

太赫兹时域光谱葛粉掺薯粉检测研究

李斌*, 杜秀洋, 刘燕德, 胡军

华东交通大学机电与车辆工程学院, 江西 南昌 330013

摘要 使用太赫兹时域光谱(THz-TDS)技术对葛粉中掺薯粉的含量进行定性、定量检测。对葛粉中掺薯粉的光谱数据进行采集,利用偏最小二乘法(PLS)建立葛粉掺薯粉定性模型以判断葛粉中是否掺薯粉,得到 PLS 的总误判率为 0%,模型相关系数为 0.925。结果表明:PLS 可实现葛粉中是否掺薯粉的定性判别。再利用 PLS 和最小二乘支持向量机(LS-SVM)算法分别建立葛粉中掺薯粉的定量模型。利用 PLS 建立的模型的相关系数为 0.932,预测集的均方根误差(RMSE)为 2.6%;利用 LS-SVM 建立的模型的相关系数为 0.957,预测集的 RMSE 为 1.6%,结果表明:利用 LS-SVM 的葛粉掺薯粉定量模型更准确,说明 THz-TDS 技术可用于对葛粉中掺薯粉进行快速、有效、无损检测。

关键词 光谱学;太赫兹时域光谱(THz-TDS);葛粉;偏最小二乘(PLS);最小二乘支持向量机(LS-SVM)

中图分类号 O439

文献标识码 A

doi: 10.3788/LOP56.203001

Detection Research of Seed Potato Starch in Kudzu Using Terahertz Time-Domain Spectroscopy

Li Bin*, Du Xiuyang, Liu Yande, Hu Jun

School of Mechatronics & Vehicle Engineering, East China Jiaotong University, Nanchang, Jiangxi 330013, China

Abstract This paper uses terahertz time-domain spectroscopy (THz-TDS) to analyze the content of seed potato starch in kudzu qualitatively and quantitatively. The spectral data of kudzu powder mixed with seed potato starch was collected, and the qualitative model was established by partial least squares method (PLS) to determine whether the starch powder was mixed with potato powder. The total false positive rate of PLS is 0%, and the model correlation coefficient is 0.925. The results show that PLS can be used to determine whether the starch powder is qualitatively differentiated. PLS and least squares support vector machine (LS-SVM) were used to establish a quantitative prediction model for the content of seed potato starch in kudzu respectively; for PLS prediction model, the determination coefficient is 0.932, and the root mean square error (RMSE) of the predicted set is 2.6%; for LS-SVM prediction model, the determination coefficient is 0.957, and the RMSE of LS-SVM of the predicted set is 1.6%. The results show that the LS-SVM quantitative prediction model is excellent. The research shows that THz-TDS can be used to rapidly and effectively detect the content of seed potato starch in kudzu qualitatively and quantitatively.

Key words spectroscopy; terahertz time-domain spectroscopy (THz-TDS); kudzu; partial least squares (PLS); least squares support vector machine (LS-SVM)

OCIS codes 300.6495; 160.2290; 110.6795; 120.3688

1 引言

葛根粉(以下简称“葛粉”)是从葛根藤根中提取的一种纯天然营养品^[1],具有生津止渴、抗菌解毒、补肾健脾、益胃安神、清心明目等功能^[2]。葛粉中含

有类黄酮的活性成分,如葛根素,对糖尿病、心血管疾病和消化系统疾病的患者具有很好的疗效,且无副作用,还富含人体所需的各类氨基酸和微量元素,是当今市场上不可或缺的畅销健康食品。然而,市场上的不法商家为赚取高额利润,在葛粉中掺杂其

收稿日期:2019-04-16;修回日期:2019-04-23;录用日期:2018-05-21

基金项目:国家自然科学基金(31760344);江西省优势科技创新团队建设计划(20153BCB24002);南方山地果园智能化管理技术与装备协同创新中心(赣教高字[2014]60号);研究生创新创业项目(YC2018-S249)

* E-mail: libingioe@126.com

他廉价的淀粉,如薯粉等,这无疑会导致葛粉市场的信誉度下降。国内外研究人员为鉴别真假葛粉进行了诸多研究,常用鉴别技术有显微镜比较法^[3]、性状比较法^[4]、感官识别法、液相色谱法^[5]、傅里叶变换红外光谱法^[6]、近红外漫反射光谱法^[7]。但上述方法均存在缺点,如:显微镜法和性状比较法操作繁琐,检测精度低;感官识别法检测不准确;液相色谱法检测成本高,操作繁琐;傅里叶变换红外光谱法和近红外光谱法检测精度低。

太赫兹时域光谱(THz-TDS)技术是一种新起的且非常有效的相干检测技术。许多极性大分子的振动和旋转能级间的转换均在太赫兹频率范围内^[8-10],因此,可通过 THz-TDS 技术定性、定量检测许多极性大分子。此外,THz-TDs 技术还具有高信噪比、高带宽、低能性、无需前处理等优点^[11-12]。在农产品检测中,Baek 等^[13]采用 THz-TDS 技术对掺杂三聚氰胺的奶粉进行定量检测,Ge 等^[14]采用 THz-TDS 技术与化学计量学结合的方法对小麦品质进行鉴别,张文涛等^[15]采用 THz-TDS 技术对红木进行检测,文献^[16-18]中采用 THz-TDS 技术对食品中的苯甲酸含量进行检测。目前,使用 THz-TDS 技术鉴别葛粉真伪还未有报道。因此,本文探讨了采用 THz-TDS 技术,结合化学计量学方法,对葛粉掺薯粉进行定性、定量快速检测的可行性,以期加强市场监管。

2 实验

2.1 实验系统

采用日本 Advantest 公司的 THz-TDS 仪器进

表 1 4 类样品中薯粉的含量

Table 1 Four classes of seed potato starch sample concentration

Sample number	Mass fraction /%	Number	Sample number	Mass fraction /%	Number
1	0	8	3	20	20
2	10	20	4	30	8

表 2 16 组样品中薯粉的含量

Table 2 Sixteen groups of seed potato starch sample concentration

Sample number	Mass fraction /%	Sample number	Mass fraction /%
1	0.5	9	12.5
2	2.0	10	14.0
3	3.5	11	15.5
4	5.0	12	17.0
5	6.5	13	18.5
6	8.0	14	20.0
7	9.5	15	30.0
8	11.0	16	40.0

行实验,利用自带的光谱采集软件 Spectroscopic Measurement 收集光谱数据。THz-TDS 仪器检测原理图如图 1 所示,激光器 1(laser 1)作用于发射系统(emitter)产生太赫兹光谱,太赫兹光谱通过光学系统作用于样品(sampled),利用同频控制器(synchronized control)、激光器 2(laser 2)和检测系统(detector)收集太赫兹光谱信息,光谱信息通过模数转换电路(A/D)接口转换为数字信号。利用快速傅里叶变换(FFT)和计算机处理光谱信息。

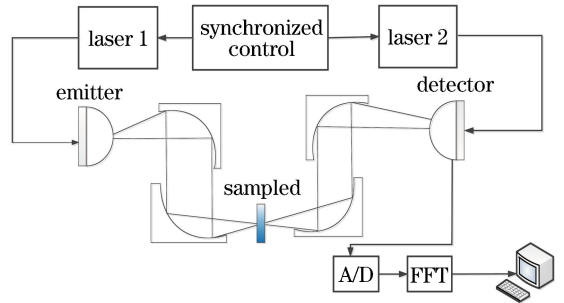


图 1 太赫兹时域光谱仪方块图

Fig. 1 Block diagram of terahertz time domain spectrometer

2.2 实验样品

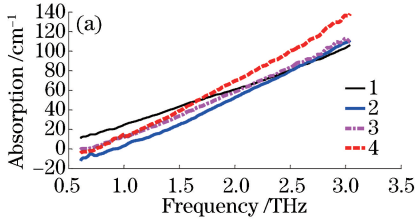
所用实验材料为农家纯天然的葛粉和薯粉。按照设计好的浓度梯度,分别称取葛粉和薯粉进行配样,并放在振荡器上振荡使其均匀混合。实验设计两组样品,第一组样品分为 4 类,用以定性判别,每一类中薯粉的含量(质量分数)如表 1 所示;第二组样品分为 16 小组,用以定量判别,每一小组中薯粉的含量表 2 所示。

将配制好的样本在压片机上进行压片,厚度在 1 mm 左右。对于第一组样品,每一类按表 1 的个数进行压片,每个压片采集 1 次光谱数据,可得 56 个光谱数据。对于第二组样品,每一个含量的样本制作 2 个压片,每个压片采集 2 次光谱数据,可得 64 个光谱数据。

2.3 光谱采集

利用 THz-TDS 仪器对两组样品进行光谱测量,扫描范围为 0~10.0 THz,分辨率为 7.6 GHz,扫描次数设置为 8192 次。在每次测量前,需对压片厚度进行重新测量并设置,以确保测量的准确性。

每个样品扫描 4 次,以平均光谱数据作为样品数据,采集到的光谱数据如图 2 所示。其中,图 2(a)为第一组光谱数据图,图 2(b)为第二组部分光谱数据图。从图中可知:样品的吸收谱没有明显的特征峰,



且 4 类样品的吸收谱存在较大差异,且随着薯粉含量增加,光谱吸收系数曲线的斜率在不断增大。因此,可根据不同薯粉含量的样品的吸收系数差异进行定性、定量分析。

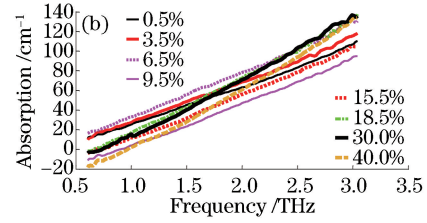


图 2 (a) 4 类样品的吸收谱;(b)部分样品的吸收谱

Fig. 2 (a) Absorption spectra of four classes of seed potato starch samples; (b) absorption spectra of partial samples

3 建模方法

3.1 偏最小二乘法

偏最小二乘法(PLS)是对光谱矩阵 \mathbf{X} 和浓度矩阵 \mathbf{Y} 同时进行分解,并在分解时考虑两者相互关系,该方法为主成分分析、多元线性回归和典型相关分析的完美结合。具体公式为

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^T + \boldsymbol{\alpha}_Y, \mathbf{X} = \mathbf{T}\mathbf{P}^T + \boldsymbol{\alpha}_X, \quad (1)$$

式中: \mathbf{U} 和 \mathbf{T} 分别为 \mathbf{Y} 和 \mathbf{X} 的得分矩阵; \mathbf{P} 和 \mathbf{Q} 为 \mathbf{X} 和 \mathbf{Y} 的载荷矩阵; $\boldsymbol{\alpha}_X$ 和 $\boldsymbol{\alpha}_Y$ 分别为 \mathbf{X} 和 \mathbf{Y} 的 PLS 拟合残差矩阵。

将 \mathbf{T} 和 \mathbf{U} 进行线性回归,即

$$\mathbf{U} = \mathbf{T}\mathbf{B}, \mathbf{B} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{Y}. \quad (2)$$

在预测时,首先根据 \mathbf{P} 求出样品未知光谱阵 \mathbf{X}_1 的得分 \mathbf{T}_1 ,根据(1)式和(2)式可求出浓度预测值为

$$\mathbf{Y}_1 = \mathbf{T}_1(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{Y}\mathbf{Q}. \quad (3)$$

3.2 最小二乘支持向量机

最小二乘支持向量机(LS-SVM)采用最小二乘线性系统作为损失函数,通过求解一组线性方程组代替传统支持向量机(SVM)采用的较复杂的二次规划方法,具有缩短训练时间、减少计算复杂程度及提高泛化能力的优点。

LS-SVM 算法的目标优化函数 H 为

$$\min H(\mathbf{w}, \boldsymbol{\alpha}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{1}{2}\beta \sum_{i=1}^n \alpha_i^2, \quad (4)$$

式中: \mathbf{w} 为权重向量; β 为正则化参数; α_i 为误差; n 为校正集样本数, i 为样品编号。

约束条件为

$$y_i = \mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_i) + b_i + \alpha_i, \quad (5)$$

式中: b_i 为样品固化矩阵元; $\boldsymbol{\phi}$ 为关系函数; \mathbf{x}_i 为校正级的输入变量。

定义 Lagrange 函数,即

$$L(\mathbf{w}, \mathbf{b}, \mathbf{A}, \boldsymbol{\alpha}) = H(\mathbf{w}, \boldsymbol{\alpha}) -$$

$$\sum_{i=1}^n A_i [\mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_i) + b_i + \alpha_i - y_i], \quad (6)$$

式中: y_i 为校正级的输出变量; A_i 为 Lagrange 系数。

优化问题可转化为解线性方程,即

$$\begin{bmatrix} \mathbf{0} & \mathbf{L}^T \\ \mathbf{L} & \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) + \frac{1}{\beta}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{A} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{y} \end{bmatrix}, \quad (7)$$

式中: \mathbf{b} 为样品固化矩阵; $\mathbf{L}^T = [1, 1, \dots, 1]$; \mathbf{I} 为单位矩阵; $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = [\boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}_j)]$, $i, j = 1, 2, \dots, n$; $\mathbf{A}^T = [A_1, A_2, \dots, A_n]$; $\mathbf{y}^T = [y_1, y_2, \dots, y_n]$ 。可根据(7)式求解出 \mathbf{A} 和 \mathbf{b} 。

对于未知变量样本 \mathbf{x} , LS-SVM 的预测值为

$$\mathbf{y}(\mathbf{x}) = \sum_{i=1}^n A_i \mathbf{K}(\mathbf{x}, \mathbf{x}_i) + \mathbf{b}. \quad (8)$$

4 结果与讨论

4.1 定性分析

将采集到的 4 类样品光谱数据导入 The Unscrambler 中,光谱数据选择 0.608~3.04 THz 范围内的吸收系数。对原始数据进行 PLS 分析。

图 3 为主成分因子决定图。从图中可知:预测和建模均方根误差(RMSE)随主成分数的增加而逐渐下降,当主成分数为 5 时, RMSE 基本保持不变,因此可确定最佳主成分数为 5。

对 4 类样品进行预测集和建模集的分配,其中预测集和建模集的分配比例为 1:3, PLS 建模所得模型散点图如图 4 所示。其中: S_1 为第一类样品和第二类样品的分界线,预测值为 1.5; S_2 为第二类样

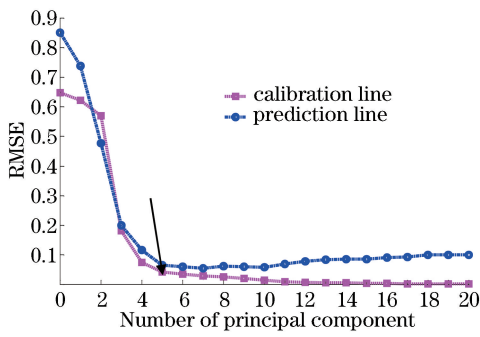


图3 定性分析模型 RMSE 随主成分因子数变化图
Fig. 3 Tendency chart of RMSE with principal component factor in qualitative prediction model



图5 定量分析模型中 RMSE 随主成分因子数变化图
Fig. 5 Tendency chart of RMSE with principal component factor in quantitative prediction model

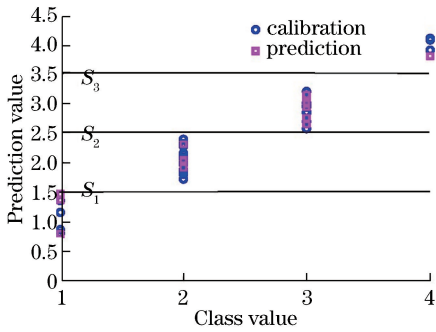


图4 4类样品的分类变量与预测值散点图
Fig. 4 Scatter plots of classified variables and predicted values for four samples

品和第三类样品的分界线,预测值为 2.5; S_3 为第三类样品和第四类样品的分界线,预测值为 3.5。从散点图可知:预测集样品中没有出现误判,但有 2 个纯葛粉处在阈值 1.5 附近,有 1 个薯粉质量分数为 20% 的样品处于阈值 2.5 附近,故总误判率为 0%。该建模集的相关系数 R_c^2 为 0.938,预测集的相关系数 R_p^2 为 0.925。故可利用 PLS 建模,以误判率和相关系数为指标,实现葛粉中是否掺薯粉的定性判别,且判别效果很好。

4.2 定量分析

将采集到的 16 组葛粉含量不同的样品的 64 个光谱数据导入 The Unscrambler 中,对其进行 PLS 建模分析。64 个光谱数据的预测集和建模集的分配比例为 1:3。把采集到的光谱数据导入 MATLAB 软件中,进行 LS-SVM 建模,64 个光谱的预测集和建模集的分配比例为 1:3。

图 5 是 PLS 建模的定量分析主成分因子决定图。从图中可知:预测和建模 RMSE 随主成分数的增加而逐渐下降,当主成分数为 7 时,预测和建模的 RMSE 均达到较低值,并且当主成分数再增加时预测和建模的 RMSE 变化较小,故确定最佳主成分数为 7。

在定量分析中,利用 PLS 建模所得模型散点图如图 6 所示。从图 6 可知:经过 PLS 建模分析后,建模集的相关系数 R_1^2 为 0.966,预测集的相关系数 R_2^2 为 0.932,建模集的 RMSE 为 1.8%,预测集的 RMSE 为 2.6%。

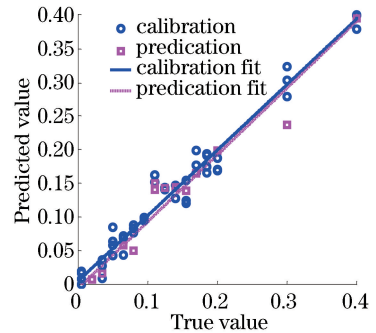


图6 PLS 预测薯粉含量散点图
Fig. 6 PLS predicted seed potato starch content scatter plot

4.3 LS-SVM 分析

在定量分析中,利用 LS-SVM 建模所得预测集模型散点图如图 7 所示。从图 7 可知:预测集的相关系数 R_3^2 为 0.957,预测集的 RMSE 为 1.6%。

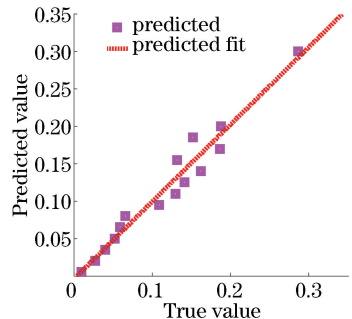


图7 LS-SVM 预测薯粉含量散点图
Fig. 7 LS-SVM predicted seed potato starch content scatter plot

4.4 讨论

采用 PLS 对样品进行定性分析, 所得误判率为 0, 结果很好, 因此没有利用其他化学计量学方法进行定性分析。采用 PLS 和 LS-SVM 两类建模方法进行定量分析, 所得结果如表 3 所示。从表 3 中可知: 采用 LS-SVM 方法预测集的均方根值和相关系数均优于 PLS 方法, 因此, 葛粉掺薯粉定量模型需采用 LS-SVM 建模。

表 3 PLS 建模模型结果与 LS-SVM 建模结果的对比

Table 3 Result contrast between PLS model and LS-SVM model

Method	RMSE /%	R^2
PLS	2.6	0.932
LS-SVM	1.6	0.957

5 结论

采用 THz-TDS 技术, 结合化学计量学方法, 对葛粉掺薯粉进行定性、定量检测。首先, 采用 PLS 方法建立葛粉掺薯粉的定性模型, 定性模型的误判率为 0%, 预测集相关系数 R_p^2 为 0.925, 结果较好。利用 PLS 和 LS-SVM 分别建立葛粉掺薯粉的定量检测模型。通过分析比较可知, 利用 LS-SVM 建模模型的预测 RMSE 为 1.6%, 预测值相关系数为 0.957, 说明该模型可较好实现葛粉掺薯粉的定量检测。本研究为葛粉掺薯粉高效、准确检测提供了一种新的途径。

参 考 文 献

- [1] Liao H B, He Z F, Wang G C, *et al.* Research development and prospect for *Pueraria DC* [J]. Science and Technology of Food Industry, 2003, 24(2): 81-83.
廖洪波, 贺稚非, 王光慈, 等. 葛根的研究进展及展望[J]. 食品工业科技, 2003, 24(2): 81-83.
- [2] Lu C Y, Chen G X, Bu X Y, *et al.* Comparisons of physio-chemical properties of commercial and laboratory-made kudzu (*Pueraria lobata*) starches [J]. Food Science, 2009, 30(21): 50-52.
卢成瑛, 陈功锡, 卜晓英, 等. 不同来源葛根淀粉的理化性质比较[J]. 食品科学, 2009, 30(21): 50-52.
- [3] Sun L, Du W, Zhou M, *et al.* Research on identification methods of kudzu vine root starch [J]. Chinese Journal of Health Laboratory Technology, 2012, 22(5): 1151-1152, 1154.
孙亮, 杜威, 周敏, 等. 葛根淀粉的鉴别方法研究[J]. 中国卫生检验杂志, 2012, 22(5): 1151-1152,

1154.

- [4] Zhao W H, Bai W D, Wang Q, *et al.* Study on rheological properties of kudzu power paster [J]. Food & Machinery, 2008, 24(4): 41-45.
赵文红, 白卫东, 王琴, 等. 葛根全粉糊流变特性初探[J]. 食品与机械, 2008, 24(4): 41-45.
- [5] Shen M J, Feng X W, Zhu Z L. Study on HPLC fingerprint and pattern recognition in radix puerariae [J]. Journal of Inner Mongolia University (Natural Science Edition), 2012, 43(1): 73-76.
沈梦洁, 冯昕轶, 朱仲良. 葛根 HPLC 指纹图谱及模式识别研究[J]. 内蒙古大学学报(自然科学版), 2012, 43(1): 73-76.
- [6] Liu J, Li J C, Chen J, *et al.* Discrimination of kudzu starch adulteration by Fourier transform infrared spectroscopy [J]. Food Science, 2011, 32(8): 226-230.
刘嘉, 李建超, 陈嘉, 等. 葛粉掺假的傅里叶变换红外光谱法鉴别研究[J]. 食品科学, 2011, 32(8): 226-230.
- [7] Chen J, Liu J, Ma Y Q, *et al.* Rapid detection of kudzu starch adulteration by near-infrared diffuse reflectance spectroscopy [J]. Food Science, 2014, 35(8): 133-136.
陈嘉, 刘嘉, 马雅钦, 等. 葛粉掺假的近红外漫反射光谱快速检测[J]. 食品科学, 2014, 35(8): 133-136.
- [8] Li P P, Zhang Y, Ge H Y, *et al.* Research advances in terahertz spectroscopy detection of harmful substances in food [J]. Science and Technology of Food Industry, 2017, 38(3): 372-375.
李鹏鹏, 张元, 葛宏义, 等. 太赫兹光谱技术在食品质量安全检测中的应用研究进展[J]. 食品工业科技, 2017, 38(3): 372-375.
- [9] Zhang Z Y, Zhang X. Research developments on the applications of terahertz time domain spectroscopy [J]. Spectroscopy and Spectral Analysis, 2016, 36(s1): 54-55.
张卓勇, 张欣. 太赫兹时域光谱技术应用研究进展[J]. 光谱学与光谱分析, 2016, 36(s1): 54-55.
- [10] Cao C, Zhang Z H, Zhao X Y, *et al.* Review of terahertz time domain and frequency domain spectroscopy [J]. Spectroscopy and Spectral Analysis, 2018, 38(9): 2688-2699.
曹灿, 张朝晖, 赵小燕, 等. 太赫兹时域光谱与频域光谱研究综述[J]. 光谱学与光谱分析, 2018, 38(9): 2688-2699.
- [11] Li B, Long Y, Liu H, *et al.* Research progress on

- terahertz technology and its application in agriculture [J]. Transactions of the Chinese Society of Agricultural Engineering, 2018, 34(2): 1-9.
- 李斌, 龙园, 刘欢, 等. 太赫兹技术及其在农业领域的应用研究进展[J]. 农业工程学报, 2018, 34(2): 1-9.
- [12] Ren F T, Zhang Y. Research advances in terahertz spectroscopy detection of harmful substances in agriculture products [J]. Electronics Quality, 2018 (5): 27-31.
- 任方涛, 张元. 太赫兹技术在农产品质量安全检测的应用研究进展[J]. 电子质量, 2018(5): 27-31.
- [13] Baek S H, Lim H B, Chun H S. Detection of melamine in foods using terahertz time-domain spectroscopy[J]. Journal of Agricultural and Food Chemistry, 2014, 62(24): 5403-5407.
- [14] Ge H Y, Jiang Y Y, Xu Z H, *et al.* Identification of wheat quality using THz spectrum [J]. Optics Express, 2014, 22(10): 12533-12544.
- [15] Zhang W T, Wang S Y, Zhan P P, *et al.* Method of identifying red wood based on terahertz time-domain spectroscopy[J]. Acta Optica Sinica, 2017, 37(2): 0230006.
- 张文涛, 王思远, 占平平, 等. 基于太赫兹时域光谱技术的红木检测方法[J]. 光学学报, 2017, 37(2): 0230006.
- [16] Liu Y D, Du X Y, Li B, *et al.* Detection of benzoic acid in arrowroot by terahertz technology[J]. Laser & Optoelectronics Progress, 2019, 56(4): 041101.
- 刘燕德, 杜秀洋, 李斌, 等. 基于太赫兹技术的葛粉中苯甲酸的检测[J]. 激光与光电子学进展, 2019, 56(4): 041101.
- [17] Rao G F, Huang L, He X W, *et al.* Comparison of tea variety discriminating models based on laser induced breakdown spectroscopy [J]. Laser & Optoelectronics Progress, 2018, 55(2): 023002.
- 饶刚福, 黄林, 何秀文, 等. 基于激光诱导击穿光谱的茶叶品种识别模型对比[J]. 激光与光电子学进展, 2018, 55(2): 023002.
- [18] Zhou K P, Bai X F, Bi W H. Detection of chemical oxygen demand in water based on multi-spectral fusion of ultraviolet and fluorescence [J]. Laser & Optoelectronics Progress, 2018, 55(11): 113003.
- 周昆鹏, 白旭芳, 毕卫红. 基于紫外-荧光多光谱融合的水质化学需氧量检测[J]. 激光与光电子学进展, 2018, 55(11): 113003.