姜萌微,丁文文,王　蕊,等. 基于偏最小二乘法的新疆杏可溶性固形物含量的无损检测[J]. 江苏农业科学,2020,48(9):237 - 241.
doi:10.15889/j.issn.1002 - 1302.2020.09.045

基于偏最小二乘法的新疆杏可溶性固形物含量的无损检测

姜萌微, 丁文文, 王　蕊, 杨嘉玲, 郭　玲
(塔里木大学植物科学学院,新疆阿拉尔 843300)

摘要:利用近红外光谱无损检测技术,以成熟的库车小白杏为研究对象,先通过一阶导数混合标准正态变量变换法进行光谱预处理,再利用偏最小二乘法建立杏可溶性固形物含量预测模型。结果表明,最佳因子数为 4,模型的相关系数为 0.964 45,校正集标准偏差为 0.215,预测集标准偏差为 0.456 时模型最佳。结果表明,基于近红外光谱技术测定杏的可溶性固形物含有具有可行性,可以为库车小白杏的无损检测提供重要参考与技术支持。

关键词:杏;可溶性固形物;近红外光谱;偏最小二乘法;无损检测
中图分类号: S662.201;TS207　　**文献标志码:** A　　**文章编号:**1002 - 1302(2020)09 - 0237 - 05

杏是新疆南部的特色果树之一,具有适应性强、管理容易、果实早熟的特点,常以杏麦间作、杏棉间作,因其经济效益高、生态效益明显等特性,被广泛种植。根据 2018 年新疆统计年鉴可知,新疆杏种植面积已达 111 246 hm^2,产量达 933 088 t,成为当地农民重要的经济来源。库车小白杏是新疆重要的地方品种,其果实卵圆形、呈黄白色,果面光滑无毛,肉质较细,纤维素含量少,品质上乘,是鲜食及加工的优质杏品种[1]。

近红外光谱(NIRS)分析技术具有简单、快速、准确、无损以及定量检测等优点[2],现已广泛地应用在桃[3]、苹果[4]、梨[5]、樱桃[6]、蜜橘[7]、葡萄[8]等新鲜果品内在品质检测。但是,目前针对杏的品质鉴定仍旧采用破损法[9],并且关于杏的采摘、分级、分选也同样依靠人工经验判断的方法。然而,依靠人工经验判断的方法易受主观因素影响,误差大、效率低。因此,采用无损检测技术鉴定新疆杏内部品质,建立一系列杏果实品质评价方法,对杏生产、加工销售过程具有重要的实践意义。新鲜果品的可溶性固形物含量是判断其成熟度、商品性的重要指标之一。本研究旨在建立基于近红外光谱的库车小白杏可溶性固形物含量的检测模型,以期为便携式光栅扫描光谱仪在 1 000 ~ 1 800 nm 波长范围内无损检测可溶性固形物含量提供理论依据。

1　材料与方法

1.1　试验材料

挑选个体均匀、无任何碰伤的 100 个成熟果实样品,将样品用流水洗净后用蒸馏水冲洗 2 次,然后放于试验室自然晾干。为排除温度及其他环境因素对试验结果的干扰,将样品置于室内,保持室温。

1.2　试验设备

SupNIR - 1500 聚光科技便携式光栅扫描光谱仪[聚光科技(杭州)股份有限公司],PAL - 1Cat. No.3810 型数字手持折射仪(日本 Atago 公司)。

1.3　试验方法

1.3.1　近红外光谱采集　采用近红外光谱无损检测仪采集近红外光谱,其波长范围为 1 000 ~ 1 800 nm,波长准确性 ± 0.2 nm,波长重复性 ≤ 0.05 nm,光谱分辨率 ≤ 12 nm,杂散光 < 0.15%。光谱扫描前检查仪器排除其他因素干扰进行调试,预热 30 min 后采集果实的光谱。采集样本光谱时,光源距果实表面的垂直距离约为 2 cm,在果实横径最大处进行光谱采集。同一果实样品完成 1 次光谱采集后将其旋转 180°采集另一果面光谱信息,共采集得到 200 个光谱信息。

1.3.2　可溶性固形物含量测定　用数字手持折射

收稿日期:2019 - 05 - 10
基金项目:国家自然科学基金(编号:31760560);塔里木大学重点科技攻关项目(编号:TDZKGG201706)。
作者简介:姜萌微(1997—),女,新疆昌吉人,主要从事园艺相关研究。E - mail:jmw970830@163.com。
通信作者:郭　玲,博士,副教授,主要从事特色园艺产品研究。E - mail:glzky@163.com。

仪测定 100 个样本的可溶性固形物含量。分别测定与采集光谱相应位置的果肉浆液可溶性固形物含量,求得每个样本平均值作为样品可溶性固形物含量的真实值。

1.3.3　光谱预处理与模型的建立　首先将采集的果实可溶性固形物含量的光谱信息与测定的可溶性固形物含量数据导入 TQAnalyst 8.0 光谱分析软件。再通过使用求导、标准正态变量变换、多元散射校正等方法对光谱进行预处理,利用设置模型参数,剔除异常光谱,选择预处理方法,从而确定主因子数。采用偏最小二乘法建立校正模型。定标模型精度的评价用定标建模的相关系数,校正集标准偏差和预测集标准偏差作为评价指标,以定标建模的交互验证均方根误差[10],确定最佳主因子数以优化模型参数[11],再通过预测集标准偏差来评价模型的预测精确度,得到最优模型。

2　结果与分析

2.1　库车小白杏可溶性固形物含量

由表 1 可知,样品的可溶性固形物含量变异范围是 13.50% ~20.30%,其包括了库车小白杏可溶性固形物含量的基本特征。通过该范围可实现基于近红外光谱可溶性固形物含量的数学模型建立及依据此范围内可溶性固形物含量信息进行模型校正。同时,样本间变异系数为 0.070 2,能够直接反映所选取样本的均一性,确保所建立模型的准确性和稳定性。

表 1　库车小白杏可溶性固形物含量

项目	样本数(个)	可溶性固形物含量(%)	均值(%)	标准偏差	变异系数
值	200	13.50 ~20.30	17.27	1.211	0.070 2

2.2　光谱数据预处理

将库车小白杏的测定样本原始光谱中 4 个异常光谱剔除,得到的光谱见图 1,横坐标为波长,波长范围是 1 000 ~1 800 nm,保证了该区域光谱波长的准确性。纵坐标为光谱吸光度,光谱吸光度范围为 0.6 ~1.9,吸光度范围越大,可用于检测样品的线性范围也越大。以波峰来看,在全波范围内该光谱曲线的特征光谱为 1 450 nm,此范围各个样本具有相同的吸收峰,但可溶性固形物含量不同吸收峰的吸光度存在差异。

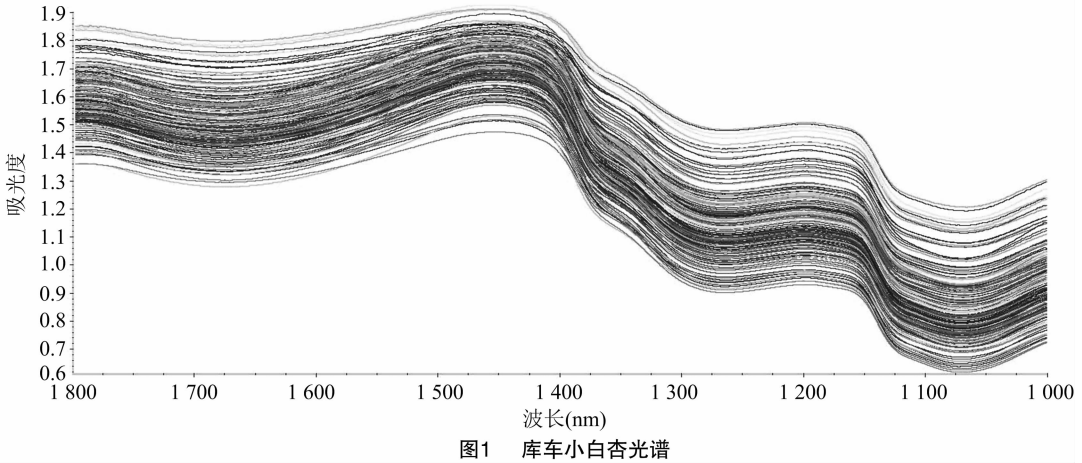

图1　库车小白杏光谱

由图 1 可知,在全波段 1 000 ~1 800 nm 范围内,光谱在 1 450 nm 附近虽有特征峰但不明显,未出现明显波峰,难以判断特征光谱区域。其中,因测定时环境掌控的差异、样品个体大小、表面光度及表皮厚度等噪声因素影响,致使原始光谱出现光谱重叠现象。

光谱数据预处理前建立的数学模型见图 2,该模型相关系数为 0.603 90、校正集标准偏差为 0.933、预测集标准偏差为 1.08、最佳因子数为 6。未处理光谱所建立的模型相关系数较低,其预测能力和稳定性较差。因此,须对原始光谱进行预处理后,才能实行进一步的光谱分析得到准确的数学模型。

采集样品光谱后,获取近红外漫反射光谱信息,利用 TQAnalyst 8.0 软件对光谱进行光谱预处理。光谱预处理方法包括标准化、平滑、求导、标准正态变量变换、多元散射校正等,这些方法可以单独使用也可以结合在一起使用[12]。比较各方法优缺点后总结得出,多元散射校正可用于消除理想中的线性散射影响,标准正态变量变换与多元散射校

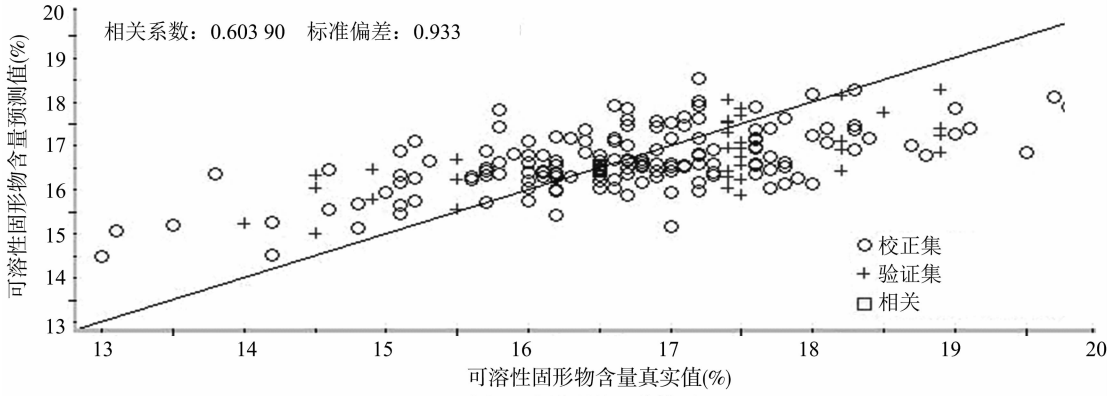

图2　光谱预处理前模型

正的目的基本相同;标准正态变量变换可用来校正样品因颗粒散射而引起的光谱的误差[13];导数处理可以消除基线漂移,提供比原光谱更高的分辨率和更清晰的光谱轮廓变化[14]。同时,对光谱分别使用多元散射校正、标准正态变量变换、求导及混合的预处理方法分别优化模型,则通过建模的相关系数越大,校正集标准偏差及预测集标准偏差越小,模型的预测能力和推广能力越好来评价预处理效果。

因此,本试验在通过使用多元散射校正、标准正态变量变换及求导的光谱预处理方法后,比较各方法光谱预处理所得模型相关系数,校正集标准偏差及预测集标准偏差。得到一阶导数混合标准正态变量变换法和一阶导数混合多元散射校正法 2 种预处理方法比单一的预处理方法具有更好的效果。

由表 2 可知,采用一阶导数混合多元散射校正法和一阶导数混合标准正态变量变换法的校正集标准偏差及预测集标准偏差分别为 0.215、0.456,均比预处理前校正集标准偏差(0.933)、预测集标准偏差(1.080)的模型预测的稳定性好,同时相关系数也较处理前更优。且预处理的 2 种方法都可降低噪声影响,校正基线漂移,提高检测精度以获得特征光谱。同时,采用一阶导数混合多元散射校正法和一阶导数混合标准正态变量变换法所建立的数学模型,二者校正集标准偏差、预测集标准偏差相同,但一阶导数混合标准正态变量变换法相关系数达到 0.964 45,其建立的模型更优。因此本研究采用一阶导数混合标准正态变量变换法作为最优光谱预处理方法。

表 2　预处理方法的比较

主成分数(个)	处理方法	相关系数	校正集标准偏差	预测集标准偏差
6	原始光谱	0.603 90	0.933	1.080
4	一阶导数 + 多元散射校正	0.964 39	0.215	0.456
4	一阶导数 + 标准正态变量变换	0.964 45	0.215	0.456

由图 3 可知,利用一阶导数混合标准正态变量变换法预处理后的光谱数据与光谱图 1 比较,波长范围为 1 000 ~ 1 800 nm 时,1 380、1 140 nm 处吸收峰具有明显的特征值。预处理后减弱光的基线不稳定等影响,比原光谱的分辨率更清晰。由此说明,该方法对降低光谱误差有一定效果。

2.3　确定最佳因子数

如何正确选取模型所需因子的数量(即主成分数)直接影响该模型的质量,选择的主成分数太小会出现拟合不足的结果;若选择的主成分数太大,导致过拟合,降低模型的质量[15]。因此,正确选取主因子数对保证定量模型预测稳定性非常重要。本研究对近红外光谱进行偏最小二乘法分析,选用交互验证方法,通过交互验证均方根误差和预测残差平方和的数值关系,确定最佳的因子数。由交互验证均方根误差和预测残差平方和越小,证明模型的稳定性越好,预测能力越强为标准。通过比较主因子数为 1 ~ 10 模型中的交互验证均方根误差值,确定其最小时对应的主因子数为最佳因子数,由图 4 可知,交互验证均方根误差随着主因子数的增加呈先上升后下降然后平缓上升的趋势,当主因子数为 4 时,交互验证均方根误差达到最低值。因此,确

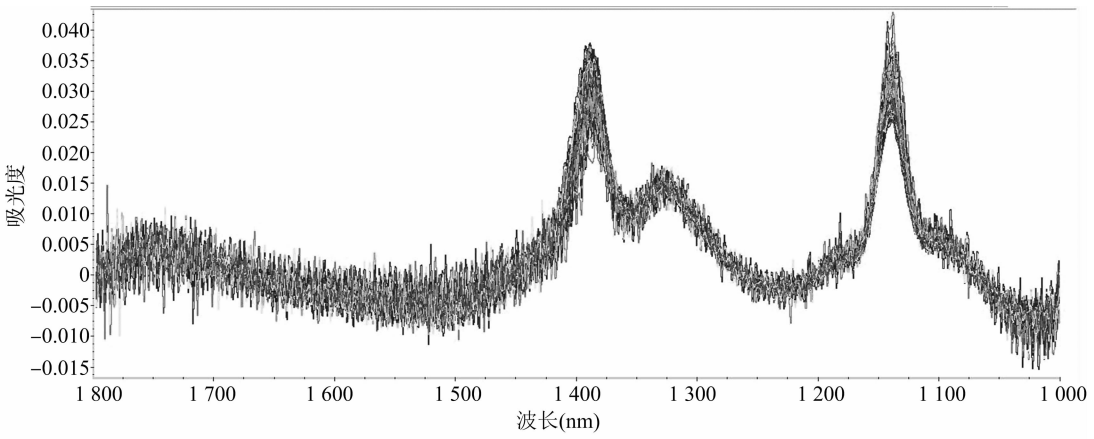

图3　经一阶导数混合标准正态变量变换法预处理后的光谱

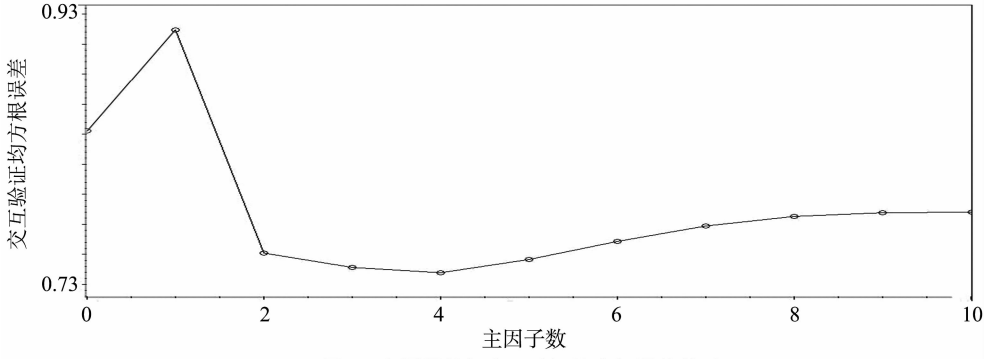

图4　主因子数与交互验证均方根误差关系

定 4 为最佳因子数。

2.4　偏最小二乘法建模

利用偏最小二乘法建立关于库车小白杏可溶性固形物含量的检测模型。偏最小二乘法采用校准试样组确定参数,在计算过程中没有经过任何矩阵求逆运算,所以计算误差小,准确度高,运算速度快[16]。运用偏最小二乘法进行数据处理与优化,通过误差最小化建立最适模型。随机选择 112 个数据作为校正集,20 个作为验证集,剔除异常数据与差异较大的数据共 68 个,在 1 000 ~ 1 800 nm 波长之间利用偏最小二乘法建立模型。如表 3 所示,20 个验证集的真实值、预测值及绝对误差。由表 3 可知,20 个验证集的可溶性固形物含量的真实值在 16.0% ~ 18.0% 之间,数据具有一定的代表性,而绝对误差范围为 ±0.85%,可以看出模型的稳定性和预测能力较好。

利用 TQAnalyst 8.0 软件处理的数据,再通过偏最小二乘法得到的最适模型见图 5,横坐标为库车小白杏可溶性固形物含量的真实值,纵坐标为预测值,这 2 个变量成线性关系,且 112 个校正集及 20 个验证集均匀地分布于直线的周围,则模型的相关

表 3　库车小白杏可溶性固形物含量的真实值、预测值及绝对误差

样品编号	真实值(%)	预测值(%)	绝对误差(%)
1	16.6	17.40	0.80
2	16.2	16.75	0.55
3	16.2	17.04	0.84
4	16.6	17.27	0.67
5	16.6	16.85	0.25
6	17.0	17.77	0.77
7	18.0	17.51	−0.49
8	17.1	17.09	−0.01
9	17.7	17.83	0.13
10	16.7	17.32	0.62
11	17.5	17.81	0.31
12	17.2	17.29	0.09
13	17.0	16.86	−0.14
14	18.1	17.97	−0.13
15	17.0	17.07	0.07
16	17.3	17.21	−0.09
17	17.7	17.22	−0.48
18	17.5	17.54	0.04
19	16.7	17.31	0.61
20	18.0	18.05	0.05

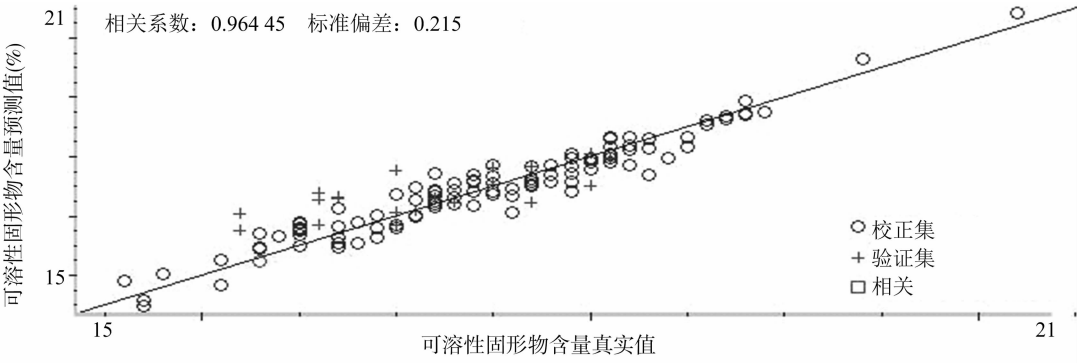

图5　偏最小二乘法建模

性较好。偏最小二乘法建模，模型相关系数为 0.964 45、校正集标准偏差为 0.215、预测集标准偏差为 0.456，得到最优模型。

3　结论与讨论

本研究利用近红外光谱分析技术检测了 100 个库车小白杏的可溶性固形物含量，波长范围为 1 000 ~ 1 800 nm，首先进行光谱预处理，再利用偏最小二乘法构建检测模型，得出以下结论：(1)采用多元散射校正、标准正态变换、一阶导数等多种方法对光谱进行预处理，可优化模型、提高预测效果。(2)建立近红外光谱可溶性固形物含量预测模型，选取最佳因子数为 4 时，模型相关系数达到最大值，为 0.964 45，校正集标准偏差和预测集标准偏差值较小，分别为 0.215、0.456，此模型稳定性好，预测效果优。(3)近红外光谱分析技术可以准确、迅速、无损地检测小白杏的可溶性固形物含量，为新疆栽培杏果品的内部品质检测提供科学依据和技术支持。

参考文献：

[1]玛依努尔·吐拉洪. 新疆阿克苏地方品种库车小白杏高效丰产栽培技术[J]. 果树实用技术与信息，2014(4)：6 - 8.

[2]徐广通，袁洪福，陆婉珍. 现代近红外光谱技术及应用进展[J]. 光谱学与光谱分析，2000，20(2)：134 - 142.

[3]唐长波，方立刚. 黄桃可溶性固形物的近红外漫反射光谱检测[J]. 江苏农业科学，2013，41(11)：331 - 333.

[4]张　猛，张昕明，高铭萱，等. 基于红外光谱的苹果糖度无损检测系统[J]. 电子技术与软件工程，2017(5)：126.

[5]孙　通，应义斌，刘魁武，等. 梨可溶性固形物含量的在线近红外光谱检测[J]. 光谱学与光谱分析，2008，28(11)：2536 - 2539.

[6]原　帅，张　娟，刘美娟，等. 樱桃含糖量的无损检测实验研究[J]. 分析试验室，2015，34(7)：760 - 764.

[7]刘燕德，罗　吉，陈兴苗. 可见/近红外光谱的南丰蜜橘可溶性固形物含量定量分析[J]. 红外与毫米波学报，2008，27(2)：119 - 122.

[8]陈　辰，鲁晓翔，张　鹏，等. 基于可见 - 近红外漫反射光谱技术的葡萄贮藏期间可溶性固形物定量预测[J]. 食品科学，2015，36(20)：109 - 114.

[9]樊丁宇，廖　康，杨　波，等. 新疆杏品种果实鲜食品质主要评价指标的选择[J]. 中国农学通报，2009，25(22)：207 - 211.

[10]王宝刚，李文生，蔡宋宋，等. 核果类水果干物质含量近红外无损检测研究[J]. 光谱实验室，2010，27(6)：2118 - 2123.

[11]姚　燕，张建强，蔡晋辉，等. 利用近红外光谱技术测定生物质的水分含量[J]. 可再生能源，2011，29(3)：46 - 49.

[12]彭云发，罗华平，王　丽，等. 3 种不同红枣水分检测方法的比较[J]. 江苏农业科学，2016，44(1)：308 - 310.

[13]尼　珍，胡昌勤，冯　芳. 近红外光谱分析中光谱预处理方法的作用及其发展[J]. 药物分析杂志，2008，28(5)：824 - 829.

[14]王姗姗，孙红男，陈　健，等. 基于近红外光谱技术检测蓝莓可溶性固形物含量[J]. 食品工业科技，2012，33(1)：340 - 341，369.

[15]张德权，陈宵娜，孙素琴，等. 羊肉嫩度傅里叶变换近红外光谱偏最小二乘法定量分析研究[J]. 光谱学与光谱分析，2008，28(11)：2550 - 2553.

[16]王镇浦，周国华，罗国安. 偏最小二乘法(PLS)及其在分析化学中的应用[J]. 分析化学，1989，17(7)：662 - 669.