

基于高光谱成像技术结合 SPA 和 GA 算法测定甜玉米种子电导率

张婷婷¹, 赵 宾¹, 杨丽明², 王建华¹, 孙 群^{1*}

1. 中国农业大学农学院植物遗传育种与种子科学系, 农业部农作物种子全程技术研究北京创新中心, 北京市作物遗传改良重点实验室, 北京 100193
2. 中国农业大学理学院, 北京 100083

摘 要 种子活力对于农业发展至关重要, 而甜玉米种子普遍存在活力较低且不耐贮藏的问题。因此, 及时准确地对甜玉米种子活力进行检测尤为重要。电导率测定法作为一种传统的种子活力检测方法, 存在对种子有一定破坏性、耗时较长、重复性不佳等缺点。针对这些问题, 尝试利用可见-近红外(VIS-NIR)高光谱成像系统结合化学计量学算法建立甜玉米种子电导率快速、无损且精确的检测方法。以高温高湿老化的绿色超人甜玉米种子为试验材料, 先通过可见-近红外高光谱成像系统采集种子的高光谱图像和进行电导率测定试验, 随后对高光谱图像进行黑白校正、提取感兴趣区域, 获取光谱反射率数据。利用多种预处理方法分别为标准正态变量变换(SNV)、二阶导(SD)、一阶导(FD)、和多元散射校正(MSC)建立甜玉米种子电导率的偏最小二乘回归(PLSR)模型, 比较分析并筛选出最适预处理方法。再通过连续投影算法(SPA)及遗传算法(GA)对 MSC 预处理后的高光谱波段进行筛选提取, 基于选出的特征波段建立 PLSR 模型, 并与全波段(Full)PLSR 模型进行对比分析, 得到与甜玉米种子电导率相关性最高的高光谱波段组合, 最终确立一种能够预测甜玉米种子电导率的方法体系。实验结果显示: 不同预处理方法(SNV, FD, SD 和 MSC)建立的 PLSR 模型性能有所差异, 其中 MSC-PLSR 模型的表现最优秀, 其校正决定系数和预测决定系数分别为 0.983 和 0.974, 相应的校正均方根误差和预测均方根误差分别为 0.165 和 0.226。进一步分析 MSC-Full-PLSR, MSC-SPA-PLSR 和 MSC-GA-PLSR 模型, 发现 GA 能够将全光谱的 853 个波段压缩至 25 个有效波段, 所建立的 MSC-GA-PLSR 模型仍表现优秀, 其校正决定系数和预测决定系数分别为 0.976 和 0.973, 相应的校正均方根误差和预测均方根误差分别为 0.194 和 0.212。实验结果表明: 基于可见-近红外(VIS-NIR)高光谱成像系统结合化学计量学算法实现对甜玉米种子电导率的预测存在一定的可行性。该研究为甜玉米种子电导率的快速、无损且精确的检测提供一定的理论支持。

关键词 高光谱技术; 电导率; 甜玉米种子; 特征波段; 偏最小二乘回归

中图分类号: S123; TP391.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)08-2608-06

引 言

种子的活力状况在极大程度上影响着农作物产量和品质^[1]。甜玉米(*Zea mays* L. *saccharata* Sturt)是玉米属(*Zea mays* L)的一个亚种, 因其携带了一种直接影响碳水化合物代谢的隐形突变基因, 使籽粒淀粉比例减少, 糖分比例增加^[2], 而深受消费者喜爱, 具有广阔的市场需求和前景^[3]。但甜玉米种子存在活力水平普遍较低且不耐贮藏的问题, 是甜玉米生产及推广的主要限制因素。因此针对甜玉米种子的

活力检测尤为重要。目前常用的传统种子活力检测方法之一为电导率测定法^[4-5], 但其存在对种子造成一定程度破坏、检测时间较长、重复性差等缺点, 已无法满足当今种业市场快速发展的需求。因此亟需一种无损、快速、精确检测种子电导率的新方法。

高光谱成像技术是一种融合了传统的图像和光谱技术的新兴光电检测技术, 即利用高光谱仪器能够获得待测样品的图像, 反映样品的外部形态和颜色信息, 同时图像上的每个像素点都包含了全波段光谱信息, 反映样品的内部结构及化学成分等特性^[6-7]。目前, 高光谱技术的潜力和优势已在诸

收稿日期: 2018-06-27, 修订日期: 2018-11-08

基金项目: 科技部“十三五”国家重点研发计划项目(2018YFD0100903), 北京市科委项目(Z151100001015004)资助

作者简介: 张婷婷, 1990年生, 中国农业大学农学院博士研究生 e-mail: ztt_cau@163.com

* 通讯联系人 e-mail: sqcau@126.com

多领域被研究及应用^[8-10]。在种子检测领域,很多学者对种子的内含物、品种、产地等方面开展了研究工作。如 Cheng 等^[11]将近红外高光谱与化学计量分析结合,实现了对花生种子含油量和蛋白质含量的预测。Gao 等^[12]利用高光谱技术实现了对不同产地麻风树种子的鉴别。Zhao 等^[13]采用高光谱成像技术对玉米种子品种进行了分类。但对传统种子电导率检测方法的优化研究较少,且将高光谱技术与化学计量学算法结合对甜玉米种子进行研究的文献也较为罕见。

本研究通过高光谱成像技术对不同老化梯度的甜玉米种子的光谱信息进行采集,随后对甜玉米种子电导率进行测定,结合多种预处理方法和偏最小二乘回归(partial least squares regression, PLSR)算法建立全波段回归模型,筛选出最佳预处理方法,再采用连续投影算法(Successive projections algorithm, SPA)和遗传算法(Genetic algorithm, GA)得到与甜玉米种子电导率密切相关的光谱波段,比较分析不同特征波段提取方法对 PLSR 模型预测电导率精度的影响,为研究精确、无损且便捷的甜玉米种子电导率检测技术提供理论依据。

1 实验部分

1.1 材料

试验材料为绿色超人甜玉米种子,于 2016 年购于北京中品开元种子有限公司,初始含水率为 10.0%。筛选其中 2 000 粒结构完整且无霉变的甜玉米种子用于后续试验。

1.2 人工加速老化

将 2 000 粒甜玉米种子平均分成 4 组,每组 500 粒(2 000 粒=4×500),其中一组放入 4℃ 冰箱保存待用,其余 3 组分别放入 3 个灭菌后的尼龙网袋中,密封于底部存有少量无菌水的玻璃干燥皿里(相对湿度为 100%),置于温度为 45℃ 的电子恒温烘箱中老化。每隔 24 h 取出一组,回干至初始含水率后进行高光谱扫描和电导率测定试验。

1.3 高光谱成像仪和光谱信息采集软件

实验采用的高光谱成像仪器同 Zhang 等^[14]。先设置系统的参数为电控平台移动速度 1.1 mm·s⁻¹,相机曝光时间 15 ms,物距 21.5 cm。随后进行黑白板图像采集,再对样品种子采集高光谱的图像信息。采集光谱信息时,先将每组甜玉米种子平均分成 20 份,每份 25 粒,以每份为一个研究对象,每份甜玉米种子的胚面朝上摆放在电控位移台上,随着平台移动,摄像头扫描整个平台。为避免室内其他光线干扰,高光谱图像采集过程在暗箱内完成。采集高光谱图像信息的软件为 Spectral Image Software (Isuzu Optics Corp., Taiwan, China)。

1.4 电导率测定

采用雷磁 DDS-307 电导率仪对每个老化梯度的甜玉米种子进行电导率测定实验^[15]。试验前将去离子水密封并置于(25±1)℃,40%RH 的恒温室中恒定 24h,随后准确称取 50 粒样品种子的重量(W),用去离子水冲洗每一粒甜玉米种子 3 次,滤纸吸干表面水分,装入 500 mL 锥形瓶中,加入 250 mL 去离子水,测定初始电导率(D₁);浸泡 24 h 后测定

浸出液电导率(D₂),计算如式(1)所示。不同老化时间的甜玉米种子电导率测定结果见表 1。

$$\text{种子浸出液电导率}[\mu\text{S}\cdot(\text{cm}\cdot\text{g})^{-1}] = (D_2 - D_1)/W \quad (1)$$

表 1 不同老化时间的甜玉米种子电导率

Table 1 Conductivity of sweet corn seeds with different ageing time

Ageing time/h	Conductivity/($\mu\text{S}\cdot\text{cm}^{-1}\cdot\text{g}^{-1}$)		
	Minimum	Maximum	Average
0	21.715 5	27.246 2	23.722 0
24	25.171 0	26.826 5	26.266 3
48	23.903 2	26.391 3	25.216 7
72	25.564 4	29.938 0	27.077 4

1.5 高光谱图像处理及数据分析

在对高光谱图像信息进行提取前,需利用 HSI Analyzer 软件(Isuzu Optics Corp., Taiwan, China)对光谱图像如式(2)进行黑白板校正。式中,I 为校正后的光谱图像;I₀为原始图像;B 为盖上镜头盖后获取的黑板图像;W 为放置聚四氟乙烯白板获取的白板图像。

$$I = \frac{I_0 - B}{W - B} \quad (2)$$

黑白板校正后,需将每份甜玉米种子从高光谱图像中分割出来。利用 ENVI4.7 软件(Research System Inc, Boulder, Co., USA),任意选取种子和背景区域,计算其光谱反射率,确定反射差值最大的波段^[16],并设定阈值将每份甜玉米种子从背景中分割出来,随后利用形态学滤波及掩膜处理获取感兴趣区域(region of interest, ROI)^[17](图 1),提取每一份种子 400.2~1 069.9 nm 处的平均反射光谱。80 份甜玉米种子的反射光谱如图 2 所示。

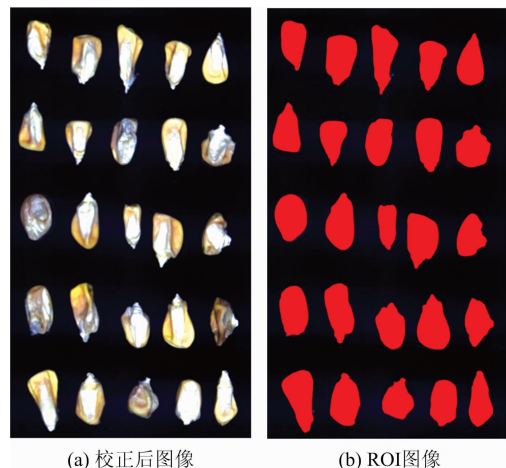


图 1 甜玉米种子(a)校正后高光谱图像及(b)波长 850 nm ROI 高光谱图像

Fig. 1 Hyperspectral images of sweet corn seeds (a) after correction and (b) at wavelength of 850 nm of ROI

随机选取 4 个老化梯度中每个梯度的 15 份甜玉米种子

建立校正集($4 \times 15 = 60$ 份), 余下的作为预测集($5 \times 4 = 20$ 份)。模型性能以校正决定系数(determination coefficients of calibration, R_c), 校正均方根误差(root mean square error of calibration, RMSEC), 预测决定系数(determination coefficients of prediction, R_p)和预测均方根误差(root mean square error of prediction, RMSEP)来评价。数据建模分析软件为 Matlab R2014a (The Math Works, Natick, USA)。

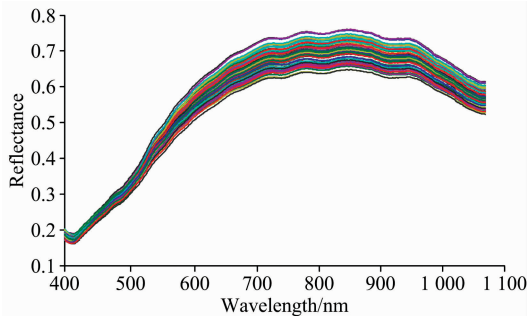


图 2 80 份甜玉米种子的反射光谱

Fig. 2 Raw reflectance spectra of 80 samples of sweet corn seeds

2 结果与讨论

2.1 基于预处理的全波段光谱 PLSR 建模分析

在采集原始光谱时, 样品状态、仪器性能及其他外界环境的干扰都会引入噪音变量, 因此需要对原始光谱信息进行预处理。通过比较分析原始及不同预处理后光谱数据建立的 PLSR 模型表现优劣, 确定最适预处理方法。其中涉及到的预处理方法为标准正态变量变换(standard normal variate, SNV)、一阶求导(first derivative, FD)、二阶求导(second derivative, SD)及多元散射校正(multiplicative scatter correction, MSC)。表 2 可见, 不同预处理方法对电导率的 PLSR 模型性能影响不同, 其中 MSC-PLSR 模型的表现最优, 校正集与预测集的 R_c 和 R_p 分别为 0.983 和 0.974, 相应的 RMSEC 和 RMSEP 分别为 0.165 和 0.226。可见, MSC 预处理方法能够对原始光谱信息进行去噪, 有效地提高信噪比, 提高对电导率预测的精度。经 MSC 预处理后的光谱曲线如图 3 所示。

表 2 基于不同预处理的 PLSR 模型结果

Table 2 PLSR model results of spectra after preprocessing with different methods

Pre-processing methods	Number of factors	Calibration sets		Prediction sets	
		R_c	RMSEC	R_p	RMSEP
None	11	0.952	0.282	0.912	0.449
SNV	9	0.961	0.216	0.938	0.283
SG1	6	0.895	0.429	0.885	0.467
SG2	4	0.769	0.716	0.456	1.260
MSC	10	0.983	0.165	0.974	0.226

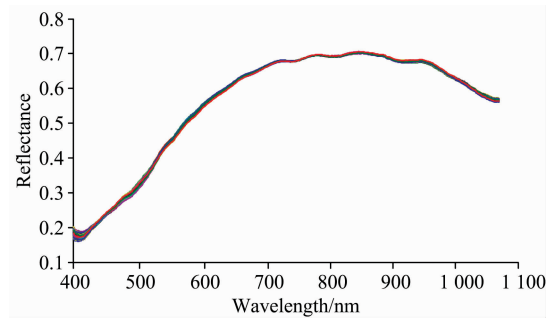


图 3 经 MSC 预处理后的反射光谱

Fig. 3 Spectrograms of raw spectra after preprocessing with MSC algorithm

2.2 电导率参数特征波段筛选

2.2.1 SPA 变量筛选法

SPA 作为一种前向循环的敏感波段选择算法, 能够在筛选出最低冗余信息变量组合的同时有效消除变量间的共线性问题, 能够最大限度的获取解释信息, 降低模型的复杂度^[18]。基于 SPA 法对甜玉米种子电导率敏感波段筛选结果见图 4。图 4(a)可见随着变量数增加, 均方根误差 RMSE 值逐渐减小, 当变量数为 35 时(空心方框), RMSE 值最小, 之后趋向平缓。图 4(b)为该 35 个波段点(空心方框)在光谱曲线上的具体分布情况。采用 SPA 法筛选出的特征波段详见表 3。

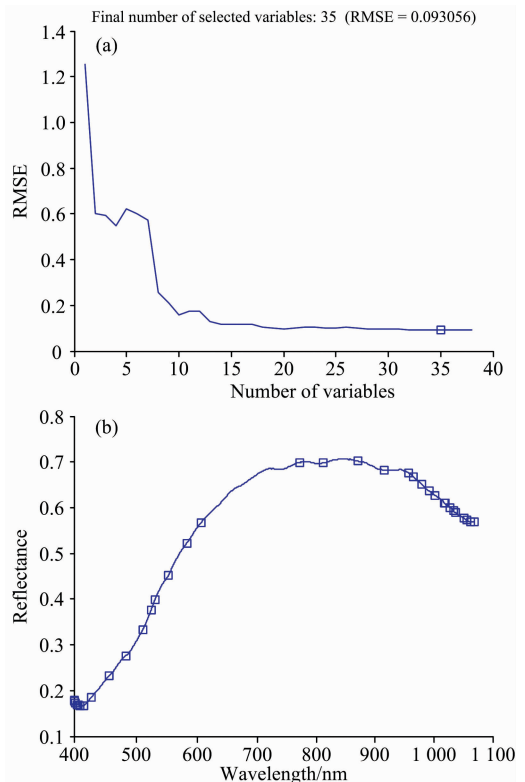


图 4 (a) SPA 法筛选的特征波段的数量;

(b) 特征波段的具体位置

Fig. 4 (a) Number of characteristic wavelengths selected by SPA algorithm; (b) Detailed position of characteristic wavelengths

些特定波段可能与植物色素有关,如叶绿素 II a 在 430, 662 和 680 nm, 叶绿素 II b 在 448 和 642 nm, 类胡萝卜素在 448 和 471 nm 附近^[21-22], 花青素的吸收峰在 535 nm 附近^[23]均存在吸收峰。近红外区域中 760~1 070 nm 的光谱特性主要是由于 O—H 键二级倍频、三级倍频和 C—H 键三级倍频振动引起的^[23-25]。另外, GA 法选出的波段中近红外波段的占比很大, 结合 GA 法更适合用于筛选与甜玉米种子电导率密切相关的特征波段, 可知在电导率测定时 760~1 070 nm 波段起到了比较大的作用。

3 结 论

提出了一种基于可见-近红外(VIS-NIR)高光谱成像系统的少量波段实现对甜玉米种子电导率预测的新方法。研究结果表明, 不同光谱预处理方法对电导率预测模型的性能影

响有所差异, 在 SNV, SG1, SG2 和 MSC 中, MSC 预处理后建立的 PLSR 模型性能最优。进一步比较 SPA 和 GA 变量筛选方法对模型的影响发现, MSC-GA-PLSR 模型仅用到 25 个特征波段且模型表现较 MSC-SPA-PLSR 优秀, 其预测集的 R_p 和 RMSEP 分别为 0.973 和 0.212。本研究为采用高光谱成像技术进行甜玉米种子电导率的精确、无损和快速的预测提供一定的理论依据。

本实验的研究对象仅为一个甜玉米品种, 后续可进一步细化至对不同品种乃至同一品种不同年份、产地的甜玉米种子进行研究。本文仅初步探究了四个电导率梯度, 后续可对老化梯度进行深入细化, 研究适用于高光谱技术预测的最小电导率梯度间隔。另外本研究是以人工加速老化的甜玉米种子为材料, 其建立的模型对自然老化的种子预测效果还有待更深入的优化研究。

References

- [1] Dumont J, Hirvonen T, Heikkinen V, et al. *Computers & Electronics in Agriculture*, 2015, 116(C): 118.
- [2] PAN Bin-rong, REN Jing-yu, ZHAO Guang-wu(潘彬荣, 任镜羽, 赵光武). *Journal of Zhejiang A&F University(浙江农林大学学报)*, 2015, 32(1): 47.
- [3] ZHANG Ting-ting, SUN Qun, YANG Lei, et al(张婷婷, 孙群, 杨磊, 等). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2017, 33(21): 275.
- [4] Anisur R, Byoung-Kwan C. *Seed Science Research*, 2016, 26(4): 285.
- [5] CHEN Jing, LI Jian-ping, LI Rong, et al(陈婧, 李建平, 李荣, 等). *Acta Agriculture Boreali-Occidentalis Sinica(西北农业学报)*, 2016, 25(6): 857.
- [6] Kamruzzaman M, Elmasry G, Sun D W, et al. *Journal of Food Engineering*, 2011, 104(3): 332.
- [7] Gowen A A, O'Donnell C P, Cullen P J, et al. *Trends in Food Science & Technology*, 2007, 18(12): 590.
- [8] ZHANG Hai-liang, CHU Bing-quan, YE Qing, et al(章海亮, 楚秉泉, 叶青, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2018, 38(2): 559.
- [9] Ferrari C, Foca G, Calvini R, et al. *Chemometrics & Intelligent Laboratory Systems*, 2015, 146: 108.
- [10] Zhang R, Li C, Zhang M, et al. *Computers & Electronics in Agriculture*, 2016, 127: 260.
- [11] Cheng J H, Jin H, Xu Z, et al. *Analytical Methods*, 2017, 9(43).
- [12] Gao J, Li X, Zhu F, et al. *Computers & Electronics in Agriculture*, 2013, 99(6): 186.
- [13] Zhao Y, Zhu S, Zhang C, et al. *RSC Advances*, 2018, 8(3): 1337.
- [14] Zhang T, Wei W, Zhao B, et al. *Sensors*, 2018, 18(3): 813.
- [15] REN Li-sha, GU Ri-liang, JIA Guang-yao, et al(任利沙, 顾日良, 贾光耀, 等). *Scientia Agricultura Sinica(中国农业科学)*, 2016, 49(16): 3108.
- [16] Yang X, Hong H, You Z, et al. *Sensors*, 2015, 15(7): 15578.
- [17] SUN Jun, LU Xin-zi, ZHANG Xiao-dong, et al(孙俊, 路心资, 张晓东, 等). *Transactions of the Chinese Society for Agricultural Machinery(农业机械学报)*, 2016, 47(6): 215.
- [18] Kamruzzaman M, Elmasry G, Sun D W, et al. *Food Chemistry*, 2013, 141(1): 389.
- [19] LIU Yan-de, XIAO Huai-chun, SUN Xu-dong, et al(刘燕德, 肖怀春, 孙旭东, 等). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2018, 34(3): 180.
- [20] Nansen C, Zhao G, Dakin N, et al. *J Photochem Photobiol B*, 2015, 145: 19.
- [21] Li J, Huang W, Zhao C, et al. *Journal of Food Engineering*, 2013, 116(2): 324.
- [22] Yang Y, Sun D, Pu H, et al. *Postharvest Biology and Technology*, 2015, 103: 55.
- [23] Dai Q, Cheng J H, Sun D W, et al. *Journal of Food Engineering*, 2015, 149: 97.
- [24] Zhang R, Li C, Zhang M, et al. *Computers & Electronics in Agriculture*, 2016, 127: 260.
- [25] Khodabakhshian R, Emadi B. *International Journal of Food Properties*, 2018(2).

Determination of Conductivity in Sweet Corn Seeds with Algorithm of GA and SPA Based on Hyperspectral Imaging Technique

ZHANG Ting-ting¹, ZHAO Bin¹, YANG Li-ming², WANG Jian-hua¹, SUN Qun^{1*}

1. Department of Plant Genetics and Breeding, College of Agronomy and Biotechnology, The Innovation Center (Beijing) of Crop Seed Sciences of Ministry of Agriculture, Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing 100193, China
2. College of Science, China Agricultural University, Beijing 100083, China

Abstract The vigor of seeds plays a vital role to the agricultural development. But the low vigor and storage-tolerance seeds are common problems for sweet corn. Therefore, it has a certain practical significance to detect the sweet corn seed vigor accurately and timely. Electrical conductivity test is a traditional method of determining the vigor of seeds. However, it is a labor-intensive, time-consuming, and destructive process, which is subject to human error. Given that, this study investigated the possibility of using visible and near-infrared (VIS/NIR) hyperspectral imaging (HSI) technique to detect the electrical conductivity of sweet corn seeds. Sweet corn seeds treated by high temperature and high humidity aging were prepared as experimental materials. The visible and near-infrared hyperspectral imaging acquisition system (400~1 000 nm) was constructed to acquire the hyperspectral images of the sweet corn seeds. After HSI spectra collection, electrical conductivity tests were conducted in sweet corn seeds. The average reflectance data of the region of interest were extracted for spectral characteristics analysis. Then different pre-processing algorithms including standard normal variate (SNV), first derivative (FD), second derivative (SD), multiplicative scatter correction (MSC) were conducted to build partial least squares regression (PLSR) models of the conductivity. Lastly, the hyperspectral effective wavelengths related to conductivity of sweet corn seeds were extracted by SPA and GA for PLSR models. The results showed that the best pre-processing algorithm was MSC method. The SPA was not performing as well as GA which selected only 25 characteristic wavebands from the all 853 spectral wavebands. The PLSR model built by using MSC and GA exhibited the optimal performance with correlation coefficient of 0.976 and 0.973 for calibration set and prediction set, respectively, and root mean squared error for calibration and prediction were 0.194 and 0.212. The results indicated that combining the visible and near-infrared hyperspectral imaging technique with MSC-GA-PLSR can be used as a feasible and reliable method for the determination of conductivity in sweet corn seeds. The result can provide a theoretical foundation for rapid detection of seed conductivity using spectral information.

Keywords Hyperspectral technology; Sweet corn seed; Vigor; Characteristic wavelength; Partial least squares regression (PLSR)

(Received Jun. 27, 2018; accepted Nov. 8, 2018)

* Corresponding author