

拉曼光谱结合偏最小二乘同时测定烟草中叶黄素和 β -胡萝卜素含量

葛 炯¹, 孙 琳², 沈晓洁¹, 沙云菲¹, 黄天雄², 杜一平², 张维冰², 谢雯燕^{1*}, 姚鹤鸣^{1*}

1. 上海烟草集团有限责任公司技术中心, 上海 200082

2. 华东理工大学化学与分子工程学院, 上海功能性材料化学重点实验室, 上海 200237

摘 要 色素是烟草中的重要组成成分, 烟草的外观质量和内在品质都与色素相关。色素主要分为绿色素、黄色素和黑色素。在生长期, 烟草中的绿色素主要是叶绿素, 成熟期的烟草中黄色素则主要为叶黄素和胡萝卜素, 黑色素则存在于成熟的烟叶中, 或在调制和发酵的过程中产生。色素的检测对于烟草原料评价和成品质量都非常重要。鉴于烟草色素基于液相色谱的传统检测方法耗时较长、样品制备过程复杂, 以及拉曼光谱操作简单、测定时间短、能提供有关分子官能团信息的特点, 开发了一种应用拉曼光谱同时定量检测烟草中叶黄素和 β -胡萝卜素的快检方法。烟草样品的有机溶剂提取物, 密封在透明玻璃瓶中, 将激光聚焦于瓶内溶液样品直接测定拉曼光谱。研究发现, 不同于常见的 514 和 785 nm 激发波长, 在短波 455 nm 激光作用下, 可获得更理想的拉曼信号, 光谱强度较高, 荧光干扰较低。对色素萃取溶剂、焦平面与光学平台间的距离等实验条件进行了优化。考虑到不同检测日期仪器操作条件的变化可对拉曼光谱产生影响, 以溶剂峰为内标峰, 对光谱进行归一化处理, 以校正因测定条件不稳定而引起的物理干扰。为了解决荧光干扰, 以及色素分子之间的拉曼光谱干扰问题, 采用偏最小二乘法(PLS)建立光谱分析的多元校正模型。建模中对波长区域以及光谱预处理方法进行了优化。研究结果表明, 光谱归一化处理显著降低了因物理因素产生的光谱干扰, 光谱求导运算对建模影响不大, 而波长选择能明显改善模型的预测能力。利用 798.2~1 752.8 cm^{-1} 波段建立叶黄素的 PLS 定量模型时, 预测效果最佳, 预测集的均方根误差(RMSEP)为 6.68 $\mu\text{g} \cdot \text{g}^{-1}$; 对于 β -胡萝卜素, 798.2~1 752.8 cm^{-1} 组合 2 254.2~2 784.5 cm^{-1} 用来建模时得到的 RMSEP 最小, 为 2.56 $\mu\text{g} \cdot \text{g}^{-1}$ 。该方法操作简便、耗时少、结果准确可靠, 为烟草样本中色素的定量分析提供了一种切实有效的新途径。

关键词 拉曼光谱; 烟草; 色素; 偏最小二乘法

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)11-3519-06

引 言

迄今为止, 人们在烟叶和烟气的混合体系中发现了 9 000 多种化学物质。色素是影响烟草香气质的主要成分之一, 它不仅直接影响烟草的外观质量, 还直接或间接的影响烟草的内在品质。烟草中色素的含量和组成与烟草的类型、种类、生长阶段以及处理方法有关。在烟草的生长期, 主要有叶绿素 a、叶绿素 b、 β -胡萝卜素等主要色素^[1]。叶绿素是烟草中主要的绿色素, 而在烟草的成熟期主要产生黄色素, 其中大量是类胡萝卜素。类胡萝卜素主要包括胡萝卜素和叶黄素, 它们占据烟草中色素含量的 1/5~1/3。而本文的主要内容就是对烟草中的叶黄素和 β -胡萝卜素进行定量分析。

色素的各类定量分析方法中^[2-3], 主要采用的是液相色谱-质谱联用技术(LC-MS)^[4]。虽然 LC-MS 法准确, 但它耗时长, 还需要复杂的样品制备过程^[5]。拉曼光谱检测速度快、操作简单, 且能提供精细的分子结构信息, 已经应用到生物、化学、环境等样品中色素的检测中^[6-7]。拉曼光谱结合化学计量学方法, 被应用于像微藻种类鉴定、水污染鉴定^[8], 餐饮中的植物油和废弃食用油的辨别^[9], 不同氮营养条件下水体的识别^[10], 藻脂类中脂肪酸的非饱和与饱和碳碳键的比值测定^[11], 体内化合物的分布和浓度测定^[12], 以及生物、化学、食品样品的检测、分类、识别等^[13-14]。

本工作对烟草提取物进行拉曼光谱检测, 建立两种色素, 叶黄素和 β -胡萝卜素的拉曼光谱定量分析方法。考虑到拉曼谱峰重叠严重, 还受到荧光光谱的干扰, 采用多元校正

收稿日期: 2018-09-20, 修订日期: 2019-01-13

基金项目: 国家自然科学基金面上项目(21475044)和中国烟草总公司科技重大专项(中烟办[2016]259)资助

作者简介: 葛 炯, 1977 年生, 上海烟草集团有限责任公司技术中心工程师 e-mail: gej@sh.tobacco.com.cn

* 通讯联系人 e-mail: xiewy@sh.tobacco.com.cn; yaohm@sh.tobacco.com.cn

方法, 偏最小二乘回归(PLS)建立叶黄素和 β -胡萝卜素的定量分析模型, 预测未知样本中两种色素的浓度。

1 实验部分

1.1 试剂

叶黄素和 β -胡萝卜素的标准物购自 Sigma 化学试剂有限公司, 正己烷($w\% \geq 97\%$)、异丙醇($w\% \geq 99.5\%$)均为分析纯, 购自上海凌峰化学试剂有限公司。所有化学试剂在使用过程中均未进行再提纯。配制样品溶液的实验用水均为超纯水($18.2 \text{ M}\Omega \cdot \text{cm}$, Sartorius arium 611DI, 德国)。

1.2 仪器

使用 DXR2xi 共聚焦显微拉曼光谱仪(ThermoFisher, 美国), 配备 EMCCD 检测器和 455 nm 的 DPSS 激光器, 激光最大输出功率 6 mW。采用 BX51 显微镜的 10 倍物镜将激光聚焦在待测样品上。光谱分辨率为 5 cm^{-1} 。在使用拉曼光谱法测定烟草萃取液之前, 用聚苯乙烯对拉曼光谱仪进行了校正。此后, 除特殊说明外, 实验数据均在相同条件下获得。紫外-可见光谱仪(Lambda 950, PerkinElmer)。

1.3 烟草样品萃取溶液的提取

称量($1.000 0 \pm 0.003 0$)g 烟末, 放入玻璃瓶, 用定量加液器加入 95% 的正己烷 10 mL。然后超声振荡 30 min, 静置 30 min, 取出澄清液过 $0.22 \mu\text{m}$ 膜, 最后取 2 mL 到玻璃色谱瓶, 待测。

1.4 拉曼光谱的采集

将装有烟草样本提取液 2 mL 的色谱瓶密封后置于光学检测平台上进行拉曼光谱采集, 扫描次数 1 000 次, 扫描波段为 $50 \sim 3\,500 \text{ cm}^{-1}$ 。每个样本选取 3 个不同的位置各采集一条光谱, 最后选用平均光谱进行分析。特别选择 1 个样品, 在全部样品共计 5 d 的检测过程中进行连续的监测, 平行测定 3 次, 其平均光谱用于全部样品光谱的归一化处理, 以校正因每次检测环境改变而产生的光谱改变, 详见 2.2 部分。

1.5 数据分析

本文探索了归一化、求导和波段选择等光谱预处理方法对建模效果的影响。其中, 归一化是计算光谱与溶剂峰($2\,877 \text{ cm}^{-1}$)的比值; 求导采用 Savitzky-Golay 方法, 窗口大小选为 9, 拟合阶数为 3。

为了建立及评价多元校正模型, 所有 141 个烟草样本被分为校正集和独立预测集。对浓度值进行排序后, 每三个连续样本中的第二个被选出来作为预测样本, 剩下的作为校正样本。因此, 校正集和预测集的样本数分别为 94 和 47。

利用 Monte-Carlo 交互检验方法选取 PLS 模型中的隐变量数。通过计算 Root Mean Square Error of Prediction(RMSEP)和 Root Mean Square Error of Calibration(RMSEC)来对模型进行评价, 计算公式如下

$$\text{均方根误差(RMSE)} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$$

式中, n 为样本数, y_i 和 \hat{y}_i 分别是第 i 个样本的参考浓度和

预测浓度, 在本文中, 光谱数据 X 和浓度 y 均做过中心化处理。

2 结果与讨论

2.1 实验条件的优化

2.1.1 激发光波长的选择

在激发光波长的选取中考虑了 455, 514 和 785 nm 这 3 种激发光波长。从图 1 展示的叶黄素和 β -胡萝卜素的紫外-可见吸收光谱来看, 与 455 nm 的激发光波长相比, 514 和 785 nm 的激发光波长看起来更适合类胡萝卜素, 因为荧光背景较弱。但是根据散射定律, 入射光的波长和拉曼强度的四次方成反比^[15], 因此在 455 nm 的激发光波长能产生更强的拉曼信号。早在 1979 年, Forrest^[16] 就建议选取 488.0 和 514.5 nm 的激发光波长来对烟草中的叶黄素与 β -胡萝卜素进行测定。综合考虑拉曼强度与荧光背景, 最终选择采用 455 nm 作为激发光波长。

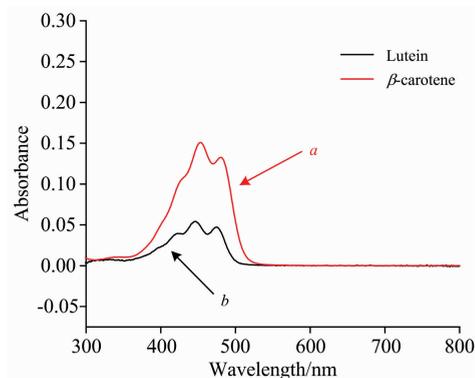


图 1 β -胡萝卜素(a)和叶黄素(b)的紫外-可见吸收光谱

Fig. 1 UV-Vis spectra of β -carotene (a) and lutein (b)

2.1.2 萃取溶剂的选择

在拉曼光谱的测量过程中, 萃取溶剂对其有很大影响^[17]。如图 2 所示, 图 2a 和 b 分别为正己烷和异丙醇标准溶液的拉曼光谱图, 结合他们分别作为萃取溶剂的烟草样本提取液(图 2c 和 d)的拉曼光谱可以看出, 选取 $1\,159$ 和 $1\,529 \text{ cm}^{-1}$ 处的谱峰可以作为叶黄素和 β -胡萝卜素的特征峰。因为采用异丙醇(图 2d)作为萃取剂得到的拉曼光谱的信号强度比正己烷(图 2c)作为萃取剂得到的拉曼光谱的信号强度低。而且从图 2d 光谱上的齿轮形状来看, 以异丙醇作为萃取剂所产生的荧光背景更强, 尽管仪器软件已经进行过背景扣除。类胡萝卜素是一种 C—C 键和 C=C 键相互交替的直链型分子结构, 当分子的有序性良好时(即分子链又长又直时), π 电子离域产生大的扩展, 使其具有较好的光学性质、宽的电子吸收带和较大的拉曼活性。非极性溶质在非极性溶剂中, 不存在诱导力、取向力、与电偶矩相关的各种静电作用。考虑到类胡萝卜素分子的有序性, 在极性较小的溶剂正己烷中, 有效共轭长度增大, 拉曼散射截面大^[18]。因此, 选择正己烷作为萃取溶剂。

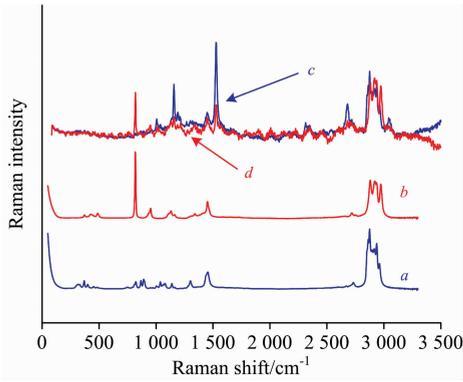


图 2 正己烷 *a* 和异丙醇 *b* 的标准溶液的拉曼光谱; 正己烷和异丙醇分别作为萃取溶剂的烟草样本提取液 *c* 和 *d* 的拉曼光谱

Fig. 2 Raman spectra of hexane *a*, isopropanol *b*, tobacco sample extracted by hexane *c* and tobacco sample extracted by isopropanol *d*

2.1.3 光学平台到焦平面间距离的选择

为了避免样品测量时溶剂挥发, 同时方便测量, 样品的萃取液被装入玻璃瓶密封后进行拉曼光谱测量。因此本工作考察了光学平台到焦平面的距离。如图 3(a)所示, 玻璃瓶被

置于光学平台上, 当不断调节光学平台与焦平面间距离的时候, 拉曼强度也在不断变化。如图 3(a)中红、蓝、绿三条线所示, 它们分别代表三种典型情况的焦平面。在图 3(b)中的上面一条拉曼光谱对应的是图 3(a)中蓝线所代表的焦平面, 明显看出, 由于玻璃瓶的影响, 拉曼效应并不完美。当焦平面运动到图 3(a)中的红线的位置时, 所得到的拉曼信号较好, 说明拉曼效应较完美。而当焦平面运动到图 3(a)中绿线的位置时, 拉曼强度有所下降, 表明随着焦平面离检测器位置的不断变远, 所接收到的拉曼信号逐渐降低。因此本工作系统地考察焦平面到光学平台间的距离, 如图 3(c)所示, 将距离调整在 600~800 μm 之间能产生较好的拉曼信号。

2.2 拉曼光谱的归一化

建立多元校正定量模型的目的是为了今后的应用, 不同时间测定的光谱有可能因为各种物理因素(如仪器操作条件等)的不同而发生变化。图 4 显示的是连续 5 d 所监测的特定样品的拉曼特征峰强度的变化情况, 主要考察叶黄素和 β-胡萝卜素在 1 159 和 1 529 cm⁻¹ 这两个特征峰处的拉曼强度的变化情况, 通过计算 5 d 内 1 159 和 1 529 cm⁻¹ 的相对标准偏差分别为 6.30% 和 7.83%, 均大于 5%, 因此可认为由激光强度和仪器采集效率变化造成的信号强度偏差在 5 d 内不可忽略。因此, 为了消除系统误差, 归一化方法显得尤为重要。

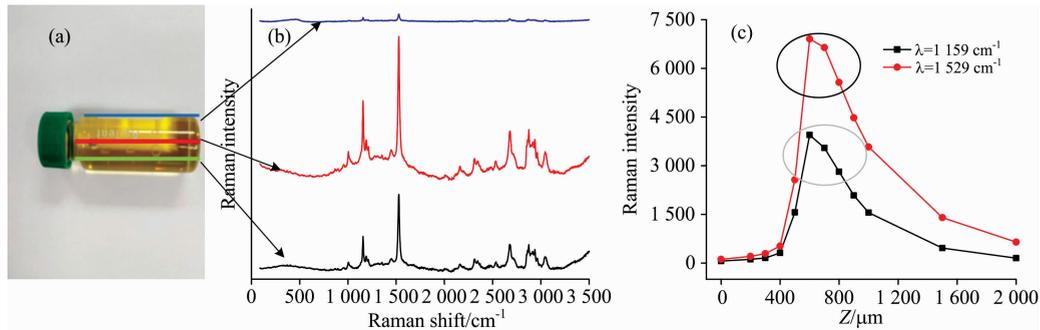


图 3 置于玻璃瓶中的烟草萃取液 (a), 典型焦平面的拉曼光谱图 (b) 和系统地考察焦平面与光学平台间不同距离所对应的特征峰的拉曼强度 (c)

Fig. 3 Extracted solution of tobacco sample in the glass bottle (a), Raman spectra of its typical focal plane (b) and Raman intensity with the distance between the focal plane and the optical platform was investigated systematically (c)

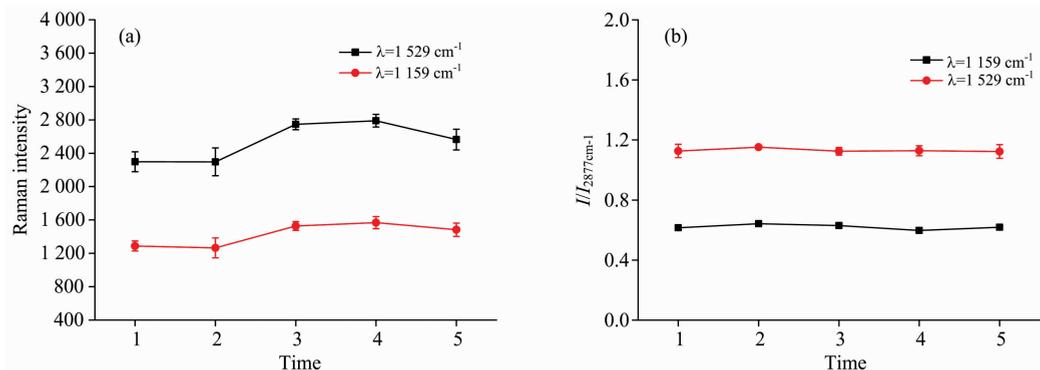


图 4 特定样品重复 5 d 在 1 159 和 1 529 cm⁻¹ 处的拉曼强度归一化前 (a) 和归一化后 (b) 的变化图

Fig. 4 Raman intensity variation of a specific sample at 1 159 and 1 529 cm⁻¹ repeated for 5 days before (a) and after (b) normalization

本文选取正己烷在 $2\ 877\ \text{cm}^{-1}$ 处的特征峰作为内标峰, 样品所有的拉曼信号强度都除以内标峰来做归一化处理。从图 4(a) 和 (b) 归一化前后强度变化情况来看, 可以看出能够显著减小变化, 并且 RSD 得到了减少, RSD 分别降低为 1.6%

和 2.4% 。表 1 结果显示, 在光谱归一化后, PLS 模型的性能显著提升。所以, 在后续的计算中拉曼光谱均做了归一化处理。

表 1 归一化对叶黄素和 β -胡萝卜素的 PLS 模型性能的影响

Table 1 The effects of normalization on the performance of the PLS models built on lutein and β -carotene

	Normalization	LVs	RMSEC/ $(\mu\text{g} \cdot \text{g}^{-1})$	RMSEP/ $(\mu\text{g} \cdot \text{g}^{-1})$	R_{cali}	R_{test}
Lutein	No	2	7.92	8.02	0.933 7	0.934 2
	Yes	4	6.56	7.08	0.955 0	0.948 4
β -carotene	No	8	1.91	3.30	0.970 9	0.924 6
	Yes	9	1.45	2.84	0.983 3	0.944 6

注: LVs 是指隐变量数

Note: LVs is the number of latent variables

2.3 PLS 对烟草中色素的拉曼光谱进行定量分析

图 5 绘出了 141 个烟草样本归一化处理后的全波段的拉曼光谱图以及叶黄素、 β -胡萝卜素和正己烷的标准拉曼光谱图。 $1\ 529\ \text{cm}^{-1}$ 处的拉曼峰对应的是 $\text{C}=\text{C}$ 键的伸缩振动^[19], 这是多烯链最重要的特征峰之一。另一个强特征峰 $1\ 159\ \text{cm}^{-1}$ 对应的是 $\text{C}-\text{C}$ ^[20] 键的伸缩振动。 $1\ 006\ \text{cm}^{-1}$ 处的特征峰是由于甲基和碳链之间的摇摆振动而产生的。根据图 5 的叶黄素和 β -胡萝卜素的标准拉曼光谱可以观察到, 在 $1\ 529$, $1\ 159$ 和 $1\ 006\ \text{cm}^{-1}$ 处有严重的峰重叠, 由于叶黄素和 β -胡萝卜素的分子结构相似, 所以传统的单变量校正曲线很难进行定量分析。因此本文采用 PLS 的多元校正方法, 因为 PLS 可以利用多个变量来处理光谱重叠问题, 从而建立校正模型。

从图 5 可以观察到, 叶黄素和 β -胡萝卜素的拉曼信号主要集中在 $798.2 \sim 1\ 752.8$ 和 $2\ 254.2 \sim 2\ 784.5\ \text{cm}^{-1}$ 这两个波段。在 PLS 建模时, 分别考虑了将这两个波段分别建模和合并起来建模。表 2 和表 3 展示了不同光谱预处理方法的结果。从表 2 可以看出, 当选取 $798.2 \sim 1\ 752.8\ \text{cm}^{-1}$ 这个波段建立校正模型时, 叶黄素的 RMSEP 最小为 $6.68\ \mu\text{g} \cdot \text{g}^{-1}$ 。而从表 3 可以观察到, 对于 β -胡萝卜素, 当选取 $798.2 \sim 1\ 752.8$ 和 $2\ 254.2 \sim 2\ 784.5\ \text{cm}^{-1}$ 这两个波段合并建模时, RMSEP 最小为 $2.56\ \mu\text{g} \cdot \text{g}^{-1}$ 。从图 5 中可以看出, 叶黄素

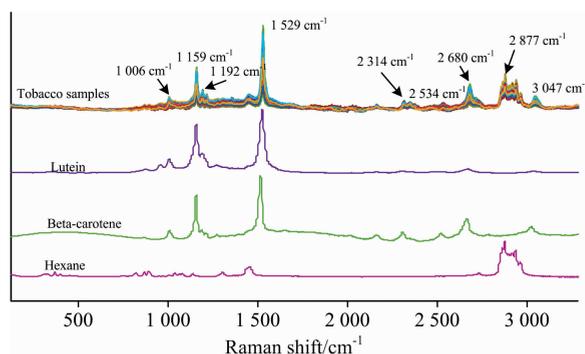


图 5 烟草样本的拉曼光谱图和叶黄素、 β -胡萝卜素以及正己烷的标准拉曼光谱图

Fig. 5 The Raman spectra of tobacco samples and the standard Raman spectra of lutein, β -carotene and hexane for comparisons

的主要特征峰集中在 $798.2 \sim 1\ 752.8\ \text{cm}^{-1}$, 而 β -胡萝卜素则主要集中在 $798.2 \sim 1\ 752.8$ 和 $2\ 254.2 \sim 2\ 784.5\ \text{cm}^{-1}$ 。因此, 叶黄素拉曼建模区域选择 $798.2 \sim 1\ 752.8\ \text{cm}^{-1}$, β -胡萝卜素的拉曼建模区域选择 $798.2 \sim 1\ 752.8$ 和 $2\ 254.2 \sim 2\ 784.5\ \text{cm}^{-1}$ 是合理的。光谱进行求导对模型的预测能力没有改善。

表 2 不同的光谱预处理方法和拉曼建模区域建立的 PLS 模型所获得的叶黄素的预测结果

Table 2 The prediction results for lutein obtained by PLS on different pretreatment methods and Raman shift region

Pretreatment	LVs	RMSEC $/(\mu\text{g} \cdot \text{g}^{-1})$	RMSEP $/(\mu\text{g} \cdot \text{g}^{-1})$	R_{cali}	R_{test}
Raw data	4	6.56	7.08	0.955	0.948
$798.2 \sim 1\ 752.8\ \text{cm}^{-1}$	5	6.31	6.68	0.959	0.954
$2\ 254.2 \sim 2\ 784.5\ \text{cm}^{-1}$	4	6.85	7.03	0.951	0.949
$798.2 \sim 1\ 752.8, 2\ 254.2 \sim 2\ 784.5\ \text{cm}^{-1}$	5	6.10	6.94	0.961	0.951
First derivative	3	7.16	7.03	0.946	0.950
Second derivative	3	7.59	7.58	0.939	0.942
$798.2 \sim 1\ 752.8\ \text{cm}^{-1}$ + First derivative	5	5.86	6.91	0.964	0.951
$798.2 \sim 1\ 752.8\ \text{cm}^{-1}$ + Second derivative	3	7.62	7.56	0.939	0.942

表 3 不同的光谱预处理方法和拉曼建模区域建立的 PLS 模型所获得的 β -胡萝卜素的预测结果Table 3 The prediction results for β -carotene obtained by PLS on different pretreatment methods and Raman shift region

Pretreatment	LVs	RMSEC / $(\mu\text{g} \cdot \text{g}^{-1})$	RMSEP / $(\mu\text{g} \cdot \text{g}^{-1})$	R_{cali}	R_{test}
Raw data	9	1.45	2.84	0.983	0.945
798.2~1 752.8 cm^{-1}	6	2.33	2.96	0.956	0.929
2 254.2~2 784.5 cm^{-1}	4	2.61	3.13	0.945	0.914
798.2~1 752.8, 2 254.2~2 784.5 cm^{-1}	7	1.83	2.56	0.973	0.945
First derivative	8	1.72	3.33	0.976	0.911
Second derivative	2	3.41	4.22	0.904	0.844
798.2~1 752.8, 2 254.2~2 784.5 cm^{-1} +First derivative	5	2.55	3.26	0.947	0.907
798.2~1 752.8, 2 254.2~2 784.5 cm^{-1} +Second derivative	2	3.42	4.27	0.903	0.841

在最优条件下, 即叶黄素选取的 798.2~1 752.8 cm^{-1} 波段, β -胡萝卜素选取的 798.2~1 752.8 和 2 254.2~2 784.5 cm^{-1} 波段, 用校正集建立 PLS 模型, 然后独立预测集用来预测样本。图 6 和图 7 绘出了校正集以及独立预测集的预测浓度和参考浓度关系散点图。从图中很明显地可以观察到, 不同的样本的预测值与参考值都相近, 表明 PLS 模型的建立可以准确可靠地预测烟草中叶黄素和胡萝卜素的浓度。

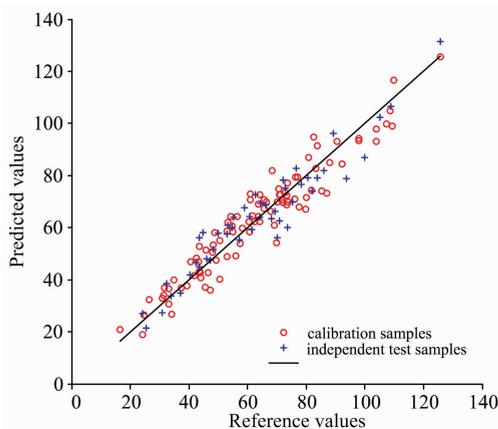
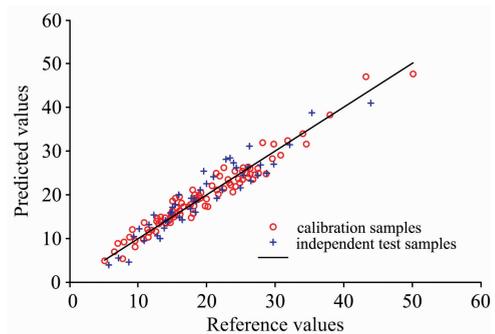


图 6 校正样本(红○)与独立预测样本(蓝十)的叶黄素浓度的预测值与参考值关系图

Fig. 6 Plot of the predicted values versus the reference values of calibration samples (red circle) and independent test samples (blue plus) for the concentration of lutein when the spectra were normalized by the solvent peak

3 结论

在优化了激发光波长, 萃取溶剂以及光学平台到焦平面

图 7 校正样本(红○)与独立预测样本(蓝十)的 β -胡萝卜素浓度的预测值与参考值关系图Fig. 7 Plot of the predicted values versus the reference values of calibration samples (red circle) and independent test samples (blue plus) for the concentration of β -carotene when the spectra were normalized by the solvent peak

的距离后得到了烟草样本的拉曼光谱图, 利用 HPLC 来测定烟草样本中叶黄素和 β -胡萝卜素的浓度的参考值。由于叶黄素和 β -胡萝卜素的主要特征峰发生严重重叠, 所以传统的单点校正曲线不能给予准确可靠的预测结果。因此在拉曼光谱和参考浓度间用 PLS 建立多元校正模型来预测未知烟草样本中的叶黄素和 β -胡萝卜素的浓度。从结果可以看出, 当拉曼光谱以溶剂峰 2 877 cm^{-1} 为内标峰进行归一化处理, 可以获得较小的预测误差, 归结为拉曼光谱仪的不稳定性。利用 798.2~1 752.8 cm^{-1} 建立叶黄素的 PLS 模型, RMSEP 为 6.68 $\mu\text{g} \cdot \text{g}^{-1}$; 利用 798.2~1 752.8 和 2 254.2~2 784.5 cm^{-1} 建立 β -胡萝卜素的 PLS 模型, RMSEP 为 2.56 $\mu\text{g} \cdot \text{g}^{-1}$ 。结果表明, 拉曼光谱结合 PLS 能够准确可靠地预测烟草样本中的叶黄素和 β -胡萝卜素的浓度。

References

- [1] NIU De-xin, LI Yan-hua, HUANG Ming-yue, et al (牛德新, 李艳华, 黄明月, 等). Journal of Agricultural Science and Technology(中国农业科技导报), 2017, 19(5): 28.
- [2] Pinzaru S C, Muller C, Tomsic S, et al. Journal of Raman Spectroscopy, 2015, 46(7): 597.
- [3] Rivera S M, Vilaro F, Zhu C F, et al. Journal of Agricultural and Food Chemistry, 2013, 61(22): 5279.
- [4] Abate-Pella D, Freund D M, Slovin J P, et al. Journal of Chromatography B, 2017, 1067: 34.
- [5] Cheng J, Su X O, Han C, et al. Sensors & Actuators B Chemical, 2018, 255(2): 2329.
- [6] Sebben J A, Espindola J D, Ranzan L, et al. Food Chemistry, 2018, 245: 1224.

- [7] He S X, Xie W Y, Zhang P, et al. *Spectrochimica Acta Part A: Molecular & Biomolecular Spectroscopy*, 2018, 190: 417.
- [8] Shao Y N, Li Y, Jiang L J, et al. *Water Research*, 2016, 104: 432.
- [9] Huang F R, Li Y P, Guo H X, et al. *Journal of Raman Spectroscopy*, 2016, 47(7): 860.
- [10] FANG Hui, JIANG Lin-jun, PAN Jian, et al(方 惠, 蒋林军, 潘 健, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2017, 37(10): 3108.
- [11] Samek O, Jonáš A, Pilát Z, et al. *Sensors*, 2010, 10(9): 8635.
- [12] Kaczor A, Turnau K, Baranska M. *Analyst*, 2011, 136(6): 1109.
- [13] Abbas A, Josefson M, Nylund G M, et al. *Analytica Chimica Acta*, 2012, 737(15): 37.
- [14] Guzmán E, Baeten V, Pierna J A F, et al. *Talanta*, 2012, 93(93): 94.
- [15] Young A T. *Physics Today*, 1982, 35(1): 42.
- [16] Forrest G, Vilcins G. *Journal of Agricultural and Food Chemistry*, 1979, 27(3): 609.
- [17] XU Sheng-nan, LIU Tian-yuan, SUN Mei-jiao, et al(徐胜楠, 刘天元, 孙美娇, 等). *Acta Physica Sinica(物理学报)*, 2014, 63(16): 167801.
- [18] Paraschuk D Y, Kobryanskii V M. *Physical Review Letters*, 2001, 87(20): 207402.
- [19] Fujii R, Chen C H, Mizoguchi T, et al. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 1998, 54(5): 727.
- [20] WU Yong-ling, LIU Tian-yuan, SUN Cheng-lin, et al(吴咏玲, 刘天元, 孙成林, 等). *Acta Physica Sinica(物理学报)*, 2013, 62(3): 037801.

Simultaneous Detection of Lutein and β -Carotene in Tobacco by Using Raman Spectroscopy Combined with Partial Least Squares

GE Jiong¹, SUN Lin², SHEN Xiao-jie¹, SHA Yun-fei¹, HUANG Tian-xiong², DU Yi-ping², ZHANG Wei-bing², XIE Wen-yan^{1*}, YAO He-ming^{1*}

1. Technical Center, Shanghai Tobacco Group Co., Ltd., Shanghai 200082, China

2. Shanghai Key Laboratory of Functional Materials Chemistry, School of Chemistry & Molecular Engineering, East China University of Science and Technology, Shanghai 200237, China

Abstract Pigments are important constituents in tobacco, that relate to appearance and quality of tobacco, meanwhile, the degradation products of pigments have great effects on the quality of aroma of tobacco. According to their color and property, pigments are generally divided into three categories; green pigment, yellow pigment and melanin. In growth, chlorophyll is the main green pigment in tobacco, while xanthophyll and carotene are the main yellow pigments in mature tobacco. Melanin consists in the progress of modulation and fermentation when tobacco leaves are ripe. Analysis of pigments in tobacco is important for evaluations of raw tobacco and the tobacco product quality. Traditional analytical methods of pigments in tobacco based on liquid chromatography, normally need long time and complex sample preparation process, while Raman spectroscopic method is simple to operate and takes short determination time, so that it could be able to provide information about molecular functional groups. Accordingly a method of simultaneous detection of lutein and β -carotene in tobacco was developed with Raman spectroscopy in the present work. A solution of organic solvent extraction from a tobacco sample was enclosed in a glass vial and was detected to collect its Raman spectrum by focusing the laser on the solution inside. The excitation light wavelength was optimized and 455 nm was selected, under which high Raman signals were obtained. Other experiment conditions, including extracting solvent of pigments and the distance between the focal plane and the optical platform were optimized. Because the spectra were measured in different dates and the measurement conditions could change, normalization was used to correct physical interference due to the changes with a selected internal standard peak from the solvent spectrum. To solve problem of spectroscopic interference due to severe overlap between Raman signals of lutein and β -carotene, partial least squares (PLS) was utilized to build multivariate calibration models between Raman intensities and concentrations of pigments. The results showed that the normalization process could help to correct the changes of the measurement conditions, derivative calculations cannot improve the models but the models can be improved significantly by selection of wavelength regions. When spectral regions were selected between 798.2 and 1752.8 cm^{-1} for lutein, regions of 798.2~1752.8 and 2254.2~2784.5 cm^{-1} for β -carotene, the built PLS models showed the best performances with the prediction errors of RMSEP being 6.68 and 2.56 $\mu\text{g} \cdot \text{g}^{-1}$, respectively. The results indicated that Raman spectra combined with PLS could supply a new way to the quantitative analysis of lutein and β -carotene in tobacco with advantages of easy operation, short time and reliable prediction.

Keywords Raman spectroscopy; Tobacco; Pigments; PLS

* Corresponding authors

(Received Sep. 20, 2018; accepted Jan. 13, 2019)