

高光谱技术无损检测单粒小麦种子生活力的特征波段筛选方法研究

张婷婷¹, 向莹莹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 100093
2. 中国农业大学理学院, 北京 100083

摘要 种子活力是种子质量的一项重要指标, 高活力的种子具有较强的抗逆性、生长优势及生产潜力。而种子活力在种子生理成熟时最高, 随后随着贮藏时间的延长而发生着自然不可逆的降低。因此, 在播种前及时、准确地对种子活力进行检测和筛选具有重要的实践意义。针对传统种子活力检测方法存在的操作过程复杂繁琐、耗时长、重复性差且对种子有破坏性等缺点, 研究尝试利用高光谱成像技术建立单粒小麦种子生活力快速、无损、精确的检测方法。以高温高湿老化后的190粒小麦种子(发芽128粒, 不发芽62粒)作为研究样本, 先利用可见-近红外(Vis-NIR)高光谱成像系统采集样本种子的光谱图像和进行标准发芽试验, 并确保光谱采集试验和标准发芽试验的小麦种子一一对应。随后提取种子光谱图像的感兴趣区域并对其光谱数据进行平均和特征分析。分别采用一阶导数(FD)、均值中心化(MC)、正交信号校正(OSC)和多元散射校正(MSC)对原始光谱数据进行预处理, 结合偏最小二乘判别分析(PLS-DA)建立全波段 PLS-DA 模型, 比较分析, 并筛选出最适预处理方法。分别利用无信息变量消除算法(UVE)、竞争性自适应重加权算法(CARS)、连续投影算法(SPA)及耦合不同变量筛选方法对特征波段进行筛选提取, 再分别基于所提取出的特征波段建立 PLS-DA 定性判别模型, 对比分析, 最终确立提取与单粒小麦种子生活力相关性最高的高光谱特征波段方法体系。结果表明: 不同光谱预处理建立的模型其表现有所差异, 在 MC, FD, OSC 和 MSC 中, 采用 MC 对原始高光谱数据进行预处理, 建立的全波段 MC-PLS-DA 判别模型, 其校正集和预测集对小麦种子生活力的整体鉴别正确率分别为 82.5% 和 83.0%, 优于原始及其他预处理后建立的全波段 PLS-DA 判别模型, 其校正集和预测集对小麦种子活种子鉴别正确率分别为 94.8% 和 90.6%。进一步对比 3 种单特征波段提取方法及其耦合分析建模中, 发现 3 种变量筛选方法耦合(UVE-CARS-SPA)的方式能够将光谱全波段的 688 个变量压缩至 8 个变量(473, 492, 811, 829, 875, 880, 947 和 969 nm), 利用所筛选出的 8 个变量建立的 MC-UVE-CARS-SPA-PLS-DA 模型获得了最优秀的鉴别效果, 其校正集和预测集对小麦种子生活力的整体鉴别正确率分别为 86.7% 和 85.1%, 较全波段模型(MC-Full-PLS-DA)分别提升了 4.2% 和 2.1%, 活种子的鉴别正确率分别为 93.8% 和 84.4%, 经过此优秀模型筛选后, 种子批最终发芽率可达到 93.1%。实验结果表明, 基于高光谱成像技术结合 UVE-CARS-SPA-PLS-DA 模型能够实现单粒小麦种子生活力的定性判别。研究工作为小麦种子活力的快速、精确且无损的检测提供理论支持。

关键词 高光谱技术; 小麦种子; 生活力; 检测; 特征波段

中图分类号: TP391.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)05-1556-07

引言

种子是农业发展的基础, 种子质量的高低直接影响着农作物的产量和品质。种子质量包括活力、净度、发芽率、纯

度等, 其中种子活力是检测种子质量的一项重要指标。而种子活力在种子生理成熟时最高, 随着贮藏时间的延长而发生着自然不可逆的种子老化, 相应的种子活力也逐渐降低^[1]。因此, 在播种前及时对种子活力进行检测和筛选意义重大。长期以来, 中国检测种子活力的方法如四唑染色实验^[2]、电

收稿日期: 2018-03-15, 修订日期: 2018-07-30

基金项目: 国家“十三五”重点研发计划(2018YFD0100904)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

导率实验^[3]和幼苗生长测定实验^[4]等,存在操作过程复杂繁琐、耗时长、重复性差且对种子有破坏性等缺点,已不能满足现代种业市场快速发展的需求^[5]。因此,亟需研究一种精确、高效且无损的种子活力检测技术。

高光谱成像技术于20世纪80年代初期兴起,是结合了光谱和图像的新一代光电无损检测技术^[6]。与机器视觉技术和近红外光谱技术相比,高光谱成像技术具有光谱分辨率高、波段多、信息量大且其信息能同时反映样品的外部形态和颜色、内部结构及化学成分等特性^[7],已在国内外很多领域被广泛研究。在农产品检测领域中,对于高光谱成像技术的研究主要集中在肉类^[8]及果蔬品质检测^[9]、种子品种检测^[10]及产地和年份的鉴别^[11]上,对其在种子活力检测的应用研究比较少。许思等^[12]分别用红外高光谱技术实现了对不同活力水平水稻种子批的分级检测。Ambrose等^[13]用400~2500 nm波段的高光谱成像技术鉴别微波热处理后及未处理的两种活力水平玉米种子批,所建立的偏最小二乘判别分析(partial least squares discriminant analysis, PLS-DA)模型校准集和预测集的鉴别率分别高达97.6%和95.6%。由上述文献可见,目前对高光谱技术在种子活力检测上的应用研究,其对象材料多为批量的种子,其结果能鉴别出一个批次种子的活力水平,却无法实现对单粒种子生活力的检测。

此外,高光谱成像技术在具有分辨率高、波段多、信息量丰富特点的同时也存在光谱信息重叠严重、无效信息和冗余波段较多的问题,这些问题会一定程度上影响后期建模效率及模型精度。因此,对高光谱信息进行特征波段的筛选是建立单粒小麦种子生活力定性反演模型的重要前提。常用的特征波段筛选方法主要有无信息变量消除算法(uninformative variables elimination algorithm, UVE)、竞争性自适应重加权算法(competitive adaptive reweighted sampling algorithm, CARS)和连续投影算法(successive projections algorithm, SPA)等。国内外学者针对不同特征波段筛选方法在定性判别模型上的运用展开了大量的研究,并获得了卓越的成果。有研究利用UVE算法剔除信息波段将全波段降低至105个波段,其偏最小二乘回归模型仍保持较高的预测精度。有报道通过CARS算法提取特征波段,将CARS与支持向量机结合建立判别模型检测番茄叶片灰霉病,正确率高达100%。Dai等^[14]以非冷冻虾和冷冻虾为研究对象,将SPA与支持向量机结合建立判别模型检测其新鲜度,模型预测集的正确率分别高达98.33%和95%。Shi等^[15]研究认为SPA算法结合偏最小二乘回归(PLSR)建立的模型,光谱波段间共线性小且模型运算效率高。可见,目前对特征波段筛选方法的应用研究,其研究对象多是单一的方法,对不同波段筛选方法的耦合研究尚处于探索阶段,特别是利用耦合波段筛选方法对单粒种子生活力进行定性判别的研究更是鲜见报道。

本工作以小麦种子为实验材料,通过高光谱成像系统采集不同生活力的单粒小麦种子光谱信息,结合多种预处理方法利用全波段光谱信息建立PLS-DA模型,筛选出最优预处理方法,并用UVE, CARS和SPA三种变量提取方法及其耦合分析筛选出与小麦种子生活力相关的特征波段,探讨

PLS-DA建模方法结合不同波段提取方法对鉴别单粒小麦种子生活力正确率的影响,为研究精确、无损且便捷的小麦种子活力检测技术提供理论依据。

1 实验部分

1.1 材料

试验材料为鲁原502小麦种子,于2016年购于山东鲁科农业科技有限公司。选取其中190粒结构完整且无霉变的种子,含水率为11.7%。

1.2 人工加速老化

实验采用高温高湿人工加速老化后的小麦种子模拟贮藏过程中自然老化的小麦种子^[16]。随机选取190粒外观一致且无破损的小麦种子,放入1个灭菌后的尼龙网里,密封于盛有少量无菌水的干燥器,放置于41℃的干燥箱中,5d后取出,回干至初始含水率进行高光谱扫描和标准发芽试验。

1.3 高光谱成像仪和分析软件

实验采用的高光谱成像系统主要包括:线扫描成像光谱仪(Imspector V10E; Spectral Imaging Ltd., Oulu, Finland), CCD相机(EM285CL; Raptor Photonics., Ireland), 变焦镜头(OLE23; Schneider, Germany), 150W卤素灯(IT 3900e; Illumination Technologies Inc., New York, USA)、控制箱、电控位移台(IRCP0076-1COMB; Isuzu Optics Corp., Taiwan, China)、计算机和暗箱等。该系统可采集的光谱范围为400~1000 nm,图像分辨率为1004×1662像素。采集光谱信息的软件为Spectral Image software (Isuzu Optics Corp., Taiwan, China)。

采集高光谱图像前,先进行黑板标定,获取校正图像,然后设置电控平台的移动速度为1.3 mm·s⁻¹、相机曝光时间为15 ms。对样品进行光谱信息采集时,将每一粒小麦种子背面(胚面)朝上,股沟面(胚乳面)朝下,摆放在电控位移台上,随着平台移动,摄像头扫描整个平台。高光谱图像采集过程在暗箱内完成。

1.4 标准发芽试验

高光谱采集190粒小麦种子的光谱信息后,立刻进行标准发芽试验。每一粒小麦种子浸于1%次氯酸钠溶液中消毒8 min后,无菌水分别冲洗3次,(25±1)℃环境下卷纸发芽^[4],第8天检查每粒种子发芽情况。当种子的胚芽和胚根长度都超过2 mm时,认定为发芽记为“1”,否则认为不发芽记为“2”^[17],并确保光谱采集试验和标准发芽试验的小麦种子一一对应。

2 高光谱图像处理及数据分析

在提取光谱信息前,对采集的高光谱图像按照式(1)进行校正。其中, I 为校正后的高光谱图像; I_0 为原始高光谱图像; B 为黑板文件图像(反射率接近0%); W 为白板文件图像(反射率接近100%)。

$$I = \frac{I_0 - B}{W - B} \quad (1)$$

黑白板校正后,需将单粒小麦种子从高光谱图像背景中分离出来。小麦种子图像分割处理过程如图 1。随机选取种子区域和背景区域,分别计算其平均光谱反射率,确定反射率差值最大的波段,得到便于设定阈值的图像^[7]。通过设定阈值将背景和小麦种子清晰地分割开,再运用形态学滤波及

掩膜处理获得感兴趣区域(region of interest, ROI)^[18]。因首尾波段区域噪声干扰较大,仅提取每粒小麦种子 430~970 nm 范围内共 688 个波段的反射光谱用于后续建模分析。所有 190 粒样品种子的反射光谱如图 2 所示。

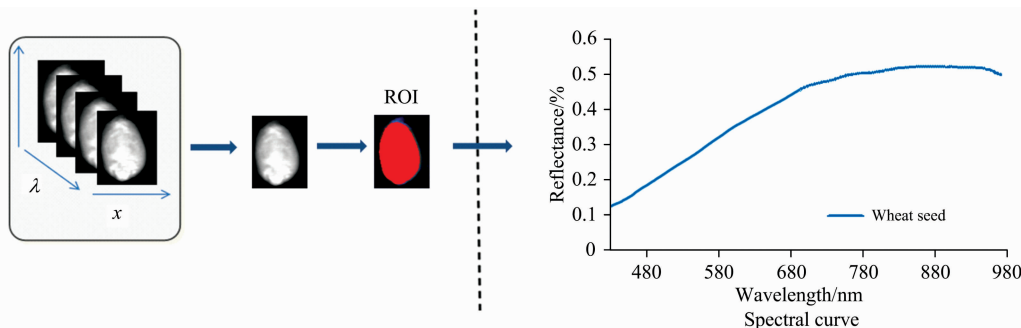


图 1 小麦种子的 ROI 高光谱图像及反射光谱曲线

(a): 背景分隔; (b): 光谱曲线

Fig. 1 Hyperspectral image of ROI and reflectance curve of a wheat seed

(a): Background segmentation; (b): Spectral curve

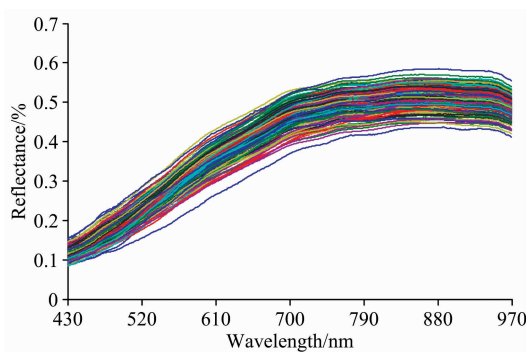


图 2 190 粒样品种子的原始反射光谱

Fig. 2 Raw reflectance spectra of 190 samples

标准发芽试验测定样品种子批发芽率为 67.4%,其中发芽种子为 128 粒,不发芽种子为 62 粒。采用随机样本生成法,随机选取其中 96 粒发芽种子和 47 粒不发芽种子构建校正集,余下 32 粒发芽种子和 15 粒不发芽种子构建预测集。研究中的光谱图像处理软件为 ENVI4.7 (Research System Inc, Boulder, Co., USA),数据分析软件为 Matlab R2014a (The Math Works, Natick, USA)。

分类模型的性能以校正集及预测集的整体鉴别正确率来评判,另外涉及到活种子鉴别正确率和最终发芽率,均按照如式(2)~式(4)计算。

$$\text{整体鉴别正确率} = \frac{G + NG}{G + NG + Gr + NG} \quad (2)$$

$$\text{活种子鉴别正确率} = \frac{G}{G + Gr} \quad (3)$$

$$\text{最终发芽率} = \frac{G}{G + NGr} \quad (4)$$

其中, G 为模型判断正确的发芽种子数量; NG 为模型判断正确的不发芽种子数量; Gr 为模型将发芽种子误判为不发芽种子的数量; NGr 为模型将不发芽种子误判为发芽种子的

数量。

3 结果与讨论

3.1 基于预处理的全波段光谱 PLS-DA 建模分析

尝试了几种常见的预处理方法如一阶求导(first derivative, FD)、均值中心化(mean centering, MC)、正交信号校正(orthogonal signal correction, OSC)及多元散射校正(multiplicative scatter correction, MSC)对原始光谱进行处理,并分别建立全光谱 PLS-DA 模型,结果见表 1。经过不同预处理后的光谱建模结果有所差异,其中 MC 对原始光谱处理后的 PLS-DA 模型表现最优,其校正集和预测集的整体鉴别正确率分别为 83.0% 和 82.5%,活种子鉴别正确率分别为 94.8% 和 90.6%。

表 1 基于不同预处理方法的 PLS-DA 模型判别正确率 (%)
Table 1 Classification accuracy (%) with PLS-DA model of raw spectrum after different pre-processing methods

Pre-processing methods	Calibration sets		Prediction sets	
	Overall	Viable	Overall	Viable
RAW	79.0	79.2	78.7	78.1
FD	88.8	91.7	76.6	75.0
MC	82.5	94.8	83.0	90.6
OSC	80.4	93.8	76.6	90.6
MSC	72.0	90.6	61.7	65.6

3.2 基于单个特征波段筛选方法的优选波段 PLS-DA 建模分析

由高光谱全波段 PLS-DA 建模结果可知,全波段光谱对单粒小麦种子生活力具有较好的判别效果,但由于其数据量庞大,含有较多的冗余信息和共线性变量,会影响模型判别

的准确率^[19]，因此将 MC 预处理后的的光谱信息进行特征波段筛选。

3.2.1 UVE 变量筛选法

基于 UVE 法的小麦种子生活力高光谱特征波段筛选结果如图 3。蓝色竖实线为变量分隔线，黄色曲线是 688 个波段变量的稳定性 C 值分布情况，红色曲线是 UVE 产生的随机噪声变量的稳定性 C 值分布情况。两条水平虚线是波段选择的阈值范围，阈值线之间的波段被当做无效信息予以剔除，余下的变量为被选出的有用信息波段。由图 3 可见，UVE 法选出的特征波段主要集中在可见光区域的 400~500 nm 及近红外区域的 760~970 nm，说明这两个波段范围内包含有更多与小麦种子生活力有关的关键光谱信息。经过 UVE 法筛选后，剔除了 585 个波段，优选出 103 个波段，特征波段个数占全波段的 15.0%。

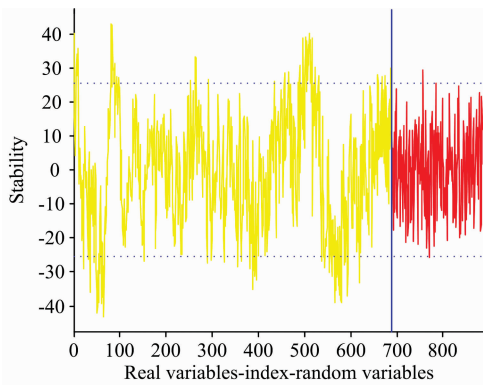


图 3 基于 UVE 法筛选特征波段

Fig. 3 Characteristic wavebands selected by UVE algorithm

3.2.2 CARS 变量筛选法

基于 CARS 法的小麦种子生活力高光谱特征波段筛选结果如图 4。由于 CARS 中蒙特卡洛采样的随机性，导致每次 CARS 法运行结果不一致，为选取相对较优的波段组合，图 4 为经过 30 次 CARS 选择后的最优结果。图 4(a) 可见，在指数衰减函数 EDP 作用下，刚开始几次采样，波段数快速降低，但随着采样次数的增加，波段数减小的速度逐渐变慢。图 4(b) 所示，随着采样次数增加，RMSECV 值整体呈现先减小再增大的趋势，当采样次数达 29 次时 RMSECV 值最低，即在 1~29 次的波段筛选过程中剔除了与小麦种子生活力无关的信息，29 次之后，RMSECV 值逐渐升高，说明一些与种子生活力有关的关键信息被剔除了。结合图 4(c) 可知，第 29 次采样时 (* 表示)，RMSECV 值最小，即此次采样所选的波段子集为预测小麦种子生活力的关键波段变量。因此，经 CARS 法筛选后，剔除了 664 个波段，优选出 24 个波段，筛出的波段个数占全波段的 3.5%。

3.2.3 SPA 变量筛选法

基于 SPA 法的小麦种子生活力高光谱特征波段筛选结果如图 5。SPA 法在优选波段时，会将不同波段子集逐一建立多元线性回归模型，并计算 RMSE 值，其中最小的 RMSE 值对应的子集即为最优波段子集。图 5 可见，RMSE 值随波段数量的增加，整体呈下降的趋势。当波段数达到 20 时，

RMSECV 值最小为 0.306，随后略有升高。经 SPA 法筛选后，剔除了 668 个波段，优选出 20 个关键波段，特征波段数占比仅为 2.9%。

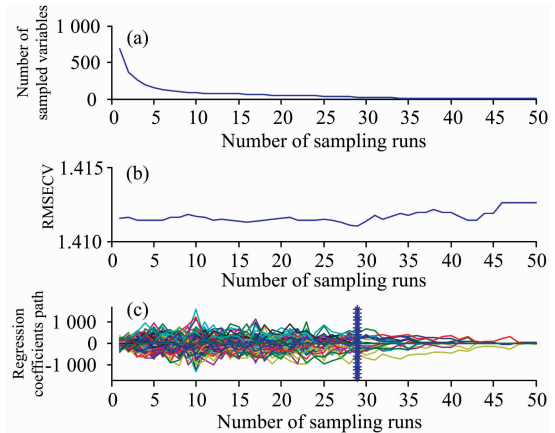


图 4 基于 CARS 法筛选特征波段

(a): 波长变量数变化; (b): 交叉验证均方根误差变化; (c): 变量回归系数趋势

Fig. 4 Characteristic wavebands selected by CARS algorithm

(a): Variation of wavelenath variables number; (b): Variation of RMSECV; (c): Trend of variable regression coefficients

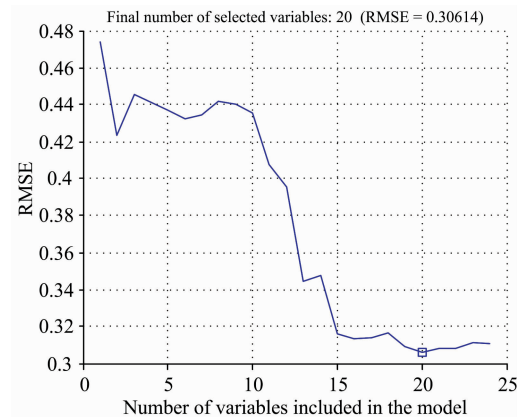


图 5 基于 SPA 法筛选特征波段

Fig. 5 Characteristic wavebands selected by SPA algorithm

3.3 PLS-DA 模型建立与比较

基于不同变量筛选方法建立的 PLS-DA 模型对单粒小麦种子生活力鉴别正确率结果见表 2。可见，在单个特征波段筛选方法分析中，UVE-PLS-DA 模型效果较全波段 PLS-D，SPA-PLS-DA 和 CARS-PLS-DA 的模型优秀，其预测集整体鉴别正确率为 83.0%，此模型筛选后，种子批最终发芽率可达到 96.2%，较全波段 PLS-DA 模型提升了 8.7%，可以较好地实现对单粒小麦种子生活力的判别及对小麦种子批发芽率的提升。在波段筛选方法耦合分析中，UVE-CARS-SPA-PLS-DA 模型的表现最优，其校正集和预测集的整体鉴别正确率分别为 86.7% 和 85.1%，经过此模型筛选后，种子批最终发芽率可达到 93.1%。结合单个特征波段提取方法中的

UVE-PLS-DA 模型比较可见, 虽然 UVE-PLS-DA 模型的最终发芽率(96.2%)高于 UVE-CARS-SPA-PLS-DA 模型, 但其对活种子的鉴别率较低(78.1%), 意味着 UVE-PLS-DA 模型在提升发芽率的同时会损失大量活种子。另外 UVE-PLS-DA 模型需用到 103 个特征波段而 UVE-CARS-SPA-PLS-DA 模型仅用到 8 个特征波段。

最佳波段筛选方法 UVE-CARS-SPA 优选出了 8 个特征波长, 分别是 473, 492, 811, 829, 875, 880, 947 和 969 nm, 所选波段集中在可见光区域 400~500 nm 及近红外区域 800~970 nm 范围内。前人研究表明 400~500 nm 处存在多条色素吸收带, 如类胡萝卜素的吸收带在 448 和 471 nm 处, 叶绿素 a 在 430 nm 存在吸收带, 叶绿素 b 在 448 nm 附近存在吸收带^[7]。另外, 在 970 nm 的近红外区域的小麦种子生活

力光谱曲线的光谱特性主要是由于 O—H 键二级倍频和 C—H 键三级倍频振动吸收而引起的。这些可能由于每一粒种子的种皮色素含量有差异且老化会显著影响种胚中的油分、水分及其他化合物, 而种皮色素和化学成分的差异可以在较大程度上反映每一粒种子的生活力情况。

上述结果充分说明, 波段提取方法耦合分析可以较好地提升模型的性能。UVE-CARS-SPA 方法是一种优秀的特征波段提取方法, 能够有效地剔除无用光谱信息, 筛选出与小麦种子生活力密切相关的特征波段, 波段数量仅占全波段的 1.2%, 大大地压缩了建模时间, 其 MC-UVE-CARS-SPA-PLS-DA 模型能够较准确地识别单粒小麦种子的生活力, 提升种子批的发芽率。

表 2 基于不同变量筛选方法的 PLS-DA 模型判别正确率 (%)

Table 2 Classification accuracy (%) with PLS-DA model based on different variable selection methods

Selection methods	Variable number	Calibration sets		Prediction sets		
		Overall	Viable	Overall	Viable	Final germination
Full	688	82.5	94.8	83.0	90.6	87.5
UVE	103	84.6	86.5	83.0	78.1	96.2
CARS	24	83.2	84.4	80.9	87.5	84.8
SPA	20	80.4	91.7	74.5	87.5	77.8
CARS-SPA	10	83.2	86.5	80.9	84.4	87.1
UVE-CARS	12	84.6	89.6	85.1	90.6	87.9
UVE-SPA	15	84.6	91.7	83.0	78.1	96.2
UVE-CARS-SPA	8	86.7	93.8	85.1	84.4	93.1

4 结 论

以单粒小麦种子为研究对象, 利用高光谱成像系统获取样品的光谱信息, 利用多种预处理方法建立光谱与单粒小麦种子生活力的 PLS-DA 模型, 随后利用不同特征波段提取方法 UVE、CARS、SPA 及其耦合分析筛选出与生活力相关的特征波段, 并进一步建模比较。结果表明, 不同光谱预处理建立的模型其性能有所差异, 在 MC, FD, OSC 和 MSC 中, MC 预处理方法建立的 PLS-DA 模型表现最优, 其校正集和预测集的整体鉴别正确率分别为 82.5% 和 83.0%。进一步对比 3 种单特征波段提取方法及其耦合分析建模中, 发现 UVE-CARS-SPA 三种波段筛选方法的耦合所提取的特征波段能较好地反映每一粒小麦种子的生活力信息, 这不仅有效地提高了模型的判别精度, 还在较大的程度上压缩了建模所

需波段数量, 简化了模型且提升了效率。最终 MC-UVE-CARS-SPA-PLS-DA 模型仅利用 8 个特征波段便获得了优秀的鉴别效果, 其校正集和预测集的整体鉴别正确率分别为 86.7% 和 85.1%, 经过此模型筛选后, 种子批最终发芽率可达到 93.1%。本研究为采用高光谱成像技术鉴别单粒小麦种子生活力及群体小麦种子活力的研究提供了理论依据。

高光谱技术实际应用于种子活力检测仍存在一些难题, 如不同品种的小麦种子甚至同一品种不同生产年份和产地的小麦种子, 其化学成分必然会有差异并会对种子光谱反射率产生影响, 相应的特征波段也会发生变化, 遂在检测种子活力时常需要重复建模, 工作量较大。另外, 本研究及前人的研究多以人工老化种子为材料, 其研究成果应用于自然老化的种子效果并不理想。如何进一步减小和消除这些外部因素对模型鲁棒性的干扰, 仍需要更精细、深入的优化研究。

References

- [1] Istrate Dan, Singiorzan Laura. Turkish Journal of Agriculture and Forestry, 2014, 36(4): 421.
- [2] Olesen M H, Nikneshan P, Shrestha S, et al. Sensors, 2015, 15(2): 4592.
- [3] Anisur R, Byoung-Kwan C. Seed Science Research, 2016, 26(4): 285.
- [4] 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.
- [5] SONG Le, WANG Qi, WANG Chun-yang, et al(宋乐, 王琦, 王纯阳, 等). Grain Storage(粮食储藏), 2015, 44(1): 20.

- [6] Yang X, Hong H, You Z, et al. *Sensors*, 2015, 15(7): 15578.
- [7] Zhang T T, Wei W S, Zhao B, et al. *Sensors*, 2018, 18(3): 813.
- [8] Tao F, Peng Y. *Journal of Food Engineering*, 2014, 126(1): 98.
- [9] Zhang B, Fan S, Li J, et al. *Food Analytical Methods*, 2015, 8(8): 2075.
- [10] Wang L, Liu D, Pu H, et al. *Food Analytical Methods*, 2015, 8(2): 515.
- [11] Gao J, Li X, Zhu F, et al. *Computers and Electronics in Agriculture*, 2013, 99(6): 186.
- [12] XU Si, ZHAO Guang-wu, DENG Fei, et al(许思, 赵光武, 邓飞, 等). *Seed(种子)*, 2016, 35(4): 34.
- [13] Ambrose A, Kandpal L M, Kim M S, et al. *Infrared Physics and Technology*, 2016, 75: 173.
- [14] Dai Q, Cheng J H, Sun D W, et al. *Journal of Food Engineering*, 2015, 149: 97.
- [15] Shi T, Chen Y, Liu H, et al. *Applied Spectroscopy*, 2014, 68(8): 831.
- [16] Balešević-Tubić S, Tatić M, Dorđević V, et al. *Romanian Biotechnological Letters*, 2011, 16(6): 6740.
- [17] Nansen C, Zhao G, Dakin N, et al. *Journal of Photochemistry and Photobiology B Biology*, 2015, 145: 19.
- [18] SUN Jun, LU Xin-zi, ZHANG Xiao-dong, et al(孙俊, 路心资, 张晓东, 等). *Transactions of the Chinese Society of Agricultural Machinery(农业机械学报)*, 2016, 47(6): 215.
- [19] WEI Wen-song, PENG Yan-kun, ZHENG Xiao-chun, et al(魏文松, 彭彦昆, 郑晓春, 等). *Acta Optica Sinica(光学学报)*, 2017, 37(11): 374.

Wavelength Variable Selection Methods for Non-Destructive Detection of the Viability of Single Wheat Kernel Based on Hyperspectral Imaging

ZHANG Ting-ting¹, XIANG Ying-ying¹, 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 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 Seeds are the basis of the agricultural industry. The viability of seeds is a very important index of seed quality, which is closely related to resistance to biotic and abiotic stress, germination percentage, plant performance, and which decreases with increasing storage period. Increased understanding of wheat seed viability would be beneficial to the wheat industry by ensuring a higher yield for farmers and reducing crop variability. Seed companies would also benefit from enhanced viability by being able to ensure a higher quality product. As the viability of seeds was gradually brought to the public attention, the rapid detection of seed viability without destroying has been a research hot spot. This study aimed at investigating the possibility of using visible and near-infrared (VIS/NIR) hyperspectral imaging (HSI) technique to discriminate viable and nonviable wheat seeds. Firstly, 190 wheat seeds treated by high temperature and high humidity aging (128 germination samples and 62 non-germination samples) 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 wheat seeds. After HSI spectra collection of the wheat seeds, a germination test was implemented to check for seed viability. We recorded a seed as germinated (yes=1) if the plumule and radicle were both over 2 mm long, and non-germinated (no=2) if not. The average reflectance data of the region of interest were extracted for spectral characteristics analysis. Secondly, different pre-processing algorithms including the first derivative (FD), orthogonal signal correction (OSC), multiplicative scatter correction (MSC), mean centering (MC) were conducted to build partial least squares discriminant analysis (PLS-DA) model of the viability of wheat seeds. Lastly, three variable selection methods including the uninformative variables elimination (UVE), competitive adaptive reweighted sampling (CARS) and successive projections algorithm (SPA) were used to screen the characteristic wavelengths related to seed viability. PLS-DA models were established by these characteristic wavelengths. The results showed that, the classification accuracies of different pre-processing algorithms were diverse. Among them, the MC method was the best pre-processing algorithm, from which the overall classification accuracy were 82.5% and 83.0%, and the viability classification accuracy were 94.8% and 90.6%, in calibration and prediction sets, respectively. Among the single variable selection methods, UVE method was superior to other two variable selection methods while maintaining an excellent performance of the model for overall classification accuracy (84.6%, 83.0%) and viability classification accuracy (86.5%, 78.1%) in the calibration and prediction sets. This model could promote the germination percentage of the seed lot from 67.4% to 96.2%. Comparing all variable selection methods comprehensively, the UVE-

CARS-SPA method selected only 8 variables (473, 492, 811, 829, 875, 880, 947 and 969 nm) from the all 688 spectral variables. The PLS-DA model built by using UVE-CARS-SPA method exhibited the optimal performance with overall accuracy of 86.7% and 85.1% for calibration and prediction, respectively, and accuracy for viable seed was 93.8% and 84.4%. After screening by this model, the germination percentage of the seed lot enhanced from 67.4% to 93.1%. The results indicated that appropriate variable selection could improve the performance of a model, simplify the classification models, and increase the classification accuracy of viable and nonviable wheat seeds. In the future, combining the visible and near-infrared hyperspectral imaging technique with MC-UVE-CARS-SPA-PLS-DA can be used as a feasible and reliable method for the determination of seed viability during the storage. The result can provide the theoretical reference for rapid detection of seed viability during grain storage using spectral information.

Keywords Hyperspectral technology; Wheat seed; Viability; Detection; Characteristic wavelength

(Received Mar. 15, 2018; accepted Jul. 30, 2018)

* Corresponding author