

基于高光谱技术及SPXY和SPA的玉米毒素检测模型建立

于慧春, 娄楠, 殷勇*, 刘云宏
(河南科技大学食品与生物工程学院, 河南 洛阳 471023)

摘要: 应用高光谱技术研究和构建霉变玉米黄曲霉毒素B₁ (aflatoxin B₁, AFB₁) 和玉米赤霉烯酮 (zearalenone, ZEN) 含量的检测方法, 通过建立霉变玉米中这两种毒素含量的预测模型, 实现对玉米霉变程度的快速、无损、准确判别。首先, 通过对比5种预处理方法, 确定标准正态变量校正法对原始光谱数据进行预处理; 然后, 采用光谱-理化值共生距离算法结合偏最小二乘回归 (partial least squares regression, PLSR) 法分析不同校正集样本预测AFB₁和ZEN含量的差异, 并分别优选出130个和140个校正集样本; 在采用均匀光谱间隔法对原始光谱变量进行初降维的基础上, 对比连续投影算法 (successive projections algorithm, SPA) 和竞争性自适应重加权算法2种变量提取法。结果表明: 经SPA分别筛选出17个特征波段且基于较少校正集样本和特征波长光谱信息建立的PLSR模型能够获得较优的预测结果, 对应AFB₁和ZEN含量预测集的相关系数和均方根误差 (root mean square error of prediction, RMSEP) (R_{pre}^2 , RMSEP) 由最初的 (0.994 4, 0.984 6) 和 (0.991 6, 2.320 9) 分别变为 (0.997 3, 0.681 5) 和 (0.997 7, 1.144 1), 在降低模型复杂度的情况下提高了预测精度, 表明该模型对这两种毒素含量能够实现较强的预测能力。因此, 利用高光谱技术对玉米AFB₁和ZEN含量实施无损检测具有可行性。

关键词: 玉米; 高光谱; 无损检测; 黄曲霉毒素B₁; 玉米赤霉烯酮

Predictive Model for Detection of Maize Toxins with Sample Set Partitioning Based on Joint x-y Distance (SPXY) Algorithm and Successive Projections Algorithm (SPA) Based on Hyperspectral Imaging Technology

YU Huichun, LOU Nan, YIN Yong*, LIU Yunhong
(College of Food and Bioengineering, Henan University of Science and Technology, Luoyang 471023, China)

Abstract: In this paper, a rapid, non-destructive and accurate method for detecting the contents of aflatoxin B₁ (AFB₁) and zearalenone (ZEN) in maize using hyperspectral imaging was established by developing predictive models. Among 5 spectral pretreatments tested, standard normal variate (SNV) was found to be the best method for preprocessing the original spectral data. Sample set partitioning based on joint x-y distance (SPXY) algorithm combined with partial least squares regression (PLSR) was used to screen the differences in the predicted contents of AFB₁ and ZEN from different calibration set samples, and 130 and 140 calibration set samples were selected for AFB₁ and ZEN, respectively. On the basis of dimensionality reduction by the uniform spectral spacing (USS) method, the two variable extraction methods: successive projections algorithm (SPA) and competitive adaptive reweighted sampling algorithm (CARS) were compared. The results showed that 17 characteristic wavelengths were selected for AFB₁ and ZEN, respectively and the PLSR model established based on fewer calibration set samples with the characteristic wavelengths had better predictive performance. The correlation coefficients (R_{pre}^2) and root mean square error of prediction (RMSEP) for AFB₁ and ZEN content were 0.997 3 and 0.681 5, and 0.997 7 and 1.144 1, respectively, while the initial R_{pre}^2 and RMSEP values were 0.994 4 and 0.984 6, and 0.991 6 and 2.320 9, respectively. The results indicated that despite reducing the reducing the complexity of the model, the predictive accuracy could be improved. Therefore, it is feasible to non-destructively predict the AFB₁ and ZEN content in maize by applying hyperspectral imaging technology.

Keywords: maize; hyperspectral imaging technology; non-destructive detection; aflatoxin B₁; zearalenone

DOI:10.7506/spkx1002-6630-2018016048

收稿日期: 2017-06-05

基金项目: 河南省科技攻关项目 (172102210256; 172102310617); 河南省教育厅自然科学研究项目 (13A550269)

第一作者简介: 于慧春 (1977—), 女, 副教授, 博士, 研究方向为农产品、食品品质无损检测技术。E-mail: yu_kin_le@126.com

*通信作者简介: 殷勇 (1966—), 男, 教授, 博士, 研究方向为农产品、食品品质无损检测技术。E-mail: yinyong@mail.haust.edu.cn

中图分类号: O433

文献标志码: A

文章编号: 1002-6630 (2018) 16-0328-08

引文格式:

于慧春, 娄楠, 殷勇, 等. 基于高光谱技术及SPXY和SPA的玉米毒素检测模型建立[J]. 食品科学, 2018, 39(16): 328-335. DOI:10.7506/spkx1002-6630-2018016048. <http://www.spkx.net.cn>

YU Huichun, LOU Nan, YIN Yong, et al. Predictive model for detection of maize toxins with sample set partitioning based on joint x-y distance (SPXY) algorithm and successive projections algorithm (SPA) based on hyperspectral imaging technology[J]. Food Science, 2018, 39(16): 328-335. (in Chinese with English abstract) DOI:10.7506/spkx1002-6630-2018016048. <http://www.spkx.net.cn>

新鲜玉米在存贮过程中, 由于其胚部大、水分含量高、带菌量多, 高温高湿环境下极易霉变^[1], 不仅给经济造成重大损失, 而且霉变玉米在代谢过程中会产生多种对人体具有极强致病性、致癌性的毒素^[2-3], 危害人畜健康。黄曲霉毒素 B_1 (aflatoxin B_1 , AFB_1) 和玉米赤霉烯酮 (zearalenone, ZEN) 是其中2种比较稳定的代谢产物, 容易在霉变玉米中积累, 导致含量升高, AFB_1 和ZEN的多少与玉米霉变情况密切相关^[4]。因此, 可以通过监测玉米中 AFB_1 和ZEN含量变化表征玉米的霉变情况, 实现对玉米霉变程度的准确评价^[5]。但是常规的 AFB_1 和ZEN检测方法存在样品处理繁琐、费时、对样品有破坏性等缺点, 难以实现简单、快速、无损检测, 无法满足实际生产的需要^[6-8]。

高光谱技术通过提取被测对象的图像和光谱信息特征并与其内外品质指标间建立联系, 从而实现对其内外综合品质的评价, 因而在果蔬品质^[9-11]、禽肉品质^[12-14]、蛋^[15]和烟叶^[16]等农产品检测领域获得广泛应用研究。在霉变玉米的高光谱检测方面, 也有学者进行了一些探索^[17-19], 但目前研究主要集中在高光谱技术对不同霉变程度玉米的定性判别方面, 而对霉变玉米中 AFB_1 和ZEN等的定量检测分析方面尚鲜有报道。袁莹^[20]和褚璇^[21]等将不同浓度的 AFB_1 溶液滴在玉米籽粒上, 然后利用高光谱技术进行检测, 但其最终的分析结果仍然只是对表面含有不同浓度 AFB_1 的玉米样品进行分类识别, 也未尝试通过高光谱技术对测试样本在特定条件下自身代谢生成的 AFB_1 含量进行预测。此外, 上述霉变玉米高光谱技术检测研究中, 对所使用的校正集样本质量也未进行考察, 大量的样本数据之间可能存在相同或者差异过小的情况, 从而影响分析结果精度和耗费建模时间。

本研究以玉米霉变过程中产生的 AFB_1 和ZEN 2种代谢产物含量为玉米霉变程度的表征指标, 通过测定不同霉变程度玉米样本的光谱信息和相应的 AFB_1 和ZEN含量, 建立基于较少校正集样本和特征波长光谱信息的 AFB_1 和ZEN含量偏最小二乘回归 (partial least squares regression, PLSR) 预测模型, 从而实现了对玉米霉变程度的准确分析, 以期为实现玉米霉变的在线、快速、精确检测提供借鉴。

1 材料与方法

1.1 材料

玉米来自于2016年10月在洛阳当地收获, 品种为中单909, 购于洛阳市中原农贸城, 由实验室自行培育出不同霉变程度的玉米。

1.2 仪器与设备

LHS-HC-100恒温恒湿培养箱 上海资一仪器设备有限公司; IST 50-3810高光谱成像光谱仪 德国Inno-spec公司; 康标达镜头 日本Computar公司; RK90000420108线性卤素灯 (500 W) 德国Esylux公司。

高光谱图像采集系统如图1所示, 该系统主要包括高光谱成像光谱仪、康标达镜头、4个500 W线性卤素灯、实验室自制传送装置和计算机等主要部件。其中高光谱成像光谱仪由摄像机和光谱仪2部分组成, 摄像机为CCD相机, 光谱摄像机为可见-近红外光谱仪, 光谱波长范围为371.05~1 023.82 nm, 可采集到的光谱数据点数为1 288个, 光谱分辨率为2.8 nm, 传送带移动速率1.25 mm/s, 曝光时间100 ms, 样本与镜头的距离为285 mm。高光谱成像仪通过USB 2.0接口数据线连接计算机, 通过SICap-STVR V1.0.x软件平台驱动控制成像仪, 及时记录和存贮高光谱数据。

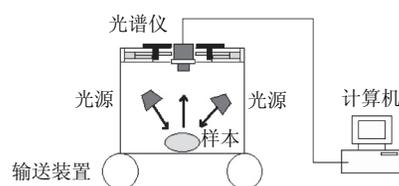


图1 高光谱图像采集系统

Fig. 1 Schematic presentation of the hyperspectral image acquisition system

1.3 方法

1.3.1 玉米样本制备

将新鲜玉米放入恒温恒湿培养箱, 在培养温度30℃、相对湿度90%^[22]的条件下进行培养。培养时间不

同,玉米霉变程度也不同,本实验用培养初始(新鲜玉米)、第2天、第4天、第6天、第8天标记玉米的不同霉变程度,5个霉变程度的玉米看作5个霉变等级,每个等级样品各取50个测试样本,每个样本含60 g(±0.1 g)玉米。

1.3.2 AFB₁和ZEN含量的测定

为了减少AFB₁和ZEN的损失,制备的样本在完成光谱数据采集后立即按照GB/T 18979—2003《食品中黄曲霉素B₁的测定》^[23]和GB/T 5009.209—2008《食品中玉米赤霉烯酮的测定》^[24]分别测定AFB₁和ZEN含量,并将其测定结果作为建模的标准参考值。

1.3.3 高光谱图像数据采集与校正

在进行高光谱图像采集时,为了满足实际玉米霉变检测的需要,将玉米样本散乱均匀平铺在规格为10 cm×10 cm的白色载物盒中,然后将装有玉米的载物盒放置在移动载物平台上,依次对玉米样本进行高光谱图像采集,设定高光谱的扫描图像大小为550×600像素。本实验共采集到250幅玉米样本高光谱图像。

同时,为了消除光源强度分布不均以及光谱暗电流噪声的影响,对每个样本按校正公式进行黑白校正,计算公式如下:

$$R = \frac{R_0 - B}{W - B}$$

式中: R 为校正后的样本高光谱图像; R_0 为样本原始高光谱图像; B 为全黑标定图像; W 为全白标定图像^[25]。样本的高光谱图像采集在SICap-STVR V1.0.x软件平台上完成,实验数据分析与处理在MATLAB 2014a(美国The Math Works公司)和遥感图像处理平台ENVI5.1(美国Boulder公司)两个软件上完成。

1.4 数据处理

1.4.1 原始光谱数据进行预处理

为了减少背景噪声、杂散光等无用信息对原始光谱数据的干扰,需要对原始光谱数据进行预处理。本实验对比标准化、变量标准化、多元散射校正(multiplicative scatter correction, MSC)、标准正态变量校正(standard normal variate, SNV)和卷积平滑(Savitzky-Golay smoothing, S-G)5种预处理方法。为保证模型校正集最大程度地表征样本均匀分布,提高模型稳定性,本实验采用光谱-理化值共生距离(sample set partitioning based on joint x-y distance, SPXY)算法^[26]划分模型的初始校正集和预测集。在此基础上,为降低或者消除校正集样本间的共线性^[27],简化模型运算量,采用SPXY算法结合PLSR法分析不同校正集样本子集预测AFB₁和ZEN含量的差异,从而进一步对划分的初始校正集样本进行优选。在采用均匀光谱间隔(uniform spectral spacing, USS)

法对原始光谱变量进行初步筛选的基础上对比连续投影算法(successive projections algorithm, SPA)^[28]和竞争性自适应重加权算(competitive adaptive reweighted sampling, CARS)法^[29]2种特征波长提取方法,最大程度地剔除原始光谱矩阵中的冗余信息,提高模型精度,减少模型运算量。

1.4.2 模型建立与评价

霉变玉米的光谱信息与其AFB₁和ZEN含量之间的关系属于非确定性问题,采用回归分析构造变量间的数理统计模型可用于该类问题的研究分析^[30]。PLSR法集主成分分析、普通多元线性回归和相关分析于一体,它在描述光谱矩阵 X 变量的同时也描述了指标矩阵 Y 变量,较好地解决了自变量的多重共线性问题,在光谱分析领域得到了广泛应用。

模型的评价指标由相关系数(包括校正集相关系数 R_{cal}^2 和预测集相关系数 R_{pre}^2)和均方根误差(root mean square error, RMSE)(包括校正集均方根误差(root mean square error calibration, RMSEC))和预测集均方根误差(root mean square error prediction, RMSEP)决定,相关系数 R_{cal}^2 和 R_{pre}^2 越大越好, RMSEC和RMSEP越小越好。

2 结果与分析

2.1 不同霉变等级的玉米AFB₁和ZEN含量

在进行AFB₁和ZEN含量检测时,对每类玉米样品的50个平行样本进行测定,以独立测定结果的绝对值差不超过算术平均值的10%为准,并将其平均值作为此类玉米样品的实际指标值,结果如表1所示。由表1可知,第4、5个等级的玉米中,这2类毒素含量已超过国家标准(AFB₁≤20 μg/kg, ZEN≤60 μg/kg)。

表1 5个霉变等级玉米样品中AFB₁和ZEN含量
Table 1 AFB₁ and ZEN values in 5 grades of moldy maize samples

霉变等级	AFB ₁ / (μg/kg)			ZEN / (μg/kg)		
	变化范围	均值	标准差	变化范围	均值	标准差
1	4.42~4.98	4.68	0.197	9.08~9.92	9.4	0.276
2	8.42~9.31	8.96	0.305	24.12~25.52	24.8	0.352
3	16.12~16.98	16.68	0.338	49.11~51.72	50.2	0.791
4	30.06~30.32	30.18	0.355	68.25~70.52	69.2	0.966
5	38.86~40.84	39.96	0.722	74.05~76.55	75.3	1.106

2.2 光谱反射值曲线

提取每个玉米样本在371.05~1 023.83 nm波长范围内的高光谱图像的平均光谱反射值,结果如图2所示,371.05~480.55 nm和999.66~1 023.83 nm两个波段受噪声的影响较大,因此剔除了这2个波段,最终采用的波段范围为481.06~999.15 nm。由图3可以看出,5个等级玉

米样品光谱曲线变化趋势基本相似，但不同霉变等级玉米的光谱值有差别，总体看来，5个等级玉米样本具有一定的可分性，这为预测建模提供了可能。

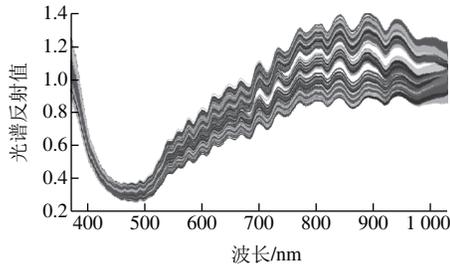


图2 250个玉米样本的光谱反射值曲线

Fig. 2 Spectral reflectance curves of 250 maize samples

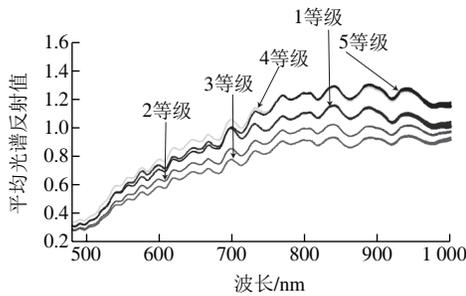


图3 5类玉米样本的平均光谱反射值曲线

Fig. 3 Average spectral reflectance curves of five grades of maize samples

2.3 光谱数据预处理

随机选取200个样本作为校正集，剩余的50个样本作为预测集。采用不同预处理方法对原始光谱数据进行预处理，并基于各预处理后的光谱数据建立PLSR模型，结果如表2所示。由表2可知，与基于原始光谱数据的预测结果相比，5种预处理方法中，基于SNV建立的PLSR模型对这2种毒素含量的预测效果最好，对应AFB₁和ZEN含量的预测集相关系数和均方根误差 (R^2_{pre} , RMSEP) 分别为 (0.994 4, 0.984 6) 和 (0.991 6, 2.320 9)，因此确定SNV为预处理方法。

表2 基于不同预处理方法的PLSR建模结果

Table 2 Results of PLSR models based on different pretreatments

预处理方法	AFB ₁				ZEN			
	校正集		预测集		校正集		预测集	
	R^2_{cal}	RMSEC	R^2_{pre}	RMSEP	R^2_{cal}	RMSEC	R^2_{pre}	RMSEP
原始	0.998 5	0.510 9	0.985 5	1.587 1	0.998 9	0.838 9	0.982 9	3.313 6
标准化	0.998 5	0.510 9	0.985 5	1.587 1	0.998 9	0.838 9	0.982 9	3.313 6
变量标准化	0.998 4	0.526 6	0.993 3	1.077 0	0.999 0	0.809 3	0.989 5	2.595 7
MSC	0.998 3	0.538 6	0.993 5	1.062 3	0.999 0	0.796 5	0.991 3	2.362 7
SNV	0.998 2	0.557 3	0.994 4	0.984 6	0.998 9	0.832 4	0.991 6	2.320 9
S-G	0.998 4	0.534 2	0.985 4	1.593 4	0.998 8	0.893 5	0.981 2	3.474 9

2.4 校正集样本的优选

2.4.1 SPXY算法划分样本集数据

校正集样本的划分在一定程度上能够决定所建模型

的预测性能，本实验采用SPXY算法对样本集进行划分。首先设定初始校正集样本个数为200个，剩余的50个样本为预测集样本，SPXY算法的详细步骤参考文献[26]，划分结果如表3所示，不同霉变等级的样品间，校正集和预测集的样本个数存在较明显的差别。

表3 SPXY划分样本集结果

Table 3 Calibration and prediction set partitioned by SPXY

等级	AFB ₁		ZEN	
	校正集个数	预测集个数	校正集个数	预测集个数
1	44	6	44	6
2	36	14	38	12
3	39	11	39	11
4	43	7	43	7
5	38	12	36	14

2.4.2 SPXY算法优选校正集样本

针对表3中初分的初始校正集样本，采用SPXY算法进行进一步优选，以尽量有效地降低样本间的共线性。确定样本数N的范围为80~200，步长为10，校正集样本优选过程中，分别基于优选出的校正集样本子集建立PLSR模型，其预测集的相关系数 R^2_{pre} 和RMSEP的变化如图4所示。

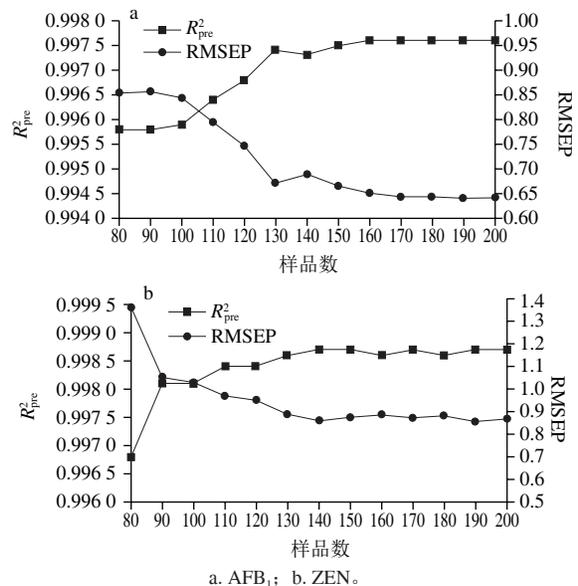


图4 R^2_{pre} 和RMSEP随校正集样本数量变化曲线
Fig. 4 Plots of R^2_{pre} and RMSEP values versus number of calibration set samples

由图4可以看出， R^2_{pre} 随样本数N的增加呈递增趋势，但总体变化不明显(0.995~0.999)，RMSEP曲线呈递减趋势，总体变化较明显。图4a中，对于RMSEP变化曲线，N在80~130范围内取值时，RMSEP变化差异明显，N在130~200范围内取值时，RMSEP的变化趋于平缓，N取值为130为该曲线的拐点，该点对应的 R^2_{pre} 和RMSEP为0.997 4和0.672 0，当N取值为190时， R^2_{pre} 达到

极大值0.997 6, RMSEP降为极小值0.641 4, 与 N 取130时对应的 R_{pre}^2 和RMSEP相比, R_{pre}^2 仅增加0.000 2, RMSEP仅减少0.030 6, 从数值上看, 两者差异较小, 但校正集样本数增加了60个, 因此, 为简化模型运算量, 综合图4a中 R_{pre}^2 和RMSEP曲线的变化趋势, 对于 AFB_1 , 最终从初始校正集中优选出130个样本组成模型校正集。同样根据图4b, 对于ZEN, 最终从初始校正集中优选出140个样本组成模型校正集, 对应的 R_{pre}^2 和RMSEP为0.998 7和0.862 1。

经SPXY算法划分的预测集样本和初始校正集样本以及经SPXY算法优选的校正集样本的 AFB_1 和ZEN含量变化情况如表4所示。由表4可知, 无论是对 AFB_1 还是对ZEN, SPXY法优选出的校正集样本的各指标值含量变化范围在初始校正集范围内, 且标准差相近, 证明SPXY算法优选后的样本具有一定的代表性。

表4 不同样本集玉米 AFB_1 和ZEN含量分布Table 4 Distribution profiles of AFB_1 and ZEN values in different sample sets

玉米类型	AFB_1 样本数	AFB_1 含量/ ($\mu\text{g}/\text{kg}$)			ZEN 样本数	ZEN 含量/ ($\mu\text{g}/\text{kg}$)		
		变化范围	均值	标准差		变化范围	均值	标准差
预测集	50	4.42~40.62	20.01	13.18	50	9.24~75.55	45.56	25.48
初始校正集	200	4.42~40.84	20.12	13.13	200	9.08~76.55	45.78	25.40
SPXY优选校正集	130	4.42~40.80	22.32	13.21	140	9.12~76.55	46.39	25.50

2.5 特征光谱的选择

2.5.1 光谱数据的降维

高光谱图像具有较高的光谱分辨率, 容易对被测物进行分辨, 但也会导致数据量的剧增和数据冗余。相邻波段间差异很小, 直接进行特征波段的提取可能会漏掉某些有用信息, 因此, 本实验首先对光谱变量进行初降维。USS法通过控制步长选出间隔均匀、相关性低、信息量大的少量波段, 从而可对光谱数据进行有效降维。在481.06~999.15 nm范围内, 分别以正整数2、3、4、5、6……20为间隔, 从1 023个经预处理后的光谱波段中抽取光谱组成新的光谱矩阵并建立PLSR模型, 根据模型的 R_{pre}^2 和RMSEP, 确定最佳的步长间隔。表5为利用不同步长间隔下抽取的光谱建立的PLSR模型对 AFB_1 和ZEN含量的预测结果。

从表5可以看出, 当间隔为8, 波段为128个时, 模型对 AFB_1 含量的预测效果较好, R_{pre}^2 和RMSEP为0.997 4和0.673 4; 当间隔为7, 波段为147个时, 模型对ZEN含量的预测效果较好, R_{pre}^2 和RMSEP为0.998 8和0.835 4。对比只采用SPXY算法优选校正集样本而未对光谱变量进行筛选所建的PLSR模型结果可知, 通过USS法进行变量筛选后所建模型对 AFB_1 含量的预测精度变化不大, 对ZEN含量的预测精度有一定的提高, 但参与建模的变量数能显著减少。

表5 基于不同间隔光谱数据建立PLSR模型预测结果

Table 5 Predictive results of PLSR models with different numbers of spectral intervals

间隔数	变量数	AFB_1			ZEN		
		因子数	R_{pre}^2	RMSEP	因子数	R_{pre}^2	RMSEP
1	1 023	18	0.997 4	0.672 0	18	0.998 7	0.862 1
2	512	17	0.997 0	0.724 4	18	0.998 4	0.940 9
3	341	17	0.997 0	0.723 0	18	0.998 6	0.887 1
4	256	17	0.997 0	0.718 4	18	0.998 5	0.930 3
5	205	16	0.997 2	0.704 2	18	0.998 5	0.910 3
6	171	16	0.997 3	0.682 1	18	0.998 4	0.969 9
7	147	16	0.997 2	0.699 1	18	0.998 8	0.835 4
8	128	16	0.997 4	0.673 4	18	0.998 5	0.925 1
9	115	16	0.997 0	0.717 9	16	0.998 0	1.062 2
10	103	16	0.997 0	0.721 6	16	0.997 6	1.181 2
...
15	69	16	0.997 1	0.711 1	16	0.997 6	1.161 0
...
20	52	16	0.997 0	0.726 4	16	0.997 6	1.170 5

2.5.2 光谱特征波段的选取

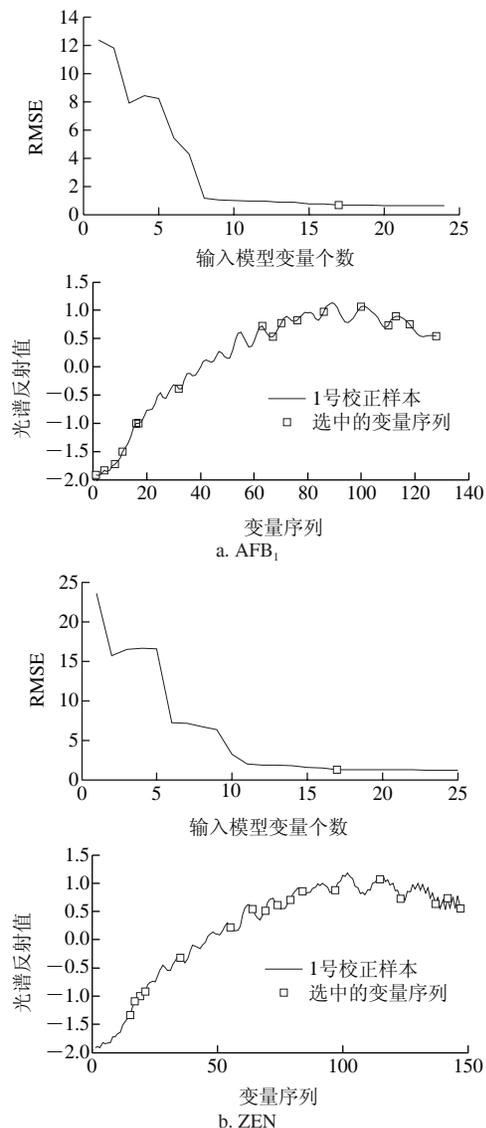


图5 基于SPA的变量筛选

Fig. 5 Wavelengths selected by SPA

为进一步实现霉变玉米在线检测的检测时间短、准确率高的要求,本实验在USS法筛选波长变量的基础上,对比了SPA和CARS 2种特征波长选择方法。

在SPA中,最小的交互验证RMSE对应的波长变量个数即为最终的选择结果,指定波段数 N 的范围为2~25,SPA的具体运算步骤可参考文献[28]。由图5a可以看出,从128个波长变量中提取17个波长时RMSE达到最小;由图5b可以看出,从147个波长变量中提取17个波长时RMSE达到最小。表6为基于SPA提取的特征波长建立的AFB₁和ZEN含量的PLSR模型预测结果。

在CARS算法中,设定蒙特卡罗采样次数为50,CARS算法的具体运算步骤参考文献[29],如图6所示,基于CARS算法针对AFB₁和ZEN含量的变量筛选过程,对这2种毒素指标值,当采样次数分别为14次和19次时,对应的特征波长个数分别为18个和19个。

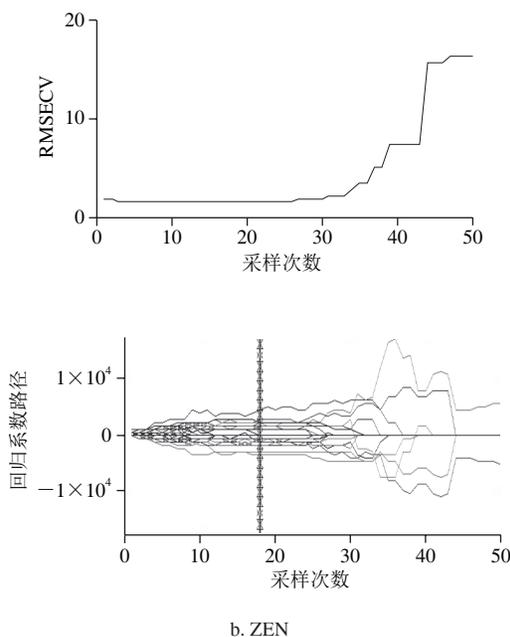
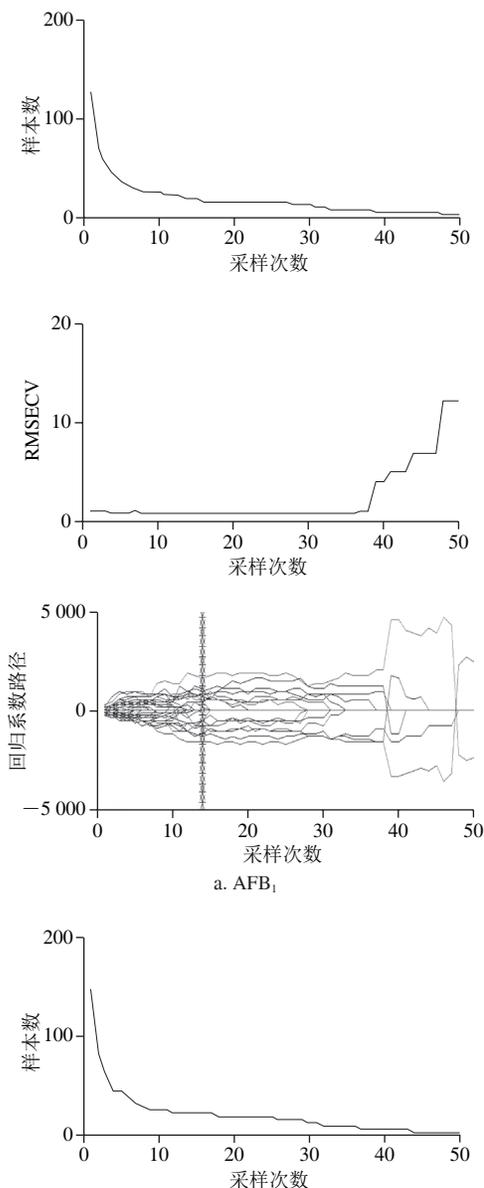


图6 基于CARS的变量筛选
Fig. 6 Wavelengths selected by CARS

表6 基于SPA特征波长提取方法的PLSR模型结果
Table 6 Parameters of PLSR models with characteristic wavelengths extracted by SPA

毒素	特征波长个数	因子数	R^2_{pre}	RMSEP
AFB ₁	17	15	0.997 3	0.681 5
ZEN	17	15	0.997 7	1.144 1

表7 基于CARS特征波长提取方法的PLSR模型结果
Table 7 Parameters of PLSR models with characteristic wavelengths extracted by CARS

毒素	特征波长个数	因子数	R^2_{pre}	RMSEP
AFB ₁	18	15	0.996 1	0.849 1
ZEN	19	15	0.997 4	1.267 6

对比表6和表7可知,基于SPA筛选出的波长所建立的PLSR模型对AFB₁和ZEN含量的预测效果较CARS算法好。

2.6 模型分析

原始光谱数据经过SNV预处理后,采用SPXY算法对校正集样本进行划分与优选,运用USS法结合SPA对光谱数据进行筛选,然后建立基于优选后的校正集样本及特征波长的PLSR模型。如图7所示,波长优选后,PLSR模型的预测能力依旧表现良好,AFB₁含量的预测精度(R^2_{pre} , RMSEP)为(0.997 3, 0.681 5),与优选之前的预测精度(0.997 4, 0.672 0)相近;ZEN含量的预测精度(R^2_{pre} , RMSEP)为(0.997 7, 1.144 1),略低于优选前的精度(0.998 7, 0.862 1)。

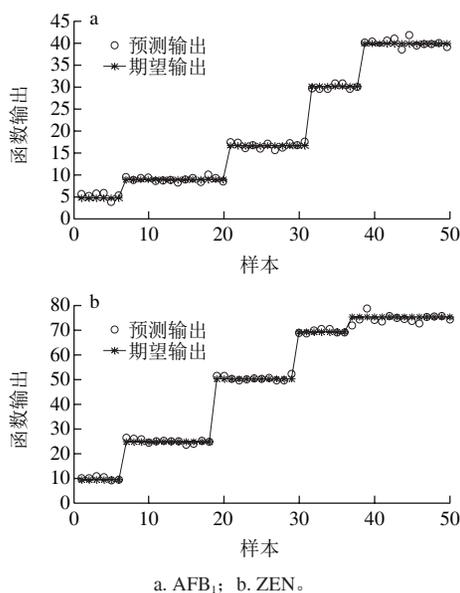


图7 基于特征波段建立PLSR模型预测结果
Fig. 7 Predictive results of PLSR models based on characteristic wavelengths

3 结论

本实验利用不同霉变时期的玉米光谱实验数据,研究了高光谱技术用于霉变玉米中AFB₁和ZEN含量检测模型的构建,并得到以下结论:

1) 对于AFB₁和ZEN含量,经SPXY法分别优选出的130个和140个校正集样本的指标值变化范围在初始校正集范围内,且标准差相近,说明SPXY算法优选的样本具有一定的代表性。

2) 基于SPXY法划分的初始校正集样本建立的PLSR模型预测精度明显高于划分之前的预测精度,且通过SPXY法对初始校正集样本进一步优选,PLSR模型精度基本不变,但校正集样本数减少为初始校正集的65%和70%,在保证模型稳健性的前提下,有效降低了校正模型的复杂性。

3) 利用USS法结合SPA进行特征波段选择,最终从1 023个波段中分别筛选出17个波段。分别建立2种毒素值基于各自特征波段的PLSR模型,结果显示AFB₁含量预测精度(0.997 3, 0.681 5)与优选前的预测精度(0.997 4, 0.672 0)相差不大;ZEN含量预测精度(0.997 7, 1.144 1)略低于优选之前的预测精度(0.998 7, 0.862 1), R^2_{pre} 从0.998 7降低到0.997 7, RMSEP从0.862 1上升到1.144 1,但波长减少为17个,在保证模型精度的前提下,实现光谱数据的压缩。

综合上述研究结果,基于SPXY算法和SPA建立的高光谱检测模型,可以实现对霉变玉米中AFB₁和ZEN含量的准确预测。

参考文献:

- [1] 矫江,王连敏,王立志,等.黑龙江省玉米霉变粒发生原因与防御对策[J].中国农学通报,2014(30):228-233. DOI:10.11924/j.issn.1000-6850.2014-1442.
- [2] 刘小云.玉米霉变引起中毒的疾病预防和治疗[J].农业科技与信息,2016(23):66-67. DOI:10.3969/j.issn.1003-6997.2016.23.048.
- [3] 董继红,高鹏飞,臧雪燕.玉米霉变对养殖业的危害及应对措施[J].畜禽业,2004(5):32-33. DOI:10.3969/j.issn.1008-0414.2004.05.023.
- [4] 赵炎,张乃建,王若兰.霉变玉米真菌毒素含量与图像颜色特征参数之间的相关性研究[J].粮食与饲料工业,2015,12(12):21-26. DOI:10.7633/j.issn.1003-6202.2015.12.006.
- [5] 殷勇,郝银凤,于慧春.基于多特征融合的电子鼻鉴别玉米霉变程度[J].农业工程学报,2016,32(12):254-260. DOI:10.11975/j.issn.1002-6819.2016.12.036.
- [6] BARBEDO J G A, TIBOLA C S, FERNANDES J M C. Detection of fusarium head blight in wheat kernels using hyperspectral imaging[J]. Biosystems Engineering, 2015, 131: 65-76. DOI:10.1016/j.biosystemseng.2015.01.003.
- [7] PEI R L, LIU J P, WANG Y. Determination of zearalenone in corn by HPLC with molecular imprinting affinity purification[J]. Advanced Materials Research, 2013, 718/720: 471-473. DOI:10.4028/www.scientific.net/AMR.718-720.471.
- [8] 谢刚,王松雪,张艳.超高效液相色谱法快速检测粮食中黄曲霉毒素的含量[J].分析化学,2013,41(2):223-228. DOI:10.3724/SP.J.1096.2013.20601.
- [9] WANG N, SUN D W, YANG Y C, et al. Recent advances in the application of hyperspectral imaging for evaluating fruit quality[J]. Food Analytical Methods, 2016, 9(1): 1-14. DOI:10.1007/s12161-015-0153-3.
- [10] SU W H, SUN D W. Potential of hyperspectral for visual authentication of sliced organic potatoes from potato and sweet potato tubers and rapid grading of the tubers according to moisture proportion[J]. Computers & Electronics in Agriculture, 2016, 125: 113-124. DOI:10.1016/j.compag.2016.04.034.
- [11] 孙俊,金夏明,毛罕平,等.基于高光谱图像光谱与纹理信息的生菜氮素含量检测[J].农业工程学报,2014,30(10):167-173. DOI:10.3969/j.issn.1002-6819.2014.10.021.
- [12] KAMRUZZAMAN M, MAKINO Y, OSHITA S, et al. Assessment of visible near-infrared hyperspectral imaging as a tool for detection of horsemeat adulteration in minced beef[J]. Food and Bioprocess Technology, 2015, 8(5): 1054-1062. DOI:10.1007/s11947-015-1470-7.
- [13] XIONG Z, SUN D W, DAI Q, et al. Application of visible hyperspectral imaging for prediction of springiness of fresh chicken meat[J]. Food Analytical Methods, 2015, 8(2): 380-391. DOI:10.1007/s12161-014-9853-3.
- [14] 刘善梅,李小昱,钟雄斌,等.基于高光谱成像技术的生鲜猪肉含水率无损检测[J].农业机械学报,2013(增刊1):165-170. DOI:10.6041/j.issn.1000-1298.2013.S1.030.
- [15] ZHANG W, PAN L Q, TU S C, et al. Non-destructive internal quality assessment of eggs using a synthesis of hyperspectral imaging and multivariate analysis[J]. Journal of Food Engineering, 2015, 157: 41-48. DOI:10.1016/j.jfoodeng.2015.02.013.
- [16] YIN Y, XIAO Y J, YU H C. An image selection method for tobacco leave grading based on image information[J]. Engineering in Agriculture, Environment and Food, 2015, 8: 148-154. DOI:10.1016/j.eaef.2015.01.005.
- [17] 袁莹,王伟,褚璇,等.光谱特征波长的SPA选取和基于SVM的玉米颗粒霉变程度定性判别[J].光谱学与光谱分析,2016,36(1):226-230. DOI:10.3964/j.issn.1000-0593(2016)01-0226-05.

- [18] KANDPAL L M, LEE S, KIM M S, et al. Short wave infrared (SWIR) hyperspectral imaging technique for examination of aflatoxin B₁ (AFB₁) on corn kernels[J]. *Food Control*, 2014, 51: 171-176. DOI:10.1016/j.foodcont.2014.11.020.
- [19] WEI W, HEITSCHMIDT G W, WINDHAM W R, et al. Feasibility of detecting aflatoxin B₁ on inoculated maize kernels surface using Vis/NIR hyperspectral imaging[J]. *Journal of Food Science*, 2015, 80(1): 116-122. DOI:10.1111/1750-3841.12728.
- [20] 袁莹, 王伟, 褚璇, 等. 基于高光谱成像技术和因子判别分析的玉米黄曲霉毒素检测研究[J]. *中国粮油学报*, 2014, 29(12): 107-110.
- [21] 褚璇, 王伟, 张录达, 等. 高光谱最优波长选择及Fisher判别分析法判别玉米颗粒表面黄曲霉毒素[J]. *光谱学与光谱分析*, 2014, 34(7): 1811-1815. DOI:10.3964/j.issn.1000-0593(2014)07-1811-05.
- [22] 王怀中, 王文志, 盛清凯, 等. 贮藏温度及贮藏期对发酵饲料霉变的影响[J]. *饲料研究*, 2016(7): 1-4; 9. DOI:10.13557/j.cnki.issn1002-2813.2016.07.001.
- [23] 国家食品药品监督管理局, 国家卫生和计划生育委员会. 食品中黄曲霉毒素B₁的测定: GB/T 18979—2003[S]. 北京: 中国标准出版社, 2003.
- [24] 国家食品药品监督管理局, 国家卫生和计划生育委员会. 食品中玉米赤霉烯酮的测定: GB/T 5009.209—2008[S]. 北京: 中国标准出版社, 2008.
- [25] SHETTY N, GISLUM R. Quantification of fructan concentration in grasses using NIR spectroscopy and PLSR[J]. *Field Crops Research*, 2011, 120(1): 31-37. DOI:10.1016/j.fcr.2010.08.008.
- [26] GALVAO R K H, ARAUJO M C U, JOSE G E, et al. A method for calibration and validation subset partitioning[J]. *Talanta*, 2005, 67(4): 736-740. DOI:10.1016/j.talanta.2005.03.025.
- [27] 褚小立. 化学计量学方法与分子光谱分析技术[M]. 北京: 化学工业出版社, 2011.
- [28] ARAÚJO M C U, SALDANHA T C B, GALVÃO R K H, et al. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis[J]. *Chemometrics and Intelligent Laboratory Systems*, 2001, 57(2): 65-73. DOI:10.1016/S0169-7439(01)00119-8.
- [29] LI H D, LIANG Y Z, XU Q S, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration[J]. *Analytica Chimica Acta*, 2009, 648(1): 77-84. DOI:10.1016/j.aca.2009.06.046.
- [30] 文韬, 洪添胜, 李立君, 等. 基于高光谱技术的霉变稻谷脂肪酸含量无损检测[J]. *农业工程学报*, 2015, 31(18): 233-239. DOI:10.11975/j.issn.1002-6819.2015.18.032.