基于近红外光谱分析技术测定库尔勒香梨硬度

盛晓慧[1]，李子文[1]，李宗朋[1]，张福艳[2]，朱婷婷[3]，王 健[1*]，尹建军[1]，宋全厚[1]

1. 中国食品发酵工业研究院有限公司，北京 100015
2. 河北衡水老白干酒业股份有限公司，河北 衡水 053000
3. 北京顺鑫农业股份有限公司牛栏山酒厂，北京 101300

摘 要 采用近红外(NIR)漫反射光谱法对新疆特色梨果库尔勒香梨的五种不同果(包括青头、粗皮、脱萼、宿萼、突顶果)的硬度进行测定。由于近红外光谱数据量大且原始光谱噪声明显、测定水果时散射严重等导致光谱建模时关键波长变量提取困难。以新疆库尔勒香梨为研究对象，为了有效地消除固体表面散射以及光程变化对NIR漫反射光谱的影响，首先采用标准正态变量变换(SNV)和多元散射校正(MSC)对库尔勒香梨的原始光谱进行预处理。为寻找适合近红外光谱检测库尔勒香梨硬度的最佳特征波长筛选方法，进行香梨近红外光谱的特征波长变量选择方法的比较与研究。研究比较了两种特征波长筛选方法对库尔勒香梨硬度偏最小二乘法(PLS)建模精度的影响。同时使用反向偏最小二乘(BiPLS)和遗传算法结合反向偏最小二乘(BiPLS-GA)在全光谱范围内筛选香梨硬度的特征波长变量，将校正均方根误差(RESMC)、预测均方根误差(RESMP)以及决定系数(R^2)作为模型的评价标准，并最终确定最优波段选择方法及最佳预测模型。基于选择的特征波长变量建立的PLS模型(BiPLS-GA)与全光谱变量建立的PLS模型进行比较发现BiPLS-GA模型仅仅使用原始变量中6.6%的信息就获得了比全变量PLS模型更好的库尔勒香梨硬度的预测结果，其中R^2，RMSEC和RMSEP分别为0.91，1.03和1.01。进一步与基于反向偏最小二乘算法(BiPLS)获得的特征变量建立的PLS模型比较发现，BiPLS-GA不仅可以去除原始光谱数据中的无信息变量，同时也能够对共线性的变量进行压缩去除，使得建模变量从301个减少到20个。极大地简化模型的同时有效地提高了模型的预测精准度和稳定性。因此该方法能够有效地用于近红外光谱数据变量的选择。证明了近红外光谱分析技术结合BiPLS-GA模型能够高效地选择出建模变量，去除与库尔勒香梨硬度无关的近红外光谱信息，显著地提高库尔勒香梨硬度定量模型的预测精度。这不仅为新疆地区特色梨果库尔勒香梨的快速、精确、无损优选分级提供一定的技术支持，同时也为基于近红外光谱分析技术预测水果内部品质的研究提供了参考。

关键词 近红外光谱技术；库尔勒香梨；反向间隔偏最小二乘；遗传算法；硬度

中图分类号：O657.3 **文献标识码**：A **DOI**：10.3964/j.issn.1000-0593(2019)09-2818-05

引 言

库尔勒香梨作为中国优质果品，因口感优良深受广大消费者的喜爱。在香梨的精准采收分选中，硬度能够反映香梨的内部品质，所以常将其作为一级指标对香梨的等外果和商品果进行区分[1]。目前一般采用农业标准(NY/T 2009—2011)进行穿孔测量香梨硬度[1]。该方法破坏了梨的内在品质，属于有损测量方法，采取的是抽样检测方式，不能对全部梨果样品逐一进行检验，而且检测速度慢，所以不适合香梨生产和分选中的大批量测量。

近些年，近红外光谱(near infrared spectroscopy，NIR)分析技术因具有快速、无损、精准的优点逐渐被用于瓜果内在品质的测定[5]。张德虎等[2]应用可见近红外光谱检测河套密瓜的糖度和硬度，结果显示其真实值和预测值间有较高的相关性。王晓明等[3]采用近红外漫反射光谱检测技术测定梨的硬度。通过偏最小二乘法(PLS)建立了梨果硬度的定量模型。王世芳等[4]采用近红外光谱检测技术测定不同冷藏期西红柿的质地，建立了西红柿质地的回归模型。以上研究证明了应用近红外光谱检测瓜果硬度是可行的，但是均没有对定

收稿日期：2018-08-17，**修订日期**：2018-12-08

基金项目：国家自然科学基金项目(31671937)，国家重点研发计划(2018YFD0400905)资助

作者简介：盛晓慧，1993年生，中国食品发酵工业研究院有限公司硕士研究生 e-mail：2012098869@qq.com

*通讯联系人 e-mail：81214112@qq.com

标模型进行深度优化，且现有的香梨硬度近红外检测精度尚难达到在商业上应用的要求。究其原因，是因为近红外技术是一种检测含氢基团在近红外谱区的合频和倍频信息的技术，而香梨的硬度是一个与其细胞结构和组织结构相关的物理指标，近红外检测香梨硬度实际上属于一种间接检测技术。由此可见，特征波长的选取以及分析其与硬度测量的关系就显得尤为重要 。

实验针对新疆库尔勒香梨的硬度进行无损分析，通过采用反向间隔偏最小二乘(BiPLS)及遗传算法结合反向间隔偏最小二乘(BiPLS-GA)从全光谱中选择特征波长，进一步探讨了光谱变量选择方法对库尔勒香梨硬度建模的影响，对比两种方法并确定最优的波段选择方法，从而达到提高模型稳定性、计算速度以及加强模型预测精度的目的，同时分析特征波长的物化意义，探讨了近红外光谱与硬度之间的关系。

1　实验部分

1.1　梨果样品的采集

从库尔勒市沙依香梨果园共采集库尔勒香梨样品 290 个，其中包括粗皮果、青头果、脱萼果、宿萼果、突顶果 5 种不同香梨果实，做好标记，储存于实验冷库 4 ℃环境中。

1.2　方法

实验前，将库尔勒香梨从冷库中取出放在实验室中 6 h，使得香梨温度与实验环境温度达到一致，光谱采集和硬度测量均在 25 ℃环境中进行。

实验使用 AOTF 分光式近红外光谱仪(中国科学院上海技术物理研究所研制)，仪器光源为卤钨灯，检测器为带制冷的 InGaAs 单点探测器，配有固体测量池。光谱范围为 11 000～4 000 cm^{-1}，进行单次扫描，利用配套软件 NIRAnalyzer 采集样品的近红外光谱信息，采用 UnscramblerX10.3 光谱分析软件(挪威 CAMO 公司)进行光谱预处理、偏最小二乘(PLS)计算，BiPLS，BiPLS-GA 等程序均在 MATLAB 环境下运行。每个香梨测量三次光谱，分别位于赤道等间距的三个位置(间隔为 120°)，取三点的平均光谱为该香梨样本的整果光谱。光谱采集结束，将对应光谱采集的三个部位削皮，采用质构仪(型号为 TMS-PRO，购于北京盈盛恒泰科技有限责任公司)与 6.0 mm 直径压力探头，测量果肉受压应力(N)。取 3 个标记部位的硬度均值作为整果硬度[6]。

1.3　建模方法

1.3.1　校正集法与验证集的划分

异常样本的存在会对模型的预测精确度产生影响，因此在建立可靠的近红外定量模型之前需要剔除掉异常样本。本实验剔除异常样本采用的是外在学生化残差—杠杆值图的方法，通过分析得到香梨的异常样本数为 6 个。在剔除了 6 个异常点的基础上，将 284 个香梨样品采用 Kennard-Stone(K-S)法，按照 2∶1 的比例来划分，得到校正集样本 190 个和验证集样本 94 个。校正集和验证集划分结果如表 1 所示。

1.3.2　模型建立方法与特征波长筛选方法

为了提高模型稳定性和精确度，分别采用 BiPLS 和 BiPLS-GA 算法在全光谱范围中进行特征波长筛选，并将筛选出的特征波长作为输入变量，采用偏最小二乘(PLS)建立香梨硬度模型。以所建模型的决定系数(R^2)、校正均方根误差(RMSEC)及预测均方根误差(RMSEP)作为模型的评价指标[9]，最终确定合适的波长筛选方法。其中，当 RMSEP 越趋近于 0，R^2 越趋近于 1，说明建立的模型效果越好[7]。同时拟采用标准正态变量变换(SNV)和多元散射校正(MSC)对香梨的原始光谱进行预处理。

表 1　库尔勒香梨校正集和验证集的划分结果

Table 1　Korla Fragrant Pear calibrationset and verification set division result*

Sample set	n^b	Hardness/N			
		Average value	Max	Min	SD[c]
Calibration	190	15.15	32.15	15.48	1.25
Validation	94	15.96	34.00	15.88	1.26

*：n^b=number of samples；SD[c]=standard deviation

2　结果与讨论

2.1　BiPLS-GA 模型建立

采用反向间隔偏最小二乘(BiPLS)将全光谱 301 个波长划分为一定数量的小区间，一次去除一个区间建立 PLS 模型，比较建模效果确定第一个应该除掉的区间，在余下的光谱区间中如此进行下去，直到剩余最后一个区间[8]。虽然经过 BiPLS 筛选波段已经去除掉全光谱中一部分无效信息，筛选出的波段内的波长变量之间仍存在共线性问题，因此，需要进一步提取光谱信息。

2.1.1　基于反向间隔偏最小二乘(BiPLS)的波段选择

由于在采用 BiPLS 筛选近红外光谱时，间隔大小能够影响波长范围的选取，间隔过小，会使得到的结果太过复杂，间隔过大，会丧失一部分有用信息。由于从理论上无法确定最佳的间隔数，所以本实验尝试采用 16～25 个间隔数，分别将全光谱分成 16～25 个子区间，研究间隔数目对于波长选择的影响。表 2 为不同间隔数的 BiPLS 波段筛选结果。

表 2　采用不同间隔数的 BiPLS 波段筛选效果

Table 2　Effect of BiPLSband filtering with different intervals

Number	PC[a]	Interval	RMSE/N	NV[b]
16	10	3，6～8，13，15	1.134 4	113
17	12	2～5，8，9，11，13，14，17	1.245 3	177
18	11	1，3，6，7，9～11，15，17	1.268 7	151
19	7	1，3，7～13，15～19	1.351 2	221
20	8	3，4，8，10，16，19	1.454 3	90
21	9	3，7～9，12，15，16，20	1.547 6	171
22	6	3，4，8，10，16，19	1.450 0	90
23	13	3，4，6，8，9，19	1.357 1	91
24	10	1，4，9，12～14，20	1.273 2	101
25	9	1，12，18，19，23，25	1.250 6	73

a：Principal component；b：Number of Variables

依据最小 RMSE 来筛选最优的子区间。表 2 显示，将全光谱 301 个波长分割为 16 个区间时，对应的 RMSE 值最小。

2.1.2 BiPLS 结合 GA 的波长筛选方法

遗传算法是一种很有效的波长选择方法，具有全局优化、易实现的特点[9]，在采用 BiPLS 从全光谱 301 个波长点中筛选出 6 个子区间，共 113 个波长点之后，再利用 GA 从这 6 个光谱区间中挑选特征波长。经 BiPLS-GA 计算之后，得到的波长变量数为 20 个。利用这 20 个波长建立的模型回归效果如表 3 所示。

表 3 不同光谱区域的建模效果

Table 3 Modeling effect of different spectral regions

Methods	Wavelengths	R^2	RMSEC /N	RMSEP /N	PC
Full spectrum	301	0.59	1.36	1.34	12
BiPLS	113	0.71	1.12	1.13	7
BiPLS-GA	20	0.91	1.03	1.01	10

如表 3 所示，通过特征波长的筛选，模型的回归效果得到明显的提高。将 BiPLS 和 GA 结合挑选波长，波长变量数极大地减少，RMSEC 和 RMSEP 进一步降低，决定系数(R^2)从最初的 0.71 增加到 0.91，说明不仅仅极大地简化了模型、提高分析速度，而且剔除掉相当一部分与香梨硬度无关的光谱信息，减少噪声的同时提高了模型的预测精度。

2.2 漫反射光谱和硬度的关系

如图 1 所示，经过 BiPLS 筛选出的特征波长主要集中在 1 090～1 180，1 375～1 655，2 040～2 130 和 2 225～2 310 nm 四个波段，其中 BiPLS-GA 筛选出来的特征波长集中在 1 100～1 180，1 500～1 655 和 2 225～2 310 nm 三个波段内，吸收峰主要出现在 1 190，1 450 和 1 940 nm 处，这些都是由于水分吸收造成的[10]。其中 1 190 nm 是 O—H 伸缩振动的合频吸收峰，1 450 nm 处为 O—H 伸缩振动的一级倍频，1 940 nm 处是 O—H 伸缩振动的二级倍频[11]。由于水分的吸收会干扰对其他成分的检测，而采用 BiPLS-GA 算法筛选出来的特征波长不包含这三个波长，从而避免了水分吸收产生的影响。

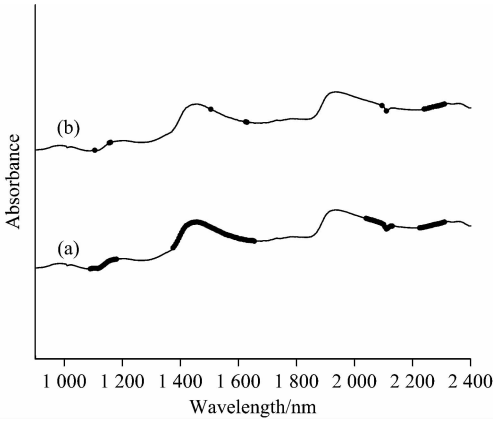

图 1 (a)BiPLS 和(b)BiPLS-GA 选择的波段

Fig. 1 Wavenumber variables selected by the (a) BiPLS and (b) BiPLS-GA for determination of hardness

有研究表明，香梨的硬度与果胶、纤维素等有机物质有关，尤其果胶是香梨成熟过程中影响硬度的关键物质，在梨果成熟过程中，原果胶的含量会不断下降，逐渐被分解转化为可溶性果胶，导致植物细胞组织间黏结性降低，使得梨果的硬度下降。成熟阶段的水果，其果胶呈现的是可溶性的状态，使细胞间结合力变得松弛，香梨质地变软。过熟香梨中的果胶发生去甲酯化变成无粘性的果胶酸，硬度加剧降低细胞进入衰老期。既然近红外检测的是有机物的吸收，因此可以通过测定果胶的吸收从而间接对硬度进行测定。

果胶作为一种富含甲氧基的化合物，其中含有大量 C—H 和 O—H 等特征官能团，在近红外区有吸收。相关文献中提到 2 250 nm 是果胶的特征吸收波长[12]，而图 1 中 BiPLS-GA 算法也挑选出 2 250 nm 作为库尔勒香梨的特征波长点。此外，BiPLS-GA 法筛选出的特征波段 1 100～1 200 nm 为 C—H 键伸缩振动的二级倍频吸收带，1 500～1 655 nm 为 C—H 键伸缩振动的一级倍频吸收带，这与 Rambo 等研究得到的果胶的特征波长是一致的[13]。因此，采用的向后间隔偏最小二乘和遗传算法得到的硬度的特征波长反映了果胶的吸收信息，也很好地解释了近红外光谱分析技术检测硬度的机理。

2.3 BiPLS-GA 模型验证

判断 BiPLS-GA 建立的定量模型的优劣，还要考察所建模型对未知样品的预测能力。据此，实验采用独立样品集对已建立的库尔勒香梨硬度的回归模型进行验证。在建立的全光谱 PLS 模型、BiPLS 模型和 BiPLS-GA 模型中导入没有参与建模的 94 个样品信息，将库尔勒香梨硬度预测值和实测值的相对偏差绘制得到以下的箱线图，如图 2 所示。相比于全光谱 PLS 模型和 BiPLS 模型的预测值与穿孔实验测得的硬度实测值之间的相对偏差，BiPLS-GA 模型的更小，平均数也更加集中。为了进一步验证 BiPLS-GA 所建模型的预测能力，对硬度的实测值和预测值在显著性水平 0.05 下进行 t 检验，该分析在 Matlab 环境下进行。结果显示，硬度的实测值和预测值之间的差异未达到显著水平($p>0.05$)，表明 BiPLS-GA 建立的模型在测量库尔勒香梨硬度方面的预测能力更强。

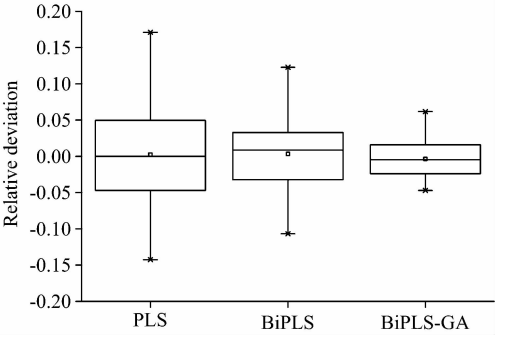

图 2 三种模型中硬度预测值和实测值的相对偏差

Fig. 2 Relative deviations of hardness predicted values and measured values in three models

表 4　库尔勒香梨分类分选指标

Table 4　Classification grade index of Korla fragrant pear

品质指标	果实类型				
	青头果	粗皮果	突顶果	宿萼果	脱萼果
硬度/N	22.00～35.00		15.00～19.00		

2.4　基于近红外光谱分析的类型果实分级品质指标评价

根据近红外模型的预测结果(见表 4)，5 类果实中，果肉硬度的差异极显著($p<0.01$)，对“青头果”和粗皮果的两类等外果有区分识别力，与突顶果、宿萼果和脱萼果 3 类商品果相比，青头果”和粗皮果具有果肉的硬度高的特征。可选择果肉硬度为识别指标，分选并剔除果肉果肉硬度大于 22 N 的两类等外果。

3　结　论

实验结果表明，BiPLS，BiPLS-GA 这两种波段筛选的方法均能在一定程度上减少建模变量，优化模型效果。其中，BiPLS-GA 筛选特征波长的 PLS 建模效果更好。经 BiPLS-GA 筛选后，建模所用的光谱变量显著减少，模型的 RMSEC 和 RMSEP 也明显降低，决定系数(R^2)提高到 0.91。筛选出的波段中包含了果胶中特征官能团的吸收带，既保留了与香梨硬度有关的特征波长，又剔除了大部分的无用信息，体现特征波长选择在提高模型精确度与稳定性方面的重要作用。同时，BiPLS-GA 建立的库尔勒香梨硬度的回归模型，具有精确、稳定的优点，能达到快速无损测定香梨硬度的精度要求。近红外的快速无损检测成为能对果实参差不齐的品质进行快速检测和区分的有效手段，同时本研究对于开发出更加精准的近红外无损检测水果的模型和设备具有借鉴作用。

References

[1] WEI Jie, MA Jian-jiang, CHEN Jiu-hong, et al(位　杰，马建江，陈久红，等). Food Science(食品科学), 2017, 38(19): 87.

[2] ZHANG De-hu, TIAN Hai-qing, LIU Chao, et al(张德虎，田海清，刘　超，等). Journal of Agricultural Mechanization Research(农机化研究), 2014, 36(2): 10.

[3] WANG Xiao-ming, ZHANG Hai-liang, LUO Wei, et al(王晓明，章海亮，罗　微，等). Chinese Journal of Agricultural Mechanization(中国农机化学报), 2015, 36(6): 120.

[4] WANG Shi-fang, SONG Hai-yan, ZHANG Zhi-yong, et al(王世芳，宋海燕，张志勇，等). Agricultural Products Processing(农产品加工), 2017, (3): 16.

[5] SONG Xue-jian, WANG Hong-jiang, ZHANG Dong-jie, et al(宋雪健，王洪江，张东杰，等). Nondestructive Testing(无损检测), 2017, 39(10): 71.

[6] LI Rui, FU Long-sheng(李　瑞，傅隆生). Journal of Agricultural Engineering(农业工程学报), 2017, 33(s1): 362.

[7] Seyed Ahmad Mireei, Seyed Saeid Mohtasebi, Morteza Sadeghi. International Journal of Food Properties, 2014, 17(6): 1199.

[8] Rungpichayapichet P, Mahayothee B, Nagle M, et al. Postharvest Biology & Technology, 2015, 111: 31.

[9] Nascimento P A M, Carvalho L C D, Júnior L C C, et al. Postharvest Biology & Technology, 2016, 111: 345.

[10] Huang X, Zou X, Zhao J, et al. Food Chemistry, 2014, 164(20): 536.

[11] Sun M, Zhang D, Li L, et al. Food Chemistry, 2017, 218: 413.

[12] Maniwara P, Nakano K, Boonyakiat D, et al. Journal of Food Engineering, 2014, 143(2): 33.

[13] Rambo M K D, Ferreira M M C. Journal of the Brazilian Chemical Society, 2015, 26(7): 612.

Determination of Korla Pear Hardness Based on Near-Infrared Spectroscopy

SHENG Xiao-hui[1], LI Zi-wen[1], LI Zong-peng[1], ZHANG Fu-yan[2], ZHU Ting-ting[3], WANG Jian[1*], YIN Jian-jun[1], SONG Quan-hou[1]

1. China National Research Institute of Food & Fermentation Industries Co., Ltd., Beijing 100015, China
2. Hebei Hengshui Laobai Dry Wine Co., Ltd., Hengshui 053000, China
3. Beijing Shunxin Agriculture Co., Ltd., Niulanshan Winery, Beijing 101300, China

Abstract Near-infrared diffuse reflectance spectroscopy was used to determine the hardness of five different fruits (including green head, rough skin, dislocated, scorpion, and apex) of Xinjiang pear fruit Korla pear. Due to the large amount of data in the near-infrared spectrum, the original spectral noise is obvious, and the scattering of fruits is serious, the key wavelength variables are difficult to extract during spectral modeling. Based on this, in order to effectively eliminate the influence of solid surface scattering and optical path variation on the NIR diffuse reflectance spectrum, it is proposed to use standard normal variable transformation (SNV) and multiple scattering correction (MSC). The original spectrum of Korla pear was pretreated. In order to find the best characteristic wavelength screening method suitable for the detection of Korla pear hardness by near-infrared spectroscopy, the comparison and research on the characteristic wavelength variable selection methods of Pear near infrared spectrum were carried out. The effects of two characteristic wavelength screening methods on the modeling accuracy of Korla pear hardness partial least squares (PLS) were compared. Simultaneously using the reverse partial least squares (BiPLS) and genetic algorithm combined with reverse partial least squares (BiPLS-GA) to screen the characteristic wavelength variable of the pear hardness in the whole spectral range, the corrected root mean square error (RESMC), The prediction root mean square error (RESMP) and the decision coefficient (R^2) were used as the evaluation criteria of the model, and the optimal band selection method and the optimal prediction model were finally determined. The PLS model based on the selected characteristic wavelength variable (BiPLS-GA) was compared with the PLS model established by the full spectral variable. It was found that the BiPLS-GA model obtains better information than the full-variable PLS model by using only 6.6% of the information in the original variable. The prediction results of Korla pear hardness, where R^2, RMSEC and RMSEP are 0.91, 1.03 and 1.01, respectively. Furthermore, compared with the PLS model established by the feature variables obtained by the reverse partial least squares algorithm (BiPLS), BiPLS-GA can not only remove the non-information variables in the original spectral data, but also compress and remove the collinear variables, reducing the number of modeling variables from 301 to 20. The model is greatly simplified while the prediction accuracy and stability of the model are effectively improved. Therefore, the method can be effectively used for the selection of near-infrared spectral data variables. It is proved that the near-infrared spectroscopy analysis technology combined with the BiPLS-GA model can efficiently select the modeling variables, remove the near-infrared spectral information unrelated to the hardness of Korla pear, and significantly improve the prediction accuracy of the Korla pear hardness quantitative model. This not only provides a certain technical support for the rapid, precise and non-destructive optimization of the characteristic pear fruit Korla pear in Xinjiang, but also provides a reference for the research of predicting the internal quality of fruit based on near-infrared spectroscopy.

Keywords Near-infrared spectroscopy; Korla fragrant pear; Backward interval partial least square; Genetic algorithm; Hardness

(Received Aug. 17, 2018; accepted Dec. 8, 2018)

* Corresponding author