

不同产地稻米的二维相关近红外光谱鉴别

戴元丰^{1,3}, 代作晓^{1*}, 郭光智¹, 王迎超²

(¹中国科学院上海技术物理研究所 上海 200083

²太仓光电技术研究所 江苏苏州 215411

³中国科学院大学 北京 100049)

摘要 为了实现针对不同产地稻米样品的无损判别,提出一种基于同步二维相关近红外光谱的欧氏距离判别方法。在分别得到 5 种稻米样品的标准二维相关谱后,通过计算测试集样品与各个标准相关谱的欧氏距离,根据最小距离原则进行归类判别,经实际测试正确率达到 95%。通过定义区分度系数还可以更精确判断对各类别样品的区分效果。为了比较二维相关方法与常用建模方法的实际效果,在同样的数据基础上进行偏最小二乘判别分析和支持向量机分类的尝试,其最高正确率分别为 75%和 70%,验证了二维相关方法的有效性。本研究为无损鉴别稻米样品提供了一种新的可能方法,并可为在线检测提供参考。

关键词 二维相关光谱; 近红外; 稻米; 欧氏距离; 鉴别分析

文章编号 1009-7848(2023)09-0331-08 **DOI:** 10.16429/j.1009-7848.2023.09.035

中国是世界上主要的水稻生产国之一,全国稻米种植区域广、种类多,土壤、环境和水质等差异形成地域因素会导致稻米的品质发生变化^[1]。具有鲜明地理标识的稻米产品因特有的口感和营养价值,进而具有更高的商业价值^[2]。在这种条件下,一些商家为了谋取更高的利益,用相近产地的稻米代替地域品牌稻米,严重侵犯了消费者的利益^[3]。正确鉴别稻米产地有重要意义。

目前,感官评价、形态性状指数、杂交亲和性鉴定、同工酶基因定量、DNA 分子标记、氯酸钾耐受性、运动细胞硅体形态性状、双峰结节鉴定等传统方法常用于水稻品种分类^[4-8]。这些传统方法均需要一定的软、硬件条件支撑,操作繁琐,耗时费力。光谱法为一种快速无损鉴别的手段,在食品科学领域得到越来越广泛的应用^[9]。近红外光(NIR)是介于可见光(Vis)与中红外(MIR)之间的电磁辐射波,它可以反映有机分子中含氢基团 X-H(例如:X 可以为 C,O,N 等)振动的各级倍频和合频的吸收情况^[10]。高地^[11]使用近红外光谱技术结合偏最小二乘判别分析(PLS-DA)建立模型,对松原地区 5 种稻米进行分类判别。钱丽丽等^[12]使用近红

外光谱技术,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稻米及五常稻米进行判别分析。林莹^[13]利用近红外光谱技术对辉南县火山岩稻米的真实性进行聚类分析。然而,上述应用大多是将样品磨粉处理后再进行光谱分析,且需要复杂的建模过程,无法满足无损、在线检测的需求。

广义二维相关光谱由 Noda 于 1993 年提出,已经成为一种强大而通用的工具,用于解释由外部扰动(如时间、温度、压力、浓度和成分)引起的细微光谱变化^[14-15]。因在针对复杂生物体系检测上的优势,二维相关光谱在食品分析领域也得到广泛的应用。Zhang 等^[16]采用二维红外相关光谱分析不同含糖量红酒的主要成分和不同厂家干红葡萄酒的挥发性残渣,实现了对不同红酒的鉴别。Sohng 等^[17]提出一种以温度变化为外扰二维相关近红外光谱分析策略,对掺假橄榄油进行分析,提高了鉴别的准确性。杨仁杰等^[18]以牛奶中掺杂物的浓度为外扰,提出一种基于欧式距离的二维相关红外谱判别方法,实现了对掺假牛奶的鉴别。

目前,针对稻米产地鉴别的二维相关光谱的研究很少。本文提出一种基于欧氏距离的二维相关近红外光谱判别方法,以温度作为外扰,针对 5 个产地的稻米样品进行鉴别。通过建立不同类别样品的标准二维相关图谱,将待测样品的二维相

收稿日期: 2022-09-23

第一作者: 戴元丰,男,博士生

通信作者: 代作晓 E-mail: tcgd_daizx@163.com

关图谱分别与标准图谱求取欧氏距离,依据距离最小归属一类的判别准则,实现对稻米产地的无损、快速鉴别。

1 材料与方法

1.1 样品与仪器

为了更好地评判鉴别方法的有效性,在样品

选择上选取了5个产地的粳米样品,分别来自黑龙江省、吉林省、宁夏回族自治区、上海市以及江苏省。每类样品各10份,每份质量为15g,放置于深10mm的样品池中。样品光谱采集在室温下(18.0±1)℃进行,环境相对湿度约为22%±2%。稻米样品的具体信息如表1所示。

表1 稻米样品信息

Table 1 Details of rice samples

类别编号	品牌名	产地	每100g营养成分			
			能量/kJ	蛋白质/g	脂肪/g	碳水化合物/g
1	京贡1号	黑龙江	1477	5.4	0	80.4
2	长粒香米	吉林	1412	5.0	1.4	75
3	吴帅	宁夏	1423	7.5	0	78.2
4	上海香米	上海	1467	5.9	0.8	78.7
5	苏软香	江苏	1401	7.3	0	75.7

光谱采集使用德国 INSION 公司的 NIR NT/H 微型光谱仪,光谱采集区间为 906~1 859 nm,分辨率为 8 nm。采用漫反射采集方式,配备 Y 型光纤,光源选用闻奕光电的 HL3000 卤素灯光源。使用恒温加热平台对样品池进行温度控制,在 30~60℃之间每隔 5℃采集动态近红外光谱。光谱采集时探头距离样品表面约 3.5 cm,在该距离下经过测试可以得到重复性较好的光谱曲线。设置积分时间为 500 ms,经过 16 次采集取平均值获得一条光谱数据。

1.2 二维相关近红外光谱

动态近红外光谱是指样品因外扰作用而处于不同状态时的光谱减去某一参考光谱之后的结果,动态光谱 $y(v, t)$ 可以表示为:

$$y(v, t) = \begin{cases} x(v, t) - \bar{x}(v), & T_{\min} \leq t \leq T_{\max} \\ 0, & \text{其它情况} \end{cases} \quad (1)$$

式中, $x(v, t)$ 为样品在外扰为 t 时变量 v 处的光谱强度,而 $\bar{x}(v)$ 则为整个外扰过程所得到的样品光谱中变量 v 处的光谱强度平均值^[19]。

同步相关光谱代表两个变量 v_1, v_2 处光谱强度随外扰而产生变化的相似性^[20]。二维相关近红外同步谱可以表示为:

$$\Phi(v_1, v_2) = \frac{1}{m-1} A^T A \quad (2)$$

式中, $\Phi(v_1, v_2)$ 为 (v_1, v_2) 处相关强度, A 为动态光谱矩阵, m 为矩阵行数,在本文中 $m=7$ 。

同步相关谱对角线上的峰称为自动峰,其强度代表了该变量处光谱强度的变化程度^[21]。非对角线上的峰称为交叉峰,峰值有正负之分,代表了在外扰作用下两个变量处光谱强度的变化是正相关还是负相关^[22]。

1.3 判别方法

首先对样品进行划分,每类稻米随机选择6份作为校正集,而剩余4份样品作为预测集。经过光谱采集和二维相关计算得到各样品的二维相关谱后,求取校正集样品二维谱的平均值作为该种类稻米的标准二维相关谱。预测集各个样品的二维相关谱分别与5个求得的标准二维相关谱求取欧式距离,将预测集样品归入与其距离最小的标准谱对应的稻米类别。具体计算公式如下:

$$D(A, B) = \sqrt{\sum_{i=1}^n \sum_{j=1}^n (a_{ij} - b_{ij})^2} \quad (3)$$

式中, A, B 分别为要求取欧氏距离的两个目标矩阵, a_{ij} 与 b_{ij} 分别为两个矩阵的元素,最终可以得到两个矩阵的欧式距离 $D(A, B)$ 。

2 结果与分析

2.1 同步二维相关近红外谱

仪器在完成背景测量和标准漫反射白板参考

光谱测量后,可以采集到样品的漫反射吸收谱。在不同温度下采集到的同一样品的光谱曲线会发生变化,经过基线校正后,某样品的 7 条在不同温度条件下的光谱如图 1 所示,光谱已经进行了基线校正预处理。随着温度的不断升高,光谱在整体趋势上呈现下降的趋势,这一点在峰值位置尤其明显。这是由于稻米所含化学物质受到温度外扰所引起的图谱变化。以平均光谱作为参考光谱,可以

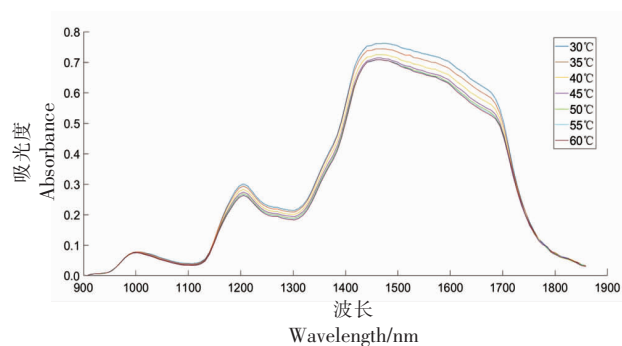
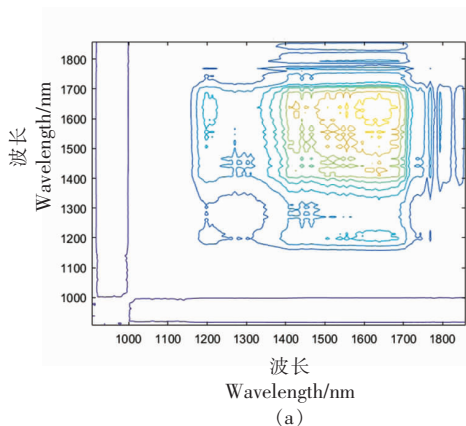


图 1 同一样品在不同温度下一维光谱示例
Fig.1 Examples of one-dimensional spectra of the same sample at different temperatures

使用 Matlab 进行相关计算,对校正集样品的二维相关谱求取平均后,图 3a~3e 分别为以等高线形式展现的 5 种样品的标准二维相关谱,颜色越深代表数值越小,颜色越明亮代表数值越大。从图中可以看出,不同类别的同步二维相关谱十分相似,副对角线上自动峰的位置也与一维光谱中峰值的位置对应,并且可以通过交叉峰来看出不同峰值间受温度外扰影响下变化的关联性,可以



得到如图 2 所示的动态光谱。从动态光谱中可以看出更加清晰地看出在 1 205,1 405,1 650 nm 附近处的下降趋势尤其明显。根据相关研究报告,近红外光谱在 1 205 nm 附近的吸收峰是由于-CH₃的-CH 键二级倍频振动造成的,而 1 405 nm 附近的吸收峰与直链淀粉分子中 O-H 基团的反对称和对称振动的一级倍频有关^[23-24]。

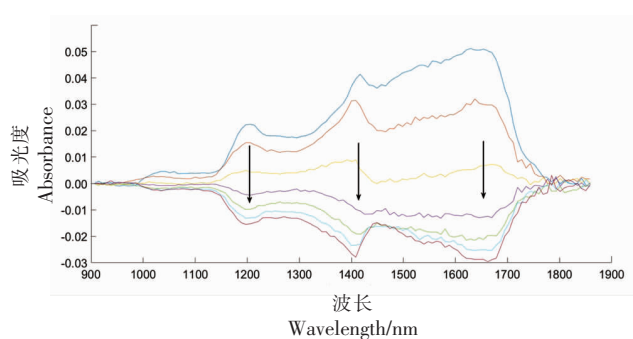
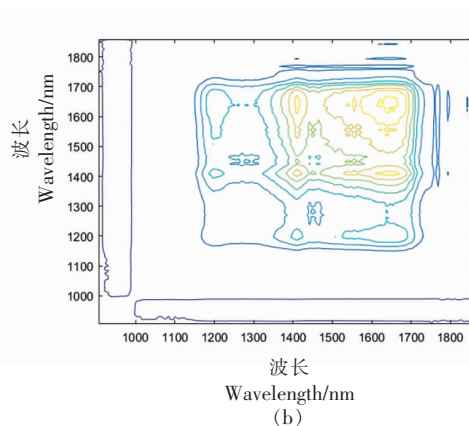


图 2 动态光谱示例
Fig.2 Examples of dynamic spectra

得到更多一维光谱中无法获得的有用信息。图 3b、3c、3e 在(1 405,1 650)及其沿着副对角线对称位置附近皆可以观察到明显的交叉峰存在,而其余两种样品的标准二维相关谱在该位置的交叉峰并不明显,说明不同样品虽然整体类似,但其内部物质含量仍存在细微差别,在受到外扰影响后得以更好地体现。



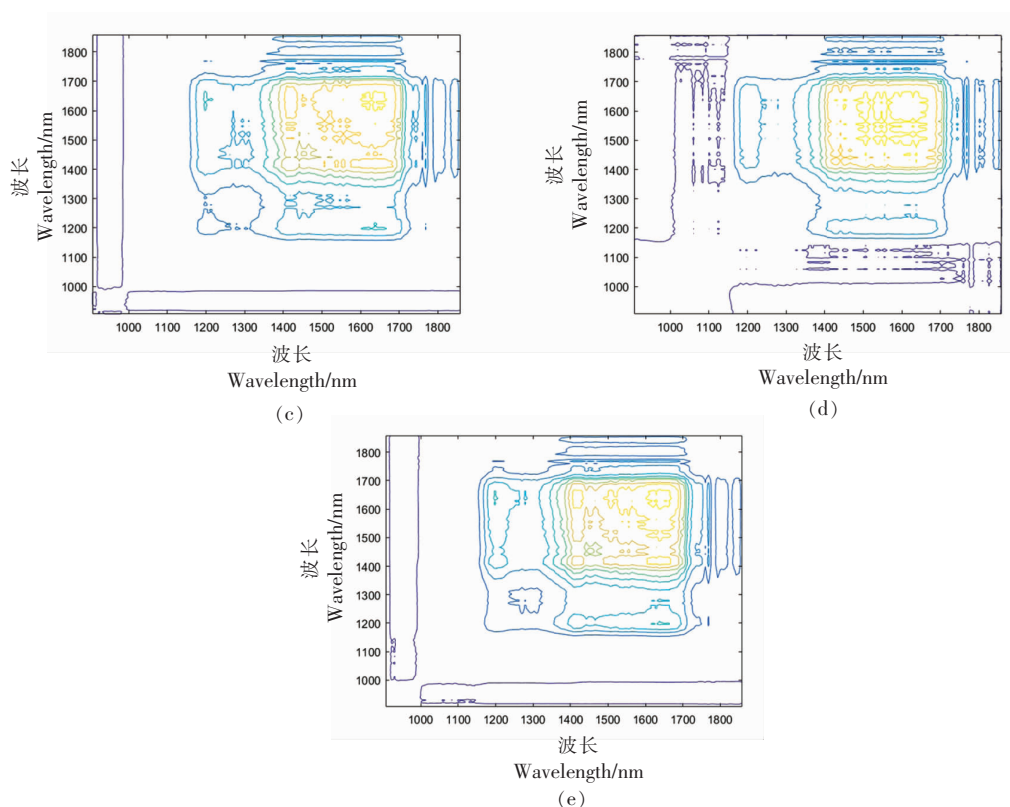


图3 不同类别样品的标准二维相关谱

Fig.3 Standard two-dimensional correlation spectra of different types of samples

2.2 判别结果

分别计算了测试集中各个样品的二维相关近红外谱与5种样品各自的标准二维相关谱之间的欧氏距离,并且根据距离最小值进行了类别归属的判断,表2展示了具体的结果。从表中可以看出,整体上除了样品2-2被误分类至类别5之外,其余所有测试集样品皆被正确分类,整体判别准确率为95%。值得注意的是,除了同类样品之外,所有其余待测样品与第3类样品的标准二维相关谱的欧式距离都是最大的,说明第3类样品与其余所有样品的差异最大。通过表1可以看出,第3类样品产地为宁夏回族自治区,其气候与地质条件与其余4处差别最大。针对样品2-2,其与各个标准相关谱的距离与同类别的其它测试集样品存在较大差异,而且并没有像其它样品一样显现出与第3类样品的显著差异。通过回溯试验记录,发现在进行该样品光谱采集时升温阶段的等待间隔存在较大误差,在动态光谱采集过程中导致样品的温度没有均匀上升,从而无法得到标准的二维

相关光谱,最终造成误判。

为了更好地体现每种样品的判别效果,定义一个区分度参数:

$$dis_i = \frac{d_i - d'_{\min}}{d_i} \quad (4)$$

式中, d_i 为第*i*个测试集样品与同类标准二维相关谱的欧氏距离, d'_{\min} 代表该样品与其余类别的标准二维相关谱欧氏距离中的最小值, dis_i 则为第*i*个测试集样品的区分度参数,该参数数值越大,代表此样品的区分度越高。

按照公式(4)来计算每类测试集样品的区分度,以整体评判该判别方式对各类别样品的整体判别效果。为了避免因为不良样品所造成的影响,在此环节去除掉样品2-2的结果。经过计算,5类样品的平均区分度参数分别为:0.31521,1.25654,2.3535,0.15369,0.4547。由此可以看出,该判别方式对第3类样品的区分度最高,其次则是第2类样品,这两种样品的平均区分度参数均大于1。其余3种样品的平均区分度较低,尤其是第4类样

表 2 判别结果
Table 2 Discrimination results

	1	2	3	4	5	判别归属
1-1	0.0032	0.0050	0.0273	0.0037	0.0073	1
1-2	0.0033	0.0075	0.0231	0.0056	0.0051	1
1-3	0.0031	0.0038	0.0261	0.0042	0.0060	1
1-4	0.0027	0.0049	0.0232	0.0063	0.0036	1
2-1	0.0060	0.0029	0.0247	0.0080	0.0056	2
2-2	0.0140	0.0163	0.0107	0.0169	0.0102	5
2-3	0.0059	0.0019	0.0269	0.0068	0.0070	2
2-4	0.0058	0.0030	0.0240	0.0080	0.0052	2
3-1	0.0384	0.0408	0.0143	0.0412	0.0346	3
3-2	0.0278	0.0303	0.0041	0.0307	0.0241	3
3-3	0.0257	0.0273	0.0057	0.0285	0.0215	3
3-4	0.0160	0.0183	0.0090	0.0188	0.0121	3
4-1	0.0058	0.0090	0.0287	0.0043	0.0100	4
4-2	0.0115	0.0096	0.0351	0.0091	0.0148	4
4-3	0.0092	0.0076	0.0328	0.0068	0.0125	4
4-4	0.0100	0.0082	0.0335	0.0075	0.0132	4
5-1	0.0071	0.0072	0.0211	0.0091	0.0042	5
5-2	0.0056	0.0059	0.0236	0.0065	0.0049	5
5-3	0.0080	0.0104	0.0167	0.0114	0.0043	5
5-4	0.0054	0.0059	0.0236	0.0063	0.0048	5

品的区分度最差。

2.3 其它常用判别方法比较

采用了其它两种常用建模方法进行了判别尝试。在同样的样品集划分情况下,对 30 ℃和 60 ℃条线下采集的样品光谱进行偏最小二乘判别分析和支持向量机分类尝试。偏最小二乘判别分析(PLS-DA)是一种线性分类方法,它结合了 PLS 回归的特性和分类技术的鉴别能力^[25]。它可以减少变量间多重共线性的影响,提高模型的判别性能^[26]。支持向量机(SVM)是一种监督学习算法,用于解决模式识别中的数据分类问题。它能解决高维问题,具有泛化误差小、易于解释、计算复杂度低等优点^[27]。当低维输入空间中的线性边界不足以区分两类时,SVM 算法在高维空间中建立一个超平面或一系列超平面进行分类^[28]。本研究选取线性核函数作为支持向量机的核函数,采用网格搜索法确定参数 C 为 7 742.645。

混淆矩阵可以直观地显示分类结果。矩阵的行和列分别代表试验类和预测类,矩阵的元素 m_{ij}

是属于 i 类且被分类为 j 类的样本数。因此,对角线元素表示正确分类的样本数,而其它元素则为被错误分类的样本数^[29]。

具体的分类结果如图 4 所示,其中(a)、(b)分别为 30,60 ℃时 PLS-DA 模型的判别结果,在 2 种温度条件下的一维光谱数据构造的模型判别准确率皆为 75%,明显低于二维相关判别方法得到的结果。该判别方法可以比较好地区分 1、2 两类样品,但对于其余类别样品的判别效果不佳。图 4c、4d 为 30,60 ℃时 SVM 模型的判别结果,准确率分别为 70%和 65%,不能对第 1 类样品外的其余样品进行很好地区分。

3 结论

本文针对不同产地的稻米样品无损鉴别问题,借助二维相关近红外光谱的技术开展了研究。对待测样品进行类别和数据集划分后,进行了二维相关分析,并提出了一种通过设定标准二维相关谱的新方法,依据欧氏距离最近归属同类的准

	1	2	3	4	5
1	4	0	0	0	0
2	0	4	0	0	0
3	0	1	3	0	0
4	0	0	2	2	0
5	0	0	0	2	2

(a)

	1	2	3	4	5
1	4	0	0	0	0
2	0	4	0	0	0
3	0	1	3	0	0
4	0	0	1	2	1
5	0	0	0	2	2

(b)

	1	2	3	4	5
1	4	0	0	0	0
2	0	3	0	0	1
3	1	0	3	0	0
4	0	0	2	2	0
5	0	0	0	2	2

(c)

	1	2	3	4	5
1	4	0	0	0	0
2	0	3	0	0	1
3	1	0	3	0	0
4	0	0	2	1	1
5	0	0	0	2	2

(d)

图4 PLS-DA 和 SVM 模型的分类结果

Fig.4 Discrimination results of PLS-DA and SVM models

则,对测试集样品进行了判别分析。最终得到的结果显示,该方法可以达到 95% 的整体判别准确率,并且针对第 2、3 类样品的判别效果尤其突出。为了更好地比较二维相关方法和常用的化学计量分析方法的判别效果,还使用 PLS-DA 和 SVM 两种策略进行了尝试。经过实际测试,两种判别方法在同样的数据条件下,正确率最高也仅为 75%,明显低于二维相关方法。因此,使用二维相关近红外光谱进行稻米样品的无损判别是一种有效的手段,它在不损伤样品的情况下可以进行准确的判别分析,而且不需要复杂的建模过程,能够为稻米的在线检测提供参考。

参 考 文 献

[1] 杨万江, 刘琦. “粮安天下”: 中国大米供求变化预测[J]. 农业现代化研究, 2019, 40(1): 44-53.
YANG W J, LIU Q. Grain security: Projections for rice supply and demand in China[J]. Research of Agricultural Modernization, 2019, 40(1): 44-53.

[2] 王亚轩, 谭峰, 辛元明, 等. 大米拉曼光谱不同预处理方法的相近产地鉴别研究[J]. 光谱学与光谱分析, 2021, 41(2): 565-571.
WANG Y X, TAN F, XIN Y M, et al. Identification of rice from similar areas with different pre-treatment methods of Raman spectrum [J]. Spec-

troscopy and Spectral Analysis, 2021, 41(2): 565-571.

- [3] 王树婷, 陈志, 张敏, 等. 国家地理标志产品洋县黑米保护的思考[J]. 陕西农业科学, 2010, 56(1): 114-118.
WANG S T, CHEN Z, ZHANG M, et al. Thoughts on the protection of Yangxian black rice, a national geographical indication product[J]. Shaanxi Journal of Agricultural Sciences, 2010, 56(1): 114-118.
- [4] 沙敏, 桂冬冬, 张正勇, 等. 拉曼光谱结合模式识别方法鉴别大米种类[J]. 中国粮油学报, 2020, 35(1): 168-174.
SHA M, GUI D D, ZHANG Z Y, et al. Identification of rice varieties by Raman spectroscopy and pattern recognition[J]. Journal of the Chinese Cereals and Oils Association, 2020, 35(1): 168-174.
- [5] 胡圣英, 任红波, 张军, 等. 大米产地溯源方法研究进展[J]. 中国农学通报, 2020, 36(14): 148-155.
HU S Y, REN H B, ZHANG J, et al. Research progress of rice origin traceability method[J]. Chinese Agricultural Science Bulletin, 2020, 36(14): 148-155.
- [6] 秦延春. DNA 分子标记及其在水稻遗传育种研究中的应用[J]. 广东农业科学, 2012, 39(7): 145-147.
QIN Y C. DNA molecular marker and application in genetics and breeding study of rice [J]. Guangdong Agricultural Sciences, 2012, 39(7): 145-147.
- [7] 袁定阳, 孙志忠, 谭炎宁, 等. 不同温度下水稻胚乳 AGPase 各同工酶基因表达特征初步分析[J]. 杂交水稻, 2011, 26(4): 61-64, 85.
YUAN D Y, SUN Z Z, TAN Y N, et al. A preliminary analysis of gene expression of seven isoforms of ADP-glucose Pyro phosphorylase (AGPase) in rice endosperm under different temperature conditions[J]. Hybrid Rice, 2011, 26(4): 61-64, 85.
- [8] 杨健, 张星灿, 华苗苗, 等. 稻米全产业链可追溯关键技术研究进展[J]. 粮油食品科技, 2020, 28(2): 43-48.
YANG J, ZHANG X C, HUA M M, et al. Progress in key technologies for traceability of rice industry chain[J]. Science and Technology of Cereals, Oils and Foods, 2020, 28(2): 43-48.
- [9] 董泳江. 光谱和光谱成像技术在食品检测中的应用研究[D]. 杭州: 浙江大学, 2017.
DONG Y J. Spectroscopy and spectral imaging technology and their applications on food detection[D].

- Hangzhou: Zhejiang University, 2017.
- [10] 彭丽君. 水稻种子鉴别的近红外光谱快速无损分析[D]. 广州: 暨南大学, 2018.
PENG L J. Rapid nondestructive analysis of rice seed identification by near infrared spectroscopy[D]. Guangzhou: Jinan University, 2018.
- [11] 高地. 基于近红外光谱技术对松原大米的品种判别及产地确证探究[D]. 长春: 吉林农业大学, 2019.
GAO D. Variety identification and origin confirmation of Songyuan rice based on near infrared spectroscopy[D]. Changchun: Jilin Agricultural University, 2019.
- [12] 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、五常稻米产地溯源[J]. 食品科学, 2018, 39(16): 321-327.
QIAN L L, SONG X J, ZHANG D J, 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.
- [13] 林莹. 基于近红外光谱技术对辉南稻米产地确证的研究[D]. 长春: 吉林农业大学, 2018.
LIN Y. Confirmation of Huinan rice origin based on near infrared spectroscopy[D]. Changchun: Jilin Agricultural University, 2018.
- [14] WANG C Y, XIANG B R, ZHANG W. Application of two-dimensional near-infrared (2D-NIR) correlation spectroscopy to the discrimination of three species of *Dendrobium*[J]. Journal of Chemometrics: A Journal of the Chemometrics Society, 2009, 23(9): 463-470.
- [15] 陆珺, 相秉仁, 刘浩. 二维相关近红外光谱及其应用[J]. 药学进展, 2007, 31(7): 303-308.
LU J, XIANG B R, LIU H. Two-dimensional near-infrared correlation spectroscopy: theory and application[J]. Progress in Pharmaceutical Sciences, 2007, 31(7): 303-308.
- [16] ZHANG Y L, CHEN J B, LEI Y, et al. Discrimination of different red wine by Fourier-transform infrared and two-dimensional infrared correlation spectroscopy[J]. Journal of Molecular Structure, 2010, 974(1/2/3): 144-150.
- [17] SOHNG W, PARK Y, JANG D, et al. Incorporation of two-dimensional correlation analysis into discriminant analysis as a potential tool for improving discrimination accuracy: Near-infrared spectroscopic discrimination of adulterated olive oils[J]. Talanta, 2020, 212: 120748.
- [18] 杨仁杰, 杨延荣, 董桂梅, 等. 基于二维红外相关谱欧氏距离判别掺杂牛奶[J]. 光谱学与光谱分析, 2014, 34(8): 2098-2101.
YANG R J, YANG Y R, DONG G M, et al. Discrimination of adulterated milk based on Euclidian distances between two-dimensional infrared correlation spectra[J]. Spectroscopy and Spectral Analysis, 2014, 34(8): 2098-2101.
- [19] NODA I, DOWREY A E, MARCOTT C, et al. Generalized two-dimensional correlation spectroscopy[J]. Applied Spectroscopy, 2000, 54(7): 236A-248A.
- [20] PARK Y, KIM M, CHUNG H, et al. The study of pH effects on phase transition of multi-stimuli responsive P (NiPAAm-co-AAc) Hydrogel using 2D-COS[J]. Polymers, 2021, 13(9): 1447.
- [21] 王月. 基于红外光谱与化学计量学对中药的鉴定方法研究[D]. 北京: 北京中医药大学, 2018.
WANG Y. Study on identification method of traditional Chinese medicine based on infrared spectroscopy and chemometrics[D]. Beijing: Beijing University of Chinese Medicine, 2018.
- [22] 陈建波. 二维相关红外光谱差异分析方法及其应用研究[D]. 北京: 清华大学, 2010.
CHEN J B. Methods and application of differentiation of two-dimensional correlation infrared spectroscopy[D]. Beijing: Tsinghua University, 2010.
- [23] SAMPAIO P S, SOARES A, CASTANHO A, et al. Optimization of rice amylose determination by NIR-spectroscopy using PLS chemometrics algorithms[J]. Food Chemistry, 2018, 242: 196-204.
- [24] BAGCHI T B, SHARMA S, CHATTOPADHYAY K. Development of NIRS models to predict protein and amylose content of brown rice and proximate compositions of rice bran[J]. Food Chemistry, 2016, 191: 21-27.
- [25] BALLABIO D, CONSONNI V. Classification tools in chemistry. Part 1: linear models. PLS-DA[J]. Analytical Methods, 2013, 5(16): 3790-3798.
- [26] WANG Y Y, LI J Q, LIU H G, et al. Attenuated total reflection-Fourier transform infrared spectroscopy (ATR-FTIR) combined with chemometrics methods for the classification of Lingzhi species[J]. Molecules, 2019, 24(12): 2210.

- [27] HE Y C, JIANG H, CHEN Q S. High-precision identification of the actual storage periods of edible oil by FT-NIR spectroscopy combined with chemometric methods [J]. *Analytical Methods*, 2020, 12 (29): 3722–3728.
- [28] ANYIDOHO E K, TEYE E, AGBEMAFLE R. Non-destructive authentication of the regional and geographical origin of cocoa beans by using a handheld NIR spectrometer and multivariate algorithm[J]. *Analytical Methods*, 2020, 12(33): 4150–4158.
- [29] HARK R R, THROCKMORTON C S, HARMON R S, et al. Multianalyzer spectroscopic data fusion for soil characterization[J]. *Applied Sciences*, 2020, 10 (23): 8723.

Identification of Rice from Different Origins by Two-dimensional Correlation Near Infrared Spectroscopy

Dai Yuanfeng^{1,3}, Dai Zuoxiao^{1*}, Guo Guangzhi¹, Wang Yingchao²

¹Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083

²Taicang Institute of Optoelectronics Technology, Suzhou 215411, Jiangsu

³University of Chinese Academy of Sciences, Beijing 100049)

Abstract In order to realize the nondestructive discrimination of rice samples from different origins, a discrimination method based on synchronous two-dimensional correlation near infrared spectroscopy and Euclidean distance was proposed. After obtaining the standard two-dimensional correlation spectra of five types of rice samples respectively, the Euclidean distance between the test set samples and each standard correlation spectrum is calculated, and the discrimination is based on the minimum distance. The accuracy of this method can reach 95%. By defining the discrimination coefficient, the discrimination effect of various types of samples can also be judged more accurately. In order to compare the practical effects of two-dimensional correlation method and common modeling methods, partial least squares discriminant analysis and support vector machine classification are tried on the basis of the same data. The highest accuracy rates are 75% and 70% respectively, which verifies the effectiveness of two-dimensional correlation method. This study provides a possible method for nondestructive identification of rice samples, and can provide a reference for on-line detection.

Keywords two dimensional correlation spectroscopy; near infrared; rice; Euclidean distance; identification analysis