Rapid determination of production date for green tea by near-infrared spectroscopy

ZHUANG Xin-gang[1], WANG Li-li[2], SHI Xue-shun[1], WANG Heng-fei[1], CHEN Qi[3], FANG Jia-xiong[2]

(1. *The 41st Research Institute of China Electronics Technology Group Corporation, Qingdao 266555, China*;
2. *Advanced Research Center for Optics, Shandong University, Ji'nan 250100, China*;
3. *State Key Laboratory of Tea and Agricultural Products Detection (Huangshan), Huangshan, 245000, China*)

Abstract: Coupled with partial least squares (PLS), near infrared (NIR) spectroscopy was applied to develop a fast and nondestructive method to identify the production date of Rizhao green tea aiming at the deficiencies of the existing methods. In the modeling process, the raw spectra were first processed by five-point smoothing and first derivative. And then, moving window back propagation artificial neural network (MW-BP-ANN) was applied to select the characteristic spectral variables. After that, the calibration model was built by PLS, and the optimum model was achieved when 9 principal component scores (PCs) were included. The performances of the calibration models were evaluated according to root mean square error of prediction ε_{RMSEP}, correlation coefficient (C_p) and residual prediction deviation (σ_{RPD}). The optimum results of the calibration model was achieved, and $\varepsilon_{RMSEP}=19.965$, $C_p=0.943$ and $\sigma_{RPD}=3.07$. The overall results sufficiently demonstrate that NIR spectroscopy combined with PLS can be efficiently applied in the rapid identification of green tea production date.

Key words: near-infrared (NIR) spectroscopy; production date of Rizhao green tea; partial least squares (PLS); five-point smoothing and first derivative

CLD number: O433.4 **Document code**: A
Article ID: 1674-8042(2018)02-0199-06 **doi**: 10.3969/j.issn.1674-8042.2018.02.016

0 Introduction

As one of the most popular beverages worldwide, green tea is widely praised and more and more favored by consumers due to its beneficial medicinal properties[1-3], clear taste[4] and strong aroma[5]. The quality of green tea mainly depends on the origins, production date and stir-frying methods. Rizhao (Shandong province, China) green tea is a new brand, which has three unique characteristics, i.e., leaves hypertrophy, high fragrance and resistant to brew. And the three characteristics are all benefit from its unique geographical location, such as large temperature difference between day and night, long wintering period and superior coastal climate condition. Nowadays, Rizhao has developed into the main producing area of green tea in Shandong province and the biggest green tea planting base in northern China. In addition, Rizhao along with Baocheng (Korea) and Shizuoka (Japan) is called the three major coastal green tea cities in the world. Compared with the southern green tea in China, Rizhao green tea is rich in amino acids, vitamins, minerals and other nutrient compositions[6].

Green tea can be generally separated into three varieties of spring tea, summer tea and autumn tea by production season. For the same region and variety, production date will have a major impact on the contents of major chemical compositions of green tea[7]. Researches indicate that spring green tea contains high concentrations of amino acids and moderate concentration of catechin. In contrast, summer green tea has high concentration of catechin and low concentrations of amino acids[8]. In general, the quality of green tea is highly correlated with production date. For example, spring green tea has the highest quality of catechins, autumn green tea takes the second place, and summer green tea is the worst, which perfectly reflect the quality status of Rizhao green tea[6]. As a result, production date

Received date: 2017-12-27

Foundation items: National Basic Research Program of China (No. JSJL2016210A001); State Key Laboratory of Sensor Technology Fund (No. SKT1507)

Corresponding author: ZHUANG Xin-gang (xingangzhuang@163.com)

largely determines the quality and price of green tea.

Near infrared spectroscopy (NIRS) has been widely used in quantitative and qualitative analyses in agricultural products, with the characteristics of rapid analysis, non-invasive detection and high accuracy[9,10]. For production date research, NIRS has been applied to identify the vinegarage[11] and rice freshness[12], et al. In green tea industry, NIRS is mainly applied for the quantitative detection of internal chemical compositions[13], qualitative identification of species[14] and origins[15]. In contrast, there is litter study on the production date of green tea. Currently, the production date of green tea was mainly identified by sensory evaluation[16] or traditional wet chemical analysis methods[17]. Wu et al. studied the feasibility of determinating the total taste scores of green tea infusion by using FT-NIR spectroscopy combined with variable selection methods[18]. Xu et al. identified the production season of green tea by high performance liquid chromatograph (HPLC) and pattern recognition algorithms, and the correct recognition rate was up to 94.8%[7]. However, there are some disadvantages in these sensory evaluation methods. Sensory evaluation methods take many steps and cast much time, and the accuracy of results is affected greatly by subjective factors of human inspectors. Although the traditional analysis methods based on chemical equipment can guarantee analytical precision, they also had many limitations, such as high cost, source waste and second pollution.

The paper was attempted to propose a fast and effective method to determinate the production date of Rizhao green tea by modeling analysis of near infrared diffuse reflectance spectra combined with multivariate calibration method.

1 Materials and methods

1.1 Sample preparation

From April to October, 100 groups of green tea samples were collected from the main producing areas (i.e., the towns of Jufeng, Houcun and Beiguo) of Rizhao, according to the seasonal distribution of total production. All samples were the same specie (Longjing 43), and each sample collected was 100 g. For experimental clarity, each sample was coded as a three-digit dummy label according to the number of production days in the year. For example, the dummy label of samples produced at May 26 is 148. All samples were processed by the same processing factory, and the spectra were collected once the progressing is completed, excluding the impacts of warehousing and stir-frying methods for the internal chemical compositions of green tea.

1.2 Spectra collection

For each sample, (30±0.1) g tea leaf was filled into a 200 mL beaker and pressed to keep the surface flat without any other pretreatment. 10 diffuse reflection spectra (1 050—2 500 nm) of each sample were obtained by using AvaSpec-NIR256/2.5TEC spectrometer (Avantes, Netherlands) equipped with a fiber-optic probe and an InGaAs detector. The distance between the probe and tea was kept at 10 mm. The NIR spectrum of each sample was obtained by taking the average of 40 scans, and each data was collected from different places. The raw data were measured in 6.4 nm intervals, which resulted in 227 variables. For each sample, the mean of 10 spectra was applied in the subsequent analysis. Then the spectral region of 1 300—2 300 nm was selected for further analysis since both ends of the spectra exhibited a high level of noise, as presented in Fig. 1(a).

1.3 Chemical analysis

The contents of polyphenols, epigallocatechin (EGC), catechin (C), epigallocatechin gallate (EGCG), epicatechin (EC), gallocatechin gallate (GCG), epicatechin gallate(ECG) and caffeine were reference measured by the ultra-performance liquid chromatography (UPLC, ACQUITY UPLC I-Class, Waters, USA). (0.2±0.000 1) g of the tea powder accurately weighted was extracted twice with 10 mL of 70% aqueous methanol each for 10 min at 70 ℃. After cooling, the extracts were centrifuged at 3 000 r/min for 10 min. Then the liquid phases of both extracts were collected in a 50 mL volumetric flask. The tea brew was filtered through a 0.22 μm membrane filter, and diluted 5 times with millipore water and analyzed immediately. The analysis results are all collected in Table 1.

As shown in Table 1, the samples of green tea with different production dates are quite different in the contents of polyphenols and caffeine, which are mainly reflected in the difference of bitter taste[19]. Relatively speaking, all standard deviations of EGC, EGCG and ECG are large, which largely determine the quality of green tea[20]. Given that the planting area of Rizhao green tea is relatively small, so the

soil and climatic conditions have no significant difference. As a result, production date is the main determinant of chemical compositions and quality of Rizhao green tea. Based on the pre-established mathematical model, NIR spectroscopy can get the contents and ratio of above chemical compositions, and further calculate the production date of Rizhao green tea.

Table 1 Reference measurements of polyphenols, caffeine and catechin in Rizhao green tea samples

Component	June	July	August	September	October	σ
Polyphenols (%)	16.954	20.215	19.312	17.879	20.883	1.624
EGC (mg/g)	42.465	42.735	42.212	42.774	36.357	2.777
C (mg/g)	8.469	9.542	9.798	8.803	10.047	0.670
EGCG (mg/g)	92.58	103.973	102.256	95.853	113.793	8.202
EC (mg/g)	10.208	10.047	10.22	10.376	9.35	0.403
GCG (mg/g)	2.319	2.5	2.468	2.16	4.443	0.941
ECG (mg/g)	23.931	27.32	26.462	23.916	30.640	2.787
Caffeine (mg/g)	29.176	32.775	31.622	28.795	34.89	2.542

1.4 Model evaluation

All 100 samples were randomly divided into calibration set and prediction set at a ratio of 7∶3. In partial least square (PLS) model building, an optimization solution was achieved by using leave-one-out cross-validation to minimize the estimated generalization error. The robustness and stability of calibration model were evaluated in terms of root mean square error of cross-validation (ε_{RMSECV}) and correlation coefficient of calibration (C_c). Root mean square error of prediction (ε_{RMSEP}), correlation coefficient of prediction (C_p) and residual predictive deviation (σ_{RPD}) were used to evaluate the performance of the prediction set in the prediction process. The σ_{RPD} and ε_{RMSECV} were calculated by

$$\sigma_{RPD} = \frac{\sigma}{\varepsilon_{RMSEP}} = \frac{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2/(n-1)}}{\sqrt{\sum_{i=1}^{n}(y_i - \hat{y})^2/n}}, \quad (1)$$

$$\varepsilon_{RMSECV} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}, \quad (2)$$

where n is the number of samples in the prediction set; $\bar{y}$ is the mean value of reference measurement result for prediction set samples; y_i is the reference measurement result for test prediction sample i, and $\hat{y}_i$ is the estimated result of the model for prediction sample i.

σ_{RPD} is generally applied to evaluate the stability and predictive ability of the calibration model. In general, if $\sigma_{RPD}>3$, the model has high stability and excellent predictive ability, which can be used in the actual detection; if $2.5<\sigma_{RPD}<3$, the model can be used in quantitative analysis, but the prediction accuracy needs to be further improved; if $\sigma_{RPD}<2.5$, the model cannot be applied in quantitative analysis[21].

1.5 Software

For the spectra collection, AvaSoft (AvaSpec NIR256/2.5TEC system) was used. And all algorithms were implemented by using a self-developed NIR analysis software ARCO-NIR, which was developed in MATLAB programming language by MATLAB 2010a (Mathworks Co., USA) under Windows 7.

2 Results and discussion

2.1 Spectral preprocessing

The spectral region of 1 300—2 300 nm was selected for further analysis, since both ends of the spectra exhibited a high level of noise. Fig. 1(a) presents the raw green tea spectra from April to October. Each spectrum is the mean spectrum for that month. And then the raw spectra were further processed by five-point smoothing and first derivate to reduce the interference of noise signal and improve the prediction accuracy of model, as show in Figs. 1(b) and (c).

Smoothing can effectively remove the high frequency noise and enhance the signal to noise ratio (SNR)[22], but it can also weak the detailed spectral information. Hence, the spectra were further processed by first derivate to enhance most of the detailed information and eliminate overlapped spectral lines[23]. Based on the preprocessed spectra, the calibration models were successively built by ARCO-NIR coupled with PLS to determinate the production

date of Rizhao green tea in prediction set.

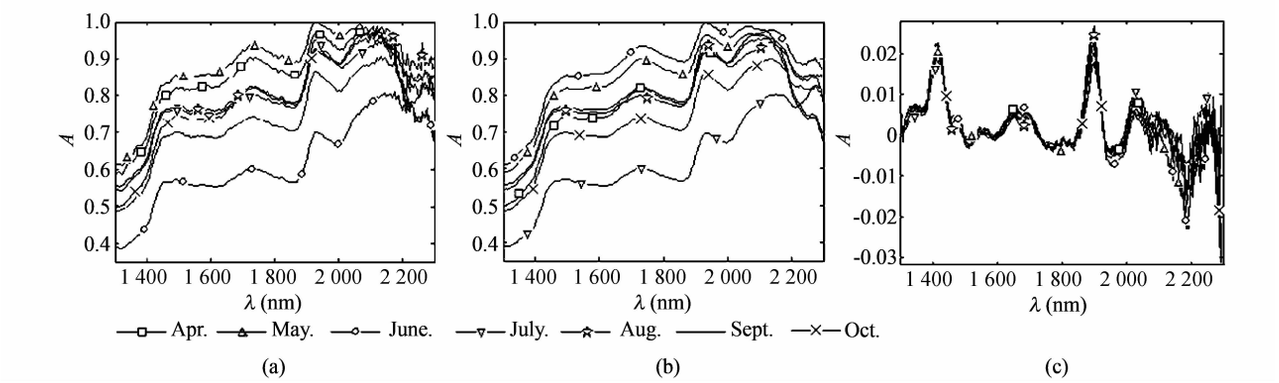

Fig. 1 Near infrared diffuse reflectance spectra of Rizhao green tea

Table 2 presents the modeling results with different processing methods. It can be observed that five-point smoothing and first derivate can significantly improve the prediction performance of calibration model with C_p, ε_{RMSEP} and σ_{RPD} up to 0.822, 87.359 and 2.16, respectively.

Table 2 C_p, ε_{RMSECV} and σ_{RPD} of preliminary model with different spectral preprocessing methods

Preprocessing method	Number of principal components	C_p	ε_{RMSEP}	σ_{RPD}
Raw	8	0.706	105.698	1.786
Five-point smoothing	8	0.737	96.866	1.949
First derivative	6	0.763	93.697	2.014
Five-point smoothing and First derivative	9	0.822	87.359	2.16

2.2 Spectral variables selection

Spectral variables selection can effectively reduce the impact of unrelated variables and improve the predictive ability of model. Traditionally, one component needs at least three wavelengths which are absorption wavelength, interference wavelength and reference wavelength. In this study, the spectra preprocessed by five-point smoothing and first derivate were further processed by moving window back propagation artificial neural network (MW-BP-ANN) to select spectral characteristic variables[24]. There are two main steps to select spectral variables. The first step is setting the width of the window which moves from the left side to the right side of the whole spectral range. And then the wavelengths covered in the window are applied to build back propagation artificial neural network calibration models as the windows moving. After a complete movement is finished, all prediction results are simultaneously calculated by computer. The second step is setting the threshold and selecting characteristic variables according to the calculation results of previous steps. The wavelengths coved by the shade in Fig. 2 are characteristic variables selected by MW-BP-ANN.

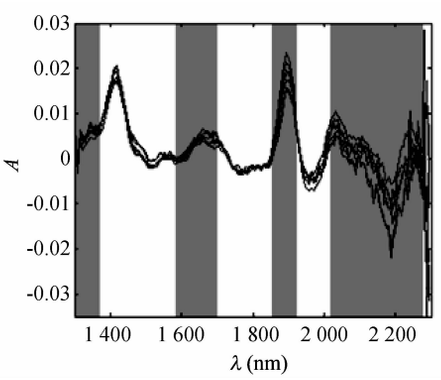

Fig. 2 Selection results of characteristic variables by MW-BP-ANN

2.3 Calibration model building

2.3.1 Selection of optimal principal components

During PLS modeling process, the most important step is to determine the optimum number of principal components (PCs), which has a great impact on the predictive power of final calibration model. The rational use of PCs is an efficient way to reduce noise and make the most of spectral information. Generally, if the number of PCs is too less, the model cannot adequately reflect the relationship between the change of absorbance and production date. But on the contrary, too many PCs can easily result in over-fitting. Because there is more chance to include noise and unrepresentative information from the calibration data with increasing number of PCs. For a nonlinear system, the optimal PCs should overcome the correlation among independent variables and reflect the non-linear relationship between the independent and dependent variables.

In this study, the number of PCs was selected by leave-one-out cross-validation and determined by the lowest ε_{RMSECV}. As presented in Fig. 3, the optimum number of PCs was 9.

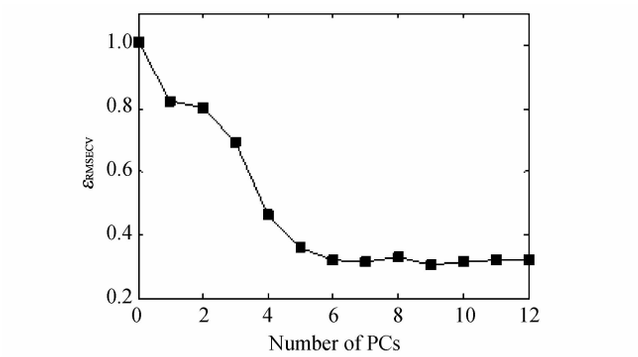

Fig. 3 Selecting optimum principal components

2.3.2 PLS model building

The characteristic spectra of the 70 samples in calibration set and corresponding production dates were taken as input and output variables to build PLS model. After that, the 30 samples in prediction sets were used to test the effectiveness of the model. The predicted results for both calibration and prediction sets were collected in Fig. 4.

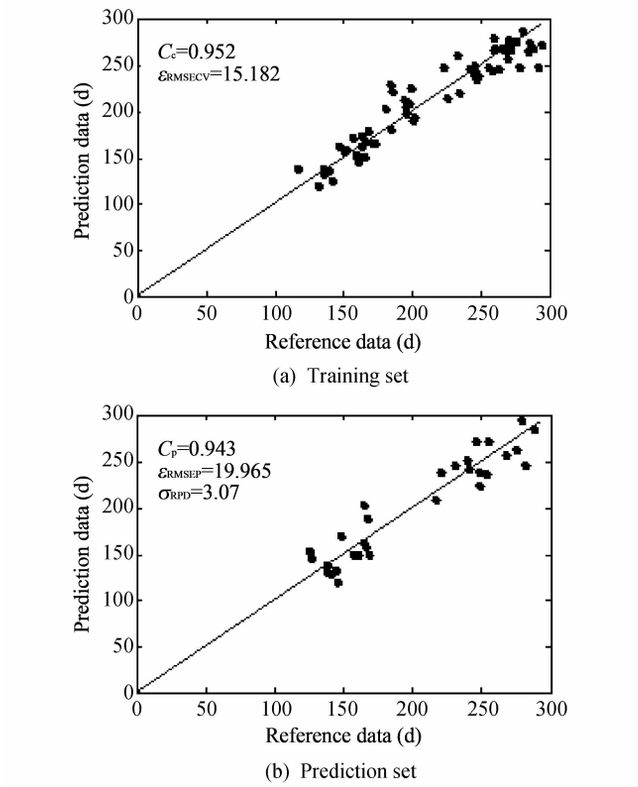

Fig. 4 Final prediction results of PLS model for Rizhao green tea prediction date

As presented in Fig. 4, C_c and ε_{RMSECV} of calibration set were 0.952 and 15.182. The values of C_p, ε_{RMSEP} and σ_{RPD} for prediction set were 0.952, 19.965 and 3.07, respectively, which were calculated by Eqs. (1) and (2). The value of σ_{RPD} is greater than 3, which indicates that the model has excellent predictive ability. Above results demonstrate that NIR spectroscopy can be applied in the prediction of production date of Rizhao green tea.

3 Conclusion

This study was carried out to develop a rapid and non-destructive inspection method to determine the production date of green tea by NIR spectroscopy and PLS. The raw spectra were firstly processed by five-point smoothing and first derivative, and then MW-BP-ANN was used to select characteristic spectral variables. It was found that the best C_p, ε_{RMSEP} and σ_{RPD} of prediction set were 0.943, 19.965 and 3.07 when the number of PCs was 9. The overall results sufficiently demonstrate that NIR spectroscopy can be successfully applied in measurement of production date of Rizhao green tea.

References

[1] Yang C S, Hong J, Hou Z, et al. Green tea polyphenols: antioxidative and prooxidative effects. Journal of nutrition, 2004, 134(11): 3181.

[2] Setiawan V W, Zhang Z F, Yu G P, et al. Protective effect of green tea on the risks of chronic gastritis and stomach cancer. International Journal of Cancer, 2001, 92(4): 600-604.

[3] Shibata K, Moriyama M, Fukushima T, et al. Green tea consumption and chronic atrophic gastritis: a cross-sectional study in a green tea production village. Journal of epidemiology, 2000, 10(5): 310-316.

[4] Narukawa M, Noga C, Ueno Y, et al. Evaluation of the bitterness of green tea catechins by a cell-based assay with the human bitter taste receptor hTAS2R39. Biochemical & Biophysical Research Communications, 2011, 405(4): 620-625.

[5] Luo F, Du X, Zeng Q G, et al. Research on the suitability for different tea cultivars processed into flat green tea. Advanced Materials Research, 2013, 726-731: 4405-4410.

[6] Li Q W. Rizhao green tea catechin quality index. Journal of Agriculture, 2014, 6: 64-66.

[7] Xu W, Song Q, Li D, et al. Discrimination of the production season of Chinese green tea by chemical analysis in combination with supervised pattern recognition. Journal of Agricultural and Food Chemistry, 2012, 60(28): 7064-7070.

[8] Chen Y, Jiang Y, Duan J, et al. Variation in catechin contents in relation to quality of 'Huang Zhi Xiang' Oolong

tea (Camellia sinensis) at various growing altitudes and seasons. Food Chemistry, 2010, 119(2): 648-652.
[9] Ikeda T, Kanaya S, Yonetani T, et al. Prediction of Japanese green tea ranking by Fourier transform near-infrared reflectance spectroscopy. Journal of Agricultural and Food, Chemistry, 2007, 55(24): 9908-9912.
[10] Nicolaï B M, Beullens K, Bobelyn E, et al. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biology and Technology, 2007, 46(2): 99-118.
[11] Li j, Li Z H, Fu X J. Study on rapid non-destructive detection of the freshness of paddy based on NIRS. Spectroscopy and Spectral Analysis, 2012, 32(32): 2126-2130.
[12] Chuang Y K, Hu Y P, Yang I C, et al. Integration of independent component analysis with near infrared spectroscopy for evaluation of rice freshness. Journal of Cereal Science, 2014, 60(1): 238-242.
[13] Chen Q, Zhao J, Chaitep S, et al. Simultaneous analysis of main catechins contents in green tea (Camellia sinensis (L.)) by Fourier transform near infrared reflectance (FT-NIR) spectroscopy. Food Chemistry, 2009, 113(4): 1272-1277.
[14] He Y, Li X, Deng X. Discrimination of varieties of tea using near infrared spectroscopy by principal component analysis and BP model. Journal of Food Engineering, 2007, 79(4): 1238-1242.
[15] Zhuang X, Wang L, Chen Q, et al. Identification of green tea origins by near-infrared (NIR) spectroscopy and different regression tools. Science China Technological Sciences, 2016, 60: 84-90.
[16] Yu H, Wang J, Zhang H, et al. Identification of green tea grade using different feature of response signal from E-nose sensors. Sensors and Actuators B: Chemical, 2008, 128(2): 455-461.
[17] Valera P, Pablos F, Gustavo González A. Classification of tea samples by their chemical composition using discriminant analysis. Talanta, 1996, 43(3): 415-419.
[18] Wu R M, Zhao J W, Chen Q S, et al. Determination of taste quality of green tea using FT-NIR spectroscopy and variable selection methods. Spectroscopy and Spectral Analysis, 2011, 31(7): 1782-1785.
[19] Zou J J, Ding L X, Liang Q, et al. Comparison on taste quality analysis of Rizhao Shandong green tea with two evaluation methods. Journal of Anhui Agricultural Sciences, 2011, 22: 13524-13526.
[20] Chen Q S, Guo Z M, Zhao J W, et al. Quantitative analysis of the catechins contents in green tea with near infrared spectroscopy and net analyte preprocessing algorithm. Journal of Infrared and Millimeter Waves, 2009, 28(5): 357-361.
[21] Li X Y, Wang J H, Huang Y W, et al. Determination of fat, protein and DM in raw milk by Portable short-wave infrared spectrometer. Spectroscopy and Spectral Analysis, 2011, 31(3): 665-668.
[22] Chen Q, Ding J, Cai J, et al. Simultaneous measurement of total acid content and soluble salt-free solids content in chinese vinegar using near-infrared spectroscopy. Journal of Food Science, 2012, 77(2): C222-C227.
[23] Wu D, Chen J, Lu B, et al. Application of near infrared spectroscopy for the rapid determination of antioxidant activity of bamboo leaf extract. Food Chemistry, 2012, 135(4): 2147-2156.
[24] Zhuang X, Wang L, Wu X, et al. Origin identification of Shandong green tea by moving window back propagation artificial neural network based on near infrared spectroscopy. Journal of Infrared and Millimeter Waves, 2016, 35(2): 200-205.

基于近红外光谱分析技术的绿茶生产日期快速鉴别

庄新港[1], 王丽丽[2], 史学舜[1], 王恒飞[1], 陈 琦[3], 方家熊[2]

(1. 中国电子科技集团公司第四十一研究所, 山东 青岛 266555;
2. 山东大学 光学高等研究中心, 山东 济南 250100;
3. 国家茶叶及农产品检测重点实验室(黄山), 安徽 黄山 245000)

摘 要: 针对现有方法检测绿茶生产日期的不足, 采用控制生产日期单一变量的方法, 利用近红外光谱分析技术结合偏最小二乘法对其进行无损伤检测。首先对原始光谱进行五点平滑和一阶微分预处理, 并利用移动窗口-BP 神经网络算法(MW-BP-ANN)提取特征光谱变量。然后采用偏最小二乘算法验证方式建立校正模型, 并采用预测均方根误差(ε_{RMSEP})、相关系数(C_p)和相对分析误差(σ_{RPD})来评价模型质量。当主成分数为 9 时获得最优模型, 3 个评价指标分别为 19.965, 0.943 和 3.07。研究结果表明, 近红外光谱结合偏最小二乘法可用于对绿茶生产日期的快速无损伤检测。

关键词: 近红外光谱; 日照绿茶生产日期; 偏最小二乘; 五点平滑和一阶微分

引用格式: ZHUANG Xin-gang, WANG Li-li, SHI Xue-shun, et al. Rapid determination of production date for green tea by near-infrared spectroscopy. Journal of Measurement Science and Instrumentation, 2018, 9(2): 199-204. [doi: 10.3969/j.issn.1674-8042.2018.02.016]