第37卷，第5期
2017年5月

光 谱 学 与 光 谱 分 析
Spectroscopy and Spectral Analysis

Vol 37，No. 5，pp1497-1501
May，2017

不知火杂柑可溶性固形物在线检测模型建立及优化

欧阳爱国，吴明明，王海阳，刘燕德*

华东交通大学机电学院，光机电技术及应用研究所，江西 南昌 330013

摘 要 应用近红外漫透射光谱检测技术对不知火杂柑的可溶性固形物(SSC)进行在线检测具有十分重要的意义。研究变量筛选方法对不知火杂柑可溶性固形物在线检测模型的影响，为实现其快速、准确的在线检测分级奠定基础。实验把形状不整、内藏瓤瓣的不知火杂柑作为研究对象，选取560～930 nm的光谱，采用偏最小二乘法(PLS)建立不知火杂柑可溶性固形物的在线检测模型，并讨论不同的光谱预处理方法(卷积平滑(S-G)、一阶微分(1st derivatives)等)，不同的变量筛选方法(移动窗口偏最小二乘法 MWPLS、遗传算法 GA、连续投影 SPA)对 PLS 所建预测模型性能的影响。经对比，多元散射校正(MSC)能有效地消除光散射的影响，遗传算法能大大地降低了建模的波长点数，缩短了建模时间，改善模型预测精度。其最优 PLS 模型的 R_P=0.956，RMSEP=0.380，R_C=0.967，RMSEC=0.340。实验表明在线检测不知火杂柑的可溶性固形物是完全可行的。

关键词 不知火杂柑；近红外漫透射光谱；在线检测；可溶性固形物；变量筛选

中图分类号：O657.33 **文献标识码**：A **DOI**：10.3964/j.issn.1000-0593(2017)05-1497-05

引 言

不知火杂柑是一种起源于日本的杂交柑橘品种。最早于2000年引入我国四川省，果肉清脆柔软多汁，风味极好，近年来深受的人们喜爱。水果内部的可溶性固形物(SSC)是决定其口感好坏的重要因素之一[1]。伴随着光谱检测技术的发展，近红外光谱检测技术已在苹果、梨、脐橙等水果的品质检测中得以广泛应用[2]。因此，基于近红外检测技术在线检测不知火杂柑的可溶性固形物是具有研究可行性的。

国内外研究人员应用近红外光谱技术对水果的内部品质进行了一系列的动态在线检测和分级[3-6]。刘燕德等[7]，应用3盏不同波长的LED作为光源，建立了PLS和偏最小二乘支持向量机(LS-SVM)模型，可溶性固形物和大小的相关系数分别达到了0.86和0.90。欧阳爱国等[8]为了提高苹果在线检测模型的精度，应用移动窗口偏最小二乘法和遗传算法、连续投影算法相结合，建立偏最小二乘回归模型，实验表明可以有效地减少变量数，提高模型预测能力。Sun等[9]对比了不同检测速度(0.3，0.5和0.7 m·s^{-1})对“翠冠”梨模型的影响，得到在检测速度为0.5 m·s^{-1}时，模型最佳。

综合以上学者的研究，都是以形状规则、表面光滑、内部品质均匀的水果为研究对象，建立了相关的检测模型。本实验则以形状不整、内藏瓤瓣的不知火杂柑(以下简称不知火)为对象，经过不同光谱预处理方法和变量筛选方法对不知火光谱数据进行处理，提高了模型的预测能力，建立了最优的PLS预测模型。

1 实验部分

1.1 材料

不知火样品共146个，从四川成都水果批发市场购买，将购买的样品用蒸馏水清理干净，晾干依次编号备用，在25 ℃的实验室存放一天。隔天采集光谱数据，把其中110个数据作为校正集，余下的46个样品作为预测集。

1.2 光谱测量

如图1所示，在线检测系统使用设备包括光谱仪(QE65000，Ocean optics INC·，USA)、12 V/100 W卤钨灯(10盏)、1 000 μm/2 m光纤、PLC、电脑、传送带等。采用

收稿日期：2014-08-20，**修订日期**：2015-05-10

基金项目：国家“863”高技术研究发展计划项目(2012AA101906)，赣鄱英才555工程领军人才培养计划项目(2011-64)，江西省光电检测工程技术研究中心资助项目(赣科发财字[2012]155号)，江西省研究生创新专项资金项目(YC2013-S166)，江西省优势科技创新团队建设计划项目(20153BCB24002)资助

作者简介：欧阳爱国，1968年生，华东交通大学机电学院光机电技术及应用研究所教授 e-mail：ouyang1968711@163.com

*通讯联系人 e-mail：jxliuyd@163.com

漫透射动态在线检测装置采集样品 3 个标记点的光谱，采集时，设置积分时间为 100 ms，平均次数为 1，传输速度为 5 个·s^{-1}，光照强度为 1 000 W。

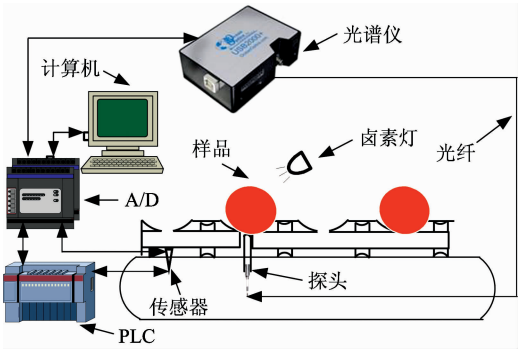

图 1 不知火杂柑在线检测原理图

Fig. 1 Schematic diagram of the online device detecting hybrid "Skiranui Tangerine" citrus

1.3 可溶性固形物含量的测定

不知火样品的化学值采用折射式数字糖度计(PR-101α，日本)进行测量。从样品的标记点切取 2～3 mm 果肉，挤出适量果汁测量其可溶性固形物值，并将不知火样品 3 个标记点的平均值作为整果的 SSC 值。

1.4 数据处理和模型评价

本实验由相关系数(R)、校正均方根误差(RMSEC)和预测样本均方根误差(RMSEP)对模型性能进行评价。一个性能良好的模型需要有较高相关系数 R 和较小的 RMSEC 和 RMSEP，并且 RMSEP 和 RMSEC 越接近越好[10]。计算方法如式(1)和式(2)

$$\mathrm{RMSEC}=\sqrt{\frac{\sum_{i=1}^{I_c}(\hat{y}_i-y_i)^2}{I_c-f-1}} \tag{1}$$

$$\mathrm{RMSEP}=\sqrt{\frac{\sum_{i=1}^{I_p}(\hat{y}_i-y_i)^2}{I_p-1}} \tag{2}$$

式中：y_i 为模型实际测量值；$\hat{y}_i$ 为模型预测值；I_c 为校正集中的样品个数；I_p 为预测集中的样品个数，f 为独立变量数。

2 结果与讨论

2.1 不知火近红外光谱分析

图 2 是采集的 146 个不知火样品原始吸光度光谱图。考虑到存在的系统误差和两端较大的噪声，选择 560～930 nm 范围内的漫透射光谱数据建模。由图 2 可见，样品光谱形状及波峰位置类似，并且由于一些含氢基团的伸缩振动[11]，光谱在 700 和 800 nm 附近都有比较明显的吸收峰。

2.2 不知火的可溶性固形物含量测定

不知火的 SSC 实际测定结果见表 1。为了使校正集建立的模型更好的适用于预测集，依据 K-S 方法，校正集 SSC 的范围应比预测集大。如表 1 所示，本实验校正集的 SSC 范围为 11.8～19.3°Brix，预测集的 SSC 范围为 12.3～17.8°Brix。

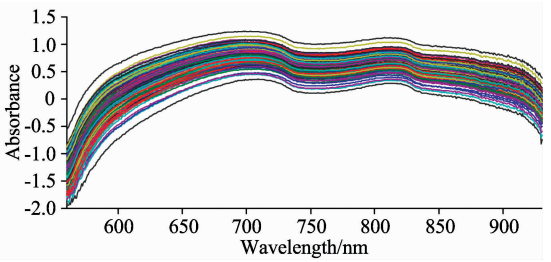

图 2 不知火杂柑原始光谱

Fig. 2 The original spectra of "Skiranui Tangerine" fruit

表 1 不知火杂柑 SSC 实际测定结果

Table 1 Statistical measurement results of SSC of "Skiranui Tangerine" fruit

Sample type	Min	Max	Average	SD	Coefficient variance/%
Calibration set(110)	11.8	19.3	14.99	1.344	8.97
Prediction set(36)	12.3	17.8	14.94	1.253	8.39

2.3 光谱预处理效果对比

实验对吸光度光谱值进行卷积平滑(S-G)、一阶微分(1st derivatives)、基线校正(Base line)、标准归一化(SNV)和多元散射校正(MSC)等预处理后分别建立 PLS 模型，建模结果见表 2。由表 2 可知，光谱经过卷积平滑、基线校正、标准归一化和多元散射校正预处理后，建立 PLS 模型得到的模型预测能力相对原始光谱直接建模均有所提高。其中，经 MSC 预处理后的建模结果最优，预测 $R_P=0.895$，RMSEP $=0.553$。因此，在进一步的进行波段筛选时采用 MSC 预处理后的光谱数据。

表 2 不同预处理方法建立的 PLS 模型结果

Table 2 The effect of different pretreatment methods on PLS Modeling results

Pretreatment method	Calibration set		Prediction set	
	R_C	RMSEC	R_P	RMSEP
Raw	0.879	0.638	0.856	0.643
S-G	0.895	0.597	0.862	0.626
1st derivatives	0.947	0.429	0.837	0.682
Base line	0.911	0.551	0.889	0.568
SNV	0.913	0.546	0.878	0.595
MSC	0.899	0.586	0.895	0.553

2.4 光谱波长筛选

本研究的目的是验证利用近红外漫透射技术在线检测不知火杂柑的可行性，并建立相应的最优预测模型。因在线检测不仅需要准确率高，同时检测效率要高，所以需要对大量的光谱数据进行筛选，从而获得较少的变量和较好的预测能力。

移动窗口偏最小二乘方法[12]是以人为设定的某一宽度的窗口在全光谱区连续移动，每移动一次，采用交互验证建立一个 PLS 模型，判断单独区间的 RMSECV 是否小于平均的 RMSECV，从而筛选出相应的波长区间。将全光谱分为

31 个子区间，将 PLS 成分数最大设为 15。如图 3 所示，RMSECV 的平均值为 0.661，当变量数为 190～195，222～235，360～365 时，区间 RMSECV 比平均值小。合并区间，共 26 个变量，其所对应的光谱区间分别为 704.46～708.25，728.68～738.50，832.23～835.96 nm。

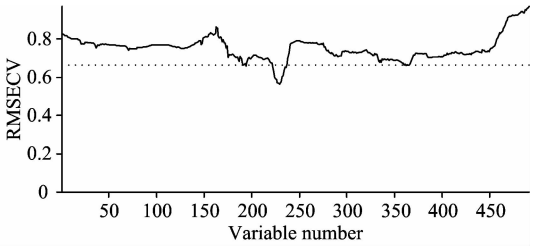

图 3　MWPLS 特征波段选择结果

Fig. 3　Results Characteristic bands selection by moving windows partial least square

遗传算法是仿照生物界择优选择和遗传原理提出来的一种算法，具有全局快速搜索，保留较优变量，剔除较差变量，建立更加简便，预测能力更强模型等优点。在对 491 个波长点的 GA 变量筛选中，根据文献[13]，主要参数设定为：群体的初值为 30，交叉概率为 0.5，变异概率为 0.01，迭代次数为 100，把 RMSECV 看作遗传算法的适应度函数。迭代 100 次结束后，从图 4 中看出当变量数为 94 时，所对应 RMSECV 值最小为 0.395。因此，这 94 个变量即为 GA 筛选出来的特征变量。由图 5 可知，在所有的 491 个变量中，频次大于 3 的变量即被选作模型的特征变量。

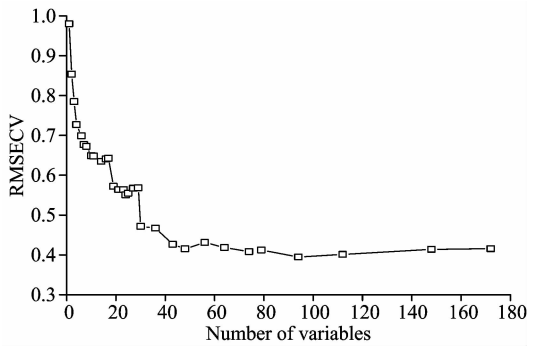

图 4　RMSECV 与变量数关系图

Fig. 4　Relation between RMSECV and number of variables

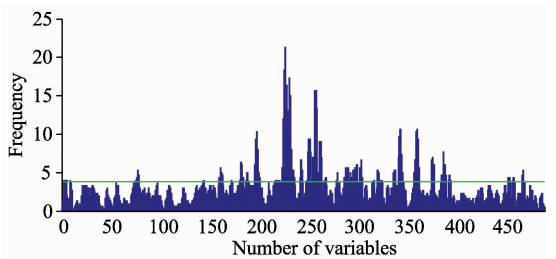

图 5　遗传算法迭代 100 次后变量的频次图

Fig. 5　Cumulative frequency of variable selection after 100 times runs by GA

连续投影算法是随机选取光谱矩阵中某一变量，然后分别计算对其他变量的投影，根据最小的 RMSECV 来决定变量的个数，一般选出的特征变量数在 20 以内。SPA 的运算步骤及变量数的选择参考文献[14]，最大特征变量设为 20，最小设为 1。图 6 为 SPA 所筛选出来的特征变量示意图，共有 8 个变量被筛选出来，其所对应的波长分别为 741.52，876.06，701.43，912.98，918.87，666.45，921.07 和 923.28 nm。

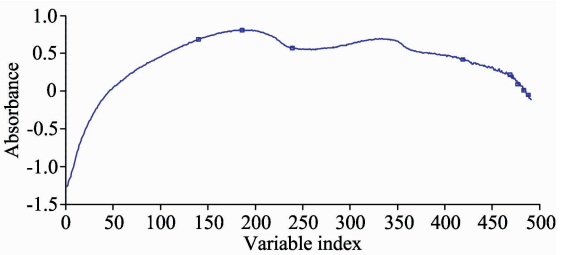

图 6　SPA 变量筛选

Fig. 6　Variables selected by SPA

表 3　不同建模方法建模结果对比

Table 3　Modeling results and comparison of different PLS

Method	Variable number	R_C	RMSEC	R_P	RMSEP
PLS	491	0.899	0.586	0.895	0.533
MWPLS	26	0.894	0.600	0.883	0.596
GA-PLS	94	0.967	0.340	0.956	0.380
SPA-PLS	8	0.874	0.650	0.820	0.711

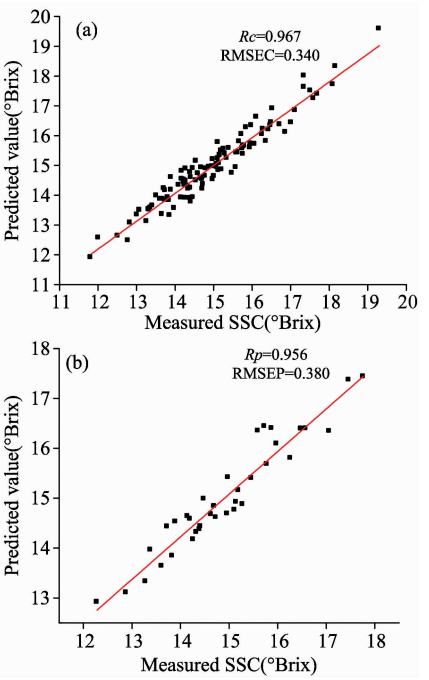

图 7　(a)GA-PLS 模型建模散点图；(b)GA-PLS 模型预测散点图

Fig. 7　Comparison of predicted values and measured values of "Skiranui Tangerine" fruit in calibration set and prediction set by GA-PLS model

2.5 模型对比

表 3 为变量筛选得到光谱数据与全光谱所建立的 PLS 模型比较，结果显示，MWPLS 和 SPA-PLS 所建立的模型相对全波段的 PLS 模型变量数少，预测能力相对有所减弱，其原因可能是部分有用的变量被剔除，以至于模型精度降低。而以 GA-PLS 所建立模型不仅使变量数从 491 个降低到了 94 个，有效的提高了运算速度，而且预测集的 RP 也从 0.895 提高到了 0.956，RMSEP 从 0.533 降低到了 0.380。因此从实验结果来看，GA 这种变量筛选方法可以使模型更优，更加简便，所建立的 GA-PLS 校正模型的预测能力能满足在线检测不知火的要求。图 7 为校正集(110)和预测集(36)不知火样品 SSC 应用 GA-PLS 所建模型的预测值与实测值的比较。

3 结 论

建立能准确在线检测不知火杂柑可溶性固形物的预测模型。比较了多种光谱预处理方法以及 MWPLS、GA、SPA 等波段筛选方法，并得出简化、改善模型中的最佳方法。预测结果显示，经 MSC 预处理后结合 GA 筛选出来的 94 个特征变量，建立的 PLS 模型预测结果最优，其 R_C 为 0.967，RMSEC 为 0.340，R_P 为 0.956，RMSEP 为 0.380。本研究表明，MSC 预处理方法能大大降低光散射的影响，利用 GA 可以明显减少建模的变量个数，简化模型，同时也提高了模型的预测能力。因此，应用近红外漫透射技术在线检测不知火杂柑的可溶性固形物是切合实际的。

References

[1] Sun Xudong, Zhang Hailiang, Liu Yande. International Journal of Agricultural and Biological Engineering, 2009, 2(1): 65.

[2] JIE Deng-fei, XIE Li-juan, RAO Xiu-qin, et al(介邓飞，谢丽娟，饶秀勤，等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2013, 29(12): 264.

[3] SUN Tong, LIN Jin-long, XU Wen-li, et al(孙 通，林金龙，许文丽，等). Journal of Jiangsu University · Natural Science Edition(江苏大学学报 · 自然科学版), 2013, 34(6): 663.

[4] MA Guang, SUN Tong(马 广，孙 通). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2013, 44(7): 170.

[5] LIU Yan-de, SHI Yu, CAI Li-jun, et al(刘燕德，施 宇，蔡丽君，等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2013, 44(9): 138.

[6] CAI Li-jun, LIU Yan-de, WAN Chang-lan(蔡丽君，刘燕德，万常斓). Journal of Northwest A&F University · Natural Science Edition(西北农林科技大学学报 · 自然科学版), 2012, 40(1): 215.

[7] LIU Yan-de, PENG Yan-ying, GAO Rong-jie, et al(刘燕德，彭彦颖，高荣杰，等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2010, 26(11): 338.

[8] OUYANG Ai-guo, XIE Xiao-qiang, LIU Yan-de(欧阳爱国，谢小强，刘燕德). Transactions of the Chinese Society for Agricultural Machinery (农业机械学报), 2014, 45(4): 220.

[9] Sun Tong, Lin Hongjian, Xu Huirong, et al. Postharvest Biology and Technology, 2009, 51(1): 86.

[10] HAN Dong-hai, CHANG Dong, SONG Shu-hui, et al(韩东海，常 冬，宋曙辉，等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2013, 44(7): 174.

[11] Bahareh J, Saeid M, Ezzedin M, et al. Computers and Electronics in Agriculture, 2012, 85: 64.

[12] DONG Xiao-ling, SU Xu-dong(董小玲，孙旭东). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2013, 33(12): 3216.

[13] ZHU Wei-xing, JIANG Hui, CHEN Quan-sheng, et al(朱伟兴，江 辉，陈全胜，等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2010, 41(10): 129.

[14] ZHANG Shu-juan, ZHANG Hai-hong, ZHAO Yan-ru, et al(张淑娟，张海红，赵艳茹，等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2012, 43(3): 108.

Establishment and Optimization of Online Model for Detecting Soluble Solids Content in Hybrid "Skiranui Tangerine" Citrus

OUYANG Ai-guo, WU Ming-ming, WANG Hai-yang, LIU Yan-de*
Institute of Optical and Electrical Machinery Technology and Application, School of Mechanical Engineering, East China Jiaotong University, Nanchang 330013, China

Abstract It is of great importance to detect soluble solids content (SSC) of online testing in hybrid "Skiranui Tangerine" citrus by using near-infrared diffuse transmittance spectra. In order to lay a good foundation for accurate and rapid online classification, this study focuses on the influence of variable methods on soluble solids content in hybrid "Skiranui Tangerine" citrus. We selected the random shape hybrid "Skiranui Tangerine" citrus with segments inside as the research object. In spectral range of 560 ~930 nm, the calibration models were developed based on partial least squares (PLS) in this experiment. Firstly, different pretreatment methods such as Savitzky-Golay, the first derivative and so on were compared with PLS Modeling results. Then moving window partial least squares (MWPLS), genetic algorithm (GA) and successive projections algorithm (SPA) were employed to improve the predictive models. After comparing the results, light scattering can be effectively eliminated by the multiplicative scatter correction (MSC). Moreover, fewer variables and model optimization were carried out with GA. The best calibration model obtained with GA-PLS method had the correlation coefficient of prediction (R_P) of 0.956, the root mean square errors of prediction (RMSEP) of 0.380, the correlation coefficient of calibration (R_C) of 0.967 and the root mean square errors of calibration (RMSEC) of 0.340. The experiment showed that online detection of SSC of "Skiranui Tangerine" is completely feasible.

Keywords Hybrid "Skiranui Tangerine" citrus; Near-infrared diffuse transmittance spectra; Online detection; Soluble solids content; Variable selection

(Received Aug. 20, 2014; accepted May 10, 2015)

* Corresponding author