

# 基于太赫兹时域光谱的食用油过氧化值定量分析研究

刘翠玲<sup>1</sup>, 杨雨菲<sup>1\*</sup>, 田芳<sup>2</sup>, 吴静珠<sup>1</sup>, 孙晓荣<sup>1</sup>

1. 北京工商大学食品安全大数据技术北京市重点实验室, 北京 100048  
2. 北京工商大学外国语学院, 北京 100048

**摘要** 针对目前太赫兹光谱技术在食用油品质检测方面存在定性多、定量难的问题, 提出一种基于衰减全反射(ATR)式太赫兹时域光谱(THz-TDS)技术的食用油过氧化值定量分析方法。首先采集不同种类、不同氧化程度食用油样本的太赫兹时域光谱图, 筛选有效信号波段并提取得到光学常数, 经预处理算法校正后的光学常数, 结合多种化学计量学方法建立定量分析模型, 实现快速、无损预测食用油的过氧化值。70个实验样本包括大豆油、菜籽油和玉米, 过氧化值覆盖范围 0.41~10.23 mmol·kg<sup>-1</sup>, 且样本的过氧化值均匀分布。采用 TeraView 公司生产配有 ATR 检测模块的 TeraPulse 4000 太赫兹脉冲光谱系统采集样本 THz-TDS 信号, 根据 THz-TDS 谱图信号特征筛选有效波段 10~86.78 cm<sup>-1</sup>用于建模分析。通过快速傅里叶变换得到频域信号并从中提取光学常数: 折射率和吸收系数, 采用 Savitzky-Golay 7 点卷积平滑分别对折射率和吸收系数进行预处理, 去除干扰信号。运用 SPXY 算法按照 3:1 比例划分校正集和预测集样本, 结合主成分回归法、偏最小二乘法两种化学计量学分析方法, 分别建立基于折射率和基于吸收系数的过氧化值分析模型。对模型评价指标均方根误差和相关系数进行分析, 基于折射率的过氧化值分析模型采用偏最小二乘法建模预测精度最理想, 选取最优主成分数为 6 时, 其校正集均方根误差 RMSEC 为 0.168%、决定系数 R<sup>2</sup> 为 0.981, 预测集均方根误差 RMSEP 为 0.231%、决定系数 r<sup>2</sup> 为 0.977; 基于吸收系数的过氧化值分析模型则采用主成分回归算法建模预测模型稳健度最好, 选取最优主成分数为 10 时, 其校正均方根误差 RMSEC 为 0.192%、校正集决定系数 R<sup>2</sup> 为 0.979, 预测均方根误差 RMSEP 为 0.262%、预测集决定系数 r<sup>2</sup> 为 0.97。该研究结果的得出, 验证了 THz-TDS 技术用于食用油过氧化值定量分析的可行性, 为食用油的品质评价找到一种高精度、性能稳定、快速无损的检测方法。

**关键词** 食用油; 太赫兹时域光谱技术; 主成分回归; 偏最小二乘; 定量分析

**中图分类号:** O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)05-1387-06

## 引言

食用油作为人体的三大营养源之一, 其品质的优劣直接影响到消费者身体健康。而油脂在加工和贮存过程中十分容易受到光照、温度、氧气等环境条件影响, 而产生氢过氧化物, 该物质进一步受环境影响氧化分解成小分子的醛、酮、酸(即酸败现象), 致使过氧化值发生改变<sup>[1-2]</sup>。

近年来, 太赫兹时域光谱(Terahertz time domain spectroscopy, THz-TDS)技术因其具有辐射波波长、穿透性好、可反映分子集体振动模式和分子间相互作用等方面的特点而得到了迅速的发展<sup>[3]</sup>。文献[4]通过采用 THz-TDS 技术采集

未加热与加热后的花生油光谱特性, 由于加热后的花生油氧化分解程度更显著, 通过比较两者谱图信息, 发现谱图吸收峰在 1~1.6 THz 之间两者差别最显著, 加热后表现为吸收性更强, 该研究结果表明 THz 光谱技术可准确反映食用油组分的变化, 并论证了食用油的氧化情况可通过 THz 光谱技术得到。文献[5]通过采集得到黑芝麻油、芝麻油、小磨香油和花生油四种食用油的 THz-TDS, 选择波段在 0.2~1.6 THz 之间的延迟特性和折射率特性作为建模输入, 将其结合多种模式识别方法鉴别出食用油的种类。

文献[6]应用衰减全反射式 THz-TDS 技术, 发现由于食用油中脂肪酸分子的低频振动和转动, 从而影响光谱图的吸收峰, 表明食用油的吸收峰与所含脂肪酸分子种类与含量有关。

收稿日期: 2020-05-08, 修订日期: 2020-08-06

基金项目: 国家自然科学基金项目(61807001), 北京市自然科学基金项目(4182017)资助

作者简介: 刘翠玲, 女, 1963年生, 北京工商大学食品安全大数据技术北京市重点实验室教授 e-mail: liuel@btbu.edu.cn

\* 通讯作者 e-mail: yangyf0919@163.com

虽然已有少数研究者在应用 THz-TDS 探索油脂品质和油料作物的检测方面取得了一些进展,但多用于食用油脂的分子结构和组分的定性分析,鲜有关注某一指标的定量研究,因此定性多、定量难的现象使得 THz 技术的发展还有很大的探索空间。

本文旨在实现衰减全反射式 THz-TDS 光谱技术用于食用油脂定量分析研究。将衰减全反射式 THz-TDS 光谱技术与化学计量学方法相结合,建立定量分析模型,以期验证 THz-TDS 光谱技术检测食用油过氧化值的可行性,并为食用油脂品质评价找到新方法。

## 1 实验部分

### 1.1 材料与仪器

由北京粮食科学研究院提供不同生产日期、不同批次的大豆油、菜籽油和玉米油,共 70 个,过氧化值覆盖范围  $0.41 \sim 8.547 \text{ mmol} \cdot \text{kg}^{-1}$ 。其中,23 个样本的过氧化值小于  $1.2 \text{ mmol} \cdot \text{kg}^{-1}$ ;21 个样本的过氧化值在  $1.2 \sim 2.4 \text{ mmol} \cdot \text{kg}^{-1}$ ;26 个样本的过氧化值大于  $2.4 \text{ mmol} \cdot \text{kg}^{-1}$ 。

实验设备来自英国剑桥 TeraView 公司生产的型号为 TeraPulse 4000 的太赫兹脉冲光谱仪。工作时仪器参数设置为:Terahertz 光源:激光门控光电导体半导体发射器,激光器:超短脉冲光纤激光器;光谱测量范围在  $2 \sim 133 \text{ cm}^{-1}$  之间,信噪比最高达到 70 dB;扫描范围:2 400 ps;快速扫描:

$30 \text{ 次} \cdot \text{s}^{-1}$ ,  $1.2 \text{ cm}^{-1}$  光谱分辨率。

光谱采集由太赫兹脉冲光谱仪配备的衰减全反射(attenuated total reflection, ATR)模块完成。ATR 能够测量固体和液体样本,具备采样面积小、样本量小等优点,并且样本制备和采集方式简单。它的工作范围在  $10 \sim 120 \text{ cm}^{-1}$  左右的电磁频谱区域,入射角  $35^\circ$ 。

### 1.2 过氧化值的测定

依照 2016 年颁布的 GB5009.227—2016《食品安全国家标准 食品中过氧化值的测定》中规定的滴定法测定食用油的过氧化值<sup>[7]</sup>。

### 1.3 光谱采集

由于太赫兹波易受温度和湿度变化的影响,本实验环境为温度和光线恒定的超净间实验室,恒定温度为  $22^\circ\text{C}$ 、恒定湿度 25%。实验过程中还需要不间断地使用空气压缩机对 TeraPulse 4000 光谱仪腔体进行吹扫,以去除空气中水蒸气对实验数据的影响。

在逐个测定样本时,均要保证在 ATR 晶体干净无污染的条件下完成。检测样本前,需检测实验环境的太赫兹信号,作为参考信息。检测样本的过程中,采用移液枪逐一取样(取样量统一为  $600 \mu\text{L}$ ),每采集完一份样本均需用酒精反复擦拭 ATR 晶体,并确保无任何液体残留后,再进行下一次样本采集工作。图 1(a)所示为 3 个不同氧化程度食用油脂样本的太赫兹时域光谱图,图中横坐标代表时延,单位 ps,纵坐标代表太赫兹信号强度。图 1(b)为局部放大图。

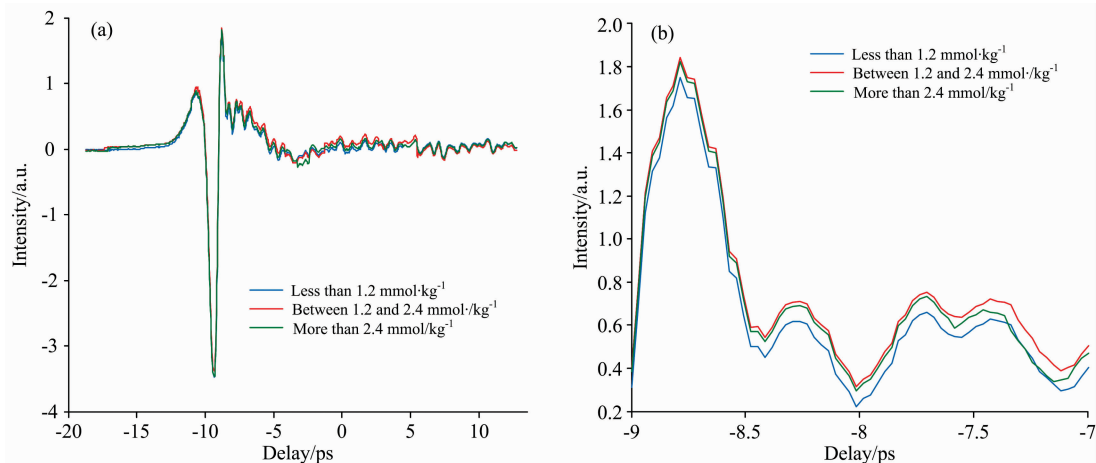


图 1 不同氧化程度食用油脂样本的太赫兹时域光谱图

(a): 全局图; (b): 局部放大图

Fig. 1 THz-TDS spectra of edible oil samples with different oxidation degrees

(a): Global picture; (b): Partial enlarged picture

如图 1 所示,不同氧化程度食用油脂的脉冲幅度在相同波数下存在差异,分析原因:虽然食用油脂内部所含化学元素种类和数量是相同的,但由于不同程度的氧化反应已引起样本组分微观化学结构发生变化。

### 1.4 分析模型评价指标

评价光谱模型的优劣,需要判断的指标有校正集决定系数  $R^2$ 、预测集决定系数  $r^2$  (决定系数取值范围  $0 \sim 1$ ),校正均方根误差(root mean square error of calibration, RMSEC),

预测均方根误差(root mean square error of prediction, RMSEP),各指标计算公式如式(1)~式(3)所示。通常认为一个理想的分析模型,其决定系数越大且接近 1,回归性越好;同时还需满足, RMSEC 和 RMSEP 的值越小且 RMSEC 值小于 RMSEP,建模结果和模型实用性越理想<sup>[8]</sup>。

$$\text{RMSEC} = \sqrt{\frac{\sum_{i=1}^n (y_{i,a} - y_{i,p})^2}{n-1}} \quad (1)$$

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^m (y_{i,a} - y_{i,p})^2}{m-1}} \quad (2)$$

$$R^2 (r^2) = 1 - \frac{\sum_{i=1}^N (y_{i,a} - y_{i,p})^2}{\sum_{i=1}^N (y_{i,a} - \bar{y}_a)^2} \quad (3)$$

式中,  $y_{i,a}$  为第  $i$  个样本的过氧化值真实值;  $y_{i,p}$  为第  $i$  个样本的模型计算值;  $\bar{y}_a$  为过氧化值真实值的均值;  $n$  为校正集的样本数;  $m$  为预测集的样本数; 当求  $R$  时  $N=n$ 、求  $r$  时  $N=m$ 。

## 2 结果与讨论

### 2.1 光谱预处理

#### 2.1.1 光学常数的提取

光学常数是表征物质宏观光学性质的物理量。光学常数

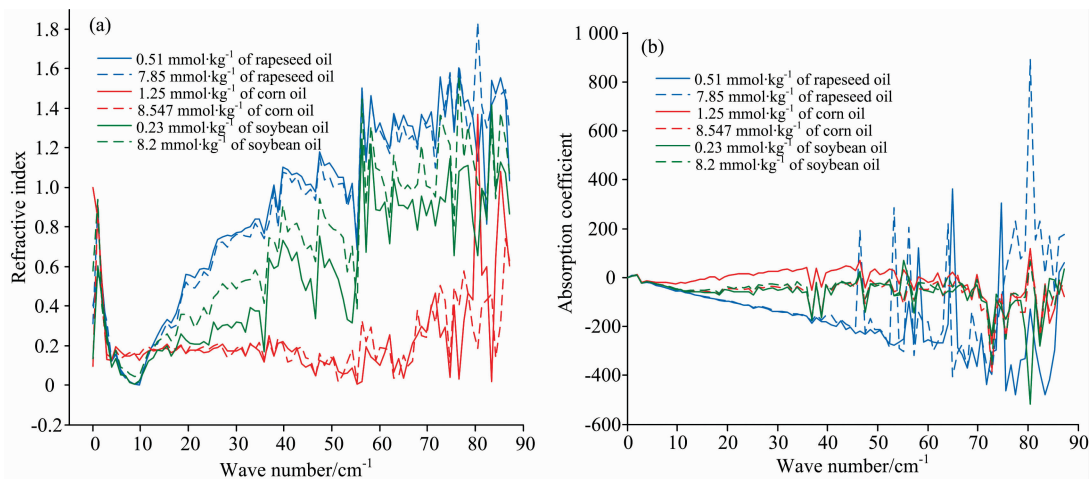


图 2 不同氧化程度菜籽油、玉米油和大豆油的光学常数谱图

(a): 折射率谱图; (b): 吸收系数谱图

Fig. 2 Optical constant spectrograms of rapeseed oil, corn oil and soybean oil with different oxidation degree

(a): Refractive index spectrums; (b): Absorption coefficient spectrums

图 2(a)中, 不同种类的食用油, 在  $10 \sim 55 \text{ cm}^{-1}$  波数范围分别分布于不同的折射率区间, 根据折射率谱分布情况可判别食用油的种类, 因此食用油的营养成分和官能团对折射率谱的分布影响较大。图 2(b)中, 菜籽油和玉米油吸收系数谱波动情况相似, 而大豆油吸收系数谱波动较大, 明显区别于其他两类食用油。但两类谱图的波动情况对过氧化值研究无明显规律, 需进一步结合化学计量学方法进行分析。

#### 2.1.2 光谱噪声的处理

由于原始光学常数谱数据含有来自高频随机噪声、基线漂移、光散射等不利因素的影响, 因此在建模时需对其进行预处理, 从而达到去除信号干扰、校正样本有效特征信息的目的。Savitzky-Golay 卷积平滑法通过移动窗口内的数据进行多项式最小二乘拟合, 可降低随机噪声、改善模型的鲁棒性<sup>[10]</sup>。针对光学常数谱图基线不稳、噪声干扰强的特点, 采用 Savitzky-Golay 7 点卷积平滑对谱图数据进行预处理。

的提取, 首先利用快速傅里叶变换对参考信息和样本的时域光谱进行转换, 得到对应的频域信号; 进而利用频域信号中的幅值和相位值, 计算得到光学常数, 即折射率和吸收系数。折射率定义为光在真空中的传播速度与传播速度之比, 也可通过不同样本在时域信号上的延迟反映; 吸收系数定义为该样本单位厚度的吸光度。折射率  $n(\omega)$  和吸收系数  $\alpha(\omega)$  的计算公式如式(4)和式(5)所示。

$$n(\omega) = 1 + \frac{\varphi(\omega)c}{\omega d} \quad (4)$$

$$\alpha(\omega) = \frac{2}{d} \ln \left\{ \frac{4n(\omega)}{\rho(\omega)[n(\omega) + 1]^2} \right\} \quad (5)$$

式中,  $\omega$  为角频率,  $d$  为样本厚度,  $c$  为太赫兹波在空气中的传播系数,  $\varphi(\omega)$  和  $\rho(\omega)$  由样本的频域信号与参考频域信号的比值函数所得,  $\varphi(\omega)$  为比值函数的相位,  $\rho(\omega)$  为比值函数的幅值<sup>[9]</sup>。

计算得到不同氧化程度菜籽油、玉米油和大豆油的光学常数谱图, 如图 2 所示。

### 2.2 数据建模

#### 2.2.1 建模分析方法

偏最小二乘(partial least square, PLS)法能够在自变量存在严重多重相关性的条件下进行回归建模, 并且可高效辨识系统信息与噪声(甚至一些非随机性的噪声), PLS 模型的建立同时兼顾样本光谱矩阵  $\mathbf{X}$  和过氧化值真实值矩阵  $\mathbf{Y}$ , 使得  $\mathbf{X}$  的主成分与  $\mathbf{Y}$  的主成分直接关联起来<sup>[11-12]</sup>。

主成分回归(principle component regression, PCR)法是一种用于解决样本变量间共线性问题的分析方法。该算法既可用于剔除干扰噪声, 对有效光谱信息进行挑选, 又可分离多重共线性的重叠信息, 提高模型的预测能力和准确性。

#### 2.2.2 基于折射率的过氧化值定量分析

采用 SPXY 法以 3 : 1 的比例划分样本集, 校正集 48 个, 预测集 16 个, 剔除异常样本 6 个。主成分数均取 6。建立基于折射率的食用油过氧化值分析模型, 预测结果如图 3 所示。

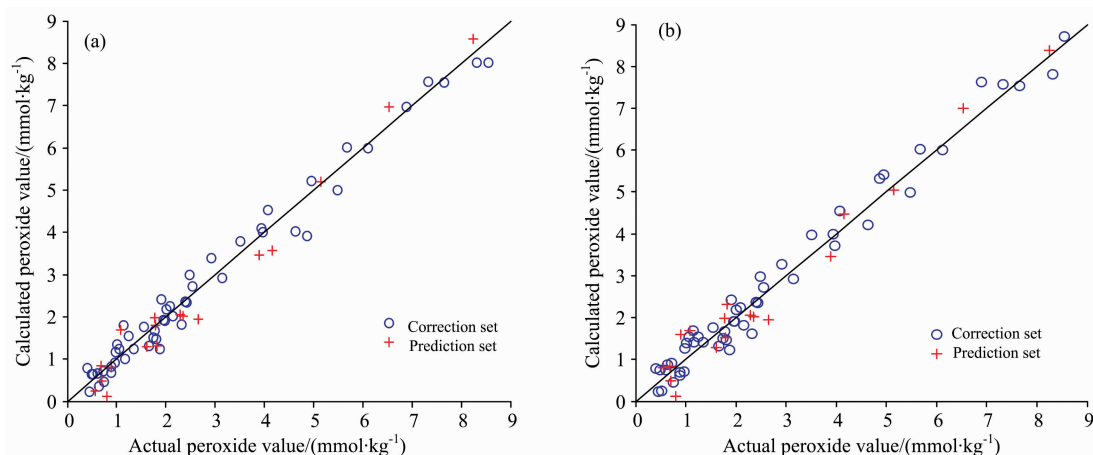


图 3 基于折射率的食用油过氧化值分析模型预测结果图

(a): PLS法; (b): PCR法

Fig. 3 Prediction results of edible oil peroxide value analysis model based on refractive index

(a): PLS method; (b): PCR method

PLS 过氧化值分析模型, RMSEC 为 0.168%,  $R^2$  为 0.981, RMSEP 为 0.231%,  $r^2$  为 0.977; PCR 过氧化值分析模型, RMSEC 为 0.329%,  $R^2$  为 0.922, RMSEP 为 0.375%,  $r^2$  为 0.892。

### 2.2.3 基于吸收系数的过氧化值定量分析

采用 SPXY 法以 3 : 1 的比例划分样本集, 校正集 48 个, 预测集 16 个, 剔除异常样本 6 个。主成分数均取 10。建立基于吸收系数的食用油过氧化值分析模型, 预测结果如图 4 所示。

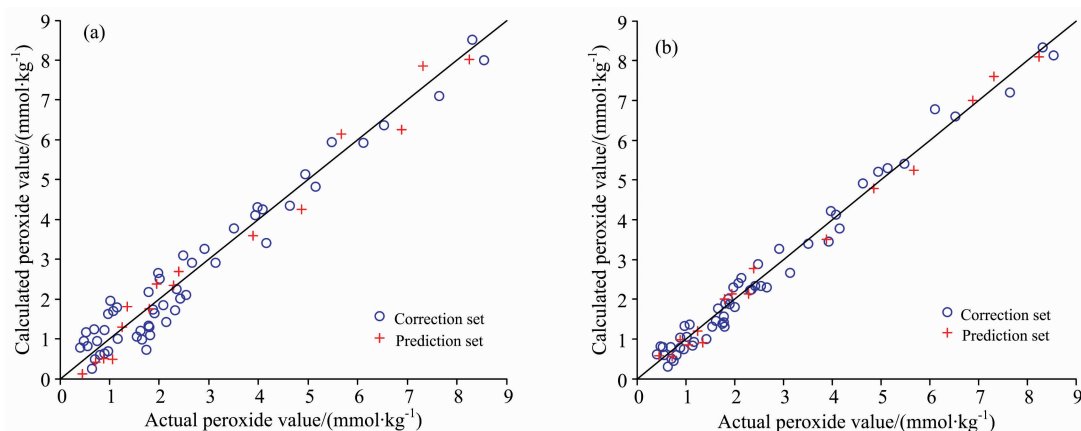


图 4 基于吸收系数的食用油过氧化值分析模型预测结果图

(a): PLS法; (b): PCR法

Fig. 4 Prediction results of edible oil peroxide value analysis model based on absorption coefficient

(a): PLS method; (b): PCR method

PLS 过氧化值分析模型, RMSEC 为 0.317%,  $R^2$  为 0.927, RMSEP 为 0.333%,  $r^2$  为 0.912; PCR 过氧化值分析模型, RMSEC 为 0.192%,  $R^2$  为 0.979, RMSEP 为 0.262%,  $r^2$  为 0.97。

### 2.2.4 建模结果对比与分析

食用油过氧化值分析模型建模结果, 如表 1 所示。

对比表 1 中模型性能评价指标可知, 对于分析食用油 THz 折射率时, 采用 SG 卷积 7 点平滑预处理算法结合 PLS 算法建立的过氧化值分析模型性能稳健、预测精度高, 建模集与预测集决定系数均大于 0.97, 模型回归性好。而分析食

表 1 食用油过氧化值定量分析结果表

Table 1 Quantitative analytical results of peroxide value						
建模数据	建模方法	主成分数	RMSEC / %	$R^2$	RMSEP / %	$r^2$
折射率	PLS	6	0.168	0.981	0.231	0.977
	PCR	6	0.329	0.922	0.375	0.892
吸收系数	PLS	10	0.317	0.927	0.333	0.912
	PCR	10	0.192	0.979	0.262	0.97

用油 THz 吸收系数时, 由于样本间的吸收系数共线性显著, 通过 SG 卷积 7 点平滑预处理原始吸收系数数据后, 采用

PCR 算法较 PLS 算法更为合适。

### 3 结 论

选用不同种类、不同生产批次的大豆油、菜籽油和玉米油样本共 70 个, 采集其衰减全反射式 THz-TDS 光谱图, 通过快速傅里叶变换提取光学常数折射率和吸收系数。经 Savitzky-Golay 7 点卷积平滑预处理后, 分别建立 PCR 模型和 PLS 模型。基于折射率的 PLS 模型预测精度高、模型稳健

性良好, 选择 6 维主成分, 其 RMSEC 为 0.168%、 $R^2$  为 0.981, RMSEP 为 0.231%、 $r^2$  为 0.977; 基于吸收系数的 PCR 模型预测误差小, 选择 10 维主成分, RMSEC 为 0.192%、 $R^2$  为 0.979, RMSEP 为 0.262%、 $r^2$  为 0.97。该研究结果的得出论证了 THz-TDS 应用于食用油过氧化值分析的可行性, 并通过对比模型性能评价指标寻找出了切实可行的食用油过氧化值分析方法, 实现了食用油品质的精确、快速无损检测。

### References

- [ 1 ] Gong Weichao, Shi Ruyi, Chen Min, et al. Journal of Food Measurement and Characterization, 2019, 13(3): 2203.
- [ 2 ] JI Min, LIU Zhong-yi, ZHANG Jian, et al(季 敏, 刘忠义, 张 剑, 等). China Oils and Fats(中国油脂), 2018, 43(5): 92.
- [ 3 ] Liang Jie, Guo Qijia, Chang Tianying, et al. Optik, 2018, 174: 7.
- [ 4 ] Benjamin S Y Ung, Brian W H Ng, Derek Abbott. 35th International Conference on Infrared, Millimeter, and Terahertz Waves, IEEE, 2010.
- [ 5 ] LIAN Fei-yu, FU Mai-xia, GE Hong-yi, et al(廉飞宇, 付麦霞, 葛宏义, 等). China Oils and Fats(中国油脂), 2017, 42(7): 69.
- [ 6 ] NIE Mei-tong, XU De-gang, WANG Yu-ye, et al(聂美彤, 徐德刚, 王与焯, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2018, 38(7): 2016.
- [ 7 ] GB 5009.227—2016 National Standard of the People's Republic of China(中华人民共和国国家标准). National Food Safety Standard-Determination of Peroxide Value in Foods(食品安全国家标准 食品中过氧化值的测定).
- [ 8 ] CHU Xiao-li(褚小立). Molecular Spectroscopy Analytical Technology Combined with Chemometrics and Its Applications(化学计量学方法与分子光谱分析技术). Beijing: Chemical Industry Press(北京: 化学工业出版社), 2011. 124.
- [ 9 ] Nakajima Shusaku, Shiraga Keiichiro, Suzuki Tetsuhito, et al. Food Chemistry, 2019, 294(5): 203.
- [10] Hu Keqing, Huyan Zongyao, Sherazi Syed Tufail Hussain, et al. Journal of Oleo Science, 2019, 68(11): 1073.
- [11] LIU Jun-xiu, DU Bin, DENG Yu-qiang, et al(刘俊秀, 杜 彬, 邓玉强, 等). Chinese Journal of Lasers(中国激光), 2019, 46(6): 614039.
- [12] Zhan Honglei, Xi Jianfeng, Zhao Kun, et al. Food Control, 2016, 67: 114.

## Study on Quantitative Analysis of Edible Oil Peroxide Value by Terahertz Time Domain Spectroscopy

LIU Cui-ling<sup>1</sup>, YANG Yu-fei<sup>1\*</sup>, TIAN Fang<sup>2</sup>, WU Jing-zhu<sup>1</sup>, SUN Xiao-rong<sup>1</sup>

1. Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China
2. School of Foreign Languages, Beijing Technology and Business University, Beijing 100048, China

**Abstract** At present, Terahertz (THz) spectroscopy techniques are mainly used for qualitative analysis, but the application of THz technology can hardly be found in quantitative analysis in the detection of edible oil's quality. This paper presents an approach to analyze edible oil quality based on Attenuated Total Reflection (ATR) and Terahertz Time Domain Spectroscopy (THz-TDS). Firstly, the THz-TDS of edible oil samples with different types and degrees of oxidation were collected, the effective signal band was filtered and the optical constants were extracted, the preprocessing algorithm corrected the optical constants, a variety of chemometrics methods were used to establish quantitative analysis models, in order to quickly and accurately predicted the peroxide value of edible oils. 70 experimental samples were used, including soybean oils, rapeseed oils and corn oils, the peroxide value ranged from 0.41 to 10.23 mmol · kg<sup>-1</sup>, and the peroxide value distribution of the samples was evenly distributed. A TeraPulse 4000 terahertz pulse spectroscopy system equipped with an ATR detection module belonging to TeraView was used to collect samples' THz-TDS signals. According to THz-TDS characteristics, the effective band 10 to 86.78 cm<sup>-1</sup> was selected for modeling analysis. The frequency domain signals were obtained by fast Fourier transform, and the optical constants were extracted; refractive index and absorption coefficient. Refractive index and absorption coefficient were

preprocessed separately through Savitzky-Golay 7-points convolution smoothing, which had achieved the purpose of removing interference signals. The SPXY algorithm was used to divide the calibration set, and prediction set samples in a 3 : 1 ratio. The peroxide value analysis models based on refractive index and absorption coefficient were established by the principal component regression algorithm and partial least square algorithm. The root mean square error and correlation coefficient of the model evaluation indexes were analyzed, the peroxide value analysis model based on the refractive index was modeled by partial least squares algorithm had ideal prediction accuracy. When the optimal principal component number was selected to be 6, RMSEC is 0.168%,  $R^2$  is 0.981, RMSEP is 0.231%,  $r^2$  is 0.977. The principal component regression algorithm modeled the peroxide value analysis model based on the absorption coefficient, and the prediction model had the best robustness. When the optimal principal component number was selected to be 10, RMSEC is 0.192%,  $R^2$  is 0.979, RMSEP is 0.262%,  $r^2$  is 0.97. This study verifies it is feasible to detect the peroxide value of edible oil by THz technology, and the more important innovation is a high-precision, stable performance, fast and non-destructive detection method for the evaluation of edible oil quality has been found. Furthermore, this research has important guiding significance for improving the safety of edible oil quality and building edible risk assessment systems.

**Keywords** Edible oil; Terahertz time domain spectroscopy technology; Principal component regression; Partial least squares; Quantitative analysis

(Received May 8, 2020; accepted Aug. 6, 2020)

\* Corresponding author