

文章编号: 1001-3806(2019)01-0025-05

## 西红柿成熟度的近红外漫透射光谱无损检测

刘燕德, 徐海, 孙旭东, 饶宇

(华东交通大学 光机电技术及应用研究所, 南昌 330013)

**摘要:** 为了证实以近红外漫透射光谱技术结合化学计量学方法能有效实现西红柿成熟度的检测,采用美国海洋光学公司的QE65000光谱仪取得了西红柿的漫透射光谱数据,采用日本柯尼卡美能达CR-10反射式色差计取得了亮度、红绿色相、黄蓝色相和总色差颜色数据;通过标准数学建模方法建立了颜色模型,并通过偏最小二乘回归校正算法建立了近红外光谱模型。结果表明,两组模型都是以红绿色相指标建模时效果最佳,在99个未知样品数据中,预测模型的颜色模型误判情况分别为未成熟0%、半成熟33.33%、成熟0%;而近红外预测模型的误判情况分别为未成熟32.14%、半成熟50%、成熟0%。该研究验证了近红外光谱技术对西红柿成熟度进行无损检测的可行性,这对实现西红柿的快速、批量分选具有一定的实际意义。

**关键词:** 光谱学; Lab颜色指标; 近红外漫透射光谱; 偏最小二乘回归算法; 定性; 西红柿

**中图分类号:** O657.3; TN247 **文献标志码:** A **doi:** 10.7510/jgjs.issn.1001-3806.2019.01.006

## Non-destructive measurement of tomato maturity by near-infrared diffuse transmission spectroscopy

LIU Yande, XU Hai, SUN Xudong, RAO Yu

(Institute of Optics Mechanics Electronics Technology and Application, East China Jiaotong University, Nanchang 330013, China)

**Abstract:** In order to prove that near-infrared diffuse transmission spectroscopy combining with chemometrics can effectively detect the maturity of tomatoes, diffuse transmission spectroscopy data of tomatoes were obtained by using QE65000 spectrometer of Ocean Optics Company of America, and the color datum of brightness, red and green, yellow blue color and total chromatic aberration were obtained by using CR-10 reflectance colorimeter of Konica Minolta of Japan. The color model was established by the standard mathematical modeling method, and the near infrared spectroscopy model was established by the partial least squares regression correction algorithm. The results show that, the red-green phase index was the best for both models. Among the 99 unknown samples, the misjudgments of the color model were immature 0%, semi-mature 33.33% and mature 0%, while the misjudgments of the near-infrared spectroscopy prediction model were immature 32.14%, semi-mature 50% and mature 0% respectively. This study verifies the feasibility of non-destructive detection of tomato maturity by near infrared spectroscopy, which has a certain practical significance for rapid and batch sorting of tomatoes.

**Key words:** spectroscopy; Lab color index; near-infrared diffuse transmission spectroscopy; partial least squares regression algorithm; qualitative; tomato

## 引言

西红柿果实营养丰富,具有丰富的番茄红素、维生

素C和B族维生素。西红柿的成熟度和其颜色有关,且一般是从未成熟的绿色阶段到成熟的红色阶段。但是现在对西红柿分选特别是二次分选还主要依靠人工分选,其成本较高、用时较长,不能很好地满足西红柿的分选需要。因此研究出一种能够针对西红柿成熟度进行无损、快速、批量分选的方法有着一定的意义。

目前,国内外研究工作者主要偏向于对西红柿品质、生长情况、储藏情况等的研究与实验,特别是对于西红柿栽培技术和病虫害防治技术研究较多<sup>[1-3]</sup>,而对于西红柿分选的研究与实验较少。

近红外光谱(near-infrared spectroscopy, NIRS)技

基金项目:国家自然科学基金资助项目(31760344);江西省优势科技创新团队建设计划资助项目(20153BCB24002);江西省高校协同创新中心建设专项资金资助项目(赣教高字[2014]60号)

作者简介:刘燕德(1967-),女,博士,教授,研究方向为光谱诊断技术。

E-mail: jxliuyd@163.com

收稿日期:2018-04-09;收到修改稿日期:2018-04-30

术作为一种快速、绿色的无损检测技术,因为其在常规光纤中良好的传输特性,使得近红外光谱在线分析技术得到很好的应用<sup>[4-6]</sup>,并且取得了较高的社会和经济效益。本文中主要提出了利用近红外光谱模型进行西红柿分选的方案,并且还对比了颜色指标模型。

## 1 材料与方 法

### 1.1 实验装置及光谱采集

实验中选用美国海洋光学公司的 QE65000 光谱仪,其采集的光谱为短波近红外光谱,波长范围:350nm ~ 1150nm;采样时间:100ms/次;光源:6个12V/100W 钨卤灯。采用漫透射方式,并在每个样品的赤道部位采集光谱。

### 1.2 样品来源及化学参考值测量

实验中所采用的西红柿样品由河北省石家庄市某果园提供。实验前,挑选表面不畸形、无损伤的正常西红柿作为实验样品,放于 22℃ 的实验室环境下保存 12h。实验中共挑选出 100 个西红柿样品,其中成熟果(红色)60 个,未成熟果(绿色)29 个,半成熟果(红绿相间)11 个。在每个西红柿的 4 个赤道部位标号,即一共有  $4 \times 100$  个样品数据,依照基于光谱变量的 Kennard-Stone (K-S)<sup>[7]</sup> 选择方法将样品集分成建模集(301 个)和预测集(99 个)2 组。其中,K-S 方法基于变量之间的欧氏距离,在特征空间中均匀选取样本。光谱采集完后,在对应部位进行西红柿表面色泽测量,实验中采用国际照明委员会(International Commission on Illumination, CIE)1976 年推荐的均匀色度空间 Lab 色系<sup>[8-13]</sup>。实验采用柯尼卡美能达 CR-10 反射式色差计(日本)测量样品的表面色泽,测量西红柿表面色泽之前,先测铁氟龙标准白板(亮度  $L = 90.6$ ,红绿色相  $a = -3.0$ ,黄蓝色相  $b = 0.5$ )作为样品的背景参比,再对水果表面进行测量,其结果通过  $L, a, b$  和总色差  $\Delta E$  这 4 个指标记录。

### 1.3 数据处理及建模方法

实验中通过美国海洋公司的光谱采集软件(SpectraSuite)采集西红柿样品的光谱,运用 Excel-2016 软件分析并建立 4 个颜色指标和对应类别的数学模型,并计算得出对应误判个数,应用偏最小二乘回归校正算法建立西红柿表面色泽的近红外校正模型。

偏最小二乘回归法是目前在近红外光谱分析中应用较为广泛的一种建模方法,也是一种新型的多元统计数据分析方法<sup>[14-15]</sup>。近些年来,它在光谱分析的理

论、方法和应用上都取得了极大的进展<sup>[16-19]</sup>。偏最小二乘回归属于多因变量到多自变量的回归建模方法,相对于普通多元回归来说,它往往能够更好地解决一些数据分析问题。

偏最小二乘法和主成分分析相似,其差别在于主成分分析只对光谱矩阵进行分解,从而消除噪声,而偏最小二乘是在其基础上还考虑了浓度阵的影响。总体来说,偏最小二乘回归即为多元线性回归、典型相关分析以及主成分分析的综合应用。其数学模型<sup>[20]</sup>为:

$$\begin{aligned} X &= TP' + E \\ Y &= UQ' + F \end{aligned} \quad (1)$$

式中, $X$  表示光谱矩阵; $Y$  表示浓度矩阵; $T$  表示  $X$  的得分; $U$  表示  $Y$  的得分; $P$  表示  $X$  的装载; $Q$  表示  $Y$  的装载; $E$  表示运用偏最小二乘模型法去拟合  $X$  所引进的误差; $F$  表示运用偏最小二乘模型法去拟合  $Y$  所引进的误差。

在理想的情况下, $X$  中误差的来源和  $Y$  中的误差来源完全相同,但实际上  $X$  中的误差和  $Y$  中误差并不相关,即  $T \neq U$ ,但在两个矩阵需要同时确定因子时, $X$  与  $Y$  的因子会具有如下关系:

$$U = bT + E \quad (2)$$

式中, $b$  表示  $U$  与  $T$  的内在关系,即回归系数。

最终经过一系列计算整理求解,得到的偏最小二乘回归方程为:

$$Y_j = a_{j0} + a_{j1}X_1 + a_{j2}X_2 + \dots + a_{jp}X_p \quad (3)$$

式中,下标  $j$  表示样本数; $a_{jp}$  表示方程的系数矩阵;下标  $p$  表示变量数。

## 2 结果与讨论

### 2.1 西红柿的颜色特征分析

水果的颜色是由果实发育过程中各种不同色素含量多少所决定的,并且在阳光、温度、水分等有不同的环境因素下,也造成了水果颜色的不同。而西红柿番茄红素含量的多少也在一定程度上反映了西红柿的成熟情况。本实验中采用了 Lab 颜色模型对西红柿进行颜色描述。Lab 是一种色彩空间,也即一种颜色模型,是在 1931 年国际照明委员会制定的颜色度量国际标准的基础上建立的,在 1976 年又经修订并被命名为 CIE Lab。这是一种与设备无关的颜色系统,也是一种基于生理特性的颜色系统,以数字化方式来描述人的视觉感应。它适用于一切光源色体或物体色的表示与计算。在 Lab 颜色空间中,一种颜色由  $L$  值、 $a$  值、 $b$  值 3 种参量表征,其满足公式:

$$\Delta E = (\Delta a^2 + \Delta b^2 + \Delta l^2)^{\frac{1}{2}} \quad (4)$$

式中,  $\Delta a$  表示红绿色相,  $+a$  表示偏红,  $-a$  表示偏绿;  $\Delta b$  表示黄蓝色相,  $+b$  表示偏黄,  $-b$  表示偏蓝;  $\Delta l$  表示明暗,  $+$  表示偏深,  $-$  表示偏浅;  $\Delta E$  表示总色差。

Table 1 Characteristics of calibration and prediction sets of tomato samples

parameter	data set	number of sample	minimum	maximum	mean	threshold 1	threshold 2
$L$	calibration set	301	-54.1	-27.1	-42.45	-34	-42
	validation set	99	-54.4	-27.9	-42.73		
$a$	calibration set	301	-9.5	33.9	13.32	-2.5	8
	validation set	99	-8.7	33	13.21		
$b$	calibration set	301	8.0	34.2	19.44	29	21
	validation set	99	10.1	36.9	19.44		
$\Delta E$	calibration set	301	39.1	63.4	51.88	45.5	51
	validation set	99	37.8	62.6	52.07		

由表 1 可知, 西红柿的亮度整体上偏暗(均为负值),  $Lab$  值数据范围都较大, 可见能够较好地地区分西红柿的红、绿颜色, 即选用  $Lab$  颜色模型对西红柿的颜色采集是合理可行的。通过数据分析可知, 阈值 1 为未成熟果与半成熟果的界限, 而阈值 2 表示半成熟果和成熟果的界限。

记录样品对应的类别标签, 其中未成熟果为 1.0; 半成熟果为 2.0; 成熟果为 3.0, 选用建模集的  $Lab$  及  $\Delta E$  数据和标签建立数学模型, 其结果见图 1。

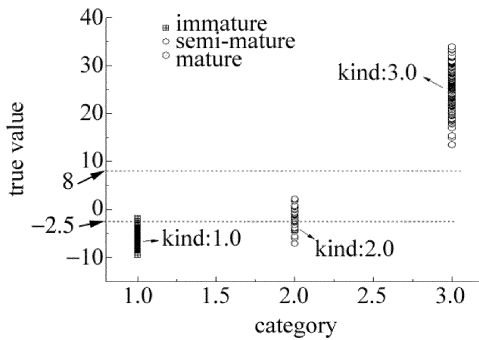


Fig. 1  $Lab$ ,  $\Delta E$  and its corresponding categories

图 1 为  $a$  颜色指标与其对应类别标签建立的散点模型。结果表明: 半成熟和成熟样品果的区分效果明显, 即能很好地判别成熟果和其它两类果; 而未成熟和半成熟样品果的分类效果不明显, 半成熟果与未成熟果的颜色数据相互包含现象严重。结合其它 3 个颜色指标对应的散点模型情况, 并经由计算分析, 西红柿样品建模集类与类之间的误判情况如表 2 所示。

通过综合比较可知,  $a$  值对应的建模效果最好, 其在识别未成熟和成熟样品果时, 误判率远远低于其它指标所建立的模型。但是, 由表 2 也能看出, 区分未成熟和半成熟果时都存在误判情况, 原因可能是一些人

在实验过程中, 通过 CR-10 色差计采集了每个西红柿各个标定面的色差信息, 即  $\Delta E, L, a, b$  4 组色差数据, 按 3:1 的比例将样品数据分为建模集和预测集, 其统计结果如表 1 所示。

Table 2 Misjudgment situation of calibration set samples

parameter	total number of sample	number of sample	number of misjudged	misjudgment rate / %
$L$ immature	301	88	44	50
		32	11	34.38
		181	1	0.55
$a$ immature	301	88	2	2.27
		32	11	34.38
		181	0	0
$b$ immature	301	88	45	51.14
		32	11	34.38
		181	0	0
$\Delta E$ immature	301	88	34	38.64
		32	22	68.75
		181	4	2.21

为测量误差和样品颜色模型不明显导致的。

## 2.2 西红柿近红外光谱特征分析

从西红柿样品中随机挑出一个成熟、一个半成熟以及一个未成熟共 3 个实验样品, 其对应的光谱信息如图 2 所示。

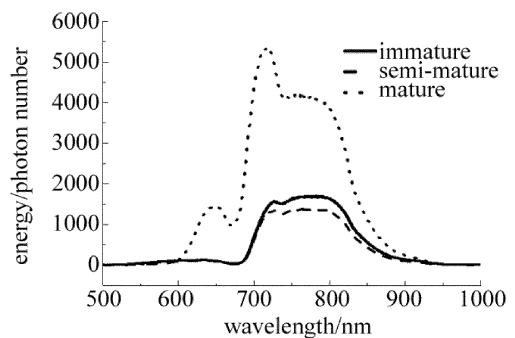


Fig. 2 NIRS of tomato samples

从图 2 中可以明显看出, 成熟果光谱与未成熟、半

成熟果光谱具有明显差异,而半成熟果与未成熟果的光谱之间没有太大的区别,只是在最大值上未成熟果略大。首先,成熟果的能量谱较未成熟果以及半成熟果更高,其中成熟果能量谱的最高值可以达到 5500 光子数,而未成熟果的光谱能量最高值仅为 1700 光子数,半成熟果的光谱能量最高值仅为 1385 光子数,产生这种差异的原因可能是成熟果内部果肉组织较未成熟果更软,使得成熟果的透光性也相对更高;而由于样本的限制,半成熟果与未成熟果差距不大。3 条光谱均存在波峰,且位置基本一致,都是在 700nm ~ 800nm 附近,而在 750nm 处存在波谷;成熟果还在 600nm ~ 700nm 存在波峰,在 650nm ~ 700nm 存在波谷,此波段对应红色光。

### 2.3 西红柿色泽的近红外光谱预测模型

偏最小二乘回归建模算法是基于全波段光谱进行主成分分析,充分利用了全波段的有效信息,其建模效果一般好于多元线性回归方法。故本实验中应用偏最小二乘回归校正算法对西红柿颜色的原始吸收光谱进行建模分析,其结果见图 3。

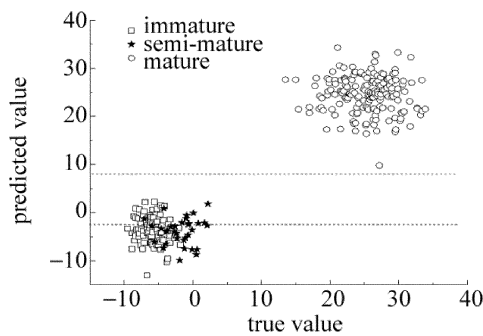


Fig. 3 Correlation diagram of NIRS modeling of color index and true value for tomato by partial least square method

图 3 是应用偏最小二乘回归方法建立的近红外光

Table 3 Misjudgement of near infrared spectrum model of tomato color and lustre index

parameter	number of sample	number of misjudged	misjudgment rate /%
<i>L</i> immature	88	53	60.22
semi-mature	32	13	40.62
mature	181	1	0.55
<i>a</i> immature	88	31	35.22
semi-mature	32	20	62.5
mature	181	0	0
<i>b</i> immature	88	65	73.86
semi-mature	32	9	28.12
mature	181	1	0.55
$\Delta E$ immature	88	32	36.36
semi-mature	32	20	62.5
mature	181	1	0.55

谱模型。图中的两条平行直线为上述表 1 中所给出的阈值。结果表明:利用偏最小二乘回归法建立的近红外光谱模型分类效果良好,尤其是在识别成熟果上表现优异,基本不发生误判,但区分未成熟果与半成熟果的结果和上述散点模型类似,误判情况明显,其结果见表 3。

表 3 结果表明,4 种颜色指标建立的近红外光谱模型都存在误判情况,其结果大致和上述散点模型类似:成熟果识别准确,而识别未成熟果和半成熟果时都存在误判;但综合比较,利用  $a$  指标建立的近红外模型效果最好。

### 2.4 不同模型对比分析

经过上述分析,选择  $a$  指标对应的近红外模型作为定标模型,为了检验定标模型的可靠性和准确性,用这一模型对 99 个未知样品的颜色进行预测分析,结果见图 4。

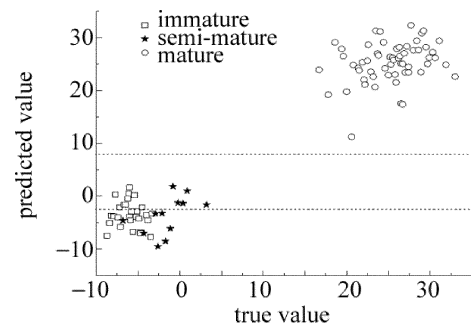


Fig. 4 Prediction effect of tomato NIRS model

将结果和标准颜色模型的预测情况进行对比,结果见表 4。

Table 4 Comparison of the prediction of near infrared spectroscopy and standard model

model	number of sample	number of misjudged	misjudgment rate /%
standard color model immature	28	0	0
semi-mature	12	4	33.33
mature	59	0	0
NIRS model immature	28	9	32.14
semi-mature	12	6	50
mature	59	0	0

从结果分析来看,近红外光谱模型在识别成熟果时预测效果较理想,但和真值建立的标准颜色模型相比,在识别其它两类时都存在较高的误判率,这可能是由于颜色指标分布过于集中而导致聚类现象不明显造成的。尽管近红外光谱模型在误判率上不如颜色模型,但是它具有以下两点优势:(1)标准颜色模型是基于采集的颜色数据,即每做一次分选,数据都要重新采

集,而利用近红外光谱模型对西红柿进行分选,光谱采集速度远远快于颜色数据采集速度,这将大大节约分选时间以及成本;(2)近红外光谱模型具有更加好的稳定性,数据采集更加准确,而颜色数据往往会因为实验人员的测量方式而影响数据采集的准确性。并且从实际角度考虑,分选的目的往往是需要检测出成熟的西红柿,从这点来看,近红外光谱模型能够很好地实现对西红柿成熟果的检测。

### 3 结 论

应用近红外光谱技术建立了西红柿成熟度的判别模型,该模型的预测效果较理想,能够很好地识别出成熟果。通过实验分析可知,近红外模型在识别西红柿是否处在半成熟这一过渡阶段的效果还不太理想,会出现将这类样品识别成未成熟果的情况。但总体来说,近红外光谱模型很适合用于大批量的西红柿成熟度分选上,能够很好地剔除非成熟果,避免在之后包装、销售等过程中的二次分选,也能减少人工分选的成本,加快分选速度。本研究可为其它农产品色泽近红外漫透射光谱在线无损检测提供参考依据。

#### 参 考 文 献

- [1] CAMPS C, DELTHEIL L, GILLI C. Qualitative and quantitative models based on handheld NIR spectroscopy to monitor the tomato fruit development during early and full season[DB/OL]. (2014-11-27) [2018-04-08]. <https://www.researchgate.net/publication/269016068>.
- [2] CAMPS C, SIMONE C, GILLI C. Assessment of tomato quality using portable NIR spectroscopy and PLSR with wavelengths selection[DB/OL]. (2014-12-01) [2018-04-08]. <https://www.researchgate.net/publication/290241088>.
- [3] ZHANG B F, YUAN L B, ZHANG B X. Study on predicting total acid content and soluble sugar of tomato juice by near infrared optical fiber spectrometer technique[J]. Spectroscopy and Spectral Analysis, 2014, 34(2): 350-353(in Chinese).
- [4] SUN T, XU H R, YING Y B. Progress in application of near infrared spectroscopy to nondestructive on-line detection of products/ food quality[J]. Spectroscopy and Spectral Analysis, 2009, 29(1): 122-126(in Chinese).
- [5] LIU Y D, WU M M, SUN X D, *et al.* Simultaneous detection of surface deficiency and soluble solids content for amygdalus persica by on-line visible-near infrared transmittance spectroscopy[J]. Transactions of the Chinese Society of Agricultural Engineering, 2016, 32(6): 289-295(in Chinese).
- [6] PAN Y Y, LIU Y D, SUN X D. Research progress of on-line detection of fruit quality using near infrared spectroscopy[J]. Laser & Infrared, 2010, 40(8): 805-810(in Chinese).
- [7] CHU X L, LIU H Y, YAN Z Ch. Near-infrared spectroscopy technology practical manual[M]. Beijing: Machinery Industry Press, 2016: 114-133(in Chinese).
- [8] RAO X Q, YING Y B. Grading a fruit by it's surface color[J]. Journal of Zhejiang University (Engineering Science Edition), 2009, 43(5): 869-871(in Chinese).
- [9] LIU Y D, CHEN X M, OUYANG A G. Non-destructive measurement of surface color of pear by visible/near-infrared diffuse reflectance spectra[J]. Journal of Infrared and Millimeter Waves, 2008, 27(4): 266-268(in Chinese).
- [10] ZHANG Y P, LI Ch, ZHAO M, *et al.* Analysis of color difference of cigarette packaging material based spectroscopy on near-infrared[J]. Tobacco Science & Technology, 2016, 49(2): 75-81(in Chinese).
- [11] QIN J, ZHANG M M, ZHAO G J, *et al.* Characterization of wood surface treated with electroless copper plating by near infrared spectroscopy technology[J]. Spectroscopy and Spectral Analysis, 2015, 35(5): 1253-1257(in Chinese).
- [12] PANG X M, YAN Z J, KAN J M. Color image segmentation based on HSI and LAB color space[J]. Journal of Guangxi University (Natural Science Edition), 2011, 36(6): 976-980(in Chinese).
- [13] JI Sh J, LUO Y, LI D H, *et al.* Effect of honey color on determination model establishment of near infrared spectrum[J]. Science and Technology of Food Industry, 2010, 31(8): 345-347(in Chinese).
- [14] YAN Y L. Fundamentals and applications of near infrared spectroscopy[M]. Beijing: China Light Industry Press, 2005: 8-39(in Chinese).
- [15] CHU X L. Chemometrics and molecular spectroscopy[M]. Beijing: Chemical Industry Press, 2011: 41-53(in Chinese).
- [16] MA C H, CUI J L. Quantitative analysis of composition in molten steel by LIBS based on improved partial least squares[J]. Laser Technology, 2016, 40(6): 876-881(in Chinese).
- [17] LIU Y D, XIE Q H, WANG H Y, *et al.* Quantitative study on phosphorus residues in navel oranges based on surface enhanced Raman spectra[J]. Laser Technology, 2017, 41(4): 545-548(in Chinese).
- [18] FU Y, ZHANG Y J, CHEN H C, *et al.* Establishment of NIR models for components determination in fresh chestnut[J]. Food Science and Technology, 2012, 37(5): 42-45(in Chinese).
- [19] OUYANG A G, XIE X Q, ZHOU Y R, *et al.* Partial least squares regression variable screening studies on apple soluble solids NIR spectral detection[J]. Spectroscopy and Spectral Analysis, 2012, 32(10): 2680-2684(in Chinese).
- [20] XU L, SHAO X G. Methods of chemometrics[M]. 2nd ed. Beijing: Science Press, 2004: 163-169(in Chinese).