

基于 LS-SVM 和 THz 光谱技术的面粉中苯甲酸检测研究

胡 军, 刘燕德*, 孙旭东, 欧阳爱国, 蔡会周, 刘洪量

华东交通大学机电与车辆工程学院, 江西 南昌 330013

摘 要 面粉(小麦粉)是中国北方大部分地区的主食, 苯甲酸是重要的酸型食品防腐剂, 为了便于食品长期保存, 往往会添加苯甲酸以便延长食品保存时间。但食用添加苯甲酸过量的小麦粉会对身体健康产生严重危害。太赫兹技术是一种新兴的检测技术, 由于处于特殊的 0.1~10 THz 的太赫兹频段, 在食品安全检测方面体现出了很强的应用潜力。主要致力于探索太赫兹光谱技术检测苯甲酸的合理性、可行性, 利用太赫兹时域光谱技术对面粉中的食品添加剂苯甲酸进行实验研究。实验获取了面粉和苯甲酸的太赫兹时域光谱和频域光谱, 其吸收系数显示苯甲酸的特征吸收峰在 1.94 THz 波段, 面粉的太赫兹吸收系数几乎以一定的斜率增加, 说明可以用 THz-TDS(Terahertz time domain spectrum)技术对面粉中的苯甲酸进行特征识别。为建立面粉中添加剂苯甲酸的定量检测模型, 实验获取了面粉中掺杂不同百分比(质量分数)苯甲酸的太赫兹时域光谱, 计算得到吸收系数谱。实验发现吸收峰幅度的变化是与苯甲酸的含量成正比的, 苯甲酸含量增加吸收峰幅度变大。首先探索了不同光谱预处理方法对太赫兹光谱的影响, 采用如平滑校正、多元散射校正、基线校正和归一化等方法对原始光谱进行校正处理。校正之后, 建立相应的 PLS (partial least squares)模型以选择最优预处理方法。然后分别建立苯甲酸浓度和太赫兹吸收系数的 MLR (multiple linear regression)、PLS 和 LS-SVM(partial least squares support vector machines)回归模型, 并对比分析不同模型的优劣。将光谱数据归一化后建立的 PLS 模型更具有优势, 预测相关系数 R_p 为 0.979, 预测均方根误差 RMSEP 为 1.30%。LS-SVM 与 PLS 和 MLR 模型相比, LS-SVM 模型可以获得更好的建模结果, LS-SVM 的预测相关系数 R_p 为 0.987, 预测均方根误差 RMSEP 为 1.10%。利用 MLR 方法仅使用 1.946 和 1.869 THz 两个波段点进行建模, 建模效果预测相关系数 R_p 为 0.955, 预测均方根误差 RMSEP 为 1.90%。通过该研究为面粉中苯甲酸添加剂的无损检测提供了新的解决方案, 也为其他类型的添加剂的检测提供了方法指导, 对促进面粉行业的健康发展具有重要的意义。

关键词 面粉; 太赫兹光谱; 偏最小二乘支持向量机; 苯甲酸; 波段比

中图分类号: O439; O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)11-3566-05

引 言

食品安全与我们每个人的生活密切相关。虽然苯甲酸常作为食品防腐剂使用, 但食用添加过量苯甲酸的食品会对人体健康造成极大伤害^[1]。

目前, 对面粉的品质检测更多是借助于化学方法检测, 国内外较常用的检测方法分为生物测定方法和理化分析法。其中, 生物测定方法主要包括免疫分析法和生物传感器法,

理化分析法主要有气相色谱法和液相色谱法等^[4]。这些检测方法大多属于有损检测, 而且检测成本高、操作复杂^[2-3]。因此, 探索一种简单、快速、实时的面粉质量和安全检测方法是迫切的。THz 的波谱范围介于远红外和微波之间, 频率范围为 0.1~10 THz 之间, 各种有机分子的弱相互作用, 低频振动吸收频率均位于 THz 频段, 具有独特的检测优势^[5]。太赫兹波具有安全性高、透视性好以及波谱分辨能力强等特点, 因此太赫兹光谱技术在很多领域都具有广阔的应用前景^[6-7]。

收稿日期: 2019-03-22, **修订日期:** 2019-07-14

基金项目: 国家“十二五”(863)计划项目(SS2012AA101306), 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项目(赣教高字[2014]60号), 国家自然科学基金项目(2002017018), 江西省研究生创新资金项目(YC-2019-B106)资助

作者简介: 胡 军, 1992年生, 华东交通大学机电与车辆工程学院博士研究生, 教师 e-mail: 1825868002@qq.com

* 通讯联系人 e-mail: jxliuyd@163.com

近年来,利用太赫兹时域光谱检测技术开展食品方面的检测研究已成为研究热点。Liu 等^[8]采用太赫兹光谱与化学计量工具相结合的方法,利用太赫兹光谱测定脂肪酸的差异来鉴别掺假食品,实验结果验证了太赫兹光谱鉴别掺假食品具有可行性。Lu 等^[9]采用太赫兹时域谱(THz-TDS)与化学计量学相结合,对具有相似化学结构和性质的谷氨酸和谷氨酰胺二元混合物进行定性和定量的分析,成功地对混合物中两种氨基酸组分 THz 谱进行解析。Ge 等^[10]通过太赫兹光谱技术获取了黄曲霉毒素 B1 在频率范围 0.4~1.6 THz 光谱信息并且建立了回归预测模型,这极大的证明了太赫兹对黄曲霉毒素 B1 检测的可行性。

目前,关于通过太赫兹光谱法定量检测添加到面粉中的苯甲酸的研究很少。本文利用太赫兹光谱检测技术建立 MLR^[11-12]、PLS^[13]和 LS-SVM^[14-16]检测模型,评估上述模型并探索最佳定量测定模型。

1 实验部分

1.1 材料

本实验所用面粉购买于某大型超市,苯甲酸样品购买于阿拉丁试剂官网,其纯度为分析纯度大于或等于 99.7%。按照设计的浓度梯度配制样品(浓度分别为 0.04%, 0.08%, 0.1%, 0.2%, 0.4%, 0.5%, 1%, 1.5%, ..., 20%),苯甲酸含量的浓度范围在 0.040%~19.99%不等。本次实验样品经过研磨、烘干、称量、混合、压片等步骤进行制备。按照上述方法依次制备 44 组不同浓度梯度面粉和苯甲酸混合样品。每个浓度梯度样品制备 4 个,共得到混合样品 176 个,另外分别制备一组面粉和一组苯甲酸样品作为对照。所有样本采用 K-S(kennard-stone)算法将样本按照 3:1 左右的比例划分为建模集和预测集,分别建立对应模型。如表 1 为面粉中苯甲酸含量真值在建模集和预测集中的分布统计结果。

表 1 面粉中苯甲酸含量真值在建模集和预测集的分布统计

Table 1 Benzoic acid content distribution of modeling and prediction

样品集分类	数量	最小值 / %	最大值 / %	平均值 / %	标准差
全部样本	176	0.048	19.999	9.5445	6.162
建模集	132	0.048	19.999	9.7596	6.122
预测集	44	0.048	19.999	9.5973	6.135

1.2 光谱采集

本实验采用的检测装置是日本 Advantest 公司的 TAS7500 太赫兹时域光谱仪,频谱测量范围设置为 0.1~5.0 THz,仪器的分辨率设置为 7.6 GHz,扫描次数为 4 048 次·点⁻¹,该装置包括两个超短脉冲光纤激光器,脉冲中心波长为 1 550 nm,最大输出功率为 50 mW,系统扫描采样率 8 ms·次⁻¹。为了减少随机误差对实验结果所造成的影响,对每个样本进行了 4 次测量。

1.3 参数提取方法

根据 Timothy^[17]和 Duvillaret^[18]等提出的光学参数提取模型,采用快速傅里叶变换(FFT)获取了 THz 脉冲在频率上的频谱分布,可表述为式(1), $A(\omega)$ 表示电场幅值, $\phi(\omega)$ 为电场的相位, $E(t)$ 为太赫兹时域波形。

$$E(\omega) = A(\omega)\exp[-i\phi(\omega)] = \int E(t)\exp(-i\omega t)dt \quad (1)$$

$$n(\omega) = \frac{\varphi(\omega)c}{\omega L} + 1 \quad (2)$$

$$\alpha(\omega) = \frac{2k(\omega)\omega}{c} = \frac{2}{d} \ln \left[\frac{4n(\omega)}{\rho(\omega)(n(\omega)+1)^2} \right] \quad (3)$$

其中, ω 是频率, $k(\omega)$ 为消光系数, $\rho(\omega)$ 为幅值比函数, $\varphi(\omega)$ 为参考信号和样本信号的相位差, d 为样品厚度, c 为真空中的光速。式(2)和式(3)分别为计算检测样品的折射率、吸收系数的公式。

1.4 模型评价方法

本文对面粉样品的太赫兹光谱进行校正处理,建立 PLS 模型,并通过所建立的 PLS 模型评价其预处理方法,选择最佳预处理方法。将最佳预处理后的太赫兹光谱分别建立相应的 PLS 和 LS-SVM 光谱检测模型。本模型评价的关键是建模集和预测的相关系数和均方根误差参数。建模集参数预测集 R_c 和 RMSEC 以及预测集参数 R_p 和 RMSEP 系数共同决定了检测模型的质量好坏。检测模型的相关系数越高,均方根的误差越小,则模型的精度越高。RMSEC 的数值和 RMSEP 的数值越接近,则建立的模型就越稳定。利用 Matlab2014b 软件进行相关分析和建立 LS-SVM 模型。

1.5 相关分析

采用两波长比值的相关分析方法,找出本计算定量分析的最佳波长对。后续分析共采用 1.0~3.0 THz 频段的光谱合计 264 个变量。波长对的决定系数 R^2 计算如式(4)所示,计算评价指标 R^2 值越大越好。

$$R^2 = \frac{r_1^2 + r_2^2 - 2r_1 \times r_2 \times r_x}{1 - r_x^2} \quad (4)$$

其中, γ 为苯甲酸的浓度值, r_1 为 γ 与 x_1 间的相关系数, r_2 为 γ 与 x_2 间的相关系数, r_x 为波长 x_1 和波长 x_2 间的相关系数, R^2 为波长 x_1 和波长 x_2 的决定系数。

1.6 最小二乘支持向量机

最小二乘支持向量机是一种针对小样本建立的统计学方法^[14-16],常用的核函数是线性核函数(Link-kernel)及径向基核函数(RBF-kernel),其公式如式(5)和式(6)所示。其中 x_i 表示为样本点, x_j 表示为核函数中心点, γ 为分布参数, σ^2 是内核参数,表示径向基函数的方差。

$$K(x_i, x_j) = x_i x_j \quad (5)$$

$$K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2) \quad (6)$$

2 结果与讨论

2.1 纯面粉与添加剂样品的太赫兹光谱响应特性分析

太赫兹光谱中信息丰富,包含了吸收系数、折射率、介电常数、相位角等太赫兹光学参数,可多维度反映物质的内部信息。如图 1 为面粉、纯苯甲酸以及其混合物的太赫兹光

谱吸收系数谱图, 考虑到前端和后端存在较多的噪声干扰, 为便于后期数据处理, 截取光谱的频率范围为 1.0~3.0 THz 频段的光谱。纯面粉的光谱吸收系数谱图接近直线, 随着频率的增加, 面粉样品的吸收系数曲线缓慢上升。纯品的苯甲酸在 1.94 THz 有明显的峰值, 在 2.46 THz 处也有一个较弱的吸收峰。不同浓度苯甲酸下的太赫兹光谱吸收系数的吸收峰位吻合, 在 1.94 THz 的位置有较强的吸收峰, 并且可以观测到吸收强度随面粉中苯甲酸浓度的增加而增加。混合样品的原始太赫兹光谱波形整体接近一致, 但在一定的波段内吸收峰的强度有区别。由于面粉中苯甲酸浓度越高, 样品对太赫兹光谱的吸收也就越强烈。

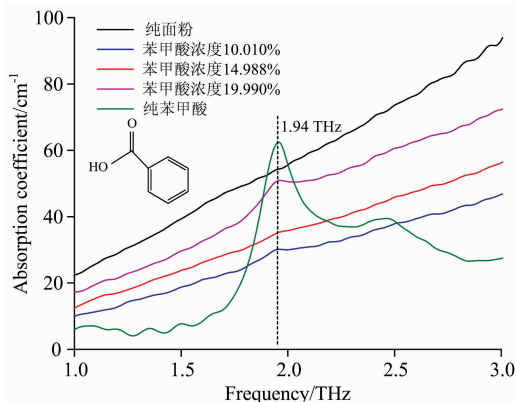


图 1 面粉、纯苯甲酸以及其混合物的太赫兹光谱吸收系数谱图

Fig. 1 Absorbance spectra of mixture, wheat flour and BA samples in the 1.0~3.0 THz frequency region

2.2 纯面粉和添加剂样品的太赫兹光谱相关分析

为找到面粉苯甲酸混合样品太赫兹吸收系数中表现出最大差异相关性的两个波段点, 在 1.0~3.0 THz 的区域计算所有可能的波段比组合。选择决定系数 (R^2) 最高的频率对作为最佳波段比。如图 2 所示为波段比相关系数的等值线图。观察到最高决定系数为 0.916, 其分别对应 1.946 和 1.869 THz。

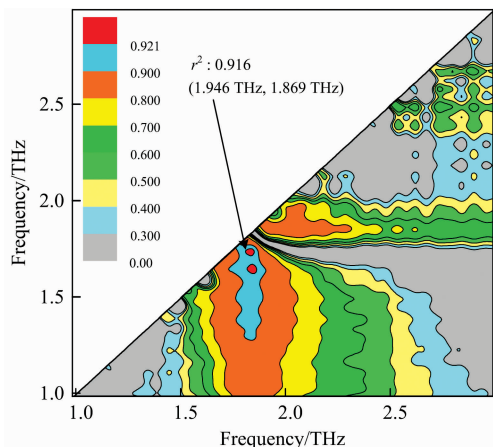


图 2 面粉苯甲酸混合物太赫兹光谱吸收系数在 1.0~3.0 THz 区域相关分析结果

Fig. 2 Results of correlation analysis in the regions of 1.0~3.0 THz

2.2 添加苯甲酸的面粉样品太赫兹光谱的校正处理

由于太赫兹光谱仪对环境的要求非常苛刻, 为减小在实验过程中实验仪器振动、噪声等原因导致的太赫兹光谱漂移、光散射等现象, 通过适当的预处理可消除部分干扰, 从而获得更好的建模效果。本文主要采用以下校正方法, 如平滑校正、多元散射校正、基线校正、归一化等方法进行校正处理, 通过偏最小二乘模型评估校正处理效果。

表 2 面粉的太赫兹光谱吸收系数校正处理 PLS 建模效果
Table 2 Result of THz PLS model by Correction processing

模型类型	预处理方法	主成分因子数	建模集		预测集	
			r_c	RMSEC	r_p	RMSEP
PLS	原始光谱	4	0.972	0.014	0.975	0.016
	平滑	4	0.972	0.014	0.975	0.016
	多元散射校正	3	0.976	0.013	0.976	0.014
	基线校正	5	0.980	0.012	0.970	0.014
	归一化	4	0.981	0.012	0.979	0.013

对添加苯甲酸面粉的太赫兹光谱进行不同的预处理之后建立 PLS 模型。参与建模的太赫兹光谱数据有 176 个, 将太赫兹光谱数据分成 44 个预测集和 132 个校正集。将建模集与预测集的相关系数和均方根误差进行比较, 可以评估校正处理效果的好坏。表 2 为添加苯甲酸面粉的太赫兹光谱不同预处理 PLS 建模结果, 经归一化校正处理的建模效果最佳, 归一化处理可以很好的校正由于微小光程差引起的太赫兹光谱的变化, PLS 模型预测集的相关系数为 0.979, 预测均方根误差为 1.30%。

3 混合面粉样品定量分析模型建立

3.1 添加苯甲酸面粉样品多元线性回归模型

当 R^2 值最大时, 面粉苯甲酸混合样品的太赫兹吸收系数在 1.94 和 1.86 THz 两个频率位置。利用多元线性回归算法 (MLR) 建立太赫兹光谱吸收系数与苯甲酸浓度之间的相关分析模型, 结果如表 3 所示, 以 1.94 和 1.86 THz 比值建立的 MLR 模型性能优于其他单点建立的 MLR 模型。MLR 模型的预测集相关系数为 0.955, 预测均方根误差为 1.9%。该模型只需要两个波段点即可建立模型, 模型简单, 数据计算量少, 但最佳模型的精度略低于 PLS 模型。

3.2 添加苯甲酸面粉样品太赫兹光谱 PLS 模型建立

偏最小二乘法 (PLS) 常用样品浓度真值与样品光谱矩阵之间的关系, 将矩阵分解及矩阵回归并为一。图 3 是添加苯甲酸面粉的 THz 吸收系数的回归系数, 回归系数较大的 1.94 THz 频率在 PLS 模型中起着重要作用。正相关系数与苯甲酸浓度响应呈正相关, 负相关系数与苯甲酸浓度响应呈负相关。面粉的太赫兹光谱吸收系数 PLS 建模效果, 其预测集相关系数为 0.979, 预测均方根误差为 1.30%。

3.3 添加苯甲酸面粉样品太赫兹光谱 LS-SVM 模型建立

最小二乘支持向量机^[14-16] (LS-SVM) 是基于统计学习理论而发展起来的一种机器学习方法, 其关键指标参数为输入向量、核函数种类及其相应的参数。径向基 RBF 核函数和线

表 3 面粉混合样品太赫兹光谱吸收系数 MLR 模型的建模结果

Table 3 Terahertz spectral absorption coefficient MLR modeling of BA and flour mixed samples

频率/THz	模型	建模集		预测集	
		R_c	RMSEC	R_p	RMSEP
1.869	$y=0.0029x_{1.869}+0.089$	0.018	0.061	0.155	0.063
1.946	$y=0.0507x_{1.946}+0.033$	0.204	0.060	0.239	0.061
1.946, 1.869	$y=0.8259x_{1.946}/x_{1.869}+0.089$	0.964	0.016	0.955	0.019

注： R_c 为建模集相关系数， R_p 为预测集相关系数，RMSEC 为建模集均方根误差，RMSEP 为预测集均方根误差。

性 Lin 核函数为 LS-SVM 的两种典型的核函数。表 4 为添加苯甲酸面粉的太赫兹光谱吸收系数建模效果，采用 RBF 核函数时，其参数组合为 $\gamma=16690$ ， $\sigma^2=229.418$ ，此时 LS-SVM 模型效果最佳，其预测集相关系数与预测集均方根误差分别为 0.987 和 1.1%。结果表明：RBF 核函数模型的效果总体优于 Lin 核函数。原因可能是 RBF 径向基核函数的泛化能力更强，并且可以逼近任意非线性函数。它能很好的处理面粉中苯甲酸含量与面粉样品苯甲酸太赫兹光谱数据的非线性关系。

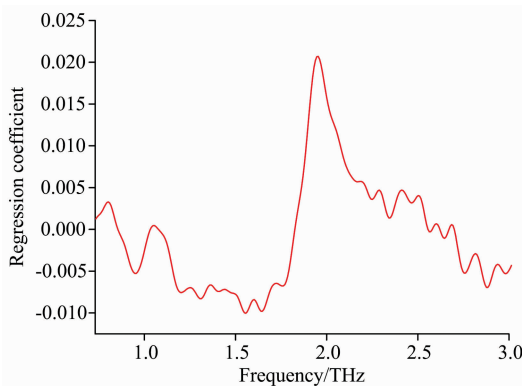


图 3 添加苯甲酸的面粉 THz 吸收系数回归系数
Fig. 3 The regression coefficient of PLS model

表 4 面粉混合样品太赫兹光谱吸收系数 LS-SVM 建模结果
Table 4 Terahertz spectral absorption coefficient LS-SVM modeling of BA and flour mixed samples

核函数类型	核函数参数	输入	R_p	RMSEP	t
Lin 核函数	$\gamma=12.670$	264	0.983	0.012	6.162
RBF 核函数	$\gamma=16690$ $\sigma^2=229.418$	264	0.987	0.011	4.368

3.4 混合面粉样品 THz 光谱 MLR, PLS 和 LS-SVM 模型对比

评估最佳 MLR, PLS 和 LS-SVM 模型的实际预测能力是利用预测集中的 44 个未知样本进行的。如图 4 为混合样品的苯甲酸浓度含量不同模型预测值与真值的拟合图。与

MLR 模型和 PLS 模型相比，LS-SVM 模型具有最高的预测精度，其预测相关系数 (R_p) 和预测均方根误差 (RMSEP) 分别为 0.987, 1.10%。结果表明：混合样品的苯甲酸浓度可以通过复杂的 LS-SVM 机器学习方法来确定。综上所述，LS-SVM 比 MLR, PLS 更适用于苯甲酸浓度的测定，因为 LS-SVM 模型具有较高的准确性，但是 MLR 用两个波段点也取得了较好的建模效果。

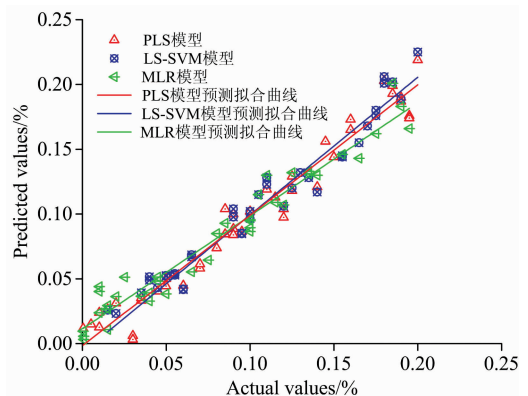


图 4 混合样品的苯甲酸浓度含量不同建模预测值结果
Fig. 4 Predicted results of MLR, PLS and LS-SVM models

4 结论

利用 THz 光谱结合 LS-SVM 的机器学习方法可以测定面粉中苯甲酸的浓度。与 PLS 模型和 MLR 模型相比，LS-SVM 模型可以获得更好的建模结果。LS-SVM 的预测相关系数 R_p 为 0.987，预测均方根误差 RMSEP 为 1.10%。苯甲酸在 1.94 THz 处呈现最大吸收峰。MLR 方法仅仅使用 1.946 和 1.869 THz 两个波段点进行建模，建模效果预测相关系数 R_p 为 0.955，预测均方根误差 RMSEP 为 1.90%。该实验成功的验证了运用太赫兹光谱检测技术对添加了苯甲酸的面粉检测的可行性，THz 技术与 LS-SVM 模型相结合，改变了传统方法检测面粉中苯甲酸添加剂时存在的费时费力、成本高昂等问题，具有较强的现实意义。

References

[1] Bilge G, Sezer B, Eseller K E, et al. Food Chemistry, 2016, 212: 183.
 [2] Huang M, Kim M S, Delwiche S R, et al. Journal of Food Engineering, 2016, 181: 10.
 [3] LI Bin, LONG Yuan, LIU Huan, et al(李斌, 龙园, 刘欢, 等). Transactions of the Chinese Society of Agricultural Engineering

- (农业工程学报), 2018, 34(2): 1.
- [4] Qin J, Xie L, Ying Y. Food Chemistry, 2015, 170: 415.
- [5] Mathanker S K, Weckler P R, Wang N. Transactions of the ASABE, 2013, 56(3): 1213.
- [6] Jiang L, Li M, Li C, et al. Journal of Infrared, Millimeter, and Terahertz Waves, 2014, 35(10): 871.
- [7] Lu S H, Li B Q, Zhai H L, et al. Food Chemistry, 2018, 246: 220.
- [8] Liu J. Optical and Quantum Electronics, 2017, 49: 1.
- [9] Lu S, Zhang X, Zhang Z, et al. Food Chemistry, 2016, 211: 494.
- [10] Ge H, Jiang Y, Lian F, et al. Food Chemistry, 2016, 209: 286.
- [11] Sousa S I V, Martins F G, Alvim-Ferraz M C M, et al. Environmental Modelling & Software, 2007, 22(1): 97.
- [12] Cheng J, Sun D. LWT-Food Science and Technology, 2015, 63(2): 892.
- [13] Ali F, Rasoolimanesh S M, Sarstedt M, et al. International Journal of Contemporary Hospitality Management, 2018, 30(1): 514.
- [14] Jiang J, Tan Q, Li W, et al. Transactions of Tianjin University, 2017, 23(3): 237.
- [15] Baek S H, Lim H B, Chun H S. Journal of Agricultural and Food Chemistry, 2014, 62(24): 5403.
- [16] Huang M, Kim M S, Delwiche S R, et al. Journal of Food Engineering, 2016, 181: 10.
- [17] Dorney T D, Baraniuk R G, Mittleman D M. J. Opt. Soc. Am. A, 2001, 18: 1562.
- [18] Duvillaret L, Garet F, Coutaz J L. Appl. Opt., 1999, 38: 409.

Quantitative Determination of Benzoic Acid in Flour Based on Terahertz Time-Domain Spectroscopy and LS-SVM

HU Jun, LIU Yan-de*, SUN Xu-dong, OUYANG Ai-guo, CAI Hui-zhou, LIU Hong-liang

School of Mechatronics Engineering, East China Jiaotong University, Nanchang 330013, China

Abstract With the further development of terahertz technology, terahertz has shown its unique advantages in food safety detection. Flour (wheat flour) is the staple food in most areas of northern China. Besides, benzoic acid(BA), as the important preservative of acid food, is often added to extend the preservation time of food. However, the excessive use of food additives would cause serious damage to human health. This paper explores the feasibility of detecting food additives through terahertz technology and conducts empirical study on benzoic acid in flour by terahertz time-domain spectroscopy (THz-TDS) technology. The terahertz time-domain and frequency domain spectrum of the mixed samples (flour and benzoic acid) were obtained. As shown by absorption coefficients, benzoic acid presented obvious absorption peak at 1.94 THz. Meanwhile, the absorption coefficient of flour increased at a certain slope, which indicated that the characteristic identification of benzoic acid in flour could be carried out by terahertz technology. In order to establish the quantitative detection model of benzoic acid additive in flour, terahertz time-domain spectra of benzoic acid doped with different percentages (mass fraction) in flour were collected, and the absorption coefficient spectrum was obtained through calculation. It was found that the absorption peak amplitude enjoys positive correlation with benzoic acid content. As for the detection method, firstly, explore the effects of different spectral pretreatment methods on THz spectroscopy, and then adopt methods like Smoothing, Multiple Scatter Correction (MSC), Baseline and Normalization to carry out correct processing. After correction, PLS model was established to select the optimal pretreatment method. Secondly, establish PLS and LS-SVM regression models for the determination of benzoic acid content in flour. The experimental results verify that PLS model established after normalization was more optimal, with correlation coefficient of prediction (r_p) of 0.979 and root mean square error of prediction (RMSEP) of 1.30%. By comparison, it was proved that the most optimal quantitative determination model of benzoic acid content in flour is LS-SVM model with correlation coefficient of prediction (r_p) of 0.987 and root mean square error of prediction (RMSEP) of 1.10% after the normalization of terahertz absorption coefficient. MLR model was established by only two bands of 1.946 and 1.869 THz with correlation coefficient of prediction (r_p) of 0.955 and root mean square error of prediction (RMSEP) of 1.90%. It is concluded that a new solution for the nondestructive detection of benzoic acid additives in flour was developed, and method guidance was provided for the detection of other types of additives, all of which have an important significance for the healthy development of flour industry.

Keywords Food additive; Terahertz spectroscopy; LS-SVM; Benzoic acid; wheat flour