

刘燕德,黎丽莎,李斌,等.多品种苹果可溶性固形物近红外无损检测通用模型研究[J].华中农业大学学报,2022,41(2):237-244.
DOI:10.13300/j.cnki.hnlkxb.2022.02.028

多品种苹果可溶性固形物近红外无损检测通用模型研究

刘燕德¹,黎丽莎¹,李斌¹,宋焯²,朱向荣³,姜延泉⁴

1. 华东交通大学机电与车辆工程学院,南昌 330013; 2. 中华全国供销合作总社济南果品研究院,济南 250220;
3. 湖南省农产品加工研究所,长沙 410125; 4. 齐鲁泉源供应链有限公司,烟台 265307

摘要 苹果可溶性固形物含量(soluble solid content,SSC)是影响果实质量的重要因素,利用近红外光谱(NIR)技术则可以实现对苹果SSC的无损检测。为获得稳健的多品种苹果无损检测通用模型,本研究将模型更新方法与变量筛选方法相结合,对红富士(Red Fuji)、青苹果(Green apple)、黄元帅(Golden Delicious)、红玫瑰(Rose)和乐淇(Lokit)等5个品种苹果的SSC进行无损检测。结果表明,更新后的新模型可以实现对5个品种苹果的SSC的高精度预测,此时模型预测均方根误差(RMSEP)为0.698%,预测相关系数(R_p)为0.904,预测偏差为0.074%,范围误差比(RPD)可达2.340。为识别和提取光谱的重要信息波段,还采用竞争性自适应重加权算法(CARS)、连续投影算法(SPA)和无信息变量消除算法(UVE)3种波段筛选方法优化模型。结果表明,CARS更能有效地选取出有效变量,建立的模型对新品种苹果的预测性能有明显改善,RMSEP为0.587%, R_p 为0.928,预测偏差减少到-0.052%,RPD=2.684。

关键词 苹果;近红外光谱;无损检测;通用模型;模型更新;波段筛选;自适应重加权;连续投影算法;无信息变量消除

中图分类号 O657.3 **文献标识码** A **文章编号** 1000-2421(2022)02-0237-08

可溶性固形物含量(soluble solid content,SSC)是影响苹果口感的重要因素,常作为苹果品质的评价指标^[1]。近红外光谱(near infrared spectrum,NIR)技术虽然能实现对苹果的SSC无损检测,但基于光谱数据开发的模型通常只对校正集同批次样本有效,对于不同品种的果实,由于内部的生物结构和物质含量不同,将导致预测效果不理想,出现预测偏差^[2]。因此,建立多品种苹果SSC的无损检测通用模型至关重要。

模型更新是建立通用模型的方法之一。以芒果(*Mangifera indica* L.)的近红外光谱无损检测为例,与单季节数据相比,合并多季节数据能够提高近红外模型的鲁棒性^[3],而该方法同样适用于李(*Prunus salicina* Lindl.)和苹果的近红外光谱检测。Louw等^[4]研究表明,加入多个品种的额外样本可以增强李质量参数预测模型的通用性,总可溶性固形物的预测结果的决定系数(R^2)为0.959,预测均方根误差(root mean square error of prediction, RMSEP)为0.453%。Peirs等^[5]在原始苹果SSC预测模型中加

入了来自不同年份的苹果,结果发现,当模型中包含更多的变量信息时,模型的准确性明显提高,模型的RMSEP从2.92%降到了0.95%。模型更新不仅能应用于水果无损检测领域,还应用于玉米种子^[6]、黄酒^[7]等产品的近红外无损检测领域中。模型更新的校正主要通过添加额外的样本信息实现,例如多季节样本信息、多品种样本信息^[8]或其他信息^[9]。再用新样本重新校正旧模型,用于新批次的预测。模型更新虽能改善模型性能,却需要新的样本加入,且需要的新样品数量往往不能直接确定^[10]。因此,需要研究模型更新所需的最佳样本数量,以尽量少的新样本实现模型优化。

近红外光谱(750~2 500 nm)由多种化学成分的特定波长的峰组成。例如对水果的水和糖进行预测,由于O—H和C—H键的第三倍频主要与光谱范围750~950 nm有关^[11-12],这意味着,若使用O—H和C—H键倍频区域以外的波长信息建立水果的干物质或SSC的预测模型可能会导致信息冗余。冗余信息干扰问题会导致建模效果无法达到最优。波段

收稿日期:2021-11-01

基金项目:国家自然科学基金项目(31760344);江西省国家科技奖后备培育项目(20192AEI91007)

刘燕德,E-mail:jxliuyd@163.com

筛选常用于解决此类问题,通过提取特征波段减轻建模负担,同时增强模型的稳健性,建立鲁棒性更优的模型^[13]。在化学计量学领域,主要存在两种类型的变量选择方法,即基于区间的筛选和基于过滤器的筛选^[14]。基于区间的筛选是在变量中选取最具预测性的子区间;基于过滤器的筛选主要根据特定标准(如最大协方差或使用用户定义的阈值)过滤出单个变量。当数据为连续信号时,如近红外光谱信号,通常采用基于区间的变量筛选方法。在NIR分析中,常用的变量筛选方法通常有竞争性自适应重加权算法(competitive adaptive reweighted sampling, CARS)^[15]、连续投影算法(successive projections algorithm, SPA)^[16]、无变量信息消除(uninformative variable elimination, UVE)^[17]、遗传算法(genetic algorithm, GA)^[18-19]、模拟退火(simulated annealing, SA)^[20]等。

由于传统建立多品种水果无损检测通用模型的过程采用的训练集样本较多,导致过多的人力和物力成本的消耗,且波段筛选算法能够在简化模型的同时增强模型的稳健性。因此,本研究依据NIR分析技术,采用将模型更新方法与波段筛选算法结合的研究策略,建立了多品种苹果SSC无损预测模型,旨在简化模型并节约建模成本的前提下,为建立水果无损检测通用模型提供一种新的思路。

1 材料与方 法

1.1 样品制备

本试验分别对5个品种的苹果进行了检测,分别为红富士(Red Fuji)、青苹果(Green apple)、黄元帅(Golden Delicious)、红玫瑰(Rose)和乐淇(Lokit)。其中,红富士产于中国陕西省延安市洛川县,共240个样品;青苹果产于中国甘肃省平凉市静宁县,共

124个样品;黄元帅产于中国辽宁省大连市瓦房店市,共124个样品;而红玫瑰和乐淇均产于新西兰,样品数均为124个。在进行数据采集前,样品均置于(20±1)℃的实验室环境中24~72 h。为获得苹果通用模型,根据大小、果皮锈蚀程度和形状,选择了品相各异的苹果。

1.2 光谱采集

光谱测量采用了深圳和而泰智能控制股份有限公司研发的商业化手持式便携式光谱检测仪(型号C-life 01A,内部搭载DLP900近红外光谱仪模块,波段范围:900~1 700 nm,光谱分辨率为3.50 nm,共228个波长点)。光谱仪内部安装卤钨灯提供光源,并内置1块可转动的漫反射标准白板,以此板为参考计算漫反射率。在每个样品赤道部位做4个均匀分布的标记,将水果的标记处正对检测口。手动按下检测按钮进行光谱信号采集,每次光谱采集前仪器会自动进行漫反射校正,并自动采集7次信号取平均作为仪器输出信号。每个样品将获得4条光谱,再取其平均光谱作为该样品的输出光谱信号。采集好的光谱信号自动存储在光谱仪中,光谱仪搭载的DLP NIRscan Nano评估模块使用USB 1.1人机界面设备(HID)进行通信,实现与PC端的数据传输。

1.3 SSC检测

将苹果的SSC作为预测模型的真值,对样品进行近红外光谱采集后,从苹果果腹的赤道处标记的4个面分别切出1 cm厚的切片。使用温度补偿糖度计(型号PAL-1; Atago Co., Tokyo, Japan)进行测定,使用前将糖度计放置于光亮处,用蒸馏水对其进行零点校正。再取苹果切片果肉部分榨汁,将果汁滴在糖度计镜面中央进行检测,每测完1个切片,用蒸馏水清洗镜面,将3次测量平均值作为其SSC值。依次测得所有样品SSC统计情况如表1所示。

表1 各批次样本SSC分布情况

Table 1 SSC distribution of each batch of samples

批次 Batch	品种 Variety	样本数 Sample number	最小值 Minimum	最大值 Maximum	均值 Mean	标准偏差 Standard deviation
1	红富士 Red Fuji	124	8.00	16.80	13.34	1.53
2	红富士 Red Fuji	116	9.50	16.30	13.40	1.55
3	青苹果 Green apple	124	8.10	16.10	12.86	1.86
4	黄元帅 Golden Delicious	124	10.40	17.50	14.06	1.45
5	红玫瑰 Rose	124	10.80	17.20	13.53	1.33
6	乐淇 Lokit	124	11.50	18.00	14.37	1.46

1.4 光谱预处理

在使用光谱仪采集光谱的过程中会遇到某些干扰(样品背景、杂散光、温度变化和光程变化等)导致基线平移、旋转以及光谱散射,为建立稳定的数学模型,需要对原始光谱进行预处理,最大限度地滤除无用及干扰信号,以消除或减弱非目标因素的影响并提高信噪比。本研究采用多元散射校正(multiplicative scatter correction, MSC)算法对采集到的漫反射光谱信号预处理,主要是利用计算得到的平均光谱对个别出现的基线偏移现象的光谱信号进行修正,从而减小由散射水平不一致带来的光谱差异的影响,进而使光谱信号与真值间的相关性得到一定程度提高^[21]。

1.5 模型的建立与评价

本研究主要采用偏最小二乘回归(partial least squares regression, PLSR)方法建立苹果SSC预测模型。PLSR是一种经典数据分析方法,通过主成分分析(principal component analysis, PCA)对数据进行降维,再计算残余方差确定主成分个数(principal components, PCs),最终建立简易的线性模型。而PCs越小,说明数据空间的维数越低,所建的模型越简易,因此,PCs亦可作为模型评价参数之一。

同时,为提高模型对不同品种苹果的预测性能,采用模型更新方法。模型更新是将新批次的一些样品光谱数据合并到旧批次样品数据中,并重新建模,从而实现旧模型的更新。将红富士中的124个样品作为第1批次,即旧模型的训练集,剩余116个红富士作为第2批次样品,作为该模型的验证集;而后的青苹果、黄元帅、红玫瑰和乐淇按品种依次记为第3、第4、第5和第6批次,并按批次顺序依次将更新的样本纳入旧模型中。利用光谱理化值共生距离(sample set partitioning based on joint X-Y distance algorithm, SPXY)算法^[22],从新批次苹果中分别选取5、10、15和20个样本作为模型更新的样本,实现用尽量少的样本数来改善模型性能的目的。

主要采用校正均方根误差(root mean square error of calibration, RMSEC)和校正相关系数(correlation coefficient of calibration, R_c)评价SSC与光谱数据之间的校正方程,利用RMSEP、预测相关系数(correlation coefficient of prediction, R_p)、预测偏差(prediction bias)以及范围误差比(ratio of performance to standard deviate, RPD)评价模型的分析性能。其中,RPD是评价模型分辨能力的重要评价参

数,RPD值越大说明模型越可靠,Chang等^[23]定义了3个类别以考虑模型的可靠性:优秀模型,RPD>2.0;可靠模型,RPD为1.4~2.0;不可靠模型,RPD<1.4。

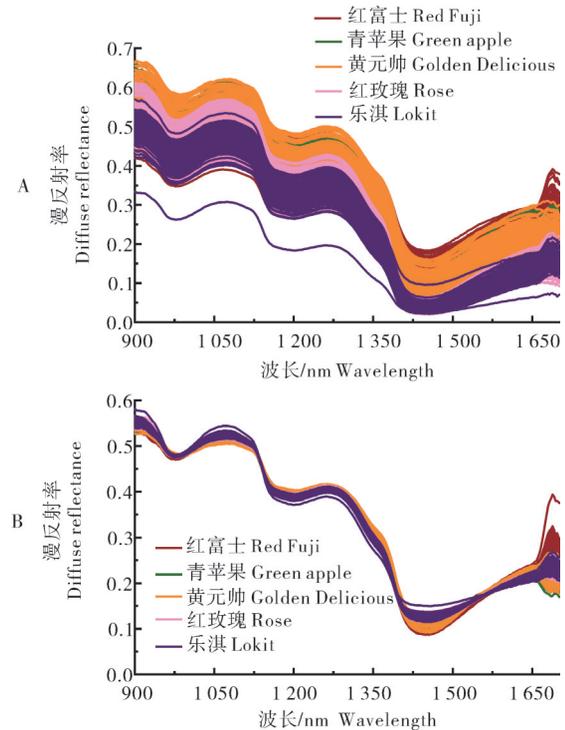
1.6 数据分析软件

本研究的数据处理过程主要使用的软件是Matlab R2014b(The Mathwork, USA)和The Unscramber 9.7(Camo Analytics, Norway),并采用Origin 2018C(OriginLab, USA)进行作图。

2 结果与分析

2.1 光谱特征分析

图1为5个品种苹果的原始漫反射率和MSC预处理后的光谱图。在900~1700 nm的光谱范围内,该区域主要承载了O—H和N—H二倍频信息,O—H、C—H和N—H三倍频信息以及C—H四倍频的信息,而SSC的化学结构主要就是由这类化学键组成,因此,利用近红外光谱可用于检测苹果的SSC。原始光谱主要在1050、1275 nm附近有2个主峰(图1A),1050 nm附近的波峰主要是由N—H三倍频、C—H与二倍频的合频的伸缩振动形成,而1275 nm附近的波峰则主要和C—H的三倍频的伸缩振动



A:原始光谱反射率 Reflectance of original spectral; B:MSC处理后的光谱反射率 Spectral reflectance after MSC.

图1 5个品种的苹果漫反射率光谱图

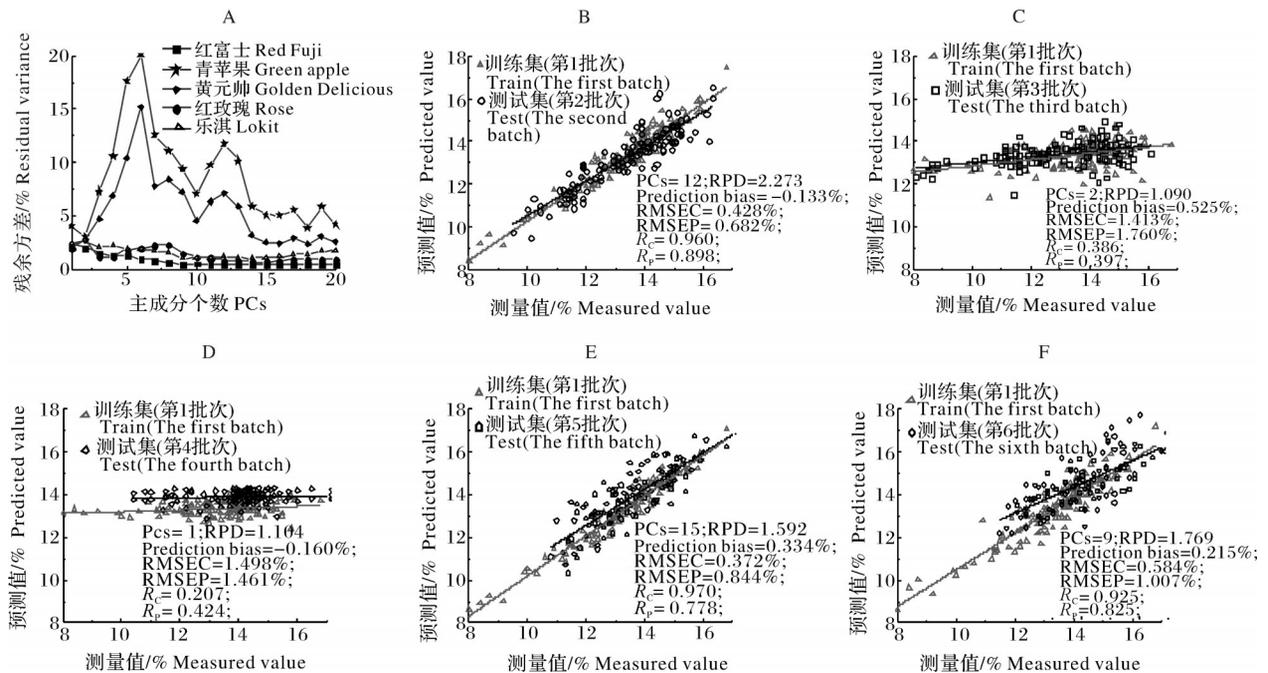
Fig.1 Diffuse reflectance spectra of five apple varieties

有关。利用MSC方法预处理后,较好地消除了由于苹果内部散射水平不同带来的光谱差异,从而增强光谱与数据之间的相关性(图1B)。

2.2 偏最小二乘回归分析

图2为未进行变量选择和模型更新的PLSR预测建模。图2A显示了选取不同主成分数时对SSC预测模型验证集残余方差的影响。本研究共选择了20个主成分并同时预测SSC,根据图2A选择方差最小时所对应的主成分数,即为最终建模时选取的主成分数,此时模型预测性能最好。各品种苹果

的SSC预测模型如图2B-F所示,可以看出,直接利用第1批次的红富士建立的模型对同品种苹果的预测性能最好(RMSEP=0.682%; RPD=2.273),对第6批次的乐淇(RMSEP=1.007%, RPD=1.769)和第5批次的红玫瑰(RMSEP=0.844%, RPD=1.592)的红玫瑰苹果预测效果其次,对第3批次的青苹果和第4批次的黄元帅建立的模型准确度低且模型不具有可靠性(RPD<1.4),这表明在第1批次的红富士上训练的模型不能直接应用到其他品种的苹果上。



A: 主成分分析 Principal component analysis; B: 红富士 SSC 预测模型 Red Fuji's SSC prediction model; C: 青苹果 SSC 预测模型 Green apple's SSC prediction model; D: 黄元帅 SSC 预测模型 Golden Delicious's SSC prediction model; E: 红玫瑰 SSC 预测模型 Rose's SSC prediction model; F: 乐淇 SSC 预测模型 Lokit's SSC prediction model.

图2 全波段光谱 PLSR 模型

Fig.2 PLSR modelling on complete spectra

2.3 模型更新

为了提高旧模型对新批次苹果的预测性能,依次将第3、第4、第5和第6批次的样本纳入第1批次模型,更新过程数据如表2所示。在依次更新过程中,PLSR模型的性能总体得到了提高。将第3批次更新样品纳入模型时,在20个新样本的模型更新情况下的模型总体预测性能最好,更新后的PLSR模型 R_p 显著增加,从0.397增加到0.918, RMSEP从1.760%减少到0.727%,预测偏差从0.525%减少到0.128%。在纳入第4批更新样本时,更新样本数为15个最佳,且同样获得了较好的优化效果: R_p 从0.424增加到0.898, RMSEP从1.461%减少到0.752%。纳入第5批更新样本时,更新样本数为20

时模型预测性能较好,此时模型 R_p 从0.778增加到0.898, RMSEP从0.844%减少到0.697%,预测偏差从0.334%减少到0.155%。在纳入第6批数据时,更新样本为5时预测模型性能最佳,此时 R_p 从0.825增加到0.904, RMSEP从1.007%减少到0.698%,预测偏差从0.215%减少到0.074%。以上结果表明,在保证建模数据尽量少的前提下,模型更新方法既可以提高模型对新品种SSC的预测性能,同时还可以保留对旧品种SSC的预测能力,最终建立了5个品种苹果的无损检测通用模型(RMSEC=0.624%; R_c =0.924; RMSEP=0.698%; R_p =0.904),且此时模型RPD=2.340,说明该模型具有较好的分辨能力。

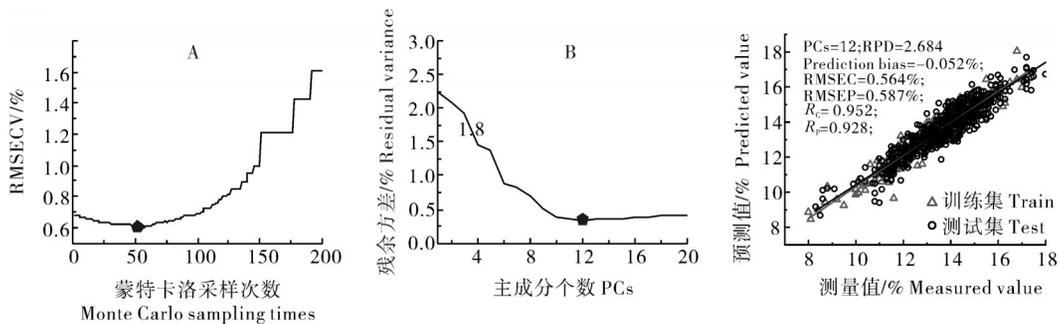
表2 更新后的PLSR模型预测结果
Table 2 Predicted results of updated PLSR model

批次 Batch	更新样本数 Number of updated samples	PCs	RMSEC/%	R_c	RMSEP/%	R_p	RPD	预测偏差/% Prediction bias
3	5	7	0.756	0.878	1.097	0.811	1.709	0.234
	10	17	0.455	0.958	0.881	0.884	2.139	0.265
	15	17	0.465	0.956	0.784	0.909	2.400	0.174
	20	17	0.472	0.956	0.727	0.918	2.522	0.128
4	5	11	0.615	0.926	0.810	0.880	2.105	0.143
	10	11	0.624	0.924	0.773	0.892	2.212	0.183
	15	11	0.628	0.923	0.752	0.898	2.273	0.165
	20	10	0.683	0.910	0.732	0.901	2.305	0.250
5	5	12	0.613	0.927	0.729	0.886	2.157	0.196
	10	12	0.615	0.927	0.715	0.891	2.203	0.188
	15	12	0.615	0.927	0.718	0.890	2.193	0.186
	20	12	0.619	0.925	0.697	0.898	2.273	0.155
6	5	12	0.624	0.924	0.698	0.904	2.340	0.074
	10	12	0.632	0.924	0.695	0.903	2.328	0.117
	15	12	0.633	0.924	0.694	0.902	2.316	0.137
	20	12	0.642	0.922	0.694	0.901	2.305	0.137

2.4 模型的简化

1) CARS 变量筛选。图 3A 显示的是采用 CARS 方法进行光谱变量筛选后的结果,采用蒙特卡罗采样方法反复迭代采样,并比较每次采样的 RMSECV 值,直至找到最小交叉验证均方根误差 (root mean square error of cross-validation, RMSECV) 值所对应的最优变量集。由图 3A 可知,第 52 次采样时所包含的变量子集的 RMSECV 最小为 0.602%,

此时对应的变量数为 68 个。而在此之后,采样次数增加导致过多有效信息波段被剔除, RMSECV 随之增加。图 3B 所示是对筛选后的光谱数据进行主成分分析,确定 PCs 为 12 时对应的方差最小,图 3C 展示了此时模型预测结果为 RMSEC=0.564%; $R_c=0.952$; RMSEP=0.587%; $R_p=0.928$, 并且根据 RPD=2.684 证明该模型是一个优秀的预测模型。



A: CARS 运行结果 CARS running result; B: 主成分分析 Principal component analysis; C: 散点图 Scatter diagram.

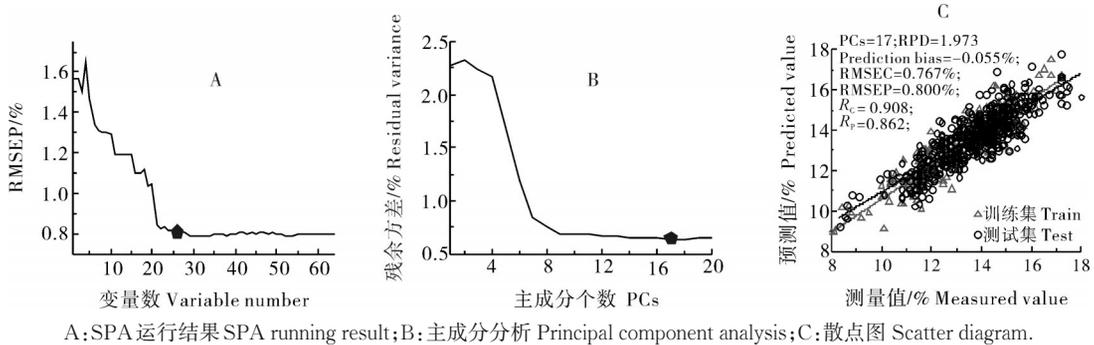
图3 CARS 筛选变量后的 PLSR 模型

Fig.3 PLSR model after CARS screening variables

2) SPA 变量筛选。SPA 变量筛选及建模过程如图 4 所示。从图 4A 可以观察到 SPA 选择过程中 RMSEP 的变化情况,在选择变量 20 个之前曲线下降迅速,表明苹果中 SSC 光谱变量应该至少选择 20 个以免产生过拟合问题;从选择 26 个变量以后,曲线明

显平缓,仅有略微波动, RMSEP 无明显变化,因此, SPA 最终选取了 26 个波段,此时 RMSEP=0.808%。再对变量进行主成分分析,确定最优 PCs 后建模结果如图 4C 所示,此时模型的 RMSEC=0.767%, RMSEP=0.800%, $R_c=0.908$, $R_p=0.862$, 虽然此时模型

预测偏差较小,但模型的预测准确度和分辨能力 (RPD=1.973)均不如CARS选取的变量所建模型, 这主要是SPA运行过程中剔除了过多的有效信息导致的。



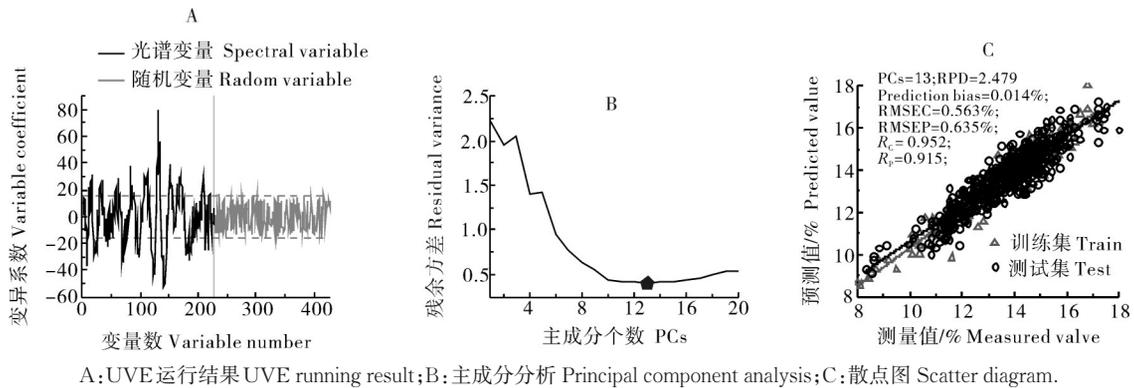
A: SPA运行结果 SPA running result; B: 主成分分析 Principal component analysis; C: 散点图 Scatter diagram.

图4 SPA筛选变量后的PLSR模型

Fig.4 PLSR model after SPA screening variables

3) UVE变量筛选。UVE选取波段的主要思想是把变异系数(coefficient of variation, CV)作为变量选择的衡量依据,运行结果如图5A所示。其中,200个随机变量纳入光谱数据中,通过交叉验证建立PLSR模型,得到变异系数矩阵。随机变量的CV值分布如图5中灰色虚线右侧所示,再将随机变量的最大CV值(16.223)作为阈值,即图5A中水平灰色虚线。当光谱变量对应的CV值低于16.223时,则剔除

该变量,最终筛选得到110个变量用于建模。经主成分分析后建立的PLSR模型如图5C所示, RMSEC=0.563%, RMSEP=0.635%, $R_c=0.952$, $R_p=0.915$ 。根据RPD=2.479可知,具有较好的预测性能,预测结果虽与CARS的建模结果相近,但由于UVE筛选的变量过多,存在冗余变量,因此,该方法并不适用于该通用模型。



A: UVE运行结果 UVE running result; B: 主成分分析 Principal component analysis; C: 散点图 Scatter diagram.

图5 UVE筛选变量后的PLSR模型

Fig.5 PLSR model after UVE screening variables

3 讨论

基于NIR技术建立苹果SSC的预测模型是苹果品质无损检测的常用技术之一。本研究首先采用了模型更新方法,将124个红富士样本作为原始建模数据。在模型更新前,原始模型对其余4种苹果的预测性能并不理想,特别是对青苹果和黄元帅2个品种几乎没有预测能力,这是由于红富士与青苹果、黄元帅内部物质的组成及含量差异较大。再将青苹果、黄元帅、红玫瑰和乐淇样本作为模型更新样本按批次加入原始数据中。在更新过程中,模型对SSC的预测性能得到了提高,当加入20个青苹果、15个黄元

帅、20个红玫瑰和5个乐淇样本时,模型对SSC的预测精度最高(RMSEP=0.698%; $R_p=0.904$),且模型的RPD=2.340,证明该模型已达到优秀模型的标准。

为进一步提高模型鲁棒性并简化模型,本研究结合了3种波段筛选算法选取光谱波段,去除冗余光谱信息。SPA是一种使矢量空间共线性最小化的变量筛选算法,其优点在于能够有效提取特征波段,从而剔除冗余的变量信息。根据样品的反射率光谱发现,900~1700 nm范围内的特征峰较少,而SPA只提取了特征峰附近部分波段,损失的信息变量较多,

从而导致模型性能下降($RMSEP=0.800\%$; $R_p=0.862$)。UVE的主要筛选机制是在原始光谱数据中加入的随机变量矩阵,并利用留一法建立PLSR模型,再根据所得的回归系数矩阵确定筛选阈值,剔除回归系数较低的波段,实现模型的简化。UVE算法虽能使模型的预测精度提高($RMSEP=0.635\%$; $R_p=0.915$),但保留的波段过多,依旧存在信息变量冗余的问题。CARS依据各波段的回归系数结合自适应重加权采样(adaptive reweighted sampling, ARS)技术选取变量,筛选出权重较大的变量集合,并对各变量集合预建模计算交叉验证均方根误差,依此作为变量集合的评价参数。CARS引入指数衰减函数控制筛选的变量数,具有较高计算效率,能够有效筛选出重要信息变量,实现提高模型稳健性的同时尽可能地简化模型($RMSEP=0.587\%$; $R_p=0.928$)。

研究表明,将模型更新与波段筛选算法结合的策略能够有效应用于多品种苹果无损检测通用模型的建立,具有一定的可行性。后续研究可以将该策略应用于其他水果或其他内部品质预测通用模型的建立,并通过改进预处理、波段筛选和建模方法提升模型性能。

参考文献 References

- [1] 李龙,彭彦昆,李永玉.苹果内外品质在线无损检测分级系统设计与试验[J].农业工程学报,2018,34(9):267-275.LI L, PENG Y K, LI Y Y. Design and experiment on grading system for online non-destructive detection of internal and external quality of apple[J]. Transactions of the CSAE, 2018, 34(9): 267-275 (in Chinese with English abstract).
- [2] TEH S L, COGGINS J L, KOSTICK S A, et al. Location, year, and tree age impact NIR-based postharvest prediction of dry matter concentration for 58 apple accessions[J/OL]. Postharvest biology and technology, 2020, 166: 111125 [2021-11-01]. <https://doi.org/10.1016/j.postharvbio.2020.111125>.
- [3] RUNGPICHAYAPICHET P, MAHAYOTHEE B, NAGLE M, et al. Robust NIRS models for non-destructive prediction of postharvest fruit ripeness and quality in mango[J]. Postharvest biology and technology, 2016, 111: 31-40.
- [4] LOUW E D, THERON K I. Robust prediction models for quality parameters in Japanese plums (*Prunus salicina* L.) using NIR spectroscopy[J]. Postharvest biology and technology, 2010, 58(3): 176-184.
- [5] PEIRS A, TIRRY J, VERLINDEN B, et al. Effect of biological variability on the robustness of NIR models for soluble solids content of apples[J]. Postharvest biology and technology, 2003, 28(2): 269-280.
- [6] 郭东生.农产品品质检测模型更新方法研究[D].无锡:江南大学,2018.GUO D S. Research on the updating method of agricultural product quality detection model[D]. Wuxi: Jiangnan University, 2018 (in Chinese with English abstract).
- [7] 陈令奕.近红外模型更新方法及在食品检测中的应用[D].无锡:江南大学,2018.CHEN L Y. Research on NIR model updating method with application in food detection[D]. Wuxi: Jiangnan University, 2018 (in Chinese with English abstract).
- [8] 刘燕德,徐海,孙旭东,等.不同品种苹果糖度近红外光谱在线检测通用模型研究[J].光谱学与光谱分析,2020,40(3):922-928.LIU Y D, XU H, SUN X D, et al. Development of multi-cultivar universal model for soluble solid content of apple online using near infrared spectroscopy[J]. Spectroscopy and spectral analysis, 2020, 40(3): 922-928 (in Chinese with English abstract).
- [9] NORDEY T, JOAS J, DAVRIEUX F, et al. Robust NIRS models for non-destructive prediction of mango internal quality[J]. Scientia horticulturae, 2017, 216: 51-57.
- [10] 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/OL]. Postharvest biology and technology, 2021, 171: 111348 [2021-11-01]. <https://doi.org/10.1016/j.postharvbio.2020.111348>.
- [11] MISHRA P, ASAARI M S M, HERRERO-LANGREO A, et al. Close range hyperspectral imaging of plants: a review[J]. Bio-systems engineering, 2017, 164: 49-67.
- [12] 余怀鑫,潘新星,朱旭,等.基于透射光谱技术的温州蜜柑含水量检测[J].华中农业大学学报,2021,40(1):86-92.YU H X, PAN X X, ZHU X, et al. Detecting water content of Satsuma mandarin based on transmission spectroscopy[J]. Journal of Huazhong Agricultural University, 2021, 40(1): 86-92 (in Chinese with English abstract).
- [13] MEHMOOD T, LILAND K H, SNIPEN L, et al. A review of variable selection methods in partial least squares regression[J]. Chemometrics and intelligent laboratory systems, 2012, 118: 62-69.
- [14] MEHMOOD T, SÆBØ S, LILAND K H. Comparison of variable selection methods in partial least squares regression[J]. Journal of chemometrics, 2020, 34(6): 62-69.
- [15] JIANG Z Q, DU Y P, CHENG F P, et al. A simple multiple linear regression model in near infrared spectroscopy for soluble solids content of pomegranate arils based on stability competitive adaptive re-weighted sampling[J]. Journal of near infrared spectroscopy, 2021, 29(3): 140-147.
- [16] CHENG J H, SUN D W, PU H B. Combining the genetic algorithm and successive projection algorithm for the selection of feature wavelengths to evaluate exudative characteristics in frozen-thawed fish muscle[J]. Food chemistry, 2016, 197: 855-863.
- [17] LI C, ZHAO T L, LI C, et al. Determination of gossypol content in cottonseeds by near infrared spectroscopy based on Monte Carlo uninformative variable elimination and nonlinear calibration methods[J]. Food chemistry, 2017, 221: 990-996.
- [18] NTURAMBIRWE J F I, NIEUWOUDT H H, PEROLD W J, et al. Non-destructive measurement of internal quality of apple fruit by a contactless NIR spectrometer with genetic algorithm model optimization[J/OL]. Scientific African, 2019, 3: e00051

- [2021-11-01]. <https://doi.org/10.1016/j.sciaf.2019.e00051>.
- [19] 王胜鹏,郑鹏程,龚自明,等.青砖茶茶汤滋味品质的近红外快速无损评价[J].华中农业大学学报,2020,39(3):113-119. WANG S P, ZHENG P C, GONG Z M, et al. Rapid evaluation of taste quality for Qingzhuang tea soup based on near infrared spectroscopy [J]. Journal of Huazhong Agricultural University, 2020, 39(3): 113-119 (in Chinese with English abstract).
- [20] SHI J Y, HU X T, ZOU X B, et al. A heuristic and parallel simulated annealing algorithm for variable selection in near-infrared spectroscopy analysis [J]. Journal of chemometrics, 2016, 30(8): 442-450.
- [21] 王建华,左玲,李志忠,等.基于高光谱技术的黑土地微量金属元素探测方法及地学意义[J].地质力学学报,2021,27(3):418-429. WANG J H, ZUO L, LI Z Z, et al. A detection method of trace metal elements in black soil based on hyperspectral technology: geological implications [J]. Journal of geomechanics, 2021, 27(3): 418-429 (in Chinese with English abstract).
- [22] GUO L B, ZHENG W N, CHEN F, et al. Meat species identification accuracy improvement using sample set portioning based on joint x - y distance and laser-induced breakdown spectroscopy [J]. Applied optics, 2021, 60(20): 5826-5831.
- [23] CHANG C W, LAIRD D A, MAUSBACH M J, et al. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties [J]. Soil science society of America journal, 2001, 65(2): 480-490.

General near-infrared model of soluble solids content in multi-variety apples

LIU Yande¹, LI Lisha¹, LI Bin¹, SONG Ye², ZHU Xiangrong³, JIANG Yanquan⁴

1. School of Mechatronics and Vehicle Engineering, East China Jiaotong University, Nanchang 330013, China;

2. Jinan Fruit Research Institute, All China Federation of Supply and Marketing Cooperatives, Ji'nan 250220, China;

3. Hunan Agricultural Products Processing Institute, Changsha 410125, China;

4. Qilu Quanyuan Supply China Co., Ltd., Yantai 265307, China

Abstract China is a big country of apple producers and exporters, so it is essential to control the quality of exported apples. The soluble solids content (SSC) in apples is an essential factor affecting the quality of apples. Non-destructive inspection of SSC in apples can be realized by near-infrared spectroscopy (NIR). The model update method combined with the variable screening method was used to perform non-destructive inspection on the SSC in five apple varieties including Red Fuji, Green apple, Golden Delicious, Rose and Lokit to obtain a general robust multi-variety apple model. The results showed that the updated new model achieved high-precision prediction of the SSC in five varieties of apples. At this time, the root mean square error of prediction (RMSEP) of the model was 0.698%, the correlation coefficient of prediction (R_p) was 0.904, the prediction deviation was 0.074%, and the ratio of performance to standard deviate (RPD) reached 2.340. The model was optimized with three waveband selection methods including competitive adaptive reweighted sampling (CARS), successive projections algorithm (SPA) and uninformative variable elimination (UVE) to identify and extract important information bands of the spectrum. The results showed that CARS selected effective wavebands more effectively. The established model significantly improved the analytical performance in new apple varieties, with RMSEP of 0.587%, R_p of 0.928, and prediction bias reduced to -0.052% , with RPD of 2.684. Therefore, the strategy of combining model updating with waveband selection has great application potential in studying and developing the general model of portable near-infrared detector for multi-variety fruits.

Keywords apple; near infrared spectroscopy; nondestructive testing; general model; model update; waveband selection; competitive adaptive reweighted sampling; successive projections algorithm; uninformative variable elimination

(责任编辑:陆文昌)