

## 基于近红外光谱两种植物油过氧化值通用模型研究

彭丹, 李林青, 刘亚丽, 毕艳兰\*, 杨国龙

河南工业大学粮油食品学院, 河南 郑州 450001

**摘要** 过氧化值的快速、准确检测对食用油脂的品质及其食品安全控制具有重要意义。近红外光谱技术是一种理想的过氧化值测量手段,但校正模型的建立需要耗费大量的资源。旨在通过近红外光谱信息与油脂过氧化物间的关系分析,探索对不同种类、不同等级植物油建立同一校正模型的可行性,以不同等级的大豆油和菜籽油为研究对象,结合二维相关光谱技术对两种植物油的近红外光谱进行分析,通过间隔偏最小二乘法选择过氧化值通用模型的最佳检测波段,考察了正交信号校正(OSC)、标准正态变量变换(SNV)和二阶导数(SD)对两种植物油过氧化值校正模型的影响,比较了主成分回归(PCR)、偏最小二乘法(PLS)和支持向量机回归(SVR)三种建模方法的预测效果,构建了大豆油(一级+三级)、菜籽油(一级+三级+四级)、一级油(大豆油+菜籽油)、三级油(大豆油+菜籽油)四种通用模型。结果显示:(1)近红外光谱能够检测植物油过氧化值的变化情况,对应的光谱信息主要分布于1700~2200 nm区域;(2)通用模型最佳的波段、预处理方法和建模方法分别为1700~2200 nm、SD法和PLS法;(3)四种通用模型中一级植物油(大豆油和菜籽油)的过氧化值通用模型具有较好预测结果,其预测均方根误差(RMSEP)、决定系数( $R^2$ )分别为0.412和0.920,与一级的大豆油和菜籽油单一模型相比,预测精度相差不大。研究表明生产工艺过程相差不大的一级植物油间有可能建立准确性高的通用模型。此外,为了扩展通用模型的性能,需要不断用新产品对模型进行及时更新。

**关键词** 近红外光谱;通用模型;化学计量学;植物油;过氧化值

**中图分类号:** O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)06-1828-05

### 引言

食用植物油是膳食中重要组成部分,为人体提供能量和必需脂肪酸,也是脂溶性维生素的重要载体。然而,油脂在加工、储运过程中易发生氧化酸败,使其风味、色泽等劣变,造成品质下降,产生的氧化产物会危害人体健康<sup>[1-2]</sup>,如促进人体衰老、肿瘤和心脑血管疾病等的发生。因此,对油脂的品质评价和氧化程度的监测十分重要。

过氧化值是衡量油脂品质和新鲜度的重要指标之一,也是评价油脂氧化程度的一个安全指标。目前,过氧化值的测量多采用国标中的滴定法(碘量法),具有检测精度高、技术成熟等优点,但费时费力,消耗大量有机溶剂,结果易受外界因素的影响,无法对油脂进行实时检测。近年来,有学者运用中红外光谱法<sup>[3]</sup>、紫外法<sup>[4]</sup>、电化学方法<sup>[5]</sup>等对油脂过氧化值进行测量,然而这些方法还不够成熟。近红外光谱技

术是一种快速、无损、可在线分析的绿色技术,已广泛应用于粮油、食品等领域,国内外也有学者采用近红外光谱对油脂氧化程度进行分析。Krzysztof等<sup>[6]</sup>采用近红外光谱结合多元数据分析分别建立橄榄油、菜籽油和葵花籽油的过氧化值模型,其相关系数均在0.97以上。Mureşan等<sup>[7]</sup>应用近红外光谱检测向日葵果仁酱中油脂的过氧化值。蒋琦等<sup>[8]</sup>建立了山茶籽油过氧化值定量模型,其预测结果小于化学分析值所允许的误差范围。这些研究表明近红外光谱可以实现油脂及含油食品氧化情况的快速检测,但大多研究仅对单一种类和等级的油脂进行分析。近红外光谱作为一种间接分析技术,必须通过校正模型来实现定性或定量分析,构建一个稳定可靠的校正模型是一项复杂的工作,需要耗费大量时间和人力,若能建立通用性模型,则能节约资源,扩大其在油脂加工、储运中的应用。现有的近红外通用模型主要针对药品的研究,也有学者尝试构建三种木本油料植物种子含油率的近红外光谱通用模型<sup>[9]</sup>,但在油脂品质的分析中却未见相关

收稿日期:2019-06-09,修订日期:2019-10-20

基金项目:国家自然科学基金项目(31601537)资助

作者简介:彭丹,1979年生,河南工业大学粮油食品学院副教授

\* 通讯联系人 e-mail: bylzy@126.com

e-mail: pengdantju@163.com

报道。本研究以大豆油(一级、三级)和菜籽油(一级、三级和四级)为研究对象,考察通用模型的最佳检测条件,研究不同植物油以及不同等级同种植物油通用模型的可行性,以期食用油脂品质的快速检测及其食品安全控制提供技术支持。

## 1 实验部分

### 1.1 材料

不同等级的大豆油(15种)、菜籽油(16种)分别购于生产厂家及超市。为了使油脂样品更具有代表性,采用烘箱加速氧化的方法,称取一定量的油样放置于60℃的烘箱中,共获得432个样本。采用Kennard-Stone算法选择2/3的样本为建模集,其余的1/3样本为预测集。油脂的过氧化值按照GB 5009.227—2016中的滴定法测定。其中,大豆油、菜籽油的过氧化值范围分别为1.00~13.44和0.58~13.63 mmol·kg<sup>-1</sup>。

### 1.2 光谱采集

采用FOSS公司的XDS型近红外光谱仪,在780~2 500 nm波长范围内对样品的吸收光谱进行测试,检测器为硅和硫化铅,光源为卤钨灯,操作温度25℃,扫描次数32次,分辨率2 nm,每个样品重复测定3次,取平均光谱。

### 1.3 数据处理

采用二维相关光谱寻找与过氧化物相关的敏感波段。以过氧化值浓度为外部微扰,进行二维相关同步谱和自相关谱分析,具体过程参考文献[10],利用Matlab R2018a软件计算。

采用间隔偏最小二乘法进行波段选择,运用正交信号校正(OSC)、标准正态变量变换(SNV)、二阶导数(SD)对原始光谱进行预处理,通过主成分回归(PCR)、偏最小二乘法(PLS)和支持向量机回归(SVR)建立过氧化值的校正模型,以决定系数( $R^2$ )、交互验证均方根误差(RMSECV)和预测均方根误差(RMSEP)为指标评价模型的性能,以上数据处理均采用CAMO公司的Unscrambler 10.4软件完成。

## 2 结果与讨论

### 2.1 植物油近红外光谱分析

图1(a)为不同过氧化值大豆油(一级、三级)和菜籽油(一级、三级和四级)的近红外光谱。由图可知,不同种类、不同等级两种植物油的近红外光谱差异较小且相互重叠,在波长1 210, 1 392, 1 412, 1 724, 1 758, 2 140和2 178 nm附近有明显吸收峰。其中,1 210 nm附近为C—H伸缩振动的二级倍频吸收,1 409~1 432, 1 449~1 469, 1 475~1 536 nm范围为过氧化物中O—H一级倍频的吸收谱带<sup>[11]</sup>; 1 700~1 800 nm区域为甲基、亚甲基和双键中C—H一级倍频的吸收<sup>[12]</sup>; 1 926 nm附近与甘油酯、醛、酮等化合物中的C=O的二级倍频吸收密切相关<sup>[11]</sup>; 在2 175 nm左右为双键的合频吸收<sup>[7]</sup>。鉴于油脂近红外光谱重叠严重以及干扰背景的存在,结合二维相关光谱进行光谱分析,以提高其光谱

表征峰的分辨率。由于波长2 200 nm以上干扰信号强、信噪比较低,故选择780~2 200 nm波长范围进行分析。

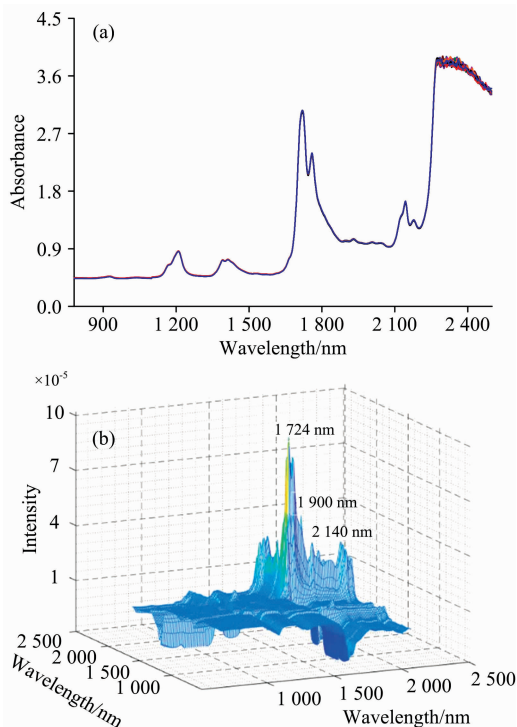


图1 两种植物油近红外光谱(a)及二维同步相关光谱立体图(b)

Fig. 1 (a) Near infrared spectra; (b) Two-dimensional synchronous correlation spectra

图1(b)为大豆油二维相关同步光谱的立体图,X轴、Y轴均为波长,Z轴为相对强度,反映光谱信号随外部扰动发生变化的协同程度,植物油中过氧化值浓度增大时,其光谱变化主要分布于1 700~2 200 nm区域,在对角线1 724, 1 900和2 140 nm处出现了3个较强的自相关峰,说明这三个波长点所对应的基团随着油脂氧化进程的变化较明显,即为与过氧化值密切相关的光谱波长。同时,在对角线两侧(1 724, 1 760), (1 704, 2 140), (1 724, 1 900)出现了3个正交叉峰及一些相对较弱的正交叉峰,表明这些波长下的光谱信号也与油脂中的过氧化物有关。同理,通过二维相关光谱对菜籽油的一维近红外光谱进行分析,所得结果与此相同。可见,近红外光谱能够鉴别植物油中过氧化值的变化情况。

### 2.2 建模条件

#### 2.2.1 波段

通用性模型建立需要大量光谱数据,这些数据中,除了待测成分外还有许多无用信息,而且相邻波段间也存在信息冗余,即使利用光谱预处理方法也很难完全消除,导致模型不稳定、准确性较差。故采用间隔偏最小二乘法进行波段选择,在波长780~2 200 nm范围内将油样光谱分割成10个区域,对未经预处理的样本数据建模,结果如图2所示。

由图可知,植物油的等级、种类对过氧化值的检测波段

选择具有一定的影响；以过氧化值模型 RMSECV 平均值为标准，两种植物油同等级（一级或三级）低于标准值的波长区域相同，而对于同种油脂不同等级的波长区域却不完全相同，如一级菜籽油与三级、四级菜籽油的波长区域不同。这可能是因为一级油的精炼程度高，具有无味、色浅、烟点高等特点；而三、四级油脂间的组成更为相似，仅经过了简单脱胶、脱酸等过程，杂质的含量较高，同时保留相对较多的

维生素 E、胡萝卜素等物质。此外，一级、三级大豆油和一级菜籽油的最佳波段数均为 9，而三级、四级菜籽油的最佳波段为 10；与波段选择前相比，两种植物油最佳波段下模型的预测准确度平均提高了 25% 左右。结合两种植物油的 RMSECV 值（图 2）和二维相关光谱 [图 1(b)]，选择 8、9 和 10 三个波长区域，即 1 700~2 200 nm 范围作为建立过氧化值通用模型的波段。

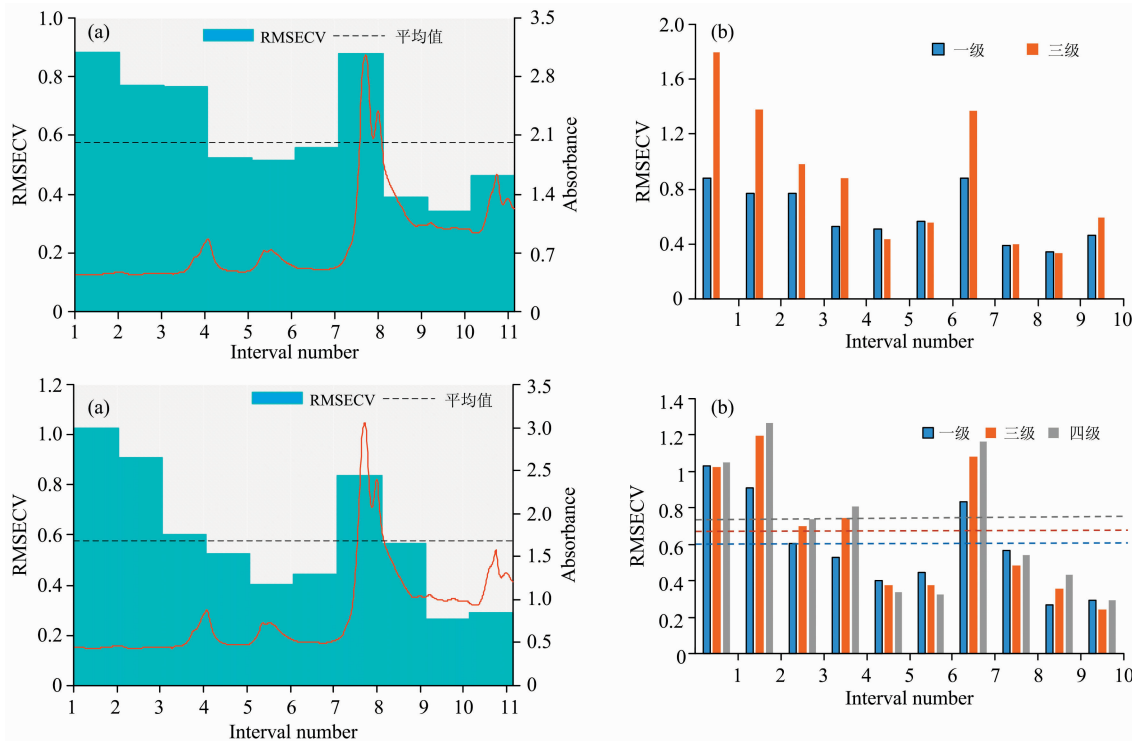


图 2 两种植物油不同波段模型的预测结果

(a): 大豆油; (b): 菜籽油

Fig. 2 Prediction results of two vegetable oils in different bands

(a): Soybean oil; (b): Rapeseed oil

### 2.2.2 预处理方法

为避免操作、仪器性能、样品物理特性变化等对光谱的干扰，采用不同预处理方法对光谱数据进行处理，在波长 1 700~2 200 nm 范围内建立两种植物油的 PLS 模型，结果如表 1 所示。

由表可知，通过 SD 处理后的模型效果最佳，与预处理前相比，大豆油、菜籽油模型的 RMSEP 分别降低了 14.0%~26.9% 和 2.3%~49.9%，这是因为 SD 法能有效增强光谱差异性，消除基线和其他背景的干扰，提高其分辨率、灵敏度以及有效信息率；OSC 法去除了与被测物质无关的信号，经 OSC 处理的一级油（大豆油和菜籽油）模型的  $R^2$  分别升高了 5.6% 和 3.8%，RMSEP 分别降低了 15.8% 和 31.3%；SNV 主要是消除样品表面散射等的影响，使被掩蔽掉的某些光谱信息显示出来，对于四级菜籽油光谱的作用显著，模型的 RMSEP 不高于 0.35。可见预处理方法能够提高过氧化值模型的准确性，本研究选择 SD 法作为通用模型的预处理方法。

表 1 不同光谱预处理后 PLS 模型的预测结果

Table 1 Prediction results of PLS models with different spectral pretreatment methods

种类	预处理方法	一级		三级		四级	
		$R^2$	RMSEP	$R^2$	RMSEP	$R^2$	RMSEP
大豆油	无	0.839	0.547	0.967	0.480		
	OSC	0.886	0.461	0.967	0.478		
	SNV	0.852	0.526	0.968	0.471		
	SD	0.924	0.400	0.975	0.413		
菜籽油	无	0.933	0.424	0.991	0.266	0.949	0.566
	OSC	0.969	0.292	0.993	0.229	0.955	0.535
	SNV	0.961	0.323	0.991	0.267	0.991	0.334
	SD	0.970	0.285	0.998	0.139	0.987	0.284

### 2.2.3 建模方法

在最佳波段范围内，经 SD 处理后分别采用 PCR，PLS 和 SVR 建立两种植物油过氧化值的定量分析模型，结果如图 3 所示。PLS，PCR 和 SVR 三种模型的平均预测准确性分

别为 0.304, 0.639 和 0.499, 除一级菜籽油 PCR 模型外, 其他模型的  $R^2$  均大于 0.85, PLS 的预测效果明显好于 PCR 和 SVR, 能够更准确预测油样的过氧化值。其原因可能是 PLS 模型将被测组分浓度信息引入到光谱矩阵的分解过程中, 计算每一个主成分前都将光谱矩阵的得分与浓度矩阵的得分进行交换, 使得光谱矩阵主成分与被测组分浓度关联<sup>[13]</sup>; 而 PCR 模型通过对光谱数据的分解来消除无用的噪声信息, 在分解光谱矩阵时没有考虑到被测组分浓度的影响; SVR 模型是一种非线性建模方法, 其主要思想是将原问题通过非线性变换转化为高维空间的线性问题并求解, 然而油样的过氧化值与光谱数据间具有较好的线性关系, 采用 SVR 法不仅增加计算量, 而且可能引起“维数灾难”, 导致模型的预测效果不佳。综上, 采用 PLS 法建立两种植物油的通用模型更为合适。

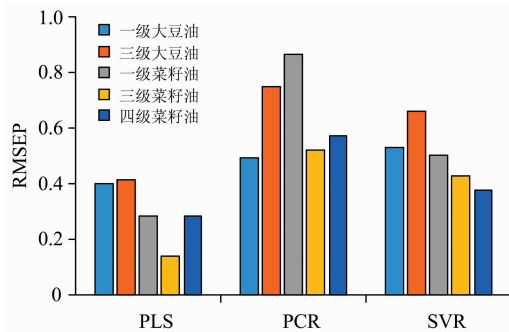


图 3 不同建模方法预测结果的比较

Fig. 3 Comparison of prediction results of different modeling methods

### 2.3 通用模型的建立与预测

建立四种过氧化值检测的通用模型, 混合样本集分别为大豆油(一级+三级)、菜籽油(一级+三级+四级)、一级油(大豆油+菜籽油)、三级油(大豆油+菜籽油), 结果如图 4 所示。

由图可知, 一级油通用型模型的预测效果最佳, RMSEP,  $R^2$  分别为 0.412 和 0.920, 与单一油脂(一级大豆油和一级菜籽油)模型结果相差不大; 而三级油通用模型的预测精度较低, 即 RMSEP 值大于 1.5,  $R^2$  低于 0.7, 这说明三级大豆油和三级菜籽油并不适合建立同一模型。对于同一种植

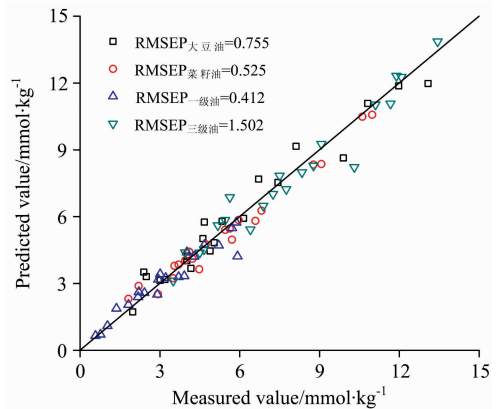


图 4 不同通用模型的预测结果

Fig. 4 Prediction results of different general models

物油, 如大豆油的通用模型相比于一级、三级大豆油模型的准确性至少降低了 82%, 菜籽油通用模型的 RMSEP 增加了 84% 以上。这表明油脂等级对过氧化值通用模型建立有较大影响。综上, 建立一级油脂(大豆油和菜籽油)的通用模型是可行的。但是, 通用模型本身也具有一定的局限性, 就是对原料或生产工艺变化相对较大的新产品难以得到理想的预测结果, 需要用新产品及时对模型进行更新, 同时扩展通用模型的性能。

### 3 结 论

不同种类、等级的植物油, 生产工艺、油脂自身特性等方面的不同, 造成其近红外光谱存在差异, 建立能同时检测不同种类、等级植物油通用模型具有一定难度。以大豆油和菜籽油为研究对象, 考察了通用模型的最佳建模条件, 研究了大豆油、菜籽油、一级油、三级油四种通用模型的可行性。通过比较得到了通用模型的最佳波段、预处理方法和建模方法分别为 1 700~2 200 nm、二阶导数和 PLS 法, 四种通用模型中一级植物油(大豆油和菜籽油)的过氧化值通用模型预测结果最佳, 与单一模型相比, 预测准确度相差不大。此外, 由于样品数量的局限, 两种一级植物油的通用模型还需进一步更新和维护。

### References

- [1] Woo Y S, Kim M J, Lee J H. Food Chemistry, 2019, 279: 216.
- [2] Budilarto E S, Kamal-Eldin A. European Journal of Lipid Science and Technology, 2015, 117(8): 1095.
- [3] Cebi N, Yilmaz M T, Sagdic O, et al. Food Chemistry, 2017, 225: 188.
- [4] Wang N, Ma T, Yu X, et al. Food Analytical Methods, 2016, 9(5): 1412.
- [5] TAO Sha, WANG Wei, CHEN Fu-sheng, et al(陶莎, 王玮, 陈复生, 等). Journal of the Chinese Cereals and Oils Association(中国粮油学报), 2015, 30(1): 129.
- [6] Krzysztof W, Igor K, Marek S, et al. Food Chemistry, 2015, 187: 416.
- [7] Mureşan V, Danthine S, Mureşan A, et al. Talanta, 2016, 155: 336.
- [8] JIANG Qi, ZHANG Yu, YANG Hua, et al(蒋琦, 张玉, 杨华, 等). Acta Agriculturae Zhejiangensis(浙江农业学报), 2017, 29(11): 1897.

- [9] LI Shui-fang, FU Hong-jun, MA Qiang, et al(李水芳, 付红军, 马强, 等). *Chemistry and Industry of Forest Products(林产化学与工业)*, 2017, 37(4): 137.
- [10] Park Y, Jin S, Noda I, et al. *Journal of Molecular Structure*, 2018, 1168: 1.
- [11] Giese E, Winkelmann O, Rohn S, et al. *Food Research International*, 2018, 106: 116.
- [12] Livia C Carvalho, Marcondes L Leite, Camilo L M Morais, et al. *LWT-Food Science and Technology*, 2019, 103: 101.
- [13] CHU Xiao-li(褚小立). *Practical Handbook for Near Infrared Spectroscopy Analysis(近红外光谱分析技术实用手册)*. Beijing: China Machine Press(北京: 机械工业出版社), 2016.

## A General Model for the Peroxidation Values of Two Vegetable Oils Based on Near Infrared Spectroscopy

PENG Dan, LI Lin-qing, LIU Ya-li, BI Yan-lan\*, YANG Guo-long

School of Food Science and Technology, Henan University of Technology, Zhengzhou 450001, China

**Abstract** The rapid and accurate detection of peroxide value is of great significance to the quality of edible oils and food safety control. Near-infrared (NIR) spectroscopy is an ideal method for measuring peroxide values, but the establishment of calibration models requires a lot of resources. In this paper, based on the relationship between near-infrared spectroscopy and lipid peroxide value, the feasibility of establishing a general calibration model for measuring peroxide value in different types and levels of vegetable oils was studied. First, different grades of soybean oil and rapeseed oil were studied. The near-infrared spectra of the two vegetable oils were analyzed by two-dimensional correlation spectroscopy, and the optimal detection band of the general model for peroxide value was selected by interval least squares method (iPLS). Then, the effects of orthogonal signal correction (OSC), standard normal variable transformation (SNV) and second derivative (SD) on the prediction precision of general models were investigated. Further, the prediction performances of three modeling methods, including principal component regression (PCR), partial least squares (PLS) and support vector machine regression (SVR), were compared in detail. At last, four general prediction models for soybean oil (including first-grade and third-grade), rapeseed oil (including first-grade, third-grade and fourth-grade), first-grade vegetable oils (including soybean oil and rapeseed oil) and third-grade vegetable oil (including soybean oil and rapeseed oil) were constructed. The experimental results showed that the change of peroxide value of vegetable oil could be detected by NIR spectroscopy technology, and its spectral changes mainly resided in the region from 1 700 to 2 200 nm. The optimal band, preprocessing method and modeling method of the general model were 1 700~2 200 nm, SD method and PLS method, respectively. Among the four general models, the one for the first-grade vegetable oils (including soybean oil and rapeseed oil) can get better performance. The root means square error of prediction (RMSEP) and the square correlation coefficient ( $R^2$ ) are 0.412 and 0.920, respectively. Compared with the models of first-grade soybean oil or rapeseed oil, the prediction results are roughly the same. It meant that it was feasible to establish a general model with high accuracy between vegetable oils with similar production processes for reducing the workload of repetitive modeling. In addition, in order to expand the versatility of the general model, it was necessary to continuously update the model with new kinds of vegetable oils.

**Keywords** Near-infrared spectroscopy; General model; Chemometrics; Vegetable oil; Peroxide value

(Received Jun. 9, 2019; accepted Oct. 20, 2019)

\* Corresponding author