

# 太赫兹光谱技术对红薯淀粉中明矾含量的检测方法研究

欧阳爱国, 郑艺蕾, 李 斌, 胡 军, 杜秀洋, 李 雄

华东交通大学机电与车辆工程学院, 水果智能光电检测技术与装备国家地方联合工程研究中心, 江西 南昌 330013

**摘 要** 明矾是一种可以改良粉条粉丝易断粗糙特性的违法添加剂, 明矾的含量过高进入人体后会直接影响身体健康。结合太赫兹光谱技术探索红薯淀粉中明矾含量快速检测方法。采用太赫兹时域光谱系统(Terahertz time domain spectroscopy, THz-TDS)于常温下获取 0.5~7 THz 范围内红薯淀粉、明矾及其混合物的光谱数据。因 0~0.5 THz 测得的频谱均为噪声, 高频段区域的吸收系数大、信噪比低, 故选取 0.5~2 THz 波段的吸收系数谱和折射率谱进行分析。发现明矾在该波段存在明显的特征吸收峰, 可作为指纹特征用于物质识别。分别采用 Savitzky-Golay 卷积平滑(SG Smoothing, SG 平滑)、基线校正(Baseline)、归一化(Normalization)等方法进行光谱预处理, 再结合偏最小二乘(partial least squares, PLS)对红薯淀粉中明矾含量建立预测模型。结果表明, 采用原始光谱、SG 平滑、Baseline、Normalization 等光谱数据建立 PLS 模型的最佳因子数(principal component factors)分别为 3, 3, 3 和 2; 校正集相关系数( $r_c$ )分别为 0.982, 0.980, 0.982 和 0.984; 预测集相关系数( $r_p$ )分别为 0.982, 0.979, 0.982 和 0.987; 校正集均方根误差(root mean square error of calibration, RMSEC)分别为 0.011, 0.012, 0.012 和 0.011; 预测集均方根误差(root mean square error of prediction, RMSEP)分别为 0.013, 0.014, 0.013 和 0.012; 可知归一化预处理后建立 PLS 模型效果最佳。为对比分析线性(PLS)与非线性(LS-SVM)两种定量模型方法的预测精度, 采用相同预处理方法后的红薯淀粉中明矾含量光谱数据建立最小二乘支持向量机(least squares support vector machine, LS-SVM)预测模型, 选用径向基函数(RBF)作为核函数。结果表明, 归一化预处理后建立 LS-SVM 模型效果最佳, 其预测集均方根误差(RMSEP)为 0.0047, 预测集相关系数( $r_p$ )为 0.997 2。发现对红薯淀粉中明矾含量建立 LS-SVM 预测模型的稳定性更好、精确度更高。采用太赫兹时域光谱结合 LS-SVM 和 PLS 对红薯淀粉中明矾含量进行定量分析。结果表明, 采用归一化预处理后的 LS-SVM 比 PLS 模型的预测效果更优, 可能是红薯淀粉与明矾混合物中含有更多的非线性信息。研究表明, 太赫兹时域光谱结合化学计量学方法可为红薯淀粉中明矾含量的定量分析提供快速精确的分析方法。

**关键词** 太赫兹时域光谱; 明矾; 红薯淀粉; 偏最小二乘; 最小二乘支持向量机

**中图分类号:** O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)03-0727-06

## 引 言

我国是红薯生产大国, 红薯在全国主要农作物生产中位居第四, 主要因其具有耐旱、易生长、产量高特性。红薯中营养成分丰富包含淀粉、纤维素、糖、蛋白质、多种维生素及矿物元素<sup>[1]</sup>。红薯内含有 10%~30% 的淀粉<sup>[2]</sup>, 人们为提高它的经济效益, 满足现在的生活需求, 将其进行深加工, 成大众喜爱的红薯粉丝、粉条。在生产中不良厂家为防止粉丝易断、粗糙, 将明矾作为增筋剂, 改善粉丝口感质感, 获

取最大利益。明矾学名为硫酸铝钾, 化学分子式为  $KAl(SO_4)_2 \cdot 12H_2O$ , 国家明文规定对于普通粉条生产过程中明矾的添加量必须少于  $300 \text{ mg} \cdot \text{kg}^{-1}$ 。明矾中的铝元素进入人体难以新陈代谢, 铝元素积累过量, 有害身体健康, 增加人体患病机率。因此对红薯淀粉中明矾的含量进行定量检测研究具有重要的现实意义。

目前, 食品中对明矾含量的测定是通过确定总铝含量, 常用的铝含量测定为铬天青 S 分光光度法<sup>[3]</sup>。该方法需要对样品采用湿式消解法进行前处理, 处理过程中使用硝酸、高氯酸消解。测定铝含量其他方法有石墨炉原子吸收法、电感

收稿日期: 2019-03-05, 修订日期: 2019-07-28

基金项目: 国家自然科学基金项目(31760344), 江西省优势科技创新团队建设计划项目(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), 江西省研究生创新专项资金项目(YC2018-S249)资助

作者简介: 欧阳爱国, 1968年生, 华东交通大学机电与车辆工程学院教授 e-mail: ouyang1968711@163.com

耦合等离子体原子发射光谱法、动力学分析法等,但都具有耗时长、环境污染大、检测步骤复杂的缺点。太赫兹波通常是指频率在 0.1~10 THz 其波长为 30  $\mu\text{m}$ ~3 mm 的电磁辐射,介于微波与红外之间<sup>[4-5]</sup>。对于有机物分子,其转动能级和振动多数分布于此波段范围。因此,太赫兹波通过样品可获取大量的物理和化学信息,太赫兹光谱中可分析得到它的复介电常数、吸收系数和折射率等参数。目前,太赫兹技术在食品中的检测研究较少。Chen<sup>[6]</sup>等使用 THz-TDS 对米粉中吡虫啉的含量进行检测,结合化学计量学方法建立预测模型,实现对米粉中吡虫啉快速检测。管爱红<sup>[7]</sup>等对红薯淀粉、明矾及混合样品进行太赫兹光谱检测,发现明矾在太赫兹波段具有特征峰。随明矾含量增大,吸收峰更明显且折射率增大。但文中未结合化学计量学方法对红薯淀粉中明矾含量建立模型,探索红薯淀粉中明矾含量的快速检测方法。

本文采用 THz-TDS 技术对红薯淀粉中明矾含量进行定量分析检测。通过获取明矾和红薯淀粉及其混合物的吸收系数谱和折射率谱,根据制备样品在 THz 波段具有的特征指纹,进行定性识别。采用三种不同预处理方法处理红薯淀粉中明矾含量的光谱数据,分别建立红薯淀粉中明矾含量的 PLS 和 LS-SVM 数学模型,对模型预测结果分析比较,获得最优模型效果,为红薯淀粉中明矾含量的快速检测提供方法。

## 1 实验部分

### 1.1 样品制备

实验所用的红薯淀粉采购于市场,红薯淀粉为固体粉末,明矾采购于上海阿拉丁生化有限公司,明矾为无色透明结晶且分析纯度大于 99.5%,使用前未做纯化处理。对红薯淀粉中明矾含量的太赫兹波段进行定性研究与定量研究,先将明矾晶体放入陶瓷研钵充分研磨成粉末状,再根据实验设计红薯淀粉与明矾二者总量为 10 g 按照不同质量比称量。使用涡旋搅拌器对配制的样品进行 3 min 的振荡,确保均匀分布。样品制备采取压片法,制片压力约为 8 t,时间约为 1 min,其厚度分布于 0.50~1.00 mm 之间,制备压片的直径为 13 mm,表面光滑无裂痕。本实验设计 37 组不同质量分数的样品,每组质量分数制备 4 个样品,一共 148 个样品。质量分数如表 1 所示。

### 1.2 仪器及光谱采集

本实验使用 TAS7500SU 太赫兹时域光谱仪由日本 Advantest 公司生产,该光谱仪测量频范围为 0.5~7 THz,频谱分辨率为 7.6 GHz,采样扫描速度为 8 ms,扫描次数设置为 8 192 次,透射模式下测量。在采集光谱前将实验室温度控制在 20  $^{\circ}\text{C}$ ,湿度控制在 25% 以下,开机后,通过无油静音空气压缩机向设备充入气体保证环境干燥,预热约 30 min 待设备内部的空气湿度降低至 5%,使用仪器开始采集光谱。对每个质量分数的 4 个样品各采集 3 条太赫兹光谱,取其平均光谱数据进行处理分析,共获得 148 条光谱数据。

### 1.3 光学参数提取

样品放置于 THz-TDS 装置中,采集参考和样品的太赫

表 1 37 组明矾样品质量分数 ( $\text{g} \cdot \text{g}^{-1}$ , %)

Table 1 Mass fraction of 37 alum samples

| 样品<br>编号 | 明矾质量<br>分数 | 样品<br>编号 | 明矾质量<br>分数 | 样品<br>编号 | 明矾质量<br>分数 |
|----------|------------|----------|------------|----------|------------|
| 1        | 0.1        | 14       | 3.0        | 27       | 10         |
| 2        | 0.2        | 15       | 3.5        | 28       | 11         |
| 3        | 0.3        | 16       | 4.0        | 29       | 12         |
| 4        | 0.4        | 17       | 4.5        | 30       | 13         |
| 5        | 0.5        | 18       | 5.0        | 31       | 14         |
| 6        | 0.6        | 19       | 6.0        | 32       | 15         |
| 7        | 0.7        | 20       | 6.5        | 33       | 16         |
| 8        | 0.8        | 21       | 7.0        | 34       | 17         |
| 9        | 0.9        | 22       | 7.5        | 35       | 18         |
| 10       | 1.0        | 23       | 8.0        | 36       | 19         |
| 11       | 1.5        | 24       | 8.5        | 37       | 20         |
| 12       | 2.0        | 25       | 9.0        |          |            |
| 13       | 2.5        | 26       | 9.5        |          |            |

兹时域谱,再通过傅里叶变换将时域信号转换为频域信号,采用 Dorney<sup>[8]</sup>和 Duvillaret<sup>[9]</sup>等提出的方法进行数据处理得到样品的吸收系数和折射率

$$n(\omega) = \frac{\varphi(\omega)c}{\omega d} + 1 \quad (1)$$

$$\partial(\omega) = \frac{2}{d} \ln \left\{ \frac{4n(\omega)}{\rho(\omega)[n(\omega) + 1]^2} \right\} \quad (2)$$

式中, $n(\omega)$ 为样品的折射率, $\partial(\omega)$ 为样品的吸收系数, $\omega$ 为频率, $d$ 为样品的厚度, $\varphi(\omega)$ 为相位, $\rho(\omega)$ 为振幅。

### 1.4 模型的建立与评价

#### 1.4.1 建模方法

偏最小二乘法是一种基于多元线性回归方法,可以解决样本数量少和多重线性相关问题<sup>[10]</sup>,采用 PLS 对红薯淀粉中明矾混合物建立模型进行定量分析,根据 PLS 模型预测的公式

$$Y = \sum_{i=1}^N \beta_i \lambda_i + B \quad (3)$$

式中, $Y$ 表示模型混合物中明矾含量的预测值, $N$ 为建模光谱变量数, $i$ 为光谱第  $i$  个变量, $\beta$ 为能量谱强度, $\lambda$ 为回归系数, $B$ 为模型的截距。

最小二乘支持向量机(LS-SVM)在支持向量机(SVM)基础上将不等式约束替换为等式约束,使用拉格朗日算子求出最优解,LS-SVM 将非线性问题转化为解线性方程组问题<sup>[11]</sup>。因此,可以有效降低求解难度,提高对问题的处理速度。本文主要选用 RBF 作为核函数,其表达式为

$$K(X_k, X_i) = \exp \left( - \frac{\|X_k - X_i\|^2}{(2\sigma^2)} \right) \quad (4)$$

式中, $X_k$ 是样本点, $X_i$ 是核函数中心点, $\sigma^2$ 是内核参数。

#### 1.4.2 模型评价

通过校正集均方根误差(RMSEC)、预测集均方根误差(RMSEP)、校正相关系数( $r_c$ )、预测相关系数( $r_p$ )等参数来评价模型的优劣。若  $R_c$  和  $R_p$  数值越接近 1,说明其校正集和预测集相关程度越高。当 RMSEC 和 RMSEP 的值越接近 0,表明模型的建模效果好和预测能力越强<sup>[12]</sup>。

## 2 结果与讨论

### 2.1 红薯淀粉和明矾的 THz 光谱

采用太赫兹时域光谱系统在室温下对红薯淀粉和明矾样品检测获得吸收系数谱、折射率谱。本实验仪器测量频谱范围为 0.5~7 THz。在 0~0.5 THz 测得的频谱均为噪声且高频段区域吸收系数大、信噪比低。明矾特征吸收峰大致位于 1 THz, 红薯淀粉无特征吸收, 故选取 0.5~2 THz 范围频谱进行分析。图 1 为明矾和红薯淀粉在 0.5~2 THz 的吸收谱

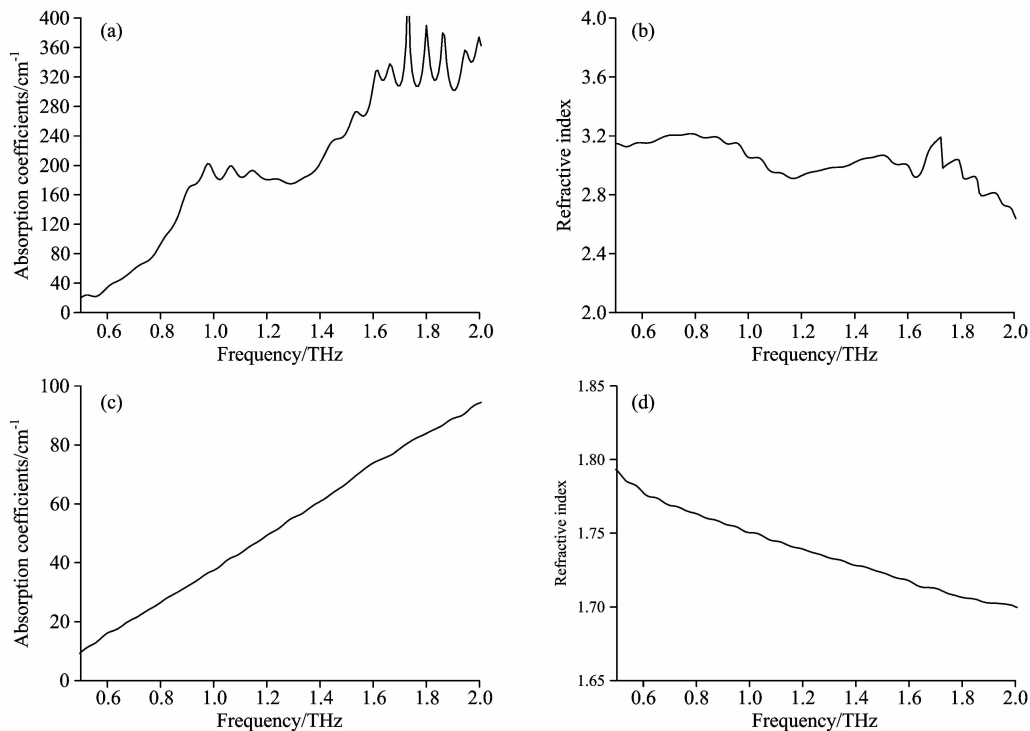


图 1 明矾和红薯淀粉的吸收谱和折射谱图

(a): 明矾吸收谱; (b): 明矾折射率谱; (c): 红薯淀粉吸收谱; (d): 红薯淀粉折射率谱

Fig. 1 Absorption and refraction spectra of alum and sweet potato starch

(a): Absorption spectra of Alum; (b): Refractive index spectrum of Alum;

(c): Absorption spectra of sweet potato starch; (d): Refractive index spectrum of sweet potato starch

### 2.2 红薯淀粉和明矾混合物的 THz 光谱

采用太赫兹时域光谱系统对红薯淀粉中不同明矾含量的混合物扫描, 再经计算获得样品的吸收系数谱。图 2 为红薯淀粉中不同明矾含量混合物的吸收谱。当明矾含量较低时, 明矾样品的光谱没有出现明显特征峰, 随着明矾含量的增大, 可清晰的观察在 0.980, 1.065 和 1.146 THz 有明显特征吸收峰。且发现明矾样品的 THz 吸收系数随着频率增加逐渐增大, 可能是样品的散射引起。

### 2.3 红薯淀粉中明矾含量的定量分析

本文经分析红薯淀粉和明矾的吸收谱和折射谱, 发现吸收谱的特征信息较丰富。建立红薯淀粉中明矾含量吸收谱数据的定量模型进行分析。由谱图可知光谱特征信息紧密分布在 1 THz, 也考虑到高频段光谱波动太大, 于是分析的光谱

和折射谱。由图 1(a)中可知, 明矾在 0.980, 1.065 和 1.146 THz 处存在吸收峰。其吸收谱在此波段内呈现上升趋势, 在 1.45~2 THz 出现几个强烈吸收峰可能是系统的信噪比低所导致。由图 1(b)所示, 明矾折射率随频率的增加而减小。还发现在三个特征吸收峰的每个相对应位置处, 其折射率均有特征变化, 表明存在反常色散现象。由图 1(c)可知, 红薯淀粉的吸收系数随频率的增加而增大, 无特征吸收。图 1(d)可看出红薯淀粉折射率则随频率的增加而平缓下滑。对于同种物质, THz 的吸收谱比折射率的谱图信息更丰富。

数据选择频段在 0.5~3 THz。

#### 2.3.1 偏最小二乘

本文对样品按 3:1 比例随机划分建模和预测, 建模集有 111 个样品和预测集有 37 个样品。为了减少采集光谱过程的噪声、背景等因素影响, 对原始光谱使用不同预处理方法分别为 SG 平滑、基线校正、归一化。将红薯淀粉中明矾含量光谱数据经不同预处理后建立 PLS 数学模型。表 2 为不同预处理方法后建立红薯淀粉中明矾含量 PLS 模型结果。结果显示, 使用归一化预处理后 PLS 模型效果最佳, 其主因子数(PCs)为 2, 预测均方根误差(RMSEP)为 0.012, 预测相关系数( $R_p$ )为 0.987。比较原始光谱与 3 种预处理方法的评价指标值, 可知光谱预处理方法对 PLS 建模效果无显著的提高。

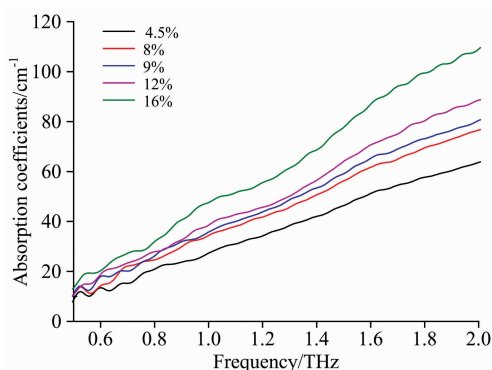


图 2 红薯淀粉中不同明矾含量混合物吸收谱

Fig. 2 Absorption spectra of mixtures with different Alum contents in sweet potato starch

表 2 不同预处理方法的 PLS 模型结果

Table 2 PLS model results of different pretreatment methods

| 预处理方法 | 最佳因子数 | $R_c$ | $R_p$ | RMSEC | RMSEP |
|-------|-------|-------|-------|-------|-------|
| 原始    | 3     | 0.982 | 0.982 | 0.011 | 0.013 |
| SG 平滑 | 3     | 0.980 | 0.979 | 0.012 | 0.014 |
| 基线校正  | 3     | 0.982 | 0.982 | 0.012 | 0.013 |
| 归一化   | 2     | 0.984 | 0.987 | 0.011 | 0.012 |

选择合适主成分因子数建模是决定 PLS 模型性能好坏的关键, 在 PLS 建模及模型预测中最大主成分因子数通常设置为 20。建模中若选取 PCs 过少, 会丢失光谱的重要信息致使模型精度变差; 若 PCs 过多, 导入较多噪声信号影响模型精度。因此, 本文采用均方根误差的最小值对应的因子数确定最佳主成分因子数。图 3 是归一化后 PLS 建模的均方根误差与主成分因子关系图, 由图 3 所示, 主因子数为 0~2 校正和预测均方根误差几乎重合呈现快速下降趋势。主因子数在 2 之后均方根误差趋于稳定的状态。因此, 最佳主成分因子数为 2。同理判断出其他预处理方法最佳因子数为 3。

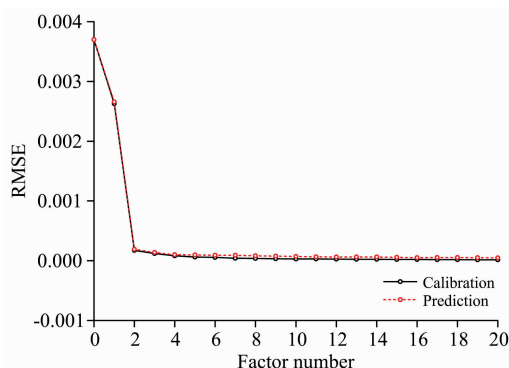


图 3 均方根误差随主成分因子变化图

Fig. 3 RMSE varies with principal component factor

图 4 是偏最小二乘建模和预测明矾含量散点图。表示其真实值与预测值的相关性。校正集的均方根误差 (RMSEC) 为 0.011、相关系数 ( $R_c$ ) 为 0.984, 预测集的均方根误差 (RMSEP) 为 0.012、相关系数 ( $R_p$ ) 为 0.987。可知该模型的

精度高, 预测能力达到较好的效果。

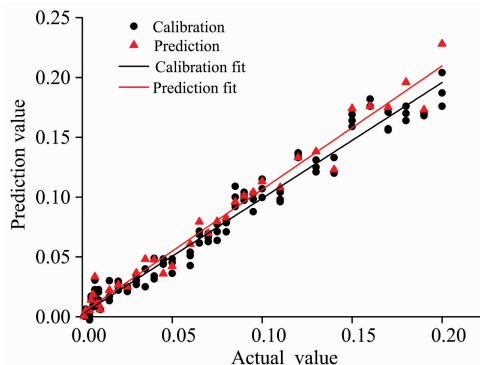


图 4 PLS 建模和预测明矾含量散点图

Fig. 4 Scatters plots of calibration and prediction results of alum content by PLS

### 2.3.2 最小二乘支持向量机

本文展开不同数学建模方法的对比, 分别采用原始光谱及 SG 平滑、基线校正、归一化方法处理红薯淀粉中明矾含量的数据作为 LS-SVM 模型的输入值。在 LS-SVM 定量模型中选择 RBF\_kernel, 其中需确定  $\gamma$  (正则化参数) 和  $\sigma^2$  (内核参数), 在一定程度上这两个参数会影响模型精确度。参数的确定采用两步搜索法, 采取大步长查找确定核参数边界范围, 再采取小步长搜索出最佳参数。表 3 为 LS-SVM 预测

表 3 LS-SVM (RBF\_kernel) 模型的预测结果

Table 3 Prediction results of LS-SVM (RBF\_kernel) model

| 预处理方法 | 参数           |            | $R_p$   | RMSEP   |
|-------|--------------|------------|---------|---------|
|       | $\gamma$     | $\sigma^2$ |         |         |
| 原始    | 17 250.866 3 | 512.174 5  | 0.996 9 | 0.005 7 |
| SG 平滑 | 12 878.598 2 | 577.169 7  | 0.997 5 | 0.005 5 |
| 基线校正  | 28 131.142 3 | 576.884 2  | 0.996 7 | 0.005 5 |
| 归一化   | 167.687 6    | 231.372 7  | 0.997 2 | 0.004 7 |

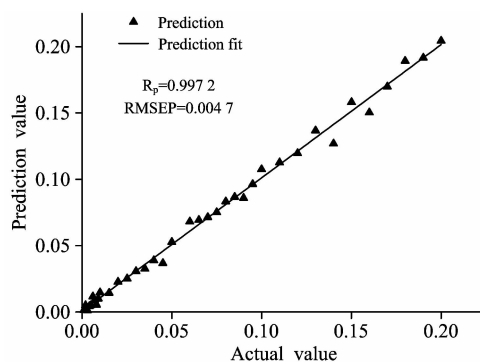


图 5 LS-SVM 明矾含量预测模型散点图

Fig. 5 Scatters plots of prediction results of alum content by LS-SVM

明矾含量结果, 由结果可知归一化处理后建模预测精度更高, 预测均方根误差 (RMSEP) 为 0.004 7, 相关系数 ( $R_p$ ) 高

达 0.997 2, 其对应的参数  $\gamma=167.6876$ ,  $\sigma^2=231.3727$ 。图 5 为归一化后 LS-SVM 预测明矾含量散点图。

### 2.3.3 模型对比分析

采用三种预处理方法分别建立红薯淀粉中明矾含量的 PLS 和 LS-SVM 模型。由表 2 和表 3 可知, 采用归一化方法建立的两种模型精度最高, 其中 PLS 建立的红薯淀粉中明矾含量预测模型的 RMSEP 为 0.012、 $R_p$  为 0.987。LS-SVM 建立的红薯淀粉中明矾含量预测模型的 RMSEP 为 0.004 7、 $R_p$  为 0.997 2。因此, PLS 和 LS-SVM 可对红薯淀粉中明矾含量进行准确检测。经对比, LS-SVM 模型整体上的预测均方根误差更低, 其相关系数更高。且该模型具有更好的处理光谱数据和样品含量之间的非线性关系能力。结果表明 LS-SVM 比 PLS 预测明矾含量效果更佳、模型更为稳定, 原因可能是红薯淀粉与明矾混合物中含有更多的非线性信息。

## 3 结 论

采用太赫兹时域光谱技术结合 PLS 和 LS-SVM 对红薯淀粉中明矾含量定性和定量分析研究。首先获得样品在 0.5~2.0 THz 的吸收谱和折射率谱。发现明矾在 0.980, 1.065 和 1.146 THz 处有明显的吸收峰, 表明可作定性识别。其次采用三种预处理方法对太赫兹吸收光谱数据进行预处理, 分别结合 PLS 和 LS-SVM 对红薯淀粉中明矾含量建立数学模型。结果表明, 原始光谱经归一化预处理后建立 LS-SVM 的预测效果更佳、模型更为稳定。其预测均方根误差为 0.004 7, 预测相关系数高达 0.997 2。研究表明, THz-TDS 可用于红薯淀粉中明矾的定性定量研究。本研究为太赫兹时域光谱技术用于食品领域的快速检测提供方法参考。

## References

- [ 1 ] DU Fang-ling, WANG Wen-liang, WANG Shou-jing(杜方岭, 王文亮, 王守经). Food and Nutrition in China(中国食物与营养), 2008, (9): 27.
- [ 2 ] LI Fei, CHEN Meng-xue, HU Ya-jie(李 飞, 陈梦雪, 胡雅婕). Agricultural Science & Technology and Equipment(农业科技与装备), 2015, (6): 55.
- [ 3 ] LIU Dong-hao(刘冬豪). China Condiment(中国调味品), 2012, 37(6): 89.
- [ 4 ] Fitch M J, Osiander R. Johns Hopkins Apl Technical Digest, 2004, 25(4): 348.
- [ 5 ] Josette El Haddad, Frederick de Miollis, Joyce Bou Sleiman et al. Analytical Chemistry, 2014, 86: 4927.
- [ 6 ] Chen Zewei, Zhang Zhuoyong, Zhu Ruohua, et al. Journal of Quantitative Spectroscopy& Radiative Transfer, 2015, 167: 1.
- [ 7 ] GUAN Ai-hong, LI Zhi, GE Hong-yi(管爱红, 李 智, 葛宏义). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2018, 38(1): 267.
- [ 8 ] Dorney T D, Baraniuk R G, Mittleman D M. Journal of the Optical Society of America A, 2001, 18(7): 1562.
- [ 9 ] Duvillaret L, Garet F, Coutaz J L. Applied Optics, 1999, 38(2): 409.
- [ 10 ] OUYANG Ai-guo, ZHANG Yu, CHENG Meng-jie, et al(欧阳爱国, 张 宇, 程梦杰). Chinese Optics(中国光学), 2017, 10(6): 752.
- [ 11 ] DING Shi-fei, QI Bing-juan, TAN Hong-yan(丁世飞, 齐丙娟, 谭红艳). Journal of University of Electronic Science and Technology of China(电子科技大学学报), 2011, 40(1): 2.
- [ 12 ] HAO Yong, SUN Xu-dong, GENG Xiang(郝 勇, 孙旭东, 耿 响). Food Science(食品科学), 2013, 34(18): 137.

# Study on the Detection Method of Alum Content in Sweet Potato Starch by Terahertz Spectroscopy

OUYANG Ai-guo, ZHENG Yi-lei, LI Bin, HU Jun, DU Xiu-yang, LI Xiong

School of Mechatronics Engineering, East China Jiaotong University, National and Local Joint Engineering Research Center of Fruit Intelligent Photoelectric Detection Technology and Equipment, Nanchang 330013, China

**Abstract** Alum is an illegal additive that can improve the fragile characteristics of vermicelli. If the content of alum is excessive, it will directly affect the health of the body. This paper combines terahertz spectroscopy to explore a rapid detection method for alum in sweet potato starch. The spectral data of sweet potato starch, alum and their mixtures in the range of 0.5~7 THz were obtained by Terahertz Time Domain Spectroscopy (THz-TDS) at room temperature. Since the spectrum measured by 0~0.5 THz is noise, the absorption coefficient of the high-band region is large, and the signal-to-noise ratio is low, the absorption coefficient spectrum and the refractive index spectrum of the 0.5~2 THz band were selected for analysis. It was found that alum has obvious characteristic absorption peaks in terahertz band, which can be used as fingerprint features for material identification. Savitzky-Golay convolution smoothing, Baseline, Normalization were used for spectral pretreatment, and combined with partial least squares(PLS) a prediction model for alum content in sweet potato starch was established. The results showed that the principal component factors of the PLS model were 3, 3, 3, 2 using the original, SG smoothing, Baseline, Normalization spectral data, respectively. The correlation coefficient of calibration( $r_c$ ) were 0.982, 0.980, 0.982, 0.984, respectively. The correlation coefficient of prediction ( $r_p$ ) were 0.982, 0.979, 0.982, and 0.987, respectively. The root mean square error of correction (RMSEC) were 0.011, 0.012, 0.012, and 0.011, respectively. The root mean square error of prediction (RMSEP) were 0.013, 0.014, 0.013, and 0.012, respectively. The PLS model had the best effect after normalization pretreatment. In order to compare and analyze the prediction accuracy of linear (PLS) and nonlinear (LS-SVM) quantitative model methods, the least square support vector machine was established using the spectral data of alum in the sweet potato starch after the same pretreatment method. For the prediction model, the radial basis function was chosen as the kernel function. The results showed that the LS-SVM model is the best after normalization preprocessing. The RMSEP of the prediction set was 0.004 7, and the correlation coefficient of the prediction set was 0.997 2. It was found that the LS-SVM prediction model for the alum content in sweet potato starch was more stable and more accurate. The content of alum in sweet potato starch was quantitatively analyzed by terahertz time domain spectroscopy combined with LS-SVM and PLS. The results showed that the LS-SVM with normalized pretreatment has better prediction effect than the PLS, which may be more nonlinear information in the mixture of sweet potato starch and alum. Studies have shown that terahertz time-domain spectroscopy combined with chemometric methods can provide a fast and accurate analytical method for the quantitative analysis of alum in sweet potato starch.

**Keywords** Terahertz time-domain spectroscopy; Alum; Sweet potato starch; Partial least squares; Least squares support vector machine

(Received Mar. 5, 2019; accepted Jul. 28, 2019)