

全透射近红外光谱的苹果整果糖度在线检测模型优化

田喜^{1,2,3}, 陈立平^{2,3}, 王庆艳^{2,3}, 李江波^{2,3}, 杨一^{2,3}, 樊书祥^{2,3}, 黄文倩^{2,3*}

1. 中国农业大学信息与电气工程学院, 北京 100083
2. 北京市农林科学院智能装备技术研究中心, 北京 100097
3. 北京市农林科学院信息技术研究中心, 北京 100097

摘要 光谱质量、样本个体差异、检测系统和建模算法等多种因素共同决定水果糖度检测模型的预测精度和稳定性。采用自主研发的短积分全透射近红外在线检测系统以5 ms积分时间和 $0.5 \text{ m} \cdot \text{s}^{-1}$ 运行速度在线获取了“富士”苹果全透射光谱信号。不同姿态获取的透射光谱强度差异明显,但曲线走势相近,均在920 nm波段具有最大的光谱强度,在850 nm波段存在波谷。采用移动平均平滑、标准正态变量变换和多元散射校正等预处理方法有效去除原始光谱的随机噪声和基线偏差,减小了样本检测姿态引起的光谱差异。为分析不同检测姿态对苹果整果糖度预测模型的影响,构建了单一姿态局限模型和多姿态通用模型,结果表明基于全位点平均透射光谱构建的单一姿态局限模型对检测姿态具有很大的局限性,而多姿态通用模型预测能力较单一检测姿态相当,但却对不同的检测姿态具有更强的适用能力。为进一步提高光谱信号质量,优化模型预测能力,采用信号强度阈值优选方法实现了苹果整果糖度预测模型优化,发现移除中央位点获取的透射光谱信号,有利于提高苹果整果糖度预测模型精度。多姿态通用信号强度优化模型综合考虑不同姿态获取的光谱信息有效性,有效提升了通用信号强度优化模型的预测能力和稳定性,当多姿态通用模型中信号强度阈值为5 000时,模型预测性能最佳,其预测参数 R_p , RMSEP和RPD分别为0.79, 0.84%和1.58。表明短积分全透射近红外在线检测系统用于不同姿态苹果糖度预测是可行的,多姿态通用模型的建立,扩大了模型在不同姿态的预测稳健性,短积分光谱采集方式结合信号强度阈值优选方法提升了光谱信号的质量和模型的预测能力。

关键词 在线检测; 全透射光谱; 通用模型; 苹果; 糖度

中图分类号: S37 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)06-1907-08

引言

苹果是世界四大水果之一,当前中国已成为世界最大的苹果生产国和消费国。随着我国苹果消费快速增长,消费者和种植者对苹果的内部品质也越来越关注。糖度是体现苹果内部质量的关键因素之一,直接影响了苹果售价。稳定、准确地对苹果糖度进行无损检测分级是目前苹果产业急需解决的问题。传统的苹果糖度测定方法属于损伤性检测,效率较低,检测范围有限,不适用于苹果品质分级。随着计算机和光电传感技术的发展,近红外光谱技术因无损、快速、准确的优势在水果品质检测领域快速发展和应用。

根据光源和光谱仪之间的位置关系不同,近红外无损检测通常可分为漫反射和透射二种模式^[1]。漫反射模式中光源和光谱仪位于同一侧,该模式在水果品质检测中应用较早,以静态检测为主。苹果表面较为光滑,漫反射模式会产生镜面反射信号,这些无用的镜面反射信号也会被光谱仪采集,降低水果品质的检测精度,这也是水果在线检测中漫反射模式研究较少的一个主要原因。另一方面,受光照强度和维管束分布影响,苹果糖度阳面高于阴面,近果皮处高于内部,近萼端部高于近梗端部,所以苹果属于内部组分分布不均的生物体。然而,漫反射模式只能获取光照区域苹果浅层果肉的光谱信息,无法对整果的综合品质做出有效评价。与漫反射模式不同,透射模式中光谱仪和光源分布在样本的两侧,

收稿日期: 2021-05-31, 修订日期: 2021-07-22

基金项目: 国家重点研发计划项目(2019YFD1101103), 青年北京学者(陈立平)项目, 国家自然科学基金项目(31901402), 2022年度农产品质量安全追溯技术及应用国家工程实验室建设项目资助

作者简介: 田喜, 1988年生, 中国农业大学信息与电气工程学院博士研究生 e-mail: tianx@sina.com

* 通讯作者 e-mail: huangwq@nrcita.org.cn

光谱仪可以采集到穿过样本的光谱信号,这些透射光谱携带了水果内部深层的组织信息^[2]。此外,透射模式也没有镜面反射等无用信号的干扰,是水果内部品质在线检测高效且极具前景的检测模式。透射模式也可以细分漫透射和全透射二种方式,漫透射方式可以获取部分水果内部组分信息;刘燕德等采用漫透射方式开发了水果动态在线分选设备,构建了不同品种苹果糖度在线检测通用模型,证明了漫透射方式在线检测苹果内部品质的可行性^[3]。全透射方式中光谱仪和光源呈直线式分布在样本两侧,入射光照射水果后,光谱仪获取整个水果的透射光谱信号,然而全透射方式对光谱仪的灵敏性要求较高,当前研究报道还较少。

尺寸、颜色、纹理、年份、产区等生物个体差异和环境温度、检测速度、检测姿态、检测位点等系统差异对预测模型性能均有影响^[4]。其中,检测姿态直接影响了光谱信号质量和预测模型的稳定性。Fan 等^[5]分析了苹果三种检测姿态的漫透射光谱信号质量,发现果梗-花萼轴竖直向上姿态获取的漫透射光谱信号最稳定。Xia 等^[6]基于苹果六种姿态获取的漫反射光谱构建了苹果整果糖度预测模型,发现果梗-花萼竖直向上姿态的整果糖度预测精度最高。建模算法对预测模型的预测能力和适用能力起到了重要影响;刘燕德^[3]等融合三个品种苹果光谱数据构建了苹果糖度在线检测通用模型。Tian 等^[7]分析了色素对苹果糖度预测模型的影响,基于移除色素特征波长后二次挑选的苹果糖度特征波段构建了更稳定的糖度预测模型。Guo 等^[8]将苹果表面颜色信息与漫反射光谱数据融合,提高了苹果糖度的预测精度。此外,光谱信号质量直接影响了光谱信噪比。苹果等水果是类球形生物体,入射光在苹果表面照射的位点不同,使得光在苹果内部传输的光程不同,而且苹果不同组织果实内部细胞结构、组成成分和光学传输特性也不相同,这些原因使得光谱仪在不同照射位点获取的全透射光谱信号强度具有差异性。总体而言,构建一个预测精度高、稳定性好的水果糖度检测模型需要综合考虑到光谱质量、样本个体差异、检测系统和建模算法等多种因素^[9-12]。

本研究采用课题组自主开发的短积分全透射近红外光谱

采集系统,基于构建的单一姿态和多姿态通用的在线检测模型,分析了检测姿态对苹果糖度预测精度的影响,并结合信号强度阈值优选方法,筛选了有利于苹果糖度预测的全透射光谱,建立了适用于单一姿态和通用姿态的苹果糖度最优检测模型。

1 实验部分

1.1 样品

试验用山东栖霞富士采购于北京岳各庄水果批发市场,手工挑选无表面缺陷、果型正常、直径 80~90 mm 苹果样本 218 个。试验前放于 22 °C 实验室环境下 24 h,以减少样本温度带来的试验干扰。

1.2 光谱采集

采用自主研发的短积分全透射光谱采集系统进行苹果样本光谱采集(图 1),该系统由一个卤素灯(FUJI, JCR, 150 W, 15 V)、一个光谱仪(光谱范围 615~1 044 nm,光谱间隔 0.5 nm)、一个传送系统、一个光电传感器和一个计算机组成。光源和光谱仪分别呈直线式分布于样本两侧,光谱仪采集样本全透射光谱信号,以便消除水果镜面反射干扰。水果传送通道采用了双层遮光帘以避免外界环境光的干扰。采用硬件触发光谱仪进行光谱采集,因光电传感器到光谱仪距离恒定,在匀速运输条件下,水果遮挡光电传感器后则可以计算水果传送到光谱仪所用的时间,光谱仪在延迟固定传送时间后开始以设定的积分时间连续采集透射光谱信号,通过计算苹果遮挡光电传感器时间,判断光谱仪的工作时长,从而避免光谱仪采集到无用信息,提高光谱的有效性。该系统具有积分时间短、信号灵敏度高的优势,可以在水果通过光谱仪时,以极短的积分时间从单个样本不同位点连续获取多条全透射光谱信号。苹果不同组织果实内部结构、组成成分和光学传输特性不同,因此不同位点获取的透射光谱信号携带了不同的组织信息,而短积分模式在增加光谱数据处理的灵活性的同时,也有利于局部组织的成分分析。

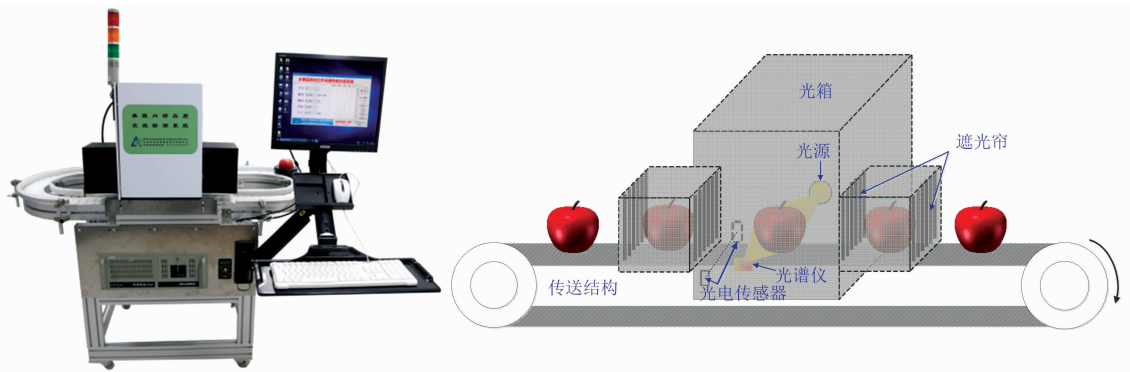


图 1 短积分全透射光谱采集系统

Fig. 1 Schematic of the on-line transmittance spectra measurement system

设定系统运行参数如下: 传送皮带运行速度 $0.5 \text{ m} \cdot \text{s}^{-1}$, 光谱仪积分时间 5 ms , 设备预热 30 后, 分别以如下三个姿态下采集苹果全透射光谱数据(图 2):

- 姿态 1(T1): 果梗-花萼轴竖直, 且垂直运动方向
 姿态 2(T2): 果梗-花萼轴水平, 且平行运动方向
 姿态 3(T3): 果梗-花萼轴水平, 且垂直运动方向

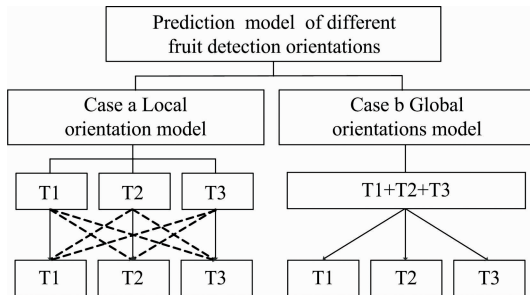


图 2 不同检测姿态的苹果糖度预测模型构建
 Fig. 2 Sugar prediction model based on different detection orientations

1.3 糖度测定

将苹果切成小块, 用双层纱布包裹后全部放入压汁机压出果汁, 摇匀后立即采用糖度计(型号 PR-101 α ; Atago Co., Tokyo, Japan)测量果汁的糖度值。重复测定三次, 取其平均值作为该样本整果糖度的真实值。

1.4 不同检测姿态的通用模型构建

偏最小二乘回归(partial least squares, PLS)是目前基于近红外光谱分析水果内部品质中广泛应用的一种建模方法。为了全面评价苹果检测姿态对苹果整果糖度预测模型的影响, 采用 PLS 构建了单一姿态局限模型和多姿态通用模型。

局限模型(Case a): 分别计算单一检测姿态获取的全位点平均透射光谱曲线, 采用 PLS 算法基于单一检测姿态校正集平均光谱数据构建单一姿态预测模型, 并采用单一姿态模型分别评价其他姿态下预测集样本的预测精度;

通用模型(Case b): 合并 T1、T2 和 T3 三种检测姿态校正集平均光谱数据, 采用 PLS 算法构建多姿态通用预测模型, 并采用通用模型分别评价三种姿态预测集样本的预测精度。

1.5 全透射光谱信号强度优选模型构建

所用的短积分全透射光谱采集系统可以 5 ms 积分时间, 在苹果通过光谱仪时, 连续不间断的从单个苹果不同位点获取多条全透射光谱信号。苹果不同位点入射光的光程不同; 苹果是类球形内部组分分布不均匀的生物体, 苹果不同位置果实内部细胞结构、组成成分和光学传输特性也不相同, 使得光谱仪在苹果不同位置获取的全透射光谱信号强度不相同。为了有效的提高短积分全透射光谱的信号质量, 提升苹果糖度预测模型精度, 发挥短积分全透射光谱采集系统信号处理灵活的优势, 提出一种基于信号强度优选的苹果整果糖度在线检测模型优化方法。方法要点如下: (1)确定全透射光谱曲线中信号强度最大的波段位置 W ; (2)确定波段 W 下信号阈值 N ; (3)筛选波段 W 下信号阈值大于 N 的全透射多

位点光谱曲线; (4)计算筛选的多位点光谱曲线平均光谱曲线; (5)采用 PLS 算法构建信号强度阈值为 N 时的模型预测; (6)按照设定步长提高信号强度阈值 N , 分析苹果整果糖度预测精度随信号强度阈值的变化趋势, 优选最优预测模型。当信号强度阈值 N 大于样本集中任一样本在波段 W 下光谱最大信号时, 该样本没有可用于建模的光谱数据, 此时模型终止运行。

1.6 模型预测精度的客观性分析

所有 218 个样本按照 3:1 比例随机划分为校正集和预测集, 其中校正集共计 162 个用于建立预测模型, 预测集共计 52 个用于评价预测模型性能。为了消除样本随机分布导致的模型精度不稳定问题, 便于客观地评价模型预测能力, 将样本随机划分了 20 次, 并基于每次样本划分结果构建预测模型, 最终模型预测精度用 20 次建模的平均值和标准偏差表示。

1.7 模型预测性能评价

所有模型的预测性能通过校正相关系数(correlation coefficient of calibration, R_c)、预测相关系数(correlation coefficient of prediction, R_p)、校正均方根误差(root mean square error of calibration, RMSEC)、预测均方根误差(root mean square error of prediction, RMSEP)、残差预测偏差(residual prediction deviation, RPD)来评价。

2 结果与讨论

2.1 糖度统计

苹果样本糖度范围在 $8.1\% \sim 14.6\%$, 平均值为 11.49% , 标准偏差为 1.33% , 样本集覆盖了足够大的范围, 这有利于提高所建预测模型的适用性。

2.2 光谱特征分析

图 3(a-c)分别显示了苹果不同检测姿态、多位点光谱采集示意图和单个苹果获取的多位点光谱曲线。650~1 000 nm 波段范围内光谱数据质量较好, 所以选用这一区间光谱数据用于后续建模分析。苹果样本直径 $80 \sim 90 \text{ mm}$, 运行速度 $0.5 \text{ m} \cdot \text{s}^{-1}$, 积分时间 5 ms , 单个样本可获取光谱数量约 $30 \sim 35$ 条。从图中可以看出, 苹果通过光谱仪时, 不同的检测姿态入射光照射在苹果表面的位点也不相同; 然而三种检测姿态下全透射光谱曲线的走势大致相同, 在 920 nm 附近具有最大的光谱信号强度, 在 850 nm 附近存在波谷。图 4 为同一苹果三种不同检测姿态的平均光谱曲线, 其中 T1 和 T2 姿态平均光谱强度相近, 而 T3 姿态光谱信号强度最低, 这可能是 T3 姿态下入射光穿透苹果的光程最长和苹果核的阻挡作用导致的。图 5 显示了三种不同检测姿态下同一苹果所有光谱采集位点 920 nm 波段信号强度的动态变化, 三种检测姿态不同位点光谱强度都有由高到低再升高的趋势, 这也反映了光谱仪获取的苹果检测位点是由边缘到中央再到边缘的过程。分析光谱的信号强度变化发现, T3 姿态下光谱强度变化最剧烈, 由边缘检测位点的最高值转变为中央检测位点的最低值。结合采集姿态和采集位点结构示意图, 分析苹果单果光谱采集的动态变化发现, 在 T1 姿态中, 透射光谱主要获取

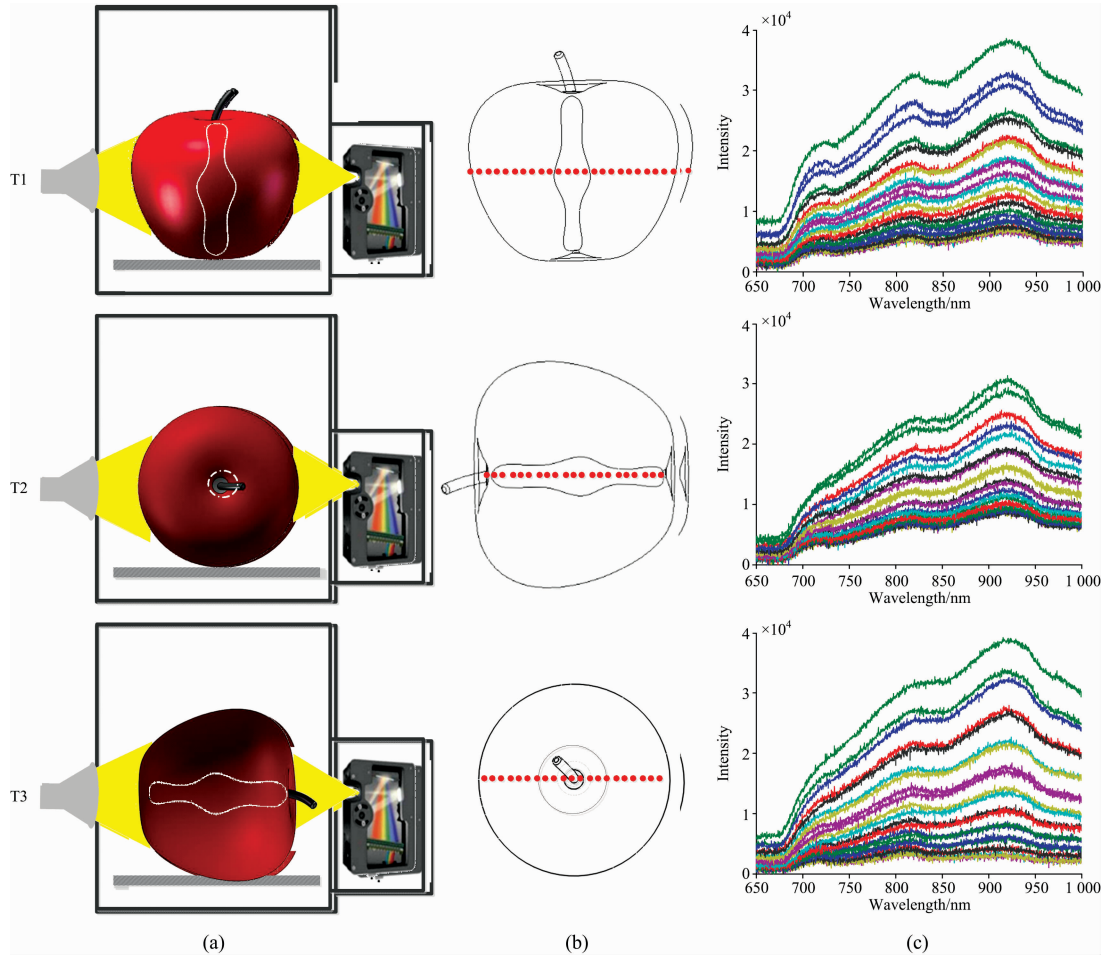


图 3 不同检测姿态下苹果透射光谱采集示意图

(a): 不同的检测姿态; (b): 检测位点示意图; (c): 同一苹果的不同姿态原始光谱

Fig. 3 The schematic diagram of different detection orientations

(a): Different orientations; (b): Spectral collection positions; (c): Raw multi-point transmittance spectra

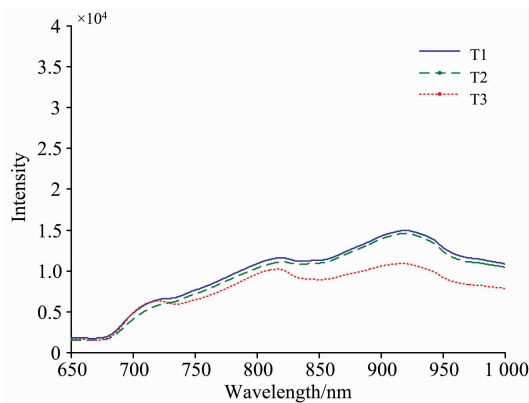


图 4 同一苹果三种不同检测姿态的平均光谱曲线
Fig. 4 Average spectral curves of a whole apple with three different detection orientations

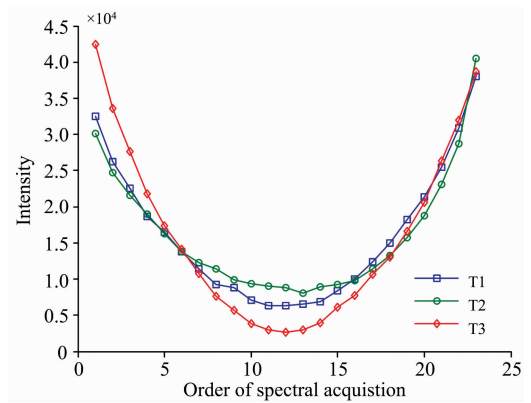


图 5 不同检测姿态下 920 nm 波段处光谱强度变化
Fig. 5 Spectral intensity changes at 920 nm for different detection orientations

于苹果赤道线上,其中中央采集位点光谱是穿透苹果果核区域而采集得到,此时穿透苹果的人射光程最长,透射光谱的信号强度最低。此外,苹果果核结构较为坚硬,光在这一

区域的穿透能力较弱,进一步削弱了透射光的信号强度。T2 姿态中,透射光谱主要沿着苹果果梗-花萼轴获取,苹果果核结构始终阻挡着部分入射光的行进。T3 姿态中,透射光谱主

要获取于苹果的果肩和果梗区域, 其中中央采集位点光谱源于果梗区域, 其信号强度最低, 再次证明了 T3 姿态下果梗-花萼轴严重削弱了果梗区域位点的全透射光谱强度。

2.3 光谱预处理分析

由于获取的原始光谱可能含有无用的背景信息或噪声, 预处理方法往往可以提高模型的性能。采用 21 点移动平均平滑(smoothing)、标准正态变量变换(standard normal variable, SNV)、和多元散射校正(multiplicative scattercorrection, MSC)等不同的预处理方法对原始光谱数据进行了预处理。

2.4 检测姿态对整果糖度预测影响

表 1 显示了基于单一检测姿态获取的全位点平均光谱及预处理后光谱数据集构建的苹果整果糖度预测模型结果(局限模型)。比较校正集和预测集模型参数发现, 基于原始光谱所构建模型出现了过度拟合问题(R_c 和 R_p 相差较大)。比较预处理前后光谱所建模型发现, 预处理后光谱所建模型没有表现过拟合, 且预测偏差 RMSEP 和 RPD 均优于原始光谱。Smoothing 预处理方法所建的 T1, T2 和 T3 局限模型的

RMSEP 和 RPD 分别为 $0.80\% \pm 0.07\%$ 和 1.45 ± 0.20 , $0.89\% \pm 0.08\%$ 和 1.36 ± 0.16 , $0.88\% \pm 0.17\%$ 和 1.42 ± 0.18 , 预测性能均优于 SNV 和 MSC 预处理方法, 且 T1 姿态检测精度优于 T2 和 T3 姿态, 因此 T1 姿态更适用于固定姿态下苹果整果糖度品质检测。

基于单一姿态构建的局限模型预测其他检测姿态光谱时, 其模型预测精度显著下降。以 T1 局限模型为例, 预测集 RMSEP 和 RPD 由 T1 的 $0.80\% \pm 0.07\%$ 和 1.45 ± 0.20 到 T2 的 $1.15\% \pm 0.13\%$ 和 1.11 ± 0.13 、T3 的 $2.13\% \pm 0.33\%$ 、 0.75 ± 0.06 , 从预测结果可以看出局限模型的鲁棒性不够好, 说明不同检测姿态获取的光谱信息存在很大差异, 局限模型对光谱检测姿态的变化比较敏感, 证明了单一姿态构建的局限模型对检测姿态具有很大的局限性。然而, 对苹果内部品质检测设备的开发而言, 苹果属于易损伤水果, 更适用于人工放果或吸盘自动放果的单果托盘式检测, 这种上果方式可以有效的保证姿态的一致性, 也有利于提高苹果糖度预测模型的精度。

表 1 不同检测姿态局限模型预测结果
Table 1 The detection result of restricted model

检测姿态	预处理方法	校正集		预测集								
		R_c	RMSEC /%	T1			T2			T3		
				R_p	RMSEP/%	RPD	R_p	RMSEP/%	RPD	R_p	RMSEP/%	RPD
T1	Raw	0.99	0.85	0.74	0.91	1.15	0.69	1.23	0.89	0.57	2.3	0.69
	Smoothing	0.88	0.76	0.81	0.80	1.45	0.74	1.15	1.11	0.67	2.13	0.75
	SNV	0.87	0.73	0.79	0.83	1.37	0.71	1.22	0.96	0.53	2.76	0.86
	MSC	0.88	0.85	0.75	1.28	0.96	0.73	0.93	1.28	0.70	1.25	1.04
T2	Raw	0.99	0.90	0.75	1.19	0.86	0.73	0.93	1.10	0.64	1.30	1.06
	Smoothing	0.86	0.85	0.79	1.07	1.09	0.76	0.89	1.36	0.76	1.00	1.37
	SNV	0.87	0.85	0.75	1.24	0.98	0.73	0.94	1.26	0.49	2.49	1.11
	MSC	0.88	0.85	0.75	1.28	0.96	0.73	0.93	1.28	0.70	1.25	1.04
T3	Raw	0.96	0.96	0.60	1.15	0.88	0.61	1.23	0.81	0.67	1.04	1.08
	Smoothing	0.88	0.86	0.66	1.06	1.00	0.69	1.00	1.11	0.77	0.88	1.42
	SNV	0.80	0.91	0.57	1.39	0.58	0.51	1.25	0.65	0.62	1.24	1.06
	MSC	0.80	0.91	0.56	1.53	0.74	0.49	1.36	0.86	0.62	1.26	1.06

2.5 不同检测姿态通用模型开发

考虑到实际在线水果检测时, 苹果固定检测姿态对人工要求较高, 为此, 建立了相对单一姿态局限模型而言的多姿态通用模型, 分析通用模型下不同检测姿态对预测结果的影响。表 2 显示了通用模型性能, Smoothing 预处理后构建的 T1、T2 和 T3 姿态通用模型性能最佳, 其 RMSEP 和 RPD 分别为 $0.85\% \pm 0.08\%$ 和 1.23 ± 0.14 , $0.87\% \pm 0.09\%$ 和 1.27 ± 0.18 , $0.84\% \pm 0.14\%$ 和 1.48 ± 0.14 。与单一姿态的局限模型相比, 通用模型与同一姿态的局限模型对预测集样本的预测结果无明显差别, 通用模型 T1 姿态(RMSEP = $0.85\% \pm 0.08\%$) 稍差于局限模型(RMSEP = $0.80\% \pm 0.07\%$), 而通用模型 T2 姿态(RMSEP = $0.87\% \pm 0.09\%$) 和 T3 姿态(RMSEP = $0.84\% \pm 0.14\%$) 稍优于局限模型(T2: RMSEP = $0.89\% \pm 0.08\%$, T3: RMSEP = $0.88\% \pm$

0.17%)。虽然通用模型预测性能没有显著优化, 但是它对在不同检测姿态的预测集都获得了准确的结果。因此, 通用模型对检测姿态的变化不再敏感, 也具有更好的适用性和实用性。

2.6 不同检测姿态信号强度优化模型

信号强度越高越能更加详细地表示预测精度随信号强度阈值的变化。三种检测姿态全透射光谱信号在 920 nm 波段强度最大, 因此, 设定阈值优选波段 W 为 920 nm, 设定起始阈值 $N=0$, 步长 500。图 6 显示三种检测姿态下 Smoothing 预处理光谱 20 次预测模型平均性能随着光谱信号阈值变化的动态曲线。随着阈值的不断提高, 三种检测姿态的苹果整果糖度预测偏差 RMSEP 先降低后升高, RPD 先升高后降低。当信号强度阈值为 12 500, 10 000 和 9 500 时, T1, T2 和 T3 姿态所建模型 RMSEP 最低, 分别为 0.78% , 0.88%

表 2 多检测姿态通用模型预测结果
Table 2 The detection result of universal model

预处理方法	校正集		预测集								
	R_c	RMSEC / %	T1			T2			T3		
			R_p	RMSEP / %	RPD	R_p	RMSEP / %	RPD	R_p	RMSEP / %	RPD
Raw	0.94	0.83	0.77	0.86	1.20	0.78	0.85	1.26	0.77	0.86	1.32
Smoothing	0.83	0.79	0.78	0.85	1.23	0.76	0.87	1.27	0.79	0.84	1.48
SNV	0.80	0.87	0.70	1.02	0.94	0.65	1.07	0.89	0.67	1.20	1.34
MSC	0.83	0.80	0.74	0.92	1.22	0.69	0.98	1.15	0.70	1.03	1.31

和 0.83%。当信号强度阈值为 12 500, 10 000 和 9 000 时, T1, T2 和 T3 姿态所建模型 RPD 最高, 分别为 1.73, 1.65 和 1.52。即 T1, T2 和 T3 优选信号阈值分别为 12 500, 10 000 和 9 000~9 500, 由图 5 可知这些阈值区域基本分布在苹果光谱采集的中央位点区域, 说明移除苹果中央位点区域获取的透射光谱可以优化预测模型精度。结合苹果结构和不同姿态下苹果光谱采集位点示意图分析, 当苹果被移动通过光谱仪过程中, 光谱仪连续采集苹果边缘—中央—边缘的检测位点, 三种检测姿态下不同位点光谱强度由高到低再到高。因此随着信号强度阈值的不断提升, 苹果中央位点光谱被逐渐移除, 苹果整果预测模型精度逐步提升。当信号强度阈值继续提高时, 被移除的光谱由苹果中央位点向两侧边缘不断扩展, 被移除的光谱范围越来越广, 透射光谱携带的苹果组织信息越来越少, 当越来越少的苹果两侧边缘位点光谱参与苹果整果糖度模型构建时, 苹果整果预测模型预测精度逐渐降低。分析不同检测姿态下模型预测精度随信号阈值变化程度发现, T3 姿态在阈值 9 000 之前变化最剧烈, 模型提升的效果也最显著, RMSEP 和 RPD 由初始阶段的 0.88% 和 1.57 分别提升到 0.83% 和 1.65。T2 姿态在阈值 10 000 之前均比较稳定, 这可能与该姿态获取的苹果组分信息较为稳定有关。三种姿态下中央位点(直径最大区域)光谱移除有助于模型精度提高, 一方面可能是因为中央位点信号强度弱, 增加了噪声的同时, 减小了光谱信噪比, 削弱了光谱信号质量

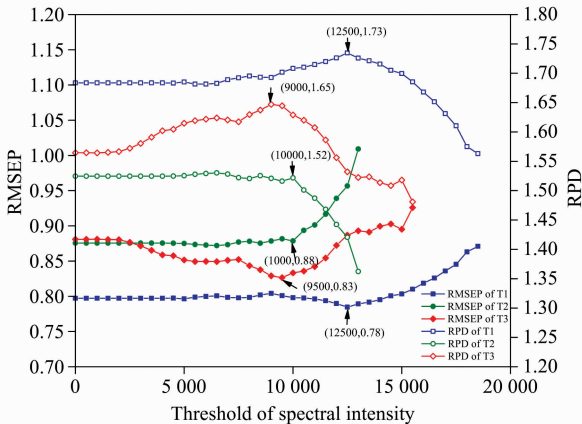


图 6 不同检测姿态下模型预测性能随光谱信号阈值变化动态曲线

Fig. 6 Dynamic curves of predictive model performance changing with spectral signal threshold in different detection orientations

导致; 另一方面也有可能是因为从中央位点获取的透射光谱穿透了苹果果核结构, 苹果核主要由籽粒和纤维膜物质组成, 这些与苹果糖度预测无关的组织信息降低了糖度模型的预测精度。

2.7 不同检测姿态通用信号强度优化模型

以全透射光谱信号强度最高的 920 nm 波段作为阈值优选波段 W, 设定起始阈值 $N=0$, 步长 500。图 7 展示了三种不同检测姿态 Smoothing 预处理光谱合并后构建的 20 次通用信号强度优化模型平均性能随信号强度变化动态曲线, 与单一姿态构建的信号强度优化模型变化趋势相同, 通用信号强度优化模型性能也随着信号强度的提升表现出先提升后降低的走势。然而通用强度优化模型中信号强度 5 000 时, 预测性能达到最优, RMSEP 和 RPD 分别为 0.84% 和 1.58, 最优信号强度阈值显著低于单一姿态下的最优信号强度阈值。不同姿态获取的透射光谱信号强度差异显著, 不同检测姿态下光谱采集位点也不相同, 通用信号强度优化模型的构建综合考虑多种姿态获取的光谱信息有效性, 删除了 920 nm 波段下信号强度小于 5 000 的无效光谱, 提升了通用信号强度优化模型的预测能力和稳定性。

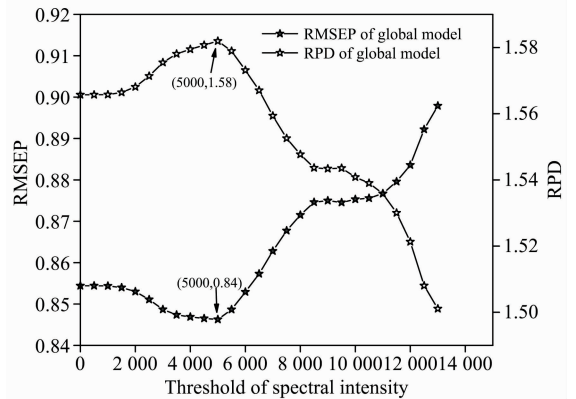


图 7 多姿态通用模型预测性能随光谱信号阈值变化动态曲线

Fig. 7 Dynamic curves of universal prediction model performance changing with spectral signal threshold in multiple orientations

3 结论

短积分全透射近红外在线检测系统可以用于不同姿态下苹果糖度通用模型开发。本研究采用移动平均平滑、标准正

态变量变换和多元散射校正等预处理方法有效去除原始光谱的随机噪声和基线偏差,减小了样本检测姿态引起的光谱差异。基于全位点平均透射光谱构建的单一姿态局限模型对检测姿态具有很大的局限性,而多姿态通用模型预测能力较单一检测姿态相当,但却对不同的检测姿态具有更强的适用能力。基于信号强度阈值优选方法的苹果整果糖度预测模型进一步优化了模型的预测能力,移除中央位点获取的透射光谱信号,有利于提高苹果整果糖度预测模型精度。不同检测姿态获取的透射光谱信号强度不同,不同检测姿态获取的苹果

组织信息也具有差异性,使得苹果整果糖度检测姿态的最优信号强度不同。不同检测姿态的通用信号强度优化模型综合考虑多种姿态获取的光谱信息有效性,有效提升了通用信号强度优化模型的预测能力和稳定性。结果表明短积分全透射近红外在线检测系统用于不同姿态苹果糖度预测是可行的,短积分多位点透射光谱采集模式提高了光谱数据处理的灵活性,信号强度阈值优选方法提升了光谱信号的质量和模型的预测能力。

References

- [1] Tian X, Fan S, Li J, et al. *Infrared Physics & Technology*, 2019, 102: 102979.
- [2] Zhang B, Gu B, Tian G, et al. *Trends in Food Science & Technology*, 2018, 81: 213.
- [3] LIU Yan-de, XU Hai, SUN Xu-dong, et al(刘燕德, 徐海, 孙旭东, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2020, 40(3): 922.
- [4] Bai Y, Xiong Y, Huang J, et al. *Postharvest Biology and Technology*, 2019, 156: 110943.
- [5] Fan S, Guo Z, Zhang B, et al. *Food Analytical Methods*, 2016, 9(5): 1333.
- [6] Xia Y, Huang W, Fan S, et al. *Infrared Physics & Technology*, 2019, 97: 467.
- [7] Tian X, Li J, Wang Q, et al. *Food Chemistry*, 2018, (239): 1055.
- [8] Guo Z, Huang W, Peng Y, et al. *Postharvest Biology and Technology*, 2016, (115): 81.
- [9] WANG Fan, LI Yong-yu, PENG Yan-kun, et al(王凡, 李永玉, 彭彦昆, 等). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2017, 33(19): 295.
- [10] GUO Zhi-ming, CHEN Quan-sheng, ZHANG Bin, et al(郭志明, 陈全胜, 张彬, 等). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2017, 33(8): 253.
- [11] Yu X, Lu Q, Gao H, et al. *Journal of Near Infrared Spectroscopy*, 2016, 24(1): 69.
- [12] Fan S, Wang Q, Tian X, et al. *Biosystems Engineering*, 2020, 193: 138.

Optimization of Online Determination Model for Sugar in a Whole Apple Using Full Transmittance Spectrum

TIAN Xi^{1, 2, 3}, CHEN Li-ping^{2, 3}, WANG Qing-yan^{2, 3}, LI Jiang-bo^{2, 3}, YANG Yi^{2, 3}, FAN Shu-xiang^{2, 3}, HUANG Wen-qian^{2, 3*}

1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

2. Intelligent Equipment Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China

3. Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China

Abstract In Vis/NIR nondestructive detection, the accuracy of the prediction model is affected by many factors such as spectral quality, biological variability, detection system and modeling method. In this study, the multi-point full-transmittance spectra (650~1 000 nm) of “Fuji” apple were acquired at a speed of $0.5 \text{ m} \cdot \text{s}^{-1}$ with an integration time of 0.5 ms using an on-line spectrum measurement system. The spectral intensity changed with the detection orientations significantly, but the spectral curves of different orientations were similar, with an obvious peak at 920 nm and an obvious valley at 850 nm. To establish a reliable, accurate, and stable sugar calibration model of intact apple, three spectra preprocessing methods, including moving average smoothing, standard normal variate, and multiplicative scatter correction (MSC), were used to reduce the influence of noise from the environment and instrumental fluctuations. In order to analyze the effect of detection orientations on prediction accuracy, a local model based on single orientation and a universal model based on global orientations were built respectively. The result showed that the prediction accuracy was limited by the detection orientation in the local model, while the universal model had better applicability for multiple detection orientation than that of the local model. In order to further improve the prediction ability, a modeling method named efficient spectrum optimization was proposed to select the spectra with a high signal-to-noise ratio and remove the inefficient transmittance spectrum by investigating the interference of transmittance spectral

intensity on the accuracy of the prediction model. The result showed that it is beneficial to optimize the prediction model after removing the spectrum collected from the central zone of the apple. The universal intensity optimization model considered the spectral quality of different orientations comprehensively. The prediction model was best with R_p , RMSEP and RPD of 0.79, 0.84% and 1.58 respectively, when the spectral intensity threshold was 5 000. Our result illustrated that the multi-potential spectrum measurement system is promising for on-line detection of apple quality. The modeling method of efficient spectrum optimization could be selective the transmittance spectrum with a high signal-to-noise ratio and optimize the prediction model.

Keywords Online detection; Full transmittance spectrum; Universal prediction model; Apple; Sugar content

(Received May 31, 2021; accepted Jul. 22, 2021)

* Corresponding author

本 刊 声 明

近期以来,一些不法分子假冒《光谱学与光谱分析》期刊社名义,以虚假网站等形式欺骗广大作者、读者。这些虚假网站公然假冒《光谱学与光谱分析》期刊名义进行大肆的征稿并骗取作者的审稿费和版面费。经部分作者及读者举报,现有关部门已就此介入调查。本刊将通过法律途径向假冒者追究相应的责任,维护本刊权益。

本刊官方网站已正式开通,网址为

<http://www.gpxygpx.com/>

在此郑重声明,本网址为《光谱学与光谱分析》期刊唯一开通运行的官方网站。本刊从未授权任何单位或个人以任何形式(包括网上网下)代理本刊征稿、审稿等业务。

希望广大读者和作者切实维护好自身的合法权益,防止受骗上当。

《光谱学与光谱分析》期刊社

2019年3月15日