

doi:10.6041/j.issn.1000-1298.2014.04.035

# 苹果可溶性固形物近红外在线光谱变量优选\*

欧阳爱国 谢小强 刘燕德

(华东交通大学机电工程学院, 南昌 330013)

**摘要:** 为简化近红外光谱模型,提高对苹果可溶性固形物含量的预测精度,将移动窗口偏最小二乘法(MWPLS)与遗传算法、连续投影算法相结合优选特征变量,建立偏最小二乘回归校正模型。其中移动窗口偏最小二乘法和遗传算法相结合优选的36个光谱变量建立的校正模型预测结果最好,可以有效筛选近红外光谱特征波长,模型预测相关系数为0.90,模型的预测均方根误差为0.70°Brix。

**关键词:** 苹果 可溶性固形物 近红外光谱 移动窗口偏最小二乘法 遗传算法 连续投影算法

**中图分类号:** O657.33 **文献标识码:** A **文章编号:** 1000-1298(2014)04-0220-06

## 引言

可溶性固形物含量(SSC)是一个组分比较复杂的水果内部品质指标<sup>[1]</sup>,也是评价水果内部品质的主要指标,很大程度上影响着水果的口感。近红外光谱分析技术因其具有不破坏样品、效率高、成本低、快速无污染等特点,已经被广泛地应用于水果的SSC检测<sup>[2-7]</sup>和建立水果的偏最小二乘回归校正模型<sup>[8-10]</sup>。但是由于某些区域的光谱信息与其SSC之间缺乏相关性,用全谱建模会造成一定程度的噪声信息,从而引起模型精度及稳定性下降,因此需要采用优化组合算法选择出光谱中最相关的光谱信息。

本文采用移动窗口偏最小二乘法结合遗传算法、连续投影算法优选光谱变量,剔除不相关的光谱变量,以选出的特征波长对应的光谱变量与可溶性固形物含量建立偏最小二乘回归校正模型,提高对苹果可溶性固形物含量的预测精度。

## 1 材料与方法

### 1.1 实验材料

所用样品由南昌某水果批发市场购回,选出没有损伤的150个苹果。对其依次进行编号及去除灰尘等处理,沿着样品赤道部位(间隔约120°)标记3点作为数据采集点。将样品置于室温20℃、湿度不变的环境下,待样品达到室温时,再进行光谱数据采集。

### 1.2 光谱采集装置

近红外在线检测装置如图1所示,主要包括光谱仪(QE65000型, Ocean optics INC, USA)、卤钨灯(12 V/100 W)、1 000 μm/2 m 光纤、PLC、A/D转换装置、输送线(速度由变频器和异步电动机控制)和计算机等。手动放置苹果样品在输送线上并传送到检测工位,由安装在链轮轴上的编码盘及接近开关检测水果位置信息,传递给光谱仪,光谱仪每触发1次,采集光谱1次;光谱数据存储由自行编写的软件完成。

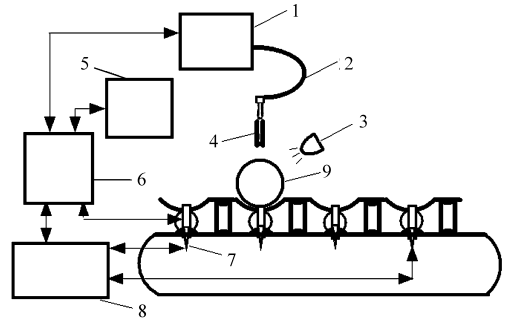


图1 近红外光谱在线检测装置示意图

Fig. 1 Sketch diagram of online near-infrared spectroscopy detection equipment

1. 光谱仪 2. 光纤 3. 卤钨灯 4. 探头 5. 计算机 6. A/D转换装置 7. 传感器 8. PLC 9. 样品

### 1.3 数据获取

在线检测装置对苹果样本采集可见近红外光谱时,参数设置为:传输速度4个/s,光谱仪积分时间

收稿日期:2013-05-05 修回日期:2013-06-26

\* 国家高技术研究发展计划(863计划)资助项目(2012AA101906)、科技部农业科技成果转化资金资助项目(2011GB2C500008)和江西省研究生创新基金资助项目(YC2012-S079)

作者简介:欧阳爱国,教授,主要从事光电检测研究,E-mail:ouyang1968711@163.com

通讯作者:刘燕德,教授,主要从事农产品智能无损检测技术研究,E-mail:jxliuyd@163.com

70 ms,平滑点数 15 点。进行光谱采集前,光源及光谱仪先预热 30 min。先在静态下采集暗电流和聚四氟乙烯球参比光谱,然后在动态下以 4 个/s 的传输速度采集苹果样本光谱。分别对苹果样本赤道上标记的 3 点采集漫反射光谱,图 2 为 1 号样品采集 3 次的吸光度光谱图,可以看出 3 次采集的特征波长是一致的,即横坐标(波长)不存在差异性,说明在线采集 3 次光谱差异主要来自于纵坐标(吸光度)的波动。为了更好的研究所有样品 3 次采集光谱的差异性,以 632.85 nm 处纵坐标的变异系数来衡量。图 3 为 150 个样品的变异系数图。可以看出,150 个样品 3 次采集光谱的变异系数在 20% 以内,说明动态在线采集光谱过程中有些小波动,为减少这种差异对后续建模的影响,取 3 次的平均光谱,作为后续的数据处理。

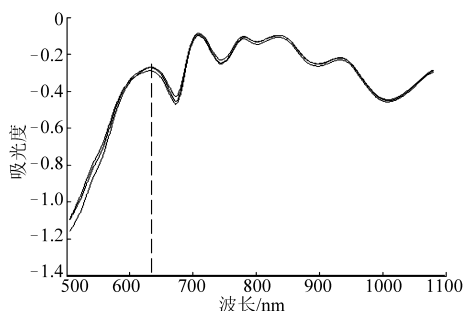


图 2 1 号样品采集 3 次的吸光度光谱  
Fig. 2 Absorbance spectra of sample No. 1 collected three times

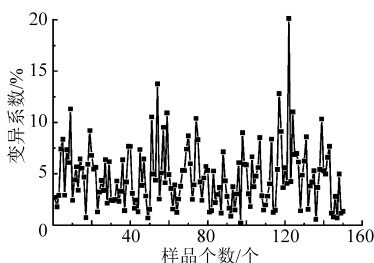


图 3 150 个样品的变异系数

Fig. 3 Coefficient of variation of 150 samples

在采集完样品的光谱数据后,测量样品的可溶性固形物含量(SSC)的真实值。样品 SSC 用 PR-101 $\alpha$  型数字式折射仪(Atago Co., Ltd., Tokyo, Japan)测量。测量前,用蒸馏水对数字手持折射仪校正零点。将样品 3 个部位标记的 3 个检测部位果肉切下榨汁,分别将果汁滴于数字折射仪的测试窗口上进行检测,取平均值作为样品 SSC 真实值。

#### 1.4 数据处理及模型评价

移动窗口偏最小二乘法<sup>[11-12]</sup>、连续投影算法<sup>[13-14]</sup>和遗传算法<sup>[15]</sup>由 Matlab R2010a 软件完成,校正模型建立及预测由 Unscrambler 8.0 软件完成。使用校正均方根误差(RMSEC)、校正相关系数  $R_c$ 、

预测样本均方根误差(RMSEP)和预测相关系数  $R_p$  评价所建模型的精确性, $R_p$  越大, RMSEP 越小,表明校正模型预测精度越好。

## 2 实验结果与分析

### 2.1 数据预处理

根据 K-S<sup>[16]</sup> 方法,校正集样品的浓度范围应该包含预测样品浓度范围,按照大概 3:1 的比例,从 150 个苹果样品中选取 110 个作为校正集,剩余 40 个为预测集,苹果样品 SSC 真实值统计特性如表 1 所示。

表 1 可溶性固形物真实值统计

Tab. 1 Measured values of soluble solids content

样品集	个数	范围	平均值 °Brix
总样品	150	8.93 ~ 16.27	12.01
校正集	110	8.93 ~ 16.27	12.02
预测集	40	10.20 ~ 13.83	11.98

图 4 为苹果的可见近红外吸光度光谱图,在后续的数据处理与分析中均采用原始吸光度光谱。

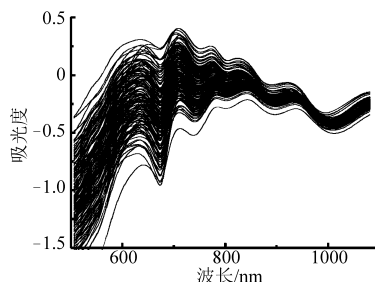


图 4 苹果可溶性固形物的可见近红外吸光度光谱  
Fig. 4 Visible-NIR absorbance spectra of apple soluble solids content

### 2.2 可见近红外光谱波段选择

苹果可见近红外光谱在表征可溶性固形物信息的同时会受到非可溶性固形物信息和环境噪声的影响,在多个波长处有吸收,谱峰较宽,致使多个谱峰重叠。本文将校正集苹果样品的可溶性固形物组分利用移动窗口偏最小二乘法(MWPLS),在一定的交叉验证均方误差(RMSECV)下,进行可见近红外光谱信息区间定位<sup>[17-18]</sup>。移动窗口偏最小二乘法参数的设置参照文献[17],设定窗口宽度为 31,所提取的 PLS 成分数上限为 15,结果如图 5 所示。图 5 中倒立峰形曲线由各变量点 RMSECV 随窗口位置变化形成,其中虚线表示全光谱含有 8 个主成分时,其交叉验证均方误差(RMSECV)为 0.69 比较合适。从图 5 可看出变量数为 99 ~ 239、526 ~ 745 时 RMSECV 值较小,对应的波长范围为 581.4 ~ 688.52、903.4 ~ 1 063.01 nm。将以上 361 个特征光

谱变量合并成新数据集作为后续的最优波长组合变量选择。

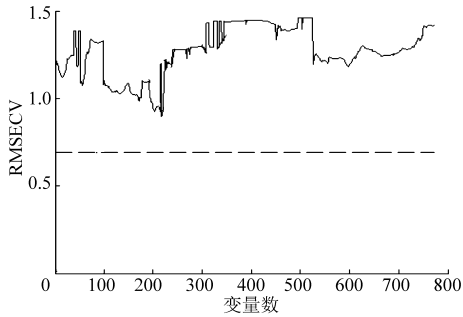


图5 移动窗口偏最小二乘法波段筛选结果

Fig. 5 Results of wavelength bands selection by moving windows partial least squares regression algorithm

### 2.3 可见近红外光谱最优波长组合变量选择

将移动窗口偏最小二乘法初步定位的 361 个波长点分别采用连续投影算法 (SPA) 和遗传算法 (GA) 筛选最优波长组合。连续投影算法是一种通过计算光谱矩阵中某一波长对其他波长的投影,以投影信息量来选择特征波长。SPA 能够有效地消除波长间的共线性影响,使模型的复杂度降低,提高建模效率。SPA 的运行参数与文献 [19-20] 相同,设定最小筛选变量数目为 1,最大筛选变量数目为 20,根据校正集的内部交叉验证均方根误差筛选出最优组合波长,其运行结果如图 6 所示。曲线表示苹果样本的可见近红外吸光度光谱图,方格表示选中的波长,共筛选出 5 个波长,具体结果见表 2。利用筛选出的 5 个波长,建立偏最小二乘法回归模型并进行预测,预测均方根误差为  $0.80^{\circ}\text{Brix}$ 。

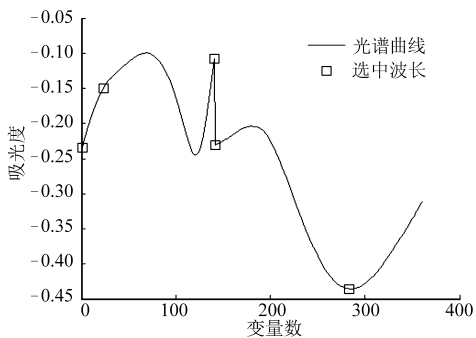


图6 连续投影算法筛选最优组合波长点结果

Fig. 6 Results of optimal wavelengths selection by SPA

遗传算法<sup>[21]</sup>是一种模拟生物进化优胜劣汰的随机优化算法,主要以适应度函数为依据,通过对群体内个体施加选择、变异、交叉等遗传操作,来实现群体中个体结构重组的迭代优化<sup>[22-23]</sup>。遗传算法参数的设置与文献[24]相同:初始群体数目为 30,交叉概率为 0.05,变异概率为 0.01,迭代次数为 100,以 RMSECV 值作为适应度函数,运行 100 次后终止迭代。运行结束后,以交叉验证均方根误差最

小原则,确定出最佳建模变量数,筛选出在迭代过程中出现频率高的波长作为特征变量。图 7 为所有变量被入选的频次图,只要变量的被选频次大于 5 就被选为最终模型的特征变量。图 8 为 RMSECV 值随变量数选取不同而变化的趋势图。从图中可以看出, RMSECV 值最低为 0.68,其所对应的变量数为 36。所以通过遗传算法共筛选出 36 个特征变量,具体结果如表 2 所示。

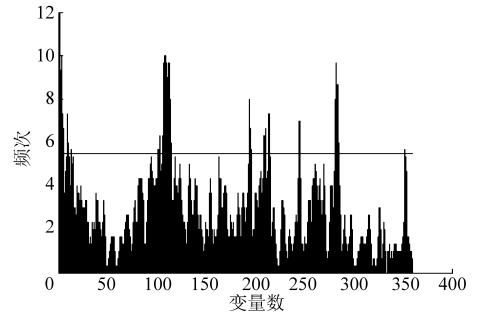


图7 由 GA 迭代 100 次后被选用变量的频次图

Fig. 7 Cumulative frequency of variable selection after 100 times runs by GA

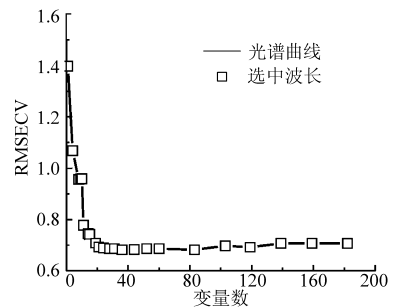


图8 不同变量数与交叉验证均方根误差的关系曲线

Fig. 8 Relation curves between RMSECV and variable numbers

### 2.4 模型建立及预测

利用全谱 771 个光谱变量、移动窗口偏最小二乘法 (MWPLS) 优选的 361 个光谱变量、移动窗口偏最小二乘法和连续投影算法组合 (MW-SPA-PLS) 优选的 5 个光谱变量、移动窗口偏最小二乘法和遗传算法组合 (MW-GA-PLS) 优选的 36 个光谱变量建立偏最小二乘回归校正模型,并且用未参与建模的 40 个样本对模型的性能进行评价,结果如表 2 所示。

在建立偏最小二乘回归校正模型过程中,主成分因子数 (PCs) 的选取尤为重要,最佳主成分因子数的确定直接关系到校正模型的预测效果。如果选取的主成分因子过少,将会丢失光谱较多的有用信息,使模型预测精度下降,易出现欠拟合现象。选取的主成分因子过多,则会将过多的干扰信息包括进来,使模型的预测误差增大,易出现过拟合现象<sup>[25]</sup>。图 9 为不同 PLS 模型的 RMSEC 与主成分因子数关

表 2 偏最小二乘回归校正模型和预测结果

Tab.2 Statistics of calibration and prediction results by different bands or wavelengths

方法	PCs	波长组合/nm	变量数	校正集		预测集	
				$R_c$	RMSEC	$R_p$	RMSEP
PLS	8	505.53 ~ 1 081.71	771	0.91	0.59	0.89	0.74
MWPLS	11	581.4 ~ 688.52, 903.4 ~ 1 063.01	361	0.93	0.54	0.88	0.74
MW-SPA-PLS	4	581.4, 599.1, 688.52, 903.4, 1 006.6	5	0.89	0.67	0.85	0.80
MW-GA-PLS	11	581.4 ~ 585.25, 588.33, 589.1, 591.41, 658.83, 659.59, 660.36, 662.64 ~ 668.74, 942.37 ~ 943.84, 952.63 ~ 954.1, 956.29, 957.03, 978.94, 979.67, 1 005.87 ~ 1 008.05, 1 056.53	36	0.93	0.54	0.90	0.70

注:PLS 为偏最小二乘法。

系,以 MW-GA-PLS 为例,随着主成分因子数的增加, RMSEC 逐渐减小,当主成分因子数为 11 时, RMSEC 出现最小值为 0.54,随着主成分因子数的继续增加,出现过拟合现象, RMSEC 逐渐增加。

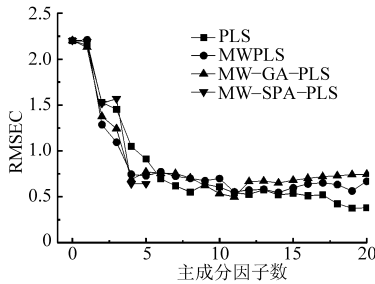


图 9 RMSEC 与模型主成分因子数关系  
Fig.9 Relationships between RMSEC and principal components

由图 9 可以看出, PLS、MWPLS、MW-SPA-PLS 和 MW-GA-PLS 校正模型的最佳主成分因子数分别为 8、11、4 和 11。通过对 PLS、MWPLS、MW-SPA-PLS 和 MW-GA-PLS 校正模型的预测结果比较, MW-GA-PLS 校正模型的预测效果最好。用未参与建模的 40 个样本对 MW-GA-PLS 校正模型的预测效果进行验证,结果如图 10 所示,预测模型最佳主成分因子数为 11,模型预测的相关系数  $R_p$  为 0.90,预测均方根误差 RMSEP 为 0.70° Brix。同时计算出这 40 个样品的预测值与真实值的相对误差,如图 11 所示,从图中可以看出,大部分样品的相对误差都低于 15%;从实验结果看, MW-GA-PLS 校正模型的预测性能能够满足实际应用。

### 3 结束语

探讨了移动窗口偏最小二乘法结合遗传算法、

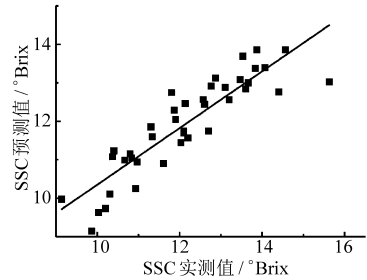


图 10 移动窗口偏最小二乘法 and 遗传算法组合筛选 36 个变量建立的偏最小二乘预测模型

Fig.10 MW-GA-PLS model with 36 variables

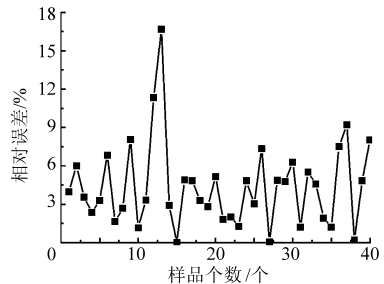


图 11 SSC 预测值与真实值误差

Fig.11 SSC error of predicted values and actual value

连续投影算法优选光谱变量建立偏最小二乘法回归模型,比较了模型的预测效果。其中移动窗口偏最小二乘法 and 遗传算法相结合优选的 36 个光谱变量建立的校正模型预测结果最好,模型预测相关系数为 0.90,模型的预测均方根误差为 0.70° Brix。研究结果表明,移动窗口偏最小二乘法与遗传算法相结合对苹果可溶性固形物近红外光谱进行变量选择,有效地减少了建模变量数,使模型更简单,同时也提高了模型的预测能力。

### 参 考 文 献

1 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.  
Sun Tong, Ying Yibin, Liu Kuiwu, et al. Online detection of soluble solids content of pear by near infrared transmission spectrum [J]. Spectroscopy and Spectral Analysis, 2008, 28(11): 2536 - 2539. (in Chinese)

- 2 孙通,许文丽,林金龙,等. 可见/近红外漫透射光谱结合 CARS 变量优选预测脐橙可溶性固形物[J]. 光谱学与光谱分析, 2012, 32(12):3229-3233.  
Sun Tong, Xu Wenli, Lin Jinlong, et al. Determination of soluble solids content in navel oranges by Vis/NIR diffuse transmission spectra combined with CARS method[J]. Spectroscopy and Spectral Analysis, 2012, 32(12):3229-3233. (in Chinese)
- 3 章海亮,孙旭东,刘燕德,等. 近红外光谱检测苹果可溶性固形物[J]. 农业工程学报, 2009, 25(增刊2):340-344.  
Zhang Hailiang, Sun Xudong, Liu Yande, et al. Measurement of soluble solid content in apples using near infrared spectroscopy [J]. Transactions of the CSAE, 2009, 25(Supp. 2):340-344. (in Chinese)
- 4 Camps C, Guillermin P, Mauget J C, et al. Discrimination of storage duration of apples stored in a cooled room and shelf-life by visible-near infrared spectroscopy[J]. Journal of Near Infrared Spectroscopy, 2007, 15(3):169-177.
- 5 朱伟兴,江辉,陈全胜,等. 梨可溶性固形物含量 NIR 与变量筛选无损检测[J]. 农业机械学报, 2010, 41(10):129-133.  
Zhu Weixing, Jiang Hui, Chen Quansheng, et al. Measurement of soluble solids content in pear by FT-NIR spectroscopy and variable selection[J]. Transactions of the Chinese Society for Agricultural Machinery, 2010, 41(10):129-133. (in Chinese)
- 6 徐惠荣,陈晓伟,应义斌. 基于多元校正法的香梨糖度可见/近红外光谱检测[J]. 农业机械学报, 2010, 41(12):126-130.  
Xu Huirong, Chen Xiaowei, Ying Yibin. Multivariate approach to determinate sugar content of fragrant pears with temperature variation by visible/NIR spectroscopy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2010, 41(12):126-130. (in Chinese)
- 7 陆辉山,傅霞萍,谢丽娟,等. 可见/近红外光估测完整柑橘水果可溶性固形物含量的研究[J]. 光谱学与光谱分析, 2007, 27(9):1727-1730.  
Lu Huishan, Fu Xiaping, Xie Lijuan, et al. Estimation of soluble solids content of intact citrus fruit by Vis/NIR spectroscopy[J]. Spectroscopy and Spectral Analysis, 2007, 27(9):1727-1730. (in Chinese)
- 8 刘燕德,罗吉,陈兴苗. 可见/近红外光谱的南丰蜜桔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.  
Liu Yande, Luo Ji, Chen Xingmiao. Analysis of soluble solid content in nanfeng mandarin fruit with visible near infrared spectroscopy[J]. Journal of Infrared and Millimeter Waves, 2008, 27(2):119-122. (in Chinese)
- 9 代芬,蔡博昆,洪添胜,等. 漫透射法无损检测荔枝可溶性固形物[J]. 农业工程学报, 2012, 28(15):287-292.  
Dai Fen, Cai Bokun, Hong Tiansheng, et al. Non-destructive measurement of soluble solid content in litchi by visible/near-infrared transmission spectroscopy[J]. Transactions of the CSAE, 2012, 28(15):287-292. (in Chinese)
- 10 田海清,应义斌,徐惠荣,等. 西瓜可溶性固形物含量近红外透射检测技术[J]. 农业机械学报, 2007, 38(5):111-113.  
Tian Haiqing, Ying Yibin, Xu Huirong, et al. Near-infrared transmittance measuring technique for soluble solids content of watermelon[J]. Transactions of the Chinese Society for Agricultural Machinery, 2007, 38(5):111-113. (in Chinese)
- 11 Zou Xiaobo, Zhao Jiwen, Malcolm J W P, et al. Variables selection methods in near-infrared spectroscopy [J]. Analytica Chimica Acta, 2010, 667(1/2):14-32.
- 12 李振庆,黄梅珍,倪一,等. 改进偏最小二乘法在近红外牛奶成分测量中的应用[J]. 光学技术, 2009, 36(1):70-73.  
Li Zhenqing, Huang Meizhen, Ni Yi, et al. Using improved PLS methods for milk components determination by near infrared spectra [J]. Optical Technique, 2009, 36(1):70-73. (in Chinese)
- 13 洪涯,洪添胜,代芬,等. 连续投影算法在砂糖橘总酸无损检测中的应用[J]. 农业工程学报, 2010, 26(增刊2):380-384.  
Hong Ya, Hong Tiansheng, Dai Fen, et al. Successive projections algorithm for variable selection in nondestructive measurement of citrus total acidity[J]. Transactions of the CSAE, 2010, 26(Supp.2):380-384. (in Chinese)
- 14 成忠,张立庆,刘赫扬,等. 连续投影算法及其在小麦近红外光谱波长选择中的应用[J]. 光谱学与光谱分析, 2010, 30(4):949-952.  
Cheng Zhong, Zhang Liqing, Liu Heyang, et al. Successive projections algorithm and its application to selecting the wheat near-infrared spectral variables[J]. Spectroscopy and Spectral Analysis, 2010, 30(4):949-952. (in Chinese)
- 15 祝诗平,王一鸣,张小超,等. 基于遗传算法的近红外光谱谱区选择方法[J]. 农业机械学报, 2004, 35(5):152-156.  
Zhu Shiping, Wang Yiming, Zhang Xiaochao, et al. Region selection method of near infrared spectrum based on genetic algorithm [J]. Transactions of the Chinese Society for Agricultural Machinery, 2004, 35(5):152-156. (in Chinese)
- 16 李华,王菊香,邢志娜,等. 改进的 K/S 算法对近红外光谱模型传递影响的研究[J]. 光谱学与光谱分析, 2011, 31(2):362-365.  
Li Hua, Wang Juxiang, Xing Zhina, et al. Influence of improved Kennard/Stone algorithm on the calibration transfer in near-infrared spectroscopy [J]. Spectroscopy and Spectral Analysis, 2011, 31(2):362-365. (in Chinese)
- 17 耿响,陈斌,叶静,等. 茶叶咖啡碱近红外光谱模型简化方法[J]. 农业工程学报, 2009, 25(10):345-349.  
Geng Xiang, Chen Bin, Ye Jing, et al. Simplification of model for tea caffeine detection by near infrared spectroscopy[J]. Transactions of the CSAE, 2009, 25(10):345-349. (in Chinese)
- 18 曹楠宁,王加华,李鹏飞,等. 基于 GA 和 SCMWPLS 算法的 NIR 光谱信息变量提取研究[J]. 光谱学与光谱分析, 2010, 30(4):915-919.  
Cao Nanning, Wang Jiahua, Li Pengfei, et al. Study on variable selection of NIR spectral information based on GA and SCMWPLS[J]. Spectroscopy and Spectral Analysis, 2010, 30(4):915-919. (in Chinese)

- 19 吴迪, 吴洪喜, 蔡景波, 等. 基于无信息变量消除法和连续投影算法的可见-近红外光谱技术白虾种分类方法研究[J]. 红外与毫米波学报, 2009, 28(6): 423 - 427.  
Wu Di, Wu Hongxi, Cai Jingbo, et al. Classifying the species of exopalaemon by using visible and near infrared spectra with uninformaton variable elimination and successive projections algorithm [J]. Journal of Infrared and Millimeter Waves, 2009, 28(6): 423 - 427. (in Chinese)
- 20 钱海波, 孙来军, 王乐凯, 等. 基于连续投影算法的小麦湿面筋近红外校正模型优化[J]. 中国农学通报, 2011, 27(18): 51 - 56.  
Qian Haibo, Sun Laijun, Wang Lekai, et al. Near infrared spectroscopy calibration model optimizing of wet gluten based on successive projections algorithm [J]. Chinese Agricultural Science Bulletin, 2011, 27(18): 51 - 56. (in Chinese)
- 21 严衍禄, 赵龙莲, 韩东海. 近红外光谱分析基础与应用[M]. 北京: 中国轻工业出版社, 2005.
- 22 邹小波, 赵杰文. 用遗传算法快速提取近红外光谱特征区域和特征波长[J]. 光学学报, 2007, 27(7): 1316 - 1321.  
Zou Xiaobo, Zhao Jiewen. Methods of characteristic wavelength region and wavelength selection based on genetic algorithm [J]. Acta Optica Sinica, 2007, 27(7): 1316 - 1321. (in Chinese)
- 23 褚小立, 袁洪福, 王艳斌, 等. 遗传算法用于偏最小二乘方法建模中的变量筛选[J]. 分析化学, 2001, 29(4): 437 - 442.  
Chu Xiaoli, Yuan Hongfu, Wang Yanbin, et al. Variable selection for partial least squares modeling by genetic algorithm [J]. Chinese Journal of Analytical Chemistry, 2001, 29(4): 437 - 442. (in Chinese)
- 24 Natalia Sorol, Eleuterio Arancibia, Santiago A Bortolato, et al. Visible/near infrared-partial least-squares analysis of Brix in sugar cane juice: a test field for variable selection methods[J]. Chemometrics and Intelligent Laboratory Systems, 2010, 102(2): 100 - 109.
- 25 刘燕德, 彭彦颖, 高荣杰, 等. 基于 LED 组合光源的水晶梨可溶性固形物和大小在线检测[J]. 农业工程学报, 2010, 26(11): 338 - 343.  
Liu Yande, Peng Yanying, Gao Rongjie, et al. Online detection of soluble solids content and size of crystal pear based on LEDs light source-detector [J]. Transactions of the CSAE, 2010, 26(11): 338 - 343. (in Chinese)

## Selection of NIR Variables for Online Detecting Soluble Solids Content of Apple

Ouyang Aiguo Xie Xiaoqiang Liu Yande

(School of Mechatronics and Electrical Engineering, East China Jiaotong University, Nanchang 330013, China)

**Abstract:** In order to improve the accuracy of online detecting soluble solids content (SSC) of apples by the method of near infrared spectroscopy, the combination of moving window partial least squares (MWPLS) and genetic algorithm (GA), successive projections algorithm (SPA) was used to select the characteristic variables, and then the partial least squares regression model was developed. The MW-GA model with the 36 selected characteristic variables obtained the best result with correlation coefficient of prediction ( $R_p$ ) of 0.90 and root mean square error of prediction (RMSEP) of 0.70° Brix, which indicated that the combination of MWPLS and GA could select the characteristic variables of near infrared spectroscopy effectively.

**Key words:** Apple Soluble solids Near infrared spectoroscapy Moving window partial least square Genetic algorithm Continuous projection algorithm