

文章编号:1671-1513(2012)01-0066-03

光谱分辨率的选取对食用油近红外模型性能影响分析

吴静珠, 刘翠玲, 邢素霞, 孙梅

(北京工商大学 计算机与信息工程学院, 北京 100048)

摘要: 近红外光谱仪的类型和工作参数的设定对测量过程和分析结果均有一定影响. 重点讨论了在傅立叶变换型光谱仪上设定不同的光谱分辨率对食用油近红外定量模型性能的影响. 实验采用 Vertex 70 光谱仪, 在 3 种光谱分辨率($4, 8, 16 \text{ cm}^{-1}$) 条件下, 采用透射式液体光纤探头采集 60 份食用油样品近红外谱图. 针对 3 组近红外光谱样品集, 分别采用蒙特卡罗采样法剔除异常样品, 根据 Kennard-Stone 法划分校正集和校验集后, 建立优化食用油棕榈酸近红外定量分析模型并作预测. 结果表明基于 16 cm^{-1} 建立的食用油近红外模型指标优于 4 cm^{-1} 和 8 cm^{-1} 所建的模型, 该结果可以为均匀液体作近红外检测时光谱分辨率的设定提供参考.

关键词: 光谱分辨率; 近红外; 蒙特卡罗采样; 食用油

中图分类号: TS207.3

文献标志码: A

近红外光谱仪的类型、工作状态和扫描时各参数的设定对测量过程和分析结果均有一定影响^[1]. 根据采用的分光技术不同, 近红外光谱仪的类型可划分为滤光片型、光栅扫描型、阵列检测型、傅立叶变换型和声光过滤调制型等. 其中, 傅立叶型光谱仪器已经成为目前实验室红外光谱仪器的主导产品^[2]. 一般情况下, 近红外光谱仪类型的选择局限于实验室所购买的仪器. 而光谱仪的工作状态(如波长准确度、信噪比等)一般在仪器上都带有硬件工作状态的自诊断软件来告知用户当前仪器是否工作在正常状态, 因此无需过多考虑. 但是光谱扫描时需要人为设定的一些参数, 如光谱分辨率, 特别需要用户引起注意, 因为光谱分辨率影响样品光谱的质量, 最终影响分析的准确性. 这类参数若在实验过程中设置得当, 会有助于用户得到准确可靠的近红外分析结果, 反之, 则会严重影响近红外分析结果. 如何设定这类参数, 需要用户反复测试摸索一些经验值, 并结合具体测试对象状态(如粉末、籽粒、均匀液态、悬浊液等)而定. 本文重点研究在傅立叶变换型光谱仪上设定不同的光谱分辨率对食用油近红外定量模型性能的影响, 探索对于食用油这

类均匀液态物质在作近红外检测时设定光谱分辨率的参考值.

1 材料与方法

1.1 实验材料

60 个植物油样品, 包括花生油、大豆油、橄榄油、芝麻油、葵花籽油、玉米胚芽油、芥花油等, 均购自当地超市.

1.2 食用油棕榈酸测定

采用气相色谱法测定 60 个食用油样品中棕榈酸含量.

1.3 近红外光谱采集

VERTEX 70 型红外光谱仪, 德国 Bruker 公司. 采集样品近红外光谱, 使用液体光纤探头采样方式, 光程 2 mm. 仪器参数设定如下: 波数范围 $4\ 000 \sim 12\ 500 \text{ cm}^{-1}$, 每个样品重复扫描次数 32 次.

一般要求仪器的分辨率比测量峰宽小 1/10 左右. 样品在近红外区域的吸收多为宽峰且重叠严重, 进行定量分析时不要求高的仪器分辨率. 在实际应用中, 通常 16 cm^{-1} 或 10 nm (在 $2\ 500 \text{ nm}$ 处) 的

收稿日期: 2011-10-28

基金项目: 北京市优秀人才培养资助项目(20081D0500300130).

作者简介: 吴静珠, 女, 副教授, 博士, 主要从事基于近红外光谱的农产品及食品检测技术方面的研究.

分辨率就可满足大部分分析对象的应用要求.但对于结构特征十分相近的复杂样品(如农产品等),要得到准确的分析结果,就要对仪器的分辨率提出一定的要求,一般也不会超过 4 cm^{-1} [2].综上所述,本实验设定分别在光谱分辨率为 $4, 8, 16\text{ cm}^{-1}$ 时扫描样品,相应地波长点个数分别为 $4\ 407, 2\ 203$ 和 $1\ 102$.显然,分辨率设得越高,扫描速度越慢,扫描时间延长,而且图谱所占的存储空间增大.因此,在不影响图谱质量的前提下,分辨率可以不取太高.图1为分辨率 16 cm^{-1} 时样品的近红外谱集.

全部样品未经任何化学处理,将光纤探头插入装有样品的小瓶中,逐一扫描样品,每次测量前均用石油醚清洗探头,避免样品间交叉污染.

图1 样品集近红外光谱图

Fig.1 Near-infrared spectroscopy of samples

2 结果与分析

实验中食用油棕榈酸近红外定量分析模型的建立采用近红外分析的常规流程,包括:异常样品剔除,校正集样品选择以及模型的建立与测试.近红外校正模型的预测精度和稳健性采用主成分数(nf),决定系数 R^2 ,交叉校验标准差(root mean square error of cross validation, RMSECV),预测均方根误差(root mean square error of prediction, RMSEP) 指标评价.

2.1 异常样品剔除

建立性能较好的近红外模型,进行异常样品的剔除是非常必要的.这里采用蒙特卡洛采样法进行异常样品的剔除[3],在MATLAB R2008a中实现.图2为在分辨率为 4 cm^{-1} ,随机采样1000次的条件下,各个样品的预测均方根误差(RMSEP)的统计信息(均值MEAN和标准偏差STD)的分布图.设定均值为2,标准偏差为0.5的条件下可将图2划分成A、B、C、D 4个区域.

其中,A区域样品的预测均方根误差的均值和

图2 60个样本的RMSEP统计信息分布图

Fig.2 Statistical information distribution of 60 samples' RMSEP

标准偏差较小,因此属于正常样本;B区域样本预测均方根误差的均值较大,属于浓度异常样本;C区域样本的误差均值和标准偏差较大,属于异常样本;D区域样本的标准偏差较大,属于光谱异常本.因此,BCD区域的样本均可作为异常本剔除.光谱分辨率为 $4, 8, 16\text{ cm}^{-1}$ 时,采用蒙特卡洛法剔除异常样本信息,如表1.从表中可以看出,除样本5外,在不同分辨率条件下,剔除的异常样本大部分都是不同的.因此这也从一个侧面反映出在不同光谱分辨率条件下,样本所含的光谱信息或者说光谱质量是有明显不同的.

表1 剔除异常样本相关信息

Tab.1 Related information of abnormal samples eliminated

分辨率/ cm^{-1}	均值	标准偏差	异常样品序号
4	2	0.5	2,5,25,29,9
8	2.4	1	2,5,17,49,51
16	1.2	0.8	5,6,10,37,51

2.2 校正集样品选择

Kennard-Stone法[4]将光谱差异较大的样品选入校正集,而其余较相近的样品进入校验集,这样可使有代表性的样品全部进入校正集,从而在一定程度上避免了校正集样品分布的不均匀.将剔除异常样本后的样本集采用Kennard-Stone法划分校正集和校验集,在MATLAB R2008a中实现,结果如表2.表2中,经过Kennard-Stone法划分得到校正集的范围涵盖了校验集的范围,因此校正集所建模型可用于校验集的预测.在不同分辨率条件下,基于光谱信息的Kennard-Stone法挑选出的校验集

样本明显不同,因此可见相同样本在不同分辨率下采集得到的光谱信息是有很大差异的。

表2 Kennard-Stone 法划分的样本集信息

Tab. 2 Classification information of sample set by Kennard-Stone method

分辨率	校正集样本个数	校正集浓度范围/%	校正集标准偏差	校验集样本个数	校验集浓度范围/%	校验集标准偏差
4 cm ⁻¹	40	7.6 ~ 16.8	2.39	15	7.6 ~ 14.4	2.26
8 cm ⁻¹	40	7.6 ~ 16.8	2.36	15	8.5 ~ 15.3	2.17
16 cm ⁻¹	40	7.6 ~ 16.85	2.49	15	10.8 ~ 14.9	1.34

2.3 模型建立与预测

采用5折交叉校验的偏最小二乘回归法(partial least squares regression method, PLS)建立食用棕榈酸的近红外定量分析模型,并对校验集作预测。上述建模和预测方法均在MATLAB R2008a中实现。结果如表3。从表中可看出,采用全谱建模时模型指标不是很好,但是仍旧能得出分辨率为16 cm⁻¹所建模型及预测结果远优于8 cm⁻¹和4 cm⁻¹。

表3 5折-pls建模分析结果

Tab. 3 Modeling analysis results by 5 fold-pls

分辨率	nf	R ²	交叉校验 标准差	预测均方根 误差
4 cm ⁻¹	7	0.8167	1.0111	0.8280
8 cm ⁻¹	15	0.7912	1.0655	1.0317
16 cm ⁻¹	14	0.9384	0.6094	0.6672

实验采用CARS(competitive adaptive reweighted sampling, CARS)^[5]波长变量挑选方法对这3组样品集进行优化建模,结果如表4。同样可以得出相同的结论:在分辨率为16 cm⁻¹所建模型的预测能力优于8 cm⁻¹和4 cm⁻¹所建模型。

表4 cars-pls建模分析结果

Tab. 4 Modeling analysis results by cras-pls

分辨率	nf	R ²	交叉校验 标准差	预测均方根 误差	保留波 点数
4 cm ⁻¹	12	0.9958	0.1536	0.6612	40
8 cm ⁻¹	12	0.9958	0.1506	0.4909	30
16 cm ⁻¹	13	0.992	0.2201	0.4672	23

本实验中,对于食用油这类均匀液体做近红外检测,分析其常规指标棕榈酸时,在分辨率为16 cm⁻¹时扫描样品建模即可获得较好的模型指标。实验表明,并非近红外光谱分辨率越高,模型性能越好。因为傅立叶变换性近红外光谱仪的分辨率由动镜移动距离决定,分辨率越高,扫描速度越慢,单位

时间内的信噪比也会降低^[6]。实验测试过程中,分辨率到底该如何设定,需视具体情况而定,如果是日常快速分析,分辨率可以较低,如果是为了研究,如建立标准光谱库或需要细节的判别分析,则可适当提高扫描光谱的分辨率。

3 结 论

本工作在Vertex 70光谱仪上,设定3种光谱分辨率(4, 8, 16 cm⁻¹)分别采集60份食用油样本的近红外光谱,重点研究了不同光谱分辨率对食用油近红外定量模型性能的影响。结果表明光谱分辨率的高低与模型性能之间并无绝对的关系,对于食用油这类均匀液态物质中的常规组分作近红外检测时,光谱分辨率的参考值可首选16 cm⁻¹。

参考文献:

- [1] 严衍祿,赵龙莲,韩东海,等. 近红外光谱分析基础与应用[M]. 北京:中国轻工业出版社,2005:240-300.
- [2] 陆婉珍. 现代近红外光谱分析技术[M]. 2版. 北京:中国石化出版社,2007:95-140.
- [3] Cao Dongsheng, Liang Yizeng, Xu Qingsong, et al. Toward better QSAR/QSPR modeling: simultaneous outlier detection and variable selection using distribution of model features[J]. Journal of Computer-Aided Molecular Design, 2011(25):67-80.
- [4] 吴静珠,王一鸣,张小超,等. 近红外光谱分析中定标集样品挑选方法研究[J]. 农业机械学报,2006,37(9):80-82.
- [5] LI Hongdong, LIANG Yizeng, XU Qingsong, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration[J]. Anal Chim Acta, 2009, 648(1):77-84.
- [6] 陆婉珍,袁洪福,褚小立. 近红外光谱仪器[M]. 北京:化学工业出版社,2010:107-114.

Study on Meat Quality Detection Techniques

FU Yan¹, GUO Pei-yuan¹, SUN Mei¹, XU Ran-ran²

(1. *School of Computer Science and Information Engineering, Beijing Technology and Business University, Beijing 100048, China*; 2. *Zolix Instruments Co. Ltd., Beijing 101149, China*)

Abstract: The artificial olfactory system (AOS) and artificial taste system (ATS) detection techniques, computer vision detection technology, near infrared spectroscopy (NIR) detection technology, and hyperspectral image detection technology were described in this paper. Hyperspectral image detection technology has become a research hotspot for meat quality detection. The basic principle of hyperspectral image detection technology and the research status in meat quality detection using hyperspectral image at home and abroad were introduced in the paper. The prospect of future research in meat quality detection using hyperspectral image detection technology was proposed to researchers on the related study as a reference.

Key words: meat quality; nondestructive detection technique; hyperspectral image detection technology

(责任编辑:王 宽)

(上接第 68 页)

Research on NIR Model of Edible Oil Impacted by Spectral Resolution Determination

WU Jing-zhu, LIU Cui-ling, XING Su-xia, SUN Mei

(*School of Computer Science and Information Engineering, Beijing Technology and Business University, Beijing 100048, China*)

Abstract: There are some effects on near infrared analysis result caused by the type of near infrared spectrometer and working parameters settings. How the different spectral resolution on Fourier transform spectrometer impact quantitative NIR model of edible oil was discussed in this paper. The experiment collected near infrared spectrum of 60 edible oil samples using Vertex 70 spectrometer, in three kinds of spectral resolution (4, 8, 16 cm^{-1}) conditions, by the transmission type liquid optical fiber probe. Monte Carlo sampling method was used to analysis three groups of near infrared spectral sample set and eliminate the abnormal samples. The calibration set and validation set were divided according to Kennard-Stone method. The quantitative models of palm acid were established, optimized and validated. Experiments shown that the model built based on 16 cm^{-1} had better performance than 4 cm^{-1} and 8 cm^{-1} . The results can provided a reference spectral resolution while detecting uniform liquid by near infrared technology.

Key words: spectral resolution; near infrared spectroscopy; Monte Carlo sampling; edible oil

(责任编辑:王 宽)