

# 香蕉贮藏中腐败基准确定与高光谱信息表征及腐败预警模型构建

薛书凝, 殷勇\*, 于慧春, 袁云霞, 马帅帅

河南科技大学食品与生物工程学院, 河南 洛阳 471023

**摘要** 为了实现香蕉在贮藏过程中的腐败预警, 对不同贮藏时间的香蕉进行高光谱数据采集。应用 Savitzky-Golar(SG)平滑对原始光谱进行降噪处理, 以获得少噪声干扰的光谱数据。提出一种基于 Wilks  $\Lambda$  统计量结合主成分信息融合的高光谱特征波长提取方法。该方法可描述为: 先对降噪后的高光谱数据进行主成分分析(PCA), 然后对获取的主成分数据分别构建 Wilks  $\Lambda$  统计量, 进而可选出 Wilks  $\Lambda$  值最小的主成分变量(第3主成分, PC3), 绘制出该主成分对应于原始变量(各波长)下的组合权重系数图, 将权重系数图中的波峰、波谷所对应的波长变量提取出来即为该法提取出的特征波长。按此方法共提取出9个特征波长。同时, 对香蕉的色差数据进行分析, 通过分析测试样本的  $L^*$ ,  $a^*$ ,  $b^*$ ,  $\Delta E$  值随贮藏时间的变化趋势, 找到数据异常点(拐点)所在位置, 并结合实际感官情况初步确定了腐败基准为第6个贮藏日。为了进一步说明所确定的腐败基准的合理性, 应用特征光谱数据做出不同贮藏时间样本平均光谱反射值曲线, 对比发现在不同特征波长下光谱反射值均在贮藏第6天达到最小值, 与色差指标分析结果一致, 进一步证明了腐败基准选取的正确性。所以, 可用第6个贮藏日的特征光谱信息来表征基准信息, 以生成腐败基准的特征光谱表征向量及其对应的协方差矩阵。最后, 以表征腐败基准的光谱特征向量为基点, 采用马氏距离(MD)构建了香蕉贮藏过程中基于高光谱特征波长信息的腐败预警模型, 并对模型进行了验证。结果表明: 随着香蕉贮藏时间的延长, 该预警模型给出的测试样本与腐败基准的距离越来越近, 吻合香蕉腐败的实际进程。因此, 所提出的高光谱特征波长提取方法、腐败基准的确定及其表征方法以及腐败预警高光谱模型是合适的, 能有效地预警香蕉腐败。

**关键词** 香蕉; 色差; 高光谱; Wilks  $\Lambda$  统计量; 马氏距离; 预警模型

**中图分类号:** TS205.9 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)12-3871-07

## 引言

香蕉是重要的水果作物, 由于香蕉口味极佳, 富含维生素等各种营养物质, 具有抗氧化性, 以及对预防疾病和维持人体健康有巨大的益处, 因此在人类饮食中占据特殊地位<sup>[1]</sup>。而香蕉作为呼吸跃变型水果, 贮藏过程中腐败现象相当严重, 造成相当大的浪费。所以, 在香蕉贮藏过程中建立一个行之有效的腐败预警模型对降低香蕉在贮藏中的浪费具有非常重要的现实意义。

高光谱成像技术以其无损、快速、分辨率高而著称, 不仅能够获取研究对象内部成分与结构的光谱信息, 还能获取其外部形状、表面缺陷的图像信息。现在, 高光谱技术已经在精准农业<sup>[2]</sup>、医学诊断<sup>[3]</sup>、军事防御<sup>[4]</sup>、农产品品质评

价<sup>[5]</sup>及食品质量控制<sup>[6]</sup>等方面得到广泛使用。但是, 香蕉的高光谱研究主要集中在冷害、成熟度及腐败进程研究, 在腐败预警方面研究尚少, 并没有建立一个有效的预警体系和腐败标准以实时监控腐败进程。因此, 拟通过对不同贮藏期的香蕉样本进行高光谱信息采集, 获取样本的光谱信息, 在有效确定腐败基准光谱信息表征向量的前提下尝试运用马氏距离构建香蕉贮藏期间可有效实现腐败预警的判别模型, 以便为降低香蕉贮藏过程中因腐败而造成经济损失提供一种预警手段。

## 1 实验部分

### 1.1 试验材料及要求

新鲜香蕉品种为广东香蕉1号, 于2019年4月8日购于

收稿日期: 2019-11-07, 修订日期: 2020-03-22

基金项目: 国家重点研发计划项目(2017YFC1600802)资助

作者简介: 薛书凝, 1996年生, 河南科技大学食品与生物工程学院硕士研究生 e-mail: 1653456810@qq.com

\* 通讯联系人 e-mail: yinyong@haust.edu.cn

当地超市,总重量为 600 斤。考虑到贮藏条件仅起到加快或延迟香蕉的品质变化过程,而不影响某一品质等级下的样本光谱信息。因此,为减少贮藏时间和试验时间,将新鲜香蕉置于贮藏库中常温存放,并对贮藏库的温湿度进行实时监控,温度变化范围为 20.07~24.07 °C、相对湿度变化范围为 90.36%~93.55%,每天采集香蕉的光谱信息和色差指标。其中随机选取 10 根香蕉进行光谱信息采集,而色差信息是对固定的 6 根香蕉进行采集的。根据实际贮藏过程中的感官情况,约在贮藏的第 6 天开始腐败,但为了确保腐败基准界定时分析数据的完整覆盖性,共采集了 8 d 的贮藏数据。

### 1.2 高光谱信息采集

高光谱图像采集系统由高光谱成像仪(IST50-3810 型,德国 Inno-Spec 公司,光谱分辨率为 2.8 nm)、4 个 500 W 的光纤卤素灯(RK90000420108 型,德国 Esylux 公司)、传送装置以及控制机(自制)构成。光谱采样范围为 371.05~1 023.82 nm,采样间隔为 0.51 nm,共采集 1 288 个波段的光谱信息。

每天从贮藏库中随机选取 10 根香蕉,从每根香蕉的中部切取薄厚均匀的五片香蕉样本,共计 50 个样本。每次试验将一个样本平铺于洁净的培养皿中,将培养皿放置于传送带上进行高光谱数据采集。采集的参数设置为:传送带传送速度为 1.20 mm·s<sup>-1</sup>,CCD 相机的曝光时间为 90 ms,物距为 350 mm。

为了更好地避免或减少相机暗电流和光源亮度分布不均匀对试验结果产生的影响,在试验前先对标准白板和全黑图像进行采集,后对采集到的高光谱数据进行黑白板校正<sup>[7]</sup>。黑白校正方法为

$$R = \frac{R_0 - R_b}{R_w - R_b} \quad (1)$$

式(1)中:  $R$  为校正后的高光谱图像;  $R_0$  为高光谱原始图像;  $R_w$  为标准白板标定图像;  $R_b$  为全黑标定图像。

### 1.3 样本色差的测定

能够表征香蕉腐败的指标众多,但色差是香蕉在贮藏过程中反映表皮颜色变化的最直观的指标,能够表征颜色感官差异性,是一种模拟人类眼睛感受颜色变化的指标,但同时以具体数据来反映颜色变化,比人感官评价更准确、具体。所以,采用了色差作为香蕉腐败进程的理化表征指标。

使用 Xrite color i5 型色差计测量香蕉表面颜色变化。色差计在使用前需用标准的白板和黑阱进行校正。颜色测定是在固定的 6 根香蕉的中段位置进行,每个样本均重复 3 次测量,取平均值作为该样本的色差值,并用亮度  $L^*$ (从黑到白)、红度  $a^*$ (从绿到红)、蓝度  $b^*$ (从蓝到黄)三个基本指标评价香蕉颜色<sup>[8]</sup>。根据测定的这三个色差指标可以通过式(2)计算出总色差  $\Delta E^{[9]}$ 。 $\Delta E$  作为反映香蕉表面颜色变化的综合指标,可以用于反映香蕉贮藏过程中总体颜色变化。

$$\Delta E = \sqrt{(L^* - L_0^*)^2 + (a^* - a_0^*)^2 + (b^* - b_0^*)^2} \quad (2)$$

式(2)中:  $\Delta E$  是总色差,  $L_0^*$ ,  $a_0^*$  和  $b_0^*$  分别是初始测得的新鲜香蕉亮度值、红度值、蓝度值,  $L^*$ ,  $a^*$  和  $b^*$  为每次测得的香蕉样本亮度值、红度值、蓝度值。

### 1.4 样本高光谱数据预处理方法

样本的高光谱信息提取主要借助 ENVI Classic 软件。首先选取样本图像中的感兴趣区域(region of interest, ROI)<sup>[10]</sup>。因样本形状不规则,所以根据样本与背景间的光谱差异,将样本从背景中提取出来,如图 1 所示的红色区域即为所分离出来的感兴趣区域,并提取出该区域的平均光谱曲线。由于在扫描初始阶段和末尾阶段存在着较大的噪声影响,可应用 ENVI 软件动态显示各波段图像,分辨出包含信息量较多的波段,除去信噪比较低的 371.05~482.60 和 880.99~1 023.82 nm 两个波段,选取 483.11~880.49 nm 为分析波段(含 781 个波段)。虽除去了信噪比较低的波段,但提取的光谱值仍受电流、光线等外界环境的影响而产生不同程度的噪声干扰,故采用 Savitzky-Golary 卷积平滑法<sup>[11]</sup>(SG 平滑)对获取的光谱进行平滑降噪处理,减少噪声的干扰。



图 1 香蕉样本图像感兴趣区域的提取

Fig. 1 Extraction of interest region from banana sample image

### 1.5 特征波长提取

为了减少冗余信息,简化计算,将降噪后的高光谱数据进行特征波长的选取。为此,提出一种基于 Wilks  $\Lambda$  统计量的权重系数高光谱特征波长提取方法进行特征波长的选取。由于 Wilks  $\Lambda$  统计量要求变量之间应相互独立<sup>[12]</sup>,所以,首先用主成分分析法(principal component analysis, PCA)对降噪后的高光谱数据进行分析,得到能够包含所有波长下光谱信息的主成分变量。对获取的主成分数据构建 Wilks  $\Lambda$  统计量<sup>[13]</sup>,并用公式(3)得到每个主成分数据下的 Wilks  $\Lambda$  值。Wilks  $\Lambda$  值是表征类别间(同一贮藏日的样本归为 1 类)差异性的统计量,其值越小说明类别间的差异性越显著。对各个主成分的 Wilks  $\Lambda$  值进行升序排列,选出 Wilks  $\Lambda$  值最小的主成分。将该主成分对应各波长下的权重系数计算出来,画出权重系数图,从而获取能够表征整体光谱信息的特征波长。值得强调的是,该方法不需要理化指标为依据,仅借助光谱信息进行特征波长提取,使所获得的特征波长不受理化指标的改变而变化,提高了选取结果的鲁棒性。

$$\Lambda = \frac{|D|}{|T|} \quad (3)$$

式(3)中:  $D$  为原始数据经 PCA 后得到的主成分矩阵中每一列的组内离差阵,  $T$  为原始数据经 PCA 后得到的主成分矩阵中每一列的总离差。

### 1.6 样本腐败基准确定方法

#### 1.6.1 基于色差变化的确定方法

香蕉在贮藏期内品质不断地发生着变化,尤其是表皮颜色变化最为明显,在实验过程中仅肉眼感官即可观测到随着贮藏时间的延长,香蕉表皮逐渐由亮黄色过渡为暗黄色,后逐渐发生褐变,出现褐色斑点,最终整体变为黑褐色。所以,根据实际贮藏过程中的感官情况,初步确定了腐败基准点为第 6 个贮藏日,并在第 6 个贮藏日之后再采集两天的贮藏数据,以确保分析数据的完整覆盖性。因此,在样本色差与光谱数据采集试验中,共进行了 8 d 的贮藏试验样本的采集。

根据不同贮藏时间样本的  $L^*$ ,  $a^*$ ,  $b^*$  值数据,通过式(2)计算反映总体颜色变化的总色差  $\Delta E$ 。分析样本  $L^*$ ,  $a^*$ ,  $b^*$ ,  $\Delta E$  值随贮藏时间变化的折线图,根据各指标的变化趋势,找到数据异常点(驻点、拐点等)所在位置,结合感官情况初步确定腐败基准点。

#### 1.6.2 基于高光谱信息的确定方法

为了进一步说明由感官及色差变化所确定的腐败基准的合理性,应用光谱数据作进一步解释。即在选取特征波长的前提下,作出不同贮藏时间下样本平均光谱反射值曲线,并对曲线最低点与色差数据进行对应分析,进一步确定了腐败基准。

### 1.7 预警模型的构建

马氏距离<sup>[14]</sup>(Mahalanobis distance, MD)可被认为是计算两个样本之间相似度的一种距离度量,不仅消除了变量之间的相关性干扰还受量纲的影响。香蕉在贮藏过程中腐败预警模型主要体现在确定腐败基准的前提下,应用马氏距离方法计算不同贮藏时间的样本与腐败基准之间的距离来表达。若距基准的距离值越小则说明香蕉越接近腐败。马氏距离计算公式如式(4)

$$d_i^2 = (x_i - y)S^{-1}(x_i - y)' \quad (4)$$

式(4)中:  $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$  表示第  $i$  个贮藏日特征光谱信息的算术平均值向量;  $y = (y_1, y_2, \dots, y_m)$  表示腐败基准日特征光谱信息表征向量;  $S$  为  $m \times m$  的矩阵,表示腐败基准日特征光谱表征信息的协方差矩阵;  $m$  表示所提取的特征波长的个数;  $d_i^2$  表示第  $i$  个贮藏日特征光谱信息与腐败基准间的马氏距离值。

实际上,在马氏距离模型构建过程中,最为核心的内容是腐败基准向量的构造和与腐败基准对应的协方差矩阵的确定。一旦它们被确定,贮藏过程中某个贮藏日到腐败基准的马氏距离即可算出。

## 2 结果与讨论

### 2.1 光谱预处理

图 2 给出了原始光谱和 SG 平滑处理后的光谱对比图。对比发现,原始光谱曲线受噪声干扰,存在许多毛刺,经过 SG 处理过后的光谱曲线明显平滑了许多,已有效的减少了光谱数据的噪声干扰,为后面预警模型的构建奠定了基础。

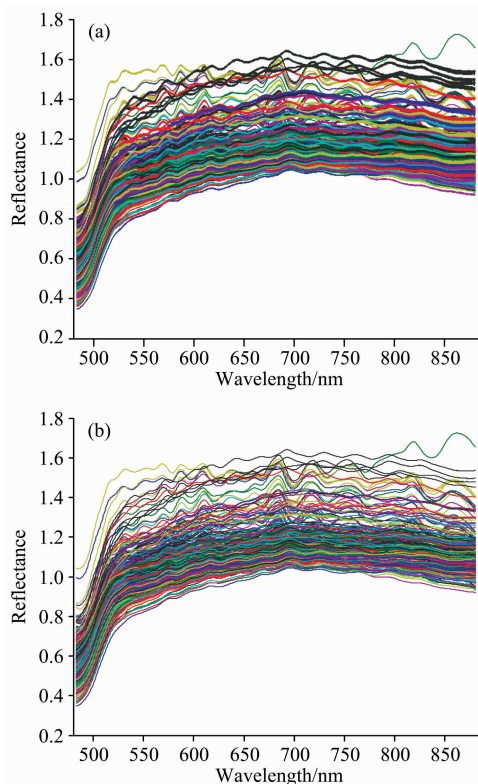


图 2 香蕉光谱图

(a): 香蕉原始光谱图; (b): SG 平滑处理后香蕉光谱图

Fig. 2 The spectra of banana

(a): Original spectra; (b): Spectra after SG smoothing

由图 2(b)可知,在 483.11~559.98 nm 波段下,光谱平滑且变化趋势明显,包含了丰富的样本信息,更能反映样本信息的差异性。而在 560.49~880.49 nm 波段下,光谱变化平缓且有较大波动,说明该波段反映的光谱信息类同,且存在较大的波动误差,不利于后续数据分析。为减少波动影响,仅将 483.11~559.98 nm 波段(151 个波长)用作后续分析。

### 2.2 特征波长提取

将经过 SG 平滑处理后的 483.11~559.98 nm 波段光谱信息经 PCA 处理,得到各个主成分数据后,运用式(3)计算各主成分变量下的 Wilks  $\Delta$  统计量值,选出 Wilks  $\Delta$  值最小

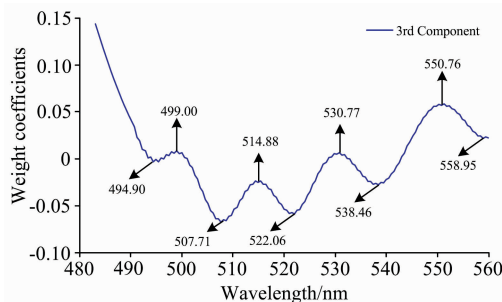


图 3 主成分 3 的权重系数图

Fig. 3 The weight coefficients diagram of the 3rd PC

的主成分变量为第 3 主成分 (the 3rd PC), 并绘出主成分 3 的权重系数图, 如图 3 所示。

图 3 给出了主成分 3 在每个波长 (151 个波长) 下对应的权重系数变化图, 而权重系数的绝对值越大则表明该波长拥有能表征类别间 (与不同贮藏日期的光谱信息对应) 差异性的信息越多, 越有利于判别贮藏过程中香蕉的质量变化情况。将权重系数图中的波峰、波谷所对应的波长提取出来即

为该法提取出的特征波长。共提取出 494.90, 499.00, 507.71, 514.88, 522.06, 530.77, 538.46, 550.76 和 558.95 nm, 9 个特征波长。

### 2.3 样本腐败基准确定

#### 2.3.1 基于色差指标的腐败基准确定

为直观地反映在贮藏期间样本的颜色变化, 图 4 给出样本  $L^*$ ,  $a^*$ ,  $b^*$ ,  $\Delta E$  值随贮藏时间变化的折线图。

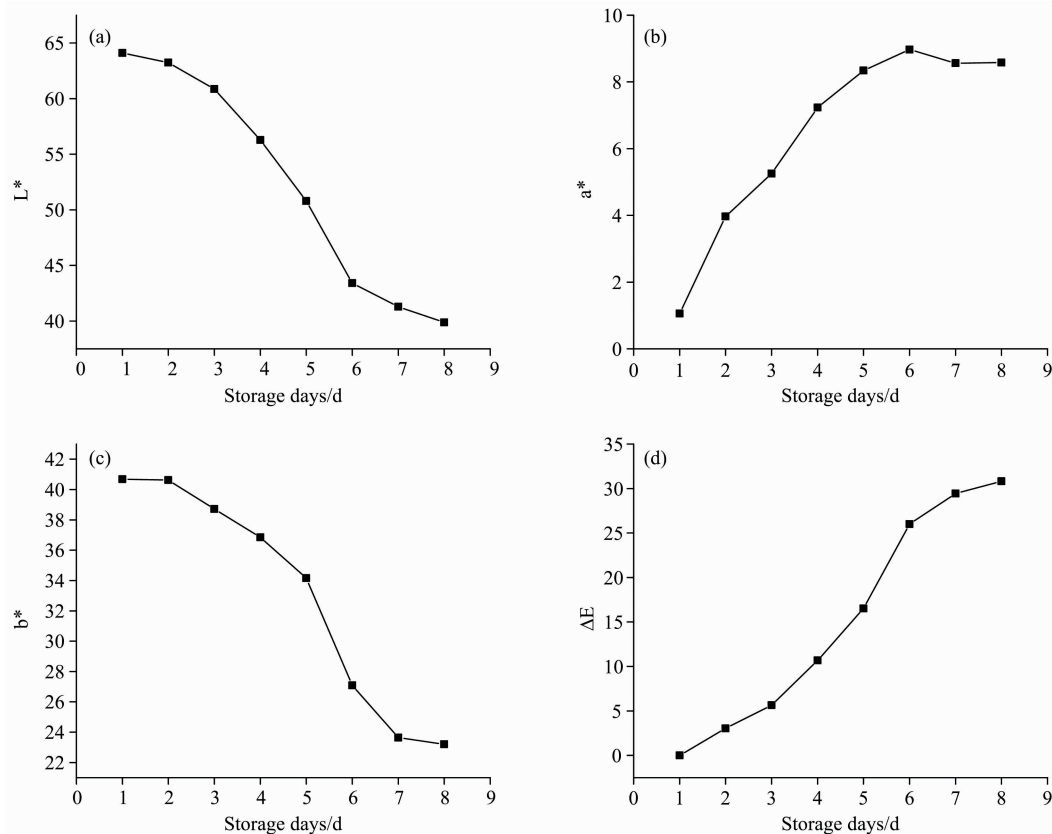


图 4 香蕉色差参数变化曲线

Fig. 4 The variation curve of banana color difference parameters

由图 4(b)可知, 在贮藏期间香蕉的  $a^*$  呈上升趋势, 但在贮藏的第 6 天出现最大值, 贮藏后期  $a^*$  基本稳定在 8.5 左右。由图 4(a)和(c)可以看出在贮藏期间香蕉的  $L^*$  和  $b^*$  呈下降趋势, 且贮藏第 1~5 天是呈上切线下降趋势, 而在贮藏第 6 天及以后出现明显的下切线下降趋势, 颜色变化出现了拐点。由图 4(d)可以看出,  $\Delta E$  在贮藏第 1~5 天是呈下切线趋势上升, 而在贮藏第 6 天及以后出现明显的上切线上升趋势, 说明在贮藏的第 6 天香蕉表皮颜色变化发生了明显的转折。所以, 可初步确定贮藏第 6 天为腐败基准点。

#### 2.3.2 基于高光谱信息的腐败基准确定

分析各个特征波长 (9 个波长) 下不同贮藏时间样本的平均光谱反射值曲线图, 以 494.90 和 550.76 nm 为例进行分析。图 5(a)和(b)分别给出 494.90 和 550.76 nm 波长下不同贮藏时间样本的平均光谱反射值曲线图。

由图 5(a)和(b)可知, 平均光谱反射值在贮藏前期呈整体下降趋势, 但在贮藏第二天有一个明显的突降, 是因为香

蕉在入库前并没有进行挑选, 且在搬运过程中有磕碰, 存在着部分样本损伤现象, 导致光谱值存在波动。而第三天上升是因为香蕉存在着后熟现象, 在前两天香蕉因没有完全成熟, 表皮颜色还可观察到些许绿色, 所以光谱值偏低, 在贮藏第三天因为香蕉完全成熟, 光谱值出现上升, 在贮藏第三天以后, 香蕉开始腐败, 光谱反射值开始依次下降。但对比图 5(a)和(b)发现, 在不同波长下光谱反射值均在贮藏第 6 天达到最小值, 与色差指标分析结果一致, 这就进一步确定了腐败基准点即为第 6 个贮藏日。

### 2.4 样本预警模型构建

由 2.3 确定腐败基准日为第 6 个贮藏日, 共提取 9 个特征波长。由于大样本更能体现统计特征以及反映整体质量属性, 且所表征的结果可靠性高。所以, 采用腐败基准日 50 个样本特征波长下光谱值的平均值作为腐败基准的表征向量。那么基于腐败基准的特征光谱信息表征向量  $\mathbf{y}$  和对应的协方差矩阵  $\mathbf{S}$  可分别给出如下

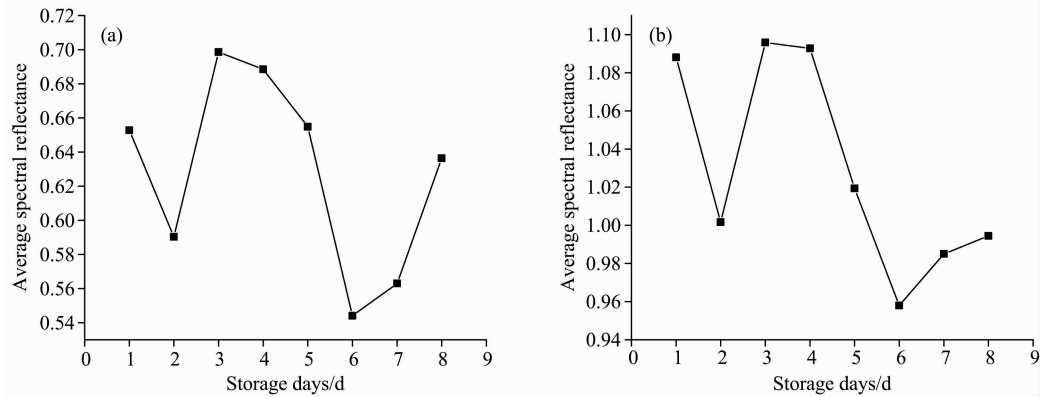


图 5 494.90 nm (a)和 550.76 nm (b)波长下香蕉的平均光谱反射值曲线

Fig. 5 Average spectral reflectance curves of banana at 494.90 nm (a) and 550.76 nm (b) wavelengths

$$y = [-0.8529 \quad -0.8528 \quad -0.7670 \quad -0.7522 \quad -0.6894 \quad -0.6894 \quad -0.6258 \quad -0.6645 \quad -0.6192]$$

$$S = \begin{bmatrix} 0.2799 & 0.2778 & 0.2959 & 0.2969 & 0.3106 & 0.3021 & 0.3117 & 0.3018 & 0.3120 \\ 0.2778 & 0.2770 & 0.2972 & 0.2984 & 0.3117 & 0.3026 & 0.3125 & 0.3025 & 0.3130 \\ 0.2959 & 0.2972 & 0.3243 & 0.3289 & 0.3459 & 0.3377 & 0.3509 & 0.3409 & 0.3536 \\ 0.2969 & 0.2984 & 0.3289 & 0.3393 & 0.3620 & 0.3583 & 0.3740 & 0.3661 & 0.3808 \\ 0.3106 & 0.3117 & 0.3459 & 0.3620 & 0.3920 & 0.3925 & 0.4114 & 0.4057 & 0.4229 \\ 0.3021 & 0.3026 & 0.3377 & 0.3583 & 0.3925 & 0.3983 & 0.4207 & 0.4163 & 0.4347 \\ 0.3117 & 0.3125 & 0.3509 & 0.3740 & 0.4114 & 0.4207 & 0.4503 & 0.4445 & 0.4641 \\ 0.3018 & 0.3025 & 0.3409 & 0.3661 & 0.4057 & 0.4163 & 0.4445 & 0.4412 & 0.4614 \\ 0.3120 & 0.3130 & 0.3536 & 0.3808 & 0.4229 & 0.4347 & 0.4641 & 0.4614 & 0.4829 \end{bmatrix}$$

分别计算不同贮藏日下 50 个样本至腐败基准的马氏距离值，这些值可视为统计意义上的“期望值”。为直观的反映出随贮藏时间的延长样本的腐败进程，图 6 给出了这些期望值的变化曲线图。

图 7(a)和(b)分别给出两次随机选取的计算结果。

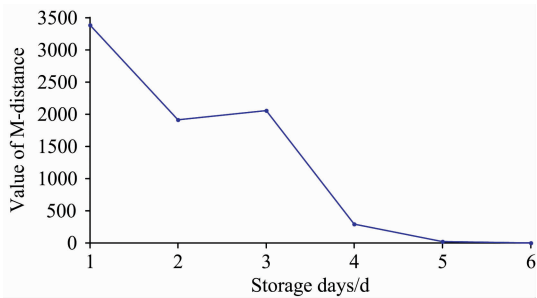


图 6 不同贮藏时间下 50 个样本与的马氏距离折线图

Fig. 6 The Mahalanobis distance curve of the 50 samples and the spoilage benchmark during storage

由图 6 可以看出，随着贮藏时间的延长，样本与腐败基准的距离越来越接近，说明香蕉也越接近腐败。而第 2 个贮藏日至基准的马氏距离有点降低，正是上面所指出的该日样本有损伤的体现，这反而更能充分地说明 MD 模型计算结果的精准性。所以，应用光谱信息构建马氏距离腐败预警模型对香蕉贮藏期间的腐败进行预警是可行的。

考虑到实际应用中一般仅测试 3~5 个平行样本来表征结果，所以，从每天采集的 50 个样本中随机选取 5 个平行样本，计算不同贮藏时间下 5 个平行样本至腐败基准的马氏距

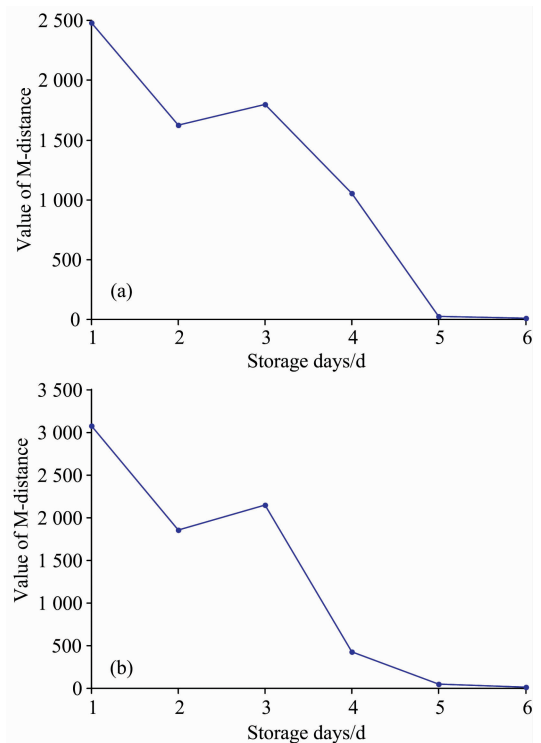


图 7 不同贮藏时间 5 个平行样本的马氏距离折线图

Fig. 7 The Mahalanobis distance curve of 5 parallel samples with different storage time

由图 7 可知,两次随机选取的不同贮藏时间 5 个平行样本至腐败基准的距离变化趋势与“期望值”结果相一致,这充分说明了基准的确定以及对应给出的表征向量和协方差矩阵是合理的与可靠的。上述结果有力地揭示了应用马氏距离对香蕉贮藏过程进行腐败预警是可行的,也是有效的。

### 3 结 论

通过对不同贮藏时间香蕉样本高光谱数据的采集,采用 SG 平滑处理,经基于 Wilk  $\Lambda$  统计量的权重系数法提取出了 9 个特征波长。根据色差变化和高光谱平均光谱值变化确定了腐败基准,构建了基于特征光谱信息的样本与腐败基准的

马氏距离预警模型。结果表明,提出的基于 Wilk  $\Lambda$  统计量的权重系数高光谱特征波长提取方法提高了特征波长提取的鲁棒性,有效除去了冗余信息,简化了运算。通过分析香蕉样本的色差指标随贮藏时间的变化趋势,初步确定了腐败基准为第 6 个贮藏日;通过分析了特征光谱数据下不同贮藏时间样本平均光谱反射值曲线,发现曲线在贮藏第 6 天存在最低点,进一步证明了腐败基准选取的正确性。同时,运用马氏距离构建的香蕉腐败预警判别模型能够较好的反映贮藏期间香蕉接近腐败的程度。因此,给出的腐败基准确定方法及用马氏距离构建的香蕉腐败预警模型对于香蕉贮藏过程中的腐败预警是有效的,为其他样品的腐败预警方法研究提供了参考。

### References

- [ 1 ] Chen J, Li F F, Li Y X, et al. *Postharvest Biology and Technology*, 2019, 158: 110999.
- [ 2 ] Hank T B, Katja B, Heike B, et al. *Surveys in Geophysics*, 2019, 40(3): 515.
- [ 3 ] Thiyaupriyadarsan M R, Suja S. J. *Med. Imaging Health Inf.*, 2019, 9(7): 1354.
- [ 4 ] Shimoni M, Haelterman R, Pemeel C. *IEEE Geoscience and Remote Sensing Magazine*, 2019, 7(2): 101.
- [ 5 ] Xiong Z, Xie A, Sun D W, et al. *Critical Reviews in Food Science & Nutrition*, 2015, 55(9): 1287.
- [ 6 ] Fu D D, Wang Q H, Ma M H, et al. *J. Food Process. Eng.*, 2019, 42(3): e13015.
- [ 7 ] SHI Ji-yong, HU Xue-tao, ZHU Yao-di, et al(石吉勇, 胡雪桃, 朱瑶迪, 等). *Chinese Journal of Food Science(中国食品学报)*, 2018, 18(2): 250.
- [ 8 ] XIAO Wei-qiang, HUANG Bing-zhi, DAI Hong-fen, et al(肖维强, 黄秉智, 戴宏芬, 等). *Food Science(食品科学)*, 2019, 40(7): 254.
- [ 9 ] YIN Hai-jiao, YANG Zhao, CHEN Ai-qiang, et al(尹海蛟, 杨 昭, 陈爱强, 等). *Food Industry Technology(食品工业科技)*, 2012, 33(7): 364.
- [10] SHI Ji-yong, HU Xue-tao, ZHU Yao-di, et al(石吉勇, 胡雪桃, 朱瑶迪, 等). *Chinese Journal of Food Science(中国食品学报)*, 2018, 18(8): 214.
- [11] Yin Y, Hao Y F, Yu H C, et al. *Food Bioprocess Technol.* 2017, 10: 2226.
- [12] Bekker A, Roux J J J, Arashi M. *Journal of Multivariate Analysis*, 2011, 102(3): 619.
- [13] YIN Yong, ZHAO Yu-zhen, YU Hui-chun(殷 勇, 赵玉珍, 于慧春). *Journal of Agricultural Engineering(农业工程学报)*, 2018, 34(15): 290.
- [14] SHI Hong-tao, LI Hong-ping, LIU Jing(时鸿涛, 李洪平, 刘 竞). *Journal of Ocean University of China(中国海洋大学学报·自然科学版)*, 2019, 49(8): 136.

# Determination of Spoilage Benchmark and Its Hyperspectral Information Representation Method as Well as Construction of Hyperspectral Based Spoilage Early Warning Model During Banana Storage

XUE Shu-ning, YIN Yong\* , YU Hui-chun, YUAN Yun-xia, MA Shuai-shuai

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

**Abstract** In order to realize the spoilage warning of bananas during storage, hyperspectral data of bananas with different storage time was collected. The original spectrum was smoothed by Savitzky-Golub (SG) so as to obtain the spectral data with less noise interference. And a method of extracting the feature wavelength from hyperspectral information based on Wilks  $\Lambda$  statistic combined with principal component information fusion is proposed. The method could be described as follows: firstly, principal component analysis (PCA) was applied to the hyperspectral data after SG treatment, then the corresponding Wilks  $\Lambda$  statistics value was calculated for the each obtained principal component, and then the principal component variable with the lowest Wilks  $\Lambda$  value (the third principal component, PC3) could be selected. And the variation curve of combination weight coefficient corresponding to the original variable (each wavelength) of the PC3 was plotted. The wavelengths variable corresponding to the wave peaks and valleys in the weight coefficient curve were picked to be as the feature wavelengths. According to this method, totaling 9 feature wavelengths were extracted. At the same time, the color difference data of the banana was analyzed. By analyzing the trend of the  $L^*$ ,  $a^*$ ,  $b^*$  and  $\Delta E$  values of the test samples with the storage time, the position of the abnormal data point (inflection point) was found. Combined with the actual sensory situation, the spoilage benchmark was initially determined as the sixth storage day. In order to further illustrate the rationality of the given spoilage benchmark, the feature spectral data was used to make the average spectral reflectance curves of different storage time samples. It was found that the spectral reflectance values reached the minimum value at the 6th day of storage for the different feature wavelengths, which was consistent with analysis results based on the color difference index, it was further determined that the spoilage benchmark was indeed the sixth storage day. Thus, the benchmark information could be characterized by the feature spectral data of the sixth storage day, and then the feature spectral representation vector and corresponding covariance matrix of the spoilage benchmark could be generated. Finally, took the hyperspectral feature vector characterizing the spoilage benchmark as a reference point, a spoilage early warning model based on feature wavelength spectrum information during banana storage was established by Mahalanobis distance (MD) and was verified effectively. The results showed that the MD between the test samples and spoilage benchmark given by the early warning model was getting closer and closer with the extension of banana storage time, which was consistent with the actual process of banana spoilage. Therefore, the proposed hyperspectral feature wavelength extraction method, determination method of spoilage benchmark, the representation method and the hyperspectral warning model are suitable and effective for early warning of banana spoilage.

**Keywords** Banana; Color difference; Hyperspectral; Wilks  $\Lambda$  statistic; Mahalanobis distance; Early warning model

(Received Nov. 7, 2019; accepted Mar. 22, 2020)

\* Corresponding author