

基于近红外光谱技术对多年际建三江、五常大米产地溯源

钱丽丽^{1,2}, 宋雪健¹, 张东杰^{1,2,*}, 张丽媛^{1,2}, 阮长青^{1,2}, 鹿保鑫^{1,2}

(1. 黑龙江八一农垦大学食品学院, 黑龙江 大庆 163319;

2. 黑龙江八一农垦大学 黑龙江省农产品加工与质量安全重点实验室, 黑龙江 大庆 163319)

摘要: 研究多年际地理标志大米产地溯源判别技术, 保护地理标志大米的品牌效益。利用近红外漫反射光谱技术对试验田样品进行与产地有关的特征波段筛选, 并在其范围内对2013—2015年来自建三江地区及五常地区的291份大米样品进行产地溯源研究。结果表明, 试验田样品在波段为 $5\,136\sim 5\,501\text{ cm}^{-1}$ 范围内产地因素差异显著。在此范围内利用因子化法建立的定性分析模型及聚类分析模型对建三江大米及五常大米的正确判别率均高于97.00%。利用偏最小二乘法建立的定量分析模型对两地区大米的正确判别率分别为95.83%、94.00%。因此, 应用该技术对多年份大米的产地溯源进行判别具有一定的可行性。

关键词: 大米; 产地溯源; 近红外漫反射光谱技术; 因子化法; 偏最小二乘法

Tracing the Geographical Origin of Sanjiang and Wuchang Rice Grown in Different Years by Near Infrared Spectroscopy

QIAN Lili^{1,2}, SONG Xuejian¹, ZHANG Dongjie^{1,2,*}, ZHANG Liyuan^{1,2}, RUAN Changqing^{1,2}, LU Baoxin^{1,2}

(1. College of Food Science, Heilongjiang Bayi Agricultural University, Daqing 163319, China; 2. Key Laboratory of Agro-Products Processing and Quality Safety of Heilongjiang Province, Heilongjiang Bayi Agricultural University, Daqing 163319, China)

Abstract: This study aimed to develop a technique to trace the geographical origin of protected geographical indication (PGI) rice grown in different years for the purpose of protecting the brand benefits of PGI rice. Near-infrared diffuse reflectance spectroscopy was used to detect rice samples from experimental fields and the characteristic bands related to the growing area were selected. The geographical origin of 291 rice samples from Sanjiang and Wuchang areas, collected during 2013 to 2015 was analyzed using the characteristic bands. There were significant differences among the experimental filed grow rice samples in terms of geographical origin in the range of $5\,136\sim 5\,501\text{ cm}^{-1}$. In this range, the qualitative analysis model and the cluster analysis model established by the factorization method showed a correct recognition rate of higher than 97.00% for Sanjiang and Wuchang rice; the correct recognition rates of the quantitative analysis model established by partial least squares (PLS) were 95.83% and 94.00%, respectively. Therefore, this method was proved feasible for tracing the geographical origin of rice harvested in different years.

Keywords: rice; geographical origin traceability; near infrared diffuse reflectance spectroscopy; factorization method; partial least squares

DOI:10.7506/spkx1002-6630-201816047

中图分类号: O657.3

文献标志码: A

文章编号: 1002-6630 (2018) 16-0321-07

引文格式:

钱丽丽, 宋雪健, 张东杰, 等. 基于近红外光谱技术对多年际建三江、五常大米产地溯源[J]. 食品科学, 2018, 39(16): 321-327. DOI:10.7506/spkx1002-6630-201816047. <http://www.spkx.net.cn>

QIAN Lili, SONG Xuejian, ZHANG Dongjie, et al. Tracing the geographical origin of Sanjiang and Wuchang rice grown in different years by near infrared spectroscopy[J]. Food Science, 2018, 39(16): 321-327. (in Chinese with English abstract)

DOI:10.7506/spkx1002-6630-201816047. <http://www.spkx.net.cn>

收稿日期: 2017-07-13

基金项目: 黑龙江省水稻现代农业科技协同创新体系-加工技术攻关实验室主任专家项目 (2017);

“十三五”国家重点研发计划重点专项 (2017YFD0401203);

黑龙江省垦总局“十三五”重点科技攻关项目 (HNK135-06-06); 黑龙江省应用技术与开发计划项目 (GA14B104)

第一作者简介: 钱丽丽 (1979—), 女, 副教授, 博士, 研究方向为农产品产地溯源。E-mail: qianlili286@163.com

*通信作者简介: 张东杰 (1966—), 男, 教授, 博士, 研究方向为农产品加工与安全。E-mail: byndzdj@126.com

地理标志作为一项知识产权已经得到国际认可并受到保护^[1-2]。素有“贡米”之称的五常大米及“中国绿色米都”建三江种植建三江大米因其籽粒饱满晶莹剔透,营养丰富,清香适口等特点被评为地理标志大米^[3],备受消费者青睐,具有广阔的市场前景。一些不法商贩以此为“良机”开始制售假大米,以谋取利益,严重损坏了地理标志大米的品牌效益,并扰乱了市场秩序。因此,为实现对地理标志大米的产地保护研究,开发一种快速鉴别地理标志大米技术是亟待解决的问题之一。目前,对大米产地溯源的研究技术主要有电子鼻技术^[4]、电子舌技术^[5]、矿物元素指纹分析技术^[6]、DNA指纹分析技术^[7]、电感耦合等离子技术^[8]、拉曼光谱技术^[9]、近红外光谱技术^[10]等。与其他技术相比较近红外光谱具有快速、高效、无损、适用范围广等特点。

近红外光谱分析技术是利用近红外谱区包含的丰富的物质信息,同时吸收带的吸收强度与分子组成或化学基团的含量有关可用于测定化学物质的成分和分析物理性质^[11],对于多组分的复杂样品,其近红外光谱也不是各组分单独光谱的简单叠加,因此,近红外光谱技术需要结合“化学计量法”对光谱信号进行处理,从而提取食品中的有效信息^[12],其研究结果较为理想。近红外光谱的采集方式主要有透射式、漫反射式、透漫射式3种,在对固体样品光谱采集时多以漫反射为主,而其他两种方式多用于液体样品的光谱采集。范积平等^[13]研究发现利用近红外漫反射光谱技术能实现对来自甘肃、青海、陕西的大黄药材进行产地鉴别。Sinelli等^[14]通过传统感官评价方法并结合近红外漫反射光谱技术对112组初榨橄榄油进行了产地溯源研究,发现采用线性判别式分析和簇独立软模式分类法对初榨橄榄油产地的判别正确率分别为71.6%、100%。也有学者将近红外漫反射技术用于小米^[15]、小麦^[16]、茶叶^[17]等产地溯源。目前,应用近红外漫反射光谱技术对大米研究多集中在品质检测及少量单一年份产地判别,鲜见应用该技术对多年际大米进行产地溯源的研究。本研究采用近红外漫反射光谱技术对多年份大米进行产地溯源研究,为地理标志大米产地保护研究提供理论研究基础。

1 材料与方法

1.1 材料与试剂

于2013—2015年采集建三江、五常地区及2015年试验田的样品,均为田间采样,采用3点随机取样的方式进行采集,每份样本采集2 kg装入尼龙网兜,并记录采样信息,所有水稻品种均为粳米。除试验田样品外,其余均归为随机样品。样本详细信息见表1、2。

表1 试验田样品信息

Table 1 Information about rice samples from experimental field

试验田地区	品种 (数量)
五常、建三江、查哈阳	空育131 (10)、龙粳26 (10)、龙粳29 (10)、龙粳31 (10)、龙粳39 (10)、龙粳40 (10)、松粳9 (10)、绥粳4 (10)、五优稻4号 (10)

表2 随机采样样品信息

Table 2 Information about randomly collected rice samples

地区	年份	品种及数量
建三江 (142 份)	2013 (50)	龙粳31 (10)、空育131 (10)、龙粳26 (4)、龙粳25 (6)、龙粳36 (8)、龙粳29 (12)
	2014 (49)	龙粳31 (11)、空育 (10)、龙粳39 (8)、龙粳29 (4)、龙粳26 (6)、垦稻18 (10)、
	2015 (43)	0318 (10)、龙粳16 (6)、龙粳31 (8)、龙粳43 (4)、龙粳46 (3)、绥粳4 (12)
五常 (149 份)	2013 (50)	五优稻4号 (5)、东农425 (5)、松粳12 (8)、松粳16 (12)、稻花香 (14)、松粳3号 (6)
	2014 (50)	吉粳803 (15)、稻花香 (16)、龙粳29 (4)、龙粳30 (5)、垦稻18 (10)
	2015 (49)	稻花香 (29)、绥粳 (5)、五优 (8)、吉粳803 (7)

1.2 仪器与设备

FC2K砵谷机 日本大竹制作所; VP-32实验碾米机 日本山本公司; FW100高速万能粉碎机 天津泰斯特仪器有限公司; TENSORII型傅里叶变换近红外光谱仪(漫反射镀金积分球、InGaAs检测器) 德国布鲁克(北京)科技有限公司。

1.3 方法

1.3.1 水稻试验田的构建

于2015年,以粳稻主产区黑龙江省(五常、建三江、查哈阳)为试验点,建立3块试验田。每块试验田种植主栽中晚熟品种9个分别为空育131、龙粳26、龙粳29、龙粳39、龙粳40、松粳9、五优稻4号、绥粳4,各试验田育苗移栽、秧田管理、施肥、灌溉均按当地生产习惯,所有品种同期播种。采用3次重复随机区组设计,小区面积不少于10 m²。四周设保护行,保护行品种与各对应品种相同。

1.3.2 样品的前处理

采集回来的样品经过晾晒、去石、脱粒等前处理后,在实验室采用统一加工方式对其进行砻谷、碾米、粉碎、过筛(100目),待测。每个样品砻谷2次。碾米的进样量为3次,碾米3次,白度为3次。

1.3.3 样品原始光谱的采集

将傅里叶变换近红外光谱仪预热30 min,打开OPUS 7.5软件经由检查信号、保存峰位,扫描背景单通道光谱,每间隔1 h扫描一次背景来消除外界信息干扰保证光谱的稳定性以减少误差。将样品粉末倒入玻璃杯中,用压样器压实,测量样品单通道采集样品光谱。环境温度为室温(25±1)℃,相对湿度为20%~30%,光谱波数范围12 000~4 000 cm⁻¹,分辨率8 cm⁻¹,扫描64次。

1.3.4 样品选取

试验田样品用于与产地因素有关的特征波段筛选。其余样品则选择各地区样品总量的2/3作为建模样品用于模型的建立, 1/3作为预测样品集用于模型的验证。各地区用于建模和预测的样品数见表3。

表3 建模与预测样品
Table 3 Modeling and prediction set samples

类别	建模样品个数	预测样品个数	总计
建三江大米	94	48	142
五常大米	99	50	149

1.3.5 试验田近红外光谱数据处理

应用OPUS 7.5软件分别对试验田样品的原始光谱进行一阶导数预处理消除噪声、基线漂移、光色散等因素的干扰及防止过拟合现象的出现, 应用SPSS 19.0软件对试验田样品进行方差分析筛选出与产地有关的波段, 并用于随机样品产地溯源研究。

1.3.6 模型的建立及验证

1.3.6.1 定性分析模型建立

采用OPUS 7.5软件在特征波段下利用因子化法和欧式距离法对随机样品进行定性分析模型建立研究, 预处理方式分为矢量归一化 (standard normal variate, SNV)、一阶导数+平滑 (5、9、13、17、21、25点, 下同)、一阶导数+SNV+平滑、二阶导数+平滑、二阶导数+SNV+平滑等, 消除无关信息的干扰, 提高模型的精度。通过比较选择性 S 值最终确定定性分析模型的建立方法, 其中 S 值表征样品之间的距离, 当 S 值小于1时, 表示两类样品“相交”, 样品未被均一鉴别; 当 S 值为1时, 表示两类样品“相切”; 当 S 值大于1时, 表示两类样品“相离”, 样品被均一鉴别, 故 S 值越大, 模型的效果越好。同时利用因子化法和欧式距离法对随机样品进行聚类分析模型建立研究, 预处理方式与上述相同。并通过比较样品之间的“距离”大小确定建模方法, 其中样品之间的“距离”越大越好。

1.3.6.2 定量分析模型建立

采用留1交叉检验的检验方法以偏最小二乘 (partial least squares, PLS) 法在特征波段进行定量分析模型的建立, 分别将建三江大米的组分赋值-1、五常大米的组分赋值为1, 以0为衡量标准, 预测值小于0为建三江大米、预测值大于0为五常大米^[18]。原始光谱的预处理方式有消除常数偏移量、减去一条直线、SNV、最小-最大归一化、多元散射校正 (multiplicative scatter correction, MSC)、内部标准、一阶导数+平滑、二阶导数+平滑、一阶导数+减去一条直线+平滑、一阶导数+SNV+平滑、一阶导数+多元散射校正+平滑, 通过交互验证均方根误差 (root mean square error of cross validation,

RMSECV) 及定向系数 (R^2) 来衡量模型的好坏, 其中 R^2 数值越接近100%则预测含量值越接近真值; RMSECV数值越小越好; 同时RMSECV先随维数的增大快速下降后略微逐渐增大, 模型效果较好, 进而确定定量分析模型。

1.3.6.3 模型验证

利用OPUS 7.5软件分别选择定性分析、聚类分析、定量分析工具栏, 调入模型, 调入预测样品光谱图, 测定得出结果。

2 结果与分析

2.1 大米近红外原始光谱分析结果

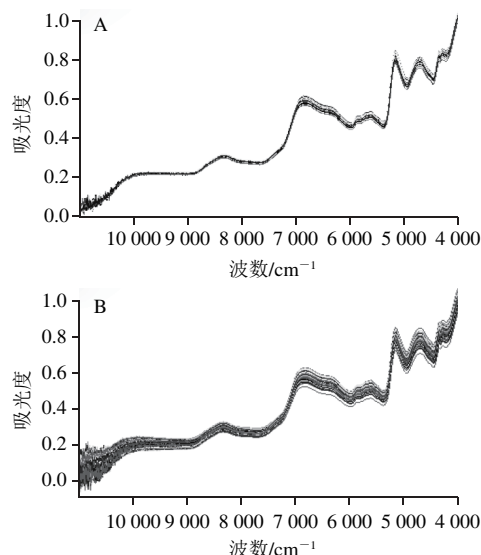


图1 试验田样品 (A) 和随机样品 (B) 近红外原始光谱图
Fig. 1 NIR Spectra of experimental field samples (A) and random samples (B)

大米近红外原始光谱如图1所示, 在波段为7 500~9 000 cm^{-1} 处 (I区) 是C—H第3组合频区, 其中8 321 cm^{-1} 附近的吸收峰是由脂肪烃中甲基 (—CH₃) 基团引起的; 5 500~7 500 cm^{-1} 处 (II区) 是C—H第2组合频区, 在6 846 cm^{-1} 附近的吸收峰是因—CH₂二级振动所引起的, 因与样品中氨基酸种类及含量有关, 所以较I区信息稍微强些; 4 000~5 500 cm^{-1} 处 (III区) 是C—H第1组合频区, 是表征蛋白质及淀粉物质中的N—H、C—H、O—H及C=O键振动的要区间, 其中5 173 cm^{-1} 处的吸收峰与其有关^[19]。不同产地及不同品种之间的近红外光谱相似, 采用一阶导数对试验田原始光谱预处理后进行方差分析, 结果表明, 不同地区来源的样品在波段5 136~5 501 cm^{-1} 处均有显著差异, 说明不同地区间样品的近红外光谱存在显著性差异。在特征波段对大米进行产地溯源研究具有较强的代表性^[20]。

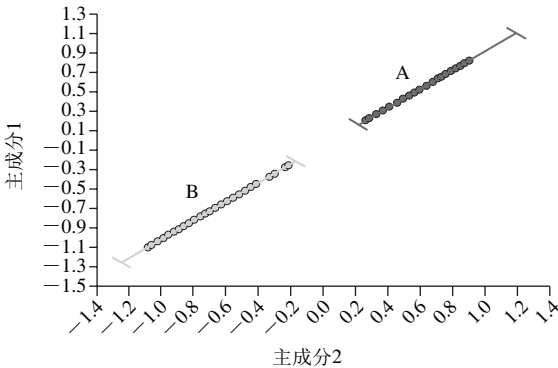
2.2 不同预处理方式对定性分析建模效果的影响

近红外光谱虽然包含了丰富的物质信息，但谱峰重叠、信号较弱、谱带较宽，难以像中红外光谱那样进行结构剖析，因此近红外光谱的定性分析主要用于物质的种属判别，即通过比较未知样品与已知样本或标准样本的光谱确定未知样品的归属^[21]。在特征波段5 136~5 501 cm⁻¹范围内分别采用因子化法和欧式距离法对随机样品的原始光谱进行定性建模分析，如表4所示。结果表明，采用因子化法结合二阶导数+SNV+5点平滑的预处理方式建立的模型效果较好，两地区的样品被均一鉴别，其中模型的S值为1.324 005，故选此方法建立定性分析模型，如图2所示。王梦东等^[22]利用近红外漫反射光谱技术结合因子化法在特征波段7 400~9 900 cm⁻¹范围内采用二阶导数+矢量归一化+21点平滑对白茶、红茶及乌龙茶进行判别研究，研究发现采用近红外漫反射光谱技术结合因子化法在3类茶叶的判别中具有可行性。张敏等^[23]研究发现，利用近红外漫反射光谱技术结合因子化法在特征波段4 482~5 238、5 369.1~6 950.5 cm⁻¹范围内采用一阶导数+矢量归一化+13点平滑的预处理方式能实现对来自6个地区的鸡血藤的产地溯源判别研究。

表4 不同预处理方式对定性分析模型效果影响
Table 4 Effects of different pretreatment methods on qualitative analysis model

预处理方式	S值	
	因子化法	欧式距离法
SNV	0.252 262	0.133 372
一阶导数+5点平滑	1.010 369	0.252 366
一阶导数+9点平滑	0.652 067	0.211 747
一阶导数+13点平滑	0.601 627	0.198 795
一阶导数+17点平滑	0.631 180	0.194 018
一阶导数+21点平滑	0.590 510	0.191 835
一阶导数+25点平滑	0.689 572	0.190 237
一阶导数+SNV+5点平滑	1.006 176	0.571 199
一阶导数+SNV+9点平滑	0.247 501	0.382 395
一阶导数+SNV+13点平滑	0.245 307	0.295 980
一阶导数+SNV+17点平滑	0.256 750	0.255 560
一阶导数+SNV+21点平滑	0.278 030	0.233 772
一阶导数+SNV+25点平滑	0.250 275	0.271 118
二阶导数+5点平滑	1.244 246	1.004 540
二阶导数+9点平滑	1.163 714	0.904 538
二阶导数+13点平滑	1.146 705	0.534 632
二阶导数+17点平滑	0.983 451	0.357 532
二阶导数+21点平滑	0.706 069	0.347 500
二阶导数+25点平滑	0.538 933	0.305 278
二阶导数+SNV+5点平滑	1.324 005	1.000 118
二阶导数+SNV+9点平滑	1.167 843	0.931 607
二阶导数+SNV+13点平滑	1.156 442	0.863 071
二阶导数+SNV+17点平滑	1.007 772	0.654 547
二阶导数+SNV+21点平滑	0.722 767	0.502 121
二阶导数+SNV+25点平滑	0.526 318	0.437 216

通过对原始光谱进行主成分分解，选取特征值较大的几个主成分得分特征变量参与模式识别，在进来保留有用信息的前提下，不仅压缩了原始光谱，而且这些变量是沿最大方差方向得到的，还起到特征信息提取的作用。而标准算法采用的是欧式距离，直接使用吸光度计算光谱距离，不能体现特征变量的变化情况。因此，采用因子化法建立的模型精度要优于欧式距离法。



A区域代表建三江大米，B区域代表五常大米。

图2 因子化法2D得分图

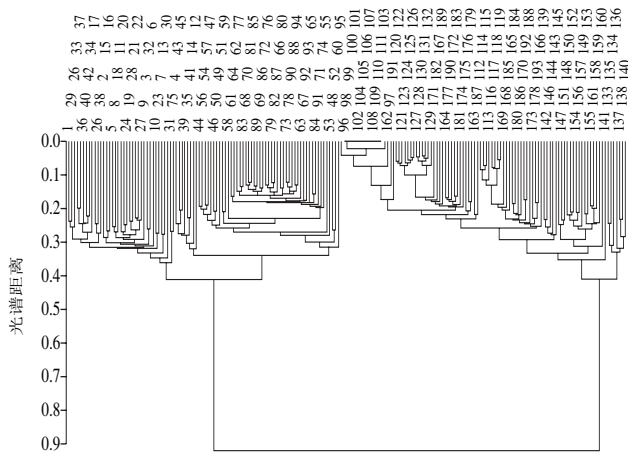
Fig. 2 2D score plot by factorization method

2.3 不同预处理方式对聚类分析建模效果的影响

表5 不同预处理方式对聚类分析模型效果影响
Table 5 Effects of different pretreatment methods on clustering analysis model

预处理方式	聚类分析结果	
	因子化法	欧式距离法
矢量归一化	—	—
一阶导数+5点平滑	—	—
一阶导数+9点平滑	—	—
一阶导数+13点平滑	—	—
一阶导数+17点平滑	—	—
一阶导数+21点平滑	—	—
一阶导数+25点平滑	—	—
一阶导数+SNV+5点平滑	0.085 00	0.043 00
一阶导数+SNV+9点平滑	—	0.037 00
一阶导数+SNV+13点平滑	—	—
一阶导数+SNV+17点平滑	—	—
一阶导数+SNV+21点平滑	—	—
一阶导数+SNV+25点平滑	—	—
二阶导数+5点平滑	0.000 30	—
二阶导数+9点平滑	0.000 13	—
二阶导数+13点平滑	0.000 65	—
二阶导数+17点平滑	0.000 46	0.000 34
二阶导数+21点平滑	—	—
二阶导数+25点平滑	—	0.028 00
二阶导数+SNV+5点平滑	0.918 00	0.062 00
二阶导数+SNV+9点平滑	0.457 00	—
二阶导数+SNV+13点平滑	0.243 00	—
二阶导数+SNV+17点平滑	0.158 00	0.043 00
二阶导数+SNV+21点平滑	0.109 00	0.028 00
二阶导数+SNV+25点平滑	—	—

注：—，两类的样品未被正确分类，数据表示两类样品之间的距离。



编号1~94为建三江大米, 95~193为五常大米。

图3 不同地区大米样品的聚类分析结果

Fig. 3 Cluster analysis of different geographical rice samples

聚类分析用于判定一系列近红外谱图的相似性, 是根据样本自身的属性, 用数学方法按照某种相似性或差异性指标确定样本之间的亲疏关系, 并按这种亲疏关系程度对样本进行聚类。与定性分析不同, 它不需要输入任何信息, 聚类分析只将相似光谱按组分类^[24]。在特征波段5 136~5 501 cm^{-1} 范围内分别采用因子化法和欧氏距离法对随机样品的原始光谱进行聚类建模分析, 如表5所示。结果表明, 采用因子化法结合二阶导数+SNV+5点平滑的预处理方式建立的模型效果较好, 两地区样品被正确分类, 且“距离”为0.918 00, 如图3所示。故选此方法建立聚类分析模型。庞艳苹等^[25]利用近红外漫反射光谱技术因子化法在4 000~8 003.8 cm^{-1} 范围内采用二阶导数+5点平滑的预处理方式对255份草莓进行产地溯源研究, 结果表明, 其模型的预测正确率高达96.70%。也有学者发现在特征波段4 000~7 500 cm^{-1} 范围内采用因子化法结合一阶导数+矢量归一化+17点平滑的预处理方法建立的模型能够实现对久保桃的产地鉴别^[26]。

2.4 不同预处理方式对定量分析建模效果的影响

建立定量分析模型的目的是对多组分样品的相关组分进行定量分析。首先需要选择能够代表待测体系的已知组分的建模样品, 然后用这些样品的近红外光谱和组分值来拟合模型, 经过检验其预测的可靠性之后, 这个模型就可用来分析未知样品的组分值^[27]。PLS法将用来寻找光谱与浓度数据矩阵之间的最佳相关函数关系, 具有更易于辨识系统信息与噪声的优点, 使其拟合的模型精度更高^[28]。在特征波段5 136~5 501 cm^{-1} 范围内采用PLS法结合交叉检验法对原始光谱进行处理, 如表6所示。结果表明, 采用一阶导数+MSC+13点平滑的预处理方式建立的模型效果较好, 其RMSECV为0.105, R^2 为98.97%, 维数为4, 故选此方法来建立定量分析模型。如图4、5所示。汪静静等^[29]研究发现利用近红外漫反射光谱技术在

特征波段7 559~8 531 cm^{-1} 范围内采用二阶导数+MSC+5点平滑建立的定量分析模型对来自3个地区的74份人参样品进行产地判别, 其正确率高达90%。赵艳丽等^[30]在波段4 007.35~7 135.33 cm^{-1} , 采用近红外漫反射光谱技术结合PLS判别分析法对3个不同产区的70份野生药用植物重建判别模型, 其预测正确率为100%。

表6 不同预处理方式对定量分析模型效果影响

Table 6 Effect of different pretreatment methods on quantitative analysis model

预处理方式	RMSECV	$R^2/\%$	维数
消除常数偏移量	0.269	96.15	5
减去一条直线	0.364	95.32	4
SNV	0.272	96.04	4
最小-最大归一化	0.366	95.24	4
MSC	0.356	96.57	4
内部标准	0.258	96.57	6
一阶导数+5点平滑	0.143	97.96	4
一阶导数+9点平滑	0.236	97.55	5
一阶导数+13点平滑	0.238	97.59	5
一阶导数+17点平滑	0.244	97.93	5
一阶导数+21点平滑	0.342	96.00	8
一阶导数+25点平滑	0.146	97.86	6
二阶导数+9点平滑	0.144	97.91	2
二阶导数+13点平滑	0.227	97.40	4
二阶导数+17点平滑	0.115	98.69	4
二阶导数+21点平滑	0.117	98.46	4
二阶导数+25点平滑	0.317	95.63	6
一阶导数+减去一条直线+5点平滑	0.107	98.86	4
一阶导数+减去一条直线+9点平滑	0.197	97.85	4
一阶导数+减去一条直线+13点平滑	0.112	98.76	4
一阶导数+减去一条直线+17点平滑	0.119	98.76	4
一阶导数+减去一条直线+21点平滑	0.323	96.50	5
一阶导数+减去一条直线+25点平滑	0.424	95.46	5
一阶导数+SNV+5点平滑	0.111	98.76	3
一阶导数+SNV+9点平滑	0.113	98.72	3
一阶导数+SNV+13点平滑	0.180	98.20	4
一阶导数+SNV+17点平滑	0.128	98.74	4
一阶导数+SNV+21点平滑	0.213	98.55	8
一阶导数+SNV+25点平滑	0.314	97.54	5
一阶导数+MSC+5点平滑	0.115	98.80	3
一阶导数+MSC+9点平滑	0.119	98.72	3
一阶导数+MSC+13点平滑	0.105	98.97	4
一阶导数+MSC+17点平滑	0.145	98.60	4
一阶导数+MSC+21点平滑	0.312	95.74	5
一阶导数+MSC+25点平滑	0.216	96.65	5

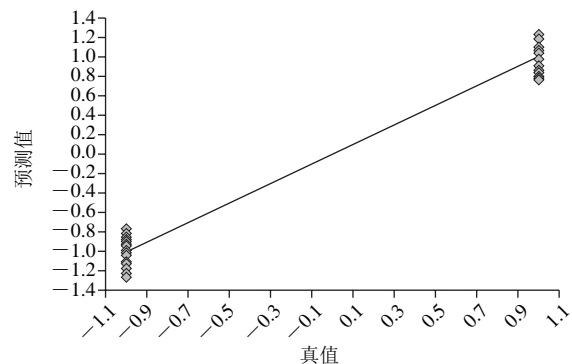


图4 地区预测值与参考值相关图

Fig. 4 Predicted value versus reference value

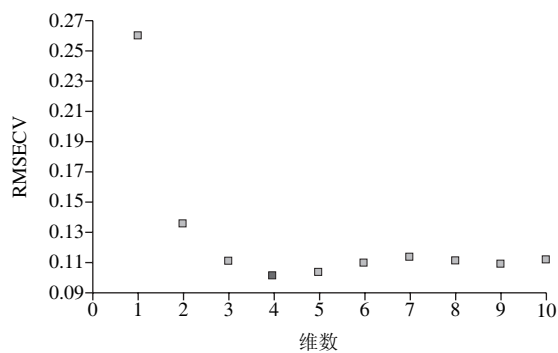


图5 RMSECV与维数的关系图

Fig. 5 RMSECV against dimensionality

2.5 大米产地溯源模型验证

将建立好的定性分析模型、聚类分析模型及定量分析模型分别带入OPUS 7.5软件中,对预测样品进行验证,结果表明,定性分析模型对建三江大米及五常大米的正确判别率分别为100%、98%。聚类分析模型对建三江大米及五常大米的正确判别率分别为97.92%、98.00%。定量分析模型对建三江大米及五常大米的正确判别率分别为95.83%、94.00%,如表7所示。

表7 模型鉴别结果

Table 7 Results of model validation

判别方法	地区	正确个数	错误个数	正确判别率/%	样品总数
定性分析	建三江	48	0	100.00	48
	五常	49	1	98.00	50
聚类分析	建三江	47	1	97.92	48
	五常	49	1	98.00	50
定量分析	建三江	46	2	95.83	48
	五常	47	3	94.00	50

3 结论

对大米进行产地溯源判别研究易受到产地、品种、施肥量等因素的影响,故为筛选与产地有关的因素实验于2015年在黑龙江五常、佳木斯、齐齐哈尔三地建立试验田,进行特殊波段的筛选,通过对试验田样品进行一阶导数处理得出特征波数范围为 $5\ 136\sim 5\ 501\ \text{cm}^{-1}$ 。在特征波段范围内对2013年至2015年来自建三江地区及五常地区的291份大米进行产地溯源研究,结果发现,采用因子化法结合二阶导数+SNV+5点平滑的预处理方式建立的定性分析模型对建三江大米及五常大米的正确判别率分别为100%、98%。采用因子化法结合二阶导数+矢量归一化+5点平滑的预处理方式建立的聚类分析模型对建三江大米及五常大米的正确判别率分别为97.92%、98.00%。采用PLS法结合一阶导数+MSC+13点平滑的预处理方式建立的定量分析模型对建三江大米及五常大米的正确判别率分别为95.83%、94.00%。其中定性分析

与聚类分析的最优处理方法相一致。故应用近红外漫反射光谱技术可以实现对多个年份大米的产地鉴别。大米产地溯源受多种因素影响,今后可以从土壤、基因型、施肥量等多种因素进行研究。

参考文献:

- [1] 冯寿波. 论地理标志的国际法律保护: 以TRIPS协议为视角[M]. 北京: 北京大学出版社, 2008.
- [2] 董炳和. 地理标志知识产权制度研究: 构建以利益分享为基础的权利体系[M]. 北京: 中国政法大学出版社, 2005: 10.
- [3] 董俊秀. 基于地理标志的黑龙江大米品牌整合策略探讨[J]. 商业经济, 2012(17): 65-66. DOI:10.3969/j.issn.1009-6043.2012.17.029.
- [4] HAN H J, LEE S H, JI Y M, et al. Discrimination of the cultivar, growing region, and geographical origin of rice (*Oryza sativa*) using a mass spectrometer-based electronic nose[J]. Food Science & Biotechnology, 2016, 25(3): 695-700. DOI:10.1007/s10068-016-0121-8.
- [5] WANG L, NIU Q F, HUI Y B, et al. Discrimination of rice with different pretreatment methods by using a voltammetric electronic tongue[J]. Sensors, 2015, 15(7): 17767-17785. DOI:10.3390/s150717767.
- [6] SHEN S G, XIA L Y, XIONG N, et al. Determination of the geographic origin of rice by element fingerprints and correlation analyses with the soil of origin[J]. Analytical Methods, 2013, 5(21): 6177-6185. DOI:10.1039/C3AY40700D.
- [7] DAI C. Application of DNA fingerprint based on SSR in rice adulteration detection and origin traceability[M]//Food Hygiene, Agriculture and Animal Science: Proceedings of the 2015 International Conference on Food Hygiene, Agriculture and Animal Science. 2015: 81-93.
- [8] CHUNG I M, KIM J K, LEE J K, et al. Discrimination of geographical origin of rice (*Oryza sativa*, L.) by multielement analysis using inductively coupled plasma atomic emission spectroscopy and multivariate analysis[J]. Journal of Cereal Science, 2015, 65: 252-259. DOI:10.1016/j.jcs.2015.08.001.
- [9] KIM Y, LEE S, CHUNG H, et al. Improving Raman spectroscopic differentiation of the geographical origin of rice by simultaneous illumination over a wide sample area[J]. Journal of Raman Spectroscopy, 2010, 40(2): 191-196. DOI:10.1002/jrs.2105.
- [10] XIA L Y, SHEN S G, LIU Z H, et al. Identification of geographical origins of rice with pattern recognition technique by near infrared spectroscopy[J]. Spectroscopy & Spectral Analysis, 2013, 33(1): 102. DOI:10.3964/j.issn.1000-0593(2013)01-0102-04.
- [11] 张小超, 吴静珠, 徐云. 近红外光谱分析技术及其在现代农业中的应用[M]. 北京: 电子工业出版社, 2012: 2.
- [12] 褚小立, 许育鹏, 陆婉珍. 用于近红外光谱分析的化学计量学方法研究与应用进展[J]. 分析化学, 2008, 36(5): 702-709. DOI:10.3321/j.issn.0253-3820.2008.05.029.
- [13] 范积平, 张柳瑛, 张贞良, 等. 不同产地大黄药材的近红外漫反射光谱法鉴别[J]. 药学实践杂志, 2005, 23(3): 148-150. DOI:10.3969/j.issn.1006-0111.2005.03.006.
- [14] SINELLI N. Application of near (NIR) infrared and mid (MIR) infrared spectroscopy as a rapid tool to classify extra virgin olive oil on the basis of fruity attribute intensity[J]. Food Research International, 2010, 43(1): 369-375. DOI:10.1016/j.foodres.2009.10.008.
- [15] 宋雪健, 钱丽丽, 周义, 等. 近红外漫反射光谱技术对小米产地溯源的研究[J]. 食品研究与开发, 2017, 38(11): 134-139. DOI:10.3969/j.issn.1005-6521.2017.11.030.

- [16] 赵海燕, 郭波莉, 魏益民, 等. 近红外光谱对小麦产地来源的判别分析[J]. 中国农业科学, 2011, 44(7): 1451-1456. DOI:10.3864/j.issn.0578-1752.2011.07.018.
- [17] HE W, ZHOU J, CHENG H, et al. Validation of origins of tea samples using partial least squares analysis and Euclidean distance method with near-infrared spectroscopy data[J]. Spectrochimica Acta Part A Molecular & Biomolecular Spectroscopy, 2012, 86(3): 399-404. DOI:10.1016/j.saa.2011.10.056.
- [18] 张晓焱. 基于近红外光谱的柑桔产地溯源及橙汁掺假识别研究[D]. 重庆: 西南大学, 2010.
- [19] 杰尔·沃克曼, 洛伊斯·文依. 近红外光谱解析实用指南[M]. 北京: 化学工业出版社, 2009: 220.
- [20] 谌蓓, 夏立娅, 窦玉蕾, 等. 基于NIR的模式识别技术在地理标志产品响水大米鉴别中的应用[J]. 河北大学学报(自然科学版), 2013, 33(5): 484-488. DOI:10.3969/j.issn.1000-1565.2013.05.008.
- [21] 张荣, 吴文娟. 近红外光谱技术的定性和定量分析[J]. 化工时刊, 2011, 25(9): 36-38. DOI:10.3969/j.issn.1002-154X.2011.09.013.
- [22] 王梦东, 王胜鹏. 适用于3类茶的定性分类及主要内含成分定量分析的近红外预测模型的建立[J]. 华中农业大学学报, 2015, 34(1): 123-127. DOI:10.13300/j.cnki.hnlkxb.2015.01.020.
- [23] 张敏, 韩正洲, 严萍, 等. 基于近红外漫反射光谱的鸡血藤药材定性分析[J]. 世界科学技术: 中医药现代化, 2014(7): 1491-1496. DOI:10.11842/wst.2014.07.009.
- [24] 温志芳, 柯立坚, 张云平, 等. NIR聚类分析结合实验室检验快速确定不明药品[J]. 中国药事, 2010, 24(10): 1015-1016.
- [25] 庞艳苹, 刘坤, 闫军颖, 等. 近红外光谱法快速鉴别成安草莓[J]. 现代食品科技, 2013(5): 1160-1162. DOI:10.13982/j.mfst.1673-9078.2013.05.007.
- [26] 庞艳苹, 夏立娅, 左永强, 等. 近红外光谱法快速鉴别真伪平谷大久保桃[J]. 安徽农业科学, 2010, 38(3): 1122-1123. DOI:10.13989/j.cnki.0517-6611.2010.03.109.
- [27] 张晓慧. 基于近红外光谱的连翘有效成分分析与产地鉴别技术研究[D]. 洛阳: 河南科技大学, 2008: 13.
- [28] IBRAHIM S, RAHIM H A. The assessment of amylose content using near infrared spectroscopy analysis of rice grain samples[C]//IEEE Conference on Open Systems. 2013: 32-37.
- [29] 汪静静, 闫述模, 杨滨. 近红外光谱技术对人参药材人参皂苷含量测定及产地识别的研究[J]. 光谱学与光谱分析, 2015, 35(7): 1885-1888. DOI:10.3964/j.issn1000-0593(2015)07-1885-04.
- [30] 赵艳丽, 张霁, 袁天军, 等. 近红外光谱快速鉴别不同产地药用植物重楼的方法研究[J]. 光谱学与光谱分析, 2014, 34(7): 1831-1835. DOI:10.3964/j.issn.1000-0593(2014)07-1831-05.