

大豆水分含量的高光谱无损检测及可视化研究

金诚谦^{1,2}, 郭 榛¹, 张 静¹, 马成业¹, 唐小涵¹, 赵 男¹, 印 祥¹

1. 山东理工大学农业工程与食品科学学院, 山东 淄博 255000
2. 农业农村部南京农业机械化研究所, 江苏 南京 210000

摘 要 采用近红外高光谱成像技术对大豆水分含量进行快速无损检测, 实现大豆水分含量可视化。采集了96个不同品种大豆样本在900~2500 nm的高光谱图像, 采用直接干燥法测量每个大豆样品的水分含量。利用系统自带的HSI Analyzer软件提取图像感兴趣区域(ROI)的平均光谱信息, 代表样品的光谱信息。利用SPXY算法划分样品校正集和预测集, 并保留938~2215 nm波段范围内的光谱数据。采用移动平滑(moving average)、S-G平滑、基线校正(baseline)、归一化(normalize)、标准正态变量变换(standard normal variate, SNV)、多元散射校正(multiple scattering correction, MSC)、去趋势(detrending)共7种光谱预处理方法, 发现Normalize方法处理后的PLSR模型效果较好。为了去除光谱冗余信息, 简化预测模型, 采用连续投影算法(SPA)、竞争性自适应加权算法(CARS)、无信息消除变量法(UVE)提取特征波长, 其中SPA、CARS和UVE三种算法优选出14、16和29个波长, 分别占总波长的6.5%、7.4%和13.4%。分别对938~2215 nm波段光谱和特征波长建立预测模型, 并将效果较优的模型与Normalize方法结合。建立的14种预测模型效果相比较, 发现SPA算法筛选的特征波长建模预测效果较好, 并优选出Normalize-SPA-PCR模型, 模型的 R^2 和 R^2_p 值较高, 分别为0.9746和0.9778, RMSEP和RMSECV值较低, 分别为0.238和0.313, 模型的稳定性和预测性较好, 可以对大豆水分含量进行准确预测。将Normalize-SPA-PCR模型作为大豆水分含量可视化预测模型, 计算高光谱图像每个像素点的水分含量, 得到灰度图像, 对灰度图像进行伪彩色变换, 得到大豆水分含量可视化彩色图像。对预测集的24个大豆品种进行可视化处理, 发现不同水分含量大豆的可视化图像颜色不同, 水分含量变化对应的颜色变化较为明显。结果表明, 高光谱成像技术结合化学计量学可以准确快速无损预测大豆水分含量, 实现大豆水分含量可视化, 为大豆收获、贮藏加工过程中水分含量检测提供了技术支持。

关键词 高光谱成像技术; 水分含量; 大豆; 无损检测; 可视化

中图分类号: O433.5 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)10-3052-06

引 言

大豆是我国重要的粮食作物和经济作物, 其品质检测一直是研究的焦点。大豆籽粒种皮薄, 发芽孔大, 吸湿返潮后, 体积膨胀, 极易生霉, 含水量直接影响大豆的贮藏期。因此入库时要严格控制水分, 长期贮藏水分不能超过12%。此外, 在育种过程中, 水分含量影响大豆种子活力, 控制和检测大豆水分含量是保证种子质量的重要环节^[1]。常用的水分检测方法有105℃恒重法、真空干燥法、定温定时烘干法和

化学法等, 这些方法检测精度和准确度较高, 但其操作过程繁琐且费时, 破坏样品, 浪费优质种质资源, 不适用于大规模无损检测。

高光谱成像技术结合了光谱技术和成像技术的优点, 可以对多个目标同时进行无损检测, 实现物质成分含量可视化, 有着连续波段多、光谱分辨率高、“图谱合一”的优点, 满足了快速无损检测的需求。近年来广泛应用于茶叶病害侵染和水分含量等品质检测^[2-4]、鱼类和肉类品质指标的检测^[5-6]、小麦籽粒蛋白质含量检测^[7]等。Nicola等^[8]使用高光谱成像技术检测单粒咖啡豆水分和脂质含量, 并使水分和脂

收稿日期: 2021-08-11, 修订日期: 2021-11-11

基金项目: 国家自然科学基金项目(32171910), 国家重点研发计划项目(2017YFD0700305), 现代农业产业技术体系建设专项资金项目(CARS-04-PS26)资助

作者简介: 金诚谦, 1973年生, 山东理工大学农业工程与食品科学学院教授, 农业农村部南京农业机械化研究所研究员

e-mail: 412114402@qq.com

质含量分布可视化。Xu 等^[9]采集单粒黄瓜种子在 400~1 000 和 1 050~2 500 nm 范围内的高光谱图像,分别基于两个波段预测单粒黄瓜种子水分含量并进行可视化分析,发现在 1 000~2 500 nm 范围内对水分含量预测效果较好。Jennyfer 等^[10]在 900~1 700 nm 波段实现单粒花生仁的水分含量检测及可视化研究,但只采用了加权回归系数法提取特征波长。Wang 等^[11]采集单个玉米籽粒胚和胚乳两侧的高光谱图像,建立的 CARS-SPA-LS-SVM 模型 R_{pre} 值为 0.931 1,表明高光谱成像技术可以快速无损检测玉米籽粒中的水分含量。朱洁等^[12]预测单粒小麦水分含量,并将单粒小麦水分含量可视化。相关研究表明,高光谱成像技术可以实现作物种子水分含量检测及可视化,目前未见高光谱成像技术检测大豆水分含量的相关报道。

本工作以 96 个品种的大豆为研究对象,利用高光谱成像技术结合化学计量学方法建立并寻找最优预测模型,在 900~2 500 nm 范围检测大豆水分含量并进行可视化研究,为大豆收获、贮藏加工过程中水分含量检测提供新的方法。

1 实验部分

1.1 大豆种子样本及样本水分含量测定

试验所用的大豆来自黑龙江龙科种业公司、辽宁东亚种业公司和临沂河东区试验农场等,包括黑农 84、绥农 88、沈农 8、东豆 1133、中黄 37、徐豆 20 等 96 个不同品种。每个品种取 100g 样品分别放置在培养皿中,在实验室静置 72h 后采集高光谱图像,随后按照 GB 5009.3—2016《食品安全国家标准食品中水分的测定》中的直接干燥法测量每个品种大豆样品的水分含量。每个品种测量三次,取平均值作为该品种大豆的水分含量。

1.2 高光谱成像系统

光谱采集仪器为近红外高光谱成像系统(中国台湾五铃光学,HSI-eSWIR-900~2 500 nm),101-0E 型电热鼓风干燥箱。

近红外高光谱成像系统由 900~2 500 nm 线扫式近红外光谱仪(芬兰 Specim, N25E-SWIR)、900~2 500 nm CCD 相机镜头(品牌:芬兰 Specim, OLES30)、900~2 500 nm 双分支卤素灯光源(中国台湾五铃光学,IRCP0078-1COMB)、暗箱、计算机等构成。

1.3 黑白标定

为了减小暗电流以及光源强度不均匀对高光谱图像的影响,需要对高光谱图像进行黑白校正^[13]。将与样品等高且反射率为 0.99 的白板(芬兰 Specim 公司)置于样品采集区域,采集的图像作为白板标定图像,记为 I_w ;盖上 CCD 相机镜头盖,采集的图像作为黑板标定图像,记为 I_d 。大豆高光谱图像黑白校正公式如式(1)所示

$$R_T = \frac{I - I_d}{I_w - I_d} \quad (1)$$

式(1)中, R_T 为校正后的样品图像, I 为原始样品图像。

1.4 光谱采集及 ROI 提取

样本扫描时,曝光时间为 2.9 ms,位移平台移动速度为

15.34 mm·s⁻¹,焦距为 30.7 mm,相机分辨率为 384×288,光源入射角度为 45°。

采用高光谱成像系统自带的 HSI Analyzer 光谱分析软件提取高光谱图像感兴趣区域(region of interest, ROI),选取半径为 200 像素的圆形区域为 ROI,提取 ROI 平均光谱作为样本光谱信息。

2 结果与讨论

2.1 样本划分

采用 SPXY(sample set partitioning based on joint X-Y distance)算法将样本按照 3:1 的比例划分为校正集和预测集^[14]。样本水分含量如表 1 所示,校正集样品水分含量范围涵盖了预测集范围,说明样本集划分合理。

表 1 大豆样品水分含量

Table 1 Moisture content of soybean samples

样本集	样本数/个	水分含量/%			
		最大值	最小值	平均值	标准偏差
校正集	72	11.06	6.12	7.86	1.63
预测集	24	10.69	6.13	7.99	1.36
总样本	96	11.06	6.12	7.90	1.58

2.2 原始光谱及光谱预处理

图 1 是大豆样品光谱反射率曲线,1 210 nm 附近明显的反射率波谷是有机物中 C—H 键的二级倍频振动带;1 450 nm 处反射率波谷与 O—H 键倍频振动有关,1 940 nm 表现 O—H 键的合频特性,都是水分含量的特征谱带。原始光谱包含背景信息和噪声,在 938 nm 之前和 2 215 nm 之后的光谱数据无法提供有效的样本信息^[15]。故保留 938~2 215 nm 共 216 个光谱带作为建模数据,带间距为 5.6 nm。

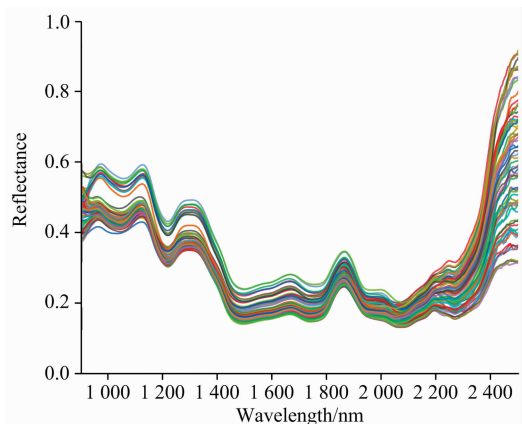


图 1 光谱反射曲线

Fig. 1 Reflectance curves of spectrum

为了减小无关信息和噪声对光谱数据的影响,结合化学计量学方法对保留的光谱数据进行预处理。PLSR 模型综合考虑自变量和因变量之间影响关系,能够在自变量存在严重的多重共线性条件下进行回归建模,因此选择 PLSR 模型比

较不同预处理方法效果^[16]。采用留一交互验证法计算交互验证集均方根误差 RMSECV 并做为模型评价指标。由表 2 可知, 未经光谱预处理和 7 种预处理的 PLSR 大豆水分含量预测模型中, 未经光谱预处理预测模型的 RMSECV 值为 0.373, Normalize 方法处理后, 模型的 RMSECV 值较小, 为 0.353, 且与未预处理相比 R_c^2 由 0.957 6 提高到 0.960 7, 但略低于经 Detrending 处理的 0.962 2。说明经预处理后模型稳定性和交互验证效果更好, 且 Normalize 方法处理后效果较好, 故选择 Normalize 方法对后期数据进行处理。

表 2 不同预处理方法 PLSR 模型

Table 2 PLSR model based on different pretreatment methods

预处理方法	PCs	校正集		交互验证集	
		R_c^2	RMSEC	R_{cv}^2	RMSECV
无	8	0.957 6	0.278	0.926 6	0.373
Moving Average	8	0.956 7	0.281	0.924 3	0.379
S-G 平滑	8	0.957 3	0.279	0.925 9	0.378
Baseline	8	0.958 7	0.275	0.930 4	0.369
Normalize	8	0.960 7	0.268	0.938 0	0.353
SNV	9	0.961 1	0.266	0.921 1	0.388
MSC	7	0.949 4	0.304	0.916 8	0.384
Detrending	8	0.962 2	0.263	0.930 3	0.364

2.3 特征波长选取

保留的 216 个光谱带中仍然包含大量冗余信息, 为了提高建模速度和模型鲁棒性, 需要对光谱数据进行压缩, 提取特征波长^[17]。应用连续投影算法^[18] (successive projections algorithm, SPA)、竞争性自适应加权算法^[19] (competitive adaptive reweighted sampling, CARS)、无信息消除变量法 (uniformative variable elimination, UVE) 分别提取特征波长。

2.3.1 连续投影算法(SPA)

SPA 筛选出 14 个特征波长, 占总波长的 6.5%, 这 14 个波长分别为 1 001, 1 296, 1 377, 1 452, 1 575, 1 726, 1 867, 1 896, 1 930, 1 952, 1 986, 2 052, 2 127 和 2 185 nm。图 2 为 SPA 筛选出的 14 个波长。

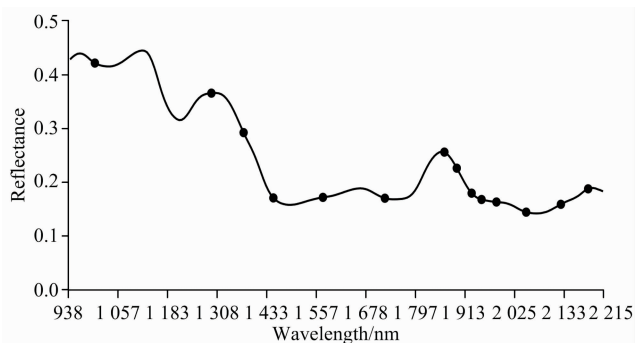


图 2 SPA 筛选特征波长

Fig. 2 Selected characteristic wavelengths by SPA

2.3.2 竞争性自适应加权算法(CARS)

图 3(a) 显示随着采样次数增加, CARS 筛选得到的变量

数逐渐减少, 且变量数变化的趋势为迅速减小到趋于平缓; 图 3(b) 显示筛选过程中交互验证错误率的变化趋势: 交互验证错误率平稳下降到最低点后曲折上升, 并在采样次数为 28 次时, 交互验证的 RMSECV 值最小, 模型的稳定性最好; 图 3(c) 为各变量在采样过程中回归系数的变化路径。经 CARS 筛选得到 16 个特征波长, 分别为: 1 308, 1 358, 1 390, 1 483, 1 672, 1 678, 1 962, 1 779, 1 832, 1 861, 1 941, 2 019, 2 025, 2 122, 2 133 和 2 138 nm, 占总波长的 7.4%。

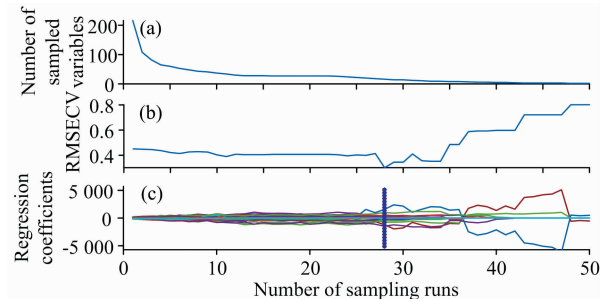


图 3 CARS 筛选过程

(a): 变量个数随样本个数的变化趋势; (b): 均方根误差;
(c): 各变量回归系数随采样次数的变化
(蓝色线表示 RMSECV 最低的位置)

Fig. 3 Selection process of CARS variables

(a): Variation trend of the number of variables with the number of samples; (b): RMSECV; (c): The change process of regression coefficient of each variable with sampling times (The blue line represents the position with the lowest RMSECV)

2.3.3 无信息消除变量法(UVE)

UVE 筛选特征波长, 当潜在变量设为 13 时, PLSR 模型的 RMSECV 值最小, 为 0.327。图 4 中, 竖虚线左右分别为 216 个波长变量, 左侧为 216 个输入变量稳定性 C 分布曲线, 右侧为 UVE 产生的 216 个随机变量稳定性 C 分布曲线; 两条水平虚线为变量选择阈值的上下限, 虚线外对应变量为筛选出的 29 个特征波长: 976, 982, 988, 994, 1 001, 1 096, 1 076, 1 082, 1 089, 1 227, 1 233, 1 239, 1 246, 1 346, 1 352, 1 358, 1 365, 1 371, 1 377, 1 396, 1 402, 1 408, 1 415, 1 421, 1 427, 1 433, 1 439, 1 446 和 1 452 nm, 占总波长的 13.4%。

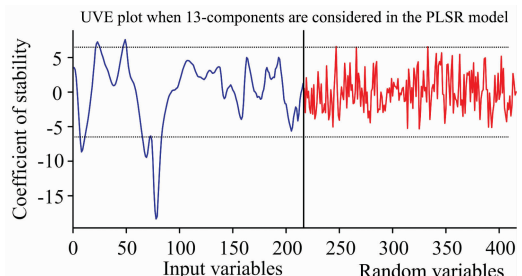


图 4 UVE-PLSR 模型的稳定性分布曲线

Fig. 4 Stability distribution curve of UVE-PLSR model

2.4 模型的优选

对 938~2 215 nm 波段光谱建立 PLSR, PCP 和 SVMR

模型。将预测集均方根误差 RMSEP 值作为评价模型预测效果的指标，RMSEP 值越低说明预测效果越好。其中，PCR 模型 RMSEP 和 RMSECV 值较低，说明基于 938~2 215 nm 波段光谱建立的 PCR 模型预测效果和稳定性较好。

为了提高建模速度和模型鲁棒性。分别对 SPA, CARS 和 UVE 三种算法筛选出来的 14, 16 和 29 个特征波长建立 PLSR, PCR 和 SVMR 模型。如表 3 所示，SPA 算法筛选出的特征波长建立的 PLSR, PCR 和 SVMR 模型较 938~2 215 nm 波段光谱建立的三种模型，RMSEP 值均有所降低，而 CARS 和 UVE 算法对模型预测效果提升并不明显甚至会降

低预测效果，但也有效降低了光谱维度。基于特征波长建立的模型中，SPA-PLSR 和 SPA-PCR 模型 RMSEP 值较低，均为 0.262，说明 SPA 算法筛选的特征波长建模预测效果较好，这可能是由于 SPA 算法能有效降低光谱共线性。

将 Normalize 方法与 SPA-PLSR 和 SPA-PCR 模型结合，发现模型的 RMSEP 值降低，经预处理后模型的预测效果进一步提升。两种模型相比较，RMSEP 值相同，但 Normalize-SPA-PCR 模型的 RMSECV 值较低，说明 Normalize-SPA-PCR 模型比 Normalize-SPA-PLSR 模型更稳定。将 Normalize-SPA-PCR 模型更适合用于大豆水分含量可视化预测。

表 3 基于不同预处理方法及特征波长筛选方法建立的模型效果

Table 3 Performance of models based on different pretreatment methods and characteristic wavelengths selecting methods

No	模型	波长数	校正集		交互验证集		预测集	
			R^2_c	RMSEC	R^2_{cv}	RMSECV	R^2_p	RMSEP
1	PLSR	216	0.957 6	0.278	0.926 6	0.373	0.957 1	0.329
2	PCR	216	0.953 7	0.291	0.930 0	0.367	0.963 7	0.303
3	SVMR	216	0.955 6	0.287	0.911 8	0.402	0.886 2	0.537
4	SPA-PLSR	14	0.967 4	0.244	0.933 7	0.358	0.972 9	0.262
5	SPA-PCR	14	0.967 7	0.243	0.934 1	0.355	0.972 9	0.262
6	SPA-SVMR	14	0.955 8	0.287	0.927 0	0.367	0.906 1	0.488
7	CARS-PLSR	16	0.982 9	0.177	0.968 8	0.254	0.952 0	0.349
8	CARS-PCR	16	0.982 5	0.179	0.964 4	0.257	0.955 8	0.335
9	CARS-SVMR	16	0.953 7	0.294	0.931 5	0.356	0.915 5	0.463
10	UVE-PLSR	29	0.964 7	0.254	0.944 0	0.350	0.953 8	0.299
11	UVE-PCR	29	0.967 3	0.244	0.944 0	0.326	0.958 5	0.324
12	UVE-SVMR	29	0.936 8	0.340	0.903 8	0.420	0.915 5	0.463
13	Normalize-SPA-PLSR	14	0.974 3	0.217	0.948 3	0.325	0.977 8	0.238
14	Normalize-SPA-PCR	14	0.974 6	0.215	0.948 9	0.313	0.977 8	0.238

2.5 大豆水分含量可视化分析

在大豆收获和加工储藏过程中无法用肉眼直接判断水分含量，而利用预测模型可以计算出高光谱图像上的每一个像素点的水分含量预测值，得到灰度图像，然后对灰度图像进行伪彩色变换，得到大豆水分含量可视化图。

图 5 是 Normalize-SPA-PCR 模型预测得到的大豆水分含量可视化图，颜色梯度条由下向上代表大豆水分含量由低变高，范围为 0~12%。图 5 为从预测集挑选的 4 个品种大

豆水分含量可视化图，4 个大豆品种按照平均水分含量高低进行排列，其中，(a) 为华豆 2 号的水分含量可视化图，水分含量为 10.40%；(b) 为垦豆 40 的水分含量可视化图，水分含量为 9.39%；(c) 为皖豆 701 的水分含量可视化图，水分含量为 7.13%；(d) 为皖豆 34 的水分含量可视化图，水分含量为 6.46%。由图 5 可知，不同品种大豆对应的水分含量可视化图颜色不同，4 幅图像颜色差异十分明显；同一图像内不同大豆的颜色也有差异，但颜色差异较小。对预测集 24 个

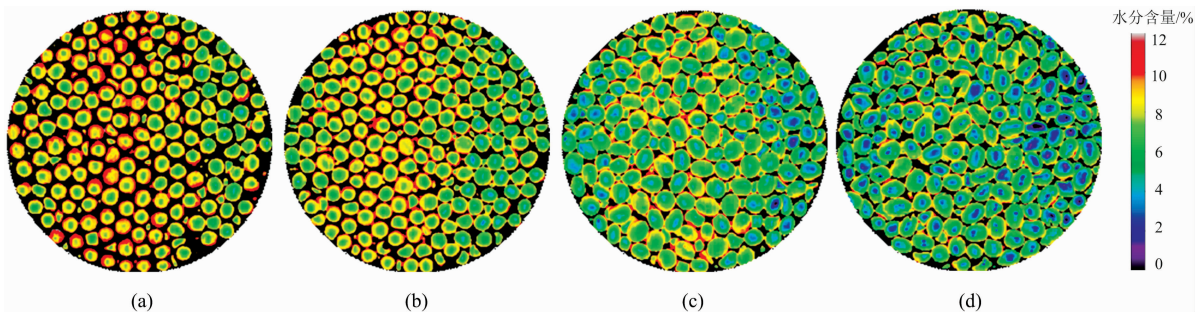


图 5 大豆水分含量可视化图

(a): 华豆 2 号; (b): 垦豆 40; (c): 皖豆 701; (d): 皖豆 34

Fig. 5 Visualization of soybean moisture content

(a): Huadou 2; (b): Kendou 40; (c): Wandou 701; (d): Wandou 34

品种的大豆高光谱图像进行可视化处理,结果表明大豆水分含量不同对应图像颜色不同,水分含量变化对应图像颜色变化较为明显,通过图像颜色变化可以判断大豆水分含量范围。

实验表明高光谱成像技术可以有效实现大豆水分含量可视化检测,与传统的水分测量方法相比,有着无损、快捷、准确的优点,并且可以得到大豆水分含量可视化图,为大豆的收获、储藏和加工提供了技术支持。

3 结 论

(1)使用 7 种预处理方法对 938~2 215 nm 波段光谱数据建立 PLSR 模型,发现经 Normalize 处理后模型的 R_c^2 值较高, RMSECV 值较低,模型稳定性较好。

(2)采用 SPA, CARS 和 UVE 三种方法提取特征波长特征波长个数分别为 14, 16 和 29 个,占光谱数据的 6.5%, 7.4%和 13.4%,有效的降低了光谱维度。

(3)将预处理方法和特征波长提取方法与建模方法相结合,比较 14 种模型效果,优选出 Normalize-SPA-PCR 模型,并用于大豆水分含量可视化预测。Normalize-SPA-PCR 模型的 R_c^2 值和 R_p^2 较高,分别为 0.974 6 和 0.977 8, RMSEP 和 RMSECV 值较低,分别为 0.238 和 0.313,模型的稳定性和预测效果较好,可快速准确无损检测大豆水分含量。

(4)对预测集 24 个高光谱图像进行可视化处理,不同水分含量大豆的可视化图像颜色不同,水分含量变化对应的颜色变化也较为明显,通过图像颜色变化判断大豆水分含量范围。

References

- [1] ZHANG Feng, LIU Mei, YANG Cui-cui, et al(张 凤, 刘 美, 杨翠翠, 等). Shandong Agricultural Sciences(山东农业科学), 2014, 46(8): 37.
- [2] Yuan Lin, Yan Peng, Han Wenyan, et al. Computers and Electronics in Agriculture, 2019, 167: 105039.
- [3] Deng Shuiguang, Xu Yifei, Li Xiaoli, et al. Computers and Electronics in Agriculture, 2015, 118: 38.
- [4] Wei Yuzhen, Wu Feiyue, Xu Jie, et al. Journal of Food Engineering, 2019, 248: 89.
- [5] Mohammed Kamruzzaman, Gamal ElMasry, Da-Wen Sun, et al. Analytica Chimica Acta, 2011, 714: 57.
- [6] Yao Xinli, Cai Fuhong, Zhu Peiyi, et al. Meat Science, 2019, 152: 73.
- [7] Nicola Caporaso, Martin B Whitworth, Ian D Fisk. Food Chemistry, 2018, 240.
- [8] Nicola Caporaso, Martin B Whitworth, Stephen Grebby, et al. Journal of Food Engineering, 2018, 227.
- [9] Xu Yunfei, Zhang Haijun, Zhang Chi, et al. Infrared Physics and Technology, 2019, 102.
- [10] Jennyfer D Rabanera, Jose D Guzman, Kevin F Yaptenco. Journal of Food Measurement and Characterization, 2021(prepublish).
- [11] Wang Zheli, Fan Shuxiang, Wu Jingzhu, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2021, 254: 119666.
- [12] ZHU Jie, HUANG Hua, ZHU Shi-ping, et al(朱 洁, 黄 华, 祝诗平, 等). Journal of the Chinese Cereals and Oils Association(中国粮油学报), 2019, 34(8): 133.
- [13] XU Ying-chao, WANG Xiang-you, YIN Xiang, et al(许英超, 王相友, 印 祥, 等). Transactions of the Chinese Society for Agricultural Machinery(农业工程学报), 2018, 49(2): 339.
- [14] Wang Qiaoyun, Wu Guangfei, Pian Feifei, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2021, 260: 119906.
- [15] Ma Chengye, Ren Zhishang, Zhang Zhehao, et al. Vibrational Spectroscopy, 2021, 114: 113230.
- [16] WANG Song-lei, WU Long-guo, WANG Cai-xia, et al(王松磊, 吴龙国, 王彩霞, 等). Journal of Optoelectronics • Laser(光电子 • 激光), 2019, 30(9): 941.
- [17] REN Zhi-shang, PENG Hui-hui, HE Zhuang-zhuang(任志尚, 彭慧慧, 贺壮壮, 等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2020, 51(S2): 466.
- [18] Mário César Ugulino Araújo, Teresa Cristina Bezerra Saldanha, Roberto Kawakami Harrop Galvão, et al. Chemometrics and Intelligent Laboratory Systems, 2001, 57(2): 65.
- [19] Li Hongdong, Liang Yizeng, Xu Qingsong, et al. Analytica Chimica Acta, 2009, 648(1): 77.

Non-Destructive Detection and Visualization of Soybean Moisture Content Using Hyperspectral Technique

JIN Cheng-qian^{1, 2}, GUO Zhen¹, ZHANG Jing¹, MA Cheng-ye^{1, 2}, TANG Xiao-han¹, ZHAO Nan¹, YIN Xiang¹

1. School of Agricultural Engineering and Food Science, Shandong University of Technology, Zibo 255000, China

2. Nanjing Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs, Nanjing 210000, China

Abstract NIR Hyperspectral imaging technology was used to detect soybean moisture content rapidly and non-destructively and realized the visualization of soybean moisture content. A total of 96 soybean samples of hyperspectral images in the region of 900~2 500 nm were acquired, and the moisture content of each soybean sample was measured by the direct drying method. The average spectral information of the region of interest(ROI)of the image was extracted by HSI Analyzer software, representing the sample's spectral information. The SPXY algorithm was used to divide the sample calibration set and prediction set, and the spectral data in the band range of 938 to 2 215 nm were retained. The spectral's pretreatment was analyzed, such as Moving Average, Smoothing S-G, Baseline, Normalize, Standard Normal Variate(SNV), Multiple Scattering Correction(MSC) and Detrending, and the PLSR model established after Normalize pretreatment had the best effect. The characteristic wavelengths were selected by successive projections algorithm(SPA), competitive adaptive reweighted sampling(CARS)and uninformative variable elimination(UVE). 14,16 and 29 characteristic wavelengths were selected by SPA, CARS and UVE, accounting for 6.5%, 7.4% and 13.4% of the total wavelengths. The prediction models were established for the spectra and characteristic wavelengths of 938~2 215 nm, and the model with better effect was combined with the Normalize method. Compared with the 14 prediction models established, it was found that the modeling and prediction effect of characteristic wavelengths selected by the SPA algorithm was good, and the Normalize-SPA-PCR model was optimized. The values of R_c^2 and R_p^2 in the model were higher, which were 0.974 6 and 0.977 8, respectively, while the values of RMSEP and RMSECV in the model were lower, which were 0.238 and 0.313, respectively. The stability and predictability of the model were good, which could be used to predict the soybean moisture content accurately. The Normalize-SPA-PCR model was used as a visual prediction model for soybean moisture content, and the moisture content of each pixel in the hyperspectral image was calculated to obtain a gray image. The gray image was transformed by pseudo-color transformation to obtain a visual color image of soybean moisture content. The 24 soybean varieties in the prediction set were visualized. The color of the visualized image was different with different moisture content, and the color of the visualized image was more evident with different moisture content. The results showed that hyperspectral imaging combined with stoichiometry could accurately, rapidly, and non-destructive predict soybean moisture content. They realized the visualization of soybean moisture content, which provided technical support for soybean moisture content detection in the process of soybean harvest, storage and processing.

Keywords Hyperspectral imaging; Moisture content; Soybean; Non-destructive detection; Visualization

(Received Aug. 11, 2021; accepted Nov. 11, 2021)