

雷竹笋硬度的近红外光谱检测模型优化

周竹^{1,2,3}, 郑剑⁴, 王允祥⁴, 曾松伟^{1,2,3}

(1. 浙江农林大学 信息工程学院, 浙江 临安 311300; 2. 浙江农林大学 浙江省林业智能监测与信息技术研究重点实验室, 浙江 临安 311300; 3. 浙江农林大学 智慧农林业研究中心, 浙江 临安 311300; 4. 浙江农林大学 农业与食品科学学院, 浙江 临安 311300)

摘要: 为了提高应用近红外光谱技术无损检测雷竹 *Phyllostachys violascens* 竹笋硬度的精度, 研究了雷竹笋硬度光谱检测模型的优化方法。首先对雷竹笋原始光谱进行正态变量变换(SNV), 然后采用后向间隔偏小二乘法(biPLS)去除部分与竹笋硬度无关的变量, 随后进一步采用竞争性自适应权重法(CARS)剔除无关变量, 最后采用连续投影算法(SPA)将光谱变量个数从 1 557 个减少为 25 个。最终, biPLS-CARS-SPA 模型的交叉验证相关系数(r_{cv}), 预测相关系数(r_p), 交叉验证均方误差(RMSECV)以及预测均方误差(RMSEP)分别为 0.984, 0.926, 0.300 N·cm⁻² 和 0.625 N·cm⁻², 优于其他几种常见的变量选择方法及其组合。研究表明, biPLS-CARS-SPA 方法所选特征变量避开了水分强吸收峰的影响, 具有实际的物理表征意义, 为竹笋硬度在线快速检测、筛选和指导切削设备的研发提供了重要的理论依据。图 6 表 3 参 20

关键词: 经济林学; 近红外光谱; 后向间隔偏小二乘法; 竞争性自适应权重法; 连续投影算法; 硬度; 雷竹笋
中图分类号: S123; O657.33 **文献标志码:** A **文章编号:** 2095-0756(2015)06-0875-08

Phyllostachys violascens model optimization for bamboo shoot firmness using near-infrared spectroscopy

ZHOU Zhu^{1,2,3}, ZHENG Jian⁴, WANG Yunxiang⁴, ZENG Songwei^{1,2,3}

(1. School of Information Engineering, Zhejiang A & F University, Lin'an 311300, Zhejiang, China; 2. Zhejiang Provincial Key Laboratory of Forestry Intelligent Monitoring and Information Technology Zhejiang A & F University, Lin'an 311300, Zhejiang, China; 3. Research Center for Smart Agriculture and Forestry, Zhejiang A & F University, Lin'an 311300, Zhejiang, China; 4. School of Agriculture and Food Science, Zhejiang A & F University, Lin'an 311300, Zhejiang, China)

Abstract: To develop a calibration model for rapid, accurate and nondestructive determination of bamboo shoots firmness with *Phyllostachys violascens* by using near infrared spectroscopy (NIRS) technology. The diffuse reflectance spectra of bamboo shoot were obtained in the wavelength range from 800 to 2 632 nm. Different informative variable selection methods were first calculated with the full spectra being pretreated using a standard normal variate (SNV) transformation. Analyses with backward interval partial least squares (biPLS), synergy interval partial least squares (siPLS), genetic algorithm (GA), successive projections algorithm (SPA), Monte Carlo uninformative variable elimination (MCUVE), and competitive adaptive reweighted sampling (CARS) were compared. Then CARS and SPA were used on the spectrum to select wavelengths in proper order. The performance of the models were tested using a correlation coefficient for cross-validation of calibration (r_{cv}),

收稿日期: 2015-01-13; 修回日期: 2015-03-23

基金项目: 浙江省自然科学基金资助项目(Y3110450, LY13C200014); 浙江省科学技术公益项目(2011C22069); 浙江农林大学智慧农林业研究中心资助项目(2013ZHL03); 浙江农林大学人才启动基金资助项目(2012FR085)

作者简介: 周竹, 讲师, 博士, 从事木材智能化检测技术等研究。E-mail: zhouzhu@zafu.edu.cn。通信作者: 郑剑, 讲师, 博士研究生, 从事农产品储藏加工与品质检测等研究。E-mail: zhengjian622@126.com

root mean square error for cross-validation of calibration (RMSECV), the correlation coefficient of prediction (r_p), and the root mean square error of prediction (RMSEP). Results showed that BiPLS combined with CARS and SPA obtained a total of 25 wavelengths or only 1.6% of the full wavelengths. The r_{cv} , RMSECV, r_p , RMSEP by biPLS-CARS-SPA were 0.984, 0.300 $N \cdot cm^{-2}$, 0.926, 0.625 $N \cdot cm^{-2}$, respectively. The good performance demonstrated that NIR spectroscopy coupled with the biPLS-CARS-SPA algorithm could be used successfully to analyze bamboo shoot firmness and revealed that the biPLS-CARS-SPA algorithm was superior to other wavelength selection methods. [Ch, 6 fig, 3 tab, 20 ref.]

Key words: cash forestry; NIR spectroscopy; biPLS; CARS; SPA; firmness; bamboo shoots

雷竹 *Phyllostachys violascens* 竹笋营养丰富, 是一种低糖、低脂、高蛋白、高纤维的绿色保健食品。硬度是竹笋品质评价的重要指标, 常被用于评判竹笋成熟度及采摘时间, 并对其采后存储、保鲜以及分类深加工具有很高的参考价值。目前, 竹笋硬度的检测主要采用感官评定和质构仪测定。其中, 感官评定主观性强、重复性差。而质构仪方法虽然能准确评定竹笋硬度, 但处理程序复杂, 操作繁琐, 检测速度慢且具有破坏性, 无法满足竹笋在储藏及加工过程中的快速检测需要。近红外光谱技术(NIRS)因其快速、无损的优点, 在肉类^[1-2]、果蔬^[3-6]等农产品品质检测中得到了广泛的研究。在农产品硬度检测方面, 国内外学者已对苹果 *Malus pumila*^[7-9], 梨 *Pyrus spp.*^[10-11], 芒果 *Mangifera indica*^[12], 猕猴桃 *Actinidia chinensis*^[13]等进行了相关研究, 这为本研究提供了借鉴。目前, NIRS 技术在竹笋硬度无损检测方面的研究较少。前人^[14]的研究表明: 硬度是一个与物质密度、细胞结构和组织结构有关的复杂的综合物理指标。与硬度有关的有机物主要是果胶、纤维素等物质, 与水分含量相比, 这些物质在待测农产品中的含量极少, 因此, 水的强吸收会掩盖果胶、纤维素等物质的吸收, 从而影响待测农产品硬度近红外光谱模型的性能。通过对竹笋光谱进行变量选择, 从而对竹笋硬度模型进行优化是提高硬度光谱检测模型性能的重要途径。本研究提出采用后向间偏最小二乘法(BiPLS)结合竞争性自适应权重法(CARS)以及连续投影算法(SPA)对雷竹笋硬度近红外光谱进行变量选择, 并与一些常见的变量选择方法及其组合进行了比较, 为快速准确建立竹笋硬度的近红外光谱模型提供了一种新的方法。

1 材料与方 法

1.1 竹笋样本制备

本研究所用的雷竹笋来自浙江省临安市天目山区, 所有竹笋长度均高于 30 cm, 基部直径大于 5 cm。为了方便光谱采集, 用打孔器在雷竹笋的节间进行打孔取样, 共获取直径 15 mm, 厚度为 10 mm 的样品 113 个。将竹笋标号放置在 4 °C 冰柜中保存待用。试验前, 将待测竹笋样本从冰柜中取出置于室温(22 °C)环境中一段时间, 使待测竹笋与室温平衡。

1.2 光谱采集

近红外漫反射光谱采集仪器为 Antaris II 傅里叶变换近红外光谱仪(Thermo Scientific Co., 美国)。通过漫反射式积分球附件进行光谱采集。扫描波长范围为 800 ~ 2 632 nm, 扫描次数 64 次, 分辨率 8 cm^{-1} 。待测样品放置于带有样品固定仿形胶垫的光谱仪承载台上, 并使样品的外表面正对光谱仪通光孔。采集时样品旋转 120°·次⁻¹, 将 3 次采集的平均光谱作为该样品的原始光谱。图 1 给出了试验样本的原始光谱图。由图 1 可知: 在光谱采集范围内的首端与末端光谱信噪比较低, 结合前期预试验的计算分析, 选择 1 000 ~ 2 500 nm 波段内的光谱进行后续分析建模。

1.3 硬度测定

竹笋样品的硬度根据国家标准(GB/T 8855-2008 新鲜水果和蔬菜取样方法)测定。光谱采集后, 在

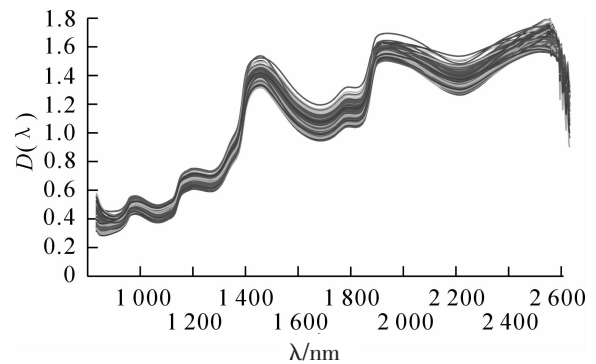


图 1 雷竹笋原始光谱

Figure 1 Spectra of bamboo shoots (*Phyllostachys violascens*)

对应的光谱采集位置采用 TA-XT2i 质构仪 (Stable Micro System Ltd, 英国) 进行硬度测量。测试探头采用的是直径 2 mm 的钢制针状压头 P2N, 探头测试深度为 5 mm, 贯入速度为 $2 \text{ mm}\cdot\text{s}^{-1}$ 。以 3 个测试点的硬度平均值作为该样本最终的硬度值。

采用蒙特卡罗采样算法^[15]剔除 5 个异常样本, 然后将 108 个样本依浓度梯度法划分为校正集和测试集。雷竹笋硬度的统计描述参数如表 1 所示。

表 1 雷竹笋硬度的统计描述参数

Table 1 Descriptive statistic data of bamboo shoots

数据集	样本数/个	平均值/($\text{N}\cdot\text{cm}^{-2}$)	测量范围/($\text{N}\cdot\text{cm}^{-2}$)	标准偏差/($\text{N}\cdot\text{cm}^{-2}$)
校正集	72	4.261	2.142~8.751	1.676
测试集	36	4.266	2.188~8.433	1.682

1.4 模型优化方法及评价

我们提出采用基于区域的波长选择方法和基于单个变量的波长选择方法组合使用来对雷竹笋硬度检测的近红外光谱模型进行优化。涉及的算法主要包括后向区间偏最小二乘法 (biPLS), 竞争性自适应重加权算法 (CARS) 以及连续投影算法 (SPA)。biPLS 是建立在间隔偏最小二乘法 (iPLS) 基础上的一种基于区域的变量选择方法。该方法将全波段范围划分为 k 个等宽的子区间, 去掉子区间 1 个·次⁻¹, 在剩余的区间内计算各个组合区间的偏最小二乘回归模型, 然后将交叉验证均方误差 (RMSECV) 值最小的组合模型时所去掉的子区间作为第一去掉子区间, 重复上述操作直至程序运行结束。最后比较各组合模型的交叉验证均方误差 RMSECV 值, 以 RMSECV 值最小对应的区间组合作为最优区间组合^[16]。CARS 算法由 Liang 等提出, 该算法模仿达尔文进化论中“适者生存”思想, 将每个变量看成一个个体, 通过自适应重加权采样技术对变量实施逐步淘汰。在此过程中还引入了指数衰减函数来控制变量的保留率, 计算效率高, 适用于高维数据的变量选择^[17]。SPA 也是一种基于单个变量的波长选择方法。该方法能够利用向量的投影分析, 寻找含有最低限度冗余信息的变量组合, 并使变量之间的共线性程度最低, 从而提高建模速度与模型性能^[18]。biPLS, CARS 和 SPA 算法的原理分别见文献[16-18]。

采用偏最小二乘法 (PLS) 建立最终模型, 模型中因子数由交叉验证法确定。模型的评价指标包括交叉验证相关系数 r_{cv} , 预测相关系数 r_p , 交叉验证均方误差 RMSECV 以及预测均方误差 RMSEP。 r_{cv} 和 r_p 越接近 1, 以及 RMSECV 和 RMSEP 越低, 则模型的预测能力和稳定性越好。上述各种算法以及最终模型的建立与评价均通过 MATLAB 2010 (The Math Works, 美国) 软件完成。

2 结果与分析

2.1 预处理方法的确定

受仪器、样品以及光谱采集环境等因素的影响, 近红外光谱中常出现噪声、谱图基线漂移和平移等现象。为了消除这些不利因素对模型的影响, 应对原始光谱 (none) 进行预处理。采用标准正态变量变换 (SNV), 去趋势变换 (detrend), 一阶导数 (FD), 二阶导数 (SD) 等方法对原始光谱进行预处理并建立 PLS 模型, 结果如表 2 所示。由表 2 可知: 标准正态变量变换预处理建立的 PLS 模型最优, 该模型的 r_{cv} , r_p 分别从无预处理时的 0.920, 0.895 提高到 0.928, 0.901; RMSECV 和 RMSEP 则分别从无预处理时的 $0.653 \text{ N}\cdot\text{cm}^{-2}$, $0.739 \text{ N}\cdot\text{cm}^{-2}$ 下降为 $0.619 \text{ N}\cdot\text{cm}^{-2}$, $0.718 \text{ N}\cdot\text{cm}^{-2}$ 。此外, 建模所用因子数也从 19 减少到

表 2 基于不同预处理方法处理的建模和预测结果

Table 2 Performance of firmness models with different preprocessing methods

预处理方法	因子数	r_{cv}	RMSECV/($\text{N}\cdot\text{cm}^{-2}$)	r_p	RMSEP/($\text{N}\cdot\text{cm}^{-2}$)
原始光谱 (none)	19	0.920	0.653	0.895	0.739
一阶导数 (FD)	20	0.921	0.639	0.901	0.728
去趋势变换 (detrend)	16	0.928	0.619	0.874	0.805
二阶导数 (SD)	6	0.892	0.754	0.832	0.921
标准正态变量变换 (SNV)	18	0.928	0.619	0.901	0.718

18。而其他预处理方法的改善效果不佳。这表明 SNV 预处理有利于减弱因竹笋样本表面不平引起的噪声干扰。可以增强模型的预测能力和稳健性。因此,在下文均采用 SNV 处理后的光谱进行分析。

2.2 基于 biPLS 的模型简化

biPLS 方法中,由于光谱区间的划分无相应的理论依据,且划分的区间越多,所需计算时间也越长,考虑到雷竹笋硬度定量分析中波长变量为 1 557 个,采用试探法依次将全波段划分为 6~50 个区间,以 RMSECV 值最小来确定最优的区间划分。经过计算分析,当将整个光谱波段划分为 48 个区间时,所建模型的 RMSECV 值最小。图 2 为竹笋硬度 biPLS 方法的特征波长选择结果,图中被填充的波段区间为被选中的波段区间,对应的波段范围为 1 129.239~1 145.200 nm, 1 162.661~1 179.588 nm, 1 198.121~1 216.104 nm, 1 235.811~1 254.952 nm, 1 297.015~1 318.115 nm, 1 437.213~1 462.343 nm, 1 490.078~1 517.105 nm, 1 710.247~1 745.947 nm, 1 825.867~1 910.636 nm, 2 290.400~2 354.892 nm, 共包含 357 个波长变量。采用 PLS 方法建模,依据 RMSECV 最小确定最优的因子数为 19,应用该模型对测试集数据进行预测,模型的 r_{cv} , r_p , RMSECV 和 RMSEP 分别为 0.963, 0.913, 0.451 $N \cdot cm^{-2}$ 和 0.675 $N \cdot cm^{-2}$ 。与全波段光谱模型相比,biPLS 模型采用的波长点由 1 557 减少到了 357 个,在保证精度的前提下简化了模型。

2.3 基于 biPLS-CARS 的模型二次简化

尽管通过 biPLS 算法剔除了大量与雷竹笋硬度检测无关的信息,且提高了模型性能。然而,biPLS 作为一种光谱变量区域选择方法,在入选的区间内,相邻的变量之间仍具有高相关性。采用 CARS 法对 biPLS 处理后的 357 个光谱变量进行变量选择,图 3 描述了具体的选择过程。其中:图 3a 为所选变量数随采样次数的变化趋势,所选变量数随采样次数的增加由快到慢递减,体现出粗选与精选 2 个阶段;图 3b 为 10 折交叉验证所得 RMSECV 的变化趋势图,在 1~132 次采样建模中, RMSECV 呈现递减趋势,表明剔除了与雷竹笋硬度检测无关的变量,133 次采样建模以后 RMSECV 开始递增,则可能剔除了雷竹笋硬度检测的关键变量;图 3c 中各线表示随着采样次数增加,各光谱变量的回归系数的变化趋势,“*”标示出最小 RMSECV 所对应的采样次数。最终, CARS 法选择了 29 个光谱变量。利用这些光谱变量的数据建立 PLS 模型,因子数为 14,模型的 r_{cv} , r_p , RMSECV 和 RMSEP 分别为 0.983, 0.926, 0.306 $N \cdot cm^{-2}$ 和 0.628 $N \cdot cm^{-2}$ 。biPLS-CARS 模型所用变量数仅为全波段变量数的 1.863%,但模型的性能却比全波段模型有了提高。

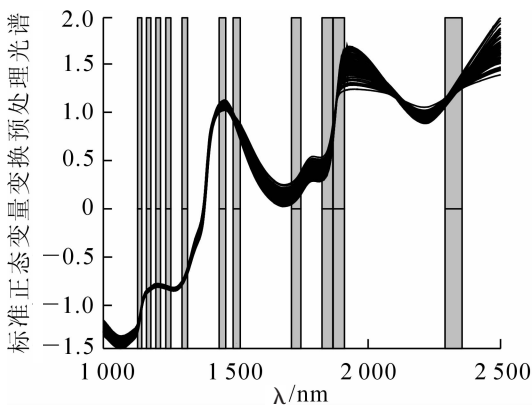


图 2 biPLS 算法选择的最优区间组合
Figure 2 Optimal spectral region selected by biPLS

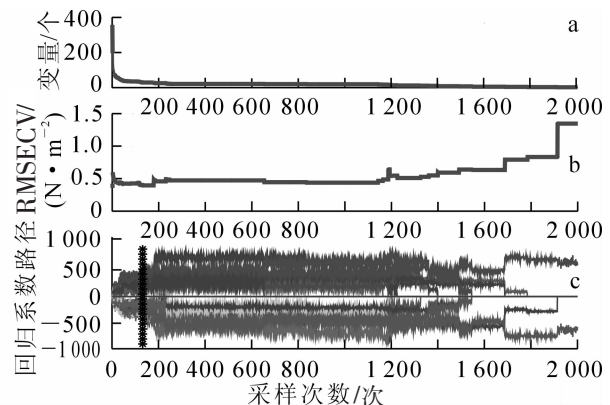


图 3 CARS 算法的变量选择过程
Figure 3 Plot of CARS for variable selection

2.4 基于 biPLS-CARS-SPA 的模型三次简化

经过 biPLS-CARS 变量选择后,变量个数从全波段时的 1 557 个减少为 29 个,然而,光谱变量仍有进一步减少的空间。SPA 利用向量的投影分析,寻找含有最低限度冗余信息且共线性最小的变量组合,可以减少建模所用变量数,提高建模的速度和性能。

图 4 给出了 SPA 方法下不同个数变量建模的 RMSCEV 值的变化情况。从图 4 中可以看出:当变量个数从 1 个增加到 23 个时, RMSCEV 值下降较快。当变量个数为 25 时, RMSCEV 值最小。利用所选

25 个光谱变量的数据建立 PLS 模型，模型的因子数为 14，校正集的 r_{cv} ，RMSECV 分别为 0.984，0.300 $N \cdot cm^{-2}$ ，测试集的 r_p ，RMSEP 分为 0.926，0.625 $N \cdot cm^{-2}$ 。模型对样本预测值和真实测量值之间的散点图如图 5 所示。与 biPLS-CARS 方法相比，该模型采用的波长个数又减少了 4 个，且模型的性能再次得以提升。

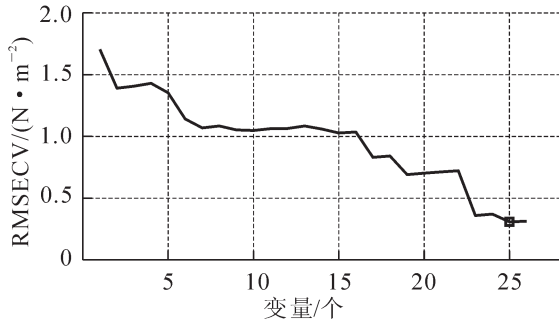


图 4 SPA 算法中变量个数与 RMSECV 的对应关系

Figure 4 RMSECV versus number of variables in SPA

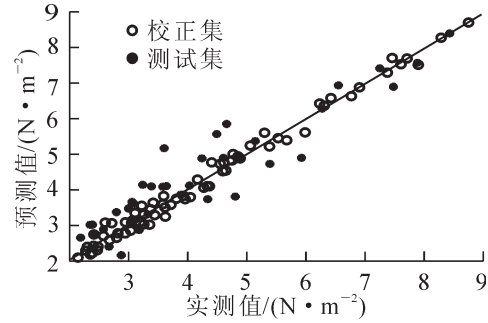


图 5 biPLS-CARS-SPA-PLS 模型的实测值和预测值的散点图

Figure 5 Reference measurement versus predicted values by biPLS-CARS-SPA-PLS

2.5 结果比较与分析

为了说明 biPLS-CARS-SPA 方法在雷竹笋硬度近红外光谱检测模型优化方面的优越性，将 biPLS-CARS-SPA 方法所建模型分别与组合区间偏最小二乘法 (siPLS)^[16]，遗传算法 (GA)^[19]，蒙特卡洛无信息变量消除法 (MCUVE)^[20]，CARS，SPA 等方法单独或组合作用下所建模型进行比较，比较的结果如表 3 所示。从表 3 可以看出：相比于 siPLS，GA，MCUVE，CARS，SPA 等方法的单独或组合作用，biPLS-CARS-SPA 方法建模所需变量最少，模型的预测精度最高，模型优化效果最好。

表 3 基于不同变量选择方法及建模方法的建模和预测结果

Table 3 Performance of model based on different spectra variable selection

建模方法	因子数	变量数	r_{cv}	RMSECV/($N \cdot cm^{-2}$)	r_p	RMSEP/($N \cdot cm^{-2}$)
PLS	18	1 557	0.928	0.619	0.901	0.718
biPLS	19	357	0.963	0.451	0.913	0.675
siPLS12-4 *	20	519	0.929	0.604	0.873	0.810
CARS-PLS	15	36	0.941	0.530	0.902	0.716
SPA-PLS	18	70	0.905	0.708	0.822	0.945
MCUVE	19	146	0.961	0.471	0.912	0.681
GA-PLS	16	199	0.935	0.584	0.882	0.781
biPLS -GA-PLS	14	94	0.959	0.469	0.920	0.649
biPLS -CARS-PLS	14	29	0.983	0.306	0.926	0.628
biPLS -SPA-PLS	16	70	0.942	0.557	0.871	0.815
biPLS -CARS-SPA-PLS	14	25	0.984	0.300	0.926	0.625
biPLS -MCUVE	17	121	0.955	0.498	0.901	0.721

说明：* 表示 siPLS12-4 将全波段划分为 12 个区间，4 个区间进行组合。

图 6 给出了雷竹的原始光谱以及文中各种方法选择的变量分布情况。据文献[20]报道，竹笋中与硬度有关的物质主要是果胶、纤维素、半纤维素等，这些物质的特征官能团主要有 C—H，O—H 以及 C=O。从图 6 中可以看到：雷竹笋在近红外光谱区的主要吸收峰位于 1 190 nm，1 450 nm 和 1 940 nm 附近，这些吸收峰主要是由水的吸收所导致的。其中 1 190 nm 附近为 O—H 的合频吸收峰，1 450 nm 附近为 O—H 伸缩振动的一级倍频吸收峰，1 940 nm 附近则是 O—H 伸缩振动的二级倍频吸收峰。由于果胶和纤维素成分含量在笋中相对较低，而水分含量较高 (达 90%)，水在近红外区的强烈吸收掩盖了果胶、纤维素等物质的吸收，这就要求必须进行雷竹笋硬度模型的优化研究。

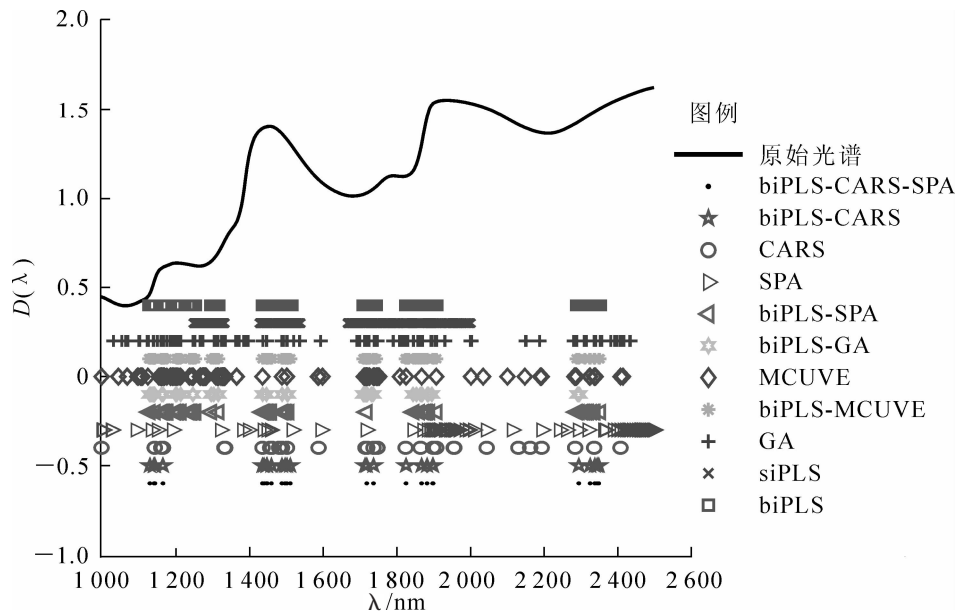


图6 不同优化方法的变量选择分布

Figure 6 Selection of variables distribution of different variable selection methods

在基于 biPLS 算法的模型简化算法选择过程中,分别采用 biPLS, siPLS, GA, CARS, MCUVE 算法对 SNV 预处理的雷竹笋光谱进行变量优选时,上述 5 种算法选择的波长区间具有一定的相似性,其中, biPLS, GA, CARS, MCUVE 算法均选择了 1 100~1 200 nm, 1 420~1 500 nm, 1 720~1 780 nm, 1 805~1 940 nm, 2 290~2 360 nm 范围内的几个或多个波长点。在这些波长变量中, 1 152 nm 附近为果胶中甲基($-\text{CH}_3$)伸缩振动的二级倍频吸收带, 1 440 nm 附近为 C—H 伸缩振动的一级倍频吸收带, 1 450 nm 附近为 O—H 伸缩振动的一级倍频带, 1 490 nm 附近为纤维素中 O—H 伸缩振动的一级倍频吸收带, 1 780 nm 附近为纤维素中 C—H 伸缩振动的一级倍频吸收带, 2 280 nm 附近为果胶中甲基($-\text{CH}_3$)伸缩振动及变形振动的合频吸收带, 1 820 nm 附近为纤维素中 O—H 伸缩振动、C=O 伸缩振动的一级倍频吸收带, 2 336 nm 为纤维素中 C—H 的伸缩振动及变形振动的组合频吸收带, 2 352 nm 则为纤维素中 C—H 伸缩振动的二级倍频吸收带^[21]。由于这些方法均选择了与竹笋硬件检测相关的波段组合并且去掉了一些无关变量,因此所建模型的性能均优于全光谱模型。然而, SPA 方法选择了过多的无关变量(雷竹笋光谱的首端与尾端),其建模效果最差。而 siPLS 方法未将果胶中与甲基($-\text{CH}_3$)有关的波段选入,其模型的性能也较差。CARS, UVE 和 GA 等选择的变量个数要少于 biPLS,但由于他们或多或少的丢失了一些与雷竹笋硬度检测相关的关键变量,因此所建模型的性能仍然要低于采用 357 个变量建模的 biPLS 模型。

基于 biPLS-CARS 算法对模型进行二次简化。由于 biPLS 方法作为一种区域变量选择方法,在区域内的相邻波长之间仍然有一定的相关性。采用 CARS, SPA, MCUVE, GA 算法对 biPLS 处理的光谱进行二次变量选择, CARS 方法选择的变量数最少,仅为 29 个,而 SPA, MCUVE 和 GA 方法选择的变量数均超过了 70 个,关键变量的丢失以及相邻波长点之间的强相关性导致 MCUVE, SPA 方法所建模型的性能比全光谱模型的性能还差。虽然 GA 所建模型性能较好,仅略低于 CARS 方法所建模型性能,但是所用变量个数(94 个)却远多于 CARS 方法(29 个)。CARS 方法保留了最多的与雷竹笋硬度检测相关的变量,较大程度去掉了无关变量,其模型性能最好。

在基于 biPLS-CARS-SPA 的模型 3 次简化过程中,采用 SPA 方法对 biPLS-CARS 处理后的变量进一步进行变量选择,变量个数从 29 个减少为 25 个,所选变量波长分布如图 6 所示。从图 6 中可以看到: biPLS-CARS-SPA 所选波长位于或接近雷竹笋硬度相关物质特征官能团吸收频带,且避开了水的强吸收波段如 1 190 nm, 1 940 nm,减小了水分对雷竹笋硬度检测模型的影响。biPLS-CARS-SPA 模型所用变量仅为全部变量的 1.605%,模型的 r_w , r_p 却比全光谱模型提高 5.957%, 2.762%, RMSECV 以及 RMSEP 分别降低了 51.543%, 13.002%。biPLS 与 CARS 以及 SPA 变量选择方法的组合使用,表明在采用近红

外光谱对雷竹笋硬度建模时,有必要在 SPA 计算之前对全波段光谱变量进行 biPLS 与 CARS 的组合变量选择,以最大程度减少冗余变量,从而提高模型性能。

3 结论

该研究利用近红外光谱技术快速检测雷竹笋的硬度并开展了模型的优化研究。在模型优化过程中,用后向间偏最小二乘法结合竞争性自适应权重法与连续投影算法(biPLS-CARS-SPA)对雷竹笋硬度近红外光谱进行特征光谱区域和特征波长的选取。结果表明:与其他模型优化方法相比,biPLS-CARS-SPA 方法显示出明显的优越性,不仅能有效减少建模所用变量个数,而且保留的特征波长避开了水分的影响,具有实际的物理表征意义。biPLS-CARS-SPA 模型建立所用变量个数为 25,因子数为 14,模型的 r_{cv} , r_p , RMSECV 和 RMSEP 分别为 0.984, 0.926, 0.300 和 $0.625 \text{ N} \cdot \text{cm}^{-2}$,模型的复杂度降低且预测精度显著提高,具有很高的实用价值。

4 参考文献

- [1] 程旒, 李小昱, 赵思明, 等. 鱼体新鲜度近红外光谱检测方法的比较研究[J]. 食品安全质量检测学报, 2013, **4**(2): 427 - 432.
CHENG Ni, LI Xiaoyu, ZHAO Siming, *et al.* Comparison of the methods to detect fish freshness based on near-infrared technology [J]. *J Food Saf Qual*, 2013, **4**(2): 427 - 432.
- [2] 汤修映, 牛力钊, 徐杨, 等. 基于可见/近红外光谱技术的牛肉含水率无损检测[J]. 农业工程学报, 2013, **29**(11): 248 - 254.
TANG Xiuying, NIU Lizhao, XU Yang, *et al.* Nondestructive determination of water content in beef using visible/near-infrared spectroscopy [J]. *Trans Chin Soc Agric Eng*, 2013, **29**(11): 248 - 254.
- [3] 李小昱, 陶海龙, 高海龙, 等. 基于多源信息融合技术的马铃薯痂疮病无损检测方法[J]. 农业工程学报, 2013, **29**(19): 277 - 284.
LI Xiaoyu, TAO Hailong, GAO Hailong, *et al.* Nondestructive detection method of potato scab based on multi-sensor information fusion technology [J]. *Trans Chin Soc Agric Eng*, 2013, **29**(19): 277 - 284.
- [4] 刘洁, 李小昱, 李培武, 等. 基于近红外光谱的板栗水分检测方法[J]. 农业工程学报, 2010, **26**(2): 338 - 341.
LIU Jie, LI Xiaoyu, LI Peiwu, *et al.* Determination of moisture in chestnuts using near infrared spectroscopy [J]. *Trans Chin Soc Agric Eng*, 2010, **26**(2): 338 - 341.
- [5] 介邓飞, 谢丽娟, 饶秀勤, 等. 近红外光谱变量筛选提高西瓜糖度预测模型精度[J]. 农业工程学报, 2013, **29**(12): 264 - 270.
JIE Dengfei, XIE Lijuan, RAO Xiuqin, *et al.* Improving accuracy of prediction model for soluble solids content of watermelon by variable selection based on near-infrared spectroscopy [J]. *Trans Chin Soc Agric Eng*, 2013, **29**(12): 264 - 270.
- [6] 周万怀, 谢丽娟, 应义斌. 全光谱匹配算法在苹果分类识别中的应用[J]. 农业工程学报, 2013, **29**(19): 285 - 292.
ZHOU Wanhuai, XIE Lijuan, YING Yibin. Application of full spectral matching algorithm in apple classification [J]. *Trans Chin Soc Agric Eng*, 2013, **29**(19): 285 - 292.
- [7] 李桂峰, 赵国建, 刘兴华, 等. 苹果硬度的傅里叶变换近红外光谱无损检测[J]. 农业机械学报, 2009, **40**(1): 120 - 123.
LI Guifeng, ZHAO Guojian, LIU Xinghua, *et al.* Using FT-NIR spectra in non-invasive measurement of apple firmness [J]. *Trans Chin Soc Agric Mach*, 2009, **40**(1): 120 - 123.
- [8] FAN Guoqiang, ZHA Jianwen, DU Ran, *et al.* Determination of soluble solids and firmness of apples by Vis/NIR transmittance [J]. *J Food Eng*, 2009, **93**(4): 416 - 420.
- [9] BLANKE M M. Non-invasive assessment of firmness and NIR sugar (TSS) measurement in apple, pear and kiwi fruit [J]. *Erwerbs-Obstbau*, 2013, **55**(1): 19 - 24.
- [10] LI Jiangbo, HUANG Wenqian, ZHAO Chunqiang, *et al.* A comparative study for the quantitative determination of soluble solids content, pH and firmness of pears by Vis/NIR spectroscopy [J]. *J Food Eng*, 2013, **116**(2): 324 - 332.

- [11] VALENTE M, LEARDI R, SELF G, *et al.* Multivariate calibration of mango firmness using vis/NIR spectroscopy and acoustic impulse method [J]. *J Food Eng*, 2009, **94**(1): 7 – 13.
- [12] 刘卉, 郭文川, 岳绒. 猕猴桃硬度近红外漫反射光谱无损检测[J]. 农业机械学报, 2011, **42**(3): 145 – 149.
LIU Hui, GUO Wenchuan, YUE Rong. Nond-destructive detection of kiwifruit firmness based on near-infrared diffused spectroscopy [J]. *Trans Chin Soc Agric Mach*, 2011, **42**(3): 145 – 149.
- [13] 王琪. 竹笋采后保鲜及软包装笋储藏品质变化的研究[D]. 杭州: 浙江工商大学, 2011.
WANG Qi. *Studies on Preservation of Harvested Bamboo Shoot and Change in Quality of Soft Package Shoot During Storage at Lower Temperature* [D]. Hangzhou: Zhejiang Gongshang University, 2011.
- [14] CAO Dongsheng, LIANG Yizeng, XU Qingsong, *et al.* A new strategy of outlier detection for QSAR/QSPR [J]. *J Comput Chem*, 2010, **31**(3): 592 – 602.
- [15] LEARDI R, NØRGAARD L. Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions [J]. *J Chemometr*, 2004, **18**(11): 486 – 497.
- [16] LI Hongdong, LIANG Yizeng, XU Qingsong, *et al.* Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration [J]. *Anal Chim Acta*, 2009, **648**(1): 77 – 84.
- [17] ARAÚJO M C U, SALDANHA T C B, GALVÃO R K H, *et al.* The successive projections algorithm for variable selection in spectroscopic multicomponent analysis [J]. *Chemometr Intell Lab*, 2001, **57**(2): 65 – 73.
- [18] LEARDI R. Application of genetic algorithm-PLS for feature selection in spectral data sets [J]. *J Chemometr*, 2000, **14**(5/6): 643 – 655.
- [19] HAN Qingjuan, WU Hailong, CAI Chenbo, *et al.* An ensemble of Monte Carlo uninformative variable elimination for wavelength selection [J]. *Anal Chim Acta*, 2008, **612**(2): 121 – 125.
- [20] 严衍禄. 近红外光谱分析基础与应用[M]. 北京: 中国轻工业出版社, 2005: 31 – 39.