

文章编号 1004-924X(2026)12-1817-13

内标校准与多变量建模的煤气多组分拉曼检测

郭松杰¹, 李俊楷¹, 邱选兵^{1,4}, 田亚莉¹, 张其生^{1,2},
东占萃³, 李 伟³, 宋黎明³, 李传亮^{1*}

- (1. 太原科技大学 应用科学学院 山西省精密测量与在线检测装备工程研究中心,
山西 太原 030024;
2. 太原科技大学启东智能装备研究中心, 江苏 启东 226200;
3. 北京首钢股份有限公司迁安钢铁公司, 河北 迁安 064400;
4. 山西晋谱光电科技有限责任公司, 山西 太原 030008)

摘要:针对转炉、焦炉等复杂工况下煤气多组分(H_2 , CO , CH_4 等)在线检测中,拉曼光谱信号易受气压波动、激光功率漂移及光路传输效率变化等多源干扰影响,导致浓度反演精度下降的问题,本文提出一种基于内标校准与多变量联合建模的煤气多组分拉曼光谱分析方法。首先,以煤气中普遍存在且化学性质稳定的 CO_2 作为内标参考组分,构建内标比 R_i 以消除系统增益的共模干扰;其次,定义系统增益校准算子 S_{factor} ,实现对气压、散射光强等系统状态变量的实时表征;最后,将内标比、气压、散射光强及系统增益校准算子等参数共同作为输入特征,采用 LightGBM 算法建立多变量非线性回归模型,对 H_2 , CO , CH_4 三种组分浓度进行高精度反演。实验结果表明,所提出的多变量联合模型对 H_2 , CO , CH_4 的均方根误差(RMSE)均低于 0.09%,平均绝对百分比误差(MAPE)均低于 0.6%,与仅采用原始光谱数据或单一内标比的模型相比, RMSE 整体降幅约为 37%~63%。该方法有效补偿了气压与散射光强变化带来的非线性干扰,显著提升了复杂工况下煤气多组分拉曼光谱的浓度反演精度,为工业煤气在线检测提供了可行方案。

关键词:拉曼光谱;煤气;内标;多变量联合建模;机器学习

中图分类号: O433.1 **文献标识码:** A

doi: 10.37188/OPE.20263412.1817 **CSTR:** 32169.14.OPE.20263412.1817

Multicomponent Raman detection of coal gas based on internal standard calibration and multivariate modeling

GUO Songjie¹, LI Junkai¹, QIU Xuanbing^{1,4}, TIAN Yali¹, ZHANG Qisheng^{1,2},
DONG Zhancui³, LI Wei³, SONG Liming³, LI Chuanliang^{1*}

- (1. School of Applied Science, Taiyuan University of Science and Technology,
Shanxi Province Engineering Research Center of Precision Measurement and Online Detection
Equipment, Taiyuan 030024, China;
2. Qidong Intelligent Equipment Research Center at Taiyuan University of Science and Technology,
Qidong 226200, China;

收稿日期: 2026-04-22; **修订日期:** 2026-05-16.

基金项目: 国家自然科学基金(No. 52506204); 山西省重点研发计划(No. 202402150301012, No. 202302150101006);
中央引导地方科技发展资金项目(No. YDZJSX2025C026); 山西省科技创新人才团队专项资助
(No. 202304051001034); 太原科技大学科研启动基金(No. 20242148); 山西省科研启动金(No. 20252013)

3. Beijing Shougang Co., Ltd., Qian'an Iron and Steel Company, Qian'an 064400, China;

4. Shanxi Jinpu Optoelectronic Technology Co., Ltd., Taiyuan 030008, China)

* Corresponding author, E-mail: clli@tyust.edu.cn

Abstract: Raman spectroscopy enables multi-component gas analysis but faces challenges in complex industrial environments such as converter and coke oven gases. In these settings, online measurements of H_2 , CO , and CH_4 suffer from reduced accuracy due to multi-source interferences, including gas pressure fluctuations, laser power drift, and variations in optical path transmission efficiency. To address these issues, this paper proposed a Raman spectroscopic analysis method based on internal standard calibration and multivariate joint modeling. First, CO_2 , which was ubiquitous and chemically stable in the gas mixture, was used as an internal reference component to construct the internal standard ratio R_i , thereby eliminating common-mode interference in system gain. Second, a system gain calibration operator S_{factor} was defined to enable real-time characterization of system state variables, including pressure and scattered light intensity. Finally, several parameters were jointly used as input features, including the internal standard ratio, gas pressure, scattered light intensity, and the system gain calibration operator. A multivariate nonlinear regression model was then established using the LightGBM algorithm. This model enabled high-precision retrieval of H_2 , CO , and CH_4 concentrations. Experimental results demonstrate that the proposed multivariate joint model achieves a root mean square error (RMSE) of below 0.09% for each of the three components. The mean absolute percentage error (MAPE) for each component is below 0.6%. Compared to models using only raw spectral data or a single internal standard ratio, the RMSE of the proposed model is reduced by approximately 37% to 63%. This method effectively compensates for the nonlinear interferences caused by variations in gas pressure and scattered light intensity, significantly improving the accuracy of concentration retrieval for multi-component gases in complex industrial conditions using Raman spectroscopy. and provides a feasible solution for online industrial gas detection.

Key words: Raman spectroscopy; coal gas; internal standard; multivariate joint modeling; machine learning

1 引 言

在钢铁冶炼过程中,煤气作为重要的二次能源和工艺介质^[1],其可燃组分 H_2 、 CO 和 CH_4 直接决定热值水平与燃烧稳定性。其中, H_2 和 CO 是热值的主要贡献者, CH_4 虽含量较低,但对燃烧安全性与火焰稳定性影响显著。因此,实现这三组分浓度的高精度、实时在线测量,对优化燃烧控制、提升能效及降低碳排放具有重要意义。针对转炉、焦炉等复杂工况下煤气组分的剧烈波动,传统单组分或间歇式检测方法难以满足过程控制需求,亟须发展多组分同步、高鲁棒性的在线检测技术^[2]。

目前,用于工业气体多组分测量的光谱技术包括非分散红外光谱(Non-Dispersive Infrared,

NDIR)^[3]、傅里叶变换红外光谱(Fourier Transform Infrared, FTIR)^[4]、光声光谱(Photoacoustic Spectroscopy, PAS)^[5]、激光吸收光谱(Laser Absorption Spectroscopy, LAS)^[6-10]等。这些技术虽能有效测量 CO 、 CH_4 等多种组分,但受限于 H_2 作为同核双原子分子无红外吸收活性的特性,均无法直接检测 H_2 。拉曼光谱^[11-13]基于散射原理,从根本上突破了 H_2 的检测限制,且所有气体分子均具有拉曼活性,因此理论上可测量的气体种类极为广泛。同时,该技术无需样品预处理、响应速度快^[14],在工业过程气体在线分析中得到了广泛关注与应用。例如,万福等^[15]采用非对称腔镜线型F-P腔增强拉曼光谱,实现了电力变压器故障特征气体的高灵敏度检测;黄保坤等^[16]利用拉曼积分球光谱仪检测燃油发动机尾气中的多种气

体成分;陈沁楠等^[17]提出基于反谐振空芯光子晶体光纤增强的拉曼光谱方法,实现了锂离子电池充放电过程中多种特征气体的原位监测。然而,在实际工业环境中,拉曼散射信号除受待测组分浓度影响外,还受到气压、散射光强等多种状态变量的耦合作用,其波动会导致光谱发生系统性偏移,若不加以校正,将显著降低浓度反演的精度。

传统的拉曼气体定量方法通常仅基于光谱强度构建线性回归模型,并假设实验环境处于稳定状态,即气压、激光功率等参数保持恒定。例如,郑传新等^[18]开发了基于拉曼光谱的转炉烟气多组分实时在线分析系统,在未考虑压强和测试条件变化的影响下,对CO、CO₂、N₂、O₂、H₂五种组分的测量误差可控制在≤0.3%以内。然而,上述方法依赖于相对稳定的工况条件,在更复杂工业环境中的适用性仍有待验证。在转炉煤气、焦炉煤气等工业拉曼光谱测量中,为提升灵敏度常对气体池增压以增加分子数密度,但压强难以长期稳定:压缩泵与控制系统的精度有限,且煤气中的粉尘、焦油等杂质易污染管路和泵体,加剧压强波动。同时,激光器功率漂移、光学元件污染及光路偏移等因素也会导致散射光强发生不可控变化。为抑制环境变量对拉曼信号的干扰,内标法被广泛采用。通过引入浓度已知或稳定的参考组分,部分抵消激光功率和光路透过率波动。Xiang等^[19]以二硫化碳为内标,校正了采样管径、温度和激光功率等干扰,使四氯化钛和二氯甲烷混合物的预测误差分别为4.25%和2.59%。杜康等^[20]在葡萄糖溶液分析中以水的1 640 cm⁻¹峰为内标,发现特征峰强度法线性更好(相关系数0.992)、检测限更低。Wang等^[21]采用腔增强拉曼光谱并引入SF₆作为内标,实现了变压器油中多种溶解气体(CO、CO₂、H₂、CH₄等)的ppm级高精度定量分析。然而,传统内标法多用于液体样品,在气体研究中应用较少,主要原因是气体中引入内标可能干扰待测体系。此外,传统内标方法通常假设参考组分浓度已知或环境条件理想,但在气压波动较大时,压强展宽效应会使峰面积与分子数密度偏离线性关系,单靠内标难以完全消除由此带来的残余误差;同时,内标法对散射光强波动的补偿效果尚不明确,有

待验证。因此,亟须引入系统状态变量与非线性建模方法,以进一步提升复杂工况下的浓度反演精度。

针对上述问题,本文提出一种基于内标校准与多变量联合建模的煤气多组分拉曼光谱分析方法。该方法首先利用CO₂作为内标参考组分,同时,引入气压和散射光强作为系统状态变量,定义系统增益校准算子,实现对压强、激光功率、光学传输效率变化等系统增益的动态漂移实时表征。在此基础上,将内标比、气压、散射光强及系统增益校准算子等参数共同作为输入,采用LightGBM算法建立多变量非线性回归模型,对煤气中H₂、CO、CH₄三种组分的浓度进行高精度反演。本文所提方法能够有效补偿气压波动和散射光强变化带来的非线性干扰,显著提升浓度预测的准确性与鲁棒性,为复杂工况下工业煤气的在线拉曼检测提供了可行的技术方案。

2 理论基础

2.1 气体拉曼散射物理模型

作为一种非弹性散射光谱技术,拉曼光谱可同时获取多组分气体的特征指纹信息,具备快速响应、可在线运行等独特优势^[22]。根据拉曼散射原理,理想状态下探测器获取的特定气体组分*i*的拉曼信号强度(谱峰积分面积) A_i 可表述为:

$$A_i = K \cdot I_0 \cdot n_i \cdot \sigma_i \cdot Q, \quad (1)$$

其中: K 表示系统常数, I_0 表示激发光功率密度, n_i 表示分子数密度, σ_i 表示组分*i*的微分散射截面, Q 表示光路收集效率。

在实际煤气在线检测工况下, I_0 受激光器功率波动影响, Q 受光学传输效率影响,而 n_i 则是压强*P*、温度*T*及组分浓度*C_i*的函数。 A_i 发生变化,会直接引起组分浓度的测量误差。为了消除非受控环境变量对浓度反演的干扰,本文采用内标校准的方法,选取煤气中普遍存在、化学性质稳定且拉曼特征峰显著的CO₂作为内标参考组分。

2.2 基于内标比的特征构建

定义待测组分*i*相对于内标组分CO₂的内标比 R_i 为:

$$R_i = \frac{A_i \cdot C_{\text{CO}_2}}{A_{\text{CO}_2}}, \quad (2)$$

其中: A_i 表示待测组分特征峰的积分面积, A_{CO_2}

表示 CO_2 特征峰的积分面积, C_{CO_2} 表示 CO_2 体积浓度。内标特征中引入 CO_2 浓度, 旨在消除内标组分浓度变化的影响, 使内标特征在保留系统共模增益校准能力的同时, 与待测组分浓度建立更直接的对应关系。由于 CO_2 在 $1\,285\text{ cm}^{-1}$ 的特征峰峰形对称、便于精确拟合, 且与煤气中 H_2 , CO , CH_4 等组分的特征峰无光谱重叠, 抗干扰能力强, 因此选择该特征峰作为内标参考。

根据拉曼散射原理, 在系统增益恒定、压强与温度稳定的理想条件下, $R_i \propto C_i$, 在该关系中, 系统增益项 (如激光功率、光学收集效率及探测器响应等) 可以通过内标归一化被有效消除, 从而使 R_i 成为表征组分浓度的稳定特征量。然而, 实际工业环境中, 气压、激光功率、光学传输效率等波动会导致系统增益动态变化, 引入非线性误差, 单纯依赖 R_i 难以完全消除多源干扰。因此, 需进一步引入气压、散射光强等系统状态变量, 结合多变量非线性建模方法进行补偿, 以实现复杂工况下多组分气体浓度的高精度反演。

2.3 基于 CO_2 散射截面的系统校准算子

为了提取仪器的实时状态特征, 本文定义了系统增益校准算子 S_{factor} 。通过引入 CO_2 的峰面积 A_{CO_2} , 经压强 P 与散射光强 I 归一化得到:

$$S_{\text{factor}} = \frac{A_{\text{CO}_2}}{P \cdot I}. \quad (3)$$

该算子反映了在单位压强与单位散射光强下的 CO_2 信号强度, 能够有效表征系统增益动态漂移。因此, 本文将 R_i 作为核心特征量, 配合压强 P 、散射光强 I 及系统增益校准算子 S_{factor} 共同输入多变量联合模型, 利用机器学习捕捉二阶非线性残差, 从而在剧烈波动的工况下实现更高精度的浓度修正。多变量联合模型的特征输入如下:

$$X = [R_i, C_{\text{CO}_2}, N_i, P, I, S_{\text{factor}}], \quad (4)$$

其中: N_i 表示以待测组分的拉曼特征峰为中心, 选取特征峰两侧有限波长区间内的原始光谱点数据。基于以上理论, 采用内标校准与多变量联合建模的处理流程如图 1 所示。

在 $0 \sim 3\,000\text{ cm}^{-1}$ 拉曼位移范围内, H_2 , CO_2 , CO , CH_4 均具有特征明显的拉曼特征峰, 这些特征峰主要来源于分子的振动或振转跃迁, 可作为多组分定量分析的依据。本文所选拉曼位移及谱带归属如表 1 所示。

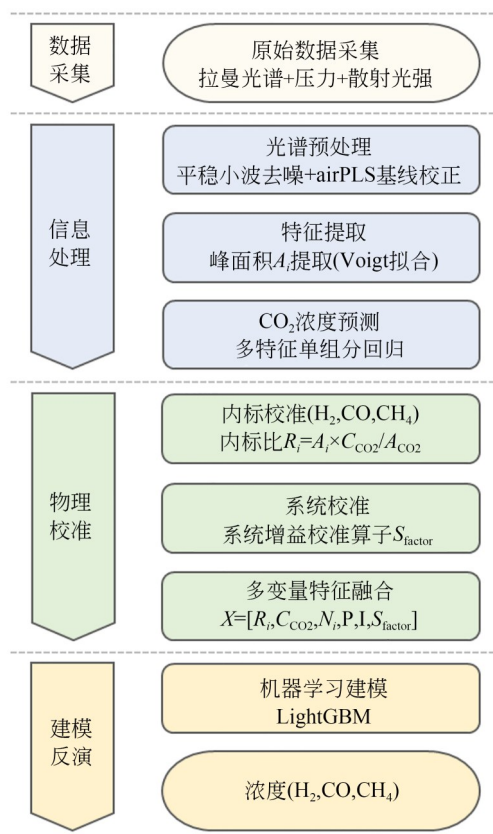


图 1 内标校准与多变量联合建模流程图

Fig. 1 Flowchart of internal standard calibration and multivariate joint modeling

表 1 煤气常见组分的拉曼位移及谱带归属

Tab. 1 Raman shifts and band assignments of common components in gas

Gas	Raman shift/ cm^{-1}	Type of spectral band
H_2	595	Pure rotation
CO_2	1 285	Bending vibration
CO_2	1 388	Fermi resonance peak
CO	2 143	Pure vibration
CH_4	2 917	Symmetric stretching vibration

3 实验

3.1 实验装置

实验装置如图 2 所示。该装置以不锈钢材质的气体池为核心, 配备进气、排气接口、激光入射与散射光收集窗口、散射光监测窗口及气压监测接口。用中心波长 532 nm 的激光器 (MGL-N-532-3W, 长春新产业) 作为激发光源; 拉曼探头 (RPB-532, 长春新产业) 内部集成独立传输的激

光发射与散射光接收光路。光谱仪(NOVA,复享光学)动态范围为15 000:1、信噪比为1 000:1,与光电探测器协同完成信号采集。自制光电探测器用于监测池内散射光强变化,计算机内置光谱预处理与浓度反演算法,可实时输出各组分浓度。

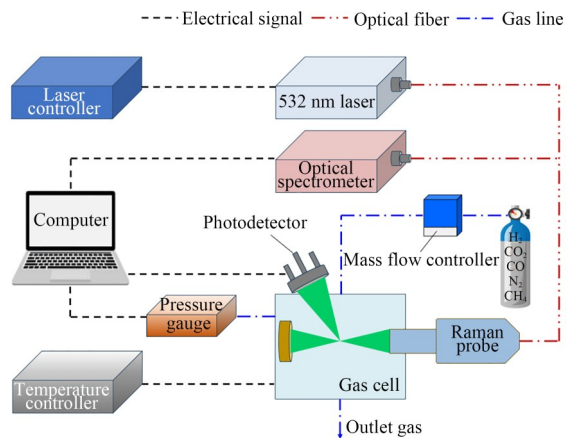


图2 实验装置图

Fig. 2 Schematic diagram of the experimental setup

3.2 标准混合气样本配置及实验条件设计

为模拟煤气在不同冶炼阶段的组分波动,实验室配置了五瓶具有代表性浓度比例的标准混合气体,其组分浓度真值如表2所示。

表2 标准混合气体组分配比

Tab. 2 Composition ratios of standard gas mixtures(%)

Gas	1	2	3	4	5
H ₂	5.28	10.09	14.71	20.21	25.00
CO ₂	2.99	6.00	9.04	11.99	15.01
CO	14.01	20.96	28.12	34.99	41.95
N ₂	4.96	10.04	15.00	22.95	5.94
CH ₄	4.01	6.02	8.06	9.86	12.10
Ar	68.75	47.90	25.08	0	0

为构建覆盖工业波动工况的多变量数据集,本实验采用正交组合策略调节系统变量。通过恒压控制系统(由质量流量控制器与压力计组成)设置2,2.5,3,3.5,4 atm共五个压强水平;通过调节激光器驱动电流,使输出功率分别稳定在1.6,1.8,2.0,2.2 W四个水平。在每个功率设定点下连续采集拉曼光谱,并同步记录光电探测

器测得的散射光强(本文中散射光强均指由光电探测器连续采集后计算得到的平均散射光强)。在每个“样气—压强—散射光强”工况节点下,记录当前压强与散射光强值,并由光谱仪同步采集拉曼光谱,光谱仪积分时间设置为5 s。实验中将气体池外壳温度维持在50 ℃恒定,主要目的是避免煤气中的水汽在光学窗口及气室内壁发生冷凝,保证窗口透过率与信号采集稳定性。同时,恒温环境可降低温度波动对分子数密度及拉曼微分散射截面的影响。

4 结果与讨论

4.1 光谱预处理

由于机械振动与环境干扰,原始拉曼光谱中存在一定随机噪声。为提升检测的准确性与灵敏度,本文采用平稳小波变换(Stationary Wavelet Transform, SWT)对原始光谱进行降噪处理。该方法在保留传统小波多分辨率分析能力的基础上,通过取消下采样操作实现平移不变性,可有效抑制阈值去噪后的伪Gibbs振荡,更好地保持拉曼特征峰的峰值强度与谱形特征。图3为原始与滤波后光谱的对比结果。在滤波处理中,具体选用db4小波基函数(Daubechies 4阶小波),分解层数设置为2层。去噪后,原始信号与去噪信号之间的残差标准偏差为24.55,表明平稳小波变换能够有效滤除高频随机噪声。可见该方法在显著抑制噪声的同时,较好地保留了拉曼特

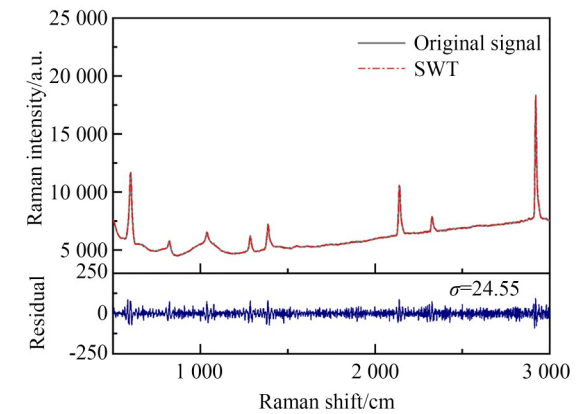


图3 原始光谱与平稳小波变换去噪后光谱的对比

Fig. 3 Comparison of raw spectrum and denoised spectrum using the stationary wavelet transform

征峰的主要结构特征。

激光器输出电流与功率的波动易引起拉曼光谱背景起伏,需在浓度反演前予以扣除,为此,本文采用基于拟合的基线校正策略。具体采用自适应迭代加权惩罚最小二乘法 (airPLS) 进行背景估计,结果如图 4 所示。鉴于光谱采集范围较宽且局部波段可能存在漂移,校正分为全局与局部两步:首先,采用 airPLS 对原始光谱进行全局基线拟合,设定较大平滑参数($\lambda=5\ 000$),通过迭代获得全谱初始基线,扣除后消除大尺度背景干扰;其次,在已校正光谱的 $725.92\sim 1\ 215.11\text{ cm}^{-1}$ 局部区域内,采用较小平滑参数($\lambda=2$)进行局部精细基线校正,以进一步改善全局基线拟合后残余背景漂移,提高整体光谱背景扣除效果。经两步校正后的光谱在有效消除低频背景的同时,较好地保持了拉曼特征峰的峰形与相对强度。

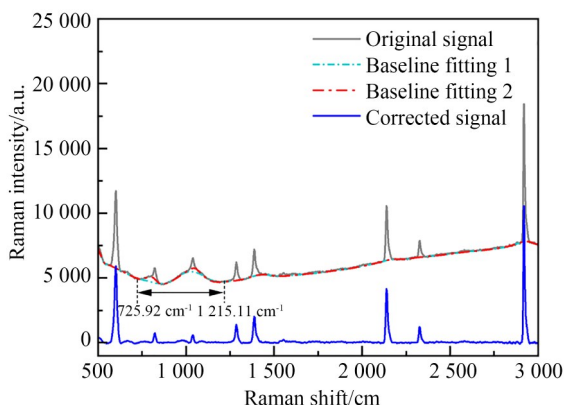


图 4 AirPLS 算法去除基线效果图

Fig. 4 Baseline removal effect diagram using the airPLS algorithm

4.2 CO₂ 浓度预测

鉴于内标法对 CO₂ 浓度的依赖性,本文在进行多组分浓度反演前,首先构建 CO₂ 浓度预测模型。不同于传统仅基于单一拉曼特征峰的定量方法,本文联合利用 CO₂ 在 $1\ 285\text{ cm}^{-1}$ 与 $1\ 388\text{ cm}^{-1}$ 处的两个特征峰,构建多特征输入模型: $X=[N_{1285}, N_{1388}, P, I]$, N_{1285} 和 N_{1388} 表示 CO₂ 两个特征峰的光谱数据。两个拉曼峰源自同一分子的不同振动模,对系统增益与浓度具有一致的响应特性,但受噪声、拟合误差及局部干扰的影响

程度各异。因此,多峰联合建模可引入多维度互补信息,有效降低随机误差对浓度反演的影响,提升模型的稳定性与鲁棒性。

在模型构建中,采用 LightGBM 算法^[23]对 CO₂ 浓度进行回归预测。该算法基于梯度提升决策树,具有训练速度快、善于捕捉非线性关系等优点,能有效应对压强展宽、光强波动及系统漂移引起的非线性问题。同时,其基于直方图的决策树策略与叶子优先生长方式降低了计算复杂度,适用于大规模数据快速训练,且对输入特征尺度不敏感,便于工程实现。实验共采集 5 种标准气、5 个压强水平、4 个激光功率水平,每个工况下获取 50 组光谱,总计 5 000 组样本。采集过程中实时监控光谱质量,剔除异常光谱(如强度突变、毛刺剧增)并立即补采,确保每个工况下 50 组光谱均为有效数据。样本按 4:1 比例划分为训练集与测试集。CO₂ 浓度预测结果如图 5 所示,模型表现出良好的预测能力,平均绝对误差 (Mean Absolute Error, MAE) 和均方根误差 (Root Mean Square Error, RMSE) 分别为 0.066 1 和 0.088 6,可满足作为内标参考的需求。

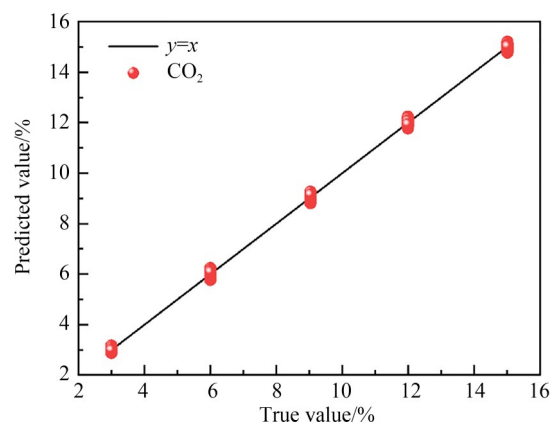


图 5 基于 LightGBM 算法的 CO₂ 浓度回归预测结果

Fig. 5 Regression prediction results of CO₂ concentration based on the LightGBM algorithm

4.3 特征面积计算

以样气 5 在 2 atm 气压、2 W 的激光功率条件下为例。提取 CO₂ ($1\ 285\text{ cm}^{-1}$) 作为内标特征,同时提取 H₂, CO 和 CH₄ 的拉曼特征峰面积。为提高积分面积计算准确度,使用 Voigt 线性函数对各特征峰进行拟合,如图 6 所示。Voigt 函数是

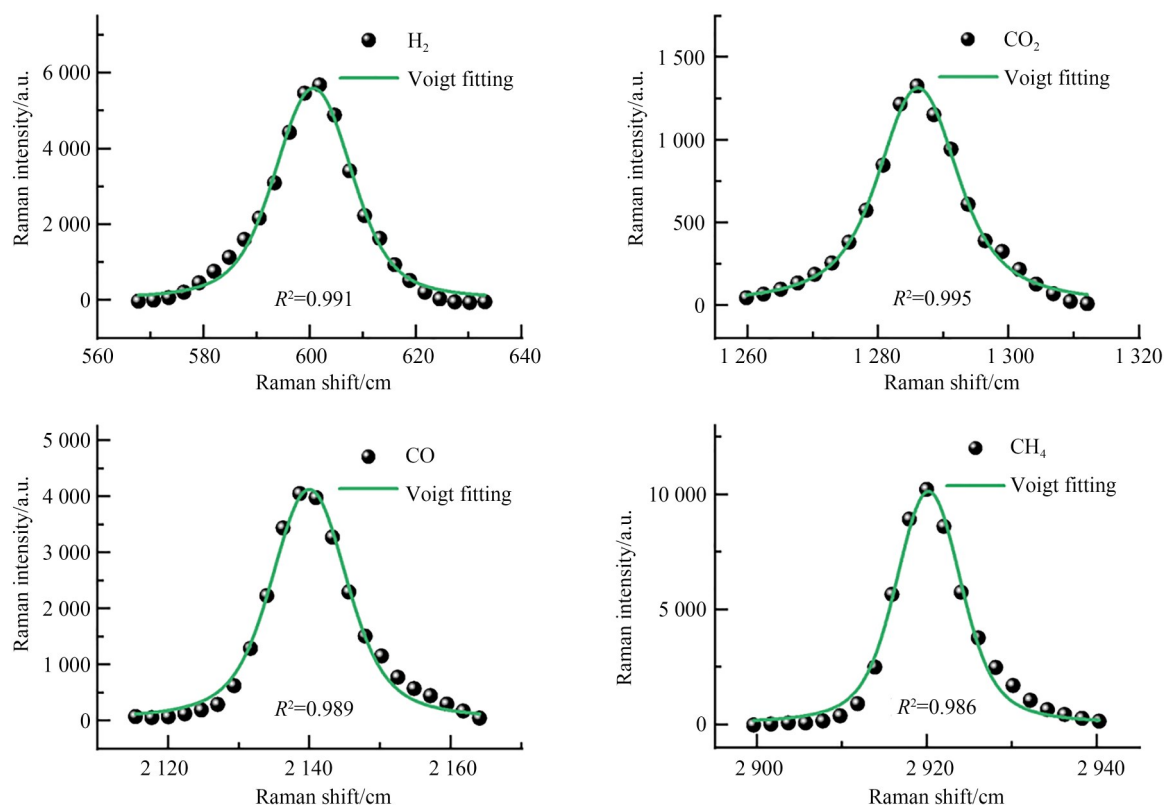


图 6 各组分拉曼特征峰的 Voigt 拟合结果

Fig. 6 Voigt fitting results of Raman characteristic peaks for each component

高斯函数与洛伦兹函数的卷积,能够较好地描述实际拉曼谱峰在压强展宽与仪器展宽共同作用下的线性特征。可以看出,五种组分的拟合决定系数 R^2 均大于 0.986,表明 Voigt 函数能够准确描述实验谱峰形状。

4.4 内标机制线性补偿能力验证与特征分析

为验证内标机制对气压与散射光强变化的一阶线性补偿能力,分别考察了特征峰面积与内标比 R_i 随气体池压强 P 和散射光强 I 的变化关系。持续向气体池中通入样气 5,将激光功率稳定在 2 W。调节压强控制系统,使气体池内压强分别稳定在 2,2.5,3,3.5,4 atm,记录四种组分 H_2 , CO_2 , CO , CH_4 的拉曼特征峰随气体池压强 P 的变化趋势,如图 7 所示。可以看出,各组分峰面积随压强增大而近似线性增加。根据拉曼散射理论,信号强度与分子数密度成正比,而分子数密度与压强呈线性关系。因此,实验结果与理论预期一致。

采用内标法对各组分进行归一化。图 8 展示

了内标比 R_i 随压强 P 的变化关系。与峰面积随压强明显变化不同,内标比 R_i 随压强的波动幅度显著减小。以 3 atm 为标准条件, H_2 , CO_2 , CO , CH_4 特征峰面积最大相对波动分别为 38.6%, 43.2%, 38.8%, 37.7%。经 CO_2 内标校准后,内标比 R_{H_2} , R_{CO} , R_{CH_4} 最大相对波动分别为:3.2%, 3.1%, 3.8%。可以看出,内标比的波动未完全消除,原因在于内标比仅消除了 CO_2 与待测组分共有的压强依赖关系,而高压条件下的压强展宽效应会导致峰面积与分子数密度之间出现微小非线性偏离,单纯依靠内标比难以完全补偿压强变化带来的非线性影响。

为进一步验证散射光强变化的影响,在恒温恒压条件下(气体池外壁 50 °C,内部压强 3 atm)向气体池持续通入样气,通过调节激光功率改变散射光强。设置四个激光功率水平(1.6, 1.8, 2.0, 2.2 W),对应散射光强分别为 1.522, 1.718, 1.916, 2.132 V。采集各功率下 H_2 , CO_2 , CO , CH_4 的拉曼特征峰信号。如图 9 所示,各组

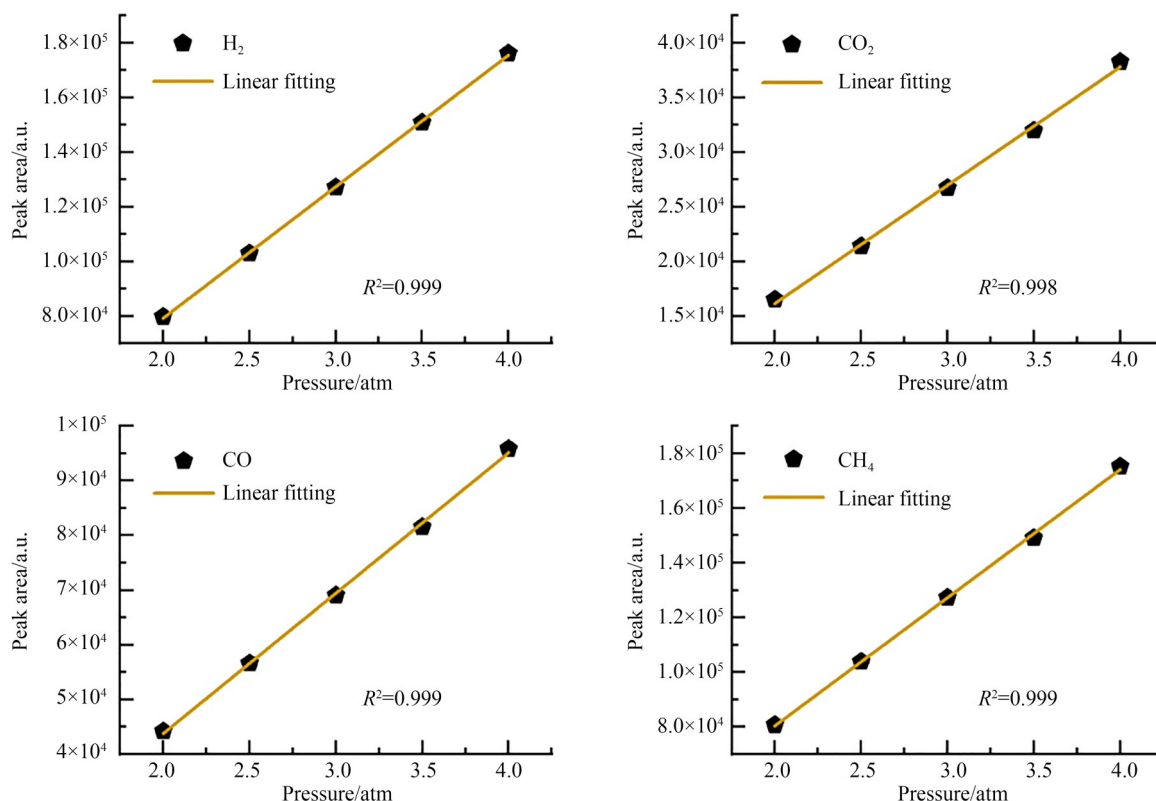


图7 不同气压下各组分的拉曼特征峰面积变化

Fig. 7 Variation of Raman characteristic peak areas of each component under different gas pressures

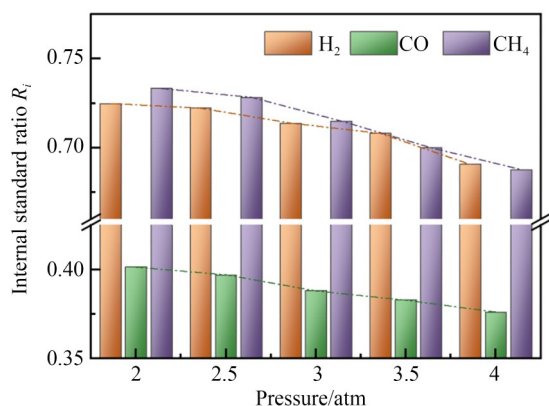


图8 不同气压下各组分的内标比变化

Fig. 8 Variation of internal standard ratios of each component under different gas pressures

分峰面积均随散射光强 I 增大而近似线性增大,表明拉曼信号对激光功率波动高度敏感。

为降低散射光强波动的影响,采用内标法对各组分进行归一化。图 10 展示了经 CO₂ 内标归一化后的内标比 R_i 随散射光强 I 的变化关系。可见,在相同压强与气体组分浓度下,各组分内标

比基本保持恒定,随 I 变化的波动幅度显著降低。以 1.916 V 为标准条件, H₂, CO₂, CO, CH₄ 特征峰面积的最大相对波动分别为 18.1%, 18.8%, 17.7%, 18.5%; 经 CO₂ 内标校准后, 内标比 R_{H_2} , R_{CO} , R_{CH_4} 的最大相对波动分别降至 1.7%, 1.6%, 1.1%。可以看出, 内标法可有效抵消激光功率波动对拉曼信号的一阶线性影响。

上述实验综合表明, 内标机制能够一定程度补偿散射光强波动与气体压强变化的一阶线性影响, 但其补偿能力有限, 最大残余误差仍达 3.8%。因此, 需进一步引入系统校准算子与多变量模型, 以捕捉内标法难以消除的系统误差及其他非线性干扰。

在验证内标机制对气压与散射光强一阶线性补偿能力的基础上, 进一步检验内标比 R_i 与组分浓度之间的线性定量关系。图 11 展示了内标比 R_i 与组分浓度之间的散点关系及拟合结果。三种组分的决定系数 R^2 均大于 0.997, 表明内标比与浓度具有良好的线性相关性, 为后续多变量模型提供了高质量的输入特征。

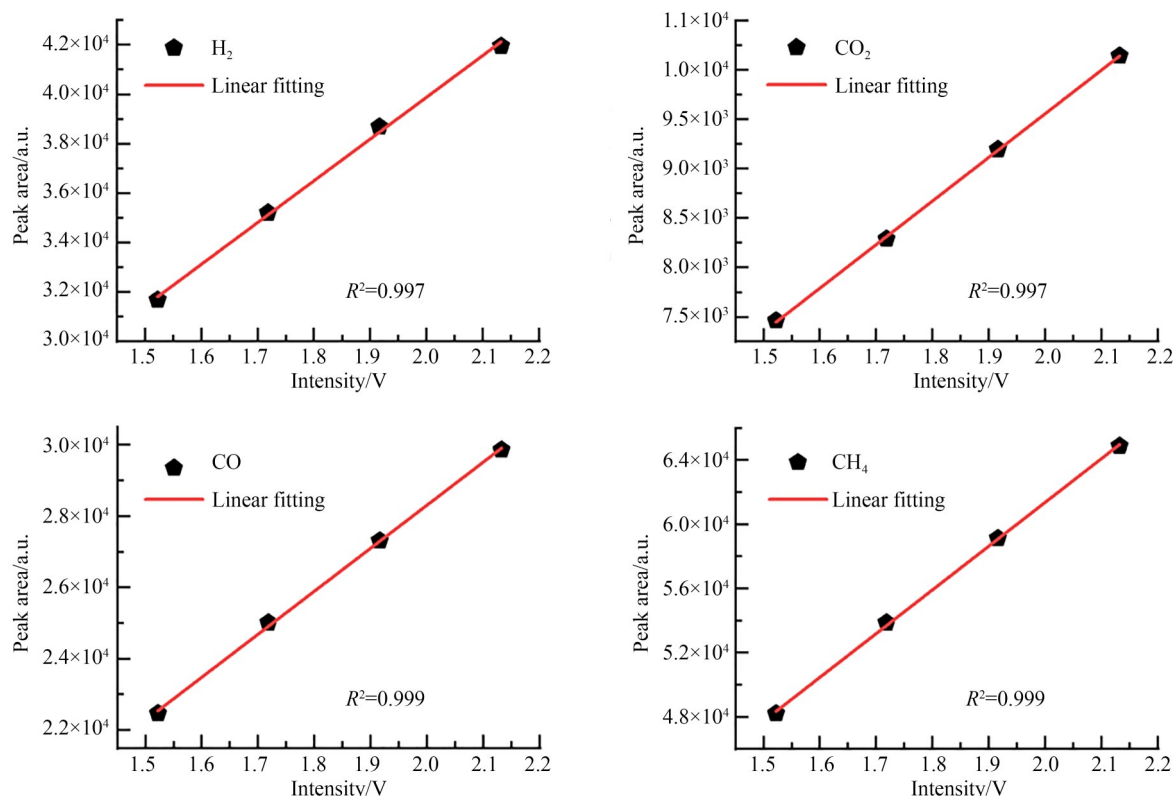


图9 不同散射光强下各组分的拉曼特征峰面积变化

Fig. 9 Variation of Raman characteristic peak areas of each component under different scattering light intensities

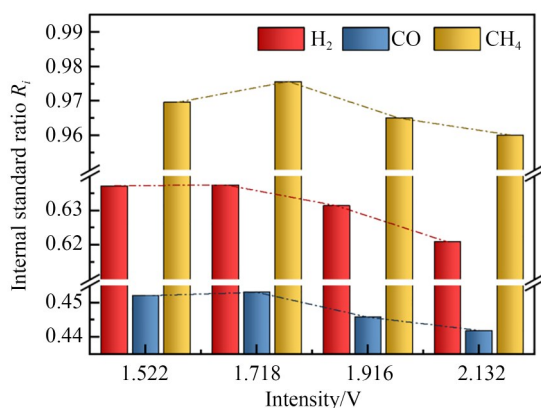


图10 不同散射光强下各组分的内标比变化

Fig. 10 Variation of internal standard ratios of each component under different scattering light intensities

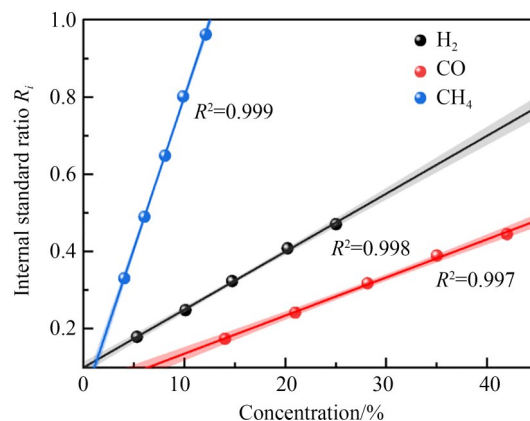


图11 内标比与组分浓度的线性拟合关系

Fig. 11 Linear fitting relationship between internal standard ratios and component concentrations

4.5 多变量模型性能评估

为进一步提取系统实时状态信息,按照公式(3)计算系统增益校准算子 S_{factor} 。并在此基础上构建多变量联合模型的输入特征向量(公式(4)),输出为各组分浓度 C_i 。基于LightGBM算法进行模型训练,采用与4.2节相同的样本。为

验证多变量联合建模方法的有效性,分别建立了四种对比模型:模型1采用特征峰光谱数据 N_i , P , I 作为输入;模型2采用内标 R_i , P , I 作为输入;模型3采用完整特征向量 X ;模型4采用完整特征向量 X ,结合 CO₂ 浓度真实值进行预测。四种

模型都采用 LightGBM 算法。模型 2、模型 3 中所用 CO_2 浓度均为 4.2 节模型的预测值,以模拟实际在线检测中 CO_2 未知的工况。各模型的测试集预测效果如图 12 所示,各组分浓度预测值和真

实值的 RMSE 如表 3 所示。

根据图 12 和表 3,对比模型 1 与模型 2,模型 2 在 H_2 、 CO 、 CH_4 三种组分上的 RMSE 分别降低了 18.5%、19.7%、21.5%。这表明,在原始特征

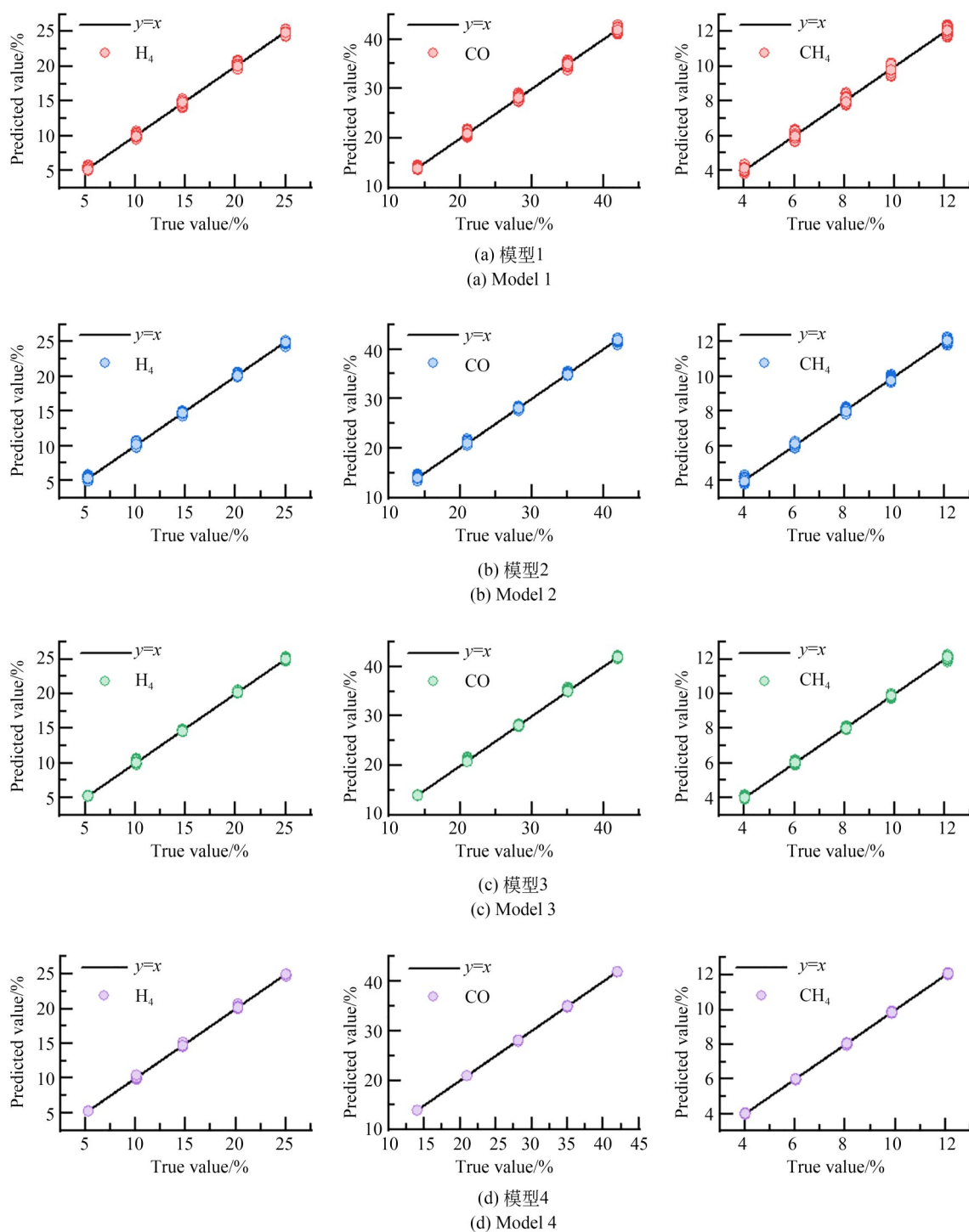


图 12 四种模型的浓度预测效果对比

Fig. 12 Comparison of concentration prediction performance of four models

表 3 各组分浓度预测值与真实值的 RMSE
Tab. 3 RMSE between predicted and true concentrations
of each component (%)

Model	H ₂	CO	CH ₄
Model 1	0.169 8	0.239 4	0.104 8
Model 2	0.138 4	0.192 2	0.082 3
Model 3	0.065 8	0.089 1	0.051 8
Model 4	0.037 2	0.050 2	0.033 1

峰数据 N_i 的基础上引入内标比 R_i , 能够有效消除压强、激光功率、光路传输效率变化等共模干扰, 显著提升浓度预测精度。对比模型 2 与模型 3, 模型 3 在 H_2 , CO , CH_4 上的 RMSE 分别进一步降低了 52.5%, 53.6%, 37.1%。这表明在内标的基础上, 引入系统增益校准算子 S_{factor} 等完整特征向量, 能够有效补偿内标法难以完全消除的压强展宽非线性效应及残余系统漂移。此外, 进一步计算了模型 3 测试集的平均绝对百分比误差 (Mean Absolute Percentage Error, MAPE), 对 H_2 , CO , CH_4 的 MAPE 分别为 0.455 9%, 0.325 7% 和 0.583 3%, 各组分相对误差均低于 0.6%。

为进一步分析 CO_2 预测误差对最终浓度反演结果的影响, 建立模型 4, 采用真实 CO_2 浓度替代预测值参与建模。结果表明, 采用真实 CO_2 浓度后, H_2 , CO 和 CH_4 的 RMSE 分别由 0.065 8%, 0.089 1% 和 0.051 8% 降低至 0.037 2%, 0.050 2% 和 0.033 1%。说明 CO_2 预测误差会通过内标特征和系统状态特征向后传播, 但两者差异整体较小, 表明所提方法对 CO_2 预测误差具有较好的鲁棒性。

5 结 论

本文针对复杂工况下煤气多组分拉曼光谱检测中气压波动、激光功率漂移及光路传输效率变化等多源干扰导致浓度反演精度下降的问题,

提出了一种基于内标校准与多变量联合建模的分析方法。实验结果表明, 内标机制对气压和散射光强波动具有一阶线性补偿能力, 经 CO_2 内标校准后, H_2 , CO , CH_4 的内标比最大相对波动分别由 38.6%, 38.8%, 37.7% 降至 3.2%, 3.1%, 3.8%, 残余误差为 3.8%, 验证了内标法可有效消除共模干扰但补偿能力有限。在此基础上, 所提出的多变量联合建模方法 (模型 3) 采用 Light-GBM 算法, 将内标比 R_i 、气压 P 、散射光强 I 及系统增益校准算子 S_{factor} 共同作为输入特征, 对 H_2 , CO , CH_4 三种组分的浓度反演 RMSE 分别为 0.065 8%, 0.089 1% 和 0.0518 %, 均低于 0.09%; MAPE 分别为 0.45 59%, 0.325 7% 和 0.583 3%, 均低于 0.6%。相比仅采用原始光谱数据 (模型 1), RMSE 分别降低 61.3%, 62.8% 和 50.6%; 相比仅采用内标比 (模型 2), RMSE 分别进一步降低 52.5%, 53.6% 和 37.1%。采用真实 CO_2 浓度对照结果表明, CO_2 预测误差虽会向后续浓度反演过程传递, 但模型 3 仍保持较高预测精度, 体现出较好的鲁棒性和适用性。综上, 本文所提方法有效补偿了气压波动与散射光强变化带来的非线性干扰, 显著提升了复杂工况下煤气多组分拉曼光谱的浓度反演精度, 为推动拉曼光谱技术在工业煤气在线检测中的实际应用提供了有力支撑。

作者贡献声明:

- 郭松杰: 测量方法提出与论文构思、撰写;
- 李俊楷: 测量实验设计与数据整理分析;
- 邱选兵: 实验数据分析与方法指导;
- 田亚莉: 实验数据分析与管理;
- 张其生: 数据结果可视化与论文指导;
- 东占萃: 论文审核与指导;
- 李 伟: 数据验证与模型优化;
- 宋黎明: 实验指导与数据分析;
- 李传亮: 论文审核与编辑写作。

参考文献:

[1] REN B L, WANG G, ZUO H B, *et al.* In-situ catalytic reforming of converter gas in converter flue based on thermochemical energy storage: Kinetics and numerical simulation [J]. *Journal of Energy Storage*, 2022, 48: 103693.

[2] 李翔, 王学谦, 李鹏飞, 等. 高炉煤气特征组分分析及其对脱硫过程的影响研究进展[J]. 化工进展, 2021, 40(12): 6629-6639.

LI X, WANG X Q, LI P F, *et al.* Progress on characteristic components analysis of blast furnace

- gas and its influence on desulfurization process [J]. *Chemical Industry and Engineering Progress*, 2021, 40(12): 6629-6639. (in Chinese)
- [3] ZHANG S S, WU H T, LIU S N, *et al.* Rotating multi-channel high-sensitivity integrating sphere gas sensing based on NDIR [J]. *Sensors and Actuators B: Chemical*, 2025, 428: 137242.
- [4] ZHU Q Q, WANG W, *et al.* Effects of biomass burning on CO, HCN, C₂H₆, C₂H₂ and H₂CO during long-term FTIR measurements in Hefei, China [J]. *Optics Express*, 2024, 32(5): 8343.
- [5] SHANG Z J, FAN Y S, *et al.* Dual-path excitation multi-quartz-enhanced photoacoustic spectroscopy [J]. *Analytical Chemistry*, 2026, 98(9): 7050-7056.
- [6] QIU X B, XIONG X H, LI L, *et al.* Development of an FDM-TDLAS sensor for long-term online monitoring of H₂S and CO to predict water wall corrosion trends [J]. *Analytical Chemistry*, 2025, 97(26): 14058-14066.
- [7] GUO S J, LI Z H, LIU Z M, *et al.* Simultaneous measurement of NH₃ and NO by mid-infrared tunable diode laser absorption spectroscopy based on machine-learning algorithms [J]. *Measurement*, 2024, 234: 114858.
- [8] 李龙, 师帅, 宫廷, 等. 燃煤锅炉高温腐蚀气体激光在线监测设备研究 [J]. *中国光学 (中英文)*, 2024, 17(5): 1060-1067.
- LI L, SHI S, GONG T, *et al.* Research on laser online monitoring equipment for high-temperature corrosive gas in coal-fired boilers [J]. *Chinese Optics*, 2024, 17(5): 1060-1067. (in Chinese)
- [9] 原康杰, 周月婷, 宫廷, 等. 基于腔衰荡的多谱线光谱分析 [J]. *中国光学 (中英文)*, 2025, 18(6): 1298-1306.
- YUAN K J, ZHOU Y T, GONG T, *et al.* Multi-spectral line spectral analysis based on cavity ring-down spectroscopy [J]. *Chinese Optics*, 2025, 18(6): 1298-1306. (in Chinese)
- [10] 杨天悦, 宫廷, 郭古青, 等. 氨气高精度激光光谱检测装置的设计及实现 [J]. *中国光学 (中英文)*, 2023, 16(5): 1129-1136.
- YANG T Y, GONG T, GUO G Q, *et al.* Design and achievement of a device for high-precision ammonia gas detection based on laser spectroscopy [J]. *Chinese Optics*, 2023, 16(5): 1129-1136. (in Chinese)
- [11] 谭雨翔, 徐玮辰, 李俊楷, 等. 焦炉煤气多组分拉曼光谱的 SPA-SVM 定量分析方法 [J]. *光电工程*, 2025, 52(9): 250189.
- TAN Y X, XU W C, LI J K, *et al.* SPA-SVM quantitative analysis method for multi component Raman spectroscopy of coke oven gas [J]. *Opto-Electronic Engineering*, 2025, 52(9): 250189. (in Chinese)
- [12] YAN X J, YANG S, KOU X L, *et al.* Optical-feedback cavity-enhanced Raman spectroscopy with continuous laser-to-cavity locking for multigas detection [J]. *Analytical Chemistry*, 2025, 97(21): 11119-11126.
- [13] 李尚锦, 丁腾龙, 张衍群, 等. 提升多次反射腔增强拉曼气体检测灵敏度 [J]. *光学精密工程*, 2025, 33(11): 1691-1699.
- LI S J, DING T L, ZHANG Y Q, *et al.* Enhancing sensitivity of multi-pass cavity enhanced Raman gas detection [J]. *Opt. Precision Eng.*, 2025, 33(11): 1691-1699. (in Chinese)
- [14] ZHU T Y, PEI Z Y, ZHAO X F, *et al.* Advanced SERS platform for uniform and reliable molecular detection [J]. *Analytical Chemistry*, 2024, 96(52): 20612-20621.
- [15] 万福, 白耀天, 王锐, 等. 变压器故障特征气体光反馈频率锁定非对称线型 F-P 腔增强拉曼光谱检测研究 [J]. *光子学报*, 2024, 53(11): 1130002.
- WAN F, BAI Y T, WANG R, *et al.* Transformer fault characteristic gas optical feedback frequency locking asymmetric linear F-P cavity enhanced Raman spectroscopy detection [J]. *Acta Photonica Sinica*, 2024, 53(11): 1130002. (in Chinese)
- [16] 黄保坤, 赵倩男, 刘叶繁, 等. 拉曼积分球原位检测燃油发动机尾气成分 [J]. *光谱学与光谱分析*, 2023, 43(10): 3310-3313.
- HUANG B K, ZHAO Q N, LIU Y F, *et al.* In situ detection of fuel engine exhaust components by Raman integrating sphere [J]. *Spectroscopy and Spectral Analysis*, 2023, 43(10): 3310-3313. (in Chinese)
- [17] 陈沁楠, 叶恩东, 王兆峰, 等. 锂电池产气空芯光纤增强拉曼光谱检测研究 [J]. *激光与光电子学进展*, 2025, 62(19): 375-385.
- CHEN Q N, YE E D, WANG Z F, *et al.* Gas production detection of lithium batteries based on hollow-core fiber-enhanced Raman spectroscopy [J]. *Laser & Optoelectronics Progress*, 2025, 62

- (19): 375-385. (in Chinese)
- [18] 郑传新, 刘晓萌, 李泉, 等. 基于拉曼光谱的转炉烟气中多组分气体在线分析方法研究[J]. 光谱学与光谱分析, 2025, 45(1): 146-151.
ZHENG C X, LIU X M, LI Q, *et al.* Online and multi-component analysis of converter flue gas by Raman spectroscopy [J]. *Spectroscopy and Spectral Analysis*, 2025, 45(1): 146-151. (in Chinese)
- [19] XIANG X Y, HE X D, XIA W T, *et al.* Quantitative determination of titanium tetrachloride and dichloromethane by Raman spectroscopy using carbon disulfide as an internal standard[J]. *Analytical Methods*, 2020, 12(7): 988-995.
- [20] 杜康, 花扬扬, 魏帅飞, 等. 基于内标法的低质量浓度葡萄糖溶液拉曼测定研究[J]. 光学学报(网络版), 2024, 1(5): 0512001.
DU K, HUA Y Y, WEI S F, *et al.* Raman determination of low mass concentration glucose solutions based on internal standard method [J]. *Acta Optica Sinica (Online)*, 2024, 1(5): 0512001. (in Chinese)
- [21] WANG P Y, CHEN W G, WANG J X, *et al.* Multigas analysis by cavity-enhanced Raman spectroscopy for power transformer diagnosis[J]. *Analytical Chemistry*, 2020, 92(8): 5969-5977.
- [22] WANG M L, WANG J X, WANG P Y, *et al.* Multi-pass cavity-enhanced Raman spectroscopy of complex natural gas components [J]. *Analytica Chimica Acta*, 2025, 1336: 343463.
- [23] LIANG J C, BU Y D, TAN K F, *et al.* Estimation of stellar atmospheric parameters with light gradient boosting machine algorithm and principal component analysis [J]. *The Astronomical Journal*, 2022, 163(4): 153.

作者简介:



郭松杰(1993—),男,山西长治人,博士,讲师,2024年于华南理工大学获得博士学位,现为太原科技大学应用科学学院讲师,主要从事拉曼光谱、激光吸收光谱等研究。E-mail: gsj@tyust.edu.cn

通讯作者:



李传亮(1983—),男,教授,博士生导师,2006年于聊城大学获得学士学位,2011年于华东师范大学获得博士学位,主要从事分子激光光谱学、高灵敏光电传感器等研究。E-mail: clli@tyust.edu.cn