

基于近红外光谱的沼液挥发性脂肪酸含量快速检测

刘金明^{1,2,3}, 郭坤林², 甄峰^{1,3}, 张鸿琼^{1,4}, 李文哲^{1,4}, 许永花^{5*}

(1. 东北农业大学工程学院, 哈尔滨 150030; 2. 黑龙江八一农垦大学电气与信息学院, 大庆 163319; 3. 中国科学院可再生能源重点实验室, 广州 510640; 4. 黑龙江省寒地农业可再生资源利用技术及装备重点实验室, 哈尔滨 150030; 5. 东北农业大学电气与信息学院, 哈尔滨 150030)

摘要: 挥发性脂肪酸 (Volatile Fatty Acids, VFA) 作为厌氧发酵过程的重要中间产物, 其在厌氧反应器中的累积能够反映出产甲烷菌的不活跃状态或厌氧发酵条件的恶化。为了实现农牧废弃物厌氧发酵过程分析和状态监控, 将近红外光谱 (Near Infrared Spectroscopy, NIRS) 与偏最小二乘 (Partial Least Squares, PLS) 相结合构建玉米秸秆和畜禽粪便厌氧发酵液乙酸、丙酸和总酸含量快速检测模型。将竞争自适应重加权采样法 (Competitive Adaptive Reweighted Sampling, CARS) 与遗传模拟退火 (Genetic Simulated Annealing, GSA) 算法相结合构建 CARS-GSA 算法对沼液中的乙酸、丙酸和总酸进行特征波长优选, 原始光谱数据 1 557 个波长点经预处理和波长优选后, 得到乙酸、丙酸和总酸特征波长变量分别为 135、101 和 245 个, 建立的回归模型验证决定系数分别为 0.988、0.923 和 0.886, 预测均方根误差 (Root Mean Squared Error of Prediction, RMSEP) 分别为 0.111、0.120 和 0.727, 相对分析误差分别为 9.685、3.685 和 3.484, 与全谱建模相比 RMSEP 分别减少了 17.78%、15.49% 和 1.22%, 能够满足农牧废弃物厌氧发酵过程发酵液中乙酸和丙酸含量的快速检测需求, 基本满足总酸的检测需求。结果表明, 通过构建 CARS-GSA 算法优选乙酸、丙酸和总酸的敏感波长变量, 参与建模的波长点数量显著减少, 有效降低了变量维度和模型复杂度, 提升了回归模型检测精度和预测能力, 为快速准确检测沼液 VFA 提供了新途径。

关键词: 厌氧发酵; 挥发性脂肪酸; 快速检测; 近红外光谱; 偏最小二乘; 遗传模拟退火算法; 竞争自适应重加权采样
doi: 10.11975/j.issn.1002-6819.2020.18.023

中图分类号: O657.33

文献标志码: A

文章编号: 1002-6819(2020)-18-0188-09

刘金明, 郭坤林, 甄峰, 等. 基于近红外光谱的沼液挥发性脂肪酸含量快速检测[J]. 农业工程学报, 2020, 36(18): 188-196. doi: 10.11975/j.issn.1002-6819.2020.18.023 http://www.tcsae.org

Liu Jinming, Guo Kunlin, Zhen Feng, et al. Rapid determination of volatile fatty acids in biogas slurry based on near infrared spectroscopy[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(18): 188-196. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2020.18.023 http://www.tcsae.org

0 引言

挥发性脂肪酸 (Volatile Fatty Acids, VFA) 作为厌氧发酵过程的重要中间产物, 为产甲烷阶段提供了底物^[1]。产甲烷菌主要利用 VFA 形成甲烷, 只有少部分甲烷由二氧化碳和氢气生成, 但二氧化碳和氢气生成甲烷时也经过高分子有机物形成 VFA 的中间过程^[2]。VFA 在厌氧反应器中的积累能反映出产甲烷菌的不活跃状态或厌氧发酵条件的恶化, 较高的 VFA 浓度对产甲烷菌有抑制作用, 过高的 VFA 浓度甚至会导致厌氧发酵发生“酸败”^[3]。在反应器运行过程中, 发酵液的 VFA 浓度常用作厌氧发酵过程的重要监控指标^[4]。通过监测发酵液中 VFA 的变化情况, 可以很好地了解有机物的降解过程以

及产甲烷菌的活性和系统的运行情况^[5]。为了对厌氧发酵状态进行有效监控, 有必要对发酵液的 VFA 含量进行快速、准确测定。

传统的 VFA 检测方法主要有精馏法、高效液相色谱法、气相色谱法和各种滴定技术^[5-6]。但传统检测方法存在处理时间长、设备操作复杂等问题, 难以满足厌氧发酵过程中通过快速测定 VFA 实现发酵过程状态监测的需求。针对厌氧发酵过程状态监测对 VFA 快速检测的需求, 相关学者深入研究了快速滴定法^[7]、新型色谱技术^[8]、电化学传感器^[9]、生物传感器^[10]和光谱分析技术^[11]在 VFA 快速检测方面的应用。光谱分析技术因其简便、快捷、无损、低成本的优势, 已在发酵液 VFA 检测方面得到了广泛应用^[12-13], 其中以近红外光谱 (Near Infrared Spectroscopy, NIRS) 定量分析技术的应用最为广泛^[14-16]。在应用 NIRS 对水体中的 VFA 含量进行快速检测方面, 主要以光谱预处理方法和多元定量校正方法的研究为主, 在 VFA 特征波长优选方面尚需进一步拓展, 以消除不相关和非线性波长点对模型精度的影响。

当前, NIRS 特征波长优选方法正朝着多种特征波长优选方法相结合的方向发展, 将区间偏最小二乘法

收稿日期: 2020-05-10 修订日期: 2020-06-28

基金项目: 中国科学院可再生能源重点实验室 (Y907k81001); 国家重点研发计划 (2019YFD1100603); 黑龙江省博士面上资助 (LBH-Z19087); 黑龙江八一农垦大学三横三纵支持计划 (ZRCQC202007); 黑龙江八一农垦大学学成人才科研启动计划 (XDB202006)

作者简介: 刘金明, 博士, 副教授, 主要从事光谱分析技术在农业领域的应用研究。Email: jinmingliu2008@126.com

*通信作者: 许永花, 副教授, 主要从事光谱分析技术方面的研究。

Email: xyhsy@126.com

(Interval Partial Least Squares, iPLS)^[17]、协同区间偏最小二乘法 (Synergy iPLS, SiPLS)^[18]、反向区间偏最小二乘法 (Backward iPLS, BiPLS)^[19]、连续投影算法^[20]、竞争自适应重加权采样法 (Competitive Adaptive Reweighted Sampling, CARS)^[21]等特征波长优选算法与遗传算法 (Genetic Algorithm, GA)^[22]、模拟退火算法 (Simulated Annealing Algorithm, SA)^[23]、粒子群优化算法^[24]等智能优化算法相结合进行 NIRS 特征波长变量优选已成为重要研究方向^[25-27]。在 NIRS 定量分析方面, GA 因其强大的特征波长优选能力已得到广泛应用^[28-29], 但 GA 存在早熟问题且进化后期搜索效率低。

遗传模拟退火 (Genetic Simulated Annealing, GSA) 算法是对 GA 的改进, 通过结合 SA 的温度参数设计适应度函数, 引入 Metropolis 扰动解选择复制策略, 有效利用 GA 强大搜索能力的同时解决了 GA 的两点不足, 在 NIRS 特征波长优选方面取得了较好的应用效果^[30]。GSA 在与 iPLS、SiPLS 和 BiPLS 相结合进行特征波长优选方面的性能也显著优于 GA^[30-32], 但在使用 iPLS、SiPLS 和 BiPLS 进行特征谱区优选时, 难以避免谱区内部存在冗余波长点。这些不相关和非线性的冗余波长点导致 GSA 编码过长, 严重影响了 GSA 特征波长点二次搜索的性能。

因此, 本文针对以秸秆和粪便为原料的厌氧发酵过程中, 沼液 VFA 快速检测的需求, 提出基于 NIRS 建立 VFA 快速检测模型, 并将 GSA 与 CARS 相结合构建 CARS-GSA 算法进行 VFA 特征波长优选, 有效解决 iPLS、SiPLS 和 BiPLS 敏感波段初步定位过程中存在冗余波长点的问题, 进而获取满足实际检测需求的有效特征波长变量, 以提高 VFA 快速检测模型的效率和精度。

1 材料与方法

1.1 样品采集与制备

试验用玉米秸秆取自东北农业大学校内试验田, 猪粪取自哈尔滨市三元畜产实业公司, 牛粪取自哈尔滨市宇峰奶牛养殖农民专业合作社, 接种物取自黑龙江省寒地农业可再生资源利用技术及装备重点实验室。将采集的玉米秸秆自然风干后一部分经铡草机切成 10 mm 的秸秆段备用, 另一部分经锤片式粉碎机 (10 mm 筛网) 粉碎成秸秆粉备用。分别以秸秆段、秸秆粉、牛粪、猪粪、秸秆粉猪粪混合物 (按总固体 (Total Solid, TS) 比 1:1) 为厌氧发酵原料, 以实验室 500 L 发酵罐常年驯化正常产气的牛粪厌氧发酵液为接种物, 进行批式厌氧发酵试验。秸秆、牛粪、猪粪和接种物的 TS 浓度分别为 86.02%、26.62%、31.22% 和 4.76%, 按 TS 接种比 1:1, 调整厌氧发酵原料和接种物添加量, 使 5 种原料对应的发酵系统起始 TS 浓度分别为 7%、6%、8%、7% 和 7%。在中温 (36±1) °C 恒温水浴槽中, 分别采用 5 和 10 L 下口瓶作为反应器, 进行 2 个批次的厌氧发酵试验, 有效发酵容积分别为 3.5 和 7 L。试验过程中每天定时对厌氧发酵反应器进行手摇搅拌 2 次, 混匀料液的同时避免浮渣

结壳。为了获取有代表性的 VFA 浓度数据样本, 采集发酵液样品主要在批式厌氧发酵前半程进行。5 L 发酵罐从装样后第 2 天开始, 每天 8:00 采集发酵液样品 40 mL 存放于 3 个 15 mL 离心管中, 共计采样 16 次。为防止料液 TS 浓度变高, 对厌氧发酵过程产生不良影响, 于第 8 天补水 300 mL。10 L 发酵罐从装样后第 2 天开始采样, 共计采样 15 次, 不需补水; 共计采集与制备发酵液样品 155 个, 于 -20 °C 冰箱冷冻保存。

1.2 光谱数据采集

发酵液冷冻样品溶解后在冷冻离心机中以 12 000 r/min 离心 10 min 后, 取上清液待测。使用 Nicolet 公司的 Antaris II 型傅里叶近红外光谱仪对采集样品进行透射光谱扫描, 光谱采集范围 4 000~10 000 cm⁻¹ (1 000~2 500 nm), 分辨率为 8.0 cm⁻¹, 样品扫描 32 次, 数据保存格式为 lg(1/T), 背景每小时扫描一次, 装样方式为 1 mm 光程石英比色皿前置通道扫描。在保持室内温湿度基本稳定的情况下, 每个样品装样 3 次, 取 3 次扫描平均值作为样品的原始光谱。原始光谱的波长数量为 1 557 个, 数据点间距为 3.86 cm⁻¹, 起始波数为 10 001.03 cm⁻¹, 结束波数为 3 999.64 cm⁻¹。

1.3 VFA 测定

使用安捷伦 GC-6890N 气相色谱仪测定厌氧发酵过程中沼液的 VFA 浓度。采用外标法建立 VFA 标准曲线, 先制备乙酸、丙酸、丁酸、异丁酸和异戊酸的混合标准溶液, 再使用去离子水稀释至 6 种不同浓度, 并测定不同浓度标准溶液各成分对应的出峰时间和积分面积。将混合溶液的保留时间与单品的保留时间进行比较, 根据已知标准溶液中各物质的浓度和积分时间绘制标准曲线。对溶解、离心并采集透射光谱数据后的厌氧发酵液样品上清液进行 VFA 含量测定。将其与 25% 偏磷酸溶液按体积比 10:1 进行混合, 然后再以 12 000 r/min 离心 10 min 后取上清液, 将上清液使用 0.45 μm 超滤膜过滤, 取滤液进行 VFA 浓度测定。

1.4 波长优选方法

1.4.1 CARS 算法

CARS 算法基于“适者生存”的原则, 将蒙特卡洛采样 (Monte-Carlo Sampling, MCS)、指数衰减函数和自适应加权采样 (Adaptive Reweighted Sampling, ARS) 相结合获取波长子集, 基于偏最小二乘 (Partial Least Squares, PLS) 回归系数绝对值的大小获取一系列变量组合, 并选择交叉验证均方根误差 (Root Mean Squared Error of Cross Validation, RMSECV) 值最小的子集作为特征波长。CARS 在迭代过程中引入 MCS 和 ARS 2 个随机因素, 难以保证每次优选结果的一致性。可以采用多次运行 CARS 算法, 每次都选中的波长点代表着光谱数据中与待测目标属性相关性高的波长点, 选定这些多次都选中波长作为特征波长, 能够建立高性能的回归模型。

1.4.2 CARS-GSA 算法

CARS-GSA 算法以 CARS 优选后的特征波长为输入, 采用 GSA 算法对 CARS 优选结果进行再优化, 以剔除

CARS 优选结果中相关性较差的波长点, 从而进一步提高建模性能。CARS-GSA 以 CARS 优选后特征波长点数为码长, 以 PLS 回归模型的 K 折 RMSECV 为目标函数, 按初始种群个数约为码长的三分之一进行二进制编码和种群初始化。“1”和“0”分别表示该波长点对应的数据“是”、“否”选中参与运算。在确定初始温度、退温操作, 并计算适应度函数值后, 执行多个轮次的 GSA 选择、交叉、变异和 Metropolis 选择复制进化操作, 完成 NIRS 特征波长点的优选。多次执行 GSA 算法对 CARS 优选结果进行再优化, 并选择多次重复选中的波长点作为特征波长变量建立 PLS 回归模型, 能够得到较高的回归模型性能。

1.5 回归模型建立及评价

在建立发酵液 VFA 快速检测模型时, 先去掉 NIRS 数据中吸光度值较大的平顶峰区域的波长点, 以消除水中羟基对光谱测量结果的影响, 再分别以乙酸、丙酸和总酸浓度为待测目标属性, 进行光谱预处理方法优选。在进行光谱预处理方法优选时, 先使用 Savitzky-Golay (SG) 平滑、多元散射校正 (Multivariate Scattering Correction, MSC)、标准正态变换 (Standard Normal Variate, SNV)、一阶导数 (First Derivative, FD) 和二阶导数及其多种方法相结合进行光谱预处理, 再使用 Sample Set Portioning Based on Joint X-Y Distances (SPXY) 法将预处理后光谱按一定比例划分为校正集样本和验证集样本, 并建立全谱下的 PLS 回归模型并计算模型性能, 基于校正集的平均相对误差确定乙酸、丙酸和总酸回归模型采用的光谱预处理方法。然后, 再采用 CARS-GSA 对沼液 VFA 进行特征波长优选, 并基于蒙特卡洛交叉验证的预测残差平方和最小确定特征波长对应的最佳主成分个数, 进而分别建立乙酸、丙酸和总酸对应的 PLS 回归校正模型。最后, 采用校正决定系数 (R_c^2)、验证决定系数 (R_p^2)、校正均方根误差 (Root Mean Squared Error of Calibration, RMSEC)、预测均方根误差 (Root Mean Squared Error of Prediction, RMSEP) 和相对分析误差 (Residual Predictive Deviation, RPD) 对校正模型的性能进行评价。 R^2 (R_c^2 和 R_p^2) 和 RPD 的值越大, 均方根误差 (Root Mean Squared Error, RMSE) (RMSEC 和 RMSEP) 的值越小, 模型的精度越高。 R^2 、RMSE 和 RPD 的计算公式为

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (2)$$

$$\text{RPD} = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \hat{y}_i)^2}} \quad (3)$$

式中 y_i 是第 i 个样本的测量值, \hat{y}_i 是第 i 个样本的预测值, \bar{y} 是测量平均值, n 是样本个数。

本文算法包括光谱预处理、样本集划分、特征波长优选及回归模型构建等全部在 Matlab R2012b 软件平台中实现。

2 结果与分析

2.1 采集数据分析

在采用安捷伦 GC-6890N 气相色谱仪测定 155 个发酵液样本的 VFA 浓度时, 得到 81 个乙酸浓度有效数据、78 个丙酸浓度有效数据和 87 个总酸浓度有效数据 (总酸浓度为乙酸、丙酸、丁酸、异丁酸和异戊酸质量分数之和)。对获得的 VFA 样本有效浓度数据进行四分位数分析, 并绘制箱线图如图 1 所示。

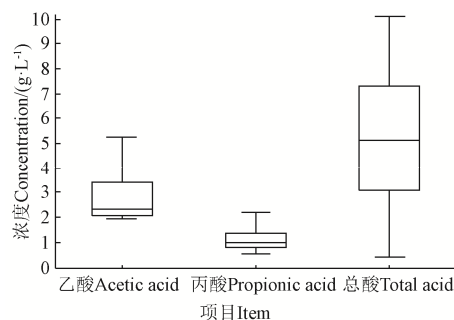


图 1 样本 VFA 浓度箱线图

Fig.1 Boxplots of Volatile Fatty Acid (VFA) concentrations for samples

由图 1 可知, 乙酸样本在低浓度区域占比较大, 丙酸样本略微偏向低浓度区域, 总酸样本分布比较均匀。乙酸样本偏离严重的原因在于厌氧发酵产乙酸、产甲烷平衡期产甲烷菌能够及时将生成的乙酸转化为甲烷和二氧化碳, 进而使平衡期阶段 (在整个发酵周期中时间占比较大) 的乙酸浓度偏低。

为消除光谱区域中平顶峰对建模结果的影响, 先剔除原始光谱数据中波数 $4\ 933.02 \sim 5\ 295.57\ \text{cm}^{-1}$ 的 95 个波长点, 再用剩余的 1 462 个有效波长点建立乙酸、丙酸和总酸回归模型, 并对不同光谱预处理方法下的回归模型性能进行评测。经计算比较后确定乙酸浓度回归模型采用的光谱预处理方法为 MSC+SG, 丙酸回归模型采用的光谱预处理方法为 SG+MSC, 总酸回归模型采用的光谱预处理方法为 FD+SNV+SG。样品原始光谱及预处理后的乙酸、丙酸和总酸光谱数据的平均光谱如图 2 所示。

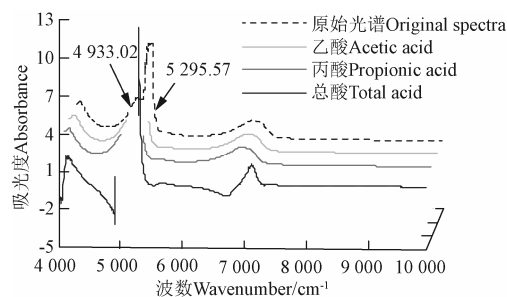


图 2 样品光谱数据

Fig.2 Spectroscopic data of the samples

对 81 个乙酸样品的原始光谱依次进行 MSC 和 SG 平滑处理后, 使用 SPXY 法划分为 60 个校正集样本和 21 个验证集样本; 对 78 个丙酸样品的原始光谱数据依次进行 SG 平滑和 MSC 处理后, 使用 SPXY 法划分为 60 个

校正集样本和 18 个验证集样本；对 87 个总酸样品的原始光谱数据依次进行 FD、SNV 和 SG 平滑处理后，使用 SPXY 法划分为 70 个校正集样本和 17 个验证集样本。乙酸、丙酸和总酸浓度值如表 1 所示。

表 1 样品 VFA 浓度

样本 Sample	成分 Component	样本数 NS	平均值 Mean	最大值 Max	最小值 Min	标准偏差 SD
校正集 Calibration set	乙酸	60	2.859	5.240	2.049	0.911
	丙酸	60	1.132	2.714	0.512	0.463
	总酸	70	5.253	10.095	0.424	2.492
验证集 Validation set	乙酸	21	2.683	5.231	2.067	1.079
	丙酸	18	1.112	2.344	0.516	0.442
	总酸	17	4.777	10.088	1.054	2.532

注：SD 是 Standard deviation 的缩写，NS 是 Number of sample 的缩写。

Note: SD is short for standard deviation, NS is short for number of sample.

2.2 特征波长优选

2.2.1 CARS 特征波长优选

在使用 CARS 优选乙酸回归模型特征波长时，先执行 500 轮次 CARS 算法，再按重复选中次数递增的方式选取 RMSEP 最小时对应的特征波长优选结果作为 CARS 的特征波长（记为 CARS500）。执行 500 次 CARS 算法共得到乙酸特征波长 383 个以波数表示，下同，选中次数最多的特征波长波数为 4 416.19 cm⁻¹，对应着乙酸-CH₃基团的组合频，选中次数为 457 次。选中次数较多的特征波长点主要分布在 4 000~4 600、4 750~4 930、5 300~5 500、5 750~6 050、6 750~7 100 和 7 500~7 800 cm⁻¹ 区域。其中 4 000~4 600 cm⁻¹ 对应着乙酸-CH₃基团的组合频，4 750~4 930 cm⁻¹ 对应着 C=O 和 -OH 基团的组合频，5 300~5 500 cm⁻¹ 对应着 -COOH 基团的一级倍频，5 750~6 050 cm⁻¹ 对应着 -CH₃基团的一级倍频，6 750~7 100 cm⁻¹ 对应着 C=O 和 -OH 基团的二级倍频，7 500~7 800 cm⁻¹ 对应着 -CH₃基团的二级倍频。CARS500 优选特征波长与乙酸平均光谱如图 3 所示。

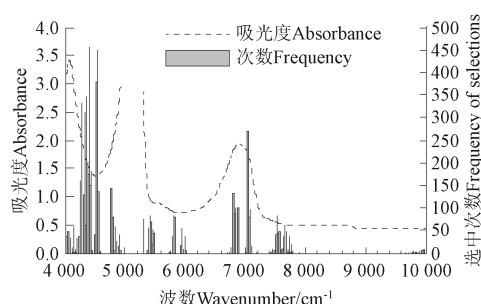


图 3 CARS500 优选乙酸特征波长

Fig.3 Characteristic wavelengths of acetic acid selected by CARS500

为分析不同重复选中次数下，CARS500 优选特征波长的建模性能，建立 RMSECV、RMSEP 和波长点个数随重复选中次数的变化关系，如图 4 所示。

由图 4 可知，RMSECV 随着选中波长点个数的减少整体上呈先迅速减少、再波浪状向前、最后跳跃式快速上升的形式，其中波长点数为 120 时，RMSECV 得到最小值 0.163，对应重复选中次数为 39 次。RMSEP 随选中

波长点个数减少整体呈锯齿型变化并逐渐增加的形式，其中重复选中次数为 30、选中波长数量为 142 时，所建 PLS 回归模型的 RMSEP 获得最小值为 0.116。采用 RMSEP 最小时对应的 142 个波长点作为 CARS500 优选的乙酸特征波长。

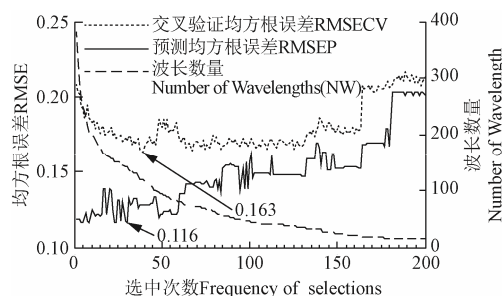


图 4 RMSE、波长数量和选中次数间的关系

Fig.4 Relationship between RMSE, Number of Wavelengths(NW) and frequency of selections

2.2.2 CARS-GSA 特征波长优选

在使用 CARS-GSA 优选发酵液乙酸特征波长时，以 CARS500 优选的 142 个波长点为码长随机生成 50 个染色体构建初始种群，执行 GSA 算法进行特征波长点二次优选。GSA 算法的初温确定系数取 100，退温系数取 0.9，进化代数取 100，交叉概率取 0.7，变异概率取 0.01，邻域解扰动位数取 10。连续执行算法 50 次，优选的乙酸特征波长中选中 35 次以上的波长共计 14 个。其中，波数 4 057.49、4 319.77、4 354.48、4 358.33、4 362.19、4 366.05、4 408.48、4 412.33、4 416.19、4 420.05、4 531.90 和 4 539.61 cm⁻¹ 对应着 -CH₃基团的组合频，波数 4 925.30 cm⁻¹ 对应着 C=O 基团的组合频，波数 5 311.00 cm⁻¹ 对应着 -COOH 基团的一级倍频。CARS-GSA 优选特征波长与乙酸平均光谱如图 5 所示。

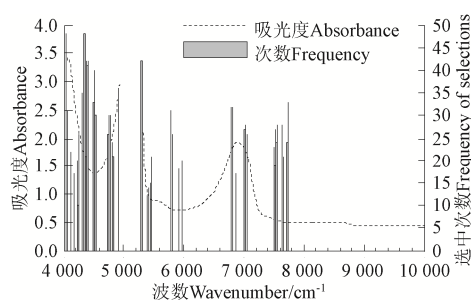


图 5 CARS-GSA 优选乙酸特征波长

Fig.5 Characteristic wavelengths of acetic acid selected by CARS-GSA

为分析 CARS-GSA 优选特征波长的建模性能，建立 RMSECV、RMSEP 与波长点个数间的对应关系，如图 6 所示。由图 6 可知，RMSECV 和 RMSEP 随选中波长点个数增加整体上呈先迅速减少、再趋于平缓、最后略有上升的趋势，但 RMSECV 的最小值要早于 RMSEP 出现。RMSECV 最小值对应的波长点数为 54、重复选中次数为 26，RMSEP 最小值对应的波长点数为 135、重复选中次数为 10，说明仅以 RMSECV 最小确定特征波长的方式容

易导致回归模型产生过拟合的问题。因此,选择 RMSEP 最小时对应的 135 个选中波长作为 CARS-GSA 优选的乙酸特征波长。由图 3 和图 5 中 RMSECV 和 RMSEP 最小值的对比可知, CARS-GSA 优选特征波长的建模性能优于 CARS-500 的建模性能。

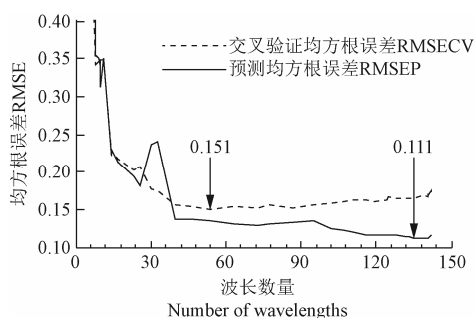


图 6 RMSE 与波长数量间的关系

Fig.6 Relationship between RMSE and NW

2.2.3 特征波长优选结果

按上述方法执行 CARS500 和 CARS-GSA 进行丙酸和总酸特征波长优选,得到 101 个丙酸特征波长和 245 个总酸特征波长。乙酸、丙酸和总酸特征波长分布情况如图 7 所示。

由图 7 可知, VFA 特征波长全部位于 $8\,000\text{ cm}^{-1}$ 以下的中低频区域,其中 $4\,000\sim 4\,933$ 、 $5\,296\sim 5\,600$ 和 $6\,600\sim 7\,200\text{ cm}^{-1}$ 区域分布的特征波长点最多,这 3 部分正好对应着光谱数据中吸收峰较强、分辨率较好的区域。丙酸特征波长在 $4\,100\sim 4\,500\text{ cm}^{-1}$ 区域有 53 个,对应着 -CH_2 和 -CH_3 基团的组合频;在 $4\,000\sim 4\,900\text{ cm}^{-1}$ 区域有 7 个,对应着 C=O 和 -OH 基团的组合频;在 $5\,300\sim 5\,320\text{ cm}^{-1}$ 区域有 2 个,对应着 -COOH 基团的一级倍频;在 $5\,670\sim 5\,700\text{ cm}^{-1}$ 区域有 9 个,对应着 -CH_2 基团的一级倍频;在 $6\,000\sim 6\,070\text{ cm}^{-1}$ 区域有 13

个,对应着 -CH_3 基团的一级倍频;在 $6\,860\sim 7\,060\text{ cm}^{-1}$ 区域有 17 个,对应着 C=O 、 -CH_2 和 -OH 的二级倍频。总酸特征波长在 $4\,000\sim 4\,720\text{ cm}^{-1}$ 区域有 139 个,对应着 C-C 、 C=O 、 -CH 、 -CH_2 和 -CH_3 基团的组合频;在 $4\,800\sim 4\,930\text{ cm}^{-1}$ 区域有 27 个,对应着 C=O 和 -OH 基团的组合频;在 $5\,300\sim 5\,380\text{ cm}^{-1}$ 区域有 19 个,对应着 -COOH 基团的一级倍频;在 $5\,930\sim 6\,010\text{ cm}^{-1}$ 区域有 11 个,对应着 -CH 、 -CH_2 和 -CH_3 基团的一级倍频;在 $6\,590\sim 6\,600\text{ cm}^{-1}$ 区域有 2 个,对应着 C=O 基团的二级倍频;在 $6\,730\sim 7\,200\text{ cm}^{-1}$ 区域有 47 个,对应着 C=O 、 -CH 、 -CH_2 、 -CH_3 和 -OH 的二级倍频。通过分析乙酸、丙酸和总酸特征波长可知, CARS-GSA 与 CARS500 优选特征波长结果具有很好的一致性, CARS-GSA 只是剔除掉 CARS500 优选特征波长中选中次数较少的相关性较差波长点。

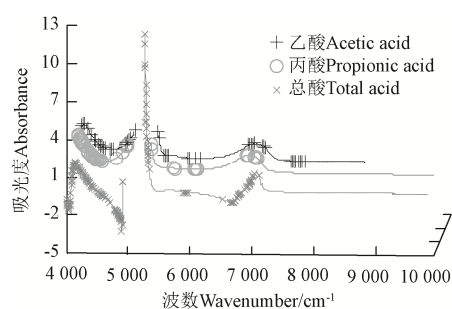


图 7 VFA 特征波长优选结果

Fig.7 Optimized results of characteristic wavelengths for VFA

2.3 回归模型评价与分析

为评测 2 种波长优选算法的性能,以 CARS500 和 CARS-GSA 优选后的特征波长变量作为 PLS 回归模型的输入,建立沼液 VFA 定量回归模型,并与全谱建模结果 (Full-PLS)、单次 CARS (运行 10 次取最佳结果) 优选特征波长的建模效果进行对比,结果如表 2 所示。

表 2 VFA PLS 回归模型评价指标

Table 2 Evaluation indicators of PLS regression models for VFA

成分 Component	方法 Methods	波长数量 NW	校正决定系数 R_c^2	验证决定系数 R_p^2	校正均方根误差 RMSEC/(g·L ⁻¹)	预测均方根误差 RMSEP/(g·L ⁻¹)	相对分析误差 RPD	主成分数 PCs
乙酸 Acetic acid	Full-PLS	1 462	0.976	0.983	0.139	0.135	8.021	13
	CARS	39	0.976	0.987	0.140	0.119	9.074	10
	CARS500	142	0.983	0.987	0.116	0.116	9.306	14
	CARS-GSA	135	0.983	0.988	0.117	0.111	9.685	13
丙酸 Propionic acid	Full-PLS	1 462	0.997	0.893	0.024	0.142	3.108	24
	CARS	30	0.942	0.912	0.108	0.130	3.413	10
	CARS500	122	0.989	0.922	0.048	0.124	3.569	22
	CARS-GSA	101	0.991	0.923	0.044	0.120	3.685	22
总酸 Total acid	Full-PLS	1 461	0.878	0.880	0.820	0.736	3.441	7
	CARS	112	0.873	0.867	0.838	0.785	3.225	5
	CARS500	258	0.859	0.884	0.876	0.729	3.472	5
	CARS-GSA	245	0.860	0.886	0.874	0.727	3.484	5

注: PCs 是 principal components 的缩写。

Note: PCs is short for principal components.

由表 2 可知,在单次 CARS 优选特征波长建立的 VFA 回归模型中,乙酸和丙酸 CARS 回归模型的性能优于全谱建模,而总酸 CARS 回归模型的性能弱于全谱建模。原因在于乙酸和丙酸的结构相对简单, CARS 能够快速定

位到相关性高的特征波长点,而总酸的结构相对复杂,不同基团对应的特征波长点数量较多,当使用 CARS 剔除波长点时可能去掉某些相关性较高的特征波长点,导致建模性能受到影响。多次执行 CARS 算法进行特征波

长优选可以解决单次 CARS 算法优选总酸特征波长建模性能较差的问题。

CARS500 优选特征波长建立的乙酸回归模型 R_c^2 和 R_p^2 都大于 0.98, RPD 大于 9.30; 建立的丙酸回归模型 R_c^2 和 R_p^2 都大于 0.92, RPD 大于 3.56; 建立的总酸回归模型 R_c^2 和 R_p^2 都大于 0.85, RPD 大于 3.47。执行 500 次 CARS 算法建立的乙酸回归模型非常成功, 建立的丙酸回归模型成功, 建立的总酸回归模型基本成功。基于多次 CARS 算法优选特征波长建立的乙酸、丙酸和总酸度回归模型与全谱建模相比, 回归模型性能都有一定的提升, RMSEP 分别降低了 14.07%、12.68% 和 0.95%, 分别剔除了全谱中 90.29%、91.66% 和 82.34% 的冗余波长点。乙酸回归模型精度提升最大的主要原因在于乙酸的分子结构相对简单, CARS500 特征波长优选算法在多次重复执行 CARS 算法时能够准确地定位到乙酸-COOH 和 -CH₃ 基团对应的相关性较高波长点。

CARS-GSA 优选特征波长建立的乙酸回归模型的 R_c^2 和 R_p^2 都大于 0.98, RPD 大于 9.68, 说明建立的乙酸回归模型非常成功; CARS-GSA 优选特征波长建立的丙酸回归模型的 R_c^2 和 R_p^2 都大于 0.92, RPD 大于 3.68, 说

明丙酸回归模型建模成功; CARS-GSA 优选特征波长建立的总酸回归模型的 R_c^2 和 R_p^2 都大于 0.85, RPD 大于 3.48, 说明总酸回归模型建模基本成功。CARS-GSA 优选特征波长建立的 VFA 浓度回归模型中, 乙酸回归模型的性能最好, 丙酸次之, 总酸最差。CARS-GSA 优选的乙酸、丙酸和总酸特征波长数量分别为 135、101 和 245, 总酸特征波长数量约为乙酸、丙酸特征波长数量之和。CARS-GSA 优选特征波长建立的乙酸、丙酸和总酸回归模型的性能都高于 CARS500 建立的相应模型, 与 CARS500 优选特征波长建立的乙酸、丙酸和总酸回归模型相比, RMSEP 分别降低了 4.31%、3.23% 和 0.27%。CARS-GSA 算法分别剔除了 CARS500 优选乙酸、丙酸和总酸特征波长中 4.93%、17.21% 和 5.04% 的冗余波长点。这一方面说明了多次执行 CARS 算法优选特征波长建立的回归模型性能优异, 另一方面也体现了 GSA 算法的全局优化性能, 能够实现优中选优。

采用 CARS-GSA 作为厌氧发酵过程中发酵液乙酸、丙酸和总酸浓度的特征波长优选方案, 以优选后的特征波长分别建立乙酸、丙酸和总酸浓度 PLS 回归模型并进行性能评测, 其结果如图 8 所示。

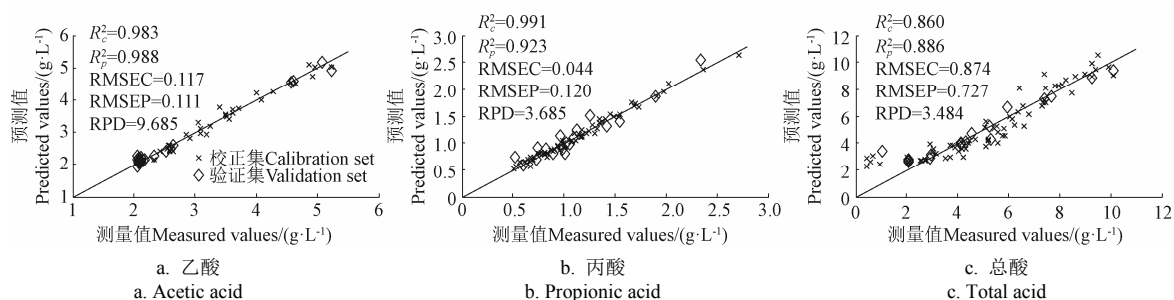


图 8 VFA 实测值与预测值分布

Fig.8 Distribution of measured and predicted values for VFA

由图 8 可知, 乙酸的实测值与预测值点基本呈对角线分布, 说明乙酸回归模型预测性能优异, 但由于采用 SPXY 法划分校正集和验证集, 导致乙酸样本在值空间上分布不是很均匀; 丙酸校正集样本点基本都位于对角线上, 验证集样本点均匀的分布于对角线两侧, 说明丙酸回归模型预测性能较好; 总酸的实测值和预测值分布于对角线两侧, 且比较分散, 说明总酸回归模型的预测性能有待提高。总酸回归模型预测精度较差的原因在于总酸浓度为乙酸、丙酸、丁酸、异丁酸和异戊酸质量分数之和, 在上述成分测量过程中客观存在且难以避免的仪器误差、操作误差等系统误差的累加导致总酸浓度测量误差较大, 从而影响了总酸回归模型的建模精度。CARS-GSA 乙酸回归模型 R_p^2 为 0.988, RMSEP 为 0.111, RPD 为 9.685, 与全谱相比 RMSEP 降低了 17.78%; CARS-GSA 丙酸回归模型 R_p^2 为 0.923, RMSEP 为 0.120, RPD 为 3.685, 与全谱相比 RMSEP 降低了 15.49%; CARS-GSA 总酸回归模型 R_p^2 为 0.886, RMSEP 为 0.727, RPD 为 3.484, 与全谱相比 RMSEP 降低了

1.22%。综上所述, 基于 CARS-GSA 特征波长优选方法建立的 VFA 回归模型可以满足农牧废弃物厌氧发酵过程中对乙酸和丙酸的快速检测需求, 基本满足对总酸的检测需求。

3 结 论

1) 本研究采用 NIRS 结合化学计量学方法进行沼液 VFA 的快速检测, 构建模型进行特征波长优选, 建立的乙酸、丙酸和总酸 PLS 回归模型验证决定系数分别为 0.988、0.923 和 0.886, 预测均方根误差分别为 0.111、0.120 和 0.727, 且 RPD 都大于 3, 能够满足农牧废弃物厌氧发酵过程中对发酵液乙酸和丙酸浓度的快速检测需求, 基本满足总酸浓度的检测需求。

2) CARS500 在发酵液 VFA 浓度 NIRS 特征波长优选方面具有良好的性能, 通过多次执行 CARS 算法并选取重复选中波长点作为特征波长的方式能够有效提高建模精度和效率, 并在一定程度上解决了 CARS 算法优选特征波长结果的随机性问题。

3) CARS-GSA 采用 GSA 对多次 CARS 优选的特征波长进行二次优化,能够有效去除 CARS500 优选波长中相关性较弱的冗余波长点,在提高建模精度和检测效率的同时,确立了乙酸、丙酸和总酸相关基团与其特征波长的对应关系和特征波长在光谱区间内的分布规律。

[参 考 文 献]

- [1] 史凤梅, 徐洪涛, 卢玢宇, 等. 温度对养殖废水厌氧发酵初期产酸的影响及其原因分析[J]. 农业工程学报, 2018, 34(增刊): 42-47.
- Shi Fengmei, Xu Hongtao, Lu Binyu, et al. Effects of temperature on production of volatile fatty acids in mesophilic anaerobic fermentation of swine wastewater and its cause analysis[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(Supp.): 42-47. (in Chinese with English abstract)
- [2] 于佳动, 赵立欣, 冯晶, 等. 序批式秸秆牛粪混合厌氧干发酵过程物料理化及渗滤特性[J]. 农业工程学报, 2019, 35(20): 228-234.
- Yu Jiadong, Zhao Lixin, Feng Jing, et al. Physicochemical and percolating characteristics of sequencing batch dry anaerobic digestion of straw-cow manure mixture[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(20): 228-234. (in Chinese with English abstract)
- [3] 宋香育, 张克强, 房芳, 等. 工艺措施对猪粪秸秆混合厌氧干发酵产气性能的影响[J]. 农业工程学报, 2017, 33(11): 233-239.
- Song Xiangyu, Zhang Keqiang, Fang Fang, et al. Influences of different technological strategies on performance of anaerobic co-digestion of pig manure with straw in solid-state[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2017, 33(11): 233-239. (in Chinese with English abstract)
- [4] Sbarciog M, Giovannini G, Chamy R, et al. Control and estimation of anaerobic digestion processes using hydrogen and volatile fatty acids measurements[J]. Water Science and Technology, 2018, 78(10): 2027-2035.
- [5] Chatterjee B, Radhakrishnan L, Mazumder D. New approach for determination of volatile fatty acid in anaerobic digester sample[J]. Environmental Engineering Science, 2018, 35(4): 333-351.
- [6] Lahav O, Morgan B E. Titration methodologies for monitoring of anaerobic digestion in developing countries: A review[J]. Journal of Chemical Technology and Biotechnology, 2004, 79(12): 1331-1341.
- [7] Hey T, Sandstrom D, Ibrahim V, et al. Evaluating 5 and 8 pH-point titrations for measuring VFA in full-scale primary sludge hydrolysate[J]. Water SA, 2013, 39(1): 17-22.
- [8] Ward A J, Bruni E, Lykkegaard M K, et al. Real time monitoring of a biogas digester with gas chromatography, near-infrared spectroscopy, and membrane-inlet mass spectrometry[J]. Bioresource Technology, 2011, 102(5): 4098-4103.
- [9] Peris M, Escuder-Gilabert L. On-line monitoring of food fermentation processes using electronic noses and electronic tongues: A review[J]. Anal Chim Acta, 2013, 804: 29-36.
- [10] Yu Z, Leng X, Zhao S, et al. A review on the applications of microbial electrolysis cells in anaerobic digestion[J]. Bioresource Technology, 2018, 255: 340-348.
- [11] Nguyen Duc, Gadhamshetty Venkataramana, Nitayavardhana Saoharit, et al. Automatic process control in anaerobic digestion technology: A critical review[J]. Bioresource Technology, 2015, 193: 513-522.
- [12] Bruni E, Ward A J, Kocks M, et al. Comprehensive monitoring of a biogas process during pulse loads with ammonia[J]. Biomass & Bioenergy, 2013, 56: 211-220.
- [13] Falk H M, Reichling P, Andersen C, et al. Online monitoring of concentration and dynamics of volatile fatty acids in anaerobic digestion processes with mid-infrared spectroscopy[J]. Bioprocess and Biosystems Engineering, 2015, 38(2): 237-249.
- [14] Stockl Andrea, Lichti Fabian. Near-infrared spectroscopy (NIRS) for a real time monitoring of the biogas process[J]. Bioresource Technology, 2018, 247: 1249-1252.
- [15] Nespeca Maurilio Gustavo, Rodrigues Caroline Varella, Santana Kamili Oliveira, et al. Determination of alcohols and volatile organic acids in anaerobic bioreactors for H₂ production by near infrared spectroscopy[J]. International Journal of Hydrogen Energy, 2017, 42(32): 20480-20493.
- [16] Li L, Peng X Y, Wang X M, et al. Anaerobic digestion of food waste: A review focusing on process stability[J]. Bioresource Technology, 2018, 248: 20-28.
- [17] Rato Tiago J, Reis Marco S. Multiresolution interval partial least squares: A framework for waveband selection and resolution optimization[J]. Chemometrics and Intelligent Laboratory Systems, 2019, 186: 41-54.
- [18] 邹小波, 张俊俊, 黄晓玮, 等. 基于音频和近红外光谱融合技术的西瓜成熟度判别[J]. 农业工程学报, 2019, 35(9): 301-307.
- Zou Xiaobo, Zhang Junjun, Huang Xiaowei, et al. Distinguishing watermelon maturity based on acoustic

- characteristics and near infrared spectroscopy fusion technology[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(9): 301-307. (in Chinese with English abstract)
- [19] Zareef M, Chen Q S, Ouyang Q, et al. Prediction of amino acids, caffeine, theaflavins and water extract in black tea using FT-NIR spectroscopy coupled chemometrics algorithms[J]. Analytical Methods, 2018, 10(25): 3023-3031.
- [20] 张亚坤, 罗斌, 宋鹏, 等. 基于近红外光谱的大豆叶片可溶性蛋白含量快速检测[J]. 农业工程学报, 2018, 34(18): 187-193.
- Zhang Yakun, Luo Bin, Song Peng, et al. Rapid determination of soluble protein content for soybean leaves based on near infrared spectroscopy[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(18): 187-193. (in Chinese with English abstract)
- [21] 王巧华, 梅璐, 马美湖, 等. 利用机器视觉与近红外光谱技术的皮蛋无损检测与分级[J]. 农业工程学报, 2019, 35(24): 314-321.
- Wang Qiaohua, Mei Lu, Ma Meihu, et al. Nondestructive testing and grading of preserved duck eggs based on machine vision and near-infrared spectroscopy[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(24): 314-321. (in Chinese with English abstract)
- [22] Jiang H, Xu W, Chen Q. Comparison of algorithms for wavelength variables selection from near-infrared (NIR) spectra for quantitative monitoring of yeast (*Saccharomyces cerevisiae*) cultivations[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019, 214: 366-371.
- [23] Ye Dandan, Sun Laijun, Zou Borui, et al. Non-destructive prediction of protein content in wheat using NIRS[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2018, 189: 463-472.
- [24] Zhu Y W, Chen X Y, Wang S M, et al. Simultaneous measurement of contents of liquiritin and glycyrrhizic acid in liquorice based on near infrared spectroscopy[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2018, 196: 209-214.
- [25] 杨玮, 李民赞, 郑立华, 等. 冬枣氮素含量预测模型中特征波长选择方法的应用[J]. 农业工程学报, 2015, 31(增刊 2): 164-168.
- Yang Wei, Li Minzan, Zheng Lihua, et al. Application of spectral screening method on prediction model of nitrogen content of jujube leaves[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2015, 31(Supp.2): 164-168. (in Chinese with English abstract)
- [26] 朱瑶迪, 张佳烨, 李苗云, 等. 肽聚糖对肉制品中产气荚膜梭菌芽孢萌发率影响及预测[J]. 农业工程学报, 2020, 36(4): 287-293.
- Zhu Yaodi, Zhang Jiaye, Li Miaoyun, et al. Effect of different peptidoglycan on *Clostridium perfringens* spore germination and quantitative prediction[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(4): 287-293. (in Chinese with English abstract)
- [27] Yun Y H, Bin J, Liu D L, et al. A hybrid variable selection strategy based on continuous shrinkage of variable space in multivariate calibration[J]. Analytica Chimica Acta, 2019, 1058: 58-69.
- [28] Yang M, Xu D, Chen S, et al. Evaluation of machine learning approaches to predict soil organic matter and pH using vis-NIR spectra[J]. Sensors, 2019, 19(2): 263.
- [29] Song J, Li G, Yang X. Optimizing GA-PLS model of soluble solids content in Fukumoto navel orange based on Vis-NIR transmittance spectroscopy using discrete wavelet transform[J]. Journal of the Science of Food and Agriculture, 2019, 99(11): 4898-4903.
- [30] 刘金明, 初晓冬, 王智, 等. 玉米秸秆纤维素和半纤维素 NIRS 特征波长优选[J]. 光谱学与光谱分析, 2019, 39(3): 743-750.
- Liu Jinming, Chu Xiaodong, Wang Zhi, et al. Optimization of characteristic wavelength variables of near infrared spectroscopy for detecting contents of cellulose and hemicellulose in corn stover[J]. Spectroscopy and Spectral Analysis, 2019, 39(3): 743-750. (in Chinese with English abstract)
- [31] 刘金明, 程秋爽, 甄峰, 等. 基于 GSA 的厌氧发酵原料碳氮比 NIRS 快速检测[J]. 农业机械学报, 2019, 50(11): 323-330.
- Liu Jinming, Cheng Qiushuang, Zhen Feng, et al. Rapid determination of C/N ratio for anaerobic fermentation feedstocks using near infrared spectroscopy based on GSA[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(11): 323-330. (in Chinese with English abstract)
- [32] Liu Jinming, Li Nan, Zhen Feng, et al. Rapid detection of carbon-nitrogen ratio for anaerobic fermentation feedstocks using near-infrared spectroscopy combined with BiPLS and GSA[J]. Applied Optics, 2019, 58(18): 5090-5097.

Rapid determination of volatile fatty acids in biogas slurry based on near infrared spectroscopy

Liu Jinming^{1,2,3}, Guo Kunlin², Zhen Feng^{1,3}, Zhang Hongqiong^{1,4}, Li Wenzhe^{1,4}, Xu Yonghua^{5*}

(1. College of Engineering, Northeast Agricultural University, Harbin 150030, China; 2. College of Electrical and Information, Heilongjiang Bayi Agricultural University, Daqing 163319, China; 3. CAS Key Laboratory of Renewable Energy, Guangzhou 510640, China; 4. Heilongjiang Key Laboratory of Technology and Equipment for the Utilization of Agricultural Renewable Resources, Harbin 150030, China; 5. School of Electrical and Information, Northeast Agricultural University, Harbin 150030, China)

Abstract: Volatile Fatty Acids (VFA), serving as important intermediate products in Anaerobic Digestion (AD), have been considered as the key variables in most AD monitoring strategies, as they respond to incoming imbalances, indicating the buffer capacity of digesters to process disturbance and imminent digester failure that caused by sudden operational changes. In order to ensure efficient operation of AD while improve the utilization rate of raw materials, it is necessary to accurately monitor and evaluate the operation state of biogas engineering, via detecting the concentrations of VFA in the process of biogas production with corn stover and animal manure as feedstocks. Previously, the rapid detection models of Acetic Acid (AA), Propionic Acid (PA) and Total Acid (TA) in biogas slurry have been constructed, using the Near Infrared Spectroscopy (NIRS) technique combined with the Partial Least Squares (PLS), aiming to overcome the time consuming and high-cost in the traditional chemical analysis method. However, a prediction model can trigger the high complexity and low accuracy, due to the spectroscopic data generally includes quantities of invalid redundant information. In this study, an integrated algorithm was presented, based on the Competitive Adaptive Reweighted Sampling (CARS) and genetic simulated annealing algorithm (GSA), to optimize the characteristic wavelength variables of AA, PA, and TA, and thereby to improve the efficiency and precision of NIRS detection models. An AD experiment was carried out with corn stover, pig manure and cow manure as feedstocks, where 155 samples of biogas slurry were collected. The NIRS data of biogas slurry was acquired in a transmittance mode using the AntarisTM II FT-NIR spectrophotometer equipped with a quartz cuvette. A Gas Chromatography (GC) system was used to measure the VFA of biogas slurry, where 81 valid data of AA, 78 valid data of PA, and 87 valid data of TA were obtained to establish the regression model. One segment of the spectrum with 95 wavelength points was removed from 4 933.02 to 5 295.57 cm^{-1} , and 1462 wavelength variables remained, mainly due to the saturation of spectrum can be caused by the strong combination band of -OH from water. The spectral preprocessing methods were selected, according to the mean relative error of calibration set. Correspondingly, the samples were divided into the calibration set and validation set, using Sample Set Portioning based on Joint X-Y Distances (SPXY) algorithm. The number of characteristic wavelength variables for AA, PA, and TA were 135, 101, and 245, respectively. The PLS regression models were established with the characteristic wavelengths of AA, PA, and TA, where the results were the coefficients of multiple determination for prediction is 0.988, root mean squared error of prediction (RMSEP) of 0.111, and the residual predictive deviation (RPD) of 9.685 for AA, coefficients of multiple determination for prediction is 0.922, RMSEP of 0.120, and RPD of 3.685 for PA, coefficients of multiple determination for prediction is 0.886, RMSEP of 0.727, and RPD of 3.484 for TA. Meanwhile, compared with the whole spectrum model, the RMSEP in the CARS-GSA model decreased by 17.78%, 15.49%, and 1.22%, respectively, showing that the number of wavelengths significantly decreased after the optimization, whereas, the performance of regressive model was obviously higher than that of the whole wavelengths. The results demonstrate that the CARS-GSA model can fulfil the requirement of rapid detection for AA and PA concentrations in biogas slurry during anaerobic fermentation with agricultural waste as feedstocks, while basically meet the detection requirement of TA concentration. The CARS-GSA model also can be used to enhance the forecasting capability of the model, while reduce its complexity. The findings can provide a new way to improve the accuracy and robustness of prediction model, base on optimizing sensitive wavelengths for AA, PA, and TA, further for rapid and accurate measurement of VFA concentrations in biogas slurry.

Keywords: anaerobic digestion; volatile fatty acids; rapid determination; near infrared spectroscopy; partial least squares; genetic simulated annealing algorithm; competitive adaptive reweighted sampling