doi:10.6041/j.issn.1000-1298.2026.04.012

基于可见-近红外光谱技术的阳光玫瑰糖度在线检测方法

臧　宇[1,2]　张译之[2]　边鸿迪[2]　郝浩源[2]　李　楠[1]　李江波[2]
（1.上海大学机电工程与自动化学院，上海 200444；2.北京市农林科学院智能装备技术研究中心，北京 100097）

摘要：基于可见-近红外光谱技术，开展了阳光玫瑰葡萄糖度的在线快速检测研究。系统分析了光谱采集速度、采集形式及预处理方法对糖度预测模型的影响，并在保证精度的前提下实现模型简化。数据采集中，将整串葡萄分为前、中、后 3 个区域采集光谱并取平均作为整串光谱信息。采用偏最小二乘回归（PLSR）和支持向量回归（SVR）建立模型，结果表明，Savitzky－Golay 平滑结合标准正态变换（SG＋SNV）预处理效果最佳，显著提升信噪比与预测准确性。整串检测优于分段检测，SVR 模型预测均方根误差为 0.49°Brix、相关系数为 0.91，PLSR 模型预测均方根误差为 0.45°Brix、相关系数为 0.94。随着光谱采集速度从 0.15 m/s 增至 0.6 m/s，预测精度逐渐降低。特征选择方面，无信息变量消除（UVE）在 980 个波长中筛选出 314 个有效波长，构建的 PLSR 模型预测均方根误差为 0.41°Brix、相关系数为 0.90，显著降低模型复杂度。综上，基于可见-近红外光谱的整串葡萄糖度在线检测方法具有较高准确性与稳定性，为串状水果的商业化在线分选提供了理论与方法参考。

关键词：近红外光谱分析；在线检测；阳光玫瑰葡萄；糖度分析；模型构建
中图分类号：TP391　　**文献标识码**：A　　**文章编号**：1000-1298(2026)04-0119-09　　**OSID**：

Online Detection Method of Shine Muscat Grape Brix Based on Visible－Near Infrared Spectroscopy Technology

ZANG Yu[1,2]　ZHANG Yizhi[2]　BIAN Hongdi[2]　HAO Haoyuan[2]　LI Nan[1]　LI Jiangbo[2]
(1. *School of Mechatronic Engineering and Automation*, *Shanghai University*, *Shanghai* 200444, *China*
2. *Intelligent Equipment Research Center*, *Beijing Academy of Agricultural and Forestry Sciences*, *Beijing* 100097, *China*)

Abstract: Rapid and non-destructive method for online determination of soluble solids content (SSC) in Shine Muscat grapes was developed by using visible－near infrared (Vis－NIR) spectroscopy. The influences of spectral acquisition speed, sampling mode, and preprocessing methods on model performance were systematically investigated, and model simplification was achieved while maintaining prediction accuracy. During spectral acquisition, each grape bunch was divided into three equal segments (front, middle, and rear), and the averaged spectra were used to represent the overall optical response of the whole bunch. Partial least squares regression (PLSR) and support vector regression (SVR) models were constructed to evaluate the effects of different preprocessing approaches. Among them, the combination of Savitzky－Golay smoothing and standard normal variate transformation (SG＋SNV) yielded the best results by effectively correcting baseline drift and scattering noise, thereby enhancing spectral quality and model precision. Whole-bunch detection achieved superior performance compared with segmented detection, with SVR producing root mean square error of prediction of 0.49°Brix and correlation coefficient of 0.91, and PLSR yielding root mean square error of prediction of 0.45°Brix and correlation coefficient of 0.94. As the acquisition speed was increased from 0.15 m/s to 0.6 m/s, the prediction accuracy was gradually declined. For wavelength selection, uninformative variable elimination (UVE) demonstrated the best performance, extracting 314 informative wavelengths from 980 while maintaining high accuracy (PLSR: root mean square error of prediction was 0.41°Brix, correlation coefficient was 0.90) and reducing model complexity. Overall, the proposed Vis－NIR-based approach enabled accurate and stable online prediction of grape SSC at the whole-bunch level, offering a practical

收稿日期：2025－09－23　修回日期：2025－11－26
基金项目：国家自然科学基金项目（W2412103）和北京市农林科学院财政追加专项（CZZJ202501）
作者简介：臧宇（1997—），男，硕士生，主要从事农产品品质快速无损检测研究，E-mail：mailto:13296382052@163.com
通信作者：李江波（1982—），男，研究员，博士生导师，主要从事农产品无损检测分级和智能包装研究，E-mail：lijb@nercita.org.cn

theoretical foundation for the commercial development of intelligent online sorting systems for bunch-type fruits.

Key words: near-infrared spectroscopy; online detection; shine muscat grape; brix analysis; model specification

0 引言

根据联合国粮食及农业组织(FAO)统计数据,2023 年中国葡萄年产量达 1.429×10^7 t,约占全球总产量的 21.6%,其中鲜食葡萄占比 70%。作为鲜食品种的重要代表,阳光玫瑰以年增长率超过 15% 的速度迅速占领市场。与此同时,高端葡萄市场(单价超过 30 元/kg)消费量保持 12.3% 的年均增速,显著体现出消费升级趋势。糖度作为葡萄品质的核心指标,亟需为种植端与流通端提供精准化检测手段。传统破坏性检测方法流程繁琐且难以实现无损批量化检测,因此,快速、无损检测技术开发对促进葡萄产业发展具有重要意义。

近年来,可见-近红外(Vis-NIR)技术因其无损、快速、低成本优势,已成为葡萄内部品质评价的主流方案。多项研究表明,结合合理的预处理与化学计量以及机器学习建模,Vis-NIR 可对可溶性固形物(SSC)、可滴定酸(TA)、pH 值、成熟度等关键指标进行稳健预测,并具备向现场与在线场景迁移的潜力。例如,SU 等[1]针对储藏期鲜食葡萄以 Vis-NIR + 特征波长选择 + PLSR/SVR 快速量化 SSC,预测相关系数约 0.94,验证了流通中应用的可行性;ZHOU 等[2]提出基于 Vis-NIR 的“综合质量指标(AQI)”构建思路,提升多指标聚合评价能力;ELSHERBINY 等[3]以 Vis-NIR 结合梯度提升等算法,同时预测 TSS、TA 与成熟度比值(TSS/TA),体现了机器学习在光谱建模中的增益;刘文政等[4]利用可见-近红外光谱技术对鲜食葡萄成熟品质关键指标进行了无损检测;ZHANG 等[5]开发手持式 IoT Vis-NIR 系统实现葡萄 SSC 的实时现场检测,支撑大规模应用落地;LYU 等[6]利用田间近红外场景光谱实现 TSS 的原位无损估计,进一步缩小了实验室与生产场景的差距。尽管如此,面向成串葡萄与分选线的大批量在线检测仍需在光路稳定性、遮挡散射抑制与跨批次鲁棒建模方面继续突破,这也印证了开展基于 Vis-NIR 的在线糖度快速检测研究的必要性与工程价值。

基于可见-近红外光谱技术,本文开展阳光玫瑰葡萄糖度在线快速检测研究。系统分析光谱采集速度、采集方式及预处理方法对糖度预测模型的影响,在确保精度的前提下对模型进行简化;并研究不同近红外光谱数据预处理方法对预测准确性的效果。以期为基于光谱技术的葡萄糖度快速在线检测提供理论支持,并为商业化分选系统开发提供参考。

1 实验与方法

1.1 葡萄样本

以阳光玫瑰为研究对象,样本于 2025 年 4 月购自北京市丰台区新发地农产品批发市场。单串样本长度为 200 ~ 220 mm,宽度为 100 ~ 120 mm。共使用 96 串葡萄样本。实验前,样本经纯净水冲洗并擦拭干净后,编号并置于实验室(温度 20℃)环境下 24 h,以减小温度对光谱数据的影响,样本如图 1 所示。



图 1 葡萄实验样本

Fig. 1 Experimental grape samples

1.2 光谱数据采集

所采用可见-近红外漫透射在线采集系统如图 2 所示。该系统主要包括减速电机、输送带、光电传感器、编码器、SE2120-300-VNIR 型光谱仪(光谱范围 350 ~ 1 020 nm),卤素灯(USHIO,JCR,15 V,60 W)、光箱和计算机等。数据采集时,光谱仪的积分时间设置为 3 ms,卤素灯分布在传送带两侧。光纤探头分布在传送带下方,通过葡萄托盘下方通光孔接收光谱信息。每串葡萄样本平放在托盘上,分别以 0.15、0.3、0.6 m/s 3 个恒定速度分 3 次通过光箱并由计算机记录光谱信息。由于采用短积分,且

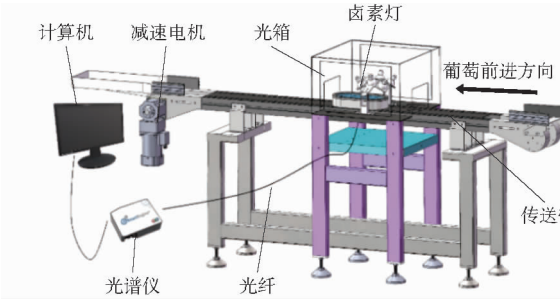

图 2 近红外光谱采集系统

Fig. 2 Near-infrared spectral acquisition system

葡萄样本较长，葡萄样本在通过光谱仪光纤头时，葡萄将按前、中、后 3 个区域等分，分 3 次测得多条漫透射光谱曲线[7]。

1.3　理化实验

葡萄样本光谱数据采集完成后，将整串葡萄按长度均匀划分为前、中、后 3 个区域，分别取各区域全部葡萄颗粒，并使用手压式榨汁器分别进行榨汁处理。随后取少量果汁滴于日本爱拓糖度计（ATAGO PAL－1 型）测量窗口，每个区域样本分别测量 2 次，以减小测量误差，取 2 次测量结果平均值作为该区域糖度。将 3 个区域榨取的果汁充分混合均匀后，采用同样方法进行 2 次测量，并取平均值作为整串葡萄糖度。为保证测量结果准确性，每次测量结束后均使用纯净水彻底冲洗糖度计液体槽，并用无纺布擦拭干净后再进行下一次测量。

1.4　异常样本剔除

光谱数据具有较多波段和较高维度，直接对原始数据进行分析易引入大量冗余信息和噪声，采用主成分分析（Principal component analysis，PCA）法[8]将高维数据降维至二维特征空间，保留原始数据主要信息的同时减少冗余和噪声影响；采用马氏距离（Mahalanobis distance）法[9]对样本数据进行异常检测。马氏距离能够考虑特征变量之间的相关性及协方差结构，计算每个样本点在二维主成分特征空间中到总体均值的距离以表征样本偏离总体中心的程度。马氏距离值越大，表示该样本与总体数据的差异越明显，说明其可能是潜在的异常数据。此处采用蒙特卡罗阈值法[10]，以置信水平 99% 确定异常检测的阈值。当自由度为 2、置信水平 99% 时其对应的阈值为 3.03。当样本点马氏距离超过此临界阈值时，即被判定为异常样本。基于上述方法，共对 3 个不同检测区域、3 种不同速度条件下共 864 个光谱样本点进行异常检测分析，共有 25 个样本点马氏距离超过设定阈值被识别为异常数据。图 3 为异常数据剔除的可视化结果。

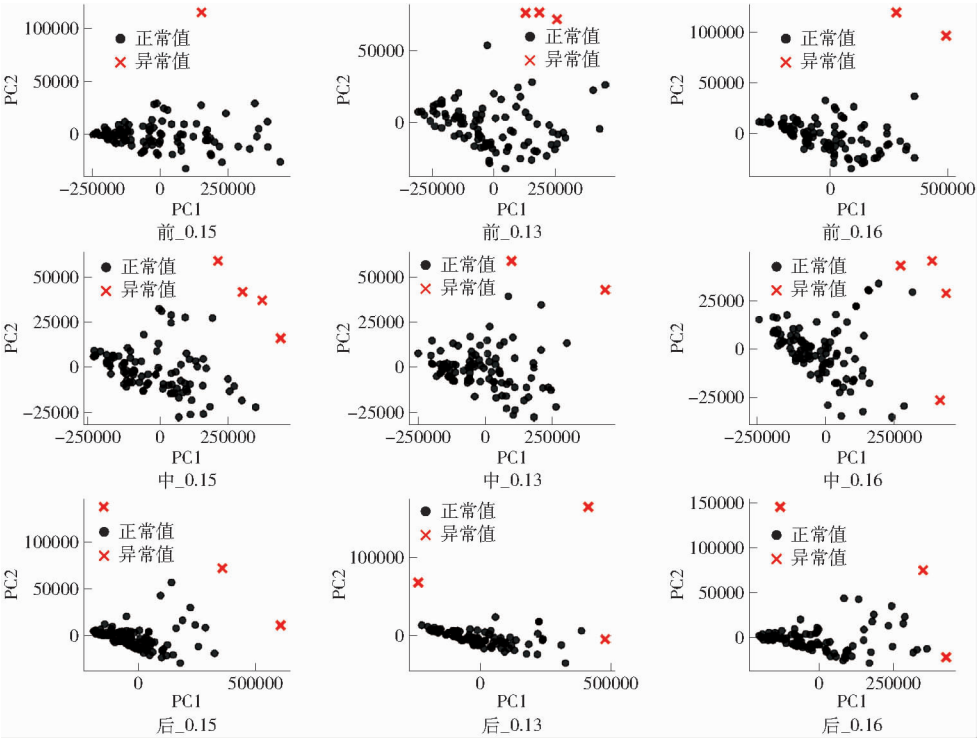

图 3　异常值剔除可视化结果

Fig. 3　Outlier removal results

1.5　光谱数据预处理

近红外光谱数据的预处理是构建稳健、高精度定量分析模型的关键基础环节。原始光谱数据通常包含大量与分析目标无关的信息，如杂散光干扰、仪器噪声及样品物理特性差异等，这些因素会显著降低模型的预测性能。采用 6 种不同预处理方法：多元散射校正（Multiplicative scatter correction，MSC）、标准正态变换（Standard normal variate transformation，SNV）、Savitzky－Golay（SG）平滑、标准化、SG 平滑与 SNV 组合法以及 SG 平滑与 MSC 组合法[11]。不同预处理方法的作用不同。针对葡萄颗粒间缝隙导致的入射光散射效应，MSC 和 SNV 可有效校正由此引起的基线漂移（Baseline shift）和乘性散射干扰（Multiplicative scatter effects）。针对光谱采集过程中不可避免的电子噪声及环境杂光干扰，SG 平滑基于局部多项式最小二乘拟合原

理,在保留原始信号特征(如峰位、峰高、峰形)的前提下,通过滑动窗口拟合替代原始值,显著提升信噪比(Signal-to-noise ratio, SNR)。相较于移动平均(Moving average)等平滑方法,SG 平滑在特征保留能力上具有显著优势。标准化可消除因异常样本产生的异常影响并提升模型的泛化能力。图 4 为原始光谱与经 SG、SNV 预处理后的光谱对比。由图 4 可以看出,原始光谱在经预处理后,光谱曲线更为平滑,光谱基线保持一致,消除各条光谱之间的强度差异。

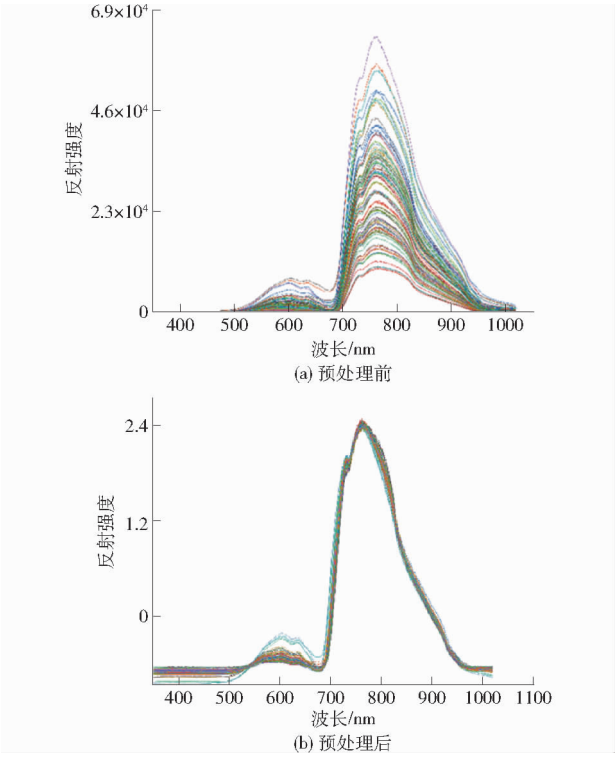

图 4 预处理前后对比结果

Fig. 4 Comparison before and after preprocessing

1.6 数据集划分

为构建稳健的糖度预测模型,采用(Kennard - Stone, K - S)算法以比例 3∶1 划分校正集与预测集[12]。分别在 3 种传送带速度(0.15、0.30、0.60 m/s)和4 种检测区域(前部、中部、后部、整串)下,统计校正集与预测集的样本数量及糖度范围情况,具体如表 1 所示。在校正集中,每个速度和区域均包含 69 个样本,糖度范围分布较广,覆盖从低至高的多个水平,均显示出良好的覆盖性。在预测集中,每个区域均包含 23 个样本,预测集糖度范围完全被校正集所覆盖,确保模型在未见样本上的有效性和代表性。对比校正集与预测集糖度平均值与标准偏差也表明两者在分布上的一致性。

表 1 校正集与预测集划分

Tab. 1 Division of calibration and prediction sets

	积分区域	0.15 m/s				0.30 m/s				0.60 m/s			
		前部	中部	后部	整串	前部	中部	后部	整串	前部	中部	后部	整串
校正集	最大值/°Brix	17.9	17.3	17.3	16.7	17.0	18.0	19.4	17.0	18.0	18.8	17.4	16.8
	最小值/°Brix	12.1	12.9	10.0	12.3	11.2	12.0	12.6	12.1	12.8	12.0	12.6	12.9
	平均值/°Brix	14.8	14.7	15.1	14.8	14.8	15.1	14.9	14.7	15.2	14.9	14.8	14.9
	标准偏差/°Brix	1.2	1.0	1.2	1.1	1.1	1.1	1.1	1.4	1.1	0.9	0.9	1.0
预测集	最大值/°Brix	17.3	16.2	17.3	16.8	17.8	17.0	16.4	17.1	16.7	16.4	17.2	17.4
	最小值/°Brix	13.3	13.5	13.0	12.5	13.0	13.3	13.1	12.4	12.7	12.7	13.7	12.9
	平均值/°Brix	14.8	14.7	15.4	14.8	14.7	15.4	14.7	14.7	15.0	14.9	15.2	14.9
	标准偏差/°Brix	1.1	0.6	0.8	1.0	1.0	1.0	0.8	1.2	1.0	0.8	0.8	1.1

1.7 数据建模

将分别建立偏最小二乘回归(Partial least squares regression,PLSR)和支持向量回归(Support vector regression, SVR)模型[13]2 类模型。PLSR 算法是一种融合主成分分析与多元线性回归优势的监督学习方法,专为处理高维、多重共线性的小样本数据而设计。该方法通过提取隐变量(Latent variables, LVs)捕捉自变量(光谱数据)与因变量(糖度信息)之间的协同变化特征,从而建立稳健的线性定量模型。与之相比,SVR 算法通过径向基核函数(RBF)将近红外光谱数据映射到高维特征空间,并在该空间内构建具有最大间隔的非线性回归超平面,以有效表征光谱数据与糖度浓度之间的复杂非线性关系。采用网格搜索法与五折交叉验证,优化 SVR 的关键参数-正则化系数(C)和核函数参数(γ),以充分平衡模型的拟合能力和泛化性能。与线性方法(如 PLSR)相比,SVR 模型能更好地适应小样本、高维数据的非线性特性。

1.8 评价指标

在建模分析时,使用相关系数(R)和均方根误差(RMSE)作为评价指标。模型评估中,相关系数(R)是衡量模型预测值与实际值之间线性相关性指标,R 越接近 1,模型拟合度越高。校正均方根误差(RMSEC)和预测均方根误差(RMSEP)分别衡量模型在训练集与预测集误差,RMSEC 通常小于

RMSEP，因为模型在训练数据上往往拟合更好。R 与 RMSEC、RMSEP 呈负相关关系，随 R（即训练集相关系数 R_c）增大，RMSEC 和 RMSEP 减小，表明模型在训练和预测阶段精度提高。预测集相关系数（R_p）反映模型泛化能力，R_p 越高，模型对未知数据预测效果越好，误差减小。较高的 R_c 和 R_p 通常伴随着较低的 RMSEC 和 RMSEP，表示模型整体预测性能较强。

2　结果与分析

2.1　不同传送速度对模型影响

在 3 种速度 0.15、0.3、0.6 m/s 下，分别对前、中、后 3 个积分区域建立 PLSR 线性模型与 SVR 非线性模型。由于 SVR 模型对特征尺度高度敏感，在建立非线性模型前对所有数据进行了标准化处理。结果如表 2 所示。

表 2　不同速度预测精度对比

Tab.2　Comparison of prediction accuracy at different speeds

积分区域	模型	速度/(m·s^{-1})	LVs	γ	C	RMSEC/°Brix	RMSEP/°Brix	R_c	R_p
前部	PLSR	0.15	5			0.43	0.48	0.87	0.82
		0.30	6			0.45	0.56	0.85	0.75
		0.60	5			0.49	0.61	0.91	0.71
	SVR	0.15		0.01	100	0.37	0.52	0.94	0.89
		0.30		0.01	100	0.42	0.50	0.93	0.88
		0.60		0.10	10	0.49	0.54	0.91	0.82
中部	PLSR	0.15	5			0.33	0.41	0.90	0.85
		0.30	5			0.34	0.45	0.91	0.85
		0.60	7			0.47	0.58	0.84	0.78
	SVR	0.15		0.10	100	0.30	0.36	0.95	0.90
		0.30		0.01	100	0.33	0.43	0.92	0.88
		0.60		0.10	100	0.40	0.51	0.90	0.85
后部	PLSR	0.15	5			0.40	0.44	0.91	0.86
		0.30	5			0.45	0.53	0.85	0.80
		0.60	7			0.56	0.73	0.83	0.76
	SVR	0.15		0.01	100	0.45	0.48	0.90	0.84
		0.30		0.10	10	0.43	0.49	0.91	0.88
		0.60		0.01	100	0.58	0.62	0.87	0.71

葡萄样本在光谱采集过程中始终处于连续运动状态，当样品运动速度提高时，光谱系统的积分采集时间相应缩短，导致单位时间内进入探测器的光子数量减少，从而降低光谱信号的信噪比。此外，随着样品运动速度增加，机械振动及随机噪声也会显著增强，这些因素共同作用，加剧了光谱数据中的随机误差和系统误差，进而对最终预测精度产生不利影响[14]。由表 2 可知，当葡萄样品运动速度由 0.3 m/s 提升至 0.6 m/s 时，模型预测精度出现了明显下降。这一现象表明速度过高可能严重影响光谱数据可靠性，进而制约葡萄在线分选的有效性。相比之下，当运动速度从 0.15 m/s 提升至 0.3 m/s 时，尽管检测精度有所降低，但预测精度仍保持在可接受范围内，考虑到在实际商业化分选场景中较高的分选速度对于提高生产效率具有重要实际意义，综合预测精度与生产效率之间的平衡关系，在后续分析中选取采集速度 0.3 m/s 来进一步探讨和评估不同光谱预处理方法对各类建模方法预测精度的影响。同时，选取预测表现最佳的模型开展进一步的参数优化研究，以期进一步提高在线分选的准确性与稳定性。

2.2　不同预处理方法下模型对比

对 5 种不同预处理方式进行对比，探究葡萄样本在光谱采样速度 0.3 m/s 下对葡萄串样本前、中和后 3 个积分区域预测精度的影响，结果如表 3 所示。

从表 3 可以看出，在所有评估的 5 种预处理方式中，首先对原始光谱数据进行 SG 平滑处理，再进行 SNV 联合处理方式表现出最佳的预测精度。这主要是由于葡萄串中颗粒间存在明显的散射效应，在动态采集光谱时，这种效应更为突出，因此对这种样本进行散射矫正预处理是必不可少的[15]。需要注意的是，若直接对原始光谱数据进行 SNV 或 MSC 散射矫正而省略了 SG 平滑处理步骤，预测效果反而有所下降。造成这种现象的原因可能是原始光谱中存在一定程度高频噪声，而 SNV 与 MSC 2 种方法都涉及到对单个样本光谱的整体归一化操作，这种处理方式可能会在一定程度上放大局部噪声的影

表 3 不同预处理方式对比结果

Tab. 3 Comparison of different preprocessing methods

积分区域	预处理	模型	LVs	γ	C	RMSEC/°Brix	RMSEP/°Brix	R_c	R_p
前部	无	PLSR	6			0.45	0.56	0.85	0.75
		SVR		0.01	100	0.42	0.50	0.93	0.88
	SG	PLSR	5			0.43	0.52	0.86	0.78
		SVR		0.10	100	0.42	0.51	0.92	0.87
	SNV	PLSR	5			0.41	0.49	0.89	0.80
		SVR		0.01	100	0.35	0.46	0.91	0.87
	MSC	PLSR	7			0.33	0.45	0.90	0.80
		SVR		0.01	100	0.46	0.41	0.93	0.81
	SG + SNV	PLSR	5			0.32	0.45	0.89	0.84
		SVR		0.10	100	0.49	0.54	0.86	0.81
	SG + MSC	PLSR	6			0.32	0.41	0.91	0.81
		SVR		0.10	100	0.37	0.63	0.93	0.89
中部	无	PLSR	5			0.34	0.45	0.91	0.85
		SVR		0.01	100	0.33	0.43	0.92	0.88
	SG	PLSR	5			0.27	0.40	0.90	0.85
		SVR		0.10	100	0.50	0.75	0.87	0.81
	SNV	PLSR	8			0.35	0.44	0.91	0.89
		SVR		0.01	100	0.38	0.47	0.95	0.88
	MSC	PLSR	7			0.43	0.58	0.86	0.79
		SVR		0.01	100	0.38	0.50	0.93	0.88
	SG + SNV	PLSR	5			0.35	0.42	0.97	0.94
		SVR		0.10	100	0.32	0.47	0.95	0.90
	SG + MSC	PLSR	6			0.42	0.56	0.89	0.82
		SVR		0.10	100	0.41	0.62	0.94	0.88
后部	无	PLSR	5			0.45	0.53	0.85	0.80
		SVR		0.10	10	0.43	0.49	0.91	0.88
	SG	PLSR	5			0.36	0.48	0.95	0.91
		SVR		0.10	100	0.62	0.36	0.87	0.82
	SNV	PLSR	8			0.43	0.56	0.83	0.77
		SVR		0.01	100	0.33	0.61	0.89	0.73
	MSC	PLSR	7			0.41	0.49	0.93	0.87
		SVR		0.01	100	0.42	0.64	0.85	0.79
	SG + SNV	PLSR	5			0.28	0.42	0.96	0.91
		SVR		0.10	100	0.35	0.47	0.92	0.86
	SG + MSC	PLSR	6			0.49	0.57	0.88	0.80
		SVR		0.10	100	0.33	0.63	0.91	0.84

响。因此,为达到更好地散射矫正效果,在散射矫正之前应对光谱数据进行平滑处理,以减少高频噪声干扰。此外,进一步探讨了不同预处理方式对整串葡萄糖度预测模型精度的影响,结果如表 4 所示。

在对前、中、后 3 个积分区域的光谱进行多次积分并实施光谱平均融合建模后,5 种预处理方式均在不同程度提升了模型的预测能力。其中,仍以 SG 平滑联合 SNV 矫正的方式表现最佳。此外,采用多段光谱融合的方法不仅能够实现令人满意的预测精度,同时也表现出比单纯的分段积分模型更高的鲁棒性,更适合实际应用中的动态葡萄糖度快速检测需求。

PLSR 与 SVR 2 种模型对葡萄糖度预测都有不错的预测精度,相比而言,无论是分段预测还是整串预测,PLSR 模型的预测精度和稳定性略高于 SVR 模型。这说明在全光谱建模的前提下,光谱信息与糖度信息存在一定线性关系;另一方面,SVR 模型对噪声的敏感程度高于 PLSR 模型。在 SG + SNV 预处理前提下,PLSR 模型预测前部的R_p和 RMSEP 为 0.84 和 0.45°Brix,中部的R_p和 RMSEP 为 0.94 和 0.42°Brix,后部的R_p和 RMSEP 为 0.91 和 0.52°Brix,整果预测的R_p和 RMSEP 为 0.94 和 0.45°Brix,SVR 模型的R_p和 RMSEP 分别为前部 0.88 和 0.41°Brix,中部为 0.89 和 0.42°Brix,后部为 0.82 和 0.57°Brix,

表 4　整串葡萄糖度预测

Tab. 4　Whole-bunch grape sugar content prediction

预处理	模型	LVs	γ	C	RMSEC/°Brix	RMSEP/°Brix	R_c	R_p
无	PLSR	5			0.31	0.46	0.94	0.88
	SVR		0.01	100	0.43	0.51	0.94	0.85
SG	PLSR	5			0.34	0.45	0.93	0.88
	SVR		0.10	100	0.50	0.61	0.92	0.88
SNV	PLSR	8			0.33	0.47	0.93	0.86
	SVR		0.01	100	0.41	0.57	0.91	0.89
MSC	PLSR	7			0.34	0.48	0.93	0.85
	SVR		0.01	100	0.39	0.55	0.94	0.89
SG + SNV	PLSR	5			0.34	0.45	0.97	0.94
	SVR		0.10	100	0.39	0.49	0.95	0.91
SG + MSC	PLSR	6			0.35	0.47	0.93	0.87
	SVR		0.10	100	0.45	0.66	0.90	0.82

整串预测的R_p和 RMSEP 为 0.91 和 0.49°Brix。对于分段预测而言，中部预测的效果略优于前部与后部预测，其原因为葡萄串中部果实比前部与后部更为紧实，杂散光更不易透过缝隙从而引入噪声，分段预测相比整串预测，整串预测在具备良好的预测精度的同时稳定性略优于分段预测。这是由于多条积分光谱进行平均融合时，高频噪声得到了有效抑制，从而提高了光谱信号信噪比，进一步提升了模型预测精度。对于在线快速检测，多条近红外光谱融合预测整串葡萄糖度的方式有更好的鲁棒性，并能从全局反映葡萄串内部糖度信息。

2.3　特征波长选择

PLSR 模型定量预测糖度能力较好，但各检测区域所建立的模型都存在预测精度不稳定问题，表现为训练集相关系数(R_c)与预测集相关系数(R_p)之间存在较大差异。考虑到参与建模的光谱变量过多可能会导致模型过拟合，通过特征波长选择优化模型。通过波长筛选，可剔除不必要特征，从而使模型更简洁高效。由于 0.3 m/s 为 3 组速度中的最佳检测速度，SG + SNV 为最优预处理方法，因此特征变量筛选仅基于检测速度 0.3 m/s，积分区域为整串预测，经 SG + SNV 预处理后的光谱数据进行 PLSR 模型优化。

采用 4 种常见特征波长选择方法：无信息变量消除(Uninformative variable elimination, UVE)[16]、变量重要性投影(Variable importance in projection, VIP)[17]、连续投影算法(Successive Projections Algorithm, SPA)[18]和竞争性自适应加权采样(Competitive adaptive reweighted sampling, CARS)[19]，并评估其对整串葡萄糖度定量预测的影响。表 5 为 4 种特征变量选择模型对整串葡萄糖度定量预测结果。

表 5　4 种特征变量模型 PLSR 预测结果

Tab. 5　PLSR prediction results of four feature variable selection models

选择方法	变量选择数	LVs	RMSEC/°Brix	RMSEP/°Brix	R_c	R_p
CARS	151	6	0.53	0.55	0.86	0.84
UVE	314	5	0.34	0.41	0.91	0.90
VIP	223	7	0.52	0.61	0.78	0.71
SPA	10	8	0.44	0.56	0.86	0.80

CARS 方法筛选出 151 个特征变量，模型 RMSEC 和 RMSEP 分别为 0.53、0.55°Brix，训练集与预测集相关系数分别为 0.86 和 0.84。UVE 方法筛选出 314 个特征变量，RMSEC 和 RMSEP 分别为 0.34、0.41°Brix，训练集与预测集相关系数最高，分别达到 0.91 和 0.90。UVE 能有效去除冗余波长并提升模型稳定性，但在特征筛选过程中可能误删部分与糖度相关的次要信息，导致模型对未知样本的预测精度略有下降。在在线快速检测中，计算速度是影响检测效率的重要指标，模型精简与信息保留之间需保持平衡。适度的精度损失可换取更高的运算效率与实时性，因而在实际应用中是可接受的。UVE 方法与 CARS 方法有共同特征，即对噪声不敏感，这可能是 UVE 方法与 CARS 方法在光谱特征变量选择表现更优的原因，与此同时，UVE 方法较 CARS 算法有更强的鲁棒性[20]。VIP 选择 223 个变量，RMSEC 和 RMSEP 分别为 0.52°Brix 和 0.61°Brix，训练集和预测集相关系数分别为 0.78 和 0.71，VIP 方法通过衡量每个特征在主成分上的重要性来选择特征，但如果特征间存在高度相关性，VIP 方法会选择多个相关性较强的变量，这会加重多重共线性问题，在挑选特征变量时，会引入冗余信息，影响模型的预测能力，这可能是 VIP 方法表现不如其他 3 种算法的原因。SPA 方法选择 10 个变量，RMSEC 和

RMSEP 分别为 0.44°Brix 和 0.56°Brix，训练集和预测集相关系数分别为 0.86 和 0.80。SPA 方法通过逐步选择具有最大投影能力的特征，通常会选择少量特征，由于 SPA 策略较为严格，其在特征选择时能显著减少冗余特征，从而避免多重共线性的问题，但是选择的特征变量过少，可能导致模型无法捕捉到光谱数据所有的重要信息，进而也会影响模型的精度。为直观展现 4 种特征变量选择方法对光谱特征的选择情况，图 5 为不同方法所选择变量在光谱上分布。

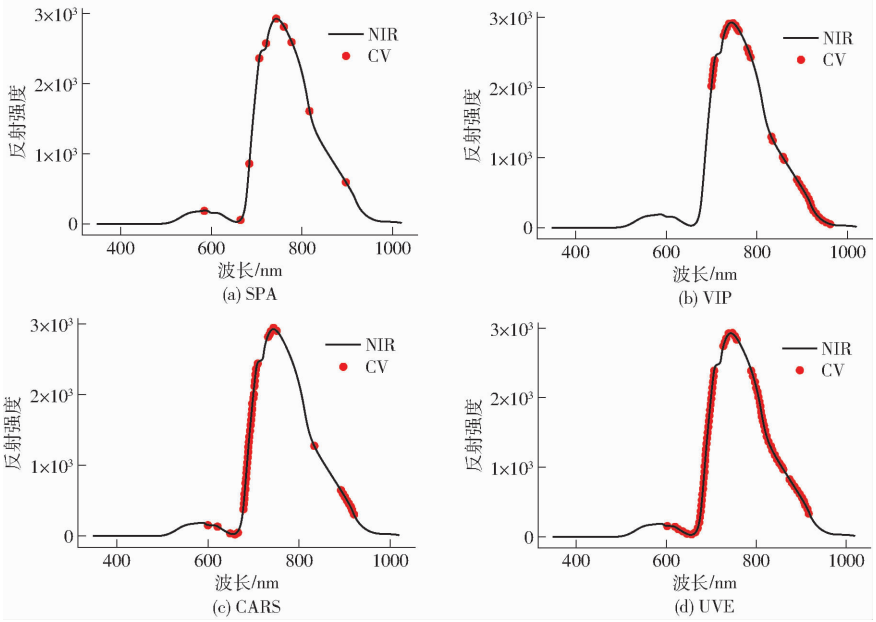

图 5 4 种特征变量筛选方法筛选波段分布情况

Fig. 5 Wavelength distribution selected by four feature variable selection methods

由图 5 可以看出，UVE 和 CARS 方法选择的波长范围较宽且较均匀，它们对信息变量的提取较为全面，SPA 方法选择的波长集中于特定区域，该方法对关键变量选择更加严格和精准，但同时也可能忽略了一些有用的光谱信息。VIP 方法的变量选择则介于二者之间，分布较为集中且稍显冗余。图 6 为

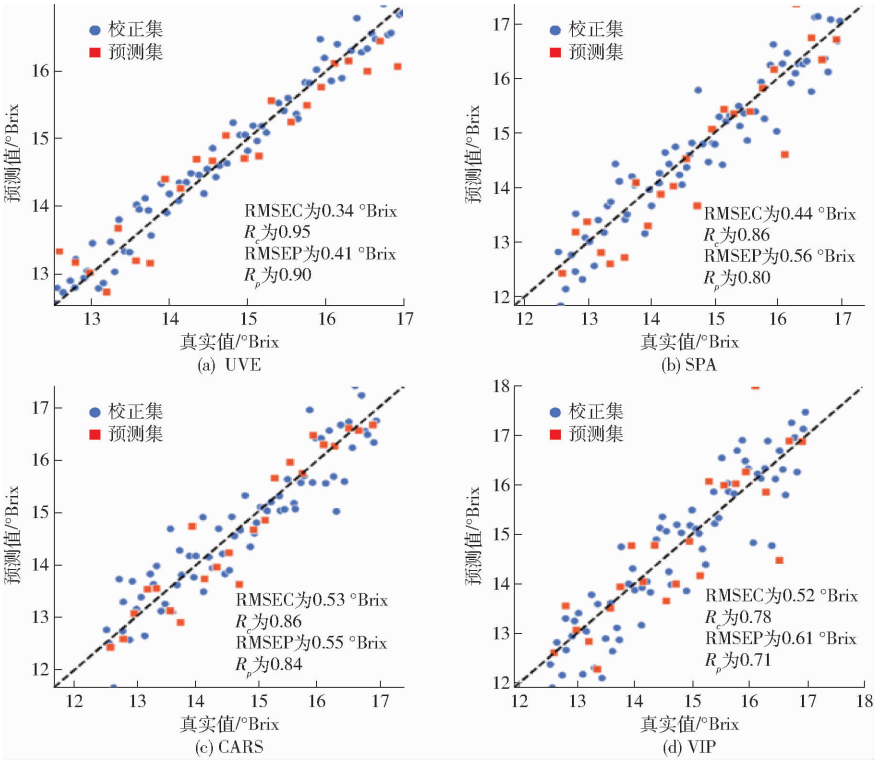

图 6 4 种特征变量选择方法预测结果散点图

Fig. 6 Scatter plots of prediction results from four feature variable selection methods

采用 4 种特征波长选择方法优化后的 PLSR 模型对整串葡萄糖度预测集与校正集散点图。由图 6 可以看出，UVE 方法预测结果数据点更靠近理想预测线，其预测能力最为准确和稳定，SPA 方法虽然模型简洁，但预测点离理想线略远，预测精度稍有降低，VIP 方法预测点散布范围最大，这与表 5 中所得定量指标结果一致。

3　结论

（1）基于可见-近红外光谱技术，实现了阳光玫瑰糖度的在线快速预测，分别构建了用于分段预测和整串预测的偏最小二乘回归（PLSR）模型和支持向量回归（SVR）模型。通过对不同光谱采集速度下的模型预测精度进行对比发现，随着检测速度提高，检测精度呈逐渐下降趋势。综合考虑检测精度与实际在线分选的效率，确定最佳光谱采集速度为 0.3 m/s。

（2）对比了不同光谱预处理方法在 2 种预测模型以及不同检测区域下的表现。结果表明，在分段预测方法中，葡萄中部区域的预测效果优于其他区域；而从整体预测角度来看，整串检测方法更适合在线快速检测。

（3）对 5 种预处理方法在 4 个检测区域的预测性能进行对比后发现，SG 平滑结合 SNV 的组合预处理效果最优，能显著改善光谱质量并降低模型预测误差。当检测速度为 0.3 m/s 时，检测区域为整串 PLSR 模型预测效果最佳，其预测均方根误差为 0.45°Brix，相关系数达到 0.94。针对整串葡萄 PLSR 预测模型，进一步研究了无信息变量消除法（UVE）的变量选择效果。综合来看，UVE 方法在保持模型精度基本稳定的前提下（RMSEP 为 0.41°Brix，R_p 为 0.90），实现了特征变量的有效压缩与建模效率的提升，为整串葡萄品质的快速在线检测提供了理论支持。

参　考　文　献

[1] SU Y, HE K, LIU W, et al. Detection of soluble solid content in table grapes during storage based on visible-near-infrared spectroscopy [J]. Food Innovation and Advances, 2025, 4(1): 10 - 18.

[2] ZHOU X, YANG J, SU Y, et al. Aggregation and assessment of grape quality parameters with visible-near-infrared spectroscopy: introducing a novel quantitative index [J]. Postharvest Biology and Technology, 2024, 218: 113131.

[3] ELSHERBINY O, EL-HENDAWY S, ELSAYED S, et al. Incorporation of visible/near-infrared spectroscopy and machine learning models for indirect assessment of grape ripening indicators [J]. Scientific Reports, 2025, 15(1): 12345.

[4] 刘文政，周雪健，平凤娇，等. 基于可见-近红外光谱的鲜食葡萄成熟品质关键指标检测 [J]. 农业机械学报，2024，55(2)：372 - 383.
LIU Wenzheng, ZHOU Xuejian, PING Fengjiao, et al. Detection of key indicators of ripening quality in table grapes based on visible-near-infrared spectroscopy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(2): 372 - 383. (in Chinese)

[5] ZHANG X, QIN Z, ZHAO R, et al. A handheld IoT Vis/NIR spectroscopic system to assess the soluble solids content of wine grapes [J]. Sensors, 2025, 25(14): 4523.

[6] LYU H, GRAFTON M, RAMILAN T, et al. Non-destructive and on-site estimation of grape total soluble solids by field spectroscopy and stack ensemble learning [J]. European Journal of Agronomy, 2025, 167: 127558.

[7] 王贺功，黄文倩，蔡仲磊，等. 基于全透射可见-近红外光谱的西瓜糖度在线检测研究 [J]. 光谱学与光谱分析，2024，44(6)：1710 - 1717.
WANG Hegong, HUANG Wenqian, CAI Zhonglei, et al. Online detection of sugar content in watermelon based on full-transmission visible and near-infrared spectroscopy[J]. Spectroscopy and Spectral Analysis, 2024, 44(6): 1710 - 1717. (in Chinese)

[8] MUNCAN J, MIYAZAKI M, KUROKI S, et al. Adaptive spectral model for abnormality detection based on physiological status monitoring of dairy cows [J]. Talanta, 2023, 253: 123893.

[9] LUO L, TUO X, ZHANG G, et al. Accuracy improvement of near infrared quantitative analysis model for fermented grains based on abnormal sample removal [J]. Journal of Food Safety and Quality, 2022, 13(9): 3017 - 3025.

[10] ZHANG L, WANG D, GAO R, et al. Improvement on enhanced Monte-Carlo outlier detection method [J]. Chemometrics and Intelligent Laboratory Systems, 2016, 151: 89 - 94.

[11] SCHOOT M, KAPPER C, VAN KOLLENBURG G H, et al. Investigating the need for preprocessing of near-infrared spectroscopic data as a function of sample size [J]. Chemometrics and Intelligent Laboratory Systems, 2020, 204: 104105.

[12] FERREIRA R D A, TEIXEIRA G, PETERNELLI L A. Kennard-Stone method outperforms the random sampling in the selection of calibration samples in SNPs and NIR data [J]. Ciência Rural, 2022, 52(5): e20201072.

[13] XU S, LU B, BALDEA M, et al. An improved variable selection method for support vector regression in NIR spectral modeling [J]. Journal of Process Control, 2018, 67: 83 - 93.

（下转第 150 页）

[20] 季宇寒,李寒,张漫,等.基于激光雷达的巡检机器人导航系统研究[J].农业机械学报,2018,49(2):14 - 21.
JI Yuhan, LI Han, ZHANG Man, et al. Navigation system for inspection robot based on LiDAR[J]. Transactions of the Chinese Society for Agricultural Machinery,2018,49(2):14 - 21. (in Chinese)
[21] 张亮,刘智宇,曹晶瑛,等.扫地机器人增强位姿融合的 Cartographer 算法及系统实现[J].软件学报,2020,31(9):2678 - 2690.
ZHANG Liang, LIU Zhiyu, CAO Jingying, et al. Cartographer algorithm and system implementation based on enhanced pose fusion for sweeping robot[J]. Journal of Software,2020,31(9):2678 - 2690. (in Chinese)
[22] 侯华,王曹,杨沛钊,等.基于改进的无迹卡尔曼滤波长基线定位算法研究[J].计算机应用与软件,2024,41(9):314 - 318,376.
HOU Hua, WANG Cao, YANG Peizhao, et al. Long baseline positioning algorithm based on improved untraced Kalman filter [J]. Computer Application and Software,2024,41(9):314 - 318,376. (in Chinese)
[23] 蒋林,李云飞,汤勃,等.基于语义物体尺寸链的改进自适应蒙特卡洛动态定位研究[J].机械工程学报,2024,60(15):49 - 59.
JIANG Lin, LI Yunfei, TANG Bo, et al. Dynamic localization research on improved AMCL based on the dimensional chain of semantic object[J]. Journal of Mechanical Engineering,2024,60(15):49 - 59. (in Chinese)
[24] 邢强,陈中,冷钊莹,等.基于实时交通信息的电动汽车路径规划和充电导航策略[J].中国电机工程学报,2020,40(2):534 - 550.
XING Qiang, CHEN Zhong, LENG Zhaoying, et al. Route planning and charging navigation strategy for electric vehicles based on real-time traffic information[J]. Proceedings of the CSEE,2020,40(2):534 - 550. (in Chinese)
[25] JIANG S, WANG S, YI Z, et al. Autonomous navigation system of greenhouse mobile robot based on 3D LIDAR and 2D LIDAR SLAM[J]. Frontiers in Plant Science, 2022, 13: 815218.
[26] 齐泽中, 周俊, 叶子渭, 等. 单侧制动式转向履带底盘自主行走局部路径规划方法[J]. 农业机械学报, 2024, 55(6): 52 - 59.
QI Zezhong, ZHOU Jun, YE Ziwei, et al. Local path planning method for autonomous walking of one-sided braking steering tracked chassis[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(6): 52 - 59. (in Chinese)

(上接第 127 页)

[14] XIA Y, HUANG W, FAN S, et al. Effect of fruit moving speed on online prediction of soluble solids content of apple using Vis/NIR diffuse transmission [J]. Journal of Food Process Engineering, 2018, 41(8): e12915.
[15] GRISANTI E, TOTSKA M, HUBER S, et al. Dynamic localized SNV, peak SNV, and partial peak SNV: novel standardization methods for preprocessing of spectroscopic data used in predictive modeling [J]. Journal of Spectroscopy, 2018, 2018: 1 - 14.
[16] JIANG W, LU C, ZHANG Y, et al. Moving-window-improved Monte Carlo uninformative variable elimination combining successive projections algorithm for near-infrared spectroscopy (NIRS) [J]. Journal of Spectroscopy, 2020, 2020: 1 - 12.
[17] RAHI S, MOBLI H, JAMSHIDI B, et al. Achieving a robust Vis/NIR model for microbial contamination detection of Persian leek by spectral analysis based on genetic, iPLS algorithms and VIP scores [J]. Postharvest Biology and Technology, 2021, 175: 111413.
[18] SUN Z, FAN J, WANG J, et al. Assessment of the human albumin in acid precipitation process using NIRS and multi-variable selection methods combined with SPA [J]. Journal of Molecular Structure, 2020, 1199: 126942.
[19] 刘燕德, 施宇, 蔡丽君, 等. 基于 CARS 算法的脐橙可溶性固形物近红外在线检测 [J]. 农业机械学报, 2013, 44(9): 138 - 144.
LIU Yande, SHI Yu, CAI Lijun, et al. Online NIR detection model optimization of soluble solids content in navel orange based on CARS[J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(9):138 - 144. (in Chinese)
[20] GUO Z, ZOU Y, SUN C, et al. Nondestructive determination of edible quality and watercore degree of apples by portable vis/NIR transmittance system combined with CARS - CNN [J]. Journal of Food Measurement and Characterization, 2024, 18(6): 4058 - 4073.