

## 小米米粉碱消值的高光谱快速预测

王国梁<sup>1,2</sup>, 余克强<sup>3</sup>, 成 锴<sup>2</sup>, 刘 鑫<sup>2</sup>, 王文俊<sup>1</sup>, 李 洪<sup>2</sup>, 郭二虎<sup>2</sup>, 李志伟<sup>1\*</sup>

1. 山西农业大学农业工程学院, 山西 太谷 030801
2. 山西农业大学谷子研究所, 山西 长治 046000
3. 西北农林科技大学机械与电子工程学院, 陕西 杨凌 712100

**摘 要** 小米米粉的主要成分是淀粉, 其食味品质决定小米米粉的市场价值。糊化特性是小米米粉的重要物理特性, 而碱消值是能够直接反应其糊化特性的主要特征指标。通过小米米粉碱消值的差异, 可以间接反映直链淀粉含量, 当碱消值降低时, 相反, 糊化温度和直链淀粉含量却很高, 而小米米粉口感粘糯性越差。采用高光谱技术结合化学计量学方法, 建立快速检测小米米粉碱消值预测模型, 旨在探索一种快速、无损、低成本预测小米米粉碱消值的方法。实验采集小米米粉高光谱数据, 在被测样品感兴趣区域(ROI)按像素点逐一选择, 提取高光谱数据矩阵, 并进行均值运算, 得到每个样品在各个光谱波段的平均光谱值。利用粘度测定仪(RVA)测定小米米粉碱消值指标。光谱数据采用全波段、竞争性自适应重加权采样法(CARS)及随机蛙跳(RF)法选择特征波段处理, 建立偏最小二乘回归(PLSR)模型; 全波段建立预测模型  $R_p$  值最高为 0.77, 说明能够利用小米米粉高光谱反射率反演小米米粉的碱消值, 而采用其他两种计算方法所得  $R_p$  值分别为 0.72 和 0.7, 与前者较为接近, 也反映了采用 CARS 和 RF 建立的回归模型具有可行性。为提高预测精度, 采用 Savitzky-Golay(S-G)法、多元散射校正(MSC)和 S-G+MSC 对数据预处理。可以看出采用 MSC 预处理光谱数据建立 PLSR 模型性能较好( $R_p=0.83$ )。对 MSC 预处理后的数据再次 CARS 和 RF 法选择特征波段, 建立 PLSR 模型, 与未进行预处理的回归模型相比,  $R_p$  值变化不大, 这也说明 CARS 和 RF 具有一定的稳定性, 可以作为小米米粉高光谱反射率预测碱消值的参考方法。结果表明: 为实现对小米米粉碱消值的快速、无损检测, 通过运用高光谱技术能够利用小米米粉高光谱反射率预测碱消值, 进而为小米米粉品质升级、加工及碱消值传感器的开发提供参考依据和数据支撑。

**关键词** 高光谱; 小米米粉; 碱消值; 化学计量学方法

**中图分类号:** O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)10-3189-05

### 引 言

随着现代人对生活品质的提升以及健康产业兴起, 小米产业作为山西省“有机旱作”农业发展的龙头产业, 其深加工产品市场价值逐渐被大众所普遍肯定, 而小米米粉又以其独特的食味品质优势, 现已成为婴幼儿辅食佳品。小米米粉中主要成分是淀粉, 淀粉在一定条件下具有晶体性不可逆的特点, 在加入足量水情况下, 将其混合液加热、升温, 淀粉吸水膨胀、溶解呈现出这种特性, 而这一现象称为糊化现象。

其中碱消值是能够直接反应糊化特性的主要特征指标, 碱消值越低, 糊化温度越高, 直链淀粉含量越高, 而小米米粉口感粘糯性变差<sup>[1-2]</sup>, 最终影响小米产品的市场价格。通常利用人工评级、冷碱糊化法和粘度速测仪(rapid visco analyser, RVA)测定淀粉粘度特征值来评价小米米粉的糊化特性及品质<sup>[3]</sup>; 而人工评级人为因素大, 结果不可靠; 冷碱糊化法和 RVA 粘度测定耗时, 破坏样品理化特性, 并且制备样品操作程序繁琐, 成本太高, 因此, 样品批量处理和实现快速检测存在一定困难。

高光谱技术是一种通过精确采集被测物每个像素点的连

**收稿日期:** 2020-09-25, **修订日期:** 2021-01-14

**基金项目:** 国家重点研发计划项目(2017YFD0701501), 山西省重点研发计划项目(201903D211005), 国家现代农业产业技术体系建设专项(CARS-06-13.5-A21), 山西农业大学科技创新基金项目(2017YJ12), 山西省优秀博士来晋工作奖励资金项目(SXYBKY2019018)资助

**作者简介:** 王国梁, 1985年生, 山西农业大学农业工程学院博士研究生 e-mail: wangguoliangwz@126.com

\* 通讯作者 e-mail: lizhiweitong@163.com

续光谱信息,并对被测样品的感兴趣区域(region of interest, ROI)数据提取和处理,从而反映被测物组成成分和分子结构的无损检测技术,具有检测效率高,不破坏被测样品的理化特性,检测时间短并具有线上实时分析等特点。目前高光谱技术已广泛被应用在土壤和农产品的分类、成分反演等农业检测相关工作中<sup>[3-14]</sup>。Shao 等利用高光谱成像结合化学计量学方法对不同种类的米粉进行分类、判别,运用竞争性自适应重加权采样法(competitive adaptive reweighted sampling, CARS)建立最小二乘支持向量机回归模型,对品种分类的结果最好<sup>[4]</sup>;Chen 等采用连续投影算法结合多元线性回归对谷子蛋白质、糖类和粗脂肪含量进行近红外光谱特性分析,结果表明利用谷子近红外光谱数据能够反演谷子中蛋白质和糖类含量<sup>[5]</sup>;Chen 等运用随机蛙跳(random frog, RF)算法提取特征波段结合高斯过程回归预测木材含水模型效果最好<sup>[6]</sup>。当前,小米米粉的高光谱研究主要集中在品种分类和成分反演等方面<sup>[4-5]</sup>,而针对小米米粉糊化特性的研究相关报道仍然较少。

以晋东南地区“羊肥小米”主产地采集到的小米米粉作为实验样本,运用高光谱仪器采集小米米粉光谱数据,RVA 仪器测定小米米粉碱消值指标,结合化学计量学方法,建立快速检测小米米粉碱消值预测模型,旨在寻求一种快速、无损、低成本预测小米米粉碱消值的方法。通过探索小米米粉糊化特性与高光谱之间的联系,间接反映小米米粉中直链淀粉占比,为小米米粉品质评级及加工提供参考依据,进而为小米米粉碱消值传感器的开发提供理论支撑。

## 1 实验部分

### 1.1 样本制备

样本材料“羊肥小米”收获于山西省长治市武乡县。在大田收获季采集实验样本材料过程中,采用棋盘式取样法取样,取样点数 1 800 个,单点采样面积 2 m<sup>2</sup>,将相邻 5 个采样点采集到的材料混匀,得到实验样本 358 份经清选、自然晾晒、去壳、研磨、过 80 目筛,取得待测小米米粉样品,并采用四分法取样<sup>[7]</sup>,被测样品先用于小米米粉高光谱数据采集,然后进行实验室碱消值测定。

### 1.2 仪器与数据采集

采用美国 Headwall Photonics 公司的高光谱成像仪采集光谱数据。该设备工作组件主要包括:高光谱成像仪(光圈 1.4,焦距 25 mm)、钨丝灯、可移动式扫描平台、电脑等。数据采集设备参数为:光谱通道步长 4.715 nm,采集范围 900~1 700 nm,共包含 170 个波段。为能采集到清晰图像,调整平台移动速度为 16 mm·s<sup>-1</sup>,曝光时间 0.9 ms,被测样品上表面距镜头 300 mm。

实验前,采集暗背景(反射率 0.1%)图像 B,扫描白背景(反射率 99.9%)图像 W,采集样品的绝对图像 I,根据式(1)校正为相对图像 X<sup>[8]</sup>。每采集 3 幅高光谱图像,重复一次光谱校正。

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

进行高光谱数据采集时,将被测样品装入直径 5 cm,高度 1 cm 的圆形试验器皿内,采样过程中要保证样品表面处理平整、紧实,将被测样品置于推扫平台上,每份样品推扫三次,采集光谱数据,同一样品 3 次推扫,取均值、编号,按实验时序保存。

### 1.3 小米米粉碱消值测定

采用澳大利亚 Newport 公司生产粘度分析仪,根据 GB/T 24852—2010 方法测得小米米粉碱消值<sup>[9]</sup>。

### 1.4 小米米粉高光谱数据处理方法

光谱数据采集时,在被测样品感兴趣区域 ROI,按像素点逐一选择,提取光谱数据矩阵,并进行均值运算,得到每个样品在各个光谱波段的平均光谱值。利用 Kennard-Stone 算法(K-S)将光谱数据按照 2:1 分为训练集和预测集<sup>[10-11]</sup>。对光谱数据全波段、CARS<sup>[4,12]</sup>和 RF<sup>[6,13]</sup>算法选择特征波段后建立偏最小二乘回归(partial least square regression, PLSR)预测模型,模型评估利用相关系数(correlation coefficient, R)值。为消除噪声对光谱数据的影响,提高模型预测精度,进一步对原始光谱数据分别运用 S-G,多元散射校正(multiplicative scatter correction, MSC)和 S-G+MSC 进行预处理<sup>[14]</sup>,再利用 CARS 和 RF 法选择特征波段,建立 PLSR 模型。光谱数据处理软件主要有 The Unscrambler X 10.1 (CAMO, Norway) 和 MATLAB 2018a (The MathWorks, USA)。

## 2 结果与讨论

### 2.1 小米米粉碱消值统计分析

称重(3±0.01) g 小米米粉样品(按 12%湿基校准)与(25±0.1) mL 水(按 12%湿基校准)在铝盒中混匀,置于 RVA 分析仪中进行碱消值检测,3 次重复测量碱消值,取每份样品 3 次测量均值为该实验最终值。小米米粉碱消值统计结果见表 1。

表 1 小米米粉碱消值统计结果

Table 1 Statistic results of alkali spreading values in millet flour

Sample Number	Minimum /°C	Maximum /°C	Mean /°C	Standard deviation
358	75.25	84.25	79.65	1.98

### 2.2 小米米粉光谱特性

图 1 为小米米粉平均光谱曲线,由于靠近光谱量程首尾两端(≤900 nm 和 ≥1 700 nm),光谱反射率噪声较大,故截取光谱可用范围为 950~1 650 nm,光谱数降为 148 条,在每幅图像中的 ROI 提取光谱数据,经均值运算后组成 358×148 待数据处理的光谱数据矩阵。从图 1 可看出,曲线吸收峰值主要集中在 980, 1 210 和 1 450 nm 附近,其中 980 和 1 210 nm 处主要受被测样品淀粉影响,1 450 nm 处受被测样品中水分子吸收影响<sup>[12]</sup>。利用小米米粉全波段绘制成的光谱曲线变化平滑,但会存在数据重叠、平移和微小噪声的问

题。为提高精度，对特征波段提取后的数据矩阵进行预处理。

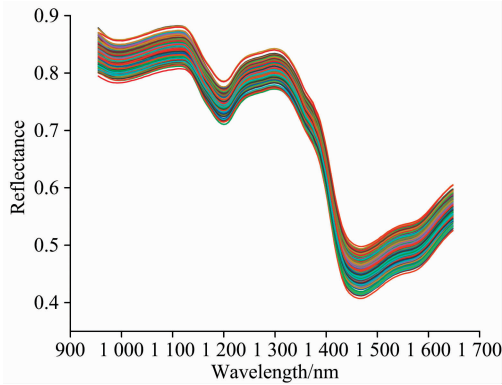


图 1 小米米粉平均光谱曲线

Fig. 1 Average spectral curves of millet flour

## 2.3 特征波段的选择

### 2.3.1 CARS 特征波段选择

CARS 是借鉴达尔文进化论“优胜劣汰”的理论思想，数

据处理采用该算法进行光谱特征波段快速选择<sup>[10]</sup>。图 2 为 CARS 选择特征波段的过程和主要参数的变化曲线，其中图 2(a)表示被测样品数随采样次数增加的变化趋势；图 2(b)中可以看出通过交叉检验(cross validation, CV)得到随采样次数增加均方根误差(root mean square error cross validation, RMSECV)的变化曲线，该曲线在 1—26 平滑下降，27—50 逐步阶梯状回升，最小值出现在 26，此点处 RMSECV 达到最小值；从图 2(c)中回归系数(regression coefficients, RC)，变量最优选择出现在图 2(c)中蓝色星号竖线标注的位置，在这点处尽可能多的保留了变量信息并选择出特征波段。经过 CARS 特征波段选择，共有 16 条特征波段被选出，分别为 964, 1 006, 1 053, 1 166, 1 171, 1 195, 1 213, 1 218, 1 223, 1 246, 1 298, 1 308, 1 350, 1 378, 1 444 和 1 529 nm。其中 964, 1 006 和 1 053 nm 在 O—H 三级倍频附近(980 nm)，1 166, 1 171, 1 195, 1 213, 1 218, 1 223, 1 246, 1 298, 1 308, 1 350 和 1 378 nm 在 C—H 二级倍频及组合频附近(1 220, 1 360 nm)，1 444 nm 此处 O—H 键与样品中水分子的吸收峰重合<sup>[12]</sup>(1 450 nm)与淀粉的近红外特性相应。

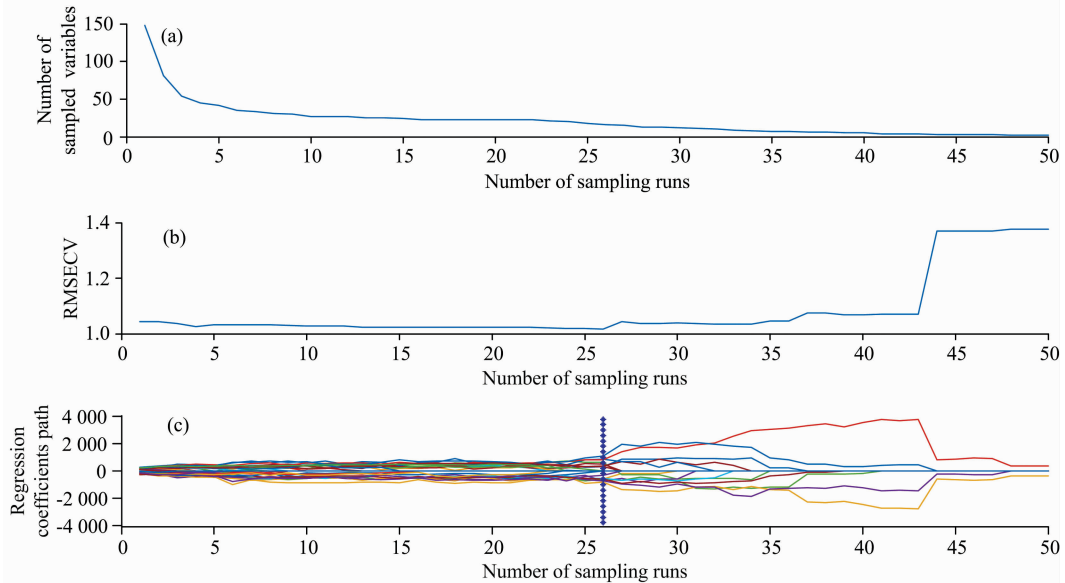


图 2 利用 CARS 算法的特征变量选择

(a): 变量变化趋势; (b): RMSECV 变化曲线; (c): 回归系数变化趋势

Fig. 2 Selection of key variables using CARS algorithm

(a): Changing trend of the number of sampled variables; (b): Variation of root-mean-square error of cross-validation values;

(c): Regression coefficients of each variable with the increasing of sampling runs

### 2.3.2 RF 特征波段选择

RF 是以统计学思想为理论基础，通过重复计算各变量被选概率来评价变量，具有不易过拟合优点<sup>[11]</sup>。RF 算法运行前，先设置相关参数(迭代参数、运行次数、初始跳转变量、候选变量集被选概率等)，依据相关文献<sup>[13]</sup>选择 0.4 作为特征变量阈值，运行次数 50 次，将运算后的均值作为变量评价的标准，每个波长被选择的可能性(selection probability, SP)如图 3 所示。在 SP 曲线中，可以发现有些特征值具有极高的 SP 值，这些峰值与小米米粉碱消值有极高的相关性。最终选出 10 条特征波段，分别为 964, 1 053, 1 171,

1 195, 1 218, 1 223, 1 242, 1 270, 1 275 和 1 279 nm。采用 RF 与 CARS 选出特征波段大部分重叠，这些特征波段与 C—H 和 O—H 键的倍频吸收有关系，并与淀粉的近红外特性相应。执行 CARS 和 RF 运行结果虽均为随机，但为保证波段选择的有效性，两种算法通过多次运算(均设定 50 次)，从而达到减少随机因素的影响<sup>[10]</sup>。

## 2.4 小米米粉碱消值的 PLSR 回归模型建立

表 2 列出采用全波段(RAW)和特征波段(CARS、RF 提取)建立的小米米粉碱消值偏最小二乘回归模型。从结果看出用 RAW 建立的 PLSR,  $R_p$  值最高为 0.77; 采用 CARS、

RF 提取特征波段建立的 PLSR,  $R_p$  值分别为 0.72 和 0.7,  $R_p$  值降低; 究其原因在于这两种方法只提取到了部分与 O—H 键和 C—H 键光谱特性有关的光谱数据<sup>[12-13]</sup>, 而导致其他相关数据信息的丢弃。通过采用 CARS、RF 提取特征波段建立的 PLSR,  $R_p$  值与运用 RAW 建立的 PLSR 接近, 这反映了采用 CARS 和 RF 建模具有可行性, 可以对小米米粉高光谱反射率进行碱消值的预测。

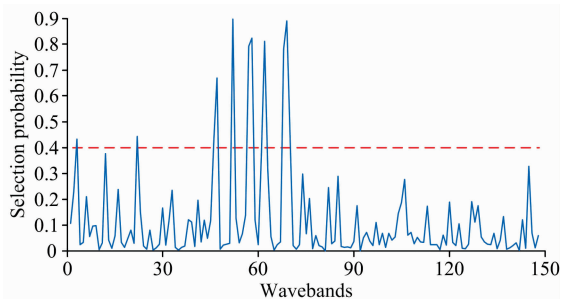


图 3 利用 RF 算法的光谱波段被选概率

Fig. 3 Selection probabilities of each wavelength using RF algorithm

表 2 不同特征波段提取方法 PLSR 模型结果

Table 2 PLSR modeling results of different methods based on key wavelengths extraction

Pretreatment	Number of variables	$R_c$	RMSEC	$R_p$	RMSEP
RAW	148	0.73	0.99	0.77	0.83
CARS	16	0.74	0.96	0.72	0.93
RF	10	0.7	1.03	0.7	0.90

为了提高模型预测精度, 基于表 2 运算结果, 对光谱数据预处理, 从表 3 中可看出, MSC 处理后光谱数据<sup>[14]</sup>  $R_p$  值达到 0.83, 精度最高。因此, 对原始光谱数据采用 MSC 预处理所得到的回归模型, 数据集的拟合程度较高, 训练集与校正集的拟合情况如图 4 所示。

表 3 不同预处理方法对原始数据 PLSR 模型分析结果

Table 3 Analysis results of PLSR models by different pretreatments

Pretreatment	$R_c$	RMSEC	$R_p$	RMSEP
S-G	0.71	1.00	0.74	0.91
MSC	0.78	0.85	0.83	0.83
S-G+MSC	0.73	0.94	0.71	1.06

对 MSC 预处理的光谱数据分别进行二次 CARS 和 RF 特征波段提取, 从表 4 中可看出, 两种算法各选择 10 条特征波段, 前者选出 964, 978, 992, 1 053, 1 166, 1 218, 1 246, 1 270, 1 275 和 1 553 nm, 后者选出 978, 992, 1 053, 1 171, 1 195, 1 218, 1 242, 1 246, 1 270 和 1 275 nm, 从以上数据

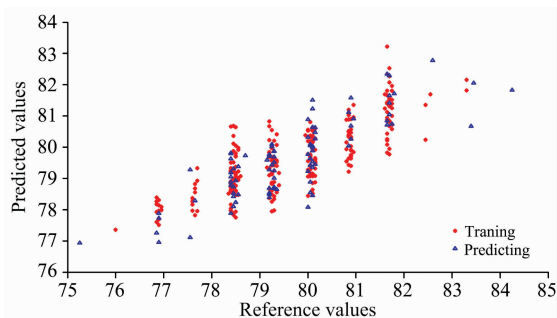


图 4 采用 MSC 预处理下的训练集和预测集拟合情况

Fig. 4 The fit of training set and prediction set pretreated by MSC

结果可以看出经 MSC 预处理后提取的特征波段数量减少, 但与未进行预处理的回归模型相比,  $R_p$  值变化不大, 这也说明 CARS 和 RF 具有一定的稳定性, 可以作为小米米粉高光谱反射率预测碱消值的参考方法。

表 4 采用 MSC 预处理下的不同特征波长提取方法 PLSR 预测模型结果

Table 4 PLSR predictive modeling results of different key wavelengths extractions pretreated by MSC

Pretreatment	Number of variables	$R_c$	RMSEC	$R_p$	RMSEP
CARS	10	0.74	0.99	0.7	0.83
RF	10	0.71	1.03	0.66	0.87

从以上数据处理结果看, 在不破坏小米米粉内部理化特性的情况下, 采用 MSC 预处理光谱数据, 建立快速检测小米米粉碱消值预测模型, 能够反演小米米粉碱消值。从而反映小米米粉中直链淀粉的占比, 进而为小米米粉品质评级、加工及小米米粉碱消值传感器的开发提供参考依据和数据支撑。

### 3 结 论

采用高光谱技术可以实现小米米粉碱消值快速检测, 进一步为小米米粉品质评级及加工提供参考依据, 主要结论为:

(1) 采用全光谱波段、CARS 和 RF 提取特征波段建立的小米米粉碱消值偏最小二乘回归模型, 结果表明三者  $R_p$  值较为相近, 全波段最高  $R_p$  为 0.77, 同时采用 CARS 和 RF 建模具有可行性, 能够利用小米米粉高光谱反射率反演其碱消值。

(2) 采用 MSC 预处理原始数据并作出 PLSR 模型分析,  $R_p$  值达到 0.83。对经 MSC 预处理的数据集, 再次作 CARS 和 RF 特征波段提取, 建立回归模型。实验结果能够为小米米粉品质评级、加工及碱消值传感器的开发提供参考依据和数据支撑。

### References

[1] Li H Y, Prakash S, Nicholson T M, et al. Food Chemistry, 2016, 196: 702.

- [ 2 ] LIU Ya-chao, LI Yong-yu, PENG Yan-kun, et al(刘亚超, 李永玉, 彭彦昆, 等). Chinese Journal of Analytical Chemistry(分析化学), 2019, 47(5): 785.
- [ 3 ] Liu S Y, Tommy Z, Wang X Y, et al. Food Hydrocolloids, 2019, 94: 217.
- [ 4 ] Shao Y Y, Xuan G T, Hu Z C, et al. IFAC-PapersOnLine, 2018, 51(17): 96.
- [ 5 ] Chen J, Ren X, Zhang Q, et al. Journal of Cereal Science, 2013, 58(2): 241.
- [ 6 ] Chen J U, Li G H. Infrared Physics & Technology, 2020, 105: e103225.
- [ 7 ] BU Xiao-pu, PENG Yan-kun, WANG Wen-xiu, et al(卜晓朴, 彭彦昆, 王文秀, 等). Food Science(食品科学), 2018, 39(16): 227.
- [ 8 ] Xia Y, Xu Y F, Li J B, et al. Artificial Intelligence in Agriculture, 2019, 1: 35.
- [ 9 ] General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China Standardization Administration(中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会). GB/T 24852—2010. Determination of the Pasting Properties of Rice—Rapid Visco Analyzer Method(大米及米粉糊化特性测定—快速粘度仪法).
- [10] Yu K Q, Zhao Y R, Liu Z Y, et al. Food Bioprocess Technology, 2014, 7: 3077.
- [11] Sun J J, Yang W D, Feng M C, et al. RSC Advances, 2020, 10(28): 16245.
- [12] Lu X, Wei S, Chen B C, et al. LWT-Food Science and Technology, 2015, 61: 590.
- [13] Kittisak P, Vasu U, Panmanas S. Power Technology, 2019, 345: 608.
- [14] Lei F F, Yang Y H, Zhang J, et al. Chemometrics and Intelligent Laboratory Systems, 2019, 191: 158.

## Hyperspectral Technique Coupled With Chemometrics Methods for Predicting Alkali Spreading Value of Millet Flour

WANG Guo-liang<sup>1,2</sup>, YU Ke-qiang<sup>3</sup>, CHENG Kai<sup>2</sup>, LIU Xin<sup>2</sup>, WANG Wen-jun<sup>1</sup>, LI Hong<sup>2</sup>, GUO Er-hu<sup>2</sup>, LI Zhi-wei<sup>1\*</sup>

1. College of Agricultural Engineering, Shanxi Agricultural University, Taigu 030801, China

2. Millet Research Institute, Shanxi Agricultural University, Changzhi 046000, China

3. College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling 712100, China

**Abstract** As the main ingredient of millet flour, the quality of starch determined the market price of millet flour. Gelatinization characteristic is one of the most important physical characteristics of millet flour, and the alkali spreading value is the main index that reflects the gelatinization characteristic directly. The differences in the alkali spreading the value of millet flour show the quality of amylose content. When the alkali spreading value becomes lower, on the contrary, the gelatinization temperature and amylose content become higher, eventually the lower the waxy of millet flour. This study employed the hyperspectral technique could with chemometrics methods to develop an approach for detecting the alkali spreading the value of millet flour, whose aim is to explore a rapid, nondestructive and low-cost method for predicting the alkali spreading the value of millet flour. First, the hyperspectral data of millet flour were collected, then the hyperspectral data matrix in the region of interest (ROI) in each pixel was computed. The results were meant in each wavelength of every sample. Then we used the rapid visco analyser (RVA) to measured the alkali spreading the value of millet flour. In the data processing, partial least square regression (PLSR) models were made after using competitive adaptive reweighted sampling(CARS) and random frog (RF) to extracted key wavelengths. The results showed that the highest predicted  $R_p$  was 0.77 in the PLSR of the full wavelengths, and that explained that the reflectance of millet flour could invert the alkali spreading the value of millet flour. The  $R_p$  in the other two methods were 0.72 and 0.7, and both were close to the previous result, these illustrated it was feasible to build the PLSR using CARS and RF. In order to improve the predicting accuracy, the full wavelengths were preprocessed by Savitzky-Golay (S-G), multiplicative scatter correction(MSC) and S-G+MSC. The performance of the PLSR model was better by using MSC predicted the full wavelengths ( $R_p=0.83$ ). Then built the PLSR model again after extracting key wavelengths using CARS and RF, compared with the models without pretreatment, the  $R_p$  does not change much, which also shows that CARS and RF have a certain stability and can be used as reference methods for predicting the alkali spreading the value of the hyperspectral reflectance of millet flour. The results showed that the reflectance of millet flour could predict its alkali spreading value by using hyperspectral. This could supply a rapid, nondestructive and low-cost method of the alkali spreading value of millet flour, then provided the theoretical foundation for the rating, processing and alkali spreading value sensor of millet flour.

**Keywords** Hyperspectral; Millet of flour; The alkali spreading value; Chemometrics methods

\* Corresponding author

(Received Sep. 25, 2020; accepted Jan. 14, 2021)