

小麦种子自然老化程度的近红外光谱无损识别

吴静珠¹, 李 慧¹, 张鹤冬¹, 毛文华^{2*}, 刘翠玲¹, 孙晓荣¹

1. 北京工商大学, 食品安全大数据技术北京市重点实验室, 北京 100048

2. 中国农业机械化科学研究院, 北京 100083

摘 要 应用近红外光谱技术无损分析小麦种子短期自然老化过程中主要化学成分的变化趋势, 并结合支持向量机建立快速判别小麦种子自然老化程度的分析模型。本实验应用 VERTEX 70 傅里叶变换红外光谱仪, 以大样品杯旋转采样方式跟踪采集了 45 份小麦种子在自然老化初期、4 个月、7 个月、9 个月的近红外光谱。标准差可以用来表征数据离散程度, 因此本实验通过计算每份样本在 4 个自然老化阶段的光谱标准差来筛选与自然老化时间显著相关的谱区。为避免单个样本由于偶然因素导致的离散度值异常, 实验统计了 45 份样本的光谱标准差均值, 根据均值光谱得到如下谱峰: 8 362, 6 950, 7 563, 5 319, 4 998 和 4 478 cm^{-1} 处。解析谱峰所在区域对应的化学基团归属可得: 6 950 cm^{-1} 处对应的是液态水中 O—H 伸缩振动的一级倍频且该处离散度值较大, 因此小麦种子在短期自然老化阶段中水分变化较为显著; 5 319, 4 998 和 4 478 cm^{-1} 处离散度值较 6 950 cm^{-1} 处小, 对应的是蛋白质仲酰胺、伯酰胺和酰胺的合频和倍频信息, 因此蛋白质变化较水分而言相对平缓; 8 362 和 7 563 cm^{-1} 处反映的主要是 C—H 振动的二级倍频信息且离散度值较大, 而种子中蛋白质、淀粉等均具有 C—H 官能团, 因此蛋白和淀粉等成分综合变化较为显著。在上述分析基础上, 本文采用多分类支持向量机结合近红外光谱建立快速识别小麦种子四种自然老化程度的定性模型。将 180 份样本光谱按照 3 : 1 随机抽取 135 个样本作为训练集, 其余样本作为测试集。选择核函数为径向基函数, 通过网格搜索法进行参数寻优得到惩罚参数为 8, 核参数为 0.008 974 2 时, 训练集和测试集的识别正确率可达 99.26% 和 99.78%。实验结果表明: 近红外光谱技术结合支持向量机可快速判别小麦种子短期自然老化程度, 为种子贮藏过程中生理特性变化的无损监测及开发利用提供便捷的检测手段。

关键词 小麦种子; 自然老化; 近红外光谱; 支持向量机; 标准差

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)03-0751-05

引 言

小麦是我国主要的粮食作物和重要的战略储备粮品种, 在粮食的生产, 流通和消费中具有重要地位。新收获的小麦经过后熟期后, 随着存储时间的延长会发生不同程度的老化, 导致种用品质和食用品质劣化, 既影响了农业生产又影响了食用价值。老化是小麦种子在贮藏过程中普遍存在的一种自然现象, 不但影响种子的萌发、幼苗生长及后期种子的质量与品质, 而且对种质资源的保存、开发和利用都产生严重的影响^[1]。因此, 监测小麦种子老化程度, 解析老化过程中种子的生理特性变化, 对于小麦种子贮藏及开发利用是至

关重要的。

现行小麦种子老化检测主要根据小麦种子品质性状的变化^[2], 如感官品质、过氧化氢酶活动度、降落数值、脂肪酸值、发芽率等指标判定。上述指标均能在一定程度上反映出小麦品质劣变程度与储藏时间的相关性, 但在实际检测过程中存在操作繁琐、耗时、且具有破坏试样等缺点, 难以适应现代农业生产发展提出的快速、无损、便捷的检测需求。

近年来, 近红外光谱技术 (near infrared spectroscopy, NIR)^[3-4] 以其快速、多组分、非破坏性等技术优势在种子质量快检领域崭露头角^[5-7]。Ashabahebwa Ambrose 等^[8] 实验表明近红外技术在无损判别玉米种子生活力领域较拉曼技术更具优势。有研究以白菜、芥菜、菜心等常见芸苔属种子为

收稿日期: 2018-06-19, 修订日期: 2018-10-25

基金项目: 国家重点研发计划 (2018YFD0101004), 北京市自然科学基金面上项目 (4182017), 科技创新服务能力建设-基本科研业务费 (PXM2018_014213_000033) 资助

作者简介: 吴静珠, 1979 年生, 北京工商大学副教授 e-mail: pubwu@163.com

* 通讯联系人 e-mail: mwh-924@163.com

实验材料,检测经人工老化后种子的发芽势、发芽率和活力指数等指标,探究种子老化机理,并采用 NIR 技术较好地预测了人工老化后种子的活力水平。Song 等^[9]利用 NIR 技术准确鉴别了经 γ 射线照射的人工老化处理后的水稻种子活力水平。目前应用 NIR 技术鉴别种子活力的研究报道中,采用的都是人工加速老化替代自然老化来节省实验时间。但是人工老化和自然老化两种老化形成的环境有着显著的差异,对种子内部的生理生化活动产生的影响也有着显著差异。有研究报道^[10]经人工/自然老化后的种子在苗期后的生长发育特性存在本质上的差异,NIR 技术能否有效鉴别和区分自然老化种子的活力水平还有待进一步验证。

因此,本工作拟通过不同老化阶段的小麦种子的近红外光谱来解析种子主要成分的变化,探索采用近红外光谱技术判别小麦种子自然老化程度的可行性,为进一步采用近红外技术无损解析种子自然老化过程中生理生化指标变化规律及活力下降主因提供理论研究和支撑。

1 实验部分

1.1 样本制备

实验收集的 45 份小麦样本由中国农业科学院作物科学研究所提供(2016 年收获)。将小麦样本装在纱网中存放于室温条件下进行自然老化,分别在老化初期(2016 年 11 月)、4 个月(2017 年 3 月)、7 个月(2017 年 6 月)和 9 个月(2017 年 8 月)取样。

1.2 近红外光谱数据采集

采用德国布鲁克公司的 VERTEX 70 傅里叶变换红外光谱仪(大样品杯旋转采样),采集该批小麦样本在老化初期、4 个月、7 个月和 9 个月的近红外光谱,共计 180 份。装样前仔细筛查剔除夹杂物和空粒。光谱仪参数设定如下:范围为 $4\ 000\sim 10\ 000\ \text{cm}^{-1}$,分辨率为 $8\ \text{cm}^{-1}$,采样点数为 1 557,每个样品扫描 64 次后取平均值。

1.3 数据处理

支持向量机^[11-12]是建立在统计学习理论 VC 维理论和结构风险最小化原理基础上的机器学习方法,能够较好地解决小样本、非线性和高维数的识别问题,具有很好的泛化能力。本实验采用的支持向量机算法参考中国台湾大学林智仁(Lin Chih-Jen)教授等开发的 LIBSVM 工具包(www.csie.ntu.edu.tw/~cjlin/libsvm/)实现,其他计算程序自行编写,在 Matlab2014a 环境运行。

2 结果与讨论

2.1 近红外光谱数据预处理

4 个老化时间节点采集的 180 份小麦样本近红外光谱如图 1(a)所示。可以观察到 4 类小麦样本原始光谱吸光度趋势相似,但是由于样品颗粒大小、光散射和光程变化等因素的影响,样品光谱的基线漂移严重。为消除样品状态、仪器状态等对光谱分析的干扰,采用标准正态化处理(standard normal variate, SNV)对原始光谱进行预处理,如图 1(b)所示。

从图 1(b)可以看出经过 SNV 预处理后光谱谱峰变化更为清晰,基线偏移量得到明显修正。但是仪器噪声在光谱两端表现较为明显,尤其是在 $9\ 000\sim 10\ 000\ \text{cm}^{-1}$ 区间,光谱毛刺较多。

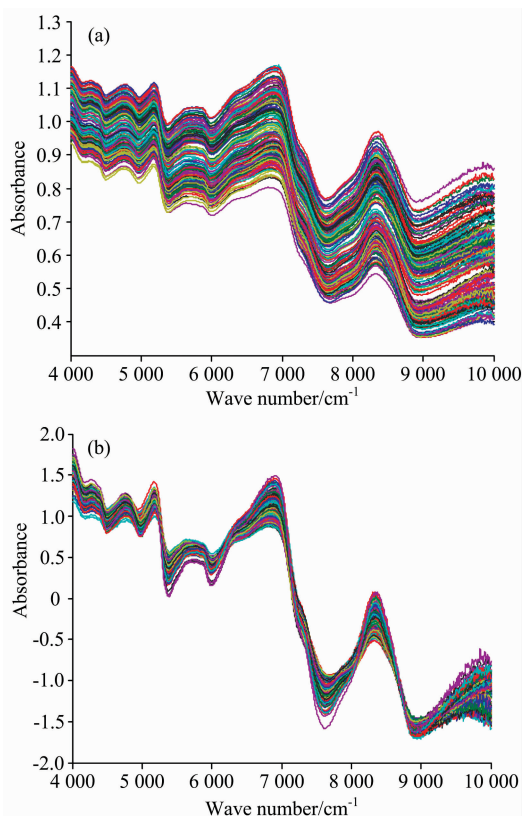


图 1 小麦种子近红外光谱预处理

(a): 原始光谱; (b): SNV 预处理后的光谱

Fig. 1 Preprocessing results of wheat NIR spectra

(a): Original spectra; (b): Spectra preprocessed by SNV

2.2 小麦种子自然老化过程光谱定性解析

小麦种子的水分、粗蛋白和淀粉等主要成分的分子结构中都存在含氢基团,因此小麦种子具有丰富的近红外光谱信息。为了明确每个样本在各波长点处的吸光度随老化阶段的具体变化情况,实验首先统计了每个样本在 4 个老化阶段随波长变化的标准差。标准差是用来表征数据离散程度的统计量,标准差越大,数据的离散程度越大,该波长点处吸光度波动明显。因此根据离散度大小就可以筛选出与自然老化时间显著相关的谱区。为避免单个样本由于偶然因素导致的离散度值异常,实验统计了 45 个样本随波长变化的光谱离散度均值,并归一化至 $[-1, 1]$ 范围内,如图 2 所示。由图 2 可以观察到按照离散度大小排序,谱峰主要分布以 $8\ 362$, $6\ 950$, $7\ 563$, $5\ 319$, $4\ 998$ 和 $4\ 478\ \text{cm}^{-1}$ 为中心的区域。在 $10\ 000\ \text{cm}^{-1}$ 处离散度最大,但其处在仪器噪声较大的区域,为避免仪器噪声对分析结果准确性的影响,因此不对该谱区分析。

$6\ 950\ \text{cm}^{-1}$ 附近表征液态水中 O—H 伸缩振动的一级倍频($6\ 944\ \text{cm}^{-1}$)信息^[15],因此该特征谱区反映出种子在自然

老化阶段中水分的显著变化。分析其原因：种子样本放置于室内开放环境下进行自然老化，随着存储时间延长，小麦逐渐干燥，水分减少，导致其近红外光谱在该区间出现了明显波动。

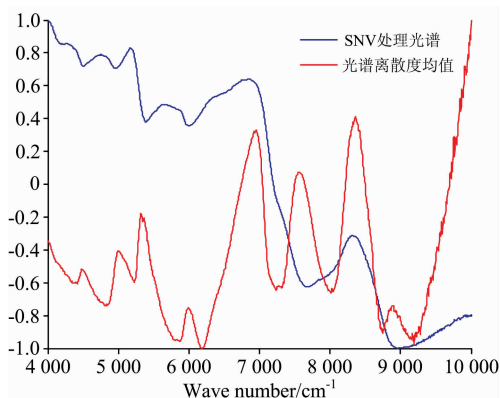


图 2 小麦种子不同老化阶段的光谱离散度均值曲线

Fig. 2 The spectra dispersion average curve of the aging wheat seeds

5 319, 4 998 和 4 478 cm^{-1} 附近区域主要反映的是蛋白质信息。蛋白质仲酰胺 CONH 中 C=O 伸缩振动的二级倍频吸收位于 5 208 cm^{-1} 附近，蛋白质伯酰胺 CONH₂ 中 N—H 伸缩振动与酰胺 II 谱带合频吸收位于 5 051 cm^{-1} 附近；蛋白质中 N—H 伸缩振动和酰胺 I 谱带的组合频、蛋白质中 N—H 弯曲振动的二级倍频吸收位于 4 854 cm^{-1} 附近；蛋白质中 C=O 伸缩振动与酰胺 III 谱带的合频吸收位于 4 587 cm^{-1} 附近等^[13]。因此该特征谱区显著反映了种子在自然老化阶段中蛋白质变化。

8 362 和 7 563 cm^{-1} 处反映的主要是 C—H 振动的二级倍频信息^[13]。种子中的蛋白质、淀粉等均具有 C—H 官能团，且该离散度值较大，因此可能是种子老化过程中蛋白质、淀粉等共同作用导致近红外光谱在该区间出现了显著波动。分析其原因：小麦种子中的蛋白质在贮藏不当或老化过程中会变性，亲水能力以及蛋白质分子间的凝聚力都会有所降低，在籽粒吸湿回潮后还会水解成游离氨基酸，导致其近红外光谱在该区间出现了明显波动；淀粉则会因水解酶的作用产生可溶性糖，可溶性糖分不断上升，因此小麦种子的近红外光谱在该区间出现了显著的波动。

综上解析上述波动显著谱区离散度值的大小以及对应的化学基团归属可得，小麦种子在短期自然老化阶段中水分波动变化较为明显，蛋白质较水分变化相对平缓，但是蛋白质和淀粉等成分的综合变化较为强烈。

2.3 小麦种子自然老化程度近红外无损判别

通过上述小麦种子在自然老化阶段的近红外光谱解析可得，小麦种子在老化过程中水分、蛋白质、淀粉等化学成分的变化可以显著反映在相应近红外特征谱区，可为应用近红外技术识别小麦种子的自然老化程度提供切实可行的理论基础。

支持向量机在小样本、高维度的模式识别分析中具有显

著的优势，因此本实验采用 4 分类 SVM 建立无损判别小麦种子 4 种自然老化程度的近红外模型。核函数和参数的选择是构造具有良好性能的 SVM 定性识别模型的关键要素。本实验采用径向基函数(radial basis function, RBF)作为核函数，参数(C, γ)是影响 SVM 性能的关键因素^[12]，其中 C 为惩罚因子， γ 为核参数。将 180 份样本光谱按照 3 : 1 的比例，随机抽取 135 个样本作为训练集，其余样本作为测试集，采用网格法搜索法对 c 和 γ 进行参数寻优，参数 c 和 γ 的搜索范围分别设定为 $[2^{-5}, 2^{10}]$ 、 $[2^{-15}, 2^0]$ 。对每组(C, γ)建立的模型用 3 折交叉验证方法得到训练集验证分类准确率，最终选定使训练集验证分类准确率最高的参数组合作为最优值。得到的最优参数分别为 $c=8$, $\gamma=0.008\ 974\ 2$ ，支持向量数为 51，利用最优参数建立小麦自然老化程度的 SVM 识别模型对测试集样本进行验证的结果如表 1 所示。

表 1 SVM 分类模型识别结果

Table 1 Recognition results of SVM classification model

样本类别	训练集识别率/%	测试集			
		实际数量	识别率/%	误识数	误识类型
原始样本	100	11	90.91	1	老化 4 个月
老化 4 个月	100	13	100	0	
老化 6 个月	100	9	100	0	
老化 8 个月	100	12	100	0	
4 类总体	99.26	45	97.78	0	

从表 1 可以看出，SVM 分类识别模型对于 4 类不同老化时间的小麦样本识别效果较好，总体识别率达到 99.26%。测试集样本实际类别和预测类别关系图如图 3 所示。

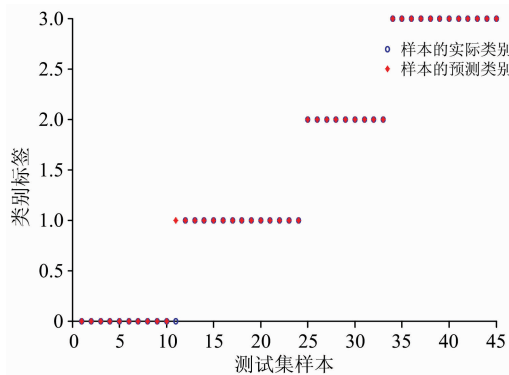


图 3 测试集样本分类结果

Fig. 3 The classification result of the test set

从图 3 可以观察到测试集 45 个样本中仅有一个分类错误，即将原始样本误判为老化 4 个月样本，分析原因在于，小麦种子的自然老化过程是缓慢进行的，存放 4 个月时小麦的老化程度与未老化样本差别不明显。而在老化 4 个月、7 个月、9 个月时，种子老化过程中的理化活动引起了种子成分的显著变化，因此尽管间隔只有 2 个月，但是可以准确的区分，说明近红外的方法用于小麦短期自然老化程度的判别也具有较高的准确度和灵敏度。

3 结 论

跟踪采集了 45 份小麦种子在短期自然老化阶段的近红外光谱, 通过光谱离散度均值分析筛选得到了与老化程度显著相关的近红外特征谱区, 通过谱区对应的化学基团归属解析可得, 小麦种子在短期自然老化阶段中以水分的显著变化为主, 蛋白和淀粉等变化次之。针对不同老化阶段的样本集

近红外光谱, 采用标准正态化预处理后采用 4 分类支持向量机方法建立了无损识别小麦种子自然老化程度的定性分析模型, 对于 4 类样本分类识别率准确率可达 97.78%。实验结果表明采用近红外光谱无损判别小麦种子老化程度在理论上和实践上都是切实可行的, 但是应用近红外光谱解析小麦种子在自然老化阶段中主要化学成分变化规律及其对种子活力的影响还有待后续深入分析。

References

- [1] LI Zhen-hua, WANG Jian-hua(李振华, 王建华). *Scientia Agricultura Sinica*(中国农业科学), 2015, 48(4): 646.
- [2] QU Chang-rong(屈长荣). *Seed Testing Technology*(种子检验技术). Tianjin: Tianjin University Press(天津: 天津大学出版社), 2011.
- [3] YAN Yan-lu, CHEN Bin, ZHU Da-zhou, et al(严衍禄, 陈斌, 朱大洲, 等). *Near Infrared Spectroscopy—Principle, Technology and Application*(近红外光谱分析的原理、技术与应用). Beijing: China Light Industry Press(北京: 中国轻工业出版社), 2013.
- [4] CHU Xiao-li(褚小立). *Molecular Spectroscopy Analytical Technology Combined with Chemometrics and Its Application*(化学计量学方法与分子光谱分析技术). Beijing: Chemical Industry Press(北京: 化学工业出版社), 2011. 262.
- [5] Lidia E A, Charles R H J. *Talanta*, 2014, 121: 288.
- [6] Jia Shiqiang, An Dong, Liu Zhe, et al. *Journal of Cereal Science*, 2015, 63: 21.
- [7] Maria Kyraleou, Christos Pappas, Eleni Voskidi, et al. *Industrial Crops and Products*, 2015, 74: 784.
- [8] Ashabahebwa Ambrose, Santosh Lohumi, Wang-Hee Lee, et al. *Sensors and Actuators B Chemical*, 2016, 224: 500.
- [9] Song Le, Wang Qi, Wang Chunyang, et al. *Journal of Stored Products Research*, 2015, 62: 46.
- [10] DUAN Yong-hong, LI Xiao-xiang, LI Wei-hong(段永红, 李小湘, 李卫红). *Seed*(种子), 2009, 1: 101.
- [11] Cortes C, Vapnik V N. *Machine Learning*, 1995, 20(3): 273.
- [12] Alves J C, Poppi R J. *Talanta*, 2013, 104(2): 155.
- [13] Jerry Workman, Lois Weyer. *Practical Guide to Interpretive Near-Infrared Spectroscopy*(近红外光谱解析实用指南). Translated by CHU Xiao-li, XU Yu-peng, TIAN Gao-you(褚小立, 许育鹏, 田高友, 译). Beijing: Chemical Industry Press(北京: 化学工业出版社), 2009.

Nondestructive Determination of Natural Aging Stage of Wheat Seeds Using Near Infrared Spectroscopy

WU Jing-zhu¹, LI Hui¹, ZHANG He-dong¹, MAO Wen-hua^{2*}, LIU Cui-ling¹, SUN Xiao-rong¹

1. Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China

2. Chinese Academy of Agricultural Mechanization Sciences, Beijing 100083, China

Abstract To study the variation trend of major chemical composition of wheat seeds during short-time natural aging, the nondestructive technology based on near infrared spectroscopy (NIR) and support vector machines (SVM) is applied to evaluate the natural aging stage at the same time. There are 45 wheat samples collected in the experiment. The samples are scanned at the beginning and after natural aging for 4 months, 7 months and 9 months respectively by VERTEX 70 Fourier transform infrared spectrometer in large sample cup rotation sampling mode. The spectral standard deviations of each sample at four natural aging stages are calculated firstly. The standard deviations represent the statistical quantity of data dispersion. The obvious variation regions are screened according to the standard deviations calculated from the spectrums of 4 aging stages. To avoid abnormal discrete degree value caused by accidental factors, the averages of 45 samples spectrum discrete degree are calculated. The spectral peaks are mainly distributed in the area of 8 362, 6 950, 7 563, 5 319, 4 998 and 4 478 cm^{-1} according to the standard deviation. The region nearby 6 950 cm^{-1} reflects stretching vibration of O—H in liquid water, and the standard deviation value is greater. This illustrates the moisture changes remarkably during natural aging stage. The region nearby 5 319, 4 998 and 4 478 cm^{-1} reflect vibration information of primary amide, secondary amide and amide in protein. The standard deviation values at these peaks are all lower than the value of 6 950 cm^{-1} , so the protein changes more slowly than moisture during aging stage. The region

nearby $8\ 362$ and $7\ 563\text{cm}^{-1}$ reflect secondary vibration information of C—H and the standard deviation value is greater. There are C—H group in protein, starch, etc. of wheat seeds. It shows that comprehensive changes of protein, starch and other components are relatively strong. According to the above analysis, the multi-classification model has been built based on NIR and SVM to determine the 4 types natural aging stages. The sample set is divided into two parts randomly according to the ratio of 3 : 1. The number of train sample is 135 and the number of test sample is 45. The best parameters of SVM are selected by grid searching. While the kernel function is RBF function, the penalty parameter is 8 and kernel parameter is 0.008 974 2, and the recognition rate of training set and test set reach to 99.26% and 99.78%. The results show that NIR technology combined with SVM can be applied to determine the natural aging stage of wheat seeds, which also provides a convenient and fast tool to monitor physiological characteristics changes during wheat seeds storage.

Keywords Wheat seeds; Natural aging; Near infrared spectroscopy; Support vector machines; Standard deviation

(Received Jun. 19, 2018; accepted Oct. 25, 2018)

* Corresponding author