemical wastes need to be treated. Therefore, a supplement method is required to improve the detection efficiency and reduce the consumption of chemicals. Spectroscopy is an ideal choice because of the advantages of high detection speed, easy operation and no extra consumption.

In addition to the conventional spectroscopic methods such as infrared and ultraviolet spectroscopy, terahertz spectroscopy shows a great potential as an important method for chemical detection. Terahertz (THz) refers to the frequency between infrared light and microwave radiation, in the range from 0.1 THz to 10 THz. Research shows fingerprint features of molecules (especially large molecules) are concentrate in this band [9,10]. This feature makes it possible to facilitate chemical detection by using THz spectroscopy. Compared to the neighboring bands, THz waves exhibit larger penetration depths for dielectric materials than the infrared light and have a higher spatial resolution than the microwave radiation

* Corresponding author at: School of Mechano-Electronic Engineering, Xidian University, Xi'an 710071, China.
E-mail address: xidianzl@outlook.com (Z. Li).

https://doi.org/10.1016/j.ijleo.2017.12.050
0030-4026/

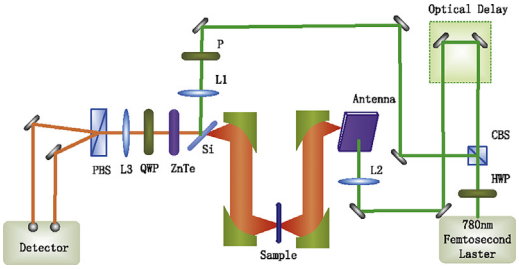

Fig. 1. Spectral measurements with THz-TDS.

[11]. Moreover, THz spectroscopy is considered as a completely safe detection technology because of its low-energy and nonionizing; it can also provide a high signal-to-noise ratio and excellent dynamic range [12]. Over the last 10 years, THz spectroscopy technique has been applied to numerous detection, such as the detection of genetically modified organisms [13,14], medical detection [15,16] and food safety detection [17,18]. In recent years, THz-TDS combined with chemometrics analysis is increasingly used for quantitative determination. In previous studies [19–21], PLS is a commonly used chemometric. However, most studies focused on PLS regression based on a full spectrum, where the existence of insignificant and irrelevant information may weaken the accuracy.

In this study, we aimed to (1) obtain the absorbance spectra of herbal samples which mixed with different concentrations of Auramine O by a THz-TDS transmission system; (2) quantitative analysis the Auramine O by an improved PLS [22]: VIP-SPLS; (3) compare the performances of VIP-SPLS and original PLS.

2. Material and methods

2.1. Sample preparation

The pure Auramine O was purchased from Regal Biology Technology (Shanghai, China). Herb Pollen Typhae was obtained from the herbal medicine market in Guilin, China. All of the materials were naturally dried and powdered. There would be a serious scattering effect if the size of powder was similar to the wavelength of the incident wave. So, the powder was filtrated through a 100-mesh sieve to ensure the size of powder is much smaller than the wavelength. In order to remove moisture absorbed from the air, all the powder were dried in a vacuum drying oven at 50 °C for 2 h. The mixtures of Auramine O and Pollen Typhae were obtained by mixing ratio of 1:99, 3:97, 5:95, 7:93, 11:89, 13:87, 15:85, 17:83, 21:79, 23:77, 25:75, 27:73 and 30:70 separately. Each of the mixed powder was pressed into circular tablets with 1 mm thickness, 12 mm diameter and 180 mg weight. These tablets were used as experimental samples for the concentration measurement of Auramine O. For each concentration, 27 samples were put in calibration set randomly, the remaining 9 were put in the prediction set.

2.2. Instrumentation

Our THz-TDS transmission system is shown in Fig. 1. A femtosecond laser (TOPTICA Photonics Inc., Germany) provided 100 fs pulse width with 780 nm central wavelength, 80 MHz repetition rate and 140 mW average power. The beam produced by the laser was split into a pump beam and a probe beam by a cubic beam splitter (CBS). THz waves were generated by the photoconductive dipole antenna at the irradiating of the pump beam and focused on the sample. The probe beam was focused onto a ZnTe crystal with the THz waves transmitted through the sample. Subsequently, the waves were guided to the detection antenna. The measurement was carried out at room temperature. The system was placed in an airtight box and filled with dry air to reduce the interference of moisture in the air. Dry air had been injecting into the box without interruption to ensure the relative humidity of the experimental system is less than 3%. The system has a high signal-to-noise ratio in the range between 0.2 THz and 1.6 THz, and our analysis would be carried out in this band.

2.3. Spectral processing

A terahertz wave transmitted through the drying air was used as the time-domain reference. After truncating the echo pules, the time-domain spectra were converted to frequency-domain spectra by fast Fourier transformation (FFT). THz absorbance spectra of the samples can be calculated from the frequency-domain THz spectra of the reference ($E_r(\omega)$) and the samples ($E_s(\omega)$):

$$Absorbance(\omega) = \log_{10}\left|\frac{E_r(\omega)}{E_s(\omega)}\right|^2 \quad (1)$$

In this formula, variable ω is the frequency.

2.4. Chemometrics

R (correlation coefficient) and *RMSE* (root-mean-square error) are evaluation parameters. R_c (correlation coefficient of calibration model) and *RMSEC* (root-mean-square error of calibration) were measure of how good a model fitted the calibration data set. R_p (correlation coefficient of the prediction set) and *RMSEP* (root mean square error of the prediction set) are used to evaluate prediction precision. *R* and *RMSE* can be calculated by the following formulas:

$$R = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})} \cdot \sqrt{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}} \tag{2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}} \tag{3}$$

where y_i and $\bar{y}$ are the reference concentrations of the *i*th sample and the average values of the reference concentrations respectively; $\hat{y}_i$ and $\bar{\hat{y}}$ are the predicted concentrations of the *i*th sample and the average values of the predicted concentrations respectively; *N* is the number of the samples.

Three multiple regression analysis were used and compared in our work. The absorbance of the samples was used as the input parameters of the quantitative analysis.

PLS is a commonly used quantitative analysis model, which have been proved to be effective in many applications. It uses the absorbance values within a given frequency range, extracts the spectrum features, and then establish the correlation between the instrumental measurements and the values of the interest property [23,24]. Word et al. [25] have gave a detailed description of PLS.

Stacked partial least squares (SPLS) divides the whole interval into several subintervals, and combines predictors from all subinterval models by application of conventional PLS regression. According to the description of Ni [26], regression analysis used in SPLS is similar to that in original PLS. On this basis, the operation of SPLS is expressed as the following steps:

(1) Evenly split the calibration spectra into *M* disjoint subintervals.

$$X = [X_1, X_2, \ldots, X_M] \tag{4}$$

(2) Calculate PLS regression vectors and *RMSECV* of the subintervals.
(3) Calculate weight of each PLS regression vector by dividing the corresponding reciprocal of the cross-validation error by the sum of all the reciprocals:

$$\Omega_k = \frac{1/RMSEC_k^2}{\sum_{i=1}^{M} 1/RMSEC_i^2} \tag{5}$$

(4) The stacked regression vector is obtained as:

$$\beta_{SPLS} = [\Omega_1\beta_1, \Omega_2\beta_2, \ldots, \Omega_M\beta_M] \tag{6}$$

(5) The values of the interest property could be obtained by multiplying absorbance matrix with stacked regression vector:

$$Y = X \cdot \beta_{SPLS}^T \tag{7}$$

(6) Search the best number of subintervals to minimize *RMSEC*.

VIP-SPLS, as the name, is built on the basis of SPLS. In this method, SPLS is combined with variable importance in the projection (*VIP*), which was proposed to assess the importance of input variables in PLS firstly by Word et al. [25,27,28]. The VIP score for the *j*th variable can be calculated in the following equation:

$$VIP_j = p\sum_{k=1}^{h} b_k^2 t_k^T t_k (\omega_{jk}/\left\|\omega_k\right\|)^2 / \sum_{k=1}^{h} b_k^2 t_k^T t_k \tag{8}$$

where *p* is the number of spectral variables, t_k, w_k, and b_k respectively stand for the *k*th column vector of *T* (score matrix of *X*), *W* (weight matrix of *X*) and *Q* (load matrix of *Y*) in original PLS. Feature points of the absorbance are re-sorted from largest to smallest according to their *VIP* scores. The next operation is consistent with step 2 to step 6 of SPLS.

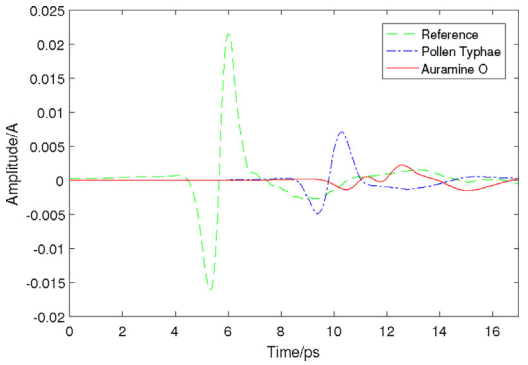

Fig. 2. The time domain spectra of Auramine O, Pollen Typhae and dry air.

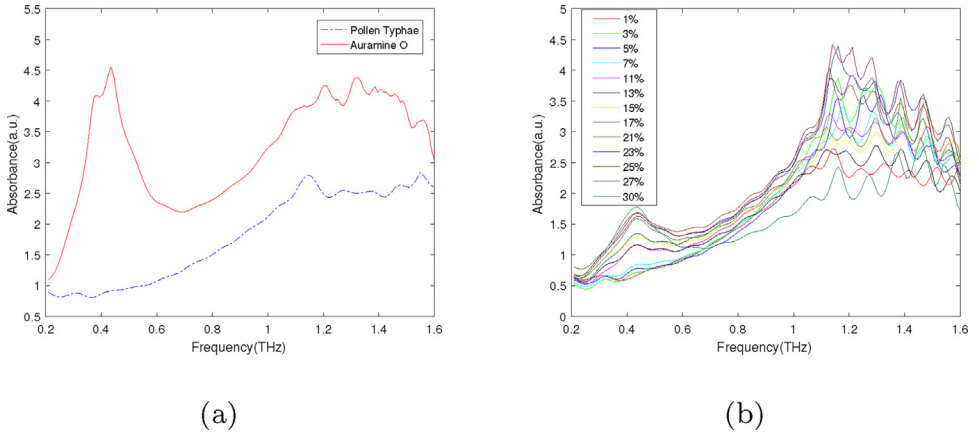

Fig. 3. Absorbance spectra: (a) the absorbance spectrum of Auramine O and Pollen Typhae in the band of 0.2–1.6 THz, (b) the THz absorbance spectra of Auramine O samples with different concentrations.

3. Results and discussion

3.1. THz spectra of samples

The time domain spectra of Auramine O, Pollen Typhae and dry air (as a reference signal) were obtained by the THz-TDS transmission system in the same experimental environment, and are shown in Fig. 2. The absorption spectra of Auramine O and Pollen Typhae were subsequently obtained by calculating the absorbance. According to the performance of the THz-TDS transmission system, we chose the frequency band between 0.2 THz and 1.6 THz (Fig. 3(a)). There are three chief absorption peaks of Auramine O located at 0.43 THz, 1.21 THz and 1.32 THz. Whereas Pollen Typhae has only one chief absorption peak at 1.15 THz. Fig. 3(b) displays the absorbance of Auramine O mixed in Pollen Typhae with different concentration. It could be seen that samples with different concentrations have different absorbance, and generally match Lambert-Beer law that absorbance is proportional to the concentrations of the attenuating species in the case of uniform thickness. On this basis, it is possible to establish the correlation between the absorbance and the concentrations of Auramine O by some commonly used multiple regression analysis such as PLS. However, because of the background noise, baseline drift or other interfering signals, some irregular curves and overlap changes appeared. For this reason, it is necessary to use some methods to select the useful interval to improve the quantitative analysis model.

3.2. Quantitative analysis by PLS and SPLS

PLS and SPLS were used to quantify the Auramine O mixed in herb Pollen Typhae based on the whole THz absorbance spectra. One-leave-out cross-validation was used in the optimization process of the quantitative model, and *RMSECV* was used as the basis for judgment. The number of PLS factor which leads the minimum value of *RMSEV* is selected as the optimum number. R_c, *RMSEC*, R_p and *RMSEP* are used to evaluate the performance of these models. As has been mentioned, irregular curves and overlap changes in the spectra interfere the quantitative analysis. As a result, a poor result was obtained by PLS (R_c = 0.9894, R_p = 0.9355, *RMSEC* = 0.1758, *RMSEP* = 0.3948). If only selected a sub interval to get rid of redundant information, useful information contained in the excluded interval would be lost in the mean while. In view of this, Ni [26] proposed SPLS

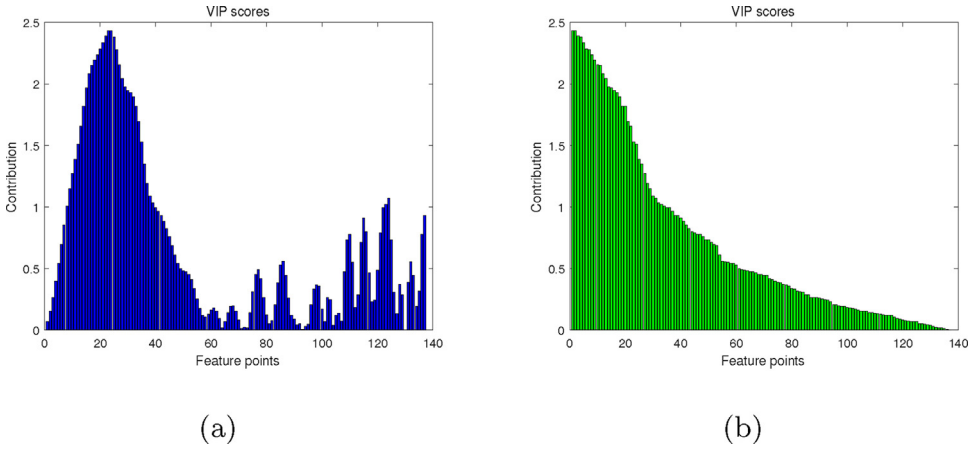

Fig. 4. VIP scores: (a) VIP scores of the features, (b) VIP scores after sorting.

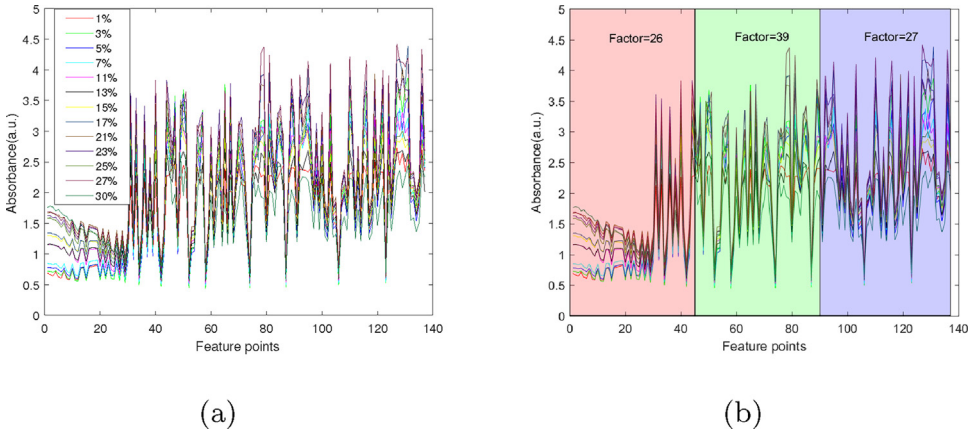

Fig. 5. Sorting and partitioning of VIP-SPLS: (a) absorbance spectra after sorting, (b) optimal partition.

to solve the problem of information leakage. Compared to PLS, SPLS has greater values of R_c and R_p in this work (0.9944 and 0.9413 respectively). Smaller values of *RMSEC* and *RMSEP* were obtained in the meanwhile (0.1068 and 0.3368 respectively). These evaluation parameters represent a certain degree of performance improvement by using SPLS.

3.3. *Quantitative analysis by VIP-SPLS*

Contribution was referred to the importance degree of useful information. For the quantitative analysis, contributions of each feature points in the spectra are different. These feature points are not distributed regularly according to their contributions. Accordingly, there may be large contribution differences between adjacent points. In the process of SPLS, adjacent points with large contribution differences may be assigned to a same subinterval and be involved in the operation with a same weight. This leads such a problem: useful information is weakened by weighting, while the noise is amplified. To avoid this problem, we re-sorted the feature points from largest to smallest according to the contribution value before the partition is performed. Then, the feature points in a subinterval would have similar contribution.

VIP scores provide a way to calculate the specific value of the contribution. The contributions of the feature points in the spectra are shown in Fig. 4(a). According to the sorting of VIP scores (Fig. 4(a)), the absorbance spectra were re-sorted as Fig. 5(a), and the intervals number ranging from 5 to 30 were tested. The weights were obtained by the sum contribution of each subinterval divided by the sum contribution of the whole interval. The regression vector of VIP-SPLS was obtained by weighted combination of the PLS regression vectors of the subintervals. The VIP-SPLS calibration can get the best *RMSEC* and R_c when intervals number set as 3 (Fig. 5(b)). Fig. 6 presents the correlation between the actual content and predicted content of Auramine O of PLS, SPLS and VIP-SPLS. Correspondingly, the performance comparison of these models is shown Table 1. Compared to SPLS, the R_c and R_p of VIP-SPLS risen to 0.9958 and 0.9464 respectively. This indicates a better correlation between the actual content and predicted content of Auramine O. In the meanwhile, deviation between the actual content and predicted content became smaller, and reflected in the decreased values of *RMSEC* (0.0921) and *RMSEP* (0.3237). Such indicates that the analytical results are improved by using VIP-SPLS.

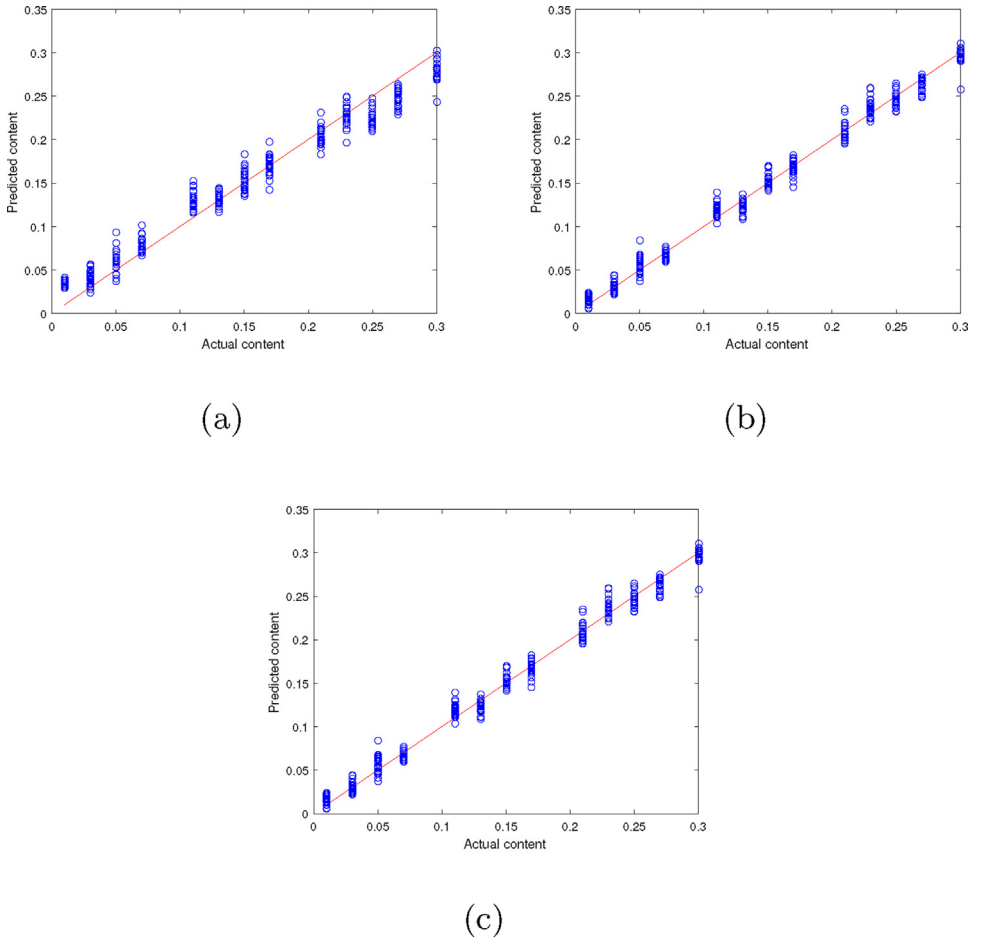

Fig. 6. Correlation statistics between the actual values and calculated values: (a) PLS regression, (b) SPLS regression, (c) VIP-SPLS regression.

Table 1
The performances comparison.

	Model	Parameters	R_c	*RMSEC*	R_p	*RMSEP*
1	PLS	Number of intervals = 1 Factors = 14	0.9894	0.1758	0.9355	0.3948
2	SPLS	Number of intervals = 3 Factor1 = 21 Factor2 = 20 Factor3 = 7	0.9944	0.1068	0.9413	0.3368
3	VIP-SPLS	Number of intervals = 3 Factor1 = 26 Factor2 = 39 Factor3 = 27	0.9958	0.0921	0.9464	0.3237

4. Conclusions

THz-TDS transmission system combined with chemometrics was used to implement a quantitative analysis of Auramine O in medicinal herb Pollen Typhae. For the powder samples, absorbance is generally proportional to the concentrations of Auramine O. The correlation between the absorbance and the concentration was established by the original PLS, SPLS and VIP-SPLS. VIP-SPLS is built on the basis of SPLS and VIP. It cannot only get rid of redundant information, but can also take effective use of concentration related information contained in each feature point. In quantitative analysis, VIP-SPLS obtained a better performance than PLS and SPLS. This study shows that THz-TDS can be used for the quantitative analysis

of harmful additives including Auramine O in medicinal herbs, and VIP-SPLS can be considered as an excellent multiple regression analysis in this process.

In addition, there were no complex pretreatment and no additional chemical consumption in the measurement. As a result, this technique can also be applied in other fields which require fast, convenient and nonpolluting quantitative analysis such as food safety, pharmaceutical industry, and environment monitoring.

Conflict of interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgements

This work is supported by Guangxi Colleges and Universities Scientific Research (Grant No. KY2015LX297).

References

[1] IARC, Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Humans, Suppl. 7, vols. 1-42, International Agency for Research on Cancer, Lyon, France, 1987.
[2] S. Gao, T. Zhang, S. Cui, Y. Guo, Quality investigation on commercial Pollen Typhae decoction pieces, Chin. Pharm. Aff. 28 (2014) 292–295, http://dx.doi.org/10.1016/j.phytol.2015.09.003.
[3] D. Li, Q. Zhu, D. Lv, B. Zheng, Y. Liu, Y. Chai, F. Lu, Silver-nanoparticle-based surface-enhanced Raman scattering wiper for the detection of dye adulteration of medicinal herbs, Anal. Bioanal. Chem. 407 (2015) 547–556, http://dx.doi.org/10.1007/s00216-015-8776-1.
[4] C. Tatebe, X. Zhong, T. Ohtsuki, H. Kubota, K. Sato, H. Akiyama, A simple and rapid chromatographic method to determine unauthorized basic colorants (rhodamine B, auramine O, and pararosaniline) in processed foods, Food Sci. Nutr. 2 (2014) 6031–6039, http://dx.doi.org/10.1002/fsn3.127.
[5] X. Lu, H. Shen, Determination of food color of basic orange and basic flavine o in bean products by LC–MS/MS, Food Ferment. Ind. 37 (2011) 166–169, http://dx.doi.org/10.13995/j.cnki.11-1802/ts.2011.08.008.
[6] M. Tripathi, S.K. Khanna, M. Das, Surveillance on use of synthetic colours in eatables vis a vis Prevention of Food Adulteration Act of India, Food Control 18 (2007) 211–219, http://dx.doi.org/10.1016/j.foodcont.2005.09.016.
[7] Q. Lin, X.-y. Zheng, S.-k. He, M. Dai, Y. Xie, Simultaneous high performance liquid chromatographic determination of chrysoidine, Auramine O and Safranine T in food, Food Sci. 30 (2009) 194–196, http://dx.doi.org/10.3321/j.issn:1002-6630.2009.14.038.
[8] H. Zhai, J. Li, Z. Chen, Q. Zhou, Y. Pan, Rapid determination of Auramine O in yellow croaker by microchip capillary electrophoresis, Chin. J. Appl. Chem. 30 (2013) 481–485, http://dx.doi.org/10.3724/SP.J.1095.2013.20273.
[9] B. Ferguson, X.C. Zhang, Materials for terahertz science and technology, Nat. Mater. 1 (2002) 26–33, http://dx.doi.org/10.1038/nmat708.
[10] A. Markelz, S. Whitmire, J. Hillebrecht, R. Birge, Thz time domain spectroscopy of biomolecular conformational modes, Phys. Med. Biol. 47 (2002) 3797–3805, http://dx.doi.org/10.1088/0031-9155/47/21/318.
[11] M. Pfleger, H. Roitner, H. Puehringer, K. Wiesauer, H. Gruen, S. Katletz, Advanced birefringence measurements in standard terahertz time-domain spectroscopy, Appl. Opt. 53 (2014) 3183–3190, http://dx.doi.org/10.1364/AO.53.003183.
[12] X.C. Zhang, J. Xu, Introduction to THz Wave Photonics, Springer US, 2010.
[13] W. Xu, L. Xie, Z. Ye, W. Gao, Y. Yao, M. Chen, J. Qin, Y. Ying, Discrimination of transgenic rice containing the Cry1Ab protein using terahertz spectroscopy and chemometrics, Sci. Rep. 5 (2015) 11115, http://dx.doi.org/10.1038/srep11115.
[14] T. Chen, Z. Li, X. Yin, F. Hu, C. Hu, Discrimination of genetically modified sugar beets based on terahertz spectroscopy, Spectrochim. Acta A – Mol. Biomol. Spectrosc. 153 (2016) 586–590.
[15] G. Takebe, Y. Kawada, K. Akiyama, H. Takahashi, H. Takamoto, M. Hiramatsu, Evaluation of drug crystallinity in aqueous suspension using terahertz time-domain attenuated total reflection spectroscopy, J. Pharm. Sci. 102 (11) (2013) 4065–4071.
[16] F. Hu, E. Guo, X. Xu, P. Li, X. Xu, S. Yin, Y. Wang, T. Chen, X. Yin, W. Zhang, Real-timely monitoring the interaction between bovine serum albumin and drugs in aqueous with terahertz metamaterial biosensor, Opt. Commun. 388 (2017) 62–67, http://dx.doi.org/10.1016/j.optcom.2016.11.006.
[17] G. Ok, H.J. Kim, H.S. Chun, S.-W. Choi, Foreign-body detection in dry food using continuous sub-terahertz wave imaging, Food Control 42 (2014) 284–289, http://dx.doi.org/10.1016/j.foodcont.2014.02.021.
[18] J. Qin, L. Xie, Y. Ying, Determination of tetracycline hydrochloride by terahertz spectroscopy with PLSR model, Food Chem. 170 (2015) 415–422.
[19] J.B. Sleiman, J.B. Perraud, B. Bousquet, N. Palka, J.P. Guillet, P. Mounaix, Chemical imaging and quantification of RDX/PETN mixtures by PLS applied on terahertz time-domain spectroscopy, in: 2015 40th International Conference on Infrared, Millimeter and Terahertz Waves (IRMMW-THz), August 23–28, Chinese Univ Hong Kong, Hong Kong, Peoples R China, 2015.
[20] J.B. Sleiman, B. Bousquet, N. Palka, P. Mounaix, Quantitative analysis of hexahydro-1,3,5-trinitro-1,3,5, Triazine/pentaerythritol tetranitrate (RDX-PETN) mixtures by terahertz time domain spectroscopy, Appl. Spectrosc. 69 (2015) 1464–1471, http://dx.doi.org/10.1366/15-07937.
[21] J. Qin, L. Xie, Y. Ying, Feasibility of terahertz time-domain spectroscopy to detect tetracyclines hydrochloride in infant milk powder, Anal. Chem. 86 (2014) 11750–11757, http://dx.doi.org/10.1021/ac503212q.
[22] H. Zhang, Fundamental research on novel modeling methods and applications off near infrared spectoscopy (Ph.D. thesis), Zhejiang University, 2015.
[23] Y. Hua, H. Zhang, H. Zhou, Quantitative determination of cyfluthrin in N-hexane by terahertz time-domain spectroscopy with chemometrics methods, IEEE Trans. Instrum. Measur. 59 (2010) 1414–1423, http://dx.doi.org/10.1109/TIM.2010.2041020.
[24] T.S. Groberio, J.J. Zacca, M. Talhavini, J.W.B. Braga, Quantification of cocaine hydrochloride in seized drug samples by infrared spectroscopy and PLSR, J. Braz. Chem. Soc. 25 (2014) 1696–1703, http://dx.doi.org/10.5935/0103-5053.20140164.
[25] S. Wold, M. Sjostrom, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2001) 109–130, http://dx.doi.org/10.1016/S0169-7439(01)00155-1.
[26] W. Ni, S.D. Brown, R. Man, Stacked partial least squares regression analysis for spectral calibration and prediction, J. Chemom. 23 (2009) 505–517, http://dx.doi.org/10.1002/cem.1246.
[27] S. Wold, E. Johansson, M. Cocchi, PLS: partial least squares projections to latent structures, in: 3D-QSAR in Drug Design, Theory, Methods, and Applications, ESCOM Science Publishers, Leiden, 1993, pp. 523–550, http://dx.doi.org/10.1002/0470845015.cpa012.
[28] I. Chong, C. Jun, Performance of some variable selection methods when multicollinearity is present, Chemom. Intell. Lab. Syst. 78 (2005) 103–112, http://dx.doi.org/10.1016/j.chemolab.2004.12.011.