DOI:10.13652/j.spjx.1003.5788.2025.80106

基于可见近红外光谱技术的新梅贮藏品质无损预测

文 龙 张 慧 赖丽思

（新疆大学智能制造产业学院,新疆 乌鲁木齐 830017）

摘要:[**目的**]利用可见近红外(Vis-NIR)光谱技术构建新梅果实可溶性固形物含量(SSC)和硬度预测模型,实现新梅内部品质的无损、准确评估。[**方法**]以280个新梅样本为研究对象,利用近红外系统采集不同贮藏期新梅的Vis-NIR反射光谱,采用主成分马氏距离(PCA-MD)法剔除新梅异常样本,再按照K-S(kennard stone)算法将数据集按3∶1的数量划分为校正集和预测集,统一采用偏最小二乘(PLS)和支持向量回归(SVR)对比分析平滑去噪(SG)、标准正态变换(SNV)、一阶导数(1st-D)和归一化(Norm)对原始光谱预处理的效果。使用竞争自适应重加权采样(CARS)、自举软收缩算法(BOSS)和区间组合优化(ICO)对新梅近红外光谱特征波长进行筛选,结合PLS和SVR算法构建新梅品质的回归预测模型。[**结果**]针对新梅SSC构建的SNV-CARS-SVR预测模型表现最佳,其预测集决定系数(R_p^2)、预测集均方根误差(RMSEP)及残差预测偏差(RPD)分别为0.866,0.651和1.956;针对新梅硬度构建的Norm-BOSS-PLS模型的预测效果最好,其R_p^2、RMSEP和RPD分别为0.894,0.740和2.207。[**结论**]利用Vis-NIR光谱技术对新梅SSC和硬度进行无损预测具有一定的可行性和良好的应用潜力。

关键词:新梅;可见近红外;可溶性固形物含量;硬度;机器学习;无损检测

Non-destructive prediction about storage quality of prunes in Xinjiang using visible and near-infrared spectroscopy technology

WEN Long ZHANG Hui LAI Lisi

(*College of Intelligent Manufacturing Modern Industry*, *Xinjiang University*, *Urumqi*, *Xinjiang* 830017, *China*)

Abstract: [**Objective**] To establish a prediction model for the soluble solid content (SSC) and firmness of prunes in Xinjiang using visible and near-infrared (Vis-NIR) spectroscopy, enabling nondestructive and accurate evaluation of the internal quality of the fruit. [**Methods**] Taking 280 prunes in Xinjiang as research samples, the study collects Vis-NIR reflectance spectra at different storage periods with a near-infrared system. Principal component analysis-Mahalanobis distance (PCA-MD) is applied to eliminate abnormal samples. Then, the remaining dataset is divided into a calibration set and a prediction set at a ratio of 3∶1 using the Kennard-Stone (K-S) algorithm. Partial least squares (PLS) and support vector regression (SVR) models are employed to compare the effects of different preprocessing methods on the raw spectra, including Savitzky-Golay smoothing (SG), standard normal variable transform (SNV), first-order derivative (1st-D), and normalization (Norm). Characteristic wavelengths of the prunes' spectra are selected using competitive adaptive reweighted sampling (CARS), bootstrapping soft shrinkage (BOSS), and interval combination optimization (ICO). Based on the selected features, regression models for the quality prediction of prunes are established using PLS and SVR algorithms. [**Results**] For the SSC prediction of prunes in Xinjiang, the SNV-CARS-SVR model demonstrates the best performance, with a determination coefficient (R_p^2) of 0.866, a root mean square error of prediction (RMSEP) of 0.651, and a residual predictive deviation (RPD) of 1.956. For firmness prediction, the Norm-BOSS-PLS

基金项目:新疆维吾尔自治区自然科学基金项目(编号:2022D01C674)

通信作者:张慧(1992—),女,新疆大学副教授,博士。E-mail:hui@xju.edu.cn

收稿日期:2025-02-13 **改回日期:**2025-08-10

引用格式:文龙,张慧,赖丽思. 基于可见近红外光谱技术的新梅贮藏品质无损预测[J]. 食品与机械,2026,42(3):86-96.

Citation:WEN Long, ZHANG Hui, LAI Lisi. Non-destructive prediction about storage quality of prunes in Xinjiang using visible and near-infrared spectroscopy technology[J]. Food & Machinery, 2026, 42(3): 86-96.

model achieves the best results, with an R_p^2 of 0.894, an RMSEP of 0.740, and a RPD of 2.207. [**Conclusion**] The nondestructive prediction of the SSC and firmness of prunes in Xinjiang using Vis-NIR spectroscopy is demonstrated to be feasible and holds considerable potential for practical application.

Keywords: prune in Xinjiang; visible-near infrared; soluble solid content; firmness; machine learning; non-destruction detection

新梅是新疆特色果品,以其色泽艳丽、口感清甜、果香浓郁及营养丰富而深受消费者喜爱。在市场流通中,新梅果实需符合一定的质量标准,这些标准不仅包括外观特征,如色泽和果形,更重要的是内部品质指标,如可溶性固形物含量(SSC)、硬度和酸度等。这些内部品质直接影响消费者的接受度和满意度,同时决定了新梅的贮藏期和适宜的贮藏条件。优质的新梅果实在完全成熟时,表皮颜色呈深紫红色,果肉淡黄色且质地柔软。由于新梅果实极易腐败,其常温货架期通常不超过7 d,极短的保鲜期对其贮藏和市场供应提出了更高的要求[1]。因此,预测SSC和硬度等关键品质指标在新梅贮藏过程中尤为重要,不仅有助于科学管理贮藏条件和时间,还能延长果实货架期,减少过度成熟及腐烂带来的损失。目前,新梅品质的检测主要依赖传统化学分析方法,虽然能够提供准确的结果,但存在操作复杂、时间长、费用高且具有破坏性等问题,难以满足现代果品快速检测和高效供应链的需求,迫切需要一种无损检测方法。

无损检测技术在农产品应用广泛,具有检测速度快、检测成本低、实时在线检测等优点,常用的有机器视觉、核磁共振、高光谱、近红外光谱、电子鼻和电学特性检测等。相较于其他检测手段,近红外光谱技术在果蔬检测领域展现出了显著的优势,具有操作简单、高效快速、绿色环保和成本低等特点,在国内外得到了众多学者的青睐。

近红外光谱分析易受仪器性能、样品背景等因素的影响,常出现谱图偏移、漂移等现象,导致光谱数据中包含大量噪声和散射等干扰信息。这些干扰不仅降低了模型的精度,还影响了测试结果的可靠性[2]。因此,为了提高模型的稳定性和预测精度,必须对光谱数据进行有效的预处理,以提取关键信息并削弱或消除噪声的影响。赵杰文等[3]采用正交信号校正法和净分析物预处理法对苹果近红外光谱进行预处理,并结合偏最小二乘(PLS)建立SSC预测模型。结果表明,经过预处理的模型性能均优于原始光谱模型,在提高预测精度的同时有效优化了模型性能。占可等[4]采用一阶导数(1st-D)、多元散射校正(MSC)、小波变换(WT)和标准正态变换(SNV)对解冻的小龙虾虾尾、虾仁及虾糜近红外光谱进行预处理,并结合PLS与卷积神经网络(CNN)构建预测模型。结果表明,预处理对模型的预测精度具有显著影响,其中,虾仁光谱经WT预处理结合CNN模型的预测能力最佳。Purwanto等[5]采用PLS结合不同的预处理方法建立了"Gedong Gincu"芒果SSC预测模型,其中构建3点平滑预处理后得到模型性能最佳,相关系数(R)为0.82,预测集均方根误差(RMSEP)为1.28 °Brix,相对预测偏差(RPD)为1.52。表明合理地选择近红外光谱预处理方法是提高预测模型准确性和稳定性的关键步骤,能显著提高模型对水果内部品质的预测性能。

近红外光谱仪采集的光谱数据覆盖全波段,其中包含一些无关信号。如果将所有数据直接输入模型,可能会降低检测的效率和准确性。波长选择通过从海量光谱变量中提取最具信息量的子集,不仅可以最小化误差,还能有效排除噪声和不可靠数据,从而显著提升预测的准确性。目前,多种特征提取方法已被广泛应用于光谱数据的处理,如无信息变量消除(UVE)、竞争自适应重加权采样(CARS)、遗传(GA)、逐步投影(SPA)等算法。此外,还有一些基于Bagging优化的BOSS光谱选择方法以及区间组合优化方法也常被应用于光谱特征提取[6-8]。罗澍寰等[9]基于Vis-NIR光谱技术结合CARS+LS-SVM预测模型,从1 401个光谱变量中优选出21个变量去除冗余信息的同时,很好地保留了涵盖特征信息的数据,实现了梨总酸含量的定量无损检测。母雯竹等[10]采用CARS结合SPA算法提取白酒固态发酵副产物黄水的近红外光谱特征,构建的PLS模型相比未使用特征提取或仅使用单一特征提取方法所建立的模型性能均有明显提升。在单一使用CARS时,R^2为0.965 4,RMSEP为0.201 2%;而结合SPA后,R^2提升至0.973 8,RMSEP降至0.174 8%。这种联合特征提取的方法最大程度减少了建模所需变量,将光谱维度从2 203个减少到了126个,不仅显著提高了预测精度,还大幅提升了建模效率。基于近红外光谱技术,选择合适的特征提取方法结合定量分析模型,可实现对水果内部品质的高效且精准预测。

由于新梅在贮藏和流通过程中面临的品质检测难题,研究拟采用Vis-NIR光谱技术获取其光谱数据,并结合化学计量学方法测定其SSC和硬度品质指标,通过对光谱数据进行预处理和特征波段提取,提取与品质指标相关的特征波段,通过PLS和SVR进行新梅的预测模型构建与对比分析,验证Vis-NIR光谱技术在贮藏期新梅品质预测中的可行性,以期为今后利用Vis-NIR光谱技术的新梅内部品质检测技术与装备研发提供参考。

1　材料和方法

1.1　材料

选用新疆伽师新梅，于2024年8月在乌鲁木齐市九鼎市场采购。挑选发育正常、外形良好、无机械损伤的新梅共280个，单果质量（45.00±20.00）g，果径（15.00±3.85）mm，逐一编号后置于HWS-350恒温恒湿箱（绍兴市思阳仪器制造有限公司）内，在室温（20~25 ℃，相对湿度15%~25%）下静置24 h，以使样本内部物质达到稳定状态。鉴于新梅在室温下的保存期为3~7 d[1]，将新梅样本分为7组，每组40个，试验周期设定为7 d。每日对样本进行光谱数据采集和理化指标测定，记录贮藏期间新梅的采后品质变化。

1.2　试验设备与光谱数据采集方法

搭建的近红外反射光谱无损检测平台如图1所示，主要由暗箱、光谱仪和计算机构成。暗箱内部铺设黑色吸光布，以减少试验过程中杂散光对光谱采集的干扰。光源采用HL-100型卤素灯（上海复享光学股份有限公司），与PG2000-Pro型背照式光谱仪（上海复享光学股份有限公司）通过Y型反射光纤（广州瑞科光电科技有限公司）连接。Y型光纤的两个分支汇聚后形成一根光纤，通过暗箱侧面开孔引入，固定于反射光纤支架上，并垂直对准待测新梅，确保与新梅赤道部最高点的距离为180 mm。为减小温度变化对光学器件和暗电流的影响，设备在试验前预热30 min。待测新梅置于暗箱底部的果托上，光谱采集时，新梅按照果萼—果梗轴线水平放置，对应标记采集点。试验中，光谱信号通过Y型光纤传输至光谱仪，并利用Morpho 3.2软件进行数据采集与处理。光谱仪的波长范围为367.00~1 052.58 nm，信噪比为800∶1，积分时间为100 ms，扫描平均次数为5次，滑动平均宽度设定为5。采集前，通过白色聚四氟乙烯圆柱体进行黑白校正。在光谱采集过程中，待测新梅每次旋转120°，并在赤道部对应标记点采集光谱，共采集3次，取其平均值作为该样本的反射光谱数据，确保数据的准确性与重复性。

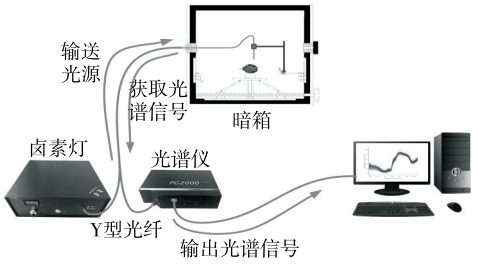


图1　新梅近红外反射光谱无损检测平台

Figure 1　Nondestructive detection platform for near-infrared reflectance spectroscopy of prunes in Xinjiang

1.3　新梅可溶性固形物含量和硬度测量

采集新梅近红外光谱数据后立即使用GY4型水果硬度测试仪（乐清市艾德堡仪器有限公司）测定其硬度。依据NY/T 2009—2011《水果硬度的测定》，在每个样本中部阴阳两面各选一点，削去薄层果皮后进行测量。新梅硬度取阴阳两面测量的平均值。采集硬度后，从新梅赤道部位均匀地分3个点处切取10 mm厚果肉，手动榨汁并摇匀，用滴管滴至PAL-1型迷你数显折射仪（爱宕科学仪器有限公司）的折光仪镜面上进行SSC测量[11]。

1.4　近红外光谱数据分析方法

1.4.1　光谱数据预处理　光谱采样时，会受到背景噪声、仪器误差、基线漂移等的干扰，因此在建模之前需要对原始光谱数据采用合适的预处理，来降低或消除无关信息的影响，提高模型的精确度和稳定性。SG采用局部多项式分解数据并使用最小二乘拟合技术平滑处理光谱，对采集到的光谱信号进行了降噪，能够增加样本信号的信噪比；SNV通过减去平均频谱并除以标准差对每个频谱进行标准化，用于校正散射以及光程变化的影响，消除基线偏移，增强光谱特征；1st-D用于基线校正，消除测量仪器背景或漂移对信号的影响，提高光谱分辨率；Norm将数据的范围固定到一个标准区间内消除量纲对于模型结果的影响，增强特征权重，消除光程差异[12]。研究拟对比多种预处理方法，基于建模评价指标选择最优策略。

1.4.2　特征波长提取方法　CARS算法是模仿达尔文生物进化理论中的“适者生存”原则，采用基于指数递减函数和自适应重加权采样技术，选出模型中回归系数绝对值大的波长变量，利用交叉验证筛选出最低交叉验证均方根误差（RMSECV）的子集，确定为最优变量子集。该方法适合高维数据的变量筛选，可有效选择与新梅可溶性固形物含量和硬度相关的最优波长组合，消除数据冗余和降低模型训练样本的复杂度[13]。使用CARS算法时，设置最大潜在变量数为15，蒙特卡罗采样次数为1 000次。

BOSS算法能够选择具有共线性的信息变量。先在变量空间中利用BBS生成K个子集，所有变量都分配有相等的权重。然后用所有子集构建K个PLS回归子模型，并挑选出最小RMSECV的最佳模型，再对归一化回归向量进行求和以获得变量的新权重。新的子集由加权自举抽样（WBS）根据新的权重重新生成，最终选择在迭代期间具有最低RMSECV的子集作为最佳变量集[7]。使用BOSS算法时，设置潜在变量数为10，交叉折叠5层，采样次数1 000。

ICO算法将全光谱划分为固定数量的等宽区间，然后在模型群体分析指导下以软收缩的方式迭代搜索最优区间组合，其中随机抽样方法采用WBS，最后进行局部搜索以优化所选区间的宽度[8]。使用ICO算法时，将整个光谱

划分为40个相等的子区间,并从500个随机区间组合中提取10%的最佳区间。

利用CARS、BOSS和ICO 3种特征提取算法,提取与新梅SSC和硬度变化相关的有效光谱特征波段,并通过不同建模算法,比较不同特征提取方法在最优预处理条件下的建模效果,确定最佳的特征波段提取方法。

1.5　预测模型与评价指标

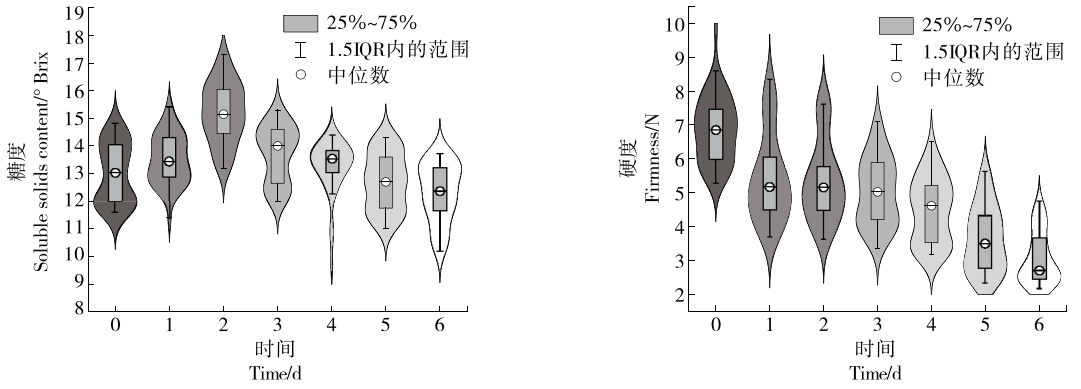
1.5.1　预测模型方法　PLS是化学计量学中一类应用广泛的回归方法,用于通过潜在变量来建立观测变量集之间的相关性模型,它可以描述光谱矩阵$\boldsymbol{X}$和理化矩阵$\boldsymbol{Y}$的信息,充分拟合样品光谱数据与理化值之间的最大相关性,实现定量分析。该方法可以有效地解决变量数量多、数据噪声大、变量间多重相关等问题[14]。SVR算法使用核函数将原始数据通过非线性变换转换到高维特征空间,从而构造一个最优超平面,使得SVR算法能够以线性的方式解决非线性的分类问题,同时具有较优的推广能力和较低的复杂度,能够很好地处理高维数据。其中核函数的选择对于SVR的回归性能有较大影响,径向基核函数(RBF)能捕捉到数据的复杂关系且参数调节相对简单,因此采用RBF作为SVR的核函数[15]。

1.5.2　评价指标　模型性能评价的统计学指标有校正集决定系数(R_c^2)、校正集均方根误差(RMSEC)、预测集决定系数(R_p^2)、预测集均方根误差(RMSEP)和残差预测偏差(RPD)。

2　结果与分析

2.1　新梅贮藏期间SSC和硬度变化趋势

如图2(a)所示,贮藏2 d时新梅SSC变化较大且有一定数量的离群值,这与新梅贮藏初期的生理代谢活跃有关。贮藏2~4 d,SSC的变化相对稳定,中位数逐渐下降,整体SSC呈逐步减少趋势。其中,贮藏0 d时新梅SSC的中位数明显高于贮藏6 d时的,表明贮藏时间对新梅贮藏期的SSC有显著影响。如图2(b)所示,贮藏0 d时新梅的硬度值存在较大差异。贮藏1~2 d硬度箱线的宽度明显收窄,变化趋于稳定,整体硬度逐渐减少。其中,贮藏0 d时新梅的硬度明显高于贮藏6 d时的,表明随着贮藏时间的延长,新梅硬度显著下降。总的来说,新梅SSC在初期贮藏阶段呈先上升的趋势,随着贮藏时间的延长,新梅组织逐渐变得松软,SSC进一步增加,风味也随之提升,而硬度则呈逐步下降的趋势。通过变异系数(CV)可以衡量测量值在样本间的相对波动性[16]。表1为不同时间下新梅SSC和硬度的变异系数值,新梅糖度的CV值变化范围为6.69%~8.57%,尽管不同贮藏时间段间略有波动,但总体波动幅度较小,表明糖度数据分布较为稳定。相比之下,新梅硬度的CV值为14.72%~26.58%,且贮藏6 d时达到峰值,反映出随着贮藏时间的延长,新梅硬度的波动性显著增加,表明样品硬度的稳定性在贮藏后期有所下降。



(a) SSC　　(b) 硬度

图2　不同贮藏时间下新梅SSC和硬度的变化

Figure 2　Changes in SSC and firmness of prunes in Xinjiang at different storage periods

2.2　新梅贮藏期间光谱变化情况

图3是新梅在不同贮藏时间下,每个时期40个样本的平均光谱曲线。可以看出,各光谱曲线的波形呈现出相同的趋势,而光谱反射率随不同贮藏时间有明显差异。贮藏前期反射率值明显高于贮藏后期的,表明样本内的物质成分基本一致,但不同贮藏时间的化合物含量发生了变化。新梅在不同贮藏时期的生理特性会发生变化,这反映在其光谱在各波段的吸收程度不同,这种光谱吸收程度的变化与果实本身的生理变化,包括内部物质含量的变化,有着密切的关系。不同贮藏时间的样本内部化学物质基本相同,但个别化合物含量不同,光谱反射率的差异为使用近红外光谱技术建立回归预测模型提供了前提。考虑到光谱首端有较为明显的噪声信号,取400.00~1 052.58 nm波段进行后续分析。

表 1　不同贮藏时间下新梅 SSC 和硬度的变异系数

Table 1　Coefficient of variation of SSC and firmness of prunes in Xinjiang at different storage periods

贮藏时间/d	变异系数/%	
	SSC	硬度
0	8.37	14.72
1	6.69	24.05
2	7.22	22.17
3	7.46	20.05
4	7.25	21.50
5	8.08	26.58
6	8.57	25.20

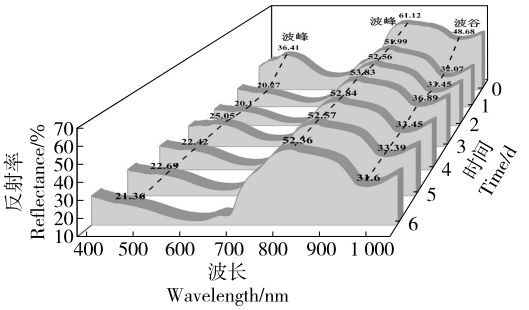
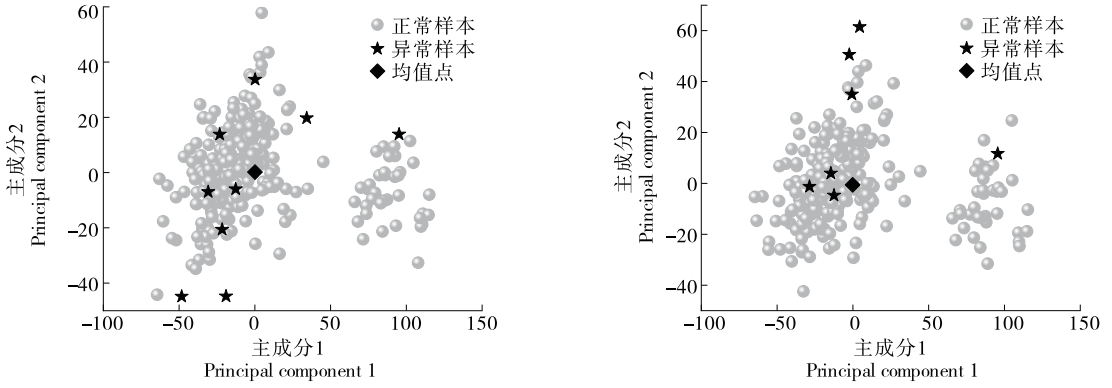


图 3　不同贮藏期新梅平均光谱 3D 曲线

Figure 3　3D curves of average spectra of prunes in Xinjiang at different storage periods

2.3　异常样本剔除和数据集划分

在采集近红外光谱数据过程中，周围环境变化、人为操作失误或仪器运行波动等因素可能导致光谱数据出现异常[17]。针对 280 个新梅样本的 SSC 和硬度定量分析，采用 PCA-MD 法剔除异常样本[18]。此方法能够有效提高近红外光谱技术对新梅果实定量分析的准确性和稳定性，结果如图 4 所示（圆圈表示正常样本，星号表示异常样本，菱形表示均值点）。图 4(a)检测到样本 29、37、93、95、138、148、264、268 和 280 为异常值；图 4(b)检测到样本 33、44、46、50、51、52 和 89 为异常值。随后对剔除后留下的 271 个新梅样本 SSC 数据和 273 个新梅样本硬度数据进行数据集划分。K-S(kennard stone)能够更好地评估模型的泛化性能，减少对特定数据集的依赖性，减轻过拟合的风险[7]，因此采用 K-S 算法将数据集按 3∶1 划分为校正集和预测集。



(a) SSC　　(b) 硬度

图 4　主成分分析空间中新梅 SSC 硬度的离群值检测结果

Figure 4　Outlier detection results for SSC and firmness of prunes in Xinjiang in PCA space

2.4　新梅光谱预处理方法的确定

利用 SG、SNV、1st-D 和 Norm 4 种不同预处理方式对新梅光谱数据进行处理，并将预处理后的数据与 SSC 和硬度进行 PLS 和 SVR 建模分析，结果如表 2 所示。由表 2 可看出，在新梅 SSC 的预测中，经过 SNV 预处理后建立的 SVR 预测模型表现优于使用其他 3 种预处理手段和原始光谱数据构建的模型性能，R_p^2 为 0.824，RMSEP 为 0.757，RPD 为 1.702，与表现最差的 1st-D-PLS 模型相比，R_p^2 和 RPD 分别提升了 0.121 和 0.331，RMSEP 减少了 0.192。表明 SNV 预处理能够有效减少漫反射光谱中因样品颗粒分布不均、颗粒大小不一致等因素引起的散射影响，从而提高了光谱数据在不同样品间的一致性和可比性[19]。

在新梅硬度的预测中，与 SSC 光谱预处理情况相同，采用 1st-D 预处理后建立的预测模型效果最差，而经过 Norm 预处理后建立的 PLS 预测模型表现最好，R_p^2 为 0.848，RMSEP 为 0.858，RPD 为 1.828，表明 Norm 预处理能有效减少光谱因微小光程差带来的影响[11]。因此，后续选择 SNV 和 Norm 分别作为 SVR 和 PLS 模型预测 SSC 和硬度的最佳预处理方法。

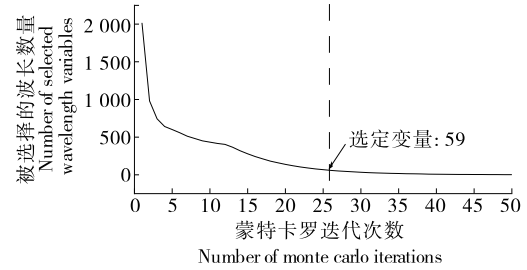
2.5　新梅光谱有效特征波长的提取

2.5.1　基于 CARS 方法的特征波长提取结果　图 5 和图 6 展示了预处理后的光谱数据进行 CARS 筛选新梅 SSC 和

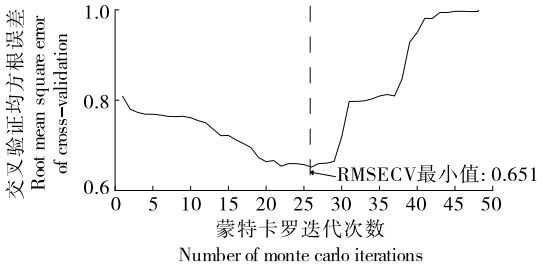
表2　预处理方法对PLS和SVR模型性能的影响

Table 2　Effects of preprocessing methods on the performance of PLS and SVR models

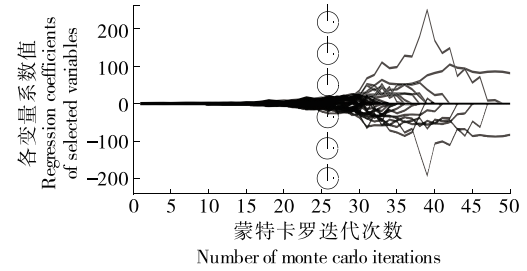
参数	模型	预处理方法	校正集		预测集		RPD	参数	模型	预处理方法	校正集		预测集		RPD
			R_c^2	RMSEC	R_p^2	RMSEP					R_c^2	RMSEC	R_p^2	RMSEP	
SSC	PLS	RAW	0.813	0.766	0.793	0.719	1.635	硬度	PLS	RAW	0.863	0.894	0.821	0.902	1.689
		SG	0.813	0.766	0.793	0.719	1.635			SG	0.862	0.896	0.822	0.902	1.689
		SNV	0.825	0.717	0.815	0.749	1.719			SNV	0.886	0.816	0.747	1.060	1.503
		1st-D	0.915	0.510	0.703	0.949	1.371			1st-D	0.954	0.534	0.755	1.092	1.401
		Norm	0.818	0.735	0.800	0.758	1.665			Norm	0.865	0.886	0.848	0.858	1.828
	SVR	RAW	0.841	0.743	0.782	0.743	1.583		SVR	RAW	0.862	0.979	0.789	0.979	1.557
		SG	0.831	0.749	0.777	0.749	1.570			SG	0.862	0.979	0.789	0.979	1.557
		SNV	0.965	0.757	0.824	0.757	1.702			SNV	0.891	1.059	0.749	1.059	1.504
		1st-D	1.000	0.912	0.726	0.912	1.427			1st-D	0.990	1.154	0.723	1.154	1.326
		Norm	0.925	0.809	0.768	0.809	1.559			Norm	0.877	0.908	0.834	0.908	1.728



(a) 特征波长选择数



(b) RMSECV 变化轨迹

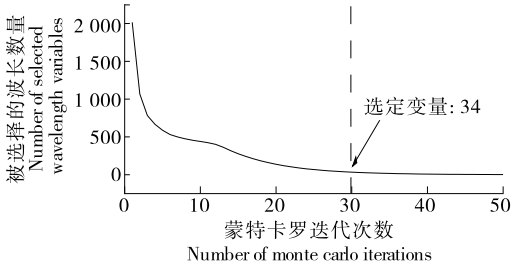


(c) 波长回归系数走势

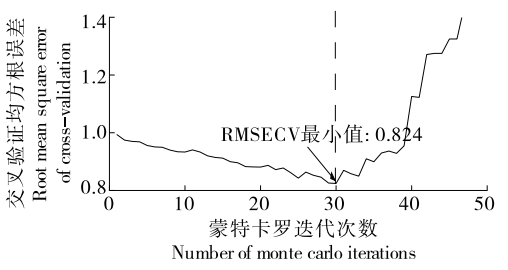
每条曲线代表不同采样次数下特征变量回归系数的变化,“○”标记的点表示RMSECV值最小时的采样次数,即模型预测性能最佳的时刻

图5　基于CARS算法筛选新梅SSC的特征波段过程

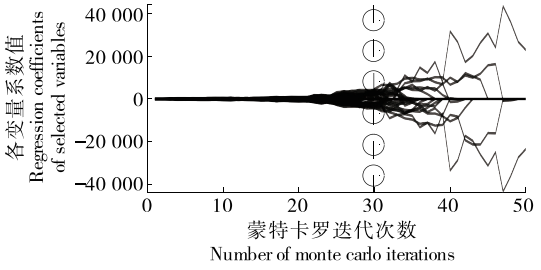
Figure 5　Feature band selection process for SSC of prunes in Xinjiang based on the CARS algorithm



(a) 特征波长选择数



(b) 交叉验证均方根误差



(c) 波长回归系数走势

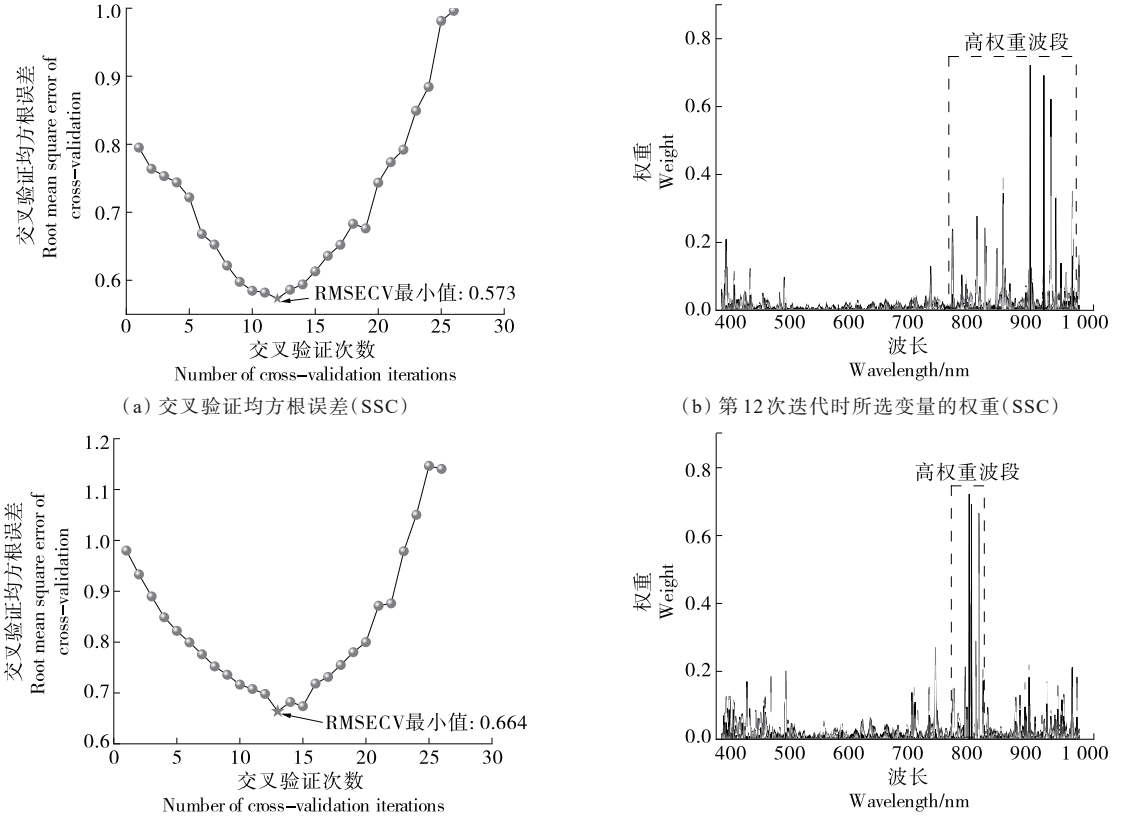
每条曲线代表不同采样次数下特征变量回归系数的变化,“○”标记的点表示RMSECV值最小时的采样次数,即模型预测性能最佳的时刻

图6　基于CARS算法筛选新梅硬度的特征波段过程

Figure 6　Feature band selection process for firmness of prunes in Xinjiang based on the CARS algorithm

硬度特征波段筛选的过程。如图5所示，随着采样次数增加，RMSECV值先下降后上升。在初期，剔除大量无效变量有助于提升模型的预测性能，会使RMSECV值持续降低。然而，随着采样次数的增加，会使某些有效变量被过度剔除，从而导致模型性能下降，RMSECV值再次增大。采样次数为26时，RMSECV值最小(0.651)，此时筛选出的特征波段为59个，占全光谱波段数(2 013个)的2.93%。对于硬度的特征变量提取过程，与SSC的分析类似。通过结合RMSECV变化趋势和回归系数路径，发现第30次采样时RMSECV值最小(0.824)，对应的特征波段为34个，占全光谱波段数的1.69%。

2.5.2　基于BOSS方法的特征波长提取结果　从图7可以看出，通过多次交叉验证对变量进行逐步筛选和优化，模型的预测性能得到了显著提升。在交叉验证初期，最小误差值呈下降趋势，表明模型通过剔除冗余变量有效提高了预测精度。然而，随着交叉验证次数的进一步增加，关键变量的过度剔除，模型预测能力逐渐下降，最小误差值出现反弹趋势。具体而言，从图7(a)可以看出，在第12次迭代时，最小误差值达到0.573，此时筛选出的最佳变量子集包含50个特征波段，占全光谱波段的0.248%。图7(b)进一步展示了BOSS方法在第12次迭代中选择的50个变量的权重分布情况。可以观察到，在960 nm附近的变量权重最大，表明该波长区域的信息量最为丰富。此外，810~1 040 nm范围内的多个波长变量也具有较高权重，表明该区域包含丰富的光谱信息。从图7(c)可知，在第13次迭代时，模型的最小误差值达到0.664，此时筛选出的最佳变量子集同样包含50个特征波段，占全光谱波段的0.248%。图7(d)显示了BOSS方法在第13次迭代中选择的50个变量的权重分布情况，其中850 nm附近的变量权重最大，信息量最为显著。此外，840~880 nm范围内也存在多个高权重波长，表明该区域包含丰富的光谱信息。



(a) 交叉验证均方根误差(SSC)　(b) 第12次迭代时所选变量的权重(SSC)

(c) 交叉验证均方根误差(硬度)　(d) 第12次迭代时所选变量的权重(硬度)

图7　基于BOSS算法筛选新梅SSC和硬度的特征波段过程

Figure 7　Feature band selection process for SSC and firmness of prunes in Xinjiang based on the BOSS algorithm

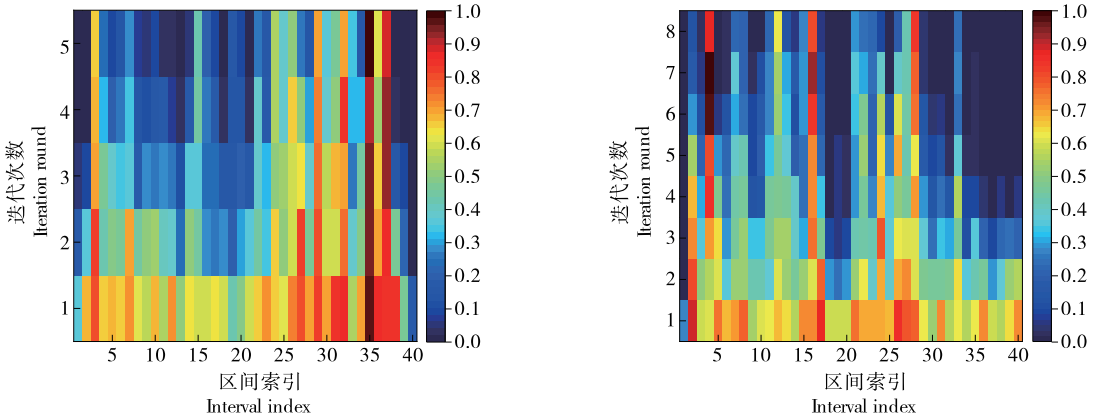
2.5.3　基于ICO方法的特征波长提取结果　图8为ICO算法迭代过程中每个波长区间的采样权重随迭代次数增加的变化情况。从图8(a)可以看出，第35个波长区间的采样权重值从第1次迭代过程中到第5次迭代过程中始终大于0.9，该区间波段直接被选中。第7个区间在第二次迭代后被排除，但在最后保留为建模的有用区间。第

35、37和38个波长区间对应光谱区间950~1 000 nm,该区间包含了O—H键拉伸振动的第二泛音[20]。经过ICO算法的选择,最终确定了以下子区间:在新梅SSC预测中,选择了第3、7、29、32、35、37和38个子区间;而在新梅硬度预测中,选择了第2、4、12、16、28和33个子区间。

图9展示了CARS、BOSS和ICO特征提取方法在新梅SSC和硬度预测中所选择特征变量的分布图。新梅SSC和硬度的特征点大部分集中在750~1 000 nm附近的波段,而在500~700 nm附近的波段,3种方法对新梅硬度的特征提取分布情况差异较大。其中ICO算法提取的特征信息较为冗余,未能全面筛选出有用的信息。从硬度特征变量分布可以看出,CARS和ICO算法特征提取增加了680 nm附近的波段的提取,同时在750~1 000 nm附近的波段中,ICO算法特征提取重要提取集中在840~860和920~940 nm。

2.6　新梅贮藏期SSC和硬度预测模型的建立与分析

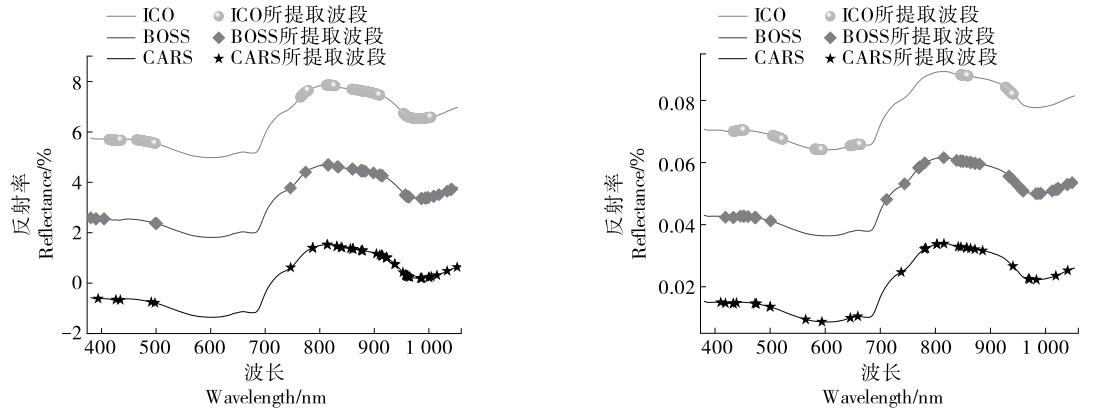
表3为新梅不同贮藏期内SSC和硬度预测模型结果。可以看出,基于SNV预处理后采用CARS构建的CARS-SVR模型对新梅SSC预测性能最优,其R_c^2为0.929,RMSEC为0.651,R_p^2为0.866,RMSEP为0.651和RPD为1.956,相比采用BOSS和ICO构建的SVR模型,R_p^2分别提高了0.022和0.039,RMSEP分别降低了0.063和0.118,RPD分别提高了0.082和0.209。其中,SNV-ICO-SVR模型是对比3种特征提取方法后建模效果最差的。不管原始全波段光谱数据是否经过SNV预处理,未经过有效特征提取过程建立的预测模型表现都不尽人意,表明全波段光谱虽然包含了丰富的信息,但也引入了大量冗余特征信息,这些冗余特征信息会干扰模型的预测,导致预测模型的性能不佳。综合来看,采用CARS算法提取特征波长建立的SVR模型,大幅度降低了计算量,还显著提升



(a) SSC　　(b) 硬度

图8　区间组合优化过程中新梅SSC和硬度特征区间的采样权重

Figure 8　Sampling weights of characteristic intervals for SSC and firmness of prunes in Xinjiang during ICO



(a) SSC　　(b) 硬度

图9　基于不同特征提取方法新梅SSC和硬度的特征变量分布结果图

Figure 9　Characteristic variable distribution for SSC and firmness of prunes in Xinjiang based on different feature extraction methods

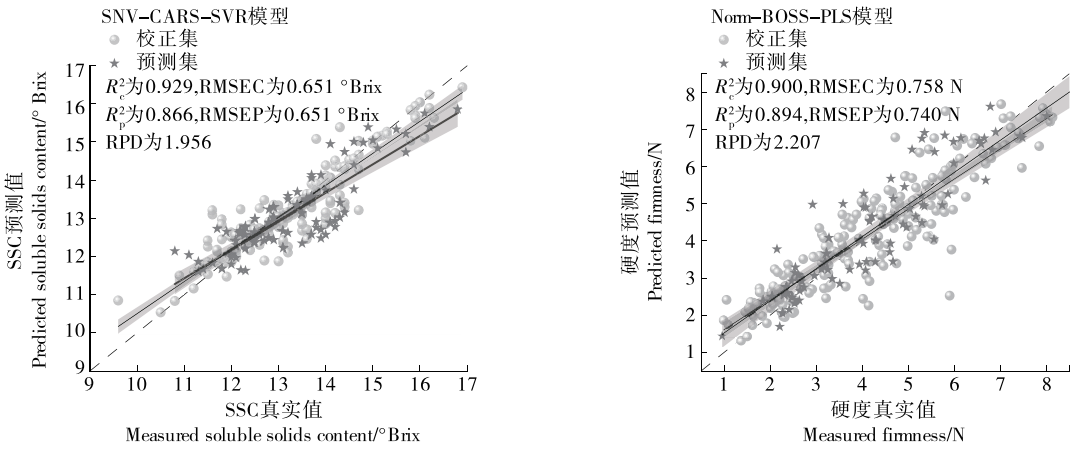
了预测的精度与稳定性，实现了简化模型的预期效果。

通过比较各个新梅硬度预测模型可以发现，虽然CARS选取的变量数量较少，但其对应模型的预测效果却不如BOSS算法，采用Norm-BOSS-PLS模型对新梅硬度的R_c^2和RMSEC分别为0.900和0.758，R_p^2和RMSEP分别为0.894和0.740，RPD达到2.207。类似地，不管原始全波段光谱数据是否经过Norm预处理，未经过有效特征提取过程建立的新梅硬度预测模型性能都最差，这同样验证了全波段光谱数据在未经有效特征提取的情况下，难以充分挖掘和利用光谱中的有效信息。图10为新梅SSC和硬度的Vis-NIR光谱回归预测模型的真实值与预测值分析结果，校正集和预测集的决定系数均在0.86以上，均方根误差低于0.65，且预测值紧密拟合在曲线周围，显示出较好的预测精度和可靠性。综上所述，选择CARS和BOSS分别作为新梅SCC SVR预测模型和新梅硬度PLS预测模型的最佳特征提取方法。

表3　新梅SSC和硬度预测模型结果表

Table 3　Results of SSC and firmness prediction models for prunes in Xinjiang

指标	模型	预处理	特征提取	变量数	校正集		预测集		RPD
					R_c^2	RMSEC	R_p^2	RMSEP	
SSC	SVR	\	\	2 013	0.841	0.743	0.782	0.743	1.583
		SNV	\	2 013	0.965	0.757	0.824	0.757	1.702
			CARS	59	0.929	0.651	0.866	0.651	1.956
			BOSS	50	0.890	0.714	0.844	0.714	1.874
			ICO	584	0.857	0.769	0.827	0.769	1.747
硬度	PLS	\	\	2 013	0.863	0.894	0.821	0.902	1.689
		Norm	\	2 013	0.865	0.886	0.848	0.858	1.828
			CARS	34	0.920	0.685	0.872	0.815	1.994
			BOSS	50	0.900	0.758	0.894	0.740	2.207
			ICO	303	0.898	0.759	0.865	0.881	1.892



（a）SSC　　（b）硬度

图10　基于不同预测模型新梅SSC和硬度的真实值与预测值散点图

Figure 10　Scatter plots of measured versus predicted values of SSC and firmness of prunes in Xinjiang based on different prediction models

在新梅SSC预测模型中，CARS是最优的特征选择方法，可在减少数据量的同时保证SVR模型预测的精度。而在新梅硬度预测模型中，相对于BOSS特征选择方法，使用CARS特征选择方法后虽然减少了数据量，但是过度的数据压缩导致重要信息丢失，使得CARS特征提取后构建的模型性能降低，因此BOSS成为了新梅硬度PLS预测模型最优的特征提取方法。沈懋生等[21]利用近红外光谱技术结合PLS建立了苹果气调贮藏期的SSC预测模型，在长波近红外波段下，MSC+SG-CARS-PLS模型取得了较好的预测精度，其R_p为0.900，RMSEP为0.478。刘燕

德等[12]利用Vis-NIR光谱漫反射技术结合PLS构建贮藏期间水蜜桃SSC和硬度预测模型,预测集R_p分别为0.820和1.003,RMSEP分别为0.841和0.829。总的来说,研究构建的贮藏期新梅SSC的SNV+CARS+SVR模型和新梅硬度的Norm-BOSS-PLS模型均表现出了较好的预测精度,表明采用Vis-NIR光谱技术预测新疆新梅贮藏品质是可行性的。

3　结论

以新疆新梅为研究对象,采集不同贮藏时间的新梅近红外光谱信息和理化指标,通过分析平滑去噪、标准正态变换、一阶导数和归一化4种预处理方法,结合3种特征筛选方法竞争自适应重加权采样、自举软收缩算法和区间组合优化,使用偏最小二乘和支持向量回归两种算法建立了新梅贮藏品质无损预测模型。结果表明,采用标准正态变换和归一化分别是预测新梅糖度和硬度的最佳方式。在构建新梅糖度无损预测模型中,比较不同特征提取方法在标准正态变换条件下的建模效果,其中通过竞争自适应重加权采样算法从全光谱2 013个波长数筛选出的59个特征波长所构建的标准正态变换+竞争自适应重加权采样+支持向量回归预测模型性能最优,其校正集决定系数和预测集决定系数分别为0.929和0.866,校正集均方根误差和预测集均方根误差均为0.651,残差预测偏差为1.956。在构建新梅硬度无损检测无损预测模型中,比较不同特征提取方法在归一化条件下的建模效果,其中通过自举软收缩算法从全光谱2 013个波长数筛选出的50个特征波长所构建的归一化—自举软收缩—偏最小二乘预测模型性能最优,其校正集决定系数和预测集决定系数分别为0.900和0.894,校正集均方根误差和预测集均方根误差分别为0.758和0.740,残差预测偏差为2.207。该研究在新梅贮藏品质无损预测方面初步验证了相关方法的可行性,但整体预测精度仍有待进一步提升。未来的研究应重点关注预处理方法与特征波长筛选策略的协同优化,以增强光谱数据的质量与代表性,并有效缓解高维波长变量之间的自相关性问题,从而提升预测模型的精度与稳健性。此外,模型性能还受限于试验样本在品种、产地、成熟度、贮藏条件以及果实自身物理特性等方面的差异。因此,后续研究应在样本多样性方面进行拓展,并建立可动态更新的建模数据库,以提升模型的泛化能力与鲁棒性,进而更好地满足果品品质快速检测在实际应用中的需求。

参考文献

[1] 新疆维吾尔自治区农业农村厅通告. 新疆新梅保鲜技术获突破保鲜期可达半年以上[EB/OL]. (2024-02-20) [2024-12-10]. https://nynct. xinjiang. gov. cn/xjnynct/c113576/202402/1c6b3e870f144c63a0929f55ef09ab4f.shtml.
Department of Agriculture and Rural Affairs of Xinjiang. Advanced preservation technology extends shelf life of prunes in Xinjiang to over six months[EB/OL]. (2024-02-20) [2024-12-10]. https://nynct. xinjiang. gov. cn/xjnynct/c113576/202402/1c6b3e870f144c63a0929f55ef09ab4f.shtml.

[2] 赵环环, 严衍禄. 噪声对近红外光谱分析的影响及相应的数学处理方法[J]. 光谱学与光谱分析, 2006, 26(5): 842-845.
ZHAO H H, YAN Y L. The effects of noise on NIR analysis and related mathematic pretreatments and models[J]. Spectroscopy and Spectral Analysis, 2006, 26(5): 842-845.

[3] 赵杰文, 张海东, 刘木华. 简化苹果糖度预测模型的近红外光谱预处理方法[J]. 光学学报, 2006(1): 136-140.
ZHAO J W, ZHANG H D, LIU M H. Preprocessing methods of near-infrared spectra for simplifying prediction model of sugar content of apples[J]. Acta Optica Sinica, 2006(1): 136-140.

[4] 占可, 陈季旺, 徐言, 等. 基于近红外光谱特征的冷冻小龙虾鲜度快速检测方法[J]. 食品科学, 2024, 45(2): 299-307.
ZHAN K, CHEN J W, XU Y, et al. A rapid detection method for freshness of frozen crayfish based on near-infrared spectroscopy [J]. Food Science, 2024, 45(2): 299-307.

[5] PURWANTO Y A, SARI H P, BUDIASTRA I W. Effects of preprocessing techniques in developing a calibration model for soluble solid and acidity in 'Gedong Gincu' mango using NIR spectroscopy[J]. International Journal of Engineering and Technology, 2015, 7(5): 1 921-1 927.

[6] LI H D, LIANG Y Z, XU Q S, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration[J]. Analytica Chimica Acta, 2009, 648 (1): 77-84.

[7] DENG B C, YUN Y H, CAO D S, et al. A bootstrapping soft shrinkage approach for variable selection in chemical modeling [J]. Analytica Chimica Acta, 2016, 908: 63-74.

[8] SONG X Z, HUANG Y, YAN H, et al. A novel algorithm for spectral interval combination optimization[J]. Analytica Chimica Acta, 2016, 948: 19-29.

[9] 罗澍寰, 孙武, 游杰, 等. 基于可见—近红外光谱法无损检测梨总酸含量[J]. 计算机与现代化, 2024(5): 80-84.
LUO S H, SUN W, YOU J, et al. Non-destructive detection of total acid content in pear based on visible-near infrared spectroscopy[J]. Computer and Modernization, 2024(5): 80-84.

[10] 母雯竹, 张贵宇, 张维, 等. 基于CARS-SPA特征提取的黄水淀粉近红外光谱定量模型优化[J]. 食品科学, 2024, 45(19): 8-14.
MU W Z, ZHANG G Y, ZHANG W, et al. Optimization of quantitative modeling of starch in Huangshui based on near-infrared spectral feature extraction using competitive adaptive reweighted sampling combined with successive projections algorithm[J]. Food Science, 2024, 45(19): 8-14.

[11] 王佳欣. 基于近红外光谱的猕猴桃可溶性固形物含量和硬

度预测模型传递方法研究[D]. 咸阳: 西北农林科技大学, 2024: 12-13.
WANG J X. Calibration transfer methods for predicting soluble solids content and firmness of kiwifruit based on near-infrared spectra[D]. Xianyang: Northwest Agriculture Forestry University, 2024: 12-13.

[12] 刘燕德, 张雨, 姜小刚, 等. 不同贮藏期水蜜桃硬度及糖度的检测研究[J]. 光谱学与光谱分析, 2021, 41(1): 243-249.
LIU Y D, ZHANG Y, JIANG X G, et al. Detection on firmness and soluble solid content of peach during different storage days [J]. Spectroscopy and Spectral Analysis, 2021, 41(1): 243-249.

[13] LIU Y, PENG Q W, YU J C, et al. Identification of tea based on CARS-SWR variable optimization of visible/near-infrared spectrum[J]. Journal of the Science of Food and Agriculture, 2020, 100(1): 371-375.

[14] MISHRA P, WOLTERING E, BROUWER B, et al. Improving moisture and soluble solids content prediction in pear fruit using near-infrared spectroscopy with variable selection and model updating approach[J]. Postharvest Biology and Technology, 2021, 171: 111348.

[15] 吕凯笛. 光谱技术结合化学计量学对黄芩及相关制剂的质量评价研究[D]. 郑州: 河南工业大学, 2024: 7-10.
LV K D. Study on the quality evaluation of Scutellaria baicalensis Georgi and related preparations based on the combination of spectral technology and chemometrics[D]. Zhengzhou: Henan University of Technology, 2024: 7-10.

[16] 倪尔冬, 林彩容, 操君喜, 等. 广西浦北茶树资源性状调研与其红茶适制性探究[J]. 食品安全质量检测学报, 2024, 15(20): 200-207.
NI E D, LIN C R, CAO J X, et al. Investigation on the characteristics of tea germplasms and suitability of black tea in Pubei County, Guangxi[J]. Journal of Food Safety and Quality, 2024, 15(20): 200-207.

[17] 毛欣然, 夏静静, 徐惟馨, 等. 手持式近红外光谱仪测定梨三种品质指标通用模型建模方法研究[J]. 光谱学与光谱分析, 2024, 44(2): 406-412.
MAO X R, XIA J J, XU W X, Study on modeling method of general model for measuring three quality indexes of pear by handheld near-infrared spectrometer[J]. Spectroscopy and Spectral Analysis, 2024, 44(2): 406-412.

[18] 罗林, 庹先国, 张贵宇, 等. 基于异常样品剔除的酒醅近红外定量分析模型的精度提升[J]. 食品安全质量检测学报, 2022, 13(9): 3 017-3 025.
LUO L, TUO X G, ZHANG G Y, 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): 3 017-3 025.

[19] 蔡杰. 滤光片型近红外光谱仪检测波长选择方法研究[D]. 长春: 吉林大学, 2024: 11-24.
CAI J. Study on the detection wavelength selection method of filter-type near-infrared spectrometer[D]. Changchun: Jilin University, 2024: 11-24.

[20] 吴莎莎, 王振杰, 江梦薇, 等. 基于多成熟度光谱信息融合的阿森泰克苹果品质预测模型研究[J]. 食品工业科技, 2024, 45(7): 294-305.
WU S S, WANG Z J, JIANG M W, et al. Prediction model of aztec apples quality based on the fusion of multi-maturity spectral information[J]. Science and Technology of Food Industry, 2024, 45(7): 294-305.

[21] 沈懋生, 赵娟. 基于近红外光谱技术检测苹果气调贮藏期可溶性固形物含量[J]. 食品安全质量检测学报, 2022, 13(17): 5 495-5 503.
SHEN M S, ZHAO J. Detection of soluble solids content in apples during controlled atmosphere storage based on near-infrared spectroscopy[J]. Journal of Food Safety and Quality, 2022, 13(17): 5 495-5 503.