

DOI: 10.13652/j.spjx.1003.5788.2025.80174

基于近红外光谱法的馥郁香型白酒基酒中 4 种主要有机酸检测模型构建

张云霞¹ 余 佶^{1,2} 李运通^{1,2} 余 冰³ 靳喜庆³ 姚茂君^{1,2}

(1. 吉首大学化学化工学院, 湖南 吉首 416000; 2. 湘西自治州馥郁香型白酒酿造与品质控制重点实验室, 湖南 吉首 416000; 3. 酒鬼酒股份有限公司, 湖南 吉首 416000)

摘要: [目的] 利用近红外光谱技术与化学计量学方法对 47 个馥郁香型白酒基酒中的 4 种主要有机酸进行快速定量分析。[方法] 采用 Kennard-Stone (K-S) 算法划分样本, 结合归一化 (Normalization)、标准正态变量变换 (SNV) 和 Savitzky-Golay 平滑等预处理策略, 通过交叉验证优化模型参数, 并评估主成分数和波段选择对主成分回归 (PCR) 和偏最小二乘法回归 (PLSR) 模型性能的影响。[结果] PCR 模型对 4 种主要有机酸 (乙酸、正丁酸、正戊酸、正己酸) 的预测相关系数均高于 0.9, 预测均方根误差均小于 7; PLSR 模型的预测相关系数均高于 0.8, 预测均方根误差均小于 8, 表明模型具有良好的泛化能力和预测精度。[结论] 近红外光谱技术结合 PCR 和 PLSR 建模方法可有效实现馥郁香型白酒基酒中 4 种有机酸的快速定量分析。

关键词: 近红外光谱; 主成分回归分析; 偏最小二乘法回归分析; 有机酸

Detection modelling of four major organic acids in the base wine of Fuyuxiangxing crude Baijiu based on near-infrared spectroscopy

ZHANG Yunxia¹ YU Ji^{1,2} LI Yuntong^{1,2} YU Bing³ JIN Xiqing³ YAO Maojun^{1,2}(1. College of Chemistry and Chemical Engineering, Jishou University, Jishou, Hunan 416000, China;
2. Xiangxi Autonomous Prefecture Key Laboratory of Fuyuxiangxing Baijiu Brewing and Quality Control, Jishou University, Jishou, Hunan 416000, China; 3. Jiugui Liquor Co., Ltd., Jishou, Hunan 416000, China)

Abstract: [Objective] To rapidly quantify four major organic acids in 47 base wines of Fuyuxiangxing crude Baijiu using near-infrared spectroscopy (NIR) and chemometrics. [Methods] The study used the Kennard-Stone (K-S) algorithm to divide the samples, combined with preprocessing strategies such as normalization, standard normal variate transformation (SNV), and Savitzky-Golay smoothing. Model parameters were optimized through cross-validation, and the effects of the number of principal components and band selection on the performance of the principal component regression (PCR) and partial least squares regression (PLSR) models were evaluated. [Results] The validation results showed that the prediction correlation coefficients (R_c^2) of the PCR models for the four major organic acids (acetic acid, n-butyric acid, n-pentanoic acid, and n-hexanoic acid) were all higher than 0.9, and the root mean square errors of prediction (RMSEPs) were all lower than 7. The R_c^2 values of the PLSR models were all higher than 0.8, and the RMSEPs were all lower than 8, indicating that the models had good generalization ability and prediction accuracy. [Conclusion] NIR combined with PCR and PLSR modeling methods can effectively achieve the rapid quantitative analysis of the four organic acids in Fuyuxiangxing crude Baijiu base wine.

Keywords: near infrared spectroscopy; principal component regression analysis; partial least squares regression analysis; organic acids

基金项目: 湖南省重点研发计划项目 (编号: 2024JK2151-3); 湘西州技术攻关“揭榜挂帅”项目 (编号: 2022JBGS0002); 吉首大学研究生校级科研项目 (编号: Jdy23006)

通信作者: 姚茂君 (1968—), 男, 吉首大学教授, 硕士。E-mail: yaomaojun@126.com

收稿日期: 2025-01-10 **改回日期:** 2025-03-20

引用格式: 张云霞, 余佶, 李运通, 等. 基于近红外光谱法的馥郁香型白酒基酒中 4 种主要有机酸检测模型构建[J]. 食品与机械, 2025, 41(4): 72-80.

Citation: ZHANG Yunxia, YU Ji, LI Yuntong, et al. Detection modelling of four major organic acids in the base wine of Fuyuxiangxing crude Baijiu based on near-infrared spectroscopy[J]. Food & Machinery, 2025, 41(4): 72-80.

馥郁香型白酒因其“前浓、中清、后酱”的口味特征而备受关注^[1-3],不同年份的馥郁香型白酒由于陈酿、贮存等因素,有机酸的含量存在显著差异,这些有机酸在酒体的风味、口感和稳定性中扮演着关键角色,且近年来成为业内研究的重点^[4-5]。有机酸是馥郁香型白酒中含量最多的香味成分之一,种类丰富,其中乙酸、正丁酸、正戊酸和正己酸为主要成分,这四大有机酸是影响馥郁香型基酒品质的关键因素,对白酒香气形成起着重要作用。传统的有机酸检测方法,如感官评价^[6]及气相色谱^[7],已被广泛应用于白酒风味分析。尽管这些方法能够提供较为准确的分析结果,但由于其操作复杂、时间消耗长且成本较高,难以满足生产过程中对快速检测的需求。

近年来,近红外光谱(NIR)技术因其高效、无损、快速的特点,不仅能够高效地完成多组分同时分析,还被广泛应用于食品与饮料领域,尤其是在品质控制和成分分析中得到了显著的关注。此外,红外光谱技术操作简便、成本低廉,且对环境友好,是一种高效的分析手段^[8-11]。张卫卫等^[12]结合近红外光谱及预处理方法研究了白酒基酒中乙醛和乙缩醛的快速检测方法,结果表明模型精度高、稳定性好,可满足生产检测需求。刘建学等^[13]用近红外光谱快速检测了浓香型白酒基酒中的己酸和乙酸,建立的模型精密度高、稳定性好,满足生产需求。Zhang等^[14]成功区分了7种不同产地的酱香型白酒,主要通过分析乳酸、乙酸、丙酸、丁酸和己酸5种有机酸的含量。尽管已有大量研究利用近红外光谱技术对白酒的品质、掺假情况及成分进行分析,但针对馥郁香型基酒中有机酸的快速检测研究仍较为有限,相关模型的建立尚处于初步阶段。

研究拟对47个不同年份的馥郁香型白酒基酒进行红外光谱分析,以气相色谱值为参考值,将乙酸、正丁酸、正戊酸、正己酸4种主要有机酸的气相数据与红外光谱数据相结合,进行定量建模,以期对馥郁香型白酒基酒测定提供一套高效、准确、快速地分析有机酸的方法。

1 材料与方法

1.1 材料与试剂

47个不同年份馥郁香型基酒样品(见表1):酒精度为52.4%~72.4%vol,酒鬼酒股份有限公司;

丙酮:色谱纯,国药集团化学试剂有限公司;

乙醇:色谱纯,上海麦克林生化科技股份有限公司;

白酒混标(甲酸乙酯、正丙醇、乙酸乙酯、仲丁醇、乙酸、异丁醇、正丁醇、2-戊酮、丙酸、异戊醇、2-甲基-1-丁醇、正戊醇、丁酸乙酯、乙酸丁酯、正丁酸、2,3-丁二醇、乙酸异戊酯、正己醇、戊酸乙酯、正戊酸、己酸乙酯、己酸、庚酸乙酯、辛酸己酯、壬酸乙酯、癸酸乙酯):色谱纯,上海安谱瑞世标准技术服务有限公司。

1.2 仪器与设备

傅里叶变换近红外光谱仪:NICOLET ISIOFT-IR型,美国赛默飞世尔公司;

表1 47个不同年份馥郁香型基酒样品

Table 1 Forty-seven samples of Fuyuxiangxing base wines from different production years

样品	生产时间	样品	生产时间
样品 1~3	2002年1月	样品 25~27	2016年4月
样品 4~6	2004年12月	样品 28~31	2018年12月
样品 7~9	2007年4月	样品 32~35	2020年11月
样品 10~12	2009年6月	样品 36~39	2022年8月
样品 13~16	2010年7月	样品 40~43	2023年6月
样品 17~20	2012年12月	样品 44~47	2023年12月
样品 21~24	2014年12月		

气相色谱仪:7890B型,美国安捷伦公司。

1.3 试验方法

1.3.1 标准溶液的配制

(1) 乙醇溶液:室温条件下,量取70 mL乙醇,用水定容至100 mL,混匀,使用酒精计校准。

(2) 26种挥发性风味物质标准品溶液:参考冯思雨等^[15]的方法并稍作改变,分别准确吸取0.2,0.4,0.6,0.8,1.0,2.0 mL标准品溶液于10 mL容量瓶中,用体积分数70%的乙醇溶液定容。

1.3.2 色谱条件 色谱柱:CP-Wax 57 CB, 50 m×0.250 mm×0.2 μm。升温程序参考Xu等^[16]的方法并稍作修改:40℃保持2 min;以4℃/min升至80℃,保持1 min;以3℃/min升至100℃;以12℃/min升至124℃,保持3 min;以14℃/min升至200℃,保持8 min。载气为氮气(纯度>99.999%),流量1.0 mL/min。分流进样,进样量1 μL,分流比20:1。

1.3.3 中红外光谱扫描条件 利用傅里叶变换近红外光谱仪采集白酒样品光谱,测量前将仪器预热30 min,在不放载玻片的情况下,以空气作参比,采用液膜法检测白酒的吸收峰。光谱范围为4 000~1 000 cm⁻¹,分辨率为8 cm⁻¹,扫描次数32次;利用配套软件OMNIC采集白酒样品的红外光谱^[17]。每个样品采集3次,取3次采集的光谱平均值作为样本的原始光谱,且在测量过程中,温度、湿度等环境条件尽量保持一致。

1.4 数据处理

数据采用Microsoft Office Excel 2019进行处理整理,为消除不同陈酿时间基酒酒精度对检测结果的影响,按式(1)对酒精度进行标准化处理,将其统一为70%vol。

$$A = \frac{D}{C} \times 70, \quad (1)$$

式中:

A——各项指标分析值;

D——各项检测值;

C——相应酒样陈酿时间酒精度,%vol;

70——统一酒精度,%vol。

利用The Unscrambler X对数据进行回归分析处理,

使用 Origin 2021 对得到的数据进行作图。

1.5 化学计量学建模

1.5.1 数据预处理 红外光谱数据预处理是为提升分析模型准确性和稳定性,利用数学或统计方法对原始数据进行优化处理的过程。其主要目标在于减少噪声、基线漂移及散射效应,强化与分析物相关的光谱特征信息^[18]。选择归一化(Normalization)^[19]、标准正态变量变换(SNV)^[20]和 Savitzky-Golay 平滑^[21]作为红外光谱数据的预处理方法。具体来说,归一化方法消除了样品间因强度差异导致的偏差,从而增强数据的可比性;SNV 通过调整光谱数据以减弱散射效应,改善基线稳定性;而 SG 平滑则在减少噪声的同时保留了光谱中的主要特征峰。预处理方法的单独应用有助于提取与目标变量相关的有效信息,从而分别提升后续模型的预测性能和泛化能力。

1.5.2 波数的选择 原始光谱数据包含大量噪声和冗余信息,可能影响模型的准确性与稳定性。针对中红外光谱的共线性问题,基于化学领域知识筛选与目标变量高度相关的特征波数,简化模型,减少计算量。通过交叉验证优化波数区间,显著提升了模型的预测准确性和泛化能力^[22]。

1.5.3 校正集和预测集的选择 为了构建回归模型,采用 Kennard-Stone(K-S)算法^[23]对 47 个酒样进行了划分。K-S 算法通过在高维空间中均匀采样,优先选择具有代表性的样本,以确保校正集和预测集的均匀分布和代表性。经过 K-S 算法的处理,将 47 个酒样划分为校正集 34 个和预测集 13 个。这种划分方法不仅确保了校正集的多样性,还能够有效评估模型在未知样本上的预测能力,为模型的泛化性能提供了保障。

1.5.4 建模方法 主成分回归^[24]是一种基于主成分分析(PCA)的回归方法,它通过线性变换将原始预测变量转化为一组新的不相关变量(主成分),进而利用这些主成分建立回归模型。PCR 的核心目标是通过降维来缓解变量过多或多重共线性带来的问题,从而提高模型的稳定性和解释能力。偏最小二乘回归(PLSR)^[25]则是一种优化回归分析的方法,它不仅通过主成分分析对自变量进行降维,还同时考虑自变量与因变量的相关性,以寻找能够最大化自变量与因变量共变的投影方向。与 PCR 不同,PLSR 的降维过程旨在增强预测性能,尤其适用于自变量间存在高相关性且样本量较少的情境。鉴于研究数据集的特性,PCR 和 PLSR 均为适宜的分析方法,能够有效应对多重共线性问题,并在提高预测精度的同时增强模型的鲁棒性。

1.5.5 模型评价方法 采用主成分回归(PCR)和偏最小二乘回归(PLSR)方法建立不同年份的馥郁香型基酒 4 种有机酸的预测模型。按文献[26—27]的方法,采用校正集样品决定系数(R^2)和交叉验证均方根误差(RMSEC)评价所建立定量校正模型的优劣,预测集样品的决定系数(R_p^2)和预测均方根误差(RMSEP)以及模型精密程度结

果评价其预测能力。

2 结果与分析

2.1 气相色谱分析

以各有机酸的质量浓度为 X ,有机酸峰面积为 Y ,进行回归计算。回归方程和相关系数见表 2,结果表明,4 种有机酸的质量浓度和峰面积具有良好的线性关系,相关系数均大于 0.999 9。

表 2 有机酸的保留时间、回归方程和相关系数

Table 2 Retention time, regression equations, and correlation coefficients for organic acids

标品名称	保留时间/min	标准曲线	相关系数
乙酸	24.210	$Y=0.247\ 41X-1.019\ 00$	0.999 9
正丁酸	28.345	$Y=0.500\ 78X-0.020\ 68$	0.999 9
正戊酸	30.023	$Y=0.560\ 75X-0.925\ 47$	0.999 9
正己酸	31.420	$Y=0.652\ 42X-1.576\ 38$	0.999 9

2.2 近红外光谱分析

47 个不同年份的馥郁香型白酒样品的近红外光谱图如图 1 所示。不同年份的白酒在近红外区域表现出显著的吸收特征,各光谱的变化趋势相似。尽管各年份的样品在相同的光谱区间内呈现出高度不一的吸收峰,但这些峰的位置基本一致。在 $3\ 395, 1\ 647\ \text{cm}^{-1}$ 处的吸收峰主要归因于白酒中乙醇和水分子中的 O—H 键的吸收作用^[28], $2\ 977\sim 2\ 900\ \text{cm}^{-1}$ 处的峰与 C—H 的拉伸振动峰有关, $1\ 500\sim 900\ \text{cm}^{-1}$ 处的红外吸收峰是由酒中的挥发性物质,有机酸和糖类引起的。 $1\ 500\sim 1\ 200\ \text{cm}^{-1}$ 处的谱带与—CH₂、CC—H 和 HC—O 的变形有关。 $1\ 200\sim 900\ \text{cm}^{-1}$ 处的吸收峰可能是由 C—C 和 C—O 引起的。 $1\ 090, 1\ 047\ \text{cm}^{-1}$ 处的吸收峰应与醇类物质 C—O 结构的延伸吸收峰有关。此外,在 $1\ 300\sim 1\ 000\ \text{cm}^{-1}$ 处的吸收峰可能是由醇和酚的 C—O 拉伸振动引起的^[29]。

2.3 校正集、预测集划分结果

校正集、预测集样品数量及乙酸、正丁酸、正戊酸、正己酸含量统计结果如表 3 所示。经 K-S 选择后,得到校正集样品 34 个、预测集样品 13 个,两项指标校正集样品范围均包含预测集样品范围,充分体现了校正集样品的代表性。

2.4 预处理

2.4.1 预处理方法选择 在红外光谱数据处理中,预处理方法对建立回归模型具有重要作用。红外光谱数据常包含基线漂移和噪声,这些因素会影响模型的精度和稳定性。通过适当的预处理,可以有效去除噪声,增强谱峰特征,从而提高模型的预测能力^[30]。研究^[31]表明,不同的预处理方法对模型效果存在显著影响,因此,选择合适的预处理方法是建立高质量模型的关键步骤。在波数为 $4\ 000\sim 1\ 000\ \text{cm}^{-1}$,主成分数为 7 的条件下,采用 4 种不同

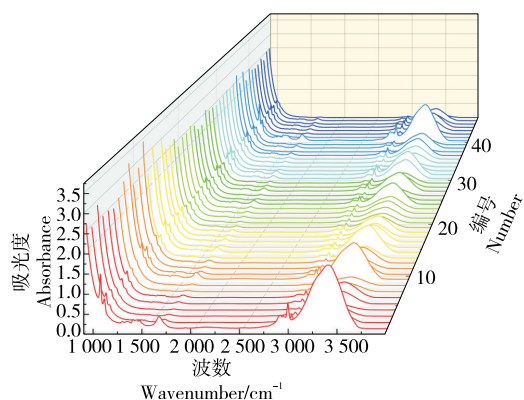


图1 47个不同年份馥郁香型白酒基酒近红外原始光谱瀑布流图

Figure 1 Near-infrared raw spectral waterfalls of samples of Fuyuxiangxing base wines from different production years

的光谱预处理方法(无预处理、S-G平滑、SNV和Normalize)结合PCR和PLSR建模,以分析乙酸、正丁酸、正戊酸和正己酸的浓度。基于原始谱图与不同预处理方式建立的PCR和PLSR模型如图2和图3所示,预测值(predicted)表示模型或算法输出的结果,参考值(reference)表示实际测量值或试验数据,作为验证预测准确性的基准。结果表明,合适的光谱预处理可有效提升定量模型预测性能,乙酸与正戊酸经SNV预处理、正丁酸和正己酸经S-G平滑预处理后的 R^2 最高, R^2 均达到0.9以上,RMSEC最低。相比之下,Normalize预处理下4种有机酸中的效果较差,未能有效改善模型的预测精度。因此,将SNV预处理作为乙酸和正戊酸的最优处理方法,将S-G平滑预处理作为正丁酸和正己酸的最优处理方法。

2.4.2 波段选择 在近红外光谱回归模型建立中,波段

选择是提高模型性能的重要步骤。近红外光谱通常包含大量的无关或噪声波段,直接使用所有波段会导致模型复杂性增加,且可能降低预测准确性。通过选择与目标变量相关性较高的波段,可以保留关键信息,排除干扰,简化模型结构,并提升模型的预测能力和计算效率^[32]。研究过程中,通过波段选择显著提高了模型的 R^2 值,并降低了预测误差,验证了波段选择在模型构建中的重要性。按上述预处理选择结果,乙酸、正丁酸、正戊酸和正己酸分别采用SNV、S-G、SNV、S-G方法进行预处理,主成分数为7进行波段选择,不同波段数下基酒中4种有机酸的PCR定量分析结果如图4所示。结果显示,乙酸、正丁酸、正戊酸和正己酸模型在1200~1000 cm^{-1} 与1700~1300 cm^{-1} 波段数组合下,预测值与参考值的 R^2 均有所提高,同时RMSEC显著降低,表明模型的预测性能得到了显著提升。不同波段数下基酒中4种有机酸的PLSR定量分析结果如图5所示。乙酸和正戊酸在1200~1000 cm^{-1} 与1700~1300 cm^{-1} 波段数组合下分别达到最佳性能,其 R^2 分别为0.9592和0.8745, RMSEC分别为14.6135和2.4637,均为最优值。正丁酸的最佳波段数组合为1700~1300 cm^{-1} 与3000~2900 cm^{-1} 对应的 R^2 为0.9373, RMSEC为4.4512。正己酸的最佳波段数为1700~1300 cm^{-1} 单一波段,其 R^2 为0.9213, RMSEC为21.1719。结果表明,波段选择对基酒中4种有机酸的定量分析具有显著影响。

综上,PCR模型中4种有机酸的最佳波段数为1200~1000 cm^{-1} 与1700~1300 cm^{-1} 组合,PLSR模型中乙酸和正戊酸的最佳波段数为1200~1000 cm^{-1} 与1700~1300 cm^{-1} 组合,正丁酸、正己酸的最佳波段数为1700~1300 cm^{-1} 与3000~2900 cm^{-1} 组合和1700~1300 cm^{-1} 。

表3 4种有机酸的校正集与预测集统计结果

Table 3 Statistical results of correction set and validation set for four organic acids

数据集	样品数	指标	平均值/($\text{mg}\cdot\text{L}^{-1}$)	最大值/($\text{mg}\cdot\text{L}^{-1}$)	最小值/($\text{mg}\cdot\text{L}^{-1}$)	标准偏差/($\text{mg}\cdot\text{L}^{-1}$)
校正集	34	乙酸	841.01	1065.54	757.64	72.36
		正丁酸	205.54	268.61	178.78	17.78
		正戊酸	64.81	86.20	55.75	6.95
		正己酸	1181.70	1463.93	1099.85	75.45
预测集	13	乙酸	817.09	856.72	780.77	23.67
		正丁酸	198.10	203.67	194.35	2.86
		正戊酸	63.56	66.97	61.24	1.68
		正己酸	1169.85	1212.98	1133.24	22.18

2.4.3 主成分数选择 在红外光谱回归模型的建立过程中,主成分数的选择对于模型的性能至关重要。过多的主成分可能导致模型复杂度过高,引入噪声,造成过拟合;而过少的主成分则可能无法充分捕捉数据中的关键

信息,导致欠拟合^[33]。因此,合理选择主成分数是确保信息提取与噪声抑制之间良好平衡的关键,以优化模型的预测能力与泛化能力。

不同主成分数下基酒中4种有机酸的PCR定量模型

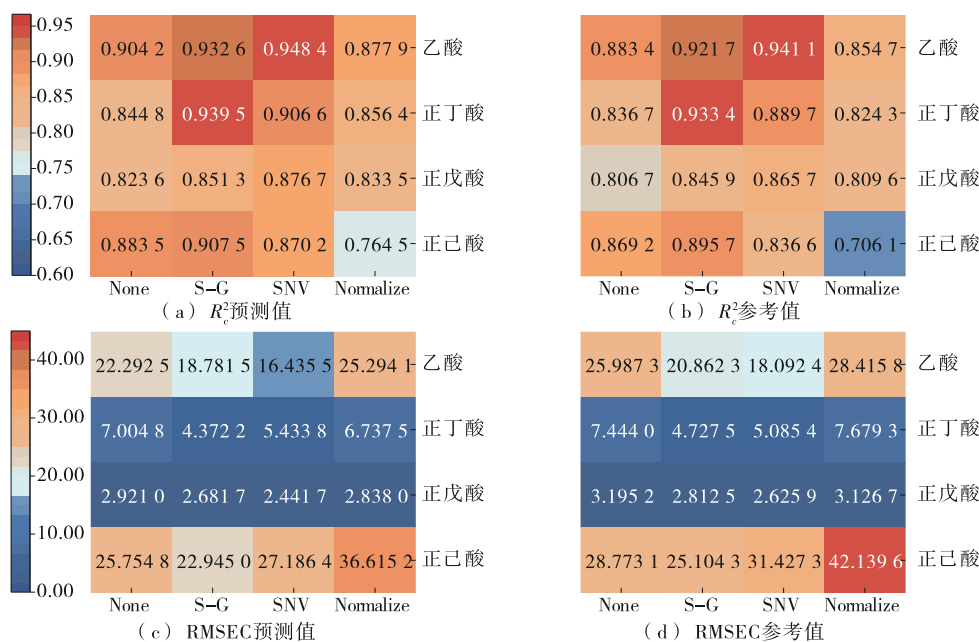


图2 主成分回归预处理方法选择

Figure 2 Selection of preprocessing methods for principal component regression analysis

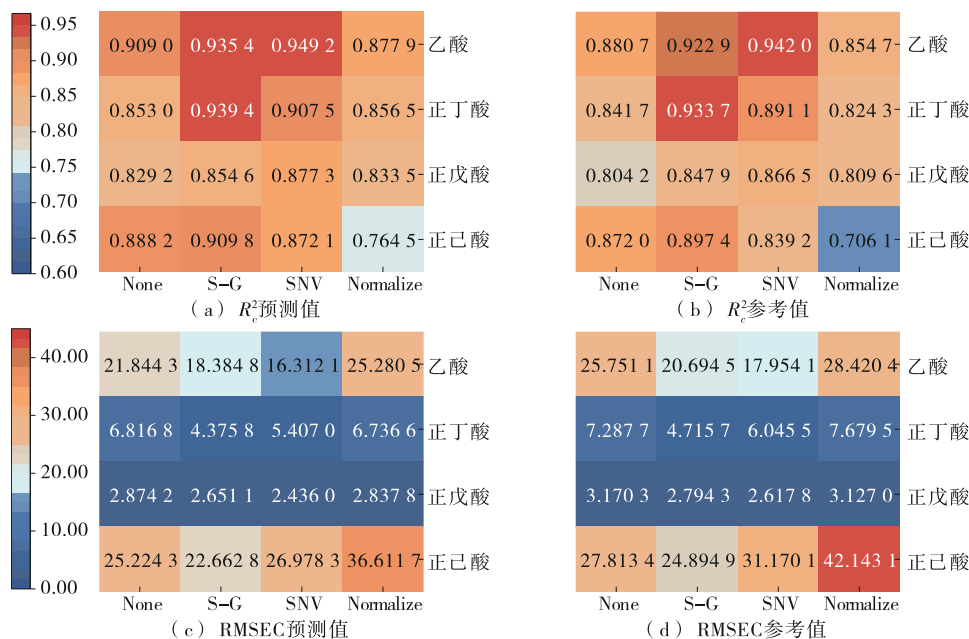


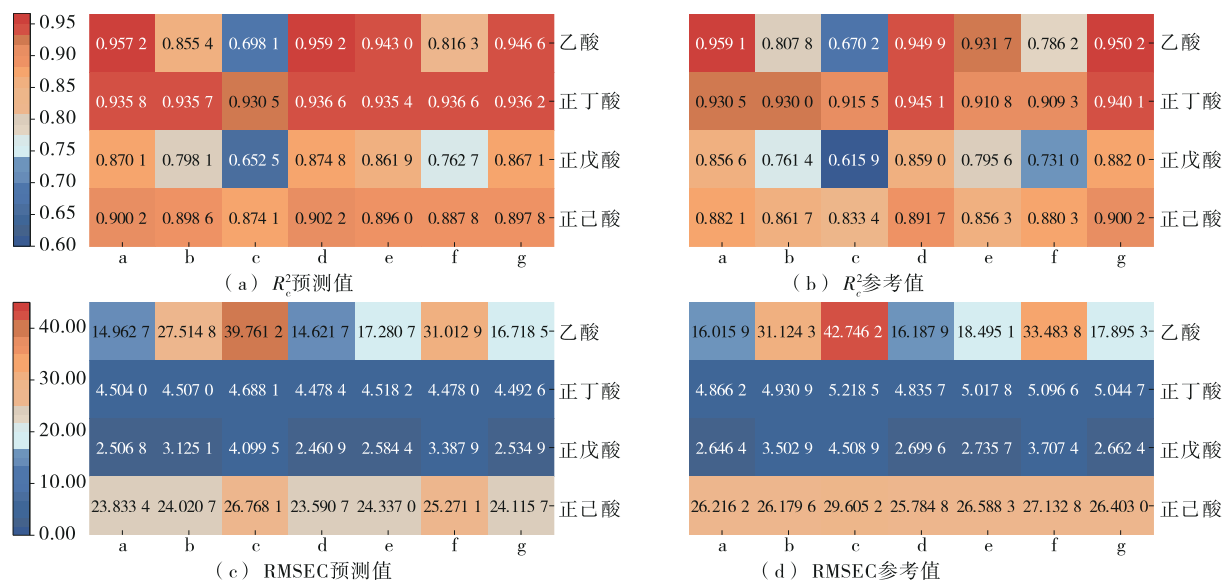
图3 偏最小二乘法回归预处理方法选择

Figure 3 Selection of preprocessing methods for partial least squares regression

结果如图6所示。乙酸、正丁酸和正戊酸在主成分数为6时表现出最佳预测性能,此时模型的RMSEC达到最低, R^2_c 均超过0.87;正己酸则主成分数为7时达到较优平衡, R^2_c 为0.902 2, RMSEC为23.590 7,展现出稳定的预测性能。不同主成分数下基酒中4种有机酸的PLSR定量模型结果如图7所示。4种有机酸在采用4个主成分时均实

现了 R^2_c 与RMSEC的最优平衡, R^2_c 均超过0.85。这表明4个主成分足以捕捉关键特征,同时有效避免了预测误差的扩大。

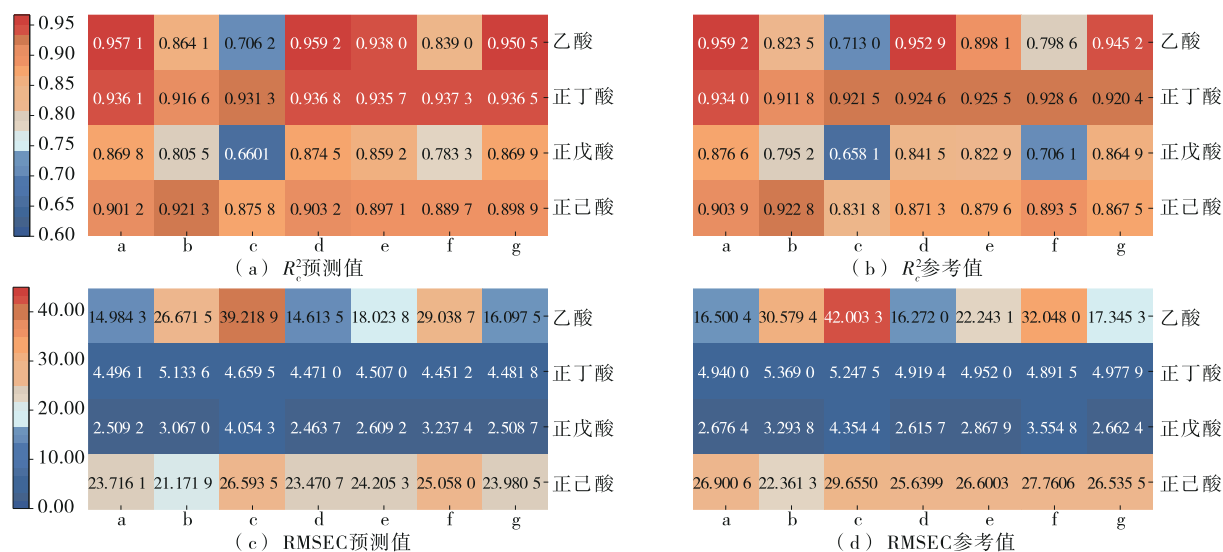
综上,乙酸、正丁酸和正戊酸的PCR模型选择主成分数为6,正己酸的PCR模型选择主成分数为7;4种有机酸的PLSR模型均选择主成分数为4。



a. 1 200~1 000 cm^{-1} b. 1 700~1 300 cm^{-1} c. 3 000~2 900 cm^{-1} d. 1 200~1 000 cm^{-1} 与 1 700~1 300 cm^{-1} 组合 e. 1 200~1 000 cm^{-1} 与 3 000~2 900 cm^{-1} 组合 f. 1 700~1 300 cm^{-1} 与 3 000~2 900 cm^{-1} 组合 g. 1 200~1 000 cm^{-1} 、1 700~1 300 cm^{-1} 及 3 000~2 900 cm^{-1} 组合

图4 主成分回归波段选择

Figure 4 Principal component regression analysis band selection



a. 1 200~1 000 cm^{-1} b. 1 700~1 300 cm^{-1} c. 3 000~2 900 cm^{-1} d. 1 200~1 000 cm^{-1} 与 1 700~1 300 cm^{-1} 组合 e. 1 200~1 000 cm^{-1} 与 3 000~2 900 cm^{-1} 组合 f. 1 700~1 300 cm^{-1} 与 3 000~2 900 cm^{-1} 组合 g. 1 200~1 000 cm^{-1} 、1 700~1 300 cm^{-1} 及 3 000~2 900 cm^{-1} 组合

图5 偏最小二乘回归波数选择

Figure 5 Partial least squares regression wave number selection

2.5 模型建立与评价

表4给出了两种模型在最优条件下的预测性能。采用PCR(主成分回归)和PLSR(偏最小二乘回归)模型,结合不同的波数选择和预处理方法,分析了乙酸、正丁酸、正戊酸和正己酸浓度预测的性能。PLSR模型在多数情况下较PCR模型提供了更高的预测精度,尤其是在乙酸和正丁酸的浓度预测中,PLSR模型的RMSEP值较低, R^2_p

值较高,说明其在这两种酸类的预测中更为精确。根据K-S法划分出的13个预测集酒样,分别代入4种有机酸的最优PCR和PLSR定量模型进行预测。

结果表明,PCR与PLSR的 R^2_p 均达到0.80以上, RMSEP均低于10,表明模型具有良好的预测能力。具体而言,乙酸、正丁酸和正己酸的PCR模型预测值 R^2_p 分别为0.937 2, 0.966 9, 0.920 3, 参考值 R^2_p 分别为0.915 8,

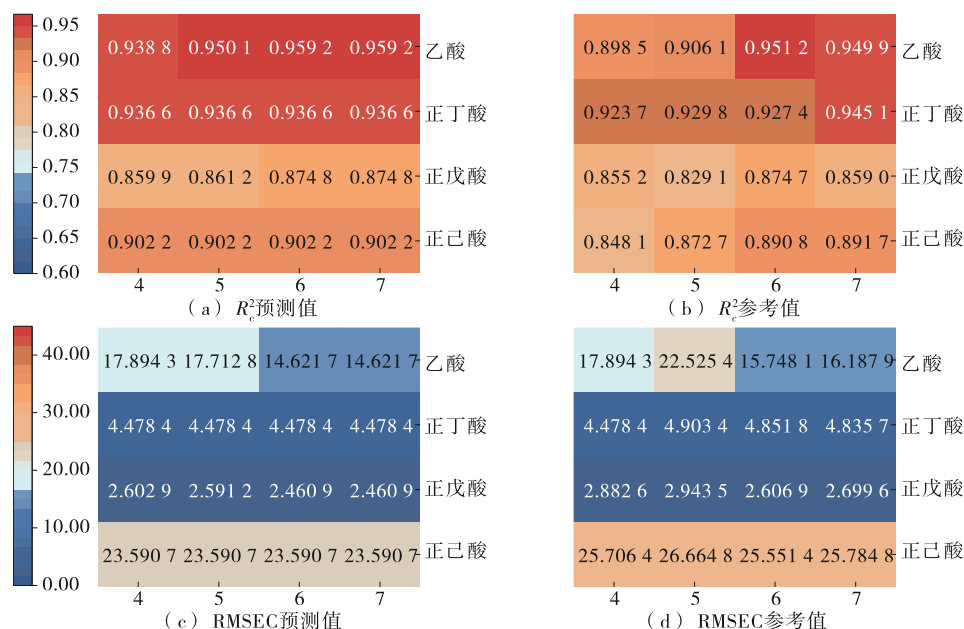


图 6 主成分回归主成分分数选择

Figure 6 Principal component regression principal component score selection

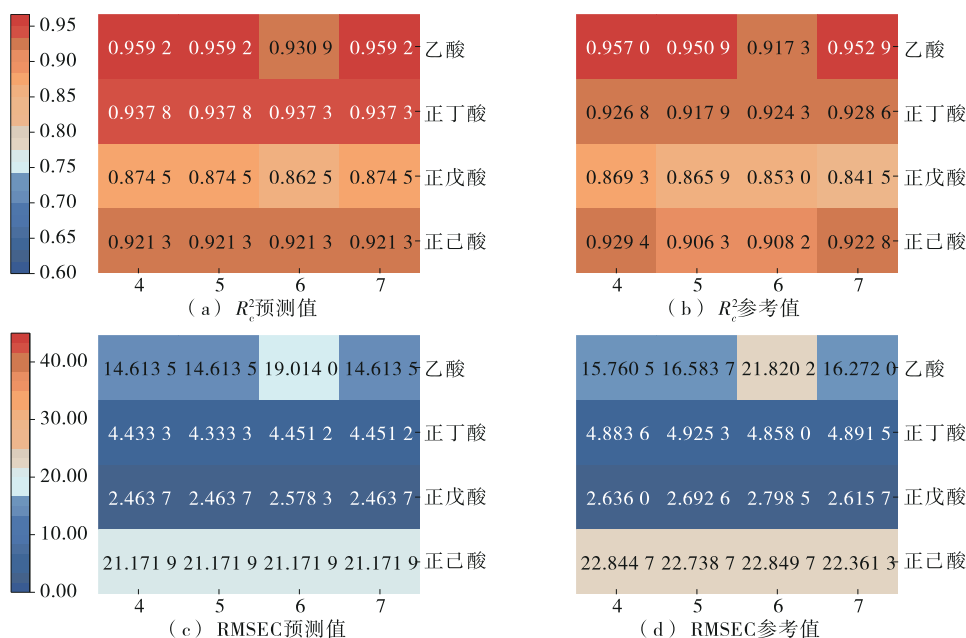


图 7 偏最小二乘法回归主成分分数选择

Figure 7 Partial least squares regression principle component selection

0.960 1, 0.904 8, 表现出较高的预测精度; PLSR 模型对 3 种有机酸的预测效果更优, 预测值 R^2_p 分别为 0.968 0, 0.893 3, 0.863 6, 参考值 R^2_p 分别为 0.954 7, 0.785 4, 0.829 0。相比之下, 正戊酸的预测值和参考值 R^2_p 在两种模型中均较低 (PCR 分别为 0.837 7 和 0.796 5, PLSR 分别为 0.825 5 和 0.775 7), 这可能由于其较低的浓度和色谱波动较大, 使得模型难以准确拟合其真实值。

3 结论

该研究对 47 个不同年份的馥郁香型基酒样品进行了定量建模, 探索了 3 种预处理方法和 2 种建模方法的最佳组合。结果表明, 主成分回归和偏最小二乘回归模型均展现了良好的预测能力, 乙酸、正丁酸和正己酸的预测效果尤为理想, 乙酸、正丁酸和正己酸的主成分回归模型预测决定系数分别为 0.937 2, 0.966 9, 0.920 3, 预测均方根

表 4 两种模型在最优条件下的预测性能
Table 4 Predictive performance of the two models under optimal conditions

指标	模型	预处理方法	波数/cm ⁻¹	主成分数	校正集				预测集			
					R ² _c		RMSEC		R ² _p		RMSEP	
					预测值	参考值	预测值	参考值	预测值	参考值	预测值	参考值
乙酸	PCR	SNV	1 200~1 000 cm ⁻¹ 与 1 700~1 300 cm ⁻¹ 组合	6	0.959 2	0.951 2	14.621 7	15.748 1	0.937 2	0.915 8	5.929 1	7.438 3
	PLSR	SNV	1 200~1 000 cm ⁻¹ 与 1 700~1 300 cm ⁻¹ 组合	4	0.959 2	0.957 0	14.613 5	15.760 5	0.968 0	0.954 7	4.236 6	5.455 9
正丁酸	PCR	S-G	1 200~1 000 cm ⁻¹ 与 1 700~1 300 cm ⁻¹ 组合	6	0.936 6	0.924 7	4.478 4	4.851 8	0.966 9	0.960 1	0.519 8	0.618 3
	PLSR	S-G	1 700~1 300 cm ⁻¹ 与 3 000~2 900 cm ⁻¹ 组合	4	0.936 6	0.927 4	4.433 3	4.883 6	0.893 3	0.785 4	0.932 6	1.433 0
正戊酸	PCR	SNV	1 200~1 000 cm ⁻¹ 与 1 700~1 300 cm ⁻¹ 组合	6	0.874 8	0.874 7	2.460 9	2.606 9	0.837 7	0.796 5	0.678 6	0.823 1
	PLSR	SNV	1 200~1 000 cm ⁻¹ 与 1 700~1 300 cm ⁻¹ 组合	4	0.874 5	0.869 3	2.463 7	2.636 0	0.825 5	0.775 7	0.703 6	0.864 1
正己酸	PCR	S-G	1 200~1 000 cm ⁻¹ 与 1 700~1 300 cm ⁻¹ 组合	7	0.902 2	0.891 7	23.590 7	25.784 8	0.920 3	0.904 8	6.262 4	7.415 3
	PLSR	S-G	1 700~1 300 cm ⁻¹	4	0.921 3	0.929 4	21.171 9	22.844 7	0.863 6	0.829 0	8.193 4	9.936 1

误差分别为 5.929 1,0.519 8,6.262 4。乙酸、正丁酸和正己酸的偏最小二乘法回归模型预测决定系数分别为 0.968 0,0.893 3,0.863 6,预测均方根误差分别为 4.236 6,0.932 6,8.193 4。主成分回归和偏最小二乘法回归模型的预测决定系数值均超过 0.8,预测均方根误差较小。相比之下,正戊酸的预测精度较低,预测决定系数值分别为 0.837 7,0.825 5,预测均方根误差分别为 0.678 6,0.7036,这可能与其较低的浓度、大幅度的色谱波动及与其他成分谱图重叠有关。综上,近红外光谱法结合主成分回归及偏最小二乘法回归适用于馥郁香型基酒有机酸的定量分析,该方法能够快速测定有机酸含量,适用于白酒实际生产过程中的分析需求,具有良好的应用前景。

参考文献

[1] 靳喜庆,殷红,湛松强,等. 馥郁香型白酒三维荧光光谱特征分析研究[J]. 中国酿造, 2023, 42(1): 203-208.
JIN X Q, YIN H, CHEN S Q, et al. Analysis of three dimensional fluorescence spectrum characteristics of Fuyu-flavor Baijiu[J]. China Brewing, 2023, 42(1): 203-208.

[2] CHEN P P, LIU Y, WU J H, et al. Sensory-directed decoding of key aroma compounds from *Jiugui-series* Baijiu, the representative of Fuyu-flavor-type Baijiu (FFTB) [J]. Journal of Food Composition and Analysis, 2022, 114: 104799.

[3] 陆梦思,余佳,余冰,等. 馥郁香型基酒陈酿过程中理化指标及挥发性成分变化分析[J]. 食品与机械, 2024, 40(5): 29-36.
LU M S, YU J, YU B, et al. Analysis of physical and chemical indicators and volatile composition changes during aging of Fuyuxiangxing crude Baijiu[J]. Food & Machinery, 2024, 40(5): 29-36.

[4] ZHANG X J, MENG L J, LU Z M, et al. Identification of age-markers based on profiling of Baijiu volatiles over a two-year maturation period: case study of Lu-flavor Baijiu[J]. LWT-Food Science and Technology, 2021, 141: 110913.

[5] WU Y S, HOU Y X, CHEN H, et al. "Key factor" for Baijiu quality: research progress on acid substances in Baijiu[J].

Foods, 2022, 11(19): 2 959.

[6] WANG L H, ZHU L, ZHENG F P, et al. Determination and comparison of flavor (retronasal) threshold values of 19 flavor compounds in Baijiu[J]. Journal of Food Science, 2021, 86(5): 2 061-2 074.

[7] ZHAO D R, SHI D M, SUN J Y, et al. Characterization of key aroma compounds in Gujingong Chinese Baijiu by gas chromatography-olfactometry, quantitative measurements, and sensory evaluation[J]. Food Research International , 2018, 105: 616-627.

[8] BEĆ K B, HUCK C W. Breakthrough potential in near-infrared spectroscopy: spectra simulation. A review of recent developments[J]. Frontiers in Chemistry, 2019, 7: 48.

[9] PASQUINI C. Near infrared spectroscopy: a mature analytical technique with new perspectives-a review[J]. Analytica Chimica Acta, 2018, 1 026: 8-36.

[10] OZAKI Y. Infrared spectroscopy: mid-infrared, near-infrared, and far-infrared/terahertz spectroscopy[J]. Analytical Sciences, 2021, 37(9): 1 193-1 212.

[11] 周旭,杨倩倩,张进,等. 基于便携式近红外光谱仪的黄桃腐败时间快速预测[J]. 食品与机械, 2024, 40(5): 101-106, 187.
ZHOU X, YANG Q Q, ZHANG J, et al. Rapid prediction of yellow peach spoilage time based on portable near infrared spectrometer[J]. Food & Machinery, 2024, 40(5): 101-106, 187.

[12] 张卫卫,刘建学,韩四海,等. 白酒基酒中醛类物质的傅里叶变换近红外光谱检测[J]. 食品科学, 2016, 37(6): 111-115.
ZHANG W W, LIU J X, HAN S H, et al. Determination of aldehydes in liquor base based on fourier transform near-infrared spectroscopy[J]. Food Science, 2016, 37(6): 111-115.

[13] 刘建学,张卫卫,韩四海,等. 白酒基酒中己酸、乙酸的近红外快速检测[J]. 食品科学, 2016, 37(4): 181-185.
LIU J X, ZHANG W W, HAN S H, et al. Rapid detection of caproic acid and acetic acid in liquor base based on fourier transform near-infrared spectroscopy[J]. Food Science, 2016, 37(4): 181-185.

- [14] ZHANG J, TIAN Z Q, MA Y Q, et al. Origin identification of the sauce-flavor Chinese Baijiu by organic acids, trace elements, and the stable carbon isotope ratio[J]. *Journal of Food Quality*, 2019, 2 019: 1-7.
- [15] 冯思雨, 田红云, 曹雯雯, 等. 白酒中甲醇、乙酸乙酯、乳酸乙酯、己酸乙酯和己酸的测定[J/OL]. *酿酒科技*. (2024-07-05) [2024-11-05]. <http://kns.cnki.net/kcms/detail/52.1051.TS.20240704.1514.002.html>.
- FENG S Y, TIAN H Y, CAO W W, et al. Determination of methanol, ethyl acetate, ethyl lactate, ethyl caproate and caproic acid in Baijiu[J/OL]. *Liquor-Making Science & Technology*. (2024-07-05) [2024-11-05]. <http://kns.cnki.net/kcms/detail/52.1051.TS.20240704.1514.002.html>.
- [16] XU M L, YU Y, RAMASWAMY H S, et al. Characterization of Chinese liquor aroma components during aging process and liquor age discrimination using gas chromatography combined with multivariable statistics[J]. *Scientific Reports*, 2017, 7: 39671.
- [17] SHEN G H, HAN L J, FAN X, et al. Classification of fish meal produced in China and Peru by online near infrared spectroscopy with characteristic wavelength variables[J]. *Journal of Near Infrared Spectroscopy*, 2017, 25(1): 63-71.
- [18] 王武力, 马海乐, 李含, 等. 基于近红外光谱技术的蜂胶萃余物中黄酮提取的原位实时监测[J]. *食品与机械*, 2024, 40(9): 15-19, 72.
- WANG W L, MA H L, LI H, et al. In-situ real-time monitoring of flavonoids extraction from propolis extracts based on near infrared spectroscopy[J]. *Food & Machinery*, 2024, 40(9): 15-19, 72.
- [19] RABATEL G, MARINI F, WALCZAK B, et al. VSN: Variable sorting for normalization[J]. *Journal of Chemometrics*, 2019, 34(2): e3164.
- [20] BI Y M, YUAN K L, XIAO W Q, et al. A local pre-processing method for near-infrared spectra, combined with spectral segmentation and standard normal variate transformation[J]. *Analytica Chimica Acta*, 2016, 909: 30-40.
- [21] JIAO Y P, LI Z C, CHEN X S, et al. Preprocessing methods for near-infrared spectrum calibration[J]. *Journal of Chemometrics*, 2020, 34(11): 1-19.
- [22] 方慧, 张昭, 王海龙, 等. 基于中红外光谱技术鉴别转基因大豆的方法研究[J]. *光谱学与光谱分析*, 2017, 37(3): 760-765.
- FANG H, ZHANG Z, WANG H L, et al. Identification of transgenic soybean varieties using mid-infrared spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2017, 37(3): 760-765.
- [23] 黄敏, 万相梅, 朱启兵, 等. 基于高光谱图像技术的菜用大豆厚度检测[J]. *食品与生物技术学报*, 2012, 31(11): 1 142-1 147.
- HUANG M, WAN X M, ZHU Q B, et al. Thickness measurement of green soybean using hyperspectral imaging technology[J]. *Journal of Food Science and Biotechnology*, 2012, 31(11): 1 142-1 147.
- [24] HU S Q, WANG L. Age discrimination of Chinese Baijiu based on midinfrared spectroscopy and chemometrics[J]. *Journal of Food Quality*, 2021, 2 021: 5527826.
- [25] GÖKTAŞ A, AKKUŞ Ö. Comparison of partial least squares with other prediction methods via generated data[J]. *Journal of Statistical Computation and Simulation*, 2020, 90(16): 3 009-3 024.
- [26] WU J H, PENG H B, LI L, et al. FT-IR combined with chemometrics in the quality evaluation of Nongxiangxing Baijiu[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2023, 284: 121790.
- [27] 董新罗, 刘建学, 韩四海, 等. 白酒基酒中酮类物质的近红外光谱检测方法[J]. *分析测试学报*, 2020, 39(11): 1 427-1 432.
- DONG X L, LIU J X, HAN S H, et al. Rapid detection of ketones in base liquor by near infrared spectroscopy[J]. *Journal of Instrumental Analysis*, 2020, 39(11): 1 427-1 432.
- [28] DONG Q, YU C, LI L, et al. Near-infrared spectroscopic study of molecular interaction in ethanol-water mixtures[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2019, 222: 117183.
- [29] SHEN F, YING Y B, LI B B, et al. Discrimination of blended Chinese rice wine ages based on near-infrared spectroscopy[J]. *International Journal of Food Properties*, 2012, 15(6): 1 262-1 275.
- [30] 李水芳, 单杨, 朱向荣, 等. 近红外光谱结合化学计量学方法检测蜂蜜产地[J]. *农业工程学报*, 2011, 27(8): 350-354.
- LI S F, SAN Y, ZHU X R, et al. Detection of geographical origin of honey using near-infrared spectroscopy and chemometrics[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2011, 27(8): 350-354.
- [31] 刘小路, 薛璐, 鲁晓翔, 等. 蓝莓可溶性固形物、总酸近红外无损检测模型的建立[J]. *食品与生物技术学报*, 2016, 35(7): 752-756.
- LIU X L, XUE L, LU X X, et al. Non-destructive testing of soluble solids and total acidity in blueberry by near-infrared diffuse reflectance spectroscopy[J]. *Journal of Food Science and Biotechnology*, 2016, 35(7): 752-756.
- [32] 余松柏, 黄张君, 吴奇霄, 等. 基于近红外光谱构建酒用高粱主要理化指标的快速无损分析模型[J]. *食品工业科技*, 2023, 44(10): 311-319.
- YU S B, HUANG Z J, WU Q X, et al. Constructing rapid and undamaged detection models for main physicochemical indexes of brewing sorghum based on near infrared spectrum[J]. *Science and Technology of Food Industry*, 2023, 44(10): 311-319.
- [33] 刘建学, 杨国迪, 韩四海, 等. 白酒基酒中典型醇的近红外预测模型构建[J]. *食品科学*, 2018, 39(2): 281-286.
- LIU J X, YANG G D, HAN S H, et al. Prediction model for typical alcohols in base liquor based on near infrared spectroscopy[J]. *Food Science*, 2018, 39(2): 281-286.