



邹华,潘圆,强海清,等.基于HPLC-DAD三维色谱指纹图谱的国产赤霞珠红葡萄酒产地识别研究[J].轻工学报,2026,41(3):10-18.
ZOU H,PAN Y,QIANG H Q,et al.Geographical origin identification of Chinese Cabernet Sauvignon red wines based on HPLC-DAD three-dimensional chromatographic fingerprinting[J].Journal of Light Industry,2026,41(3):10-18. DOI:10.12187/2026.03.002

基于HPLC-DAD三维色谱指纹图谱的国产赤霞珠红葡萄酒产地识别研究

邹华,潘圆,强海清,孟展,尹小丽,谷惠文

长江大学 化学与环境工程学院/生命科学学院,湖北 荆州 434023

摘要:【目的】丰富与发展葡萄酒真实性鉴别及产地溯源技术。【方法】采用高效液相色谱-光电二极管阵列检测(HPLC-DAD)法采集我国秦皇岛、银川和吐鲁番3个著名产地共45个赤霞珠红葡萄酒的三维色谱指纹图谱,并借助化学计量学多元曲线分辨-交替最小二乘(MCR-ALS)算法解析上述图谱数据,基于所得有效组分的相对浓度,利用主成分分析(PCA)、偏最小二乘-判别分析(PLS-DA)和支持向量机(SVM)3种机器学习算法对红葡萄酒样品进行产地区分。【结果】通过MCR-ALS解析算法共获得56个有效组分,PCA得分显示各产地赤霞珠红葡萄酒有依照产地分类的趋势,且PLS-DA和SVM模型的分类效果均良好,训练集和预测集识别准确率都可达100%。此外,基于22个差异变量建立的VIP-PLS-DA模型同样可对葡萄酒产地进行100%的准确判别。【结论】HPLC-DAD三维色谱指纹图谱技术结合机器学习算法能建立稳定可靠的识别模型,实现3个国产赤霞珠红葡萄酒产地客观、准确的鉴别。

关键词:赤霞珠红葡萄酒;HPLC-DAD;三维色谱指纹图谱;产地识别;机器学习

中图分类号:TS262.6 **文献标识码:**A **文章编号:**2096-1553(2026)03-0010-09

0 引言

作为新兴的葡萄酒生产国,中国展现出极大的葡萄酒消费潜力。据国际葡萄和葡萄酒组织(OIV)统计,2024年中国的葡萄种植面积为753 000 hm²,成为继西班牙和法国之后的世界第三大葡萄种植国^[1]。中国已形成多个主要葡萄酒产区,包括东北、京津冀、山东、黄河故道、黄土高原、内蒙古、贺兰山东麓、河西走廊、新疆、西南高山,以及广西壮族自治区和湖南省的特色产区^[2]。由于各葡萄酒

产区在地理环境和气候条件上存在显著差异,不同产区所酿造的葡萄酒表现出鲜明的“风土特征”,有助于定义葡萄酒产区的区域价值、身份和声誉^[3]。准确标注葡萄酒的产区信息,既是对消费者权益和葡萄酒市场的保护,也是推进中国葡萄酒产区保护的重要前提。然而,受经济利益驱使,葡萄酒原产地造假事件时有发生,这不仅严重损害了消费者的合法权益,也破坏了葡萄酒消费市场的声誉。因此,亟需建立有效的葡萄酒产地鉴别方法,为产地溯源提供可靠的技术支撑。

收稿日期:2025-05-06;修回日期:2025-09-21;出版日期:2026-06-15

基金项目:国家自然科学基金项目(32371501;32272409);宁夏回族自治区重点研发计划重点项目(2026BEG02035)

作者简介:邹华(1977—),女,湖北省荆州市人,长江大学实验师,主要研究方向为仪器分析化学。E-mail:70401773@qq.com

通信作者:谷惠文(1989—),男,湖南省耒阳市人,长江大学教授,博士,主要研究方向为食品安全与真实性检测及溯源。

E-mail:gruyclewee@yangtzeu.edu.cn

色谱指纹图谱技术是一种具有高灵敏性、高检测效率的分析手段,能够检测约 70% 的已知有机化合物^[4]。其中,三维色谱指纹图谱不仅能克服基于单波长或单一类别物质构建的色谱指纹图谱难以全面反映分析对象质量属性的不足,还可以借助三维数据分析算法,利用数据间的相关性与互补性,有效解决色谱峰重叠、时间漂移、基线漂移等问题,进一步结合多元统计分析方法,则能够更大限度地提取色谱指纹图谱中隐含的特征信息。例如,T. Q. Peng 等^[5]利用高效液相色谱-光电二极管阵列检测(HPLC-DAD)三维色谱指纹图谱技术,成功实现了崂山绿茶的等级区分;H. W. Gu 等^[6]采用 HPLC-DAD 三维色谱指纹图谱结合主成分分析(PCA)和正交偏最小二乘-判别分析(OPLS-DA),不仅成功区分了浙江和山东的绿茶,还筛选出 17 种产地特征标志物。然而,目前关于我国不同产地葡萄酒的三维色谱指纹图谱研究鲜见报道。

鉴于此,本研究拟基于 HPLC-DAD 三维色谱指纹图谱技术,结合机器学习算法,对我国秦皇岛、银川和吐鲁番 3 个著名产地的赤霞珠红葡萄酒进行色谱指纹图谱解析并建立 PCA、偏最小二乘-判别分析(PLS-DA)和支持向量机(SVM)产地判别模型,以期葡萄酒的真实性鉴别及产地溯源提供参考。

1 材料与方法

1.1 主要材料与试剂

甲醇(色谱级),美国 J. T. Baker 公司;甲酸(色谱级),上海阿拉丁生化科技股份有限公司;蒸馏水,广州屈臣氏食品饮料有限公司;45 瓶赤霞珠红葡萄酒,均从酿酒酒庄直接购入,产地分别是河北秦皇岛(QHD, 12 瓶)、宁夏银川(YC, 17 瓶)和新疆吐鲁番(TP, 16 瓶)。

1.2 主要仪器与设备

Ultimate 3000 型 HPLC-DAD 仪(配有 Ultimate 3000 RS 系列泵、自动进样系统、柱温箱及二极管阵列检测器),美国赛默飞世尔科技有限公司;0.22 μm 聚四氟乙烯膜,天津市科亿隆实验设备有限公司。

1.3 实验方法

1.3.1 葡萄酒样品预处理 在 N_2 氛围下,将所有

赤霞珠红葡萄酒重新分装至 50 mL 棕色小瓶中,随即密封于真空袋内抽真空,置于 $-18\text{ }^\circ\text{C}$ 冰箱中保存。在进行 HPLC-DAD 分析前,将红葡萄酒样品于 $4\text{ }^\circ\text{C}$ 下解冻,经 0.22 μm 聚四氟乙烯膜过滤后待用;分别取 45 个样本各 500 μL ,等体积混合均匀,制得质控(QC)样本;随机选取其中 37 个样本作为训练集,用于建立 PLS-DA 和 SVM 模型;剩余 8 个样本作为预测集,用于模型验证。

1.3.2 HPLC-DAD 检测条件 色谱柱为 Accucore™ aQ C18 液相色谱柱(150 mm \times 2.1 mm \times 2.6 μm);柱温为 $40\text{ }^\circ\text{C}$;DAD 的扫描波长范围为 200~600 nm,步长为 1 nm,频率为 2 Hz;流动相为 0.1% (若无特指,百分数均指体积分数,下同)甲酸溶液(A)和甲醇(B)。梯度洗脱程序设置为 0~20 min,5%~45%B;20~25 min,45%~60%B;25~28 min,60%~95%B;28~30 min,95%B;30~31 min,95%~5%B;31~40 min,5%B。流速为 0.3 mL/min;进样体积为 5 μL 。

1.4 数据处理与分析

采用 HPLC-DAD 仪自带的 Xcalibur 4.0 软件采集三维色谱指纹图谱原始数据,数据文件以 .raw 格式保存;利用 Chromeleon 软件将原始数据转换为文本格式(.txt),以便进行后续化学计量学分析^[7]。数据分析流程如下:首先,在 Matlab 环境下采用不对称最小二乘法(AsLS)对所有样本的三维色谱指纹图谱进行基线漂移校正^[8],通过 MCR-ALS GUI 2.0 图形用户界面进行多元曲线分辨率-交替最小二乘(MCR-ALS)分析,将整体三维色谱指纹图谱分解为各单一组分的色谱与光谱轮廓^[9-10];然后,对所得各化学成分色谱图进行积分处理,构建相对浓度矩阵 $\mathbf{X}(m \times n)$,其中 m 为样本数, n 为化学成分数;最后,采用 PCA、PLS-DA 和 SVM 3 种机器学习算法,基于 $\mathbf{X}(m \times n)$ 建立模型,用于葡萄酒样品的分类与产地识别。其中,PCA 和 PLS-DA 建模在 SIMCA-P 14.1 软件中进行,SVM 建模则使用 LIBSVM 软件包在 Matlab 环境中实现^[11]。

2 结果与讨论

2.1 三维色谱指纹图谱分析

根据原始 HPLC-DAD 三维色谱指纹图谱结果,

去除基线漂移严重及无色谱峰的时间段,最终对保留时间为 0.5 ~ 31.1 min, 紫外光谱范围为 242 ~ 600 nm 的指纹图谱进行解析。中国赤霞珠红葡萄酒的 HPLC-DAD 三维色谱指纹图谱如图 1 所示。由图 1a) 可知, 未处理的原始 HPLC-DAD 三维色谱指纹图谱存在明显的基线漂移。尽管引起基线漂移的原因较为复杂且难以完全消除, 但借助 AsLS 算法^[12-13]可有效改善基线漂移, 获得较为平滑的基线图谱(图 1b)), 从而避免基线漂移对数据造成不良影响。

在 MCR-ALS 解析过程中, 组分数的估算会直接影响后续定性分析的准确性。为提高分析速度及准确度, 首先根据色谱峰的保留时间将每个样本的 HPLC-DAD 三维色谱指纹图谱划分为 17 个色谱子区域; 随后, 基于奇异值分解(SVD)和欠拟合度(LOF)对各子区域的组分数进行估计; 最后, 将以纯粹变量方法提取的光谱作为初始值, 设定 10% 的允许噪声水平, 启动色谱和光谱轮廓的优化程序。在

MCR-ALS 迭代计算过程中, 对色谱和光谱轮廓均施加非负约束, 并对光谱进行归一化处理。

MCR-ALS 分辨色谱子区域 S07 的结果如图 2 所示, 该区域在原始色谱图中仅呈现一个峰。图 2a) 为 57 个红葡萄酒样本(45 个红葡萄酒样本+12 个 QC 样本)的分辨扩增色谱曲线; 由图 2b) 可知, 蓝色色谱曲线完全覆盖了黄色色谱曲线, 且后者的响应强度远低于前者; 由图 2c) 可知, 各组分的的光谱曲线存在较为明显的重叠现象。即便如此, MCR-ALS 算法仍能成功解析出 4 个组分的色谱图和光谱图。本研究进一步采用 MCR-BANDS 软件对分辨结果的旋转模糊程度进行评估^[14], 结果见表 1。由表 1 可知, 各色谱段的 R^2 均大于 99.42%, LOF 值均低于 3.0%, 表明模型具有较好的分辨率及稳健性。此外, f_{\max} 与 f_{\min} 之间的差值极小, 表明模糊旋转程度已得到极大程度的消除, 进一步验证了 MCR-ALS 解析结果的可靠性。通过上述解析流程, 最终从 45 个红葡萄酒样本的 17 个色谱子区域中获得 56 个有效组分。

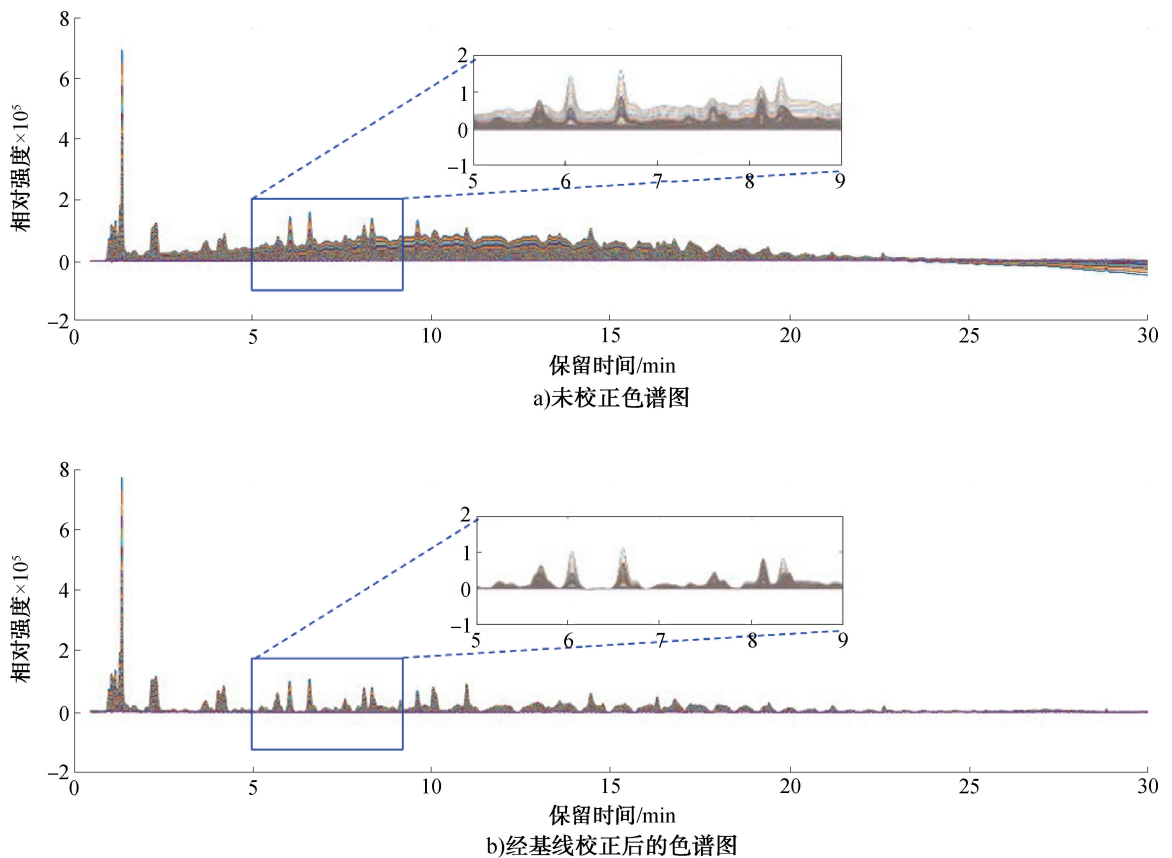


图 1 中国赤霞珠红葡萄酒的 HPLC-DAD 三维色谱指纹图谱

Fig. 1 HPLC-DAD three-dimensional chromatographic fingerprint of Chinese Cabernet Sauvignon red wine

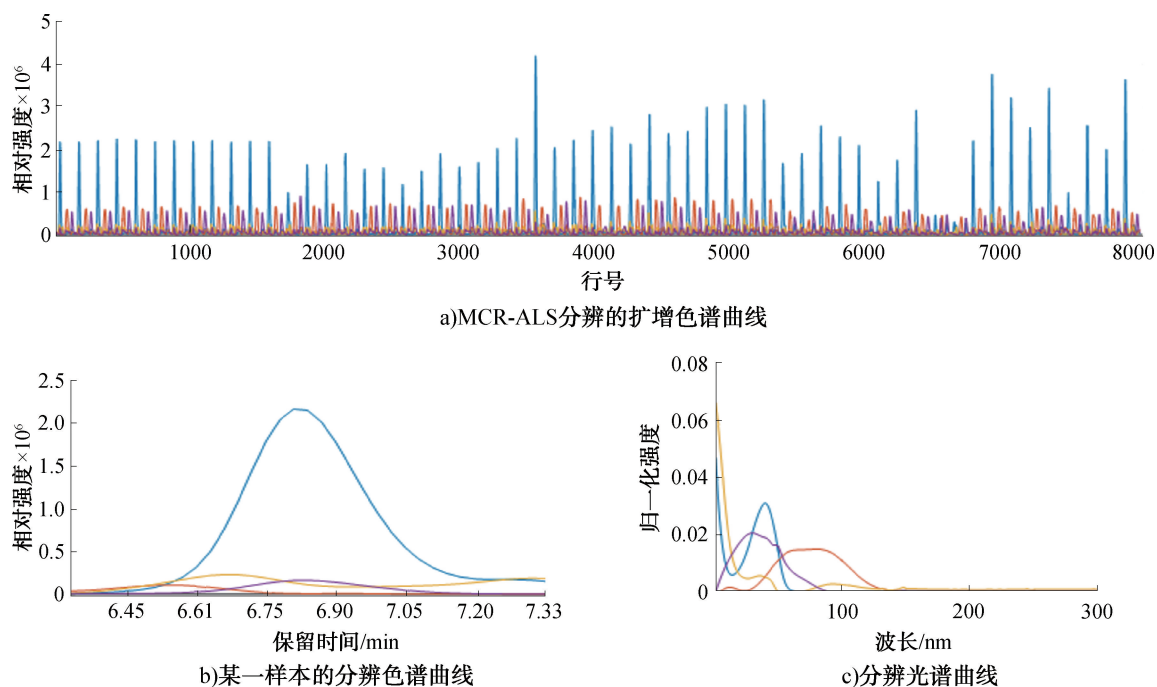


图 2 MCR-ALS 分辨色谱子区域 S07 的结果

Fig. 2 MCR-ALS resolution results for chromatographic segment S07

表 1 各色谱子区域的组分数估计结果与旋转模糊性评估结果

Table 1 Estimated number of components and rotational ambiguity assessment results for each chromatographic segment

子区域	保留时间/min	LOF 值/%	R^2 /%	化合物数	f_{\max} 与 f_{\min} 之间的差值
S01	0.8~1.5	0.580 8	99.94	5	0/0/0/0/0
S02	1.6~3.0	0.463 1	99.86	3	$3.22 \times 10^{-4} / 3.72 \times 10^{-5} / 1.48 \times 10^{-4}$
S03	3.0~4.0	1.900 8	99.42	4	0/0/0/0
S04	4.0~4.4	0	99.91	1	0
S05	5.0~5.5	2.611 8	99.48	3	$4.50 \times 10^{-3} / 5.31 \times 10^{-3} / 4.91 \times 10^{-4}$
S06	5.5~6.3	1.130 9	99.91	4	$1.68 \times 10^{-3} / -1.12 \times 10^{-2} / 3.43 \times 10^{-4} / -2.52 \times 10^{-5}$
S07	6.3~7.5	1.469 2	99.87	4	$1.24 \times 10^{-4} / 8.52 \times 10^{-5} / 1.34 \times 10^{-5} / 6.59 \times 10^{-5}$
S08	7.5~8.6	0.728 4	99.81	4	$6.11 \times 10^{-2} / 4.49 \times 10^{-2} / 3.10 \times 10^{-2} / 0.31$
S09	8.8~9.4	0.593 8	99.90	3	$3.40 \times 10^{-4} / -1.06 \times 10^{-6} / 2.80 \times 10^{-4}$
S10	9.4~10.3	0.675 8	99.86	4	$2.32 \times 10^{-5} / 6.29 \times 10^{-4} / 2.73 \times 10^{-5} / 8.39 \times 10^{-4}$
S11	10.7~11.2	0.455 9	99.97	3	$1.17 \times 10^{-3} / 2.32 \times 10^{-2} / -1.40 \times 10^{-4}$
S12	14.0~14.7	0.654 1	99.78	2	$-1.67 \times 10^{-3} / 1.53 \times 10^{-5}$
S13	15.6~17.0	0.787 6	99.90	5	$-2.66 \times 10^{-3} / 2.08 \times 10^{-3} / 9.58 \times 10^{-4} / -2.06 \times 10^{-3} / 2.40 \times 10^{-2}$
S14	17.0~18.3	0.763 8	99.43	3	$6.88 \times 10^{-11} / 5.72 \times 10^{-3} / 1.00 \times 10^{-4}$
S15	18.9~20.0	0.812 2	99.73	3	$4.00 \times 10^{-3} / 2.90 \times 10^{-4} / 2.10 \times 10^{-2}$
S16	21.1~21.5	0.958 5	99.91	2	$-3.07 \times 10^{-2} / 3.34 \times 10^{-2}$
S17	22.2~22.8	0.342 2	99.87	3	$4.07 \times 10^{-12} / 2.02 \times 10^{-10} / 7.62 \times 10^{-11}$

2.2 不同产地红葡萄酒的机器学习建模结果分析

2.2.1 PCA 结果分析 对 MCR-ALS 解析获得的相对浓度矩阵 X 进行 PCA 分析,结果如图 3 所示。

由图 3 可知,经 UV 归一化处理后,PCA 模型前 9 个主成分的累积解释方差($R^2 X_{cum}$)为 0.847,说明采用 PCA 将 56 个有效成分变量降维到 9 个主成分后仍能保留大部分有效信息;PCA 得分图中心的

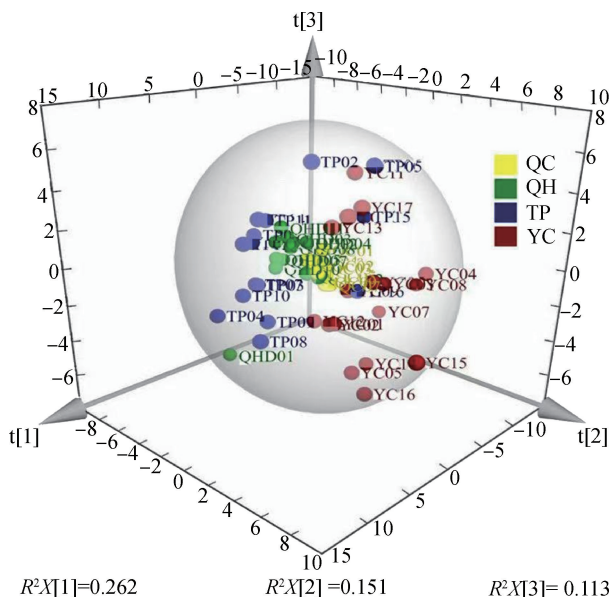


图3 MCR-ALS 解析相对浓度矩阵的 PCA 三维得分图

Fig. 3 PCA 3D score plot based on MCR-ALS resolved relative concentration matrix

QC 样本聚集较为紧密,表明所建立的 HPLC-DAD 法和 MCR-ALS 算法均具有良好的稳健性;秦皇岛产地的红葡萄酒除个别样本外,整体聚集较为紧密;而其余 2 个产地的红葡萄酒分布较为分散,且存在一定程度的重叠,但 3 个产地的红葡萄酒样本总体上仍可按地理来源区分。因此,基于 MCR-ALS 算法从 HPLC-DAD 三维色谱指纹图谱中分辨出的化学成分信息,可用于区分中国赤霞珠红葡萄酒的产地。

2.2.2 PLS-DA 结果分析 在食品真实性检测及溯源领域,PLS-DA 是一种常用的有监督模式识别方法^[15-16],可以识别不同类别样本之间的差异,并通过拆分变量的多维空间实现分类。MCR-ALS 解析相对浓度矩阵的 PLS-DA 得分图如图 4 所示。由图 4 可知,与 PCA 模型相比,银川和吐鲁番 2 个产地的样本沿 x 轴分离更为明显,且 3 个产地的样本分布在坐标系的不同象限。具体而言,秦皇岛样本主要分布于第一象限,吐鲁番样本主要分布于第四象限,银川样本则分布于第二和第三象限。对于经 UV 归一化处理的 PLS-DA 模型, X 矩阵和 Y 矩阵的累积解释方差分别为 57.7% ($R^2X_{cum} = 0.577$) 和 85.8% ($R^2Y_{cum} = 0.858$),模型的交叉验证预测能力 Q^2_{cum} 为 0.645。一般而言, R^2Y_{cum} 和 Q^2_{cum} 均大于

0.5,且越接近 1,表明模型的判别及预测能力越好。红葡萄酒产地判别的 PLS-DA 模型结果见表 2。由表 2 可知,PLS-DA 模型对训练集和预测集的判别准确率均达到 100%,因此本研究所建立的 PLS-DA 产地判别模型是有效的。

2.2.3 SVM 结果分析 采用高斯核函数的 SVM 模型作为一种非线性的有监督模式识别方法,广泛应用于食品真实性检测及溯源^[17-18]。为进一步探究不同产地葡萄酒之间的差异,并与 PLS-DA 建模结果进行比较,本研究采用与 PLS-DA 相同的训练集和预测集,建立葡萄酒产地的 SVM 判别模型。首先,对数据进行 Auto-scaling 预处理,通过网格搜索法计算 SVM 模型的惩罚因子 C 和核函数参数 γ ,优化搜索范围为 $2^{-8} \sim 2^8$,间隔为 1;同时结合留一交叉验证法选出核函数的最优参数组合,结果如图 5 所示。由图 5a)—b) 可知,随着迭代次数的累积,交叉验证准确率最高达 97.297 3%,对应的最优 C 和 γ 分别为 4 和 0.007 812 5。

红葡萄酒产地判别的 SVM 模型结果见表 3。由表 3 并结合图 5c)—d) 可知,训练集和预测集的红葡萄酒样本均能被准确识别。分类准确率均达到 100%,且模型灵敏度及特异性都等于 1。因此,SVM 模型能够准确识别秦皇岛、银川和吐鲁番 3 个产地的红葡萄酒。此外,对比 PLS-DA 和 SVM 模型

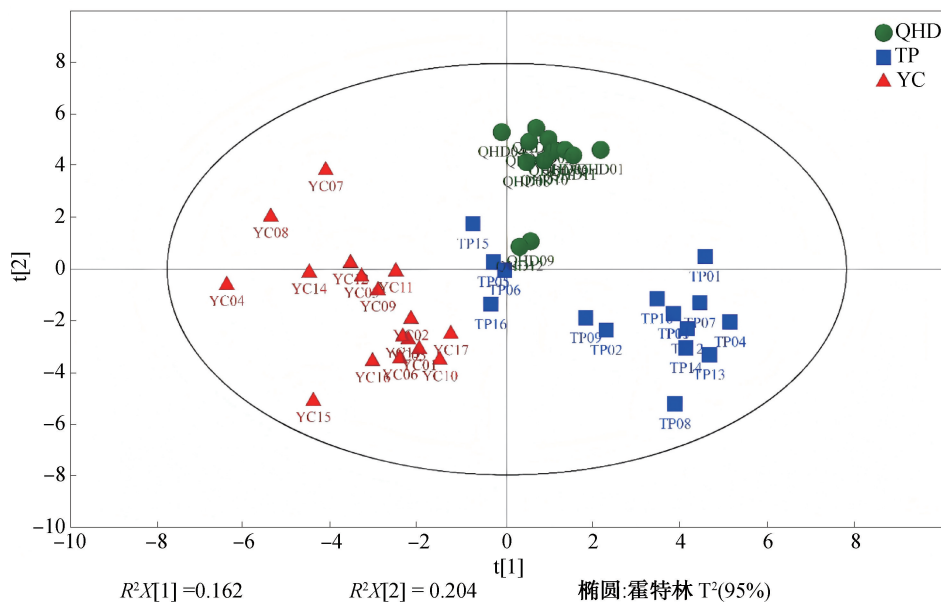


图 4 MCR-ALS 解析相对浓度矩阵的 PLS-DA 得分图

Fig. 4 PLS-DA score plot based on MCR-ALS resolved relative concentration matrix

表 2 红葡萄酒产地判别的 PLS-DA 模型结果

Table 2 Results of PLS-DA model for identifying the origins of red wines

数据集	产地	正确识别样本数	错误识别样本数	识别率/%	平均识别率/%	灵敏度	特异性
训练集	QHD	10	0	100	100	1	1
	YC	13	0	100		1	1
	TP	14	0	100		1	1
预测集	QHD	2	0	100	100	1	1
	YC	3	0	100		1	1
	TP	3	0	100		1	1

的判别结果,无论是使用线性算法的 PLS-DA 模型,还是使用非线性算法的 SVM 模型,在有监督模式下建立的产地判别模型均能对不同产地红葡萄酒的地理来源进行 100% 识别。

2.2.4 差异变量分析 为进一步明确影响建模的关键变量,以 $VIP > 1$ 和 $P < 0.05$ 为条件进行差异变量筛选,共获得 22 个差异变量,分别为: Var02、Var05、Var06、Var07、Var09、Var11、Var13、Var18、Var20、Var23、Var30、Var33、Var34、Var36、Var39、Var41、Var42、Var47、Var49、Var52、Var54、Var56。根据文献^[19-20]报道,这些差异变量主要为黄酮醇及其衍生物、黄烷-3-醇及其衍生物、氨基酸及其衍生物等葡萄在生长或发酵过程中产生的

代谢产物。基于上述差异变量,重新建立 VIP-PLS-DA 判别模型,其得分图及 200 次置换检验图如图 6 所示。由图 6a) 可知,该模型的 $R^2X_{cum} = 0.717, R^2Y_{cum} = 0.858, Q^2_{cum} = 0.712$,即对 X 矩阵的累积解释方差较原始 PLS-DA 模型有所提高,且模型预测力 Q^2_{cum} 达到 0.712。同时,新模型分类趋势与原 PLS-DA 模型基本一致。由图 6b) 可知,该模型的 $R^2 = (0, 0.259), Q^2 = (0, -0.474)$,表明模型未出现过拟合,可靠有效。此外,采用 8 个预测集样本对 VIP-PLS-DA 模型进行外部验证,结果显示所有样本均可被准确识别(见表 4)。综上所述,所筛选的差异变量可用于红葡萄酒的产地识别。

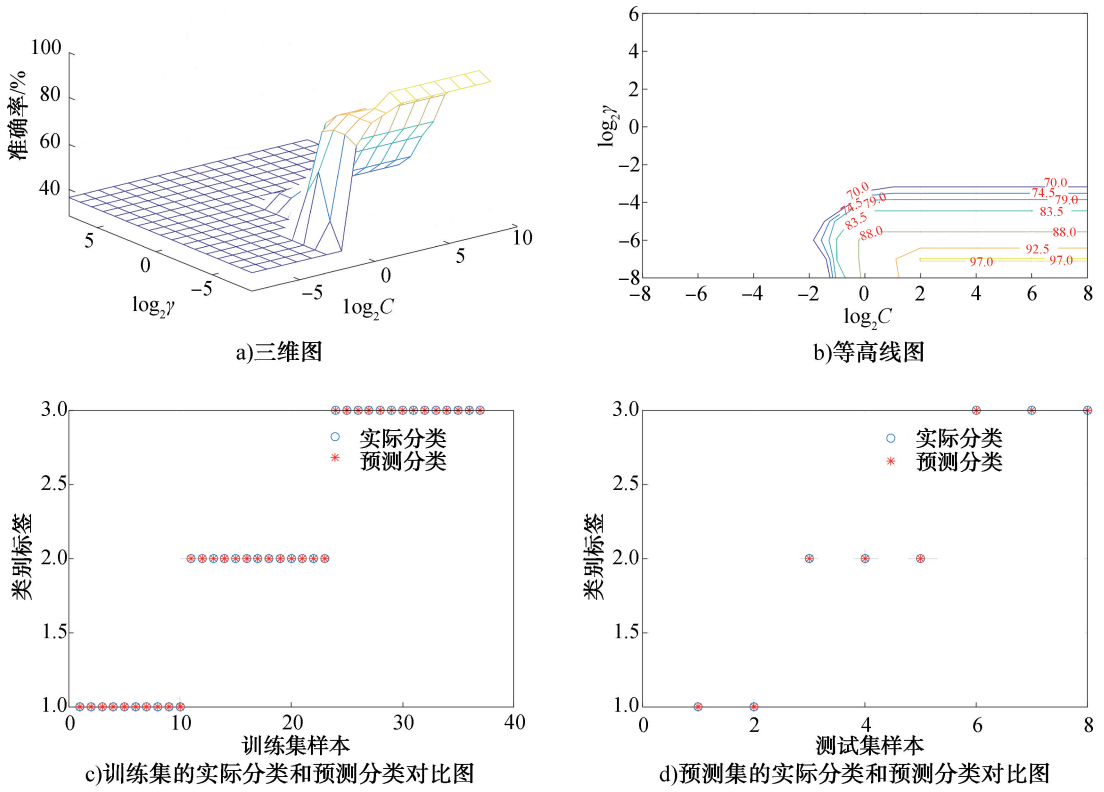


图 5 SVM 参数优化及分类判别结果

Fig. 5 Results of SVM parameter optimization and classification discrimination

表 3 红葡萄酒产地判别的 SVM 模型结果

Table 3 Results of SVM model for identifying the origins of red wines

数据集	产地	正确识别样本数	错误识别样本数	识别率/%	平均识别率/%	灵敏度	特异性
训练集	QHD	10	0	100	100	1	1
	YC	13	0	100		1	1
	TP	14	0	100		1	1
预测集	QHD	2	0	100	100	1	1
	YC	3	0	100		1	1
	TP	3	0	100		1	1

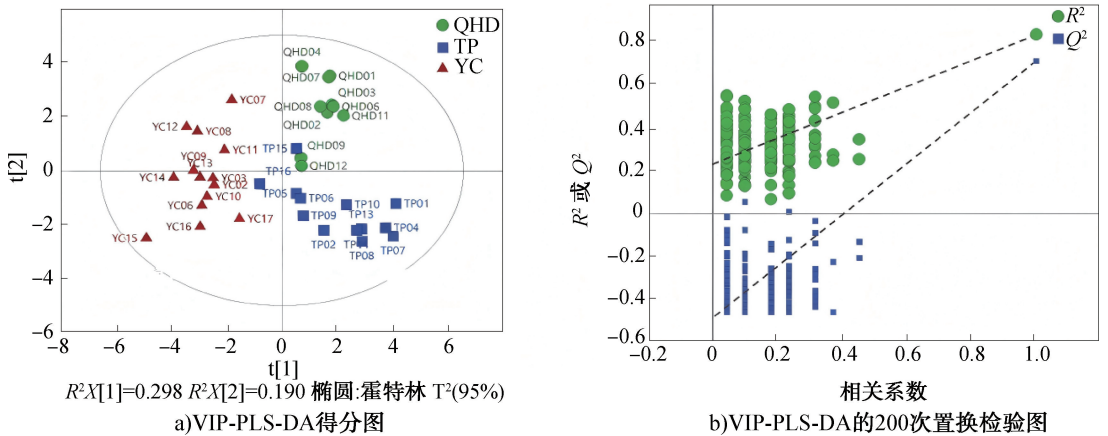


图 6 VIP-PLS-DA 判别模型得分图及 200 次置换检验图

Fig. 6 VIP-PLS-DA discriminant model score plot and 200-times permutation test plot

表 4 红葡萄酒产地判别的 VIP-PLS-DA 模型结果
Table 4 Results of the VIP-PLS-DA model for identifying the origins of red wines

数据集	产地	正确识别样本数	错误识别样本数	识别率/%	平均识别率/%	灵敏度	特异性
训练集	QHD	10	0	100	100	1	1
	YC	13	0	100		1	1
	TP	14	0	100		1	1
预测集	QHD	2	0	100	100	1	1
	YC	3	0	100		1	1
	TP	3	0	100		1	1

3 结论

本研究基于 HPLC-DAD 三维色谱指纹图谱结合多元化学计量学方法,对不同产地红葡萄酒进行了产地识别分析。通过 MCR-ALS 算法共解析出 56 个有效组分,并在此基础上建立了 PCA、PLS-DA 和 SVM 判别模型。PCA 初步显示,3 个产地样本具有一定的区分趋势,但银川与吐鲁番样本存在部分重叠。进一步采用有监督的 PLS-DA 与 SVM 模型,分类效果显著提升,验证了该方法在红葡萄酒产地判别中的可靠性与有效性。此外,以 $VIP > 1$ 且 $P < 0.05$ 为阈值,从 PLS-DA 模型中筛选出 22 个关键差异变量,据此构建的 VIP-PLS-DA 模型实现了对葡萄酒产地 100% 的准确判别。本研究证实了 HPLC-DAD 三维指纹图谱结合多种分析方法在葡萄酒产地溯源中的应用潜力,可为葡萄酒品质控制和市场监管提供支撑,对防止葡萄酒标识造假、规范葡萄酒市场具有积极意义。然而,本研究的样本量有限,难以涵盖每个产地内全部的天然变异,从而影响模型的泛化性能。为建立稳健且广泛适用的产地判别模型,未来研究应纳入更多不同年份及不同生产商的葡萄酒样本,以进一步验证该方法的有效性。

参考文献:

[1] The International Vine and Wine Organisation (OIV). State of the world vine and wine sector in 2024: Adaptation & cooperation [R/OL]. (2025-04-15) [2025-04-30]. <https://www.oiv.int/index.php/press/state-world-vine-and-wine-sector-2024-adaptation-cooperation>.

[2] 谢喜麟. 中国葡萄酒产区分布及气候变化对其影响分析(以东北产区为例) [D]. 杨凌: 西北农林科技大学,

2018.

XIE X L. Study on status of wine regions in China and impact of climate change on it (take north east for example) [D]. Yangling: Northwest A&F University, 2018.

[3] WHITE R E. The value of soil knowledge in understanding wine terroir [J]. *Frontiers in Environmental Science*, 2020, 8: 12.

[4] 冯淑敏. 高效液相色谱在食品检测中的应用探究 [J]. *食品安全导刊*, 2023 (6): 163-165.

FENG S M. Research on the practice of high performance liquid chromatography in food detection [J]. *China Food Safety Magazine*, 2023 (6): 163-165.

[5] PENG T Q, YIN X L, GU H W, et al. HPLC-DAD fingerprints combined with chemometric techniques for the authentication of plucking seasons of Laoshan green tea [J]. *Food Chemistry*, 2021, 347: 128959.

[6] GU H W, YIN X L, PENG T Q, et al. Geographical origin identification and chemical markers screening of Chinese green tea using two-dimensional fingerprints technique coupled with multivariate chemometric methods [J]. *Food Control*, 2022, 135: 108795.

[7] 彭天琴. 基于三维色谱指纹图谱的绿茶质量等级评定新方法研究 [D]. 荆州: 长江大学, 2021.

PENG T Q. Development of a new method for evaluating quality grade of tea based on three-way chromatographic fingerprints [D]. Jingzhou: Yangtze University, 2021.

[8] DONG Z Y, XU J L. Baseline estimation using optimized asymmetric least squares (O-ALS) [J]. *Measurement*, 2024, 233: 114731.

[9] JAUMOT J, DEJUAN A N, TAULER R. MCR-ALS GUI 2.0: New features and applications [J]. *Chemometrics and Intelligent Laboratory Systems*, 2015, 140: 1-12.

[10] ANZARDI M B, ARANCIBIA J A, OLIVIERI A C. Processing multi-way chromatographic data for analytical calibration, classification and discrimination: A successful marriage between separation science and chemometrics [J]. *TrAC Trends in Analytical Chemistry*, 2021, 134: 116128.

[11] CHANG C C, LIN C J. LIBSVM: A library for support vector machines [J]. *ACM transactions on intelligent systems and technology (TIST)*, 2011, 2(3): 1-27.

[12] BOS T S, KNOL W C, MOLENAAR S R A, et al. Recent

- applications of chemometrics in one-and two-dimensional chromatography[J]. *Journal of Separation Science*, 2020, 43(9/10): 1678–1727.
- [13] NIEZEN L E, SCHOENMAKERS P J, PIROK B W J. Critical comparison of background correction algorithms used in chromatography [J]. *Analytica Chimica Acta*, 2022, 1201: 339605.
- [14] ZHANG X, ZHANG Z Y, TAULER R. Evaluation of the extension of rotation ambiguity associated to multivariate curve resolution solutions by the application of the MCR-BANDS method[J]. *Talanta*, 2019, 202: 554–564.
- [15] WANG Y, GU H W, YIN X L, et al. Deep learning in food safety and authenticity detection: An integrative review and future prospects [J]. *Trends in Food Science & Technology*, 2024, 146: 104396.
- [16] JIMÉNEZ-CARVELO A M, MARTÍN-TORRES S, ORTEGA-GAVILÁN F, et al. PLS-DA vs sparse PLS-DA in food traceability. A case study: Authentication of avocado samples[J]. *Talanta*, 2021, 224: 121904.
- [17] LI S C, YU X W, ZHEN Z P, et al. Geographical origin traceability and identification of refined sugar using UPLC-QToF-MS analysis [J]. *Food Chemistry*, 2021, 348: 128701.
- [18] GU H W, ZHOU H H, LV (LYU) Y, et al. Geographical origin identification of Chinese red wines using ultraviolet-visible spectroscopy coupled with machine learning techniques [J]. *Journal of Food Composition and Analysis*, 2023, 119: 105265.
- [19] PAN Y, GU H W, LV (LYU) Y, et al. Untargeted metabolomic analysis of Chinese red wines for geographical origin traceability by UPLC-QTOF-MS coupled with chemometrics[J]. *Food Chemistry*, 2022, 394: 133473.
- [20] YIN X L, PENG Z X, PAN Y, et al. UHPLC-QTOF-MS-based untargeted metabolomic authentication of Chinese red wines according to their grape varieties [J]. *Food Research International*, 2024, 178: 113923.

Geographical origin identification of Chinese Cabernet Sauvignon red wines based on HPLC-DAD three-dimensional chromatographic fingerprinting

ZOU Hua, PAN Yuan, QIANG Haiqing, MENG Zhan, YIN Xiaoli, GU Huiwen

College of Chemistry and Environmental Engineering/College of Life Sciences, Yangtze University, Jingzhou 434023, China

Abstract: **[Objective]** This study aimed to advance the authenticity identification and origin traceability techniques of wines. **[Methods]** The three-dimensional chromatographic fingerprints of 45 Cabernet Sauvignon red wines from three famous producing regions in China (Qinhuangdao, Yinchuan, and Turpan) were acquired using high-performance liquid chromatography with diode array detection (HPLC-DAD). The data were then resolved by multivariate curve resolution-alternating least squares (MCR-ALS). Based on the relative concentrations of the resolved components, three machine learning algorithms—principal component analysis (PCA), partial least squares-discriminant analysis (PLS-DA), and support vector machine (SVM)—were applied to discriminate the geographical origins of the red wines. **[Results]** A total of 56 resolved components were obtained by MCR-ALS analysis of the three-dimensional chromatographic fingerprints. PCA score plots showed a tendency of the wines to cluster according to their geographical origins. Both PLS-DA and SVM models achieved good classification performance, with 100% recognition accuracy for both the training and prediction sets. Furthermore, a VIP-PLS-DA model based on 22 differential variables enabled accurate discrimination of the geographical origins of the wines with 100%. **[Conclusion]** The HPLC-DAD three-dimensional chromatographic fingerprinting technique combined with machine learning algorithms can establish a stable and reliable recognition model, and is expected to provide objective and accurate identification of three geographical origins of Chinese Cabernet Sauvignon red wines.

Key words: Cabernet Sauvignon red wine; HPLC-DAD; three-dimensional chromatographic fingerprinting; origin identification; machine learning

[责任编辑: 王晓波 李波]