

# 拉曼光谱结合偏最小二乘的甲醇汽油甲醇含量快速定量分析方法研究

李茂刚<sup>1</sup>, 闫春华<sup>2</sup>, 杜瑶<sup>1</sup>, 张天龙<sup>2</sup>, 李华<sup>1,2\*</sup>

1. 西安石油大学化学化工学院, 陕西 西安 710065

2. 合成与天然功能分子教育部重点实验室, 西北大学化学与材料科学学院, 陕西 西安 710127

**摘要** 甲醇汽油是一种用以替代传统汽油的新型燃料, 其品质受到甲醇含量的严重影响。因此, 甲醇汽油中甲醇含量的快速分析对其品质把控具有深远意义。基于拉曼光谱(Raman)结合偏最小二乘(PLS)建立了一种甲醇汽油中甲醇含量快速定量分析方法。采用激光拉曼光谱仪对49组甲醇汽油样品的Raman光谱进行采集, 并进行光谱解析。比较了五种光谱预处理方法对甲醇汽油原始Raman光谱的预处理效果, 并采用变量重要性投影(VIP)对小波变换(WT)预处理后的甲醇汽油Raman光谱数据进行了特征变量提取。其次, 采用五折交叉验证(5-fold cross-validation (CV))对PLS校正模型的潜变量数目(LVs)及VIP阈值进行优化。在最优输入变量和模型参数下, 分别构建了基于不同输入变量的PLS模型。研究表明, 相较于原始光谱-偏最小二乘模型(RAW-PLS)和小波变换-偏最小二乘模型(WT-PLS), 变量重要性投影-偏最小二乘模型(VIP-PLS)可以获得更好的分析性能, 其预测集决定系数( $R_p^2$ )为0.9604, 均方根误差(RMSEP)为0.0341。因此, Raman光谱结合PLS是一种快速准确的甲醇汽油中甲醇含量分析方法。

**关键词** 甲醇汽油; 拉曼光谱; 变量重要性投影; 偏最小二乘

**中图分类号:** O657.37 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)07-2099-06

## 引言

甲醇汽油是一种新型清洁燃料, 相较于传统汽油, 有着价格低廉、尾气排放量低、高辛烷值等优势<sup>[1-2]</sup>。然而其理化性质受到其甲醇含量的影响, 例如, 含量过高会导致汽化现象加剧, 增加机动车部件的磨损; 含量过低无法体现出甲醇汽油的优良特性<sup>[3]</sup>。因此, 甲醇汽油中甲醇含量的快速定量分析对于甲醇汽油品质把控显得尤为重要。传统检测方法包括色谱、质谱等虽然可以实现对甲醇汽油品质的分析, 但受限于技术特点, 其检测分析时间长、样品处理复杂等均不利于甲醇汽油生产过程中在线及现场分析的应用<sup>[4-5]</sup>。近年来, 分子光谱技术在油品分析中被广泛报道, 尤以红外光谱(infrared spectroscopy, IR)最为突出<sup>[6-8]</sup>。Raman光谱技术也是一种分子光谱分析技术, 不同于IR技术, 其可用于样品分子中C=C, N—O, S—O, N—H, Si—O等化学键振动及转动等信息的检测, 从而实现对分子结构及含量等的表征<sup>[9]</sup>。相较于传统色谱、质谱等分析方法而言, Raman光谱技术用于样品分析时具有分析速度快, 无需复杂样品前处

理、可实时在线及现场分析等技术特点。现如今, Raman光谱技术应用于油品品质分析主要包括品质鉴别及定量分析两方面。张冰等<sup>[10]</sup>采用Raman光谱结合多输出最小二乘支持向量回归机(multioutput least squares support vector regression, MLS-SVR)算法对汽油中不同组分含量进行检测, 结果表明, MLS-SVR法相较于偏最小二乘(partial least squares, PLS)法对于汽油中烯烃、芳烃、苯、乙醇、甲醇等具有更好的检测精度, 其均方根误差(root mean square error of prediction set, RMSEP)分别为0.30%, 0.27%, 0.16%, 0.12%, 0.17%。Ardila等<sup>[11]</sup>采用Raman光谱技术结合化学计量学方法对汽油分类问题进行研究, 结果表明, 对于不同厂家的汽油样品, 所构建模型的交叉验证误差均小于5%。采用Raman光谱进行样品分析时, 会获得大量的光谱信息, 如果直接采用全谱进行模型构建, 会造成模型分析时间变长、性能变差。因此, 需先进行光谱预处理及特征提取, 再构建恰当的预测模型。目前, 光谱预处理方法有多元散射校正(multiple scattering calibration, MSC)、标准正态变换(standard normal variate, SNV)及小波变换(wavelet transform, WT)等, 特征变量提取方法包括无信息变量消除

收稿日期: 2020-06-24, 修订日期: 2020-10-30

基金项目: 国家自然科学基金项目(21873076, 21675123, 21605123)资助

作者简介: 李茂刚, 1995年生, 西安石油大学化学化工学院硕士研究生 e-mail: lmg1mg1995@163.com

\* 通讯作者 e-mail: huali@nwu.edu.cn

(uninformative variable elimination, UVE)、连续投影算法(successive projection algorithm, SPA)、变量重要性投影(variable importance in projection, VIP)等<sup>[12-14]</sup>。

本文基于 Raman 光谱结合 PLS 法对甲醇汽油中甲醇含量快速定量分析方法进行研究。利用激光拉曼光谱仪对不同含量的甲醇汽油样品进行 Raman 光谱采集,探究不同光谱预处理方法对 PLS 校正模型预测性能产生的影响。采用 Kennard-Stone(KS)算法以 2:1 的比例将甲醇汽油样品划分为校正集和预测集,基于此构建初始 PLS 校正模型,采用 5-fold CV 对模型的潜变量数目(latent variables, LVs)进行优化,采用 VIP 法对甲醇汽油样品的 Raman 光谱数据进行特征变量提取并对 VIP 阈值进行优化。最终,基于优化的输入变量及 LVs 构建基于 Raman 光谱的甲醇汽油中甲醇含量快速定量分析的 PLS 校正模型,并采用该 PLS 模型对预测集

样品中甲醇含量进行预测。

## 1 实验部分

### 1.1 甲醇汽油样品制备

制备了不同甲醇含量的甲醇汽油样品 49 组。样品制备时,将甲醇(分析纯, >99.7%,天津市富宇精细化工有限公司)与汽油(陕西西安某加油站,牌号为 98#)以不同的比例进行掺混,从而获得了含有不同甲醇含量的 49 组甲醇汽油样品。样品制备均在实验室通风橱环境进行,室温为 20 °C。由于甲醇与汽油可以很好的互溶,无需加入其他助溶剂来促进其溶解。将制备好的 49 组甲醇汽油样品置于 4 °C 环境下保存待测。表 1 所示为甲醇汽油样品制备时不同甲醇汽油样品中甲醇含量信息。

表 1 甲醇汽油样品中甲醇体积分数

Table 1 Methanol volume fraction in methanol gasoline samples

编号	体积分数/%	编号	体积分数/%	编号	体积分数/%	编号	体积分数/%	编号	体积分数/%
1	1.00	11	6.00	21	17.00	31	27.00	41	37.00
2	1.50	12*	8.00	22*	18.00	32*	28.00	42*	38.00
3*	2.00	13	9.00	23	19.00	33	29.00	43	39.00
4	2.50	14	10.00	24	20.00	34	30.00	44	40.00
5	3.00	15*	11.00	25*	21.00	35*	31.00	45*	41.00
6*	3.50	16	12.00	26	22.00	36	32.00	46	42.00
7	4.00	17	13.00	27	23.00	37	33.00	47	43.00
8	4.50	18*	14.00	28*	24.00	38*	34.00	48*	44.00
9*	5.00	19	15.00	29	25.00	39	35.00	49	45.00
10	5.50	20	16.00	30	26.00	40	36.00		

注:表 1 中编号带\*的样品在 PLS 校正模型构建时划分为预测集,其余样品为校正集。

Note: the samples with \* in Table 1 are divided into prediction set, and the others are calibration set.

### 1.2 Raman 光谱采集

使用一台激光拉曼光谱系统(QEpro6500 型,美国 Ocean Optics 公司)对甲醇汽油样品的 Raman 光谱进行采集。该系统由半导体激光器(785 nm)、光纤(RPB-Y)、光谱仪(QEpro6500)及光谱处理软件(OceanView)等部件组成。样品的 Raman 光谱采集范围选择为 0~2 000  $\text{cm}^{-1}$ ,光谱仪分辨率设置为 4  $\text{cm}^{-1}$ ,激光器功率设置成 300 mW,将甲醇汽油样品置于 5 mm 的比色皿中进行 Raman 光谱采集,检测环境温度 18 °C,以空气为背景检测并将其扣除。为了减少实验误差,光谱采集时,同一样品重复检测 25 次获得 25 条光谱,每 5 条光谱平均,同一样品获得 5 条 Raman 光谱,总计获得 245 条光谱数据。

### 1.3 Raman 光谱数据处理

为了消除原始光谱数据中实验误差、仪器误差、环境因素等对模型性能造成的影响,本文探究了五种光谱预处理方法(MSC, SNV, Derivative, Normalization 和 WT)对甲醇汽油 Raman 光谱的预处理效果;采用 VIP 对 Raman 光谱进行特征变量提取;采用 5-fold CV 对 PLS 校正模型的 LVs 和 VIP 阈值进行了优化。最终,基于优化的输入变量及 LVs 构建了甲醇汽油中甲醇含量快速定量分析的 PLS 校正模型,用

于预测集样品中甲醇含量分析。以决定系数(coefficient of determination,  $R^2$ )和均方根误差(root mean square error, RMSE)作为模型性能评价指标<sup>[12]</sup>。VIP 法应用于甲醇汽油 Raman 光谱特征变量筛选过程如下:

(1) 对甲醇汽油 Raman 光谱全谱的 VIP 重要性值进行计算;

(2) 采用不同的 VIP 阈值进行光谱特征变量提取,以筛选的特征变量作为输入变量构建 PLS 校正模型;

(3) 采用 5-fold CV 对 PLS 校正模型的预测性能进行计算;

(4) 改变阈值,重复步骤(2)至(3),直至 PLS 校正模型获得较为理想的预测性能。

## 2 结果与讨论

### 2.1 甲醇汽油 Raman 光谱解析

图 1 所示为甲醇、98# 汽油、49 组不同体积分数甲醇含量的甲醇汽油的 Raman 光谱图。甲醇的 Raman 光谱特征峰主要包括 1 031 和 1 448  $\text{cm}^{-1}$ ,98# 汽油的 Raman 光谱特征峰主要包括 723, 784, 1 001, 1 029 和 1 448  $\text{cm}^{-1}$ 等。甲醇汽

油的 Raman 光谱特征峰包括了甲醇( $1\ 031$  和  $1\ 448\ \text{cm}^{-1}$ )和汽油( $723$ ,  $784$  和  $1\ 029\ \text{cm}^{-1}$ )的特征峰,其中, $1\ 031\ \text{cm}^{-1}$ 是由 C—O 键的旋转和振动引起, $1\ 448\ \text{cm}^{-1}$ 是由甲基中 C—H 键的旋转和振动引起, $723\ \text{cm}^{-1}$ 是由汽油中甲基叔丁基醚中含有的叔碳基团的旋转和振动引起, $784\ \text{cm}^{-1}$ 是由汽油中环烷的旋转和振动引起的, $1\ 001\ \text{cm}^{-1}$ 是由汽油中甲苯的旋转和振动引起的, $1\ 029\ \text{cm}^{-1}$ 是由 C—O 键的旋转和振动引起的。此外,从图 1 中可以看出,随着甲醇汽油样品中甲醇含量的不断增加,甲醇汽油样品的 Raman 光谱呈现出一定的变化趋势,例如, $1\ 031$  和  $1\ 448\ \text{cm}^{-1}$ 的特征峰强度逐渐增大。由此可得,对甲醇汽油样品的 Raman 光谱进行分析,可以获得甲醇汽油中甲醇含量的相关信息。

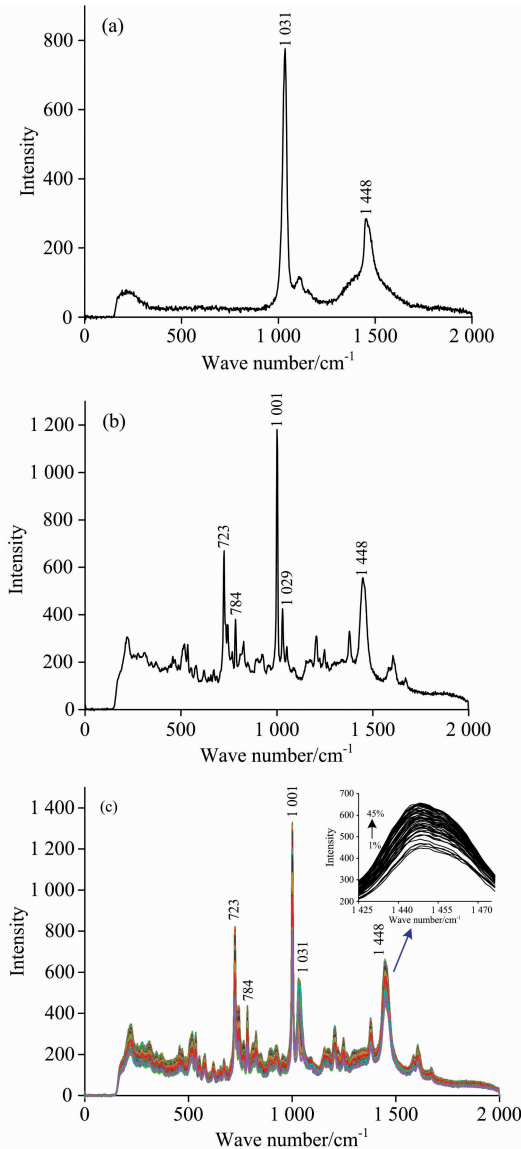


图 1 不同样品的 Raman 光谱图

(a): 甲醇; (b): 汽油; (c): 不同甲醇含量的甲醇汽油

Fig. 1 Raman spectra of different samples

(a): Methanol; (b): Gasoline;

(c): Methanol gasoline with different methanol content

## 2.2 甲醇汽油 Raman 光谱预处理方法的选择

从图 1 中可以看出,甲醇汽油的 Raman 光谱存在较为严重的基线漂移及杂散光等因素引起的光谱不平滑现象。因此,在校正模型构建前需对原始 Raman 光谱进行预处理。本文共比较了五种光谱预处理方法(MSC, SNV, Derivative, Normalization 以及 WT)对甲醇汽油 Raman 光谱的预处理效果。其中,使用 Derivative 法对甲醇汽油的 Raman 光谱进行预处理时,使用的是一阶导数法(D1st),探究了不同的平滑点数(3, 5, 7 和 9)对甲醇汽油的 Raman 光谱的预处理效果。结果表明,平滑点数为 7 时,模型预测性能最佳。使用 Normalization 法进行预处理时,具体用到的 Normalization 法为基于 Raman 光谱强度最大值的 Normalization 法。使用 WT 法进行预处理时,分别比较了基于不同小波分解层数(1, 2, 3, 4, 5, 6 和 7)与不同小波基函数(db1, db2, db3, db4 和 db5)的不同组合方式对甲醇汽油的 Raman 光谱的处理效果,结果表明,以小波基函数为 db5,小波分解层数为 6 的 WT 对甲醇汽油 Raman 光谱进行预处理,模型预测性能最佳。

表 2 所示为基于不同光谱预处理方法的 PLS 校正模型的预测性能。从表 2 中可以看出,使用原始光谱进行模型构建时,获得的校正集决定系数(coefficient of determination of cross-validation,  $R_{cv}^2$ )为 0.967 4,均方根误差(root mean square error of cross-validation, RMSECV)为 0.034 5,预测集决定系数(coefficient of determination of prediction set,  $R_p^2$ )为 0.936 2,均方根误差 RMSEP 为 0.041 3。采用 MSC, SNV, D1st, Normalization 以及 WT 法进行预处理后,PLS 校正模型校正集的  $R_{cv}^2$  和 RMSECV 基本变化不大。对于预测集, D1st 处理后光谱为输入变量的 PLS 校正模型的预测性能下降,其  $R_p^2$  从 0.936 2 下降为 0.906 8,其 RMSEP 从 0.041 3 变为 0.050 7。使用 MSC, SNV, Normalization 以及 WT 进行处理后,模型预测集的预测性能均有所提升,以 WT 预处理的光谱数据作为输入变量构建的 PLS 校正模型具有最好的预测性能,相较于原始光谱-偏最小二乘模型(raw spectrum-PLS, RAW-PLS)校正模型,其  $R_p^2$  从 0.936 2 上升为 0.960 4, RMSEP 从 0.041 3 下降为 0.034 1。因此,选择 WT 作为甲醇汽油原始 Raman 光谱数据的预处理方法。

表 2 基于不同光谱预处理方法的 PLS 校正模型预测性能比较  
Table 2 Comparison of prediction performance of PLS calibration models based on different spectral pretreatment methods

预处理方法	5-fold CV		预测集	
	$R_{cv}^2$	RMSECV/%	$R_p^2$	RMSEP/%
RAW	0.967 4	0.034 5	0.936 2	0.041 3
MSC	0.967 4	0.034 4	0.955 1	0.035 0
SNV	0.969 1	0.033 6	0.953 4	0.033 7
D1st	0.965 7	0.035 4	0.906 8	0.050 7
Normalization	0.958 7	0.038 8	0.955 2	0.033 3
WT	0.974 3	0.030 7	0.960 4	0.034 1

### 2.3 PLS 校正模型 LVs 的优化

LVs 是 PLS 校正模型的关键参数, 设置不合适则会引起模型产生“过拟合”或“欠拟合”, 这两种现象均对模型的预测性能具有负面影响。本文采用 5-fold CV 对 PLS 校正模型的 LVs 进行优化。图 2 所示为基于 5-fold CV 的 PLS 校正模型 LVs 优化。从图 2 可以看出, 当 LVs 设置为 2 时, 该模型可以获得最好的校正集预测性能, 其  $R_{CV}^2$  为 0.974 3, RMSECV 为 0.030 7。因此, 在 PLS 校正模型构建时, 将 LVs 设置为 2。

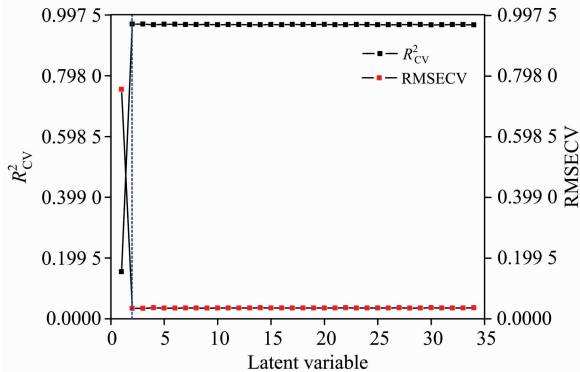


图 2 基于五折交叉验证的 PLS 校正模型预测性能随不同潜在变量的变化曲线

Fig. 2 Prediction performance curve of PLS calibration model based on 5-fold CV with different LVs

### 2.4 PLS 校正模型输入变量的优化

甲醇汽油原始 Raman 光谱中不仅存在噪声信息, 而且存在大量的冗余信息, 它们会对预测模型的性能带来负面影响。因此, 建模时不仅需对原始光谱进行预处理, 也要对其进行特征变量提取, 以确保模型的性能。本文采用 VIP 法对甲醇汽油 Raman 光谱数据特征变量进行提取, 并用 5-fold CV 对 VIP 阈值进行优化。

图 3 所示为采用 5-fold CV 进行 VIP 阈值优化的过程。从图 3 中可以看出, 随着 VIP 阈值的不断增大, 校正集的  $R_{CV}^2$  呈现出一种先缓慢增加, 而后逐渐降低的趋势, RMSECV 呈现出一种先缓慢降低, 而后逐渐增加的趋势, 预测集的  $R_p^2$  呈现出一种先缓慢增加, 而后迅速降低的趋势, RMSEP 呈现出一种先较小的波动, 而后迅速上升的趋势。结合 VIP 阈值对校正集和预测集预测性能的影响, 最终, 选择以 2.7 作为甲醇汽油 Raman 光谱数据的 VIP 变量重要性阈值对甲醇汽油 Raman 光谱数据的特征变量进行筛选。

以经过 VIP 特征变量筛选后的甲醇汽油 Raman 光谱数据作为输入变量构建 PLS 校正模型, 并对该模型进行 LVs 优化, 最终优化出的 LVs 为 2。基于优化后的 LVs 及甲醇汽油 Raman 光谱数据构建了用于甲醇汽油中甲醇含量快速定量分析的 PLS 校正模型, 该模型具有较好的预测性能, 其校正集的  $R_{CV}^2$  和 RMSECV 分别为 0.973 7 和 0.031 1, 预测集的  $R_p^2$  和 RMSEP 分别为 0.960 4 和 0.034 1。

### 2.5 基于不同输入变量的 PLS 校正模型预测性能对比

为了进一步验证基于甲醇汽油 Raman 光谱数据的 PLS

校正模型的预测性能, 分别构建了 RAW-PLS 校正模型、小波变换-偏最小二乘 (wavelet transform-PLS, WT-PLS) 校正模型以及变量重要性投影-偏最小二乘 (variable importance in projection-PLS, VIP-PLS) 校正模型。图 4 和表 3 所示为上述三种 PLS 校正模型预测性能的比较。结合图 4 和表 3 中可以看出相较于 RAW-PLS 校正模型, WT-PLS 校正模型和 VIP-PLS 校正模型的预测性能都得到了一定的提升。比较 WT-PLS 校正模型和 VIP-PLS 校正模型可以发现, 经过特征变量提取后, VIP-PLS 校正模型的预测性能产生了略微的下降, 在实际应用中基本可以忽略。在建模时间方面, VIP-PLS 校正模型的建模时间得到了很大的缩减, 因此, VIP-PLS 校正模型更适用于实际甲醇汽油中甲醇含量实时在线检

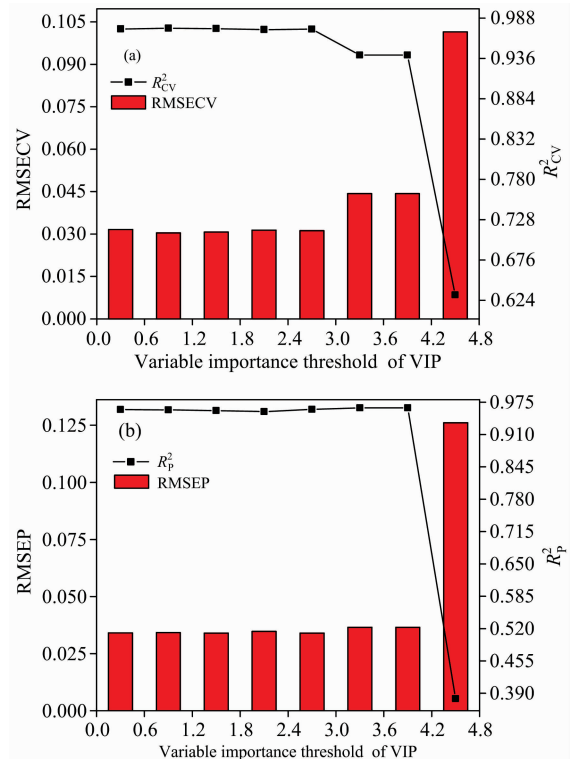
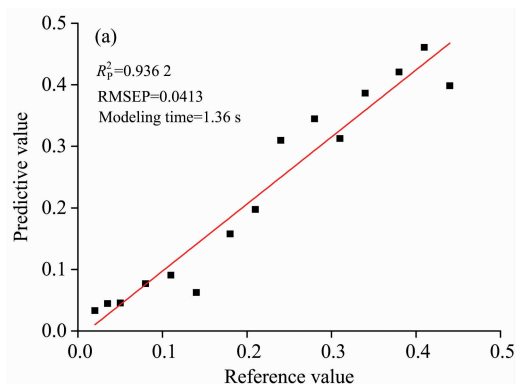


图 3 基于 PLS 校正模型的 VIP 变量重要性阈值优化图

(a): 校正集; (b): 预测集

Fig. 3 Optimization figure of VIP variable importance threshold based on PLS calibration model

(a): Calibration set; (b): Prediction set



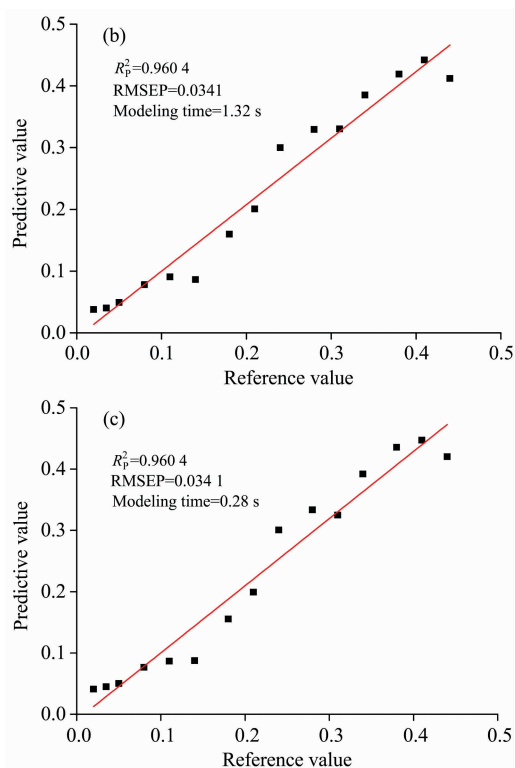


图 4 基于 Raman 光谱的三种 PLS 校正模型预测性能

(a): RAW-PLS; (b): WT-PLS; (c): VIP-PLS

Fig. 4 Prediction performance of three PLS calibration models based on Raman spectrum

(a): RAW-PLS; (b): WT-PLS; (c): VIP-PLS

测。相较于 RAW-PLS 校正模型, VIP-PLS 校正模型的建模时间从 1.36 s 缩短至 0.28 s, 变量个数从 1 044 个减少为 22

个, 校正集  $R_{CV}^2$  从 0.967 4 提升为 0.973 7, RMSECV 从 0.034 5 降低为 0.031 1, 预测集  $R_p^2$  从 0.936 2 上升为 0.960 4, RMSEP 从 0.041 3 降低为 0.034 1。

表 3 基于不同输入变量的 PLS 校正模型预测性能的比较  
Table 3 Comparison of prediction performance of the PLS calibration models based on different input variables

模型	建模时间/s	变量个数	5-fold CV		预测集	
			$R_{CV}^2$	RMSECV /%	$R_p^2$	RMSEP /%
RAW-PLS	1.36	1 044	0.967 4	0.034 5	0.936 2	0.041 3
WT-PLS	1.32	1 044	0.974 3	0.030 7	0.960 4	0.034 1
VIP-PLS	0.28	22	0.973 7	0.031 1	0.960 4	0.034 1

### 3 结论

建立了一种基于 Raman 光谱技术结合 PLS 法的甲醇汽油中甲醇含量快速定量分析方法。制备了 49 组不同甲醇含量的甲醇汽油样品并在优化的光谱采集条件下对其进行光谱采集及解析。比较了五种光谱预处理方法对原始 Raman 光谱预处理的效果, 并采用 VIP 对 WT 法预处理后的甲醇汽油 Raman 光谱数据进行了特征变量提取。采用 5-fold CV 对 PLS 校正模型的 LVs 和 VIP 阈值进行优化。在优化的输入变量和模型参数下, 分别构建了 RAW-PLS, WT-PLS 和 VIP-PLS 模型。研究表明, 相较于 RAW-PLS 和 WT-PLS 模型, VIP-PLS 模型可以获得更好的定量分析性能, 其建模时间为 0.28 s, 校正集  $R_{CV}^2$  为 0.973 7, RMSECV 为 0.031 1, 预测集  $R_p^2$  为 0.960 4, RMSEP 为 0.034 1。该方法具有检测耗时短、样品无损、结果准确等优势, 可以为甲醇汽油中甲醇含量的快速准确定量分析提供新思路与新方法。

### References

- [1] Wang C, Li Y, Xu C, et al. Fuel, 2019, 248: 76.
- [2] Vreugdenhil A J, Butler I S. Applied Spectroscopy, 1995, 49: 482.
- [3] Liu F, Hua Y, Wu H, et al. Energy & Fuels, 2018, 32(6): 6823.
- [4] Gotor R, Bell J, Rurack K. Journal of Materials Chemistry C, 2019, 7: 2250.
- [5] Avila L M, Pereira F D S A, Ignacio M D M D, et al. Fuel, 2018, 212: 236.
- [6] LI Mao-gang, YAN Chun-hua, XUE Jia, et al(李茂刚, 闫春华, 薛佳, 等). Chinese Journal of Analytical Chemistry(分析化学), 2019, 47(12): 1995.
- [7] Li J, Chu X. Energy & Fuels, 2018, 32(12): 12013.
- [8] Luan X, Jin M, Liu F. Applied Spectroscopy, 2018, 72(8): 1199.
- [9] Zheng J, Sun X, Du Y, et al. Journal of Materials Research, 2016, 31(15): 2302.
- [10] ZHANG Bing, DENG Zhi-yin, ZHENG Jing-kui, et al(张冰, 邓之银, 郑靖奎, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2015, 35(6): 1577.
- [11] Ardila J A, Soares F L F, Farias M A D S, et al. Analytical Letters, 2017, 50(7): 1126.
- [12] Farres M, Platikanov S, Tsakovski S, et al. Journal of Chemometrics, 2015, 29(10): 528.
- [13] Bian X, Li S, Lin L, et al. Analytica Chimica Acta, 2016, 925: 16.
- [14] Mahanty B. Chemometrics and Intelligent Laboratory Systems, 2018, 174: 45.

# Study on Rapid Quantitative Analysis Method of Methanol Content in Methanol Gasoline by Raman Spectroscopy and Partial Least Squares

LI Mao-gang<sup>1</sup>, YAN Chun-hua<sup>2</sup>, DU Yao<sup>1</sup>, ZHANG Tian-long<sup>2</sup>, LI Hua<sup>1, 2\*</sup>

1. College of Chemistry and Chemical Engineering, Xi'an Shiyou University, Xi'an 710065, China

2. Key Laboratory of Synthetic and Natural Functional Molecular of the Ministry of Education, College of Chemistry & Materials Science, Northwest University, Xi'an 710127, China

**Abstract** Methanol gasoline is a new fuel to replace traditional gasoline, and its quality is greatly affected by methanol content. Therefore, the rapid analysis and detection of methanol content in methanol gasoline will have far-reaching significance for its quality control. A rapid quantitative analysis method of methanol content in methanol gasoline based on Raman spectroscopy and partial least squares (PLS) was established in this work. Raman spectra of 49 methanol gasoline samples were collected by laser Raman spectrometer, and spectral analysis was carried out. The effects of five spectral pretreatment methods on the raw Raman spectra of methanol gasoline were compared. In addition, variable importance in projection (VIP) was used to extract the Raman spectra's feature variables preprocessed by wavelet transform (WT). The number of latent variables (LVs) and VIP threshold of the PLS calibration model was optimized by 5-fold cross-validation (CV). Under the optimal input variables and model parameters, PLS models based on different input variables were constructed. The results show that compared with RAW-PLS and WT-PLS, VIP-PLS can achieve better analysis performance, with the determination of the prediction set ( $R_p^2$ ) of 0.960 4 and root mean square error of prediction set (RMSEP) of 0.034 1. Therefore, Raman spectroscopy combined with PLS is a fast and accurate method for analysing methanol content in methanol gasoline.

**Keywords** Methanol gasoline; Raman spectrum; Variable importance in projection; Partial least squares

(Received Jun. 24, 2020; accepted Oct. 30, 2020)

\* Corresponding author