

可见-近红外透射光谱检测活体甘蔗蔗糖分的方法

吕雪刚^{1, 2}, 李修华^{1, 2*}, 张诗敏², 张木清¹, 蒋洪涛¹

1. 广西甘蔗生物学重点实验室, 广西 南宁 530004

2. 广西大学电气工程学院, 广西 南宁 530004

摘要 蔗糖分是衡量甘蔗品质的重要指标, 基于光谱学原理的活体甘蔗蔗糖分无损检测的方法具有重要意义。探究了不同光谱测量方式对预测结果的影响, 改进了特征提取方法, 建立了甘蔗蔗糖分预测模型。采用自行搭建的甘蔗透射光谱采集平台, 在光源与探测器夹角为 120°, 150°和 180°的条件下, 按照甘蔗表皮蜡质去除与否, 获取了 123 个甘蔗样本在 6 类不同采集方式下的甘蔗透射光谱数据集。首先采用了偏最小二乘回归(PLS)的建模精度评价了 Savitzky-Golay 平滑处理(S-G 平滑)、标准正态变化(SNV)、多元散射校正(MSC)、一阶微分(FD)等多种光谱预处理方法对模型性能的影响, 选取了综合效果最好的 SNV 进行预处理。接着分析了不同光谱测量方式对蔗糖分建模的影响。结果发现: (1)对于蜡质覆盖的影响而言, 去蜡后的光谱透过率更高、单样本不同采集位点的光谱差异更低、与蔗糖分的相关性也更高; (2)对于光谱采集角度的影响而言, 在一定范围内光谱透过率随角度的增大而降低; (3)在获取的 6 类数据集中, 去蜡且采集角度为 120°的测量方式下获取的透射光谱建模效果最好, 预测集相关系数为 0.790 6, 预测集均方根误差为 0.898 6。最后, 使用了间隔偏最小二乘法(i-PLS)、遗传算法(GA)、蚁群算法(ACO)和提出的基于全波段 PLS 建模回归系数的改进蚁群算法(VRC-ACO)筛选特征波长。VRC-ACO 挑选得到的特征波长数量最少(10 个), 且建模效果最好, 预测集相关系数达到 0.861 6, 比全波段建模结果提高 9.0%, 预测均方根误差降低到 0.746 6, 比全波段建模结果降低 20.0%。该研究为甘蔗蔗糖分无损检测及相应传感器的研发提供了理论支持。

关键词 可见-近红外光谱; 甘蔗; 蔗糖分; 特征波长; 测量方式

中图分类号: S566.1 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)12-3747-06

引言

甘蔗的蔗糖分是衡量甘蔗品质的重要指标, 活体甘蔗蔗糖分的速测对甘蔗进行及时的田间管理、甘蔗品质评价等方面都具有一定的现实意义。在实际生产中, 常用的甘蔗蔗糖含量检测方法有旋光法、二次旋光法^[1]、液相色谱法等, 而这些检测方法都需要对样本进行预先处理、耗时耗力、测定步骤繁杂等问题, 难以实现快速大批量的检测。

可见-近红外光谱分析技术相比于传统的化学分析方法显著减少了检测的时间成本和化学药品对环境影响^[2], 是一种快速高效的检测手段, 目前已经广泛的运用在食品^[3-4]、农产品内部品质检测当中^[5-7]。对于诸如哈密瓜、苹果等圆型瓜果, 已有学者对比了不同的光谱测量方式和特征提取方

法对建模效果的影响, Li 等^[8]采集了哈密瓜赤道部位和尾部在 750~950 nm 范围的漫反射光谱来预测哈密瓜的可溶性固形物, 得出哈密瓜尾部的光谱建模效果高于赤道部位的结论。Xu 等^[9]对比了回归系数、遗传算法等四种不同的特征波长筛选方法在沙田柚水分含量预测的效果, 结果表明遗传算法是柚子水分检测分析的最佳特征提取方法, 其水分预测模型的决定系数为 0.712, 均方根误差为 0.048 8。针对甘蔗这类圆柱体样本的光谱检测方法研究也有相关报道, Taira 等^[10]使用便携式近红外仪器, 获取了 180°夹角下的未去蜡甘蔗的 600~1 100 nm 的透射光谱, 建立了全波段的甘蔗蔗糖分预测模型; SteidleNeto 等^[11]通过 Y 型光纤采集了 450~1 100 nm 的甘蔗表面漫反射光谱数据对甘蔗品种进行分类, 发现使用偏最小二乘法判别分析的分类精度可达 82%。

针对甘蔗这类外皮较硬、表面覆蜡的圆柱体样本, Taira

收稿日期: 2021-06-07, 修订日期: 2021-09-25

基金项目: 国家自然科学基金项目(31760342), 广西科技重大专项(桂科 2018-266-Z01, 桂科 AA18118037)资助

作者简介: 吕雪刚, 1997 年生, 广西大学电气工程学院硕士研究生 e-mail: lyuxg1997@qq.com

* 通讯作者 e-mail: lixh@gxu.edu.cn

和 SteidleNeto 等学者进行了相应的光谱学研究，但是其样本光谱采集方式相对单一，没有对不同的光谱测量方式进行差异化研究。因此，本文探究了光谱测量角度的变化、蜡质的覆盖对甘蔗光谱的透射效果以及建模效果的影响。并在最优的光谱测量方式下，探讨了不同特征波长筛选方法的性能，以进一步提升建模精度。

1 实验部分

1.1 样本采集

2021 年 1 月 5 日至 7 日，于广西大学扶绥农科新城甘蔗试验田，随机选取了 6 个不同的甘蔗品种，每个品种均挑选茎秆笔直未倒伏，表皮没有裂口和其他明显瑕疵的蔗茎，共 70 根（其中 60 根处于成熟期，10 根处于伸长期）。蔗茎采集后存放于实验室，在室温下放置 24 h 之后，使用干毛巾清理样本表面灰尘。对成熟期的甘蔗，由于其糖分的分布呈下端高上端低的特点，特将其分割为上端和下端两部分，每部分作为一个单独的实验样本；对于伸长期的甘蔗则单独作为一个样本。最后得到甘蔗实验样本 130 个，剔除数据异常样本 7 个，最终获得 123 个有效样本。

1.2 方法

采用自行设计搭建的采集平台获取透射光谱数据，平台如图 1 所示，由光谱仪、光源、光纤、电脑等组成。光源在甘蔗蔗茎的一侧进行照射，另一侧采用光谱仪从不同角度对经过蔗茎的透射光进行检测。透射光谱的获取应保证探头与光源之间的夹角在 $(90^\circ, 180^\circ]$ 范围内，根据等角度间距的原则，确定了采集平台上三根不同角度的光谱探测通路，以此来获取 120° 、 150° 和 180° 情况下的甘蔗样本透射光谱。实验所采用的光谱仪为美国 ASD 公司生产的 FieldSpec 光谱仪（FieldSpec® 4 Hi-Res, USA），光谱范围 $350\sim2500\text{ nm}$ ，采样间隔为 1.4 nm ($350\sim1000\text{ nm}$) 和 1.1 nm ($1001\sim2500\text{ nm}$)。光源采用了一个 12 V 、 20 W 的卤素灯，采用直流稳压电源供电。为了避免光源侧的杂散光对透射光检测侧的干扰，在入射部位周围通过不透光的黑色海绵进行包裹，使得

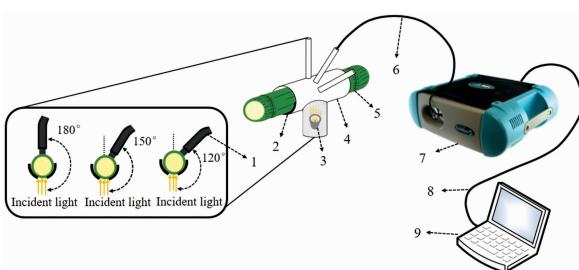


图 1 甘蔗透射光谱采集平台

1：探头；2：黑色海绵垫片；3：卤素灯；
4：样品固定桶；5：甘蔗样本；6：光纤；
7：光谱仪；8：数据线；9：计算机

Fig. 1 Transmission spectrum acquisition platform of sugarcane

1: Probe; 2: Black sponge pad; 3: Halogen lamp;
4: Sample fixing cylinder; 5: Sugarcane sample;
6: Optical fiber; 7: Spectrometer; 8: Cable; 9: Laptop

光源、入射光通路、甘蔗入射部位三者形成一个封闭的空间。测量时，ASD 光谱仪预热 30 min，并使用 10 mm 厚度的聚四氟乙烯板作为参考板，将光强数据转换为透光率。光谱数据的分析及建模分别在 Excel Professional2019 (Microsoft, USA) 及 MATLAB(R2019a, USA) 中进行。

1.3 光谱数据采集

每个甘蔗样本分别在原始状态（去蜡前）、去蜡后分别进行了透射光的测量。在去除蜡质前后，每个样本选择两处（大约在 $1/4$ 与 $3/4$ 位置处）测量点，每处通过 90° 旋转共测量得到 4 组透射光谱，共 8 组数据。去蜡前后的甘蔗及其光谱采集位点如图 2 所示。

将每个样本的 8 组数据取平均，得到测量方向与入射光方向夹角分别为 120° 、 150° 和 180° 的透射光谱。所得透射光谱波长范围为 $350\sim2500\text{ nm}$ ，但从实际数据来看，首尾波段的透射光谱幅值接近 0，且存在大量噪声，因此选取了噪声较小，幅值变化显著的 $670\sim950\text{ nm}$ 波段内的光谱数据为有效建模区域。



图 2 去蜡前后的甘蔗样本及透射光谱采集位点示意图

Fig. 2 The sugarcane samples before and after wax removal and the sketch of transmittance measurement positions

1.4 甘蔗蔗糖分测定

甘蔗样本的蔗糖分采用国标 GB/T 10499—2014 方法进行检测样液的配置与获取，并使用液相色谱仪（Thermo Fisher UltiMate 3000, USA）进行甘蔗内部蔗糖分含量测定，剔除错误样本后，共获得 123 个样本用于后续分析与建模。将所有样本按 $2:1$ 分成校正集 ($n=82$) 和预测集 ($n=41$)，二者的蔗糖分统计结果如表 1 所示，可以看出两个集合包含的数据分布趋势一致，有利于模型的建立与评价。

表 1 甘蔗样本蔗糖分统计

Table 1 Statistical characteristics of sucrose content in sugarcane samples

数据集	样本数量	最小值 /%	最大值 /%	平均值 /%	标准差
校正集	82	6.07	13.71	10.05	1.628 1
预测集	41	7.23	13.17	9.99	1.472 8

Note: % represents the unit of sucrose (the same below)

1.5 数据分析

使用 MATLAB R2019a 对表皮去蜡前后的甘蔗平均光谱进行预处理以及特征波长分析。对原始光谱进行预处理方法后，通过偏最小二乘法 (partial least squares, PLS) 回归建模的效果对比选择最优方法。接着采用了常规的间隔偏最小二乘 (interval partial least squares, i-PLS)、遗传算法 (genetic

algorithm, GA)、传统蚁群算法 (ant colony optimization, ACO) 对特征波长进行了提取；并提出了基于全波段 PLS 建模回归系数 (variable regression coefficient, VRC) 的改进蚁群算法 (VRC-ACO) 优化筛选效率。最后针对各方法筛选得到的特征波长建立了甘蔗糖分的 PLS 预测模型。模型的预测能力和稳定性由校正相关系数 (R_c)、预测相关系数 (R_p)、校正均方根误差 (RMSEC) 和预测均方根误差 (RMSEP) 这 4 个指标来评价。

2 结果与讨论

2.1 光谱预处理

由于可见-近红外光谱采集的有效信息弱，且带有噪声，为了提升后续建模的可靠性，需对原始光谱进行预处理。本文以 120° 光谱获取角度下的甘蔗光谱数据集为例，采用了 6 种方法分别对光谱进行了预处理，并采用 PLS 建模结果评价各预处理方法的优劣，结果如表 2 所示。

表 2 甘蔗样本的 120° 透射光谱预处理结果

Table 2 Spectral pretreatment results of sugarcane samples at 120° measurement angle

样本类别	预处理	校正集		预测集	
		R_c	RMSEC/%*	R_p	RMSEP/%*
原始样本	无处理	0.650	1.23	0.723	1.016
	SG	0.660	1.217	0.713	1.029
	FD	0.768	1.036	0.704	1.049
	MSC	0.804	0.963	0.779	0.943
	SNV	0.809	0.952	0.777	0.949
	SG+MSC	0.804	0.962	0.773	0.955
	SG+SNV	0.808	0.954	0.775	0.954
去蜡样本	无处理	0.634	1.251	0.749	0.967
	SG	0.703	1.151	0.760	0.950
	FD	0.784	1.005	0.790	0.899
	MSC	0.811	0.947	0.783	0.947
	SNV	0.821	0.925	0.791	0.933
	SG+MSC	0.811	0.947	0.780	0.957
	SG+SNV	0.817	0.934	0.787	0.943

由表 2 可知，在相同光谱测量角度下，SNV 预处理效果在校正集表现优于其他预处理方法，同时在预测集建模效果上，去蜡样本的预测集建模效果表现最好，原始样本的预测集建模效果仅比最高的 MSC 方法略低，其他角度下的预处理结果当中，SNV 整体表现也最优。综合分析可知，SNV 对以上数据集的干扰因素校正能力更强，更适用于本实验所对应的数据，故后续采用经过 SNV 预处理后的光谱数据进行建模处理。

2.2 不同甘蔗光谱数据测量方式的差异性分析

2.2.1 蜡质覆盖的影响

甘蔗在自然生长的情况下表皮会覆盖一层薄薄的植物蜡，和其他天然蜡质一样，是酯、游离酸、醇和碳氢化合物的混合物，同时蜡质覆盖也会改变甘蔗表皮颜色，因此本研究首先探讨了甘蔗表皮蜡质对甘蔗透射光谱预测糖分的

影响。

图 3(a) 所示为某样本在探测夹角为 180° 时，8 个采集位点去蜡前后的光谱曲线；其余夹角条件下，波形走势与此相似。可以看出，去蜡后的光谱透过率更高，不同位点的透射光谱变化也更小，使得数据稳定性有所提升。分析 6 类光谱数据与蔗糖分的相关性，结果如图 3(b) 所示，原始样本对应的 3 种角度下的透射光谱数据与蔗糖分的整体相关性较低，而去蜡后 3 种角度下两者的相关性有着明显的提升。同时由图 2 可见，未去蜡的甘蔗表面颜色分布不均匀，且颜色比去蜡后的深，是造成光谱差异的原因之一；另外，构成甘蔗表皮蜡的化学混合物质也会由于光照产生分子振动，从而产生光谱的倍频和合频吸收，进而影响透射光谱。

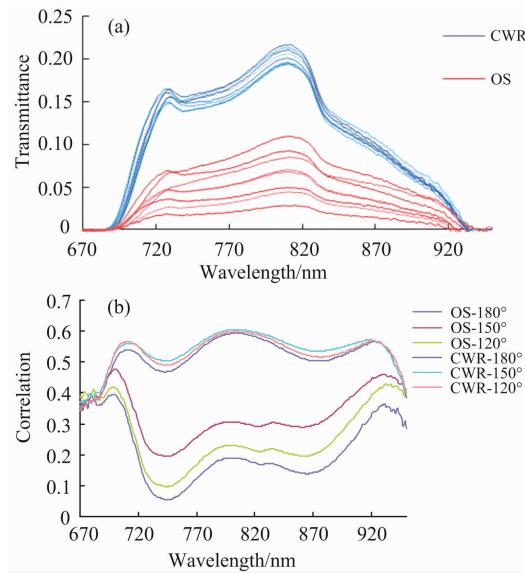


图 3 甘蔗去蜡前后的光谱差异

CWR：去蜡样本；OS：原始样本

(a)：某样本不同采集位点的透射光谱；

(b)：光谱透过率与蔗糖分的相关性

Fig. 3 Spectral difference of sugarcane before and after wax removal

CWR: Cane wax removed; OS: Original sample

(a)：Transmission spectra at different sampling positions of a sugarcane sample；(b)：Correlation between transmittance spectra and sucrose

2.2.2 不同探测角度的影响

光源与探测器的夹角不同，会影响光程长短，光线在甘蔗内部经过多角度的散射之后，携带着甘蔗内部信息的光强被光纤收集。图 4 所示为 120° 、 150° 和 180° 三个角度下所有样本的平均透射光谱，从图中的结果来看，去蜡前后均表现出 120° 夹角下的透过率最高， 150° 下的透过率略低，而 180° 下的透过率最低。

2.2.3 不同光谱测量方式对预测模型精度的影响

针对 6 类不同测量方式采集得到的光谱数据集分别建立了蔗糖分的 PLS 全波段预测模型，结果如表 3 所示。从蜡质覆盖来看，未去蜡数据集的预测模型 R_p 范围在 $0.75 \sim 0.77$ ，去蜡的 R_p 范围在 $0.77 \sim 0.79$ ，未去蜡的模型预测效果在整体

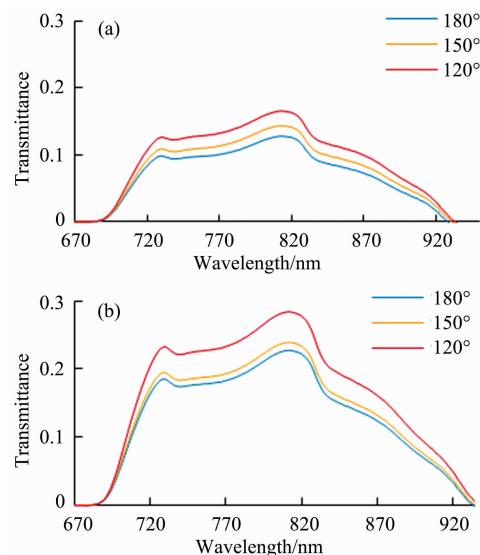


图 4 不同角度下样本的平均光谱透过率

(a): 原始样本; (b): 去蜡样本

Fig. 4 Average spectral transmittance of samples at different angles

(a): Original sample; (b): Cane wax removed

上都低于去蜡样本, 进一步证明蜡质会对建模精度造成影响。从不同的探测角度来看, 不论去蜡与否, 120° 下的建模效果都最佳, 未去蜡与去蜡的 R_p 分别为 0.776 8 和 0.790 6, RMSEP 分别为 0.948 6 和 0.933 4, 表明了光源与探测器的夹角大小也会影响透射光谱建模效果。

表 3 不同测量方式下的 PLS 建模结果**Table 3 PLS modeling results obtained by different spectral measurement methods**

样本类别	探测角度	校正集		预测集	
		R_c	RMSEC/%*	R_p	RMSEP/%*
原始样本	120°	0.809	0.952	0.777	0.949
	150°	0.810	0.949	0.752	1.010
	180°	0.792	0.989	0.765	0.967
去蜡样本	120°	0.821	0.925	0.791	0.933
	150°	0.818	0.932	0.775	0.981
	180°	0.816	0.935	0.772	0.971

2.3 特征波长的筛选

选取 6 类数据集中建模效果最好的 120° 探测夹角下经 SNV 预处理后的甘蔗去蜡透射光谱作为后续分析的数据集, 进一步对特征波长筛选方法进行研究与对比。

2.3.1 i-PLS 筛选特征波长

间隔偏最小二乘法的基本原理是将全波段区域进行区间距划分, 得到 n 个光谱区间, 而后对每个子区间内的光谱数据进行 PLS 建模分析。采用 i-PLS 对甘蔗透射光谱进行的波长筛选, 将 671~950 nm 的范围划分为 15 个区间, 在波段 747~765 nm 区间内 RMSECV 最低。i-PLS 挑选出的波长组合如表 4 所示。

2.3.2 GA 筛选特征波长

遗传算法是仿照生物自然选择和进化发展而来的自适应算法。GA 相关的参数设置如下: 初始化群体数目 200, 交叉概率 0.8, 变异概率 0.3, 迭代次数 50, 波长选择阈值为 5。算法运用前向选择的方式进行波段组合挑选, 为了防止算法的随机性对结果造成的影响, 共进行了 10 次重复运算, 挑选出建模效果最优的波长组合。GA 挑选出的波长组合如表 4 所示。

2.3.3 ACO 筛选特征波长

蚁群算法模拟了自然界当中蚁群的觅食行为, 根据路径上蚂蚁留下的信息素浓度收敛到寻找食物最短的路径。ACO 相关的参数设置如下: 初始蚂蚁数量 200, 信息素挥发引子 0.85, 显著性因子 0.1, 迭代次数 50 次, 波长选择阈值为 5。与遗传算法保持一致, 根据前向选择挑选波长组合, 并进行 10 次重复运算。ACO 挑选出的波长组合如表 4 所示。

2.3.4 VRC-ACO 筛选特征波长

传统的蚁群算法中每个波段对应的初始化信息素含量都为 1, 而统一的初始信息素含量可能导致算法收敛速度缓慢, 并且容易陷入局部最优解。因此提出一种改进的蚁群算法, 将全波段 PLS 建模当中的变量回归系数与传统蚁群算法结合, 以期改善 ACO 收敛速度慢等问题。VRC 能够在一定程度上体现每个波段在全波段 PLS 建模当中所占的重要程度, 其绝对值越高表明该变量对模型的影响越大。VRC-ACO 挑选出的波长如表 4 所示。

表 4 挑选波段结果**Table 4 Results of selected wavelengths**

特征提取方法	变量数	特征波长/nm
i-PLS	19	748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766
GA	16	677, 697, 737, 738, 741, 763, 775, 779, 780, 788, 790, 794, 833, 922, 941, 946
ACO	15	676, 703, 735, 747, 763, 768, 769, 771, 778, 786, 798, 792, 819, 821, 822
VRC-ACO	10	737, 741, 760, 762, 763, 764, 765, 780, 930, 949

从表 4 的结果来看, i-PLS 选取的特征波长数量最多, 达到了 19 个, GA 和 ACO 算法选择的特征波长数量分别为 16 个和 15 个, 而 VRC-ACO 挑选的波长数量最少, 仅有 10 个特征波长。从特征波长挑选的数量来看, VRC-ACO 的表现最优。同时相比算法改进之前, 采用 VRC-ACO 运行程序的速度与传统的 ACO 运行速度相比提升了将近 10%。图 5 为算法选择波长的频次图, 波长的挑选也有比较明显的高低峰值。

2.4 蔗糖分预测模型的建立

蔗糖分子的化学式为 $C_{12}H_{22}O_{11}$, 分子结构中含有亚甲基($-CH_2-$)和羟基($-OH$)等化学基团, 亚甲基在近红外光谱区域的四倍频吸收带在 762 nm 附近, 三倍频吸收带在

934 nm 附近^[12]; 羟基的三倍频在 980 nm 附近, 四倍频在 730 nm 附近^[13]。从表 4 的波长挑选结果可以看出, 各方法筛选得到的特征波长均包含亚甲基的四倍频吸收带 762 nm 附近的波长, 并规避了羟基倍频所对应的特征波长, 一定程度上反映了筛选结果的合理性。

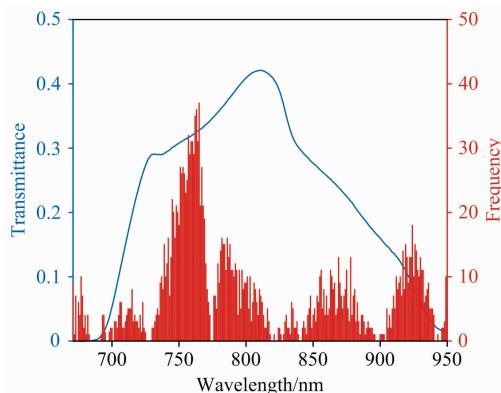


图 5 VRC-ACO 算法选择的波长频次

Fig. 5 Frequency of each wavelength selected by VRC-ACO algorithm

不同算法挑选得到的特征波长的建模结果如表 5 所示。基于 i-PLS 筛选波段的模型精度 ($R_p = 0.761\ 9$) 低于全波段的精度 ($R_p = 0.790\ 6$), 这是由于该方法选取的变量是按照连续区间进行选取, 特征波长的选取受到了较大的限制。其余 3 种筛选方法所得波段的建模精度 ($R_p > 0.84$) 均明显高于全波段的精度。从建模结果来看, 效果最好的为 VRC-ACO 方法, 该方法筛选得到的波段数最少, 仅为 10 个, 但由此建立的 PLS 模型的精度最高: R_p 达到了 0.861 6, 较全波段 PLS 预测模型提升了 8.98%; RMSEP 仅为 0.746 6, 较全波段 PLS 预测模型下降了 20%。基于 VRC-ACO 筛选得到的特征波段建立的 PLS 模型, 其校正集与预测集的散点图如图 6 所示。

表 5 不同算法挑选波长的建模结果

Table 5 Modeling results of selected wavelengths by different algorithms

特征提取方法	变量数	校正集		预测集	
		R_c	RMSEC/%*	R_p	RMSEP/%*
None	280	0.821	0.935	0.791	0.933
i-PLS	19	0.746	1.082	0.762	1.167
GA	16	0.844	0.869	0.843	0.790
ACO	15	0.848	0.858	0.840	0.798
VRC-ACO	10	0.852	0.849	0.862	0.747

3 结 论

通过自行搭建的甘蔗透射光谱采集平台获取了活体甘蔗样本去蜡前后在 120°、150° 和 180° 测量夹角下的 6 类透射光

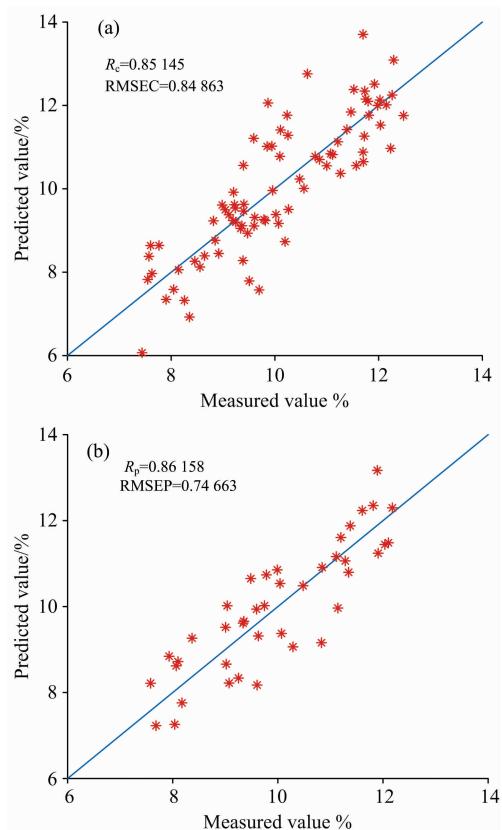


图 6 VRC-ACO 挑选的波长所建立的模型散点图

(a): 校正集; (b): 验证集

Fig. 6 Scatter plot of the model established by the wavelength selected by VRC-ACO algorithm

(a): Calibration set; (b): Prediction set

谱, 分析了光谱特征差异, 对比并改进了特征波长筛选方法, 最后建立了蔗糖分的预测模型。主要结论如下:

(1) 比较了 S-G 平滑处理、FD、MSC、SNV、S-G+MSC 和 S-G+SNV 这 6 种光谱预处理方法对模型精度优化的能力, 发现 SNV 综合表现最佳。

(2) 甘蔗表皮蜡质会对光谱透射产生干扰, 降低建模精度, 同时光源与探测器的角度变化也会影响建模效果。不同测量方式下, 120° 夹角下去蜡后的光谱数据集建立的预测模型精度最好, 预测集相关系数为 0.790 6, 预测均方根误差为 0.93。

(3) 比较了 i-PLS, GA, ACO 和本工作改进得到的 VRC-ACO 四种特征提取算法的特征波长建模预测效果, 提取出的特征波长数量分别为 19, 16, 15 和 10。基于 VRC-ACO 算法挑选出的 10 个特征波段建立的 PLS 模型预测蔗糖分的效果最优, 预测集相关系数为 0.861 6, 比全波段 PLS 的结果提升了 8.98%; 预测均方根误差为 0.746 6, 比全波段 PLS 降低了 20%; 模型输入变量减少了 96.4%。该研究为甘蔗蔗糖分无损检测及相应传感器的研发提供了理论支持。

References

- [1] LI Xuan, GAN Ning-jun, WEN Tao, et al(李 喆, 干宁军, 温 铊, 等). Applied Chemical Industry(应用化工), 2017, 46(10): 2023.
- [2] Assis C, Ramos R S, Silva L A, et al. Applied Spectroscopy, 2017, 71(8): 2001.
- [3] Gersdorff G J E V, Kirchner S M, Hensel O, et al. Meat Science, 2021, 178(4): 108525.
- [4] Minaei S, Shafiee S, Polder G, et al. Infrared Physics and Technology, 2017, 86: 218.
- [5] Liang P S, Haff R P, Hua S, et al. Biosystems Engineering, 2018, 166: 161.
- [6] Kamboj U, Guha P, Mishra S. Analytical Letters, 2017, 50(11): 1754.
- [7] Martínez-Arias R, Müller B U, Schechert A. Sugar Tech, 2017, 19(5): 526.
- [8] Li M, Han D, Liu W. Biosystems Engineering, 2019, 188: 31.
- [9] Xu S, Lu H, Ference C, et al. Biosensors, 2020, 10(4): 41.
- [10] Taira E, Ueno M, Saengprachatanarug K, et al. Journal of Near Infrared Spectroscopy, 2013, 21(4): 281.
- [11] SteidleNeto A J, Lopes D C, Toledo J V, et al. The Journal of Agricultural Science, 2018, 156(4): 537.
- [12] LU Wan-zhen(陆婉珍). Modern Near Infrared Spectroscopy Analytical Technology(现代近红外光谱分析技术). Beijing: China Petrochemical Press(北京: 中国石化出版社), 2006. 30.
- [13] HUANG Li-xian, LI Da-peng, WU Fan, et al(黄立贤, 李大鹏, 吴 凡, 等). High Power Laser and Particle Beams(强激光与粒子束), 2018, 30(1): 186.

A Method for Detecting Sucrose in Living Sugarcane With Visible-NIR Transmittance Spectroscopy

LÜ Xue-gang^{1, 2}, LI Xiu-hua^{1, 2*}, ZHANG Shi-min², ZHANG Mu-qing¹, JIANG Hong-tao¹

1. Guangxi Key Laboratory of Sugarcane Biology, Guangxi University, Nanning 530004, China

2. School of Electrical Engineering, Guangxi University, Nanning 530004, China

Abstract Sucrose content is an important indicator to measure the quality of sugarcane. It is of great significance to study the non-destructive detection method of sucrose content in living sugarcane based on the principle of spectroscopy. Sugarcane has a cylindrical shape with a hard skin and waxy surface. Different spectra detection angles and surface conditions will affect the modeling results to a certain extent. In addition, dimension reduction of characteristic wavelengths extraction is another factor that affects the model's accuracy. In this study, the effect of different spectral measurement styles on the accuracy of the sucrose content prediction model was evaluated, an improved method for characteristic wavelengths extraction was proposed, and a sucrose prediction model was eventually constructed. A spectra acquisition platform was designed to obtain the transmittance spectra of sugarcane stalks. When acquiring the transmittance spectra, there were three different acquisition angles (120° , 150° and 180°) between the incident light and the measurement probe and two surface conditions (wax-not-removed and wax-removed). Six data sets of 123 samples were obtained in total. Firstly, PLS modeling result was used to evaluate the effect of different spectral pretreatment methods, including S-G smoothing, standard normal variation (SNV), multiplicative scatter correction (MSC), first derivation (FD), etc., and the result showed that SNV had the best comprehensive performance and was selected for further study. Then the effect of different measurement styles on the modeling of sucrose content was evaluated. The result found that: (1) regarding the effect of wax coverage, the spectral transmittance after wax-removed was high, the spectral difference among different collection sites of a single sample was lower, and the correlation with sucrose was much higher; (2) regarding the effect of the spectra acquisition angle, the transmittance decreased as the angle increased in a certain range; (3) the best modeling result was obtained ($R_p = 0.790$ 6, RMSEP = 0.898 6) with the measurement style of wax-removed and 120° measurement angle. Finally, the interval partial least squares method (i-PLS), genetic algorithm (GA), ant colony algorithm (ACO) and an improved ant colony algorithm (VRC-ACO) based on full wavelengths PLS modeling variable regression coefficient proposed in this study were used to extract the characteristic wavelengths. The result showed that the number of characteristic wavelengths selected by the VRC-ACO algorithm, which had only ten wavelengths, was the least, yet the prediction accuracy was the best ($R_p = 0.861$ 6, 9.0% higher than the full-band model; RMSEP = 0.746 6, 20.0% lower than the full-band model). This research provides theoretical support for the non-destructive detection of sugarcane and the development of corresponding sensors.

Keywords Visible-Near infrared spectroscopy; Sugarcane; Sucrose content; Characteristic wavelengths; Measurement style

* Corresponding author

(Received Jun. 7, 2021; accepted Sep. 25, 2021)