

# 香蕉贮藏气体 3D 荧光表征特征选择及早期腐败预警初探

李孟丽, 殷勇\*, 袁云霞, 李欣, 刘雪茹

河南科技大学食品与生物工程学院, 河南 洛阳 471023

**摘要** 为利用 3D 荧光技术实现基于贮藏室气体荧光信息的香蕉腐败早期预警, 选取两批不同贮藏日期的香蕉贮藏室气体进行三维荧光数据采集。首先对荧光数据进行了预处理: 为消除三维荧光仪整体漂移现象, 对得到的三维荧光数据进行去除漂移处理; 利用 matlab 中 eemscat 工具包对瑞利与拉曼散射进行去除处理, 有效消除了瑞利散射和拉曼散射的不利影响; 运用 Savitzky-Golar(SG)进行数据平滑处理, 减少了噪声对荧光信号的干扰。同时, 对三维荧光数据进行初步筛选, 去除了荧光强度接近于 0 的发射波长, 以及利用三阶高斯混合分布对不同激发波长下的发射光谱进行拟合去除了离散性较大的激发波长。然后针对荧光数据的特征表征, 提出了一种基于 Wilks  $\Lambda$  统计量融合间隔偏最小二乘法(iPLS)的荧光数据特征选择策略。具体是: 用 Wilks  $\Lambda$  统计量进行特征激发波长的选取, 初步选出了 5 个特征激发波长; 根据初选的特征激发波长用 iPLS 结合 pH 值及相对电导率进行了特征发射波段的选取, 结果每个特征激发波长下均选出包含 14 个波长的特征发射波段。为进一步减少分析变量个数, 根据选出的特征发射波长, 运用 Wilks  $\Lambda$  统计量再次进行特征激发波长反选, 最终得到了 3 个特征激发波长。考虑到各特征激发波长下对应 14 个特征发射波长, 故可选出 42 个特征发射波长。最后基于香蕉在贮藏中其品质变化具有时变特点, 根据 42 个特征发射波长使用系统聚类分析法(SCA)进行香蕉腐败基准界定, 得到两批香蕉均在贮藏的第 5 天出现品质突变。因此选用第 5 天贮藏室气体荧光信息来表征香蕉的腐败情况。另外, 利用主成分分析(PCA)初步探索了用第 1 主成分实现香蕉早期腐败的预警。结果表明: 文中提出的三维荧光数据特征波长的选择策略是能够有效降低光谱数据的复杂度, 同时给出的早期腐败预警方法是可行的。

**关键词** 香蕉; 腐败预警; 三维荧光; Wilks  $\Lambda$  统计量; 间隔偏最小二乘; 系统聚类分析

**中图分类号**: TS205.9 **文献标识码**: A **DOI**: 10.3964/j.issn.1000-0593(2021)02-0558-07

## 引言

香蕉因其易于食用且含有丰富的钾、镁等矿物质元素而广受欢迎, 成为世界上进出口贸易额最大的水果之一<sup>[1]</sup>, 同时也是我国南方重要的经济作物。但是, 由于香蕉是典型的呼吸跃变型果实, 其经乙烯催熟后即迅速进入黄熟衰老阶段, 货架期一般仅有 3~5 d<sup>[2]</sup>。期间主要变化为果实硬度下降、糖度增加、果皮变黄进而出现褐变, 严重降低品质, 造成极大的经济损失<sup>[3]</sup>。因此, 在香蕉贮藏期间, 对其实现腐败变质的早期预警是十分必要的<sup>[4]</sup>。

目前研究者对香蕉的腐败过程的研究多是通过主观的感官评定或是通过检测可溶性糖含量、硬度、果皮色泽指数、pH 以及相对电导率等理化检测方法<sup>[5]</sup>, 也有研究者通过高

光谱或近红外技术对香蕉进行品质等级的研究<sup>[6]</sup>, 但这些方法实施起来要么繁琐、要么需要随机检测大量的样本(如高光谱检测), 而且因取样的随机性与个体差异性不能真正做到对贮藏库内香蕉整体质量变化的监控。在香蕉贮藏过程中, 香蕉品质的变化会引起贮藏室内空气的浮游微生物含量及气体组分的变化。同时, 考虑到贮藏室内气体因扩散而具有一致性的特点, 因此通过对贮藏室内的气体进行荧光检测, 可更加准确反映香蕉在贮藏过程中的品质变化。而目前通过环境变化对香蕉的腐败进程进行追踪的研究较少, 通过采集具有均匀一致性的贮藏库气体来实现香蕉的整体质量监控具有创新性。

三维荧光光谱技术由于其具有较高灵敏度和多组分选择性, 能够表征丰富的荧光信息, 同时具有光谱指纹特性, 因而逐渐取代二维荧光光谱技术, 在水环境、医药、食品等<sup>[7]</sup>

收稿日期: 2019-12-25, 修订日期: 2020-04-19

基金项目: 国家重点研发计划项目(2017YFC1600802)资助

作者简介: 李孟丽, 1994 年生, 河南科技大学食品与生物工程学院硕士研究生 e-mail: liml2021@126.com

\* 通讯作者 e-mail: yinyong@haust.edu.cn

领域得到迅速发展。因此,本研究拟采用三维荧光光谱技术,以香蕉贮藏期间贮藏室内空气为取样检测对象,来实现香蕉贮藏过程中品质的监测。

在香蕉的贮藏过程中,其品质变化是一个时变过程,因此要实现香蕉腐败的早期预警,最为关键的是准确确定香蕉腐败的基准。因此,论文首先对复杂的三维荧光信息进行特征选择以减少分析变量个数,选出能够表征香蕉贮藏过程中贮藏室气体变化的特征波长;然后根据选出的特征波长尝试确定了香蕉腐败的基准;最后初步探索了香蕉腐败的早期预警实现方法。目前国内也有研究者运用光谱学对果蔬的腐败进程进行研究,但是对贮藏库内果蔬的整体质量监控多集中于对样品进行随机取样检测,缺乏快速有效的检测方法。本文的研究旨在为香蕉贮藏期间的腐败情况提供一种快捷检测方法的同时,探索了香蕉腐败的早期预警方法。

## 1 实验部分

### 1.1 材料

第一批香蕉,品种为广东 1 号,总重量 300 kg,于 2019 年 4 月 8 日购于超市。第二批香蕉,品种为威廉斯,总重量 300 kg,于 2019 年 5 月 6 日购于超市。考虑到贮藏条件仅起到加快或延迟香蕉的品质变化过程,因此为了减少试验时间,两批香蕉均常温贮藏。在香蕉贮藏过程中,每天测量记录贮藏库的温湿度,第一批香蕉在贮藏期间温度波动范围在 20~24 °C,相对湿度波动范围在 90%~93%;第二批香蕉贮藏期间温度波动范围在 20~25 °C,相对湿度波动范围在 89%~94%。两批香蕉贮藏过程中每天均采集三维荧光数据以及 pH、相对电导率数据。同时,为了保证腐败基准确定的可靠性,试验数据不仅仅截止到感官上出现腐败为止(约为第 6 个贮藏日),还应持续到完全腐败,使试验数据能够覆盖贮藏过程中香蕉品质变化的全部,为此贮藏试验均持续了 8 d。

### 1.2 仪器与参数

采用上海雷磁生产的 PHS-3E 型 pH 计与 DDS-307 型电导率仪,广东环凯微生物科技有限公司生产的 HKM-II 型空气浮游微生物采样器,上海卢湘仪离心仪器有限公司生产的 TG16A 台式高速离心机,以及江苏强盛功能化学股份有限公司生产的分析纯氯化钠。

采用美国安捷伦 Cary eclipse 荧光光谱仪采集 3 维荧光光谱。激发波长  $\lambda_{ex}$  范围设置为 200~300 nm,步长为 5 nm,共包含 21 个激发波长,狭缝宽度为 5 nm;发射波长  $\lambda_{em}$  范围设置为 260~800 nm,步长为 2 nm,共包含 271 个发射波长,狭缝宽度为 5 nm,扫描速度为 fast,电压为 900 V,激发光源为氙灯,在 2 mm×2 mm 的石英荧光比色皿中测量。

### 1.3 理化数据采集

香蕉贮藏过程中能表征其腐败的指标很多,鉴于 pH 值测量操作简便,相对电导率又可以表征香蕉贮藏过程中细胞受损伤程度。因此,选取 pH 值以及相对电导率值作为理化表征指标。pH 值测定:试验当天从贮藏库随机选取 3 根香蕉,做 3 个平行样本。将每根香蕉剥皮,取中间段果肉 10 g,

放入匀浆机,加入蒸馏水 90 mL,每次匀浆时间 60 s,然后用 4 层纱布过滤以去除浮沫及滤渣。用 pH 计依次测量每个样本的 pH 值。每个样本测 3 次,以消除仪器误差,以这 3 个香蕉样本的 pH 平均值作为该贮藏日香蕉样品的 pH 值。

相对电导率测定:试验当天从贮藏库随机选取 3 根香蕉,做 3 个平行样本。每根香蕉取约 2.5 cm×2 cm 的香蕉皮 10 片,放于锥形瓶中,加入蒸馏水 100 mL,用保鲜膜封口,置于常温摇床中震荡 30 min 后取出,依次测量煮前电导率  $\alpha_1$ 。将测量好的样本重新盖上薄膜密封,用电炉煮沸 10 min,冷却至室温,测得煮后电导率  $\alpha_2$ 。相对电导率  $\alpha$  用式(1)计算。两批香蕉样品 pH 和相对电导率数据结果如表 1 和表 2 所示。

$$\alpha = \frac{\alpha_1}{\alpha_2} \times 100 \quad (1)$$

表 1 第一批香蕉 pH 值和相对电导率  
Table 1 pH value and relative conductivity  
of the first batch of banana

贮藏天数/d	pH 值	相对电导率
1	5.03	14.83
2	5.13	15.77
3	5.29	14.61
4	5.35	19.72
5	5.50	24.74
6	5.54	28.53
7	5.69	30.10
8	5.68	30.49

表 2 第二批香蕉 pH 值和相对电导率  
Table 2 pH value and relative conductivity  
of the second batch of banana

贮藏天数/d	pH 值	相对电导率
1	4.91	12.18
2	5.12	13.19
3	5.16	13.21
4	5.21	14.64
5	5.42	13.92
6	5.40	13.96
7	5.55	18.17
8	5.53	20.15

### 1.4 三维荧光数据采集

用超纯水配置 0.9% 的生理盐水, pH 值为 6.66。将 5 mL 0.9% 的生理盐水均匀覆盖在玻璃培养皿中。将采样器放置于贮藏库中央位置,放入培养皿,启动采样器。采样量为 1 000 L,采样时间 10 min。采样结束后,将样品收集到离心管中,并用生理盐水定容至 4 mL。考虑到从空气中采集到的水溶性物质为微量级,对样品进行离心浓缩处理。离心机采用 10 000 r·min<sup>-1</sup> 转速离心 1 min,弃去上清液 2 mL。将剩余 2 mL 样品混匀,进行三维荧光检测。荧光仪每次使用前预热 30 min。

## 1.5 荧光数据预处理

将采集到的三维荧光数据经过预处理及特征波长的选择之后, 利用系统聚类分析法(systematic cluster analysis,

SCA)及主成分分析(principal component analysis, PCA)进行香蕉的腐败预警。本文采用的主要方法及流程如图 1 所示。

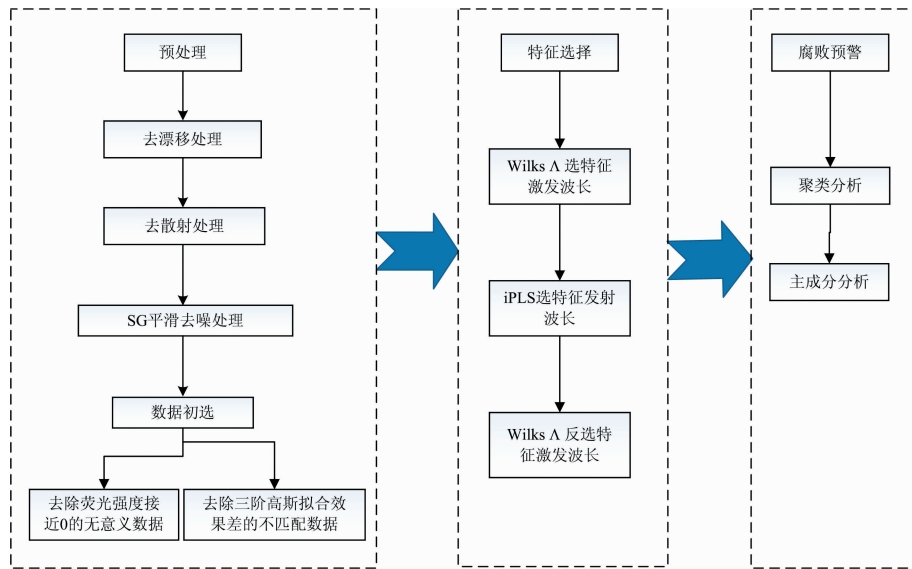


图 1 处理方法流程图

Fig. 1 Processing method flowchart

### 1.5.1 基线漂移去除

由于样品在进行三维荧光检测时其数值会出现漂移现象, 也就是相对于坐标轴轴线荧光数据整体会出现上浮(正值)或者下移(负值)。由于仪器本身设置的最大荧光强度量程为 1 000 a. u., 且由于瑞利散射的强度较大, 容易出现满量程现象(两批香蕉试验所有检测样本的瑞利散射荧光强度均达到满量程), 则可根据瑞利散射的变化情况及满量程来确定漂移校正。具体为: 计算某检测样本荧光强度最大值(瑞利散射值)与 1 000 的差值(分正差值和负差值), 然后将该样本的荧光数据全部加上该差值, 使瑞利散射的荧光强度达到 1 000, 从而达到去漂移的效果。例如某个检测样本的瑞利散射荧光强度值范围为  $-10 \sim 990$ , 便对该样本的整体荧光数据值加上 10, 以达到去漂移的效果。

### 1.5.2 散射去除及平滑处理

荧光漂移校正完成后, 用 Matlab 2014a 中 eemscat 工具箱对瑞利散射及拉曼散射进行去除处理<sup>[8]</sup>。为减少环境及仪器噪声对数据的影响, 对每个激发波长下的发射光谱再进行 SG 卷积平滑处理。

### 1.5.3 荧光波长初选

由于三维荧光光谱数据量较大且存在冗余信息, 因此需要对原始光谱数据进行特征波长的选择, 提取出最能够表达香蕉贮藏期间品质变化的特征波长, 以减少分析变量个数和冗余信息。根据三维荧光数据连续性的特点<sup>[9]</sup>, 使用三阶高斯混合分布对每个激发波长下的发射光谱曲线拟合。拟合效果越差, 表明该波长下的数据离散性、波动性较大, 因此其不能代表被检测样本, 或者说与被检测样本不匹配。

## 1.6 特征选取方法

### 1.6.1 特征激发波长选择

在多元统计分析中, Wilks  $\Lambda$  统计量<sup>[10]</sup>可以对不同样本中各变量的区分能力进行评估。 $\Lambda$  值越大, 表示变量间相似性越大;  $\Lambda$  值越小表示变量间越有差异。由于 Wilks  $\Lambda$  统计量要求输入变量之间无相关性, 因此对原始数据不同激发波长的主成分得分进行  $\Lambda$  值计算。再根据从小到大排序的  $\Lambda$  值来挑选出最合适的特征激发波长。Wilks  $\Lambda$  值的计算表达式如式(2)所示。

$$\Lambda = \frac{|D|}{|T|} \quad (2)$$

式(2)中,  $T$  为三维荧光数据矩阵分析过程中主成分得分矩阵的总离差,  $D$  为矩阵中每列数据的组内离差。将检测样本在不同激发波长下的发射光谱数据划分为一组。 $\Lambda$  值越小, 说明对应的激发波长越能反应香蕉贮藏期间品质的变化。

实际上, 在特征激发波长的选择过程中可分为初选和再选两步, 由上述方法初选激发波长后按下面 1.6.2 给出的方法选择特征发射波长, 然后根据选出的发射波长再用 Wilks  $\Lambda$  统计量反选出更少更具有代表性的特征激发波长。

### 1.6.2 特征发射波段选择

考虑到荧光光谱数据的整体性, 特征波段更能体现样本信息, 因此根据选出的特征激发波长的全发射光谱, 使用间隔偏最小二乘法(interval partial least squares, iPLS)对其进行特征发射波段的选择。该方法通过将全发射光谱分为若干等间隔的子区间, 在每个子区间上运用 PLS 分析, 在不同子区间由交叉验证得到的交叉验证均方根误差(root mean square error of cross validation, RMSECV)来衡量区间回归模型的精度。

## 1.7 基准确定及预警方法

### 1.7.1 腐败基准界定

香蕉在贮藏过程中,随着贮藏过程的进行,相邻两天或几天的相关距离越小,相关性越强,说明这两天的香蕉品质越相似;相邻两天的相关距离越大,相关性越差,说明这两天香蕉的品质变化大,因此根据不同贮藏时间的香蕉品质相似性来实现香蕉腐败基准的确定。为了确定腐败基准,先将每天试验样本自成一类,然后根据基于欧氏距离的系统聚类<sup>[11]</sup>将最接近的两类并成一个新类;通过计算新类与其他各类间的相似性,再选择最接近的两类合成新类,直到所有的样本都并成一类。最后,根据聚类结果可判别腐败基准。

1.7.2 早期预警初探

主成分分析(PCA)<sup>[12]</sup>是一种信息融合方法,也是一种数据分析空间的变换方法。对选出的特征波长进行 PCA 处理,初步探索了利用 PCA 的第 1 主成分的变化来度量香蕉贮藏期间可能出现的腐败接近度,以此实现早期预警。

2 结果与讨论

2.1 三维荧光光谱预处理

2.1.1 散射去除及平滑结果

为去除瑞利及拉曼散射,减少噪声影响,对三维荧光光谱进行去瑞利及 SG 平滑处理,图 2 给出了荧光数据预处理前后情况。由两图对比发现,图 2(b)已经有效的消除了三维荧光光谱数据中瑞利散射及拉曼散射的不利影响,同时图 2(b)比图 2(a)的荧光光谱等高线图区域边界线平滑,说明已经减少了噪声等环境因素的影响,有利于后续数据的分析。

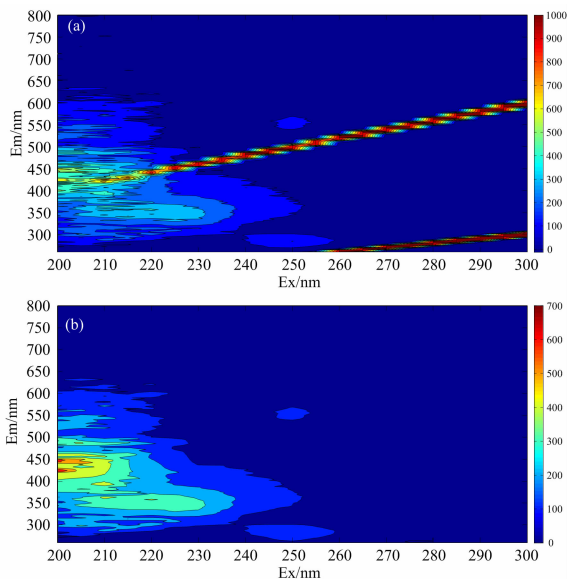


图 2 散射去除及平滑处理前后荧光光谱图

(a): 处理前; (b): 处理后

Fig. 2 Fluorescence spectrum before and after scatter removal and SG smoothing

(a): Before processing; (b): After processing

2.1.2 荧光光谱初选结果

在进行数据分析时发现,样品在发射波长 600~800 nm 的波段基本不产生荧光,其荧光强度接近于 0,因此去除后

100 个发射波长(Em: 600~800 nm)。同时对检测样本 21 个激发波长下的发射光谱进行三阶高斯混合分布拟合,结果如表 3、图 3 所示。

表 3 三阶高斯拟合结果

Table 3 Third-order Gaussian fitting result

Ex/nm	高斯拟合 R <sup>2</sup>	Ex/nm	高斯拟合 R <sup>2</sup>
200	0.81	255	0.98
205	0.68	260	0.92
210	0.74	265	0.90
215	0.81	270	0.93
220	0.78	275	0.96
225	0.90	280	0.97
230	0.95	285	0.97
235	0.96	290	0.97
240	0.94	295	0.97
245	0.98	300	0.96
250	0.98		

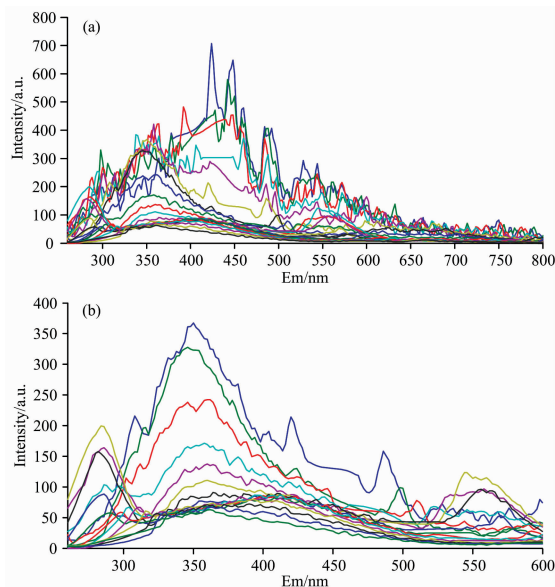


图 3 初步筛选前后的荧光激发波长下的发射光谱

(a): 初选前; (b): 初选后

Fig. 3 Emission spectrum at the fluorescence excitation wavelengths before and after preliminary filtering

(a): Before preliminary filtering; (b): After preliminary filtering

由表 3 可以看出前 5 个激发波长的拟合效果不够理想,表明其离散性较大,数据不够稳定,或许是这些激发波长本身不匹配要检测的对象,因此去除前 5 个激发波长。经过预处理后的荧光数据去除了杂乱信息的干扰和与检测样本不匹配的数据。图 3 给出了预处理前后的荧光数据不同激发波长下的发射光谱图。

2.2 特征选择

2.2.1 特征激发波长选取结果

对两批香蕉预处理后试验数据采取 Wilks  $\Lambda$  值进行特征激发波长选取,  $\Lambda$  值排序后对应的激发波长如表 4。由表 4

可以看出,综合两批次香蕉试验,初步选取 5 个  $\Delta$  值较小且共有的特征激发波长为 225, 230, 235, 240 和 255 nm。

**表 4 两批香蕉试验  $\Delta$  值排序后的激发波长 (Ex)**  
**Table 4 Excitation wavelengths sorted by  $\Delta$  values of two batches of banana**

$\Delta$ 值由低到高排序	第一批香蕉试验 Ex/nm	第二批香蕉试验 Ex/nm
1	225	230
2	240	235
3	255	240
4	230	250
5	235	225
6	260	285
7	280	255
8	270	245
9	250	260
10	300	290
11	285	275
12	290	280
13	295	300
14	245	270
15	265	265
16	275	295

### 2.2.2 特征发射波段选取结果

将初选出的特征激发波长看作不同的变量,结合每批次试验各自的 pH 值、相对电导率值,利用 iPLS 进行特征发射波段的选取。发射光谱在 260~600 nm 间共包括 171 个波长点,发射波长变量编号为 1—171,采用试凑法,将发射光谱划分为 13 个等间距的子区间,交互值为 12。图 4 给出了两批香蕉试验发射光谱特征波段选取情况。横坐标代表发射波长变量编号,纵坐标代表不同子区间的交互验证均方根误差值(RMSECV)。

由图 4(a)可以看出第一批香蕉试验的第 2, 3, 4, 5 和 6 个子区间的 RMSECV 值较低,模型精度较高,说明这几个子区间能够体现不同贮藏日期香蕉的品质变化。由图 4(b)可以看出第二批香蕉试验的第 3 个子区间的 RMSECV 值较低。考虑到选出的特征波长要能够表征品种不同、贮藏日期不同的香蕉品质变化,因此选取两批香蕉试验共有的特征发射波段作为特征区间,即第 3 个子区间: 316~342 nm, 包含 14 个发射波长。因此共选出 5 个激发波长下各 14 个发射波长,共 70 个特征波长。

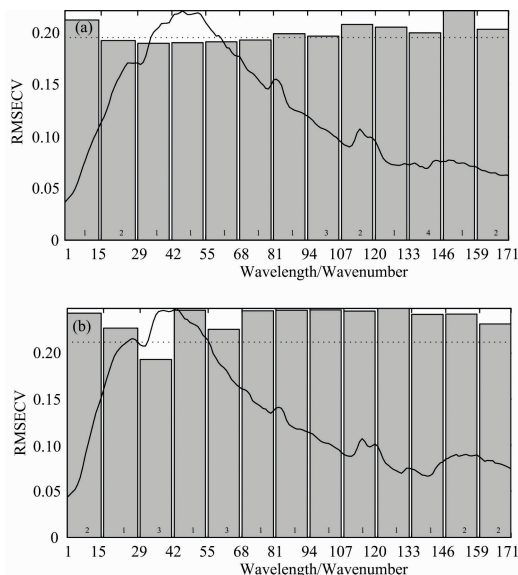
为进一步减少分析变量个数,根据筛选出的 14 个特征发射波长,使用 Wilks  $\Delta$  统计量对荧光数据预处理筛选出的 16 个激发波长再选,结果如表 5 所示。由表 5 可以看出,最后共选出了 3 个  $\Delta$  值较小的特征激发波长: 225, 230 和 235 nm。因此对应此 3 个特征激发波长下各选 14 个特征发射波长,共 42 个特征波长。

## 2.3 确定腐败基准及预警

### 2.3.1 腐败基准确定

对选出的 42 个特征波长的光谱数据进行系统聚类分析,

根据类与类之间的相似性来对两批香蕉进行贮藏期间的腐败基准确定。图 5 给出了两批香蕉试验聚类分析的结果。



**图 4 两批香蕉发射光谱特征发射波段**

(a): 第一批香蕉; (b): 第二批香蕉

**Fig. 4 Emission spectrum feature bands of the two batches of banana**

(a): The first batch; (b): The second batch

**表 5 两批香蕉试验 Wilks 再选  $\Delta$  值排序后的激发波长 (Ex)**  
**Table 5 Excitation wavelengths sorted by  $\Delta$  values of reselected Wilks of two batches of banana**

$\Delta$ 值由低到高排序	第一批香蕉试验 Ex/nm	第二批香蕉试验 Ex/nm
1	225	225
2	235	235
3	240	230
4	230	250
5	260	245
6	245	275
7	280	255
8	265	260
9	300	265
10	275	270
11	250	285
12	285	295
13	290	290
14	270	300
15	255	280
16	295	240

由图 5 可以看出,两批香蕉虽然品种不同,但系统聚类分析可以实现对其贮藏进程中的品质分类,从而确定腐败基准。由图 5(a)可以看出,根据选出的特征波长信息,第一批香蕉在贮藏的第 1, 2, 3, 4 天信息相似性高,因此将其归为一类,说明前 4 天香蕉的品质相近,而第 5 天没有与前 4 天归成一类,说明香蕉在第 5 天品质发生变化。因此第一批贮



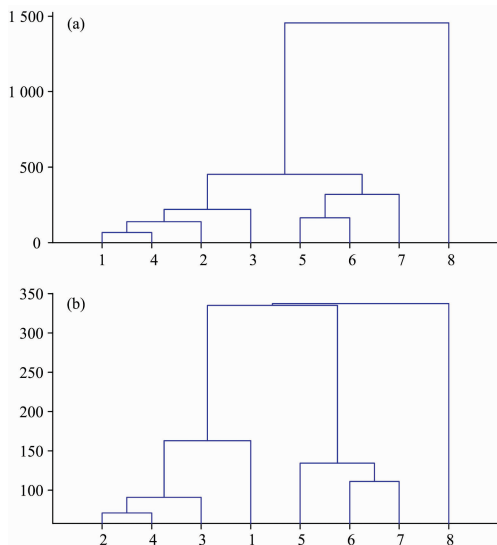


图 5 两批香蕉聚类分析结果

(a): 第一批香蕉; (b): 第二批香蕉

Fig. 5 Cluster analysis results of the two batches of banana

(a): The first batch; (b): The second batch

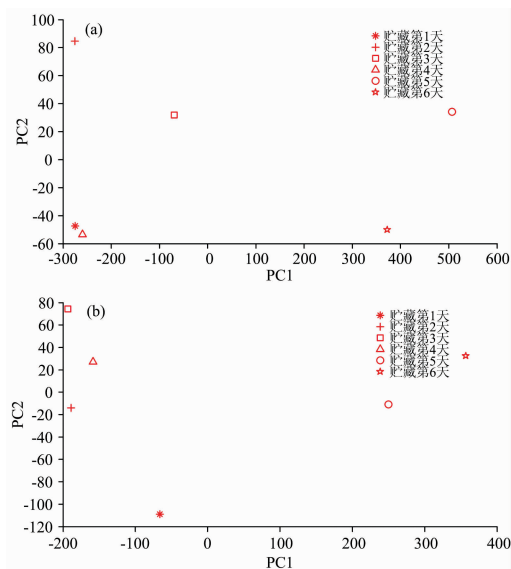


图 6 两批香蕉 PCA 结果

(a): 第一批香蕉; (b): 第二批香蕉

Fig. 6 PCA results of the two batches of banana

(a): The first batch; (b): The second batch

藏试验选用第 5 天的贮藏库气体荧光信息来表征香蕉腐败基准信息。由图 5(b)可以看出, 由于香蕉品种的不同, 第二批贮藏试验的类间距离与第一批有差异, 但使用系统聚类分析还是可以实现对贮藏期间香蕉的品质分类, 从而确定香蕉的腐败基准信息。同理, 选用第 5 天的贮藏库气体荧光信息作为第二批贮藏试验的香蕉腐败基准信息。

### 2.3.2 预警实现

对选出的特征波长的光谱数据进行 PCA 处理, 结果如图 6 所示。

由图 6 给出了两批香蕉特征波长 PCA 处理的第 1 主成分与第 2 主成分结果, 通过两批次贮藏试验数据对比可以看出两批香蕉虽然品种不同, 但是存在相同的腐败阈值。由图 6(a)和(b)可以看出腐败前 4 天与腐败后 2 天可以从第 1 主成分明显分开, 因此可以初步确定第 1 主成分的值在 0 作为香蕉腐败的阈值, 达到该值即意味着香蕉出现腐败。由此可以推断即便是不同品种、不同成熟度的香蕉在贮藏进程中都会达到该腐败阈值, 从而确定贮藏库内香蕉的腐败情况, 实现对香蕉的腐败预警。

## 3 结 论

通过使用荧光仪采集两批贮藏日期不同、品种不同的香蕉在贮藏期间贮藏室空气的 3D 荧光信息, 在对原始荧光数据进行去漂移, 散射去除及 SG 平滑处理后, 实现了荧光数据的初步筛选。然后, 根据 Wilks  $\Delta$  统计量对激发波长进行初选和反选, 并结合 iPLS 对荧光数据进行了特征发射波段选取, 最终选出 3 个不同的特征激发波长以及它们分别对应的 14 个特征发射波长。在此基础上, 利用聚类分析确定了两批香蕉的腐败基准, 并初步探索了 PCA 用于香蕉腐败的早期预警方法。结果表明, 在运用给出的特征波长选择方法所选出的特征激发波长和特征发射波长的基础上, 利用系统聚类分析法能够实现香蕉贮藏过程中腐败基准的确定, 并通过 PCA 分析基本实现了香蕉贮藏过程中产生腐败的早期预警。因此, 研究提出的激发特征波长和发射特征波长的选择方法是有效的, 所给出的早期预警方法也是可行的。另外, 以贮藏库中气体的均匀一致性取样取代香蕉个体的随机性取样, 可克服香蕉样本自身之间差异的影响, 使研究结果更加可靠、准确。因此, 该研究方法及其结果具有普遍的指导价值。

## References

- [ 1 ] Escalante-Minakata P, Ibarra-Junquera V, de Jesús Ornelas-Paz J, et al. 3 Biotech, 2017, 8(1): 30.
- [ 2 ] JIA Cai-hong, JIN Zhi-qiang, LIU Ju-hua, et al(贾彩红, 金志强, 刘菊华, 等). Chinese Journal of Tropical Crops(热带作物学报), 2012, 33(1): 189.
- [ 3 ] ZHANG Jiang-zhong, WANG Bin, LIU Lin, et al(张江周, 王斌, 刘林, 等). Chinese Journal of Tropical Crops(热带作物学报), 2018, 39(11): 2272.
- [ 4 ] Ma B, Han Y M, Cui S Y, et al. Food Control, 2020, 108: 106824.

- [ 5 ] HAI Jin-ping, TONG Han-qing, LUO Xin-ling(海金萍, 童汉清, 罗新玲). Food Science and Technology(食品科技), 2017, 42(8): 54.
- [ 6 ] LI Yuan-hong, WANG Hong-jun, DENG Jian-meng, et al(黎源鸿, 王红军, 邓建猛, 等). Modern Food Science and Technology(现代食品科技), 2017, 33(10): 268.
- [ 7 ] WU Hai-long, LI Yong, KANG Chao, et al(吴海龙, 李 勇, 康 超, 等). Chinese Journal of Analytical Chemistry(分析化学), 2015, 43(11): 1629.
- [ 8 ] Xu Jing, Liu Xiao-fei, Wang Yu-tian. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. 2016, 165: 1.
- [ 9 ] SUN Jun, ZHOU Xin, MAO Han-ping, et al(孙 俊, 周 鑫, 毛罕平, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2016, 32(19): 302.
- [10] Yin Y, Chu B, Yu H C, et al. Journal of Food Measurement and Characterization, 2014, 8(1): 29.
- [11] ZHANG Xiao-yun, WANG Fei, LI Guo-sheng, et al(张晓云, 王 飞, 李国省, 等). Chinese Journal of Analytical Chemistry(分析化学), 2019, 47(1): 99.
- [12] Jafarzadegan M, Safi-Esfahani F, Beheshti Z. Expert Systems With Applications, 2019, 137: 1.

## Feature Selection of 3D Fluorescence Data Based on Storage Room Gas and Preliminary Early Warning of Banana Spoilage

LI Meng-li, YIN Yong\*, YUAN Yun-xia, LI Xin, LIU Xue-ru

College of Food and Bioengineering, Henan University of Science and Technology, Luoyang 471023, China

**Abstract** In order to use three-dimensional (3D) fluorescence technology to realize early warning of banana spoilage based on storage room gas, the storage room gas corresponding to two batches of banana with different storage dates were tested to collect 3D fluorescence data. Firstly, the 3D fluorescence data was pre-processed; to eliminate the overall drift of the scanning data of the 3D fluorescence instrument, the drift of the obtained 3D fluorescence data was processed; the removal and interpolation of Rayleigh and Raman scattering were handled by using the eemscat toolbox in matlab platform, which effectively eliminated the adverse effects of Rayleigh scattering and Raman scattering; and the Savitzky-Golar (SG) method was employed for data smoothing to reduce the influence of noise on the fluorescence signal. Meanwhile, the 3D fluorescence data were preliminarily screened, the emission wavelengths with fluorescence intensity close to 0 were removed, and the more discrete excitation wavelengths were removed by using a third-order Gaussian mixture distribution to fit the emission spectra at different excitation wavelengths. Then, aiming at the feature selection of 3D fluorescence data, a feature wavelength selection strategy based on Wilks  $\Lambda$  statistic combined with interval partial least squares (iPLS) was proposed. The specific steps are: step 1, using Wilks  $\Lambda$  statistics to select feature excitation wavelengths, and five feature excitation wavelengths were preliminarily selected; step 2, based on the initially selected feature excitation wavelengths, the iPLS method was used to select the feature emission bands in combination with pH and relative conductivity, and feature emission band including 14 wavelengths was selected at each feature excitation wavelength; step 3, in order to further reduce the number of analysis variables, according to the selected feature emission band, Wilks  $\Lambda$  statistics was used again to select the feature excitation wavelengths inversely, and 3 feature excitation wavelengths were finally obtained. Combined with 14 emission wavelengths at each feature excitation wavelength, a total of 42 feature emission wavelengths were selected. Finally, considering the time-varying characteristic of banana quality during storage, with the help of the 42 feature emission wavelengths, systematic cluster analysis (SCA) was employed to define the benchmark for banana spoilage, and the cluster results showed that both batches of bananas had abrupt changes in quality on the 5th day of storage. Therefore, the fluorescence information of the storage room gas on the 5th day was used to characterize the banana spoilage. In addition, these feature wavelength variables were computed by principal component analysis (PCA), and the first principal component was preliminary explore to realize early warning of banana spoilage. The research results show that the selection strategy of feature wavelengths of the 3D fluorescence data proposed in this paper can effectively reduce the complexity of the spectral data so as to facilitate subsequent analysis and the early warning method of banana spoilage is also feasible.

**Keywords** Banana; Spoilage warning; Three-dimensional fluorescence; Wilks  $\Lambda$  statistic; Interval partial least squares; Systematic cluster analysis

\* Corresponding author

(Received Dec. 25, 2019; accepted Apr. 19, 2020)