

## 番茄碰伤和可溶性固形物近红外光谱同时在线检测

刘燕德, 饶宇, 孙旭东, 肖怀春, 姜小刚, 祝柯, 徐海

华东交通大学机电工程学院, 江西 南昌 330013

**摘要** 可溶性固形物和碰伤是影响番茄品质的两个主要因素。研究的目的是探索可见近红外漫透射光谱同时在线检测番茄碰伤和可溶性固形物的可行性。在单通道送果速度5个每秒条件下,采集番茄近红外漫透射光谱。对比分析碰伤与正常番茄样品的近红外漫透射光谱特性,结果表明,碰伤与正常番茄样品的近红外漫透射光谱在光强上存在明显差异,碰伤果光强要强于正常果,其原因可能是碰伤后果肉变软,透光性变强;在650和675 nm处碰伤果比正常果要多两个吸收峰,可能是碰伤后,番茄表皮颜色发生变化所致。选取贡献率占比最多的前三个主成分,对正常果与碰伤果近红外漫透射光谱主成分定性分析,正常果与碰伤果不能有效聚类,故近红外漫透射光谱主成分定性分析效果不明显,需选择建立高维近红外漫透射光谱定性判别模型。故建立了碰伤番茄样品的近红外漫透射光谱偏最小二乘定性判别模型,误判率为0%,能正确判别碰伤果,故选用碰伤番茄样品的近红外漫透射光谱偏最小二乘定性判别模型作为番茄碰伤果在线剔除分选模型。通过对未参与建模的样品进行验证,能正确识别出碰伤果。经近红外漫透射光谱偏最小二乘定性判别模型剔除碰伤果后,按照可溶性固形物指标进行分级。分别使用全部波段和606~850 nm的波段进行建模预处理,且对全部波段和606~850 nm波段光谱进行2阶导数预处理,前后平滑设为9,利用连续投影算法与遗传算法优选可溶性固形物的光谱建模变量,对比发现,利用未经算法筛选过的606~850 nm波段光谱变量进行建模,效果最好,建立了可溶性固形物在线检测模型,预测集均方根误差为0.43 Brix°。采用未参与建模的样品进行碰伤和可溶性固形物同时在线检测验证,碰伤样品的分选准确率达96%,可溶性固形物样品的分选准确率达91%。表明:番茄碰伤和可溶性固形物近红外漫透射光谱同时在线检测是可行的。

**关键词** 漫透射;近红外光谱;在线检测;碰伤;可溶性固形物

**中图分类号:** S24 **文献标识码:** R **DOI:** 10.3964/j.issn.1000-0593(2019)12-3910-06

### 引言

我国的番茄种植、加工和出口都处于持续增长的趋势,是全球最重要的番茄制品生产与出口国;可溶性固形物是番茄中糖、酸、维生素等可溶性物质的总和,是表征番茄品质重要指标之一。而番茄在采摘、运输和存储过程中极易碰伤表皮,造成番茄内部品质劣变,影响净出口<sup>[1-4]</sup>。传统的人工识别表面碰伤和可溶性固形物破损式的检测方法,效率低,难以满足大宗番茄分选需求。可见/近红外光谱技术能够快速、无损的检测水果的内部品质及表面缺陷,故提出一种基于漫透射的快速、无损、大量检测番茄碰伤及可溶性固形物

的方法,具有重要意义。

当前,对番茄品质快速分选的国内外研究报道较少,大多针对的是环境对番茄的影响以及番茄采摘系统、病害等的研究<sup>[5-6]</sup>。番茄为多汁浆果,果肉由果皮及胎座组织构成,番茄果实是由子房发育的真果,果皮是发育的子房壁,由外果皮、中果皮和内果皮组成。目前仍未有同时进行番茄碰伤和可溶性固形物的动态在线分选的研究报道。本文主要提出了综合考虑番茄碰伤与可溶性固形物的两个因素的动态在线检测方案,实验采用弹性垫圈和透光孔组合结构,依靠番茄自身构造和重量实现密封,能较有效减少杂散光的影响,同时对比了不同的分析方法对判别模型的影响。

收稿日期:2018-10-19,修订日期:2019-02-08

基金项目:国家自然科学基金项目(31760344),江西省优势科技创新团队建设计划项目(20153BCB24002),南方山地果园智能化管理技术与装备协同创新中心项目(赣教高字[2014]60号),江西省杰出青年人才资助计划项目(20171BCB23060)资助

作者简介:刘燕德,女,1967年生,华东交通大学机电工程学院教授 e-mail: jxliuyd@163.com

## 1 实验部分

### 1.1 番茄样品准备

实验样品为 320 个江西某果园正常番茄, 将番茄表面擦净, 置于 25 °C 的室温下存放 24 h, 依次采集番茄样品的横纵径、重量、可溶性固形物等物理指标如表 1。采用聚四氟乙烯球撞击标记的四个面, 用以模拟实际生产生活中碰撞挤压造成的碰伤, 得到实验用碰伤果样品。同一番茄样品碰伤前后对比与撞击示意图如图 1 所示。图 1(a)中, 左边为进行碰伤实验前, 右边为进行碰伤实验后, 番茄碰伤前后外观无明显变化; 图 1(b)中  $H=125$  mm,  $\beta$  约为 25°, 小球的质量约为 380 g, 番茄样品放于斜坡下方位置, 聚四氟乙烯球从上方位置沿斜坡滚落, 撞击番茄样品, 忽略斜坡与小球之间的摩擦力, 小球碰撞产生的能量大约为 0.6 J。

共 320 个实验样品, 其中 240 个正常果, 80 个碰伤果。将实验所得数据分为两组, 每组均按照 3:1 比例分为校正集与预测集进行建模, 由此来区分碰伤果对可溶性固形物的影响。组 1 为 240 个正常果与 80 个碰伤果, 其中校正集为 240 个, 预测集为 80 个, 且校正集包括 185 个正常果, 55 个碰伤果, 预测集包括 55 个正常果, 25 个碰伤果。组 2 为 240

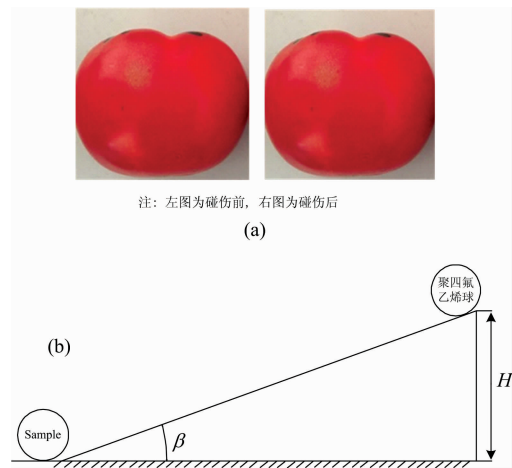


图 1 实验样品及实验

(a): 同一番茄样品碰伤前后对比; (b): 模拟碰伤实验图

Fig. 1 Experiment and samples

(a): Comparison before and after injury;

(b): Bruise experiment

个正常果, 用来建立和预测正常果的可溶性固形物混合模型。可溶性固形物真值与横纵径统计如表 1 所示。

表 1 校正集与预测集番茄样品可溶性固形物真实值与横纵径统计结果

Table 1 Statistical values of soluble solid content and diameter for tomato samples in calibration and prediction set

组别	横径/mm	纵径/mm	数量	重量/g	可溶性固形物范围/%	可溶性固形物平均值/%	可溶性固形物标准差
校正集							
1	64~85	52~74	240	148.1~297.7	3.6~7.7	4.78	0.79
2	69~78	62~70	185	154.6~219.7	4.1~7.4	4.93	0.71
预测集							
1	64~85	52~74	80	148.1~297.7	3.9~7.3	4.85	0.73
2	69~78	52~74	55	151.3~283.4	4.0~7.5	4.88	0.74

### 1.2 番茄漫透射近红外光谱采集

实验采用自主研发的漫透射式动态在线检测装置采集番茄果光谱数据, 该装置主要有光源、光谱仪、果杯等。光源为 6 个 12 V、100 W 的卤钨灯, 环绕分布在样品两侧; 采用 Ocean Optics 公司的 QE65000 光谱仪, 光谱类型为短波近红外光谱, 波长范围为 350~1 150 nm; 果杯内圈设有软塑料遮光圈, 通过果外形与重量等指标, 可以有效抑制杂散光; 动态在线采集光谱时, 果杯带着果随传动链条向前移动, 经过光源照射, 光谱仪采集光谱信息。

### 1.3 番茄碰伤与可溶性固形物含量判定

使用折射式数字糖度仪 (PR-101 $\alpha$ , 日本 ATAGO 公司) 测定番茄样品的可溶性固形物含量, 每次测量重复 3 次, 取平均值作为糖度实验值, 所得值以质量分数表示。碰伤果的判定标准为表面果肉出现异常变软现象。

### 1.4 模型建立及评价

采用 MATLAB2012a 软件导出能量光谱, 利用 BiPLS 工具包、SPA 工具包、GA 工具包进行变量筛选。光谱预处理

方法及 PLS 建模在 Unscrambler8.0 软件中实现。通过比较分析各个模型的校正集及预测集相关系数、校正集均方根误差、预测集均方根误差等多个参数来进行模型评价。若模型的预测集相关系数高, 预测集均方根误差小, 则表明模型的预测能力越好。

## 2 结果与讨论

### 2.1 正常与碰伤番茄近红外漫透射光谱特性对比分析

随机采集同一个实验样品同一位置碰伤前后的光谱信息如图 3 所示, 碰伤前后总谱型变化趋势基本一致, 而在波长为 650 与 675 nm 处, 碰伤样品比正常样品多两条峰, 可能是颜色变化造成吸收峰不同。从强度上来看, 正常果的光谱能量最高为 4 500 光子数左右, 而碰伤果最高达 9 000 光子数左右, 其原因可能是碰伤之后, 番茄外部保护层受到破坏, 果肉组织变软, 透光能力变强。两条对比光谱均在 700 与 750 nm 附近存在波峰, 在 750 nm 附近存在波谷。其有效信

息集中在 606~850 nm 之间,故选用 606~850 nm 的波段范围进行建模。

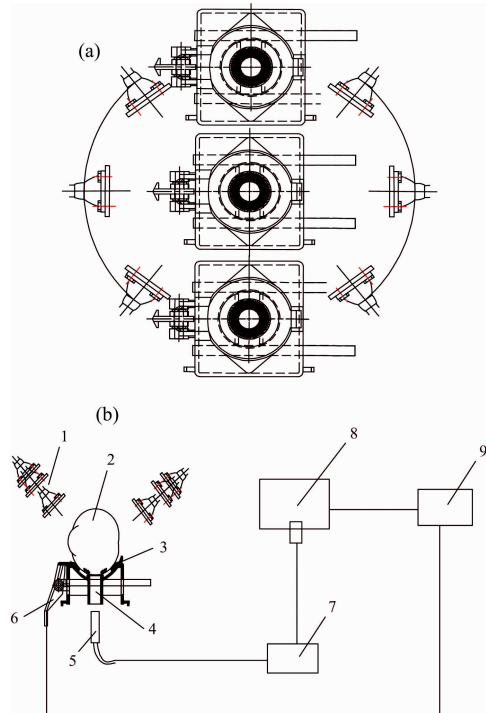


图 2 近红外漫透射在线检测装置

(a): 光源布置图; (b): 番茄在线检测光路图

1: 光源; 2: 番茄样品; 3: 遮光圈; 4: 光路;

5: 光纤; 6: 弹跳爪; 7: 光谱仪; 8: 计算机; 9: PLC

Fig. 2 Device of Near infrared diffuse transmission on-line detection

(a): Arrangement of light source;

(b): Optical path of tomatoes online check

1: Light source; 2: Tomatoes sample; 3: Lens hood; 4: Light path; 5: Lifting lever; 6: Bounce claw; 7: Spectrograph; 8: Computer; 9: PLC

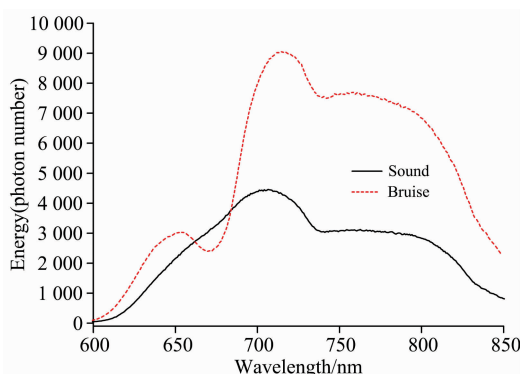


图 3 正常与碰伤果近红外漫透射光谱

Fig. 3 Near-infrared diffuse transmission spectra of perfect and bruised samples

## 2.2 碰伤果近红外漫透射光谱定性判别分析

### 2.2.1 碰伤果近红外漫透射光谱主成分定性分析

主成分分析采用的是全谱分析,将实验所采集到的光谱信息压缩为若干个主成分的线性组合,由式(1)可知, $P$ 为样本差别, $T$ 为光谱变量,主成分 $X$ 为每个样本差别与光谱变量相乘的总和。前三个主成分因子的得分散点图如图 4 所示,正常果和碰伤果的三个 PCA 得分散点存在混在一起的现象。图 5 中,PC1 表示光谱基线移动信息,PC2 和 PC3 描述了碰伤和正常光谱的一些细节信息。结合图 3,碰伤果光谱强度明显强于正常果,碰伤果的光谱基线也远大于正常果,而 PC2 和 PC3 在波长为 650, 675 和 680 nm 处,对应图 3 中碰伤果 650, 675 和 680 nm 的光谱峰,分析原因可能是成熟度颜色造成的差异。光谱曲线权重系数越大,该 PC 积累的贡献率越高。通过主成分分析不能将 320 个样品分为两类,想要改善分类效果,需在更高维空间建立定性判别模型。

$$X = P_1 T_1 + P_2 T_2 + \dots + P_{20} T_{20} \quad (1)$$

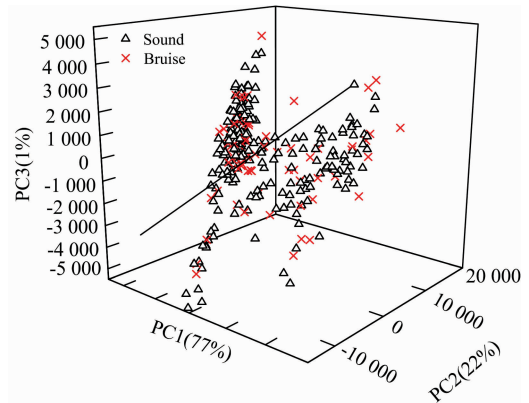


图 4 主成分得分散点图

Fig. 4 Scores plots of principal component

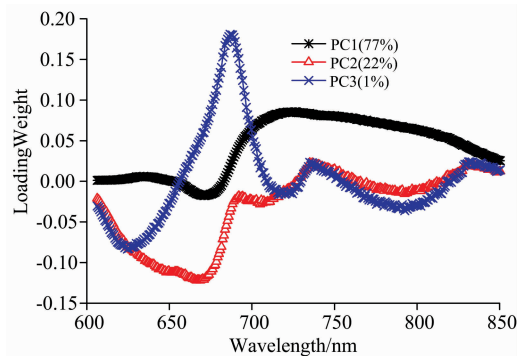


图 5 光谱曲线权重系数

Fig. 5 Loading weight coefficient of spectral curve

### 2.2.2 碰伤果近红外漫透射光谱偏最小二乘判别分析

偏最小二乘判别分析方法是在偏最小二乘法的基础上建立样本分类模型,是一种数学优化技术,通过最小化误差的平方和找到一组数据的最佳函数匹配。该方法需要按照样本的类别特性,赋予样本分类变量值<sup>[5-6]</sup>。选用 606~850 nm 建立偏最小二乘判别模型,其建模结果如图 6,采用 315 个样品进行建模,人为设定正常样品为 2,表面缺陷样品为 8,

阈值  $T=5$ 。模型建模级相关系数  $R_c$  为 0.94，建模标准偏差 0.86；模型预测级相关系数为 0.92，标准偏差为 1.06，误判率为 0%。随着主成分数的增加，预测集的均方根误差逐渐降低，当主成分数为 10 时，预测集均方根误差最小，模型误判率为零，达到最优。回归系数与近红外漫透射光谱曲线对比如图 7，光谱变量在 PLS 定量模型中的贡献率越大，回归系数越大。正回归系数对应的光谱变量越大，碰伤样品的概率越大，反之亦然。光谱变量与回归系数加权求和再加上截距  $b=4.24$ ，得 PLS-DA (partial least squares discrimination analysis) 模型预测的类别值。再通过与阈值的比较，实现碰伤样品的预测。其中，碰伤果的预测公式如式(2)所示；其中  $Y$  为阈值， $Y \leq 5$  即判定为碰伤果，否则判定为正常果。

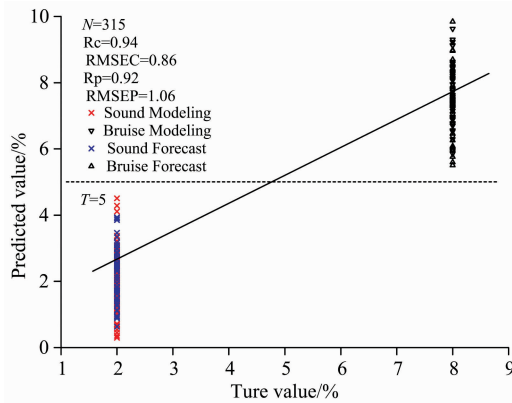


图 6 偏最小二乘判别模型  
Fig. 6 DPLS model

$$Y = \sum_{i=1}^n \beta \gamma_i + b \quad (2)$$

其中， $Y$  为模型的预测阈值； $n$  为参与建模的光谱变量数； $\beta$  为能量谱强度； $\gamma$  为回归系数； $b$  为模型的截距。

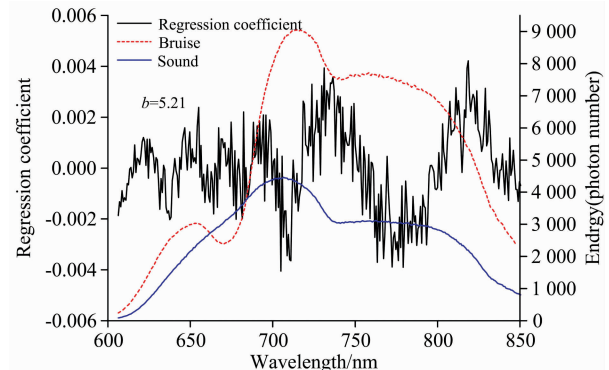


图 7 回归系数与近红外漫透射光谱对比  
Fig. 7 Comparison of regression coefficient and near-infrared diffuse transmission spectrum

### 2.3 偏最小二乘可溶性物模型建立及预测

首先将 315 个样品混合建模，其中包括 239 个正常果，76 个碰伤果。对全部样品分别采用连续投影算法 (successive projections algorithm, SPA) 与遗传算法 (genetic algorithm, GA)，连续投影算法是随机选取光谱矩阵中的某几个变量，最后分别计算对其他变量的投影，根据均方根误差最小的原则来决定变量个数；而遗传算法则是通过模拟自然进化过程，通过交叉验证的方法，寻找最优解<sup>[6-9]</sup>。

表 3 不同组别的模型统计结果

Table 3 Model statistical results of different groups

组别	筛选变量方法	变量数	因子数	样品数	$R^2$	RMSEC/%	$Q^2$	RMSEP/%
1	无	500	13	315	0.77	0.37	0.25	0.75
2	SPA	21	12	315	0.34	0.63	0.27	0.72
3	GA	81	19	315	0.55	0.52	0.25	0.73
4	无	482	7	238	0.69	0.43	0.52	0.53
5	SPA	11	4	238	0.46	0.56	0.56	0.51
6	GA	22	4	238	0.45	0.56	0.48	0.56

组别 1 采用全部样品建模校正集决定系数为 0.77，校正集均方根误差为 0.37%，其中，光谱变量为 606~850 nm 的波长，组别 2 采用连续投影算法 (SPA) 对光谱变量筛选，得到 21 个光谱变量进行建模，校正集决定系数为 0.34，校正集均方根误差为 0.63；组别 3 采用遗传算法 (GA) 对光谱变量筛选，得到 81 个光谱变量进行建模，校正集决定系数为 0.55，校正集均方根误差为 0.52。

组别 4—组别 6 先将全部样品进行预处理，光谱变量前后一致，将光谱变量进行 2 阶导数预处理，前后平滑设为 9。组别 4 采用 238 个正常果进行建模，建模校正集决定系数 0.69，校正集均方根误差为 0.43%；组别 5 采用连续投影算法 (SPA) 对光谱变量筛选，得到 11 个光谱变量进行建模，校正集决定系数为 0.46，校正集均方根误差为 0.56%；组别 6

采用遗传算法 (GA) 对光谱变量筛选，得到 22 个光谱变量进行建模，校正集决定系数为 0.45，预测集均方根误差为 0.55%。由于组别 4 明显优于其他 5 个组别，通过对比可知，碰伤果影响番茄可溶性固形物的模型预测精度，故需先剔除碰伤果再建立番茄可溶性固形物模型，如图 8 所示。通过遗传算法与连续投影算法筛选变量进行建模的方法在本实验中效果不佳。

随着主成分数不断增加，该模型交互验证均方根误差先减后增，当主成分数为 7 时，交互验证均方根误差达到最小。图 9 为经过 2 阶导数预处理的回归系数与三条近红外漫透射光谱对比图。回归系数反映了不同光谱变量在 PLS 模型中的贡献率，光谱变量在 PLS 定量模型中的权重越大，回归系数越大。正回归系数对应的光谱变量越大，样品的可溶性固形

物值越大。PLS 定量模型的截距  $b_2 = 5.03$ 。其可溶性固形物的预测公式如式(3)所示。

$$Y_{SSC} = \sum_{i=1}^n \theta_i \gamma_i + b_2 \quad (3)$$

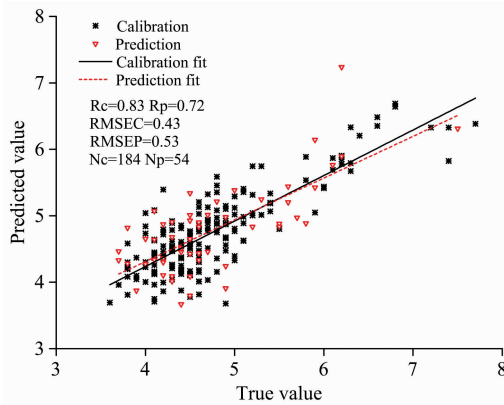


图 8 偏最小二乘回归建模和模型预测散点图

Fig. 8 Scatters PLS calibration and prediction models

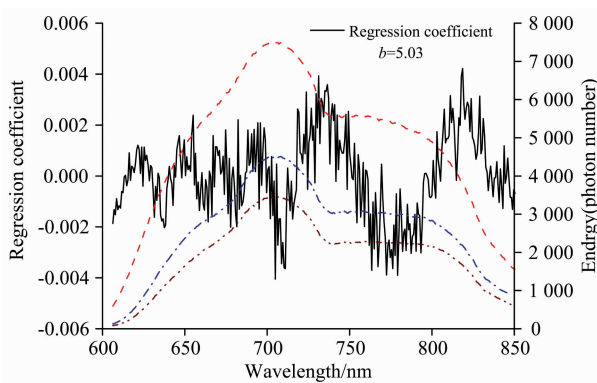


图 9 偏最小二乘模型回归系数与近红外漫透射光谱对比

Fig. 9 Comparison of PLS regression coefficient and NIR diffuse transmission spectrum

其中,  $Y$  为模型的预测可溶性固形物值;  $n$  为参与建模的光谱变量数;  $\theta$  为能量谱强度;  $\gamma$  为回归系数;  $b_2$  为模型的截距。

#### 2.4 在线分选模型验证

首先将建立的碰伤果偏最小二乘判别模型与可溶性固形物偏最小二乘模型加载到自主研发的在线检测软件中, 其基本参数为模型的回归系数和截距<sup>[8-10]</sup>。将未参与建模的 10 个番茄样品对该模型进行预测, 其中 6 个正常果, 4 个碰伤果。由于在出口中不允许存在碰伤果, 因此需先把碰伤果剔除, 再进行糖酸度分级。首先, 通过动态在线分选装置采集该样品的光谱信息, 然后用偏最小二乘判别模型预测出一个值, 并与阈值进行对比, 若大于阈值, 则认定为异常果; 若小于阈值, 则认定为正常果, 继续进行可溶性固形物含量的偏最小二乘判别模型预测。有研究表明, 糖度差异在 2% 以上会有明显不同的口感, 综合考虑模型的均方根误差, 将糖度区间设置 10% 以下、10~12%、12~14%、14% 以上。在实验中, 将 10 个预测实验样品按照序号依次进行分选实验, 每个样品按照标记朝上的方式进行上果, 每个面放置四次, 共放置 160 次, 并记录每次进入的分级口, 碰伤果判别准确率为 100%, 分选可溶性固形物时, 误入相邻的分级口 9 次, 可溶性固形物在线分选正确率为 91%。

### 3 结论

通过近红外光谱漫透射测量与分析, 验证了番茄碰伤与可溶性固形物同时在线检测是可行的, 建立了番茄碰伤果近红外漫透射偏最小二乘判别模型, 模型的准确性达到 100%, 在后续的实验中, 能够准确判别正常果与碰伤果。此外, 建立番茄可溶性固形物最小二乘回归模型时也考虑到番茄碰伤果对模型预测能力的影响, 最终建立番茄正常果的可溶性固形物的偏最小二乘回归模型, 提出了番茄碰伤与可溶性同时在线检测的方法。在后续的在线分选实验中, 正确率为 91%。为番茄生产出口在线检测分选方案提供了参考和依据。

### References

- [1] Zhiguo Li, Colin Thomas. Trends in Food Science & Technology, 2014, 35: 138.
- [2] Tian Jinjin, Zhang Zhe, Yan Lei, et al. Science and Technology of Food Industry, 2017, 7: 340.
- [3] Shimeles Tilahun, Do Su Park, Mu Hong Seo, et al. Postharvest Biology and Technology, 2018, 136: 50.
- [4] Wu Guifang, Wang Chunguang. Postharvest Biology and Technology, 2014, 98: 41.
- [5] Ding Xiaoxiao, Guo Ying, et al. Vibrational Spectroscopy, 2016, 82: 1.
- [6] Liu Yande, Shi Yu, Cai Lijun. Chinese Journal of Agricultural Machinery, 2013, 9: 138.
- [7] Xin Huang, Li Xia. Chemometrics and Intelligent Laboratory Systems, 2017, 6: 22.
- [8] Habib Motieghader, Ali Najafi, Balal Sadeghi, et al. Informatics in Medicine Unlocked, 2017, 9: 246.
- [9] Zhao Yize, Long Qi. Wiley Interdisciplinary Reviews: Computational Statistics, 2017, 9(5): doi: 10.1002/wics.1402.
- [10] Xu Shu, Lu Bo, Baldea Michael, et al. Journal of Process Control, 2018, 67: 83.

## The Online Detection Model Research of Tomatoes' Bruise and SSD

LIU Yan-de, RAO Yu, SUN Xu-dong, XIAO Huai-chun, JIANG Xiao-gang, ZHU Ke, XU Hai  
School of Mechatronics Engineering, East China Jiaotong University, Nanchang 330013, China

**Abstract** Soluble solids and bruise are the two main factors affecting tomato quality. The purpose of the study was to explore the feasibility of simultaneous on-line detection of bruises and soluble solids in tomatoes by near-infrared diffuse transmission spectroscopy. The near-infrared diffuse transmission spectra of tomato were collected under the condition of a single-channel fruit delivery speed of 5/s. The near-infrared diffuse transmission spectrum characteristics of the bruised and normal tomato samples were compared and analyzed. The results showed that there was a significant difference in the light intensity between the bruises and the normal tomato samples. The light intensity of the bruises was stronger than that of the normal fruits. The reason may be that the meat becomes soft and the light transmission becomes stronger after the bruising. The two absorption peaks are more than the normal fruit at 650 and 675 nm. The reason may be that the color of the tomato skin changes before and after the injury. The first three main scores with the highest contribution rate were selected. After qualitative analysis of principal components of near-infrared diffuse transmission spectra of normal fruits and bruises, normal fruits and bruises could not be effectively clustered. Therefore, high-dimensional near-infrared diffuse transmission spectral qualitative discriminant model was selected. By establishing the near-infrared diffuse transmission spectrum partial least squares qualitative discriminant model of the injured tomato sample, the false positive rate of the partial least squares qualitative discriminant model is 0%, which can correctly discriminate the fruit, so the near-infrared diffuse transmission spectroscopy partial least squares qualitative discriminant model of the touched tomato sample was selected as the online knockout sorting model for tomato touch injury. Validation of samples that have not been involved in modeling can correctly identify bruises. After the injurious fruit was removed by the near-infrared diffuse transmission spectroscopic partial least-squares qualitative discriminant model, the classification was based on the soluble solids index. The model is preprocessed using all the bands and the 606~850 nm band, and the second-order derivative preprocessing is performed on all the bands and the 606~850 nm band spectrum, and the front-back smoothing is set to 9, and the continuous projection algorithm and genetic algorithm are used to optimize the soluble solids. The spectral modeling variables, through comparison found that the use of non-algorithm screening 606~850 nm band spectral variables modeling, the best effect, established a soluble solids online detection model, the prediction set root mean square error of 0.43 Brix°. Simultaneous on-line detection of bruising and soluble solids using samples not involved in modeling demonstrated that the accuracy of sorting of bumped specimens was 96%, and the accuracy of sorting of soluble solids samples was 91%. The experimental results show that the simultaneous on-line detection of tomato bruising and soluble solids near-infrared diffuse transmission spectroscopy is feasible.

**Keywords** Diffuse transmission; Spectroscopy; On-line detection; Bruises; Soluble solids

(Received Oct. 19, 2018; accepted Feb. 8, 2019)