

ACCURACY IMPROVEMENT OF GEOGRAPHICAL INDICATION OF RICE BY LASER-INDUCED BREAKDOWN SPECTROSCOPY USING SUPPORT VECTOR MACHINE WITH MULTI-SPECTRAL LINE

P. Yang,* H. T. Liu, Z. L. Nie,
and X. N. Qu

UDC 543.423:633.18

Mislabeled and adulteration are problems in the food industry. Considering the frequent occurrence of safety affairs in agricultural products, it is necessary to establish a traceability system for the quality and safety of agricultural products. The aim of the present study was to establish a rapid detection method for distinguishing rice samples from ten different products of geographical indication in China using laser-induced breakdown spectroscopy (LIBS). A support vector machine (SVM) is used to calculate the recognition rate of single spectral lines and multi-spectral lines of the geographic origins of rice. The adjusting spectral weighting of the multi-spectral line composition of mineral metal elements is higher, which can effectively improve the identification rate of the origin of the rice. The results show that the classification accuracies of single spectral line recognition and multi-spectral line recognition are 90.8 and 94.6%, respectively. It can be concluded that the LIBS technique combined with SVM should be a promising tool for rapidly distinguishing different geographic origins of rice.

Keywords: laser-induced breakdown spectroscopy, multi-spectral line, support vector machine, geographic origins.

Introduction. Rice is one of the main food products in the population's diet. Being rich in carbohydrates, vitamins, and mineral elements, rice supplies varieties of micronutrients and becomes the fundamental foodstuff for human beings. Mislabeled and adulteration are problems in the food industry. Wuchang rice is originally from China, and is mostly grown under rain-fed conditions, especially in the northeast region of Heilongjiang. It commands a considerably premium price in the rice market and the phenomena of mislabeling and adulteration are becoming more and more serious in the market. They threaten the credibility of the producers and traders and the rights of the consumers. Distinguishing rice quality and geographic origin has become significant, and classification of rice using effective detection methods is necessary.

The chemical composition of rice changes depending on the variety of rice, its geographic origin, and weather conditions. Certification of geographical indication is one of the most important parameters for protecting the production and origin of agricultural products [1]. The use of a geographical indication may act as a certification that the product possesses certain qualities. It is particularly important to rapidly control the unhealthy tendency by classifying the geographic origin of the rice.

Traditional methods of classifying rice geographic origin include sensory and chemical detection. Sensory detection is subjective and time-consuming, and chemical detection requires tedious chemical pretreatment, which is also time-consuming and cannot meet the requirements of speed, environmental compatibility, and batch processing. Wang et al. [2] identified volatile components from different areas and different storage times of rice samples, combining gas chromatography with a principal components analysis (PCA) and a partial least squares discriminant analysis. The recognition rate was 96 and 100%, respectively. Maione et al. [3] presented a data mining study of samples of rice obtained from producers in Goiás (midwestern region) and Rio Grande do Sul (southern region), and built classification models capable of predicting the geographical origin of rice based on its chemical components. The recognition rate of the support vector machine (SVM), random forest, and artificial neural network classification models was 93.66, 93.83, and 90%, respectively. The above-

*To whom correspondence should be addressed.

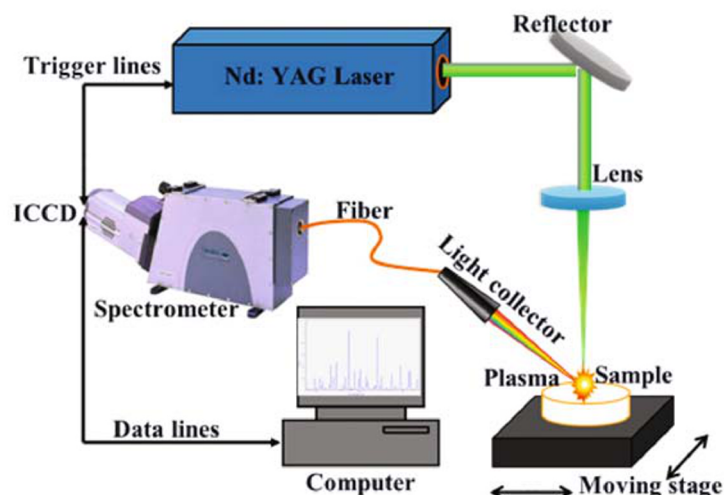


Fig. 1. Schematic diagram of the experimental setup.

mentioned methods have obtained a high degree of accuracy in rice classification, but they require solid-phase extraction and wet digestion. The detection process is complex, and rapid, real-time, and in situ detection are difficult. Therefore, studying rapid, environmentally friendly, and accurate detection methods of rice classification is significant.

A laser-induced breakdown spectroscopy (LIBS) is a spectroscopic technique for material composition analysis. With the characteristics of minimal sample preparation, rapid analysis, and multi-elemental analysis, it has been widely used in industry [4, 5], biomedical [6, 7], food [8–10], soil [11, 12], and other fields [13–15]. Combining LIBS with chemometric methods, classification or identification has some intrinsic advantages. Yu et al. [16] applied LIBS with SVM for 11 kinds of plastic classification, and the recognition accuracy was close to 100%. When Bilge et al. [17] applied the LIBS technique with PCA for beef, pork, and chicken classification, the recognition accuracy was 83.37%. Moncayo et al. [18] applied the LIBS technology and neural network for the placement of wine origins, with a recognition accuracy of 98.6%. Yan et al. [19] used LIBS coupled with image histogram of an oriented gradient features method for the analysis of 24 rice (*Oryza sativa* L.) samples, and the recognition accuracy was 81%. Therefore, the LIBS technology with chemometrics is feasible for classifying samples. However, few investigations on the products of geographical indication classification using LIBS combined with chemometric methods have been reported.

This work developed an SVM-based multi-spectral line approach to improve the accuracy of analysis for products of the classification of geographical indication. Single spectral lines and multi-spectral line recognition were compared to achieve better classification accuracy.

Experimental. The schematic diagram of the experimental setup used in this study is shown in Fig. 1. A Q-switched Nd:YAG pulsed laser (Quantel Brilliant B, wavelength: 532 nm, pulse duration: 8 ns, repetition rate: 10 Hz) was used to ablate samples. The laser beam was reflected by a dichroic mirror, and focused onto the sample surface by a UV-grade quartz lens with a focal length of 150 mm. The plasma spectrum was obtained using an echelle spectrometer (Andor Tech., Mechelle 5000, spectral range from 200 to 975 nm with a resolution of $\lambda/\Delta\lambda = 5000$) coupled with an intensified charge-coupled device (ICCD) camera (Andor Tech., iStar DH-334T). The acquisition and analysis of data were performed using a personal computer. In order to avoid the same point ablated at the sample surface, the samples were analyzed in the natural air and were automatically translated during the LIBS measurements using a double-axis motorized stage controlled by a programmable controller. It is necessary to do "bow" shape motion along X and Y directions.

Experimental Samples and Preparation Methods. Samples harvested in 2019 were from China's major rice regions (i.e., Guangdong, Hubei, Guangxi, Liaoning, Heilongjiang, Jilin, Anhui, Hunan, Jiangxi, and Jiangsu). Table 1 shows sample information including sample ID, commercial brand, and places of origin.

Considering the physical properties of the samples, such as irregularity, which affect the spectrum signal, in order to reduce the influence of moisture on the tested samples, rice samples underwent desiccation treatment in a 100°C drying cabinet, and then they were crushed. Different origins of rice samples were prepared into powder using the laboratory mill. Following this procedure, the powdered samples were sieved with 200 mesh screens. Then, the samples were formed as

TABLE 1. Rice Samples from Different Varieties and Geographical Origins

Sample	Commercial brand	Origin	Origin abbreviation
1	Golden Dolphin Thai fragrant rice	Guang Dong	GD
2	Organic rice	Hu Bei	HB
3	Guangxi Bama glutinous rice	Guang Xi	GX
4	Panfu Fengjin rice	Liao Ning	LN
5	Wuchang rice	Hei LongJiang	HLJ
6	Northeast brown rice	Ji Lin	JL
7	Anhui indica rice	An Hu	AH
8	Xiang Chi rice	Hu Nan	HN
9	Wannian kongmy	Jiang Xi	JX
10	Chongming Island rice	Jiang Su	JS

pellets for LIBS analysis with a pellet press machine by means of 15 g of the rice to 25-MPa pressure. For each rice sample, four pellets were prepared.

Results and Discussion. The LIBS spectra were produced under the following conditions: the laser pulse energy was 40 mJ, the laser frequency was 10 Hz, the gate delay and gate width were set to 1.5 and 3 μ s, respectively, and the ICCD gain was 2500. To reduce the intensity deviation, each spectral intensity was accumulated for 50 shots. One hundred spectra were recorded for each rice origin under an air background. At the same time, in order to prevent air breakdown, the focal point was 1.27 mm below the sample surface.

Rice has rich nutrients, especially carbohydrates, vitamins, mineral elements, and micronutrients. A carbohydrate is a biological molecule consisting of carbon, hydrogen, and oxygen atoms. According to the database of the National Institute of Standards and Technology (NIST), atomic spectrum database [20], and the Identification of Molecular Spectra [21], relative intensity, wavelength λ , transition probability A_{ki} , energy E , and statistical weight g of the upper and lower excitation levels for the spectral emission lines were referred to. The spectral emission lines of C, N, O, H, Mg, Mn, Ca, Si, Al, Na, and K can be referred to in the NIST atomic spectrum database. The molecular emission bands of C–N and C–C were referred to in the Identification of Molecular Spectra [21].

Figure 2 shows the LIBS spectra of 10 different rice samples. The selection of characteristic spectral lines was important to rice classification. The valuable and high-intensity spectral lines were located near the regions of ~245–290, 305–370, 370–425, 425–600, 600–800, 810–880, 516, 588, and 656 nm. Figure 2 shows 11 kinds of main elements and two kinds of molecular bands, namely, C, N, O, H, Mg, Mn, Ca, Si, Al, Na, K, C–N, and C–C, respectively, which were detected by the spectrometer with a broad spectral range.

The recognition accuracy of the single spectral line algorithm model is not high. In order to address these problems, adjusting spectral weighting (ASW) was developed for classification in [16]. The differences between rice from different regions in the case, which used a single characteristic spectral line as an algorithm model input variable, were too small to identify. In this work, combining several characteristic spectral lines, which belong to the same element and are closely adjacent, improves the classification of rice from different regions. The specific combinations are shown in Table 2.

The SVM algorithm was used in MATLAB R2010b (Math Works Corporation, USA) for data processing and obtaining classification accuracy. The SVM toolbox developed [21] (A Library for Support Vector Machines, LIBSVM) was used for data classification. The details are referred to in [22]. The radial basis function (RBF) was selected as a kernel function, and the LIBSVM program package was used. The so-called Radial Basis Function (RBF), is some kind of scalar function along the radial symmetry. It is usually defined as a monotonic function of the Euclidean distance between any point x in space and a center x_c , which can be written as $k(\|x - x_c\|)$, and is often used to calculate the similarity:

$$K(\bar{X}_i, \bar{X}_j) = \varphi(\bar{X}_i)^T \varphi(\bar{X}_j) = \exp\left(-g \|\bar{X}_i - \bar{X}_j\|^2\right), \quad g > 0. \quad (1)$$

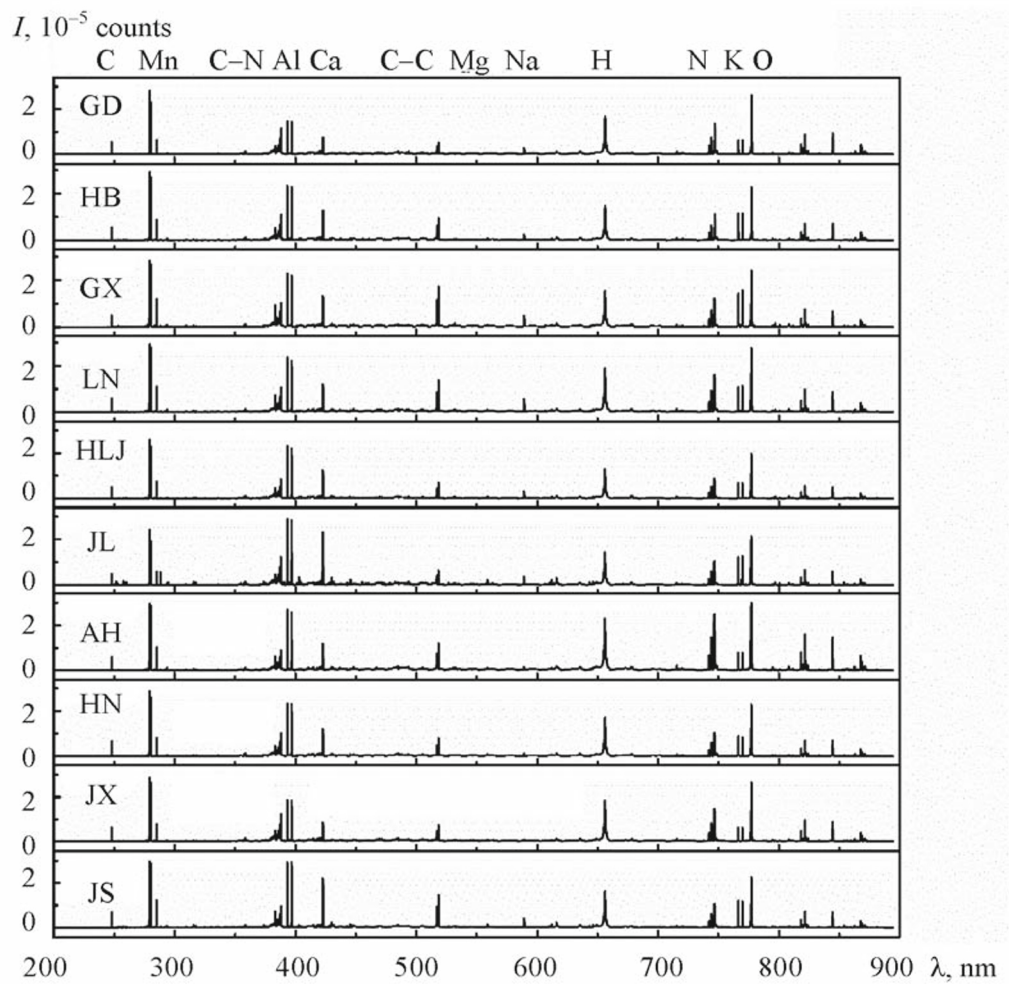


Fig. 2. Laser-induced breakdown spectroscopy spectra of 10 different rice samples.

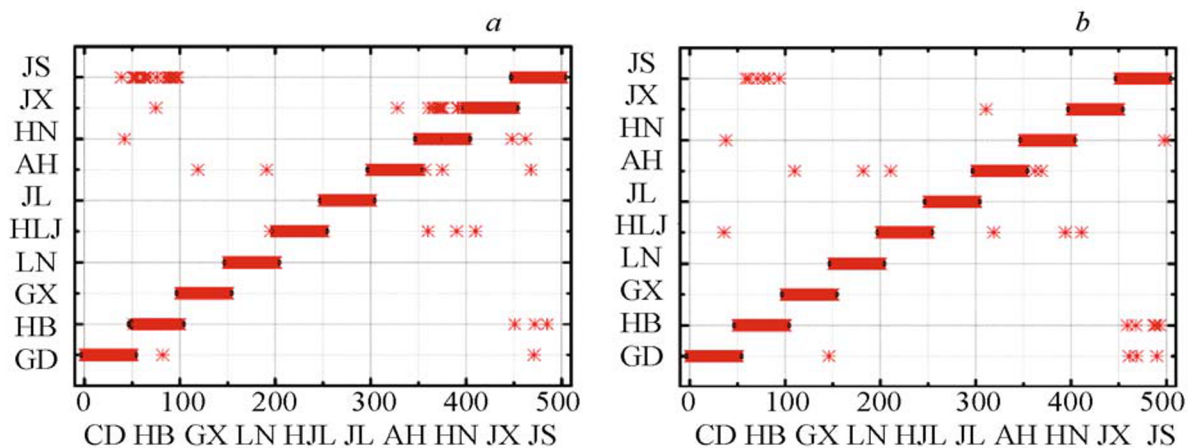


Fig. 3. Classification result of 10 kinds of rice using the support vector machine model by (a) a single spectral line and (b) a multi-spectral line combination. Actual label (●), label predicted by the SVM model (*). The successful classification is achieved when the "*" overlaps with the "●."

TABLE 2. Single Characteristic Spectral Line and Corresponding Multi-Spectral Line Combination

Single spectral line, nm	Multi-spectral line combination, nm
C–N (0,0) 388.34	C–N (0,0) 388.34, C–N (1,1) 387.14, C–N (2,2) 386.19, C–N (3,3) 385.47, C–N (4,4) 385.09
C I 247.86	C I 247.86
N I 746.83	N I 746.83, N I 744.23, N I 742.36
O I 777.19	O I 777.19, O I 777.42
H I 656.29	H I 656.29
C–C (0,0) 516.52	C–C (0,0) 516.52
Mg II 279.55	Mg II 279.55, Mg II 280.27, Mg I 285.21
Mn I 403.08	Mn I 403.08, Mn I 403.31, Mn I 403.45
Si I 288.16	Si I 288.16
Al I 394.40	Al I 394.40, Al I 396.15
Na I 588.95	Na I 588.95, Na I 589.59
K I 766.49	K I 766.49, K I 769.90

The nonlinear SVM classifier is

$$F(\bar{X}) = \text{sign} \left(\sum_{i=1}^l y_i \alpha_i \varphi(\bar{X}_i)^T \varphi(\bar{X}) + b \right). \quad (2)$$

The constraint condition is

$$\begin{cases} 0 \leq \alpha_i \leq c, i = 1, \dots, l, \\ \sum_{i=1}^l \alpha_i y_i = 0, \end{cases} \quad (3)$$

where X is the data of the spectra of the test set, $\varphi(X)$ is the test set of mapping functions, $\bar{X}_i \bar{X}_j$ is the 2-norm distance, l is the number of training sets, α_i is the Lagrange multiplier, b is the offset factor in the equation, c is the error term parameter, and g is the kernel parameter. The penalty parameter c of the error term and kernel parameter g of the RBF were optimized by a grid search combined with a cross validation method. Before the SVM model was trained, the spectral intensities for each spectrum were normalized by the calcium spectral line located at 422.67 nm.

The data set was split into training and validation sets. Fifty spectra from the original 100 spectra of each rice sample, which were selected randomly, were used to train the SVM model (training set). Other 50 spectra (validation set) were used to test the performance of the SVM model and to obtain the classification rate. Figure 3 shows the classification results of the test set spectra by the trained SVM model. Each spectrum has a label. The recognition accuracy of the multi-spectral line combination of the SVM algorithm was compared with single spectral lines.

As seen in Table 3, the average recognition rate of the test was 90.8%. The recognition rate of the multi-spectral line combination test was 94.6%. It is obvious that the accuracy rate of the multi-spectral line combination classification is higher than that of the single spectral lines. The penalty parameter c of the error term and kernel parameter g of the RBF were optimized by genetic algorithms combined with a cross validation method and fixed. The respective values of c and g were 222.8609 and 0.125 for the single spectral line, whereas for the multi-spectral line combination the values were 588.1336 and 0.047366, respectively.

Table 4 lists the classification weightings of normalized characteristic spectral lines of the single spectral line and multi-spectral line combination ASW, which were determined by a PCA algorithm. Obviously, classification weights of metal spectral lines, which play an important role in the classification trace ability, have been increased by the multi-spectral line combination. The rice absorbs many metallic mineral elements through the water, air, and fertilizer, as well

TABLE 3. Number of Spectra in Training and Validation Sets for Classifying Rice Origin Based on Single Spectral Line and Multi-Spectral Line Combination and the Classification Results

Sample	Single spectral line			Multi-spectral line combination		
	Training set	Validation set	Classification rate, %	Training set	Validation set	Classification rate, %
GD	50	50	96	50	50	96
HB	50	50	64	50	50	88
GX	50	50	98	50	50	96
LN	50	50	96	50	50	98
HLJ	50	50	100	50	50	98
JL	50	50	100	50	50	100
AH	50	50	98	50	50	96
HN	50	50	72	50	50	94
JX	50	50	96	50	50	98
JS	50	50	88	50	50	82
Average accuracy	90.8			94.6		

TABLE 4. Classification Weightings of Normalized Single Spectral Line and Multi-Spectral-Line Combination by PCA

Observed features	Weightings			Observed features	Weightings		
	Single spectral line	Multi-spectral-line combination	Trend		Single spectral line	Multi-spectral-line combination	Trend
C-N/Ca	0.2207	0.3699	↑	Mg/Ca	0.0971	0.1281	↑
C/Ca	0.0825	0.0493	↓	Mn/Ca	0.0083	0.0152	↑
N/Ca	0.0921	0.0792	↓	Si/Ca	-0.0016	-0.0003	↑
O/Ca	0.2027	0.1979	↓	Al/Ca	-0.0051	-0.0019	↑
H/Ca	0.2925	0.1447	↓	Na/Ca	-0.0043	-0.0020	↑
C-C/Ca	0.0138	0.0104	↓	K/Ca	0.0014	0.0096	↑

Note. ↑ Upward trend, ↓ down trend.

as other media in the growth process, and the trace elements in the environment continue to accumulate in the body. This causes the content of metallic mineral elements to vary from one origin to another, so that the metal mineral elements can become traceability indicators of the origin classification of the agricultural products [23]. The weight ratio shows that the multi-spectral line combination method effectively reduces the weight of nonmetal elements. The weighting of C-N/Ca is expected to be higher than the single spectral line. The reason is that the weights of N and O are affected by nitrogen and oxygen in the air. The weight of the multi-spectral line composition of mineral metal elements is higher, which can effectively improve the identification rate of rice origin.

Conclusions. A new approach to products of geographical indication classification by laser-induced breakdown spectroscopy with the multi-spectral line combination support vector machine was developed in this work, aimed at improving the classification accuracy. The results showed that the recognition accuracy rate of the multi-spectral line combination support vector machine was higher than that of a single spectral line, and the classification recognition accuracy

rate increased from 90.8 to 94.6%. The adjusting spectral weighting of the multi-spectral line composition of mineral metal elements is higher, which can effectively improve the identification rate of rice origin. It is feasible to identify products of geographical indication using laser-induced breakdown spectroscopy with a multi-spectral line combination support vector machine.

Acknowledgments. This work was financially supported by Changzhou Sci & Tech Program (Grant No. CJ20200016), by the Natural Science Foundation of Higher Education Institutions of Jiangsu Province (Grant No. 20KJB140013), by Natural Science Foundation of Jiangsu province, youth project (Grant No. BK20200190), by the University-level scientific research projects of Changzhou College of Information Technology (Grant No. CXKZ201910Q), by the Opto-electronic Information Technology Laboratory, CCIT (Grant Nos. KYPT202101Z, KYPT202102Z), 2020 Jiangsu PhD Holder Plan.

REFERENCES

1. S. Kelly, M. Baxter, S. Chapman, C. Rhodes, J. Dennis, and P. Brereton, *Eur. Food Res. Technol.*, **214**, No. 1, 72 (2002).
2. N. N. Wang, X. W. Feng, Y. P. Sun, P. S. Cong, and Z. L. Zhu, *J. Instrum. Anal.*, **32**, No. 10, 1227 (2013).
3. C. Maione, B. L. Batista, A. D. Campiglia, F. Barbosa, and R. M. Barbosa, *Comput. Electron Agric.*, **121**, 101 (2016).
4. J. M. Li, Z. H. Zhu, R. Zhou, N. Zhao, R. X. Yi, X. Y. Yang, X. Y. Li, L. B. Guo, X. Y. Zeng, and Y. F. Lu, *Anal. Chem.*, **89**, No. 15, 8134 (2017).
5. C. M. Li, Z. M. Zou, X. Y. Yang, Z. Q. Hao, L. B. Guo, X. Y. Li, Y. F. Lu, and X. Y. Zeng, *J. Anal. At. Spectrom.*, **29**, No. 8, 1432 (2014).
6. J. Wang, L. Li, P. Yang, Y. Chen, Y. N. Zhu, M. Tong, Z. Q. Hao, and X. Y. Li, *Lasers Med. Sci.*, **33**, No. 6, 1381 (2018).
7. M. A. Gondal, M. A. Shemis, A. A. Khalil, M. M. Nasr, and B. Gondal, *J. Anal. At. Spectrom.*, **31**, No. 2, 506 (2016).
8. P. Yang, R. Zhou, W. Zhang, R. X. Yi, S. S. Tang, L. B. Guo, Z. Q. Hao, X. Y. Li, Y. F. Lu, and X. Y. Zeng, *Food Chem.*, **272**, 323 (2019).
9. P. Yang, Y. N. Zhu, S. S. Tang, Z. Q. Hao, L. B. Guo, X. Y. Li, Y. F. Lu, and X. Y. Zeng, *Appl. Opt.*, **57**, No. 14, 3730 (2018).
10. T. B. Chen, L. Y. Zhang, H. Lin, M. H. Liu, J. Y. Chen, and M. Y. Yao, *J. Anal. At. Spectrom.* (2019).
11. R. X. Yi, X. Y. Yang, R. Zhou, J. M. Li, H. W. Yu, Z. Q. Hao, L. B. Guo, X. Y. Li, Y. F. Lu, and X. Y. Zeng, *Anal. Chem.*, **90**, No. 11, 7080 (2018).
12. R. X. Yi, J. M. Li, X. Y. Yang, R. Zhou, H. W. Yu, Z. Q. Hao, L. B. Guo, X. Y. Li, X. Y. Zeng, and Y. F. Lu, *Anal. Chem.*, **89**, No. 4, 2334 (2017).
13. W. Zhang, R. Zhou, P. Yang, K. Liu, J. J. Yan, P. Y. Gao, Z. Y. Tang, X. Y. Li, Y. F. Lu, and X. Y. Zeng, *Talanta*, **198**, (2019).
14. Z. Y. Tang, R. Zhou, Z. Q. Hao, S. X. Ma, W. Zhang, K. Liu, X. Y. Li, X. Y. Zeng, and Y. F. Lu, *J. Anal. At. Spectrom.* (2019).
15. X. Y. Yang, R. X. Yi, X. Y. Li, Z. F. Cui, Y. F. Lu, Z. Q. Hao, J. C. Huang, Z. X. Zhou, G. X. Yao, and W. X. Huang, *Opt. Express*, **26**, No. 23, 30456 (2018).
16. Y. Yu, L. B. Guo, Z. Q. Hao, X. Y. Li, M. Shen, Q. D. Zeng, K. H. Li, X. Y. Zeng, Y. F. Lu, and Z. Ren, *Opt. Express*, **22**, No. 4, 3895 (2014).
17. G. Bilge, H. M. Velioglu, B. Sezer, K. E. Eseller, and I. H. Boyaci, *Meat Sci.*, **119**, 118 (2016).
18. S. Moncayo, J. D. Rosales, R. Izquierdo-Hornillos, J. Anzano, and J. O. Caceres, *Talanta*, **158**, 185 (2016).
19. J. J. Yan, P. Yang, Z. Q. Hao, R. Zhou, X. Y. Li, S. S. Tang, Y. Tang, X. Y. Zeng, and Y. F. Lu, *Opt. Express*, **26**, No. 22, 28996 (2018).
20. G. Bekefi, *Principles of Laser Plasmas*, New York, Wiley-Interscience (1976).
21. R. W. B. Pearse and A. G. Gaydon. *The Identification of Molecular Spectra*, London, Chapman and Hall (1976).
22. Y. Yu, Z. Q. Hao, C. M. Li, L. B. Guo, K. H. Li, Q. D. Zeng, X. Y. Li, Z. Ren, and X. Y. Zeng, *AcPSn*, **62**, No. 21 (2013).
23. C. F. Zeng, L. F. Zhang, J. D. Xu, J. Shu, and D. H. Liu, *Sci. Technol. Food Ind.*, **36**, No. 4, 367 (2013).