

## 基于多参量的变压器油中特征气体拉曼光谱分析

陈新岗<sup>1,2</sup>, 冯煜轩<sup>1\*</sup>, 李昌鑫<sup>1</sup>, 陈姝婷<sup>1</sup>, 陈小青<sup>1</sup>, 龙尧<sup>1</sup>, 陈霖池<sup>3</sup>

1. 重庆理工大学电气与电子工程学院, 重庆 400054
2. 重庆市能源互联网工程技术研究中心, 重庆 400054
3. 重庆邮电大学, 重庆 400065

**摘要** 油中特征气体( $H_2$ ,  $CO$ ,  $CO_2$ ,  $CH_4$ ,  $C_2H_4$ ,  $C_2H_6$ ,  $C_2H_2$ )的快速准确检测是变压器在线监测的重要环节。激光拉曼光谱技术适用于特征气体的检测,能克服传统在线监测的诸多不足。在拉曼光谱图  $2\ 900 \sim 3\ 300\ cm^{-1}$  谱段,甲烷( $CH_4$ )和乙烷( $C_2H_6$ )气体特征谱峰聚集,研究此谱段中不同含量比的混合气体样本对变压器油中混合气体定量分析具有重要意义。在单一特征气体拉曼光谱检测的研究基础上,选取预处理后光谱图中特征峰的谱峰高度、半高宽以及谱峰面积多个参量作为特征因素,对变压器油中混合气体进行定量分析。以二阶微扰理论分析计算得出,甲烷拉曼光谱中存在四个特征谱峰,选取的谱段中包含以  $3\ 111$  与  $3\ 284\ cm^{-1}$  为拉曼频移中心的两峰,乙烷拉曼光谱中存在六个特征谱峰,选取谱段中存在  $3\ 111$  与  $3\ 187\ cm^{-1}$  两峰,理论上通过谱段中携带的特征谱峰信息能够计算两种气体含量;通过拉曼光谱平台检测,混合气体光谱图特征谱峰会产生平移以及聚合,对光谱图中寻峰得到的中心频移为  $2\ 902$ ,  $2\ 918$ ,  $2\ 956$  和  $3\ 022\ cm^{-1}$  的四个混合峰建立高斯函数模型,得到特征谱峰的谱峰高度、半高宽以及谱峰面积;建立偏最小二乘回归模型,以谱峰高度、半高宽、谱峰面积为自变量,两种气体含量为因变量计算分析。模型潜在因子取到  $t_6$  时,调整后的  $R^2$  为  $0.993$ ,表明自变量与因变量具有确切关系,回归模型可靠。对回归方程参数分析发现,谱峰半高宽相比谱峰面积以及谱峰高度有显著贡献,符合预期目标,混合气体光谱图中四个特征谱峰对两种气体均有影响。通过实验可总结得出,针对甲烷乙烷混合气体,在室温  $25\ ^\circ C$ ,积分时间  $15\ s$ ,积分次数  $2$ ,狭缝  $100\ \mu m$  条件下,通过获取谱峰高度、谱峰面积以及半高宽三个参量,能够准确测量气体含量,为变压器油中多种特征气体的同时检测奠定了基础。

**关键词** 拉曼光谱; 变压器; 特征气体; 定量分析; 多参量

**中图分类号:** O657.37 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)06-1916-07

### 引言

电力变压器作为电能传输和电压转换的枢纽设备,在电力系统网络中扮演着极其重要的角色,因此对其进行故障监测具有重大意义。其中,油浸式电力变压器由其良好的散热、绝缘性能和较大的传输容量在电力行业运用十分广泛。当油浸式变压器内部出现过热或局部放电等故障时,油中会分解产生以七种气体( $H_2$ ,  $CO$ ,  $CO_2$ ,  $CH_4$ ,  $C_2H_2$ ,  $C_2H_4$  和  $C_2H_6$ )为主的特征气体<sup>[1]</sup>,通过对气体组分和含量的分析,可以对变压器进行运行状态诊断以实现故障前缺陷修复和配合保护动作,保障变压器的安全运行。

气相色谱法(gas chromatography, GC)是一种较为典型的油中特征气体检测方法,能同时实现油中多种特征气体浓度的检测。但是需要对样本进行较为复杂的预处理且色谱柱性能会随着时间的增长而下降,使得其在变压器状态在线监测的应用上受到了限制<sup>[2]</sup>。相比于气相色谱法,拉曼光谱技术可以在无需预处理的情况下通过对少量样本的检测获取到可用的光谱数据,对样本的无损检测使其在变压器运行状态在线诊断中具有极大的优势。近年来,已有学者对油中气体拉曼信息进行了研究,确定了各种特征气体的谱峰归属<sup>[3]</sup>,但在多种气体混合频谱的分析上依旧存在许多困难。

本文基于变压器油中特征气体拉曼光谱检测实验平台,完成了七种混合特征气体样本的光谱数据采集。以全谱段中

收稿日期: 2019-07-24, 修订日期: 2019-11-08

基金项目: 重庆市教委科学技术研究项目(KJ1400917, KJ1709225), 重庆理工大学研究生创新项目(YCX20192057)资助

作者简介: 陈新岗, 1968年生, 重庆理工大学电气与电子工程学院教授 e-mail: chenxingang@cqut.edu.cn

\* 通讯联系人 e-mail: 568598953@qq.com

CH<sub>4</sub> 和 C<sub>2</sub>H<sub>6</sub> 所在的部分谱段作为研究对象, 提出定点寻峰<sup>[4]</sup>的方法, 并结合多原子特征气体分子拉曼理论, 进行了光谱数据特征量提取研究。考虑变压器油中特征气体含量小, 使用偏最小二乘法 (partial least squares, PLS), 建立了以混合特征气体多种拉曼光谱特征量为基础的气体定量分析模型<sup>[5]</sup>, 并验证了小测试样本数量下该模型气体定量分析的准确性和有效性。

## 1 实验部分

变压器油中特征气体拉曼光谱检测实验平台如图 1, 主要包括气路系统、光路系统和数据存储系统。气路系统中气体样品检测池由钢材料制成, 设有进行抽气、排气和充气的三个口, 用于气体样本的有效载入; 光路系统主要由各透镜、反射镜和滤波片组成, 用于形成检测光和散射光传递的通路; 数据存储系统中采用英国 Andor 公司 Solis 光谱分析软件实现光谱数据的简单存储和分析。低噪声 DPSS 激光器 (CoboltSamba™ 100, 100 mW) 产生波长为 532 nm 的单体激光通过载有气体样本的检测池, 在该处发生拉曼散射。散射光信号进入光路系统通过背照式深耗尽型 CCD 探测器 (Andor, DU416-LDC-DD) 转换成电信号并在计算机中完成光谱数据的实时采集与存储。

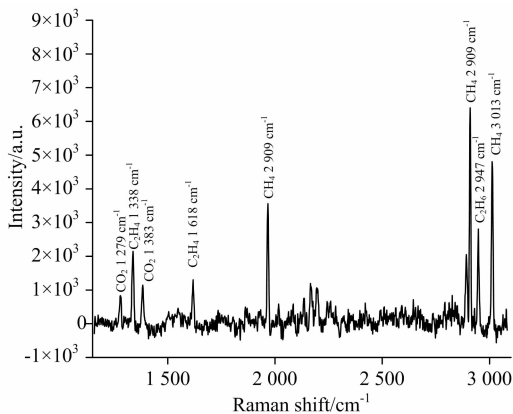


图 1 拉曼光谱检测实验平台结构示意图

Fig. 1 Structural sketch of Raman spectroscopy detection experimental platform

为了避免高温引起过高的暗电流和阅读噪声以提高 CCD 探测器的灵敏度, 检测前将其内部工作温度降至零下 60 °C<sup>[6]</sup>, 设计光路并在实验台上固定各器件位置。恒定激光功率为 30 mW, 实验室环境温度为 25 °C; 设置数据采集积分时间为 15 s, 积分次数为 2, 选择 100 μm 狭缝<sup>[7]</sup>; 每次检测前先进行洗气操作, 再将检测池通过真空泵抽成真空状态, 然后通过充气口充入配比好的气体样本, 最后再通过洗气操作完成检测池的上一次气体样本清理。

在同一环境和实验条件下, 记录每次充入样本后检测池的压强并完成样本的拉曼光谱实验数据测量。样本检测池体积固定, 实验外部环境温度保持恒定, 根据理想气体状态方程

$$PV = nRT \quad (1)$$

可知, 当理想气体体积  $V$ , 热力学温度  $T$  保持不变,  $R$  为理想气体常数, 气体的压强  $P$  与气体物质的量  $n$  成正比关系, 即实验过程中记录的检测池的压强与样本气体含量成正比关系, 关系式为

$$N = aP (a = V/RT) \quad (2)$$

## 2 光谱数据信息分析

### 2.1 气体的仿真计算及制备

按照一定比例制备七种特征气体混合样本并充入真空状态下的检测池, 完成检测并获取得到混合气体拉曼光谱。如图 2 所示, 七种混合特征气体拉曼光谱 1 200~3 030 cm<sup>-1</sup> 谱段范围内 CO<sub>2</sub>, CH<sub>4</sub>, C<sub>2</sub>H<sub>4</sub> 和 C<sub>2</sub>H<sub>6</sub> 都检测到两个可识别拉曼谱峰。其中 CH<sub>4</sub> 和 C<sub>2</sub>H<sub>6</sub> 在 2 900 cm<sup>-1</sup> 频移附近位置出现谱峰重叠现象, 因此选取 CH<sub>4</sub> 和 C<sub>2</sub>H<sub>6</sub> 混合气体样本作为研究对象, 研究多种混合特征气体拉曼频移数据处理方法, 在混合特征气体拉曼频移测定中具有代表性。

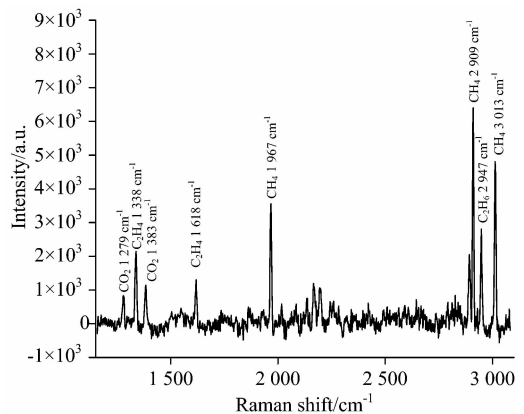


图 2 七种混合特征气体实测拉曼光谱图

Fig. 2 Raman spectrograms of seven mixed characteristic gases

基于二阶微扰理论 (second-order moller-pletset perturbation theory, MP2) 对七种特征气体分子进行结构建模优化和拉曼频率计算, 分析可知多原子特征气体分子具有多个不同的拉曼振动模式<sup>[8]</sup>, 且不同模式对应不同拉曼频移并携带不同的物理信息。CH<sub>4</sub> 和 C<sub>2</sub>H<sub>6</sub> 基于 MP2 在 6-31G(d) 基组下的分子优化构型如图 3。

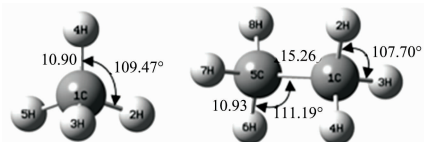


图 3 CH<sub>4</sub> 和 C<sub>2</sub>H<sub>6</sub> 优化后分子结构图

Fig. 3 Optimized molecular structure of CH<sub>4</sub> and C<sub>2</sub>H<sub>6</sub>

CH<sub>4</sub> 一共具有 9 种拉曼振动模式, 且其中 4 种振动模式具有拉曼活性<sup>[9]</sup>; C<sub>2</sub>H<sub>6</sub> 具有 18 种振动模式, 6 种振动模式具有拉曼活性, 其拉曼频率计算结果如表 1。结果为研究两种混合气体的拉曼振动特性特征谱峰的选取提供了理论依据。

选择纯度均为 99.99% 的  $\text{CH}_4$  和  $\text{C}_2\text{H}_6$  气体, 在同一环境和实验条件下, 按不同含量比例制备混合气体样本并进行测量, 制备 17 组不同浓度的混合气体。配制气体含量如表 2。

表 1  $\text{CH}_4$  和  $\text{C}_2\text{H}_6$  的仿真计算结果

分子名称	计算拉曼频移/ $\text{cm}^{-1}$	拉曼活性/ $(\text{A}4 \cdot \text{AMU}^{-1})$
$\text{CH}_4$	1 413.56	3.024 3
	1 624.74	32.033 5
	3 111.33	110.172 5
	3 284.47	19.679 6
$\text{C}_2\text{H}_6$	1 049.43	9.338 4
	1 270.65	3.553 8
	1 491.79	6.635 7
	1 568.64	38.243 3
	3 111.29	188.651 6
	3 187.77	120.104 7

表 2 甲烷乙烷配置表

序号	甲烷含量/MPa	乙烷含量/MPa
1	0.05	0.05
2	0.1	0.05
3	0.15	0.05
4	0.2	0.05
5	0.25	0.05
6	0.13	0.05
7	0.13	0.1
8	0.13	0.15
9	0.13	0.2
10	0.13	0.25
11	0.05	0.08
12	0.1	0.08
13	0.15	0.08
14	0.2	0.08
15	0.25	0.08
16	0.13	0
17	0	0.08

控制一种气体量不变, 另一种气体逐渐增加。建模计算需选取四个特征谱峰中谱峰高度、半高宽和谱峰面积多参量作为自变量, 为了避免气体含量计算模型发生局部最优解, 影响检测准确性, 至少 12 组气体作为训练样本; 设计两组单一气体, 既方便拉曼光谱预处理中基线的提取, 亦便于单一气体与混合气体光谱图的差异比较。配置 17 组气体, 先以 12 组混合气体作为训练样本, 对建立的模型参数进行分析, 若模型评估不通过, 依次增加训练样本, 直至得到准确模型。

实际检测得到的两种单一气体的光谱图如图 4, 两种单一气体的特征峰拉曼频移与仿真计算结果相差不大, 在建模计算时应以实际光谱图数据为准。

## 2.2 光谱数据预处理

实验拉曼光谱中存在激光发射噪声和暗电流噪声等信

号, 拉曼信号容易被噪声信号所淹没, 因此需要对所获得的拉曼光谱进行去噪处理, 以提高光谱的信噪比。为了减小拉曼光谱数据在处理过程中小峰丢失和峰位位移的数据失真, 本文基于式(3)采用移动平滑的方法<sup>[10]</sup>进行去噪处理。

$$S'(n) = \frac{\sum_{i=-m}^m S(n+i)}{2m+1} \quad (3)$$

式(3)中,  $S$  为原始光谱,  $S'$  为平滑后的光谱,  $2m+1$  为移动数据窗口(此处数据窗口设置为 5); 采用宽度为  $2m+1$  的窗口沿着光谱数据方向移动, 每个窗口具有基数个数据点<sup>[11]</sup>, 通过计算每个窗口中  $2m+1$  个点的数据平均值代替原中心点  $n$  位置处的数据值以实现去噪运算。

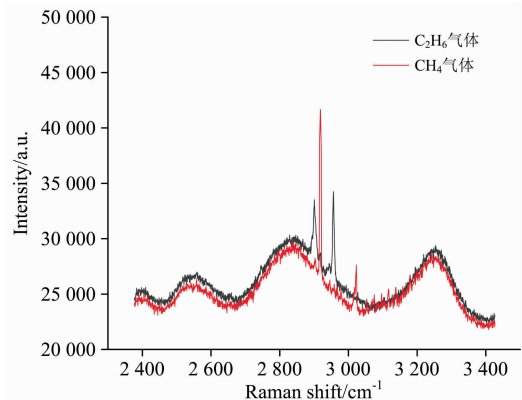


图 4 两种单一气体的光谱图

Fig. 4 Raman Spectra of  $\text{CH}_4$  and  $\text{C}_2\text{H}_6$

由于光谱中还存在荧光背景, 且拉曼检测装置在多次检测过程中, 光谱图的基线也会发生一定起伏<sup>[12]</sup>, 因此还需要进行基线扣除。根据文献<sup>[13]</sup>研究, 拉曼光谱数据图信息的提取主要集中在特征谱峰信息<sup>[14]</sup>, 对光谱数据进行基线扣除操作过程中, 考虑的重点是将不同光谱图同一频移位置的谱峰位于同一基准。不同混合比的气体光谱图中基线存在差异, 若对每张光谱图采用多项式取基线, 工作量大, 耗时多, 特征谱峰处基线差异明显。本文分别将  $\text{CH}_4$  和  $\text{C}_2\text{H}_6$  单一气体的拉曼光谱图基线作为基础, 对两基线求平均数, 得到混合气体的“中位线基线”。混合气体光谱图中特征谱峰不是简单的两种气体光谱图叠加, 但使用“中位线基线”能有效地降低四个特征谱峰处的总体误差; 对于荧光背景的消除效果良

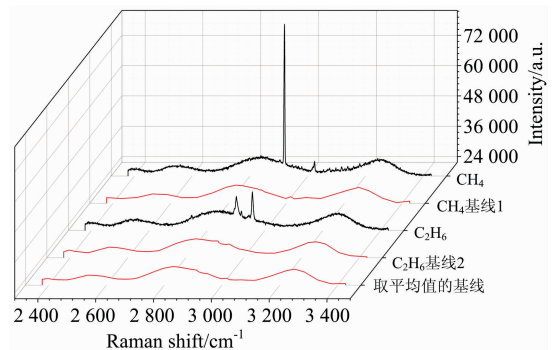


图 5 单一气体的基线以及最终基线光谱图

Fig. 5 Baseline of a single gas and final baseline

好；测试仪器问题而产生的基线漂移得到良好消除。得到两种单一气体基线以及取平均数得到的最终基线如图 5。

预处理后混合气体实验拉曼光谱如图 6，一种颜色代表一组混合气体，经过预处理后的光谱图差异集中在四个特征谱峰上，达到预期效果。

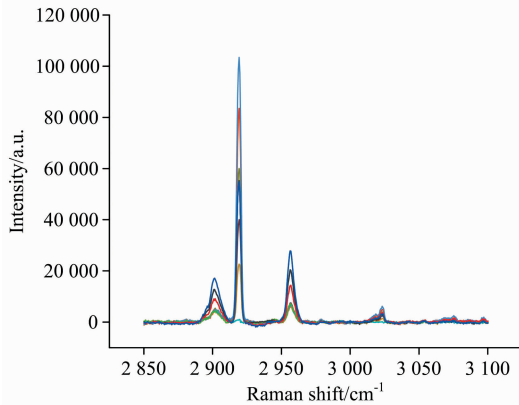


图 6 预处理后混合气体样本的拉曼光谱图

Fig. 6 Raman spectra of pretreated mixed gas samples

### 2.3 定位寻峰多参量提取

针对拉曼光谱图特性，本文以测量数据为依据，基于 Matlab 设计了定位寻峰程序。针对研究气体样本提取参考频移 ±100 cm<sup>-1</sup> 范围内光谱数据进行强度比较<sup>[15]</sup>，选取与气体样本理论拉曼频移数目相同的多个拉曼频移与文献中的参考频移进行比较，通过相对位置和偏差率判断出 CH<sub>4</sub> 和 C<sub>2</sub>H<sub>6</sub> 混合拉曼光谱中 2 902, 2 918, 2 956 和 3 022 cm<sup>-1</sup> 四个谱峰归属。

光谱图中每个谱峰都具有明确的物理意义且携带不同的物理信息，且考虑量子力学修正可知拉曼散射强度与样品单位体积内的散射分子数(样品浓度)成正比。采用式(4)中高斯函数<sup>[16]</sup>作为基本形式对定位的四个谱峰进行分峰拟合，并用高斯参量表达各不同频移位置处谱峰的面积、半高宽和峰高等参数。

$$y = y_0 + Ae^{-\frac{(x-x_c)^2}{2(\frac{FWHM}{2\sqrt{\ln 4}})^2}} \quad (4)$$

式(4)中，y 为拉曼散射强度；x 为拉曼频移；A 为拉曼谱峰高度；x<sub>c</sub> 为谱峰中心偏移量；FWHM 为谱峰半高宽。拟合相关系数 R<sup>2</sup> 为 0.9851，表明得到了较好的拟合结果。得到全部特征量参数如表 3。

表 3 混合谱图高斯分峰拟合特征参数

Table 3 Gauss peak separation fitting characteristic parameters of mixed spectrum

序号	x <sub>c</sub> =2 901 cm <sup>-1</sup> (峰 1)			x <sub>c</sub> =2 918 cm <sup>-1</sup> (峰 2)			x <sub>c</sub> =2 956 cm <sup>-1</sup> (峰 3)			x <sub>c</sub> =3 021 cm <sup>-1</sup> (峰 4)		
	S <sub>1</sub>	A <sub>1</sub>	FWHM <sub>1</sub>	S <sub>2</sub>	A <sub>2</sub>	FWHM <sub>2</sub>	S <sub>3</sub>	A <sub>3</sub>	FWHM <sub>3</sub>	S <sub>4</sub>	A <sub>4</sub>	FWHM <sub>4</sub>
1	41 478.1	4 459.5	7.778 3	83 340.9	22 785.2	3.496 87	36 368.9	6 319.1	5.441 77	7 642.3	1 635.9	3.320 8
2	47 881.8	4 943.2	7.964 36	147 301.8	40 069.8	3.540 19	44 475.3	7 506.7	5.646 55	13 432.5	2 552.4	3.839 28
3	53 061.5	5 571.9	7.791 4	224 306.5	60 160.4	3.556 84	44 435.1	7 308.1	5.865 36	23 156.4	4 093.2	3.742
4	47 026.0	4 984.1	7.648 96	311 867.8	83 642.9	3.527	37 729.8	6 583.5	5.681 02	32 326.5	5 067.6	4.215 89
5	49 338.7	5 101.1	8.384 72	389 240.9	103 636	3.544 65	40 104.0	6 663.3	5.802 49	40 658.3	6 338.4	3.978 57
6	38 846.5	4 546.1	7.596 59	198 741.4	55 177.5	3.555 4	40 022.3	6 999.3	5.598 78	19 211.9	3 182.2	4.329 68
7	82 803.6	9 131.5	7.775 23	197 374.8	54 227.3	3.554 78	82 892.9	14 462.1	5.293 5	26 415.6	3 631.5	4.361 76
8	116 912.5	12 741.8	7.987 09	199 282.9	55 245.4	3.546 14	122 300	20 466.1	5.510 58	15 238.4	3 177.7	3.659 57
9	160 207.9	17 147.8	7.993 82	198 608.6	54 813.7	3.533 03	167 599	27 784.8	5.424 35	19 723.0	3 244.0	4.492 69
10	225 263.9	25 915.2	8.161 95	197 461.3	55 645.8	3.561 89	209 384	34 019.5	5.365 47	20 814.3	3 195.1	4.717 65
11	45 514.0	5 957.0	6.373 51	90 323.0	25 696.1	3.465 31	56 568.9	9 177.4	5.776 99	5 973.7	1 757.1	3.844 96
12	57 626.5	6 390.0	8.568 18	151 715.6	42 998.6	3.444 75	55 929.6	10 549.9	5.056 06	56.1	246.8	0.247 82
13	60 602.8	6 533.5	7.659 23	230 459.3	62 933.0	3.534 81	56 031.4	10 318.4	5.278 37	20 268.5	3 869.0	4.103
14	41 496.3	5 444.0	7.405 22	300 350.6	82 797.9	3.551 41	55 831.8	10 028.9	5.371 5	29 247.7	5 151.6	3.907 67
15	35 198.0	5 282.8	6.618 77	386 214.7	102 726	3.606 12	52 226.8	9 323.1	5.283 5	42 700.4	6 432.1	4.875 27
16	0	0	0	201 566	3.536 46	55 418.6	0	0	0	18 388.9	3 728.1	3.711 73
17	61 267.8	6 828.7	7.375 65	0	0	0	59 829.8	10 163.5	5.412 77	0	0	0

由此，针对每组 CH<sub>4</sub> 和 C<sub>2</sub>H<sub>6</sub> 混合气体，选取四个特征峰，每个峰选取中心频移、谱峰面积、半高宽、峰高四个参量，作为下一步定量分析的基础。

### 3 油中气体偏最小二乘回归模型

考虑选取的自变量为谱峰高度、半高宽及谱峰面积，以要得到的两种气体含量为因变量。在单一气体定量分析的研究中，已经确定谱峰高度、半高宽及谱峰面积与气体含量相

关性较强，目的是求出回归方程各参数，因此归纳为建立一个多参量二输出的回归数学模型。

考虑到实验选取谱段中每组预测样本自变量个数以及对两个因变量，电气在线监测要求快速准确，且传统回归建模分析计算比较复杂，不适用于多种特征气体混合频谱数据的分析，因此本文采用偏最小二乘法(PLS)建立预测模型。PLS 是一种简单的多元校正建模方法，建模分析过程中融合了主成分分析、典型相关分析和线性回归分析等方法，具有更好的鲁棒性和预测稳定性。

针对拉曼光谱数据特点, 建立偏最小二乘法回归模型的步骤如下:

(1) 数据标准化。

(2) 求第一对主成分。自变量  $X$  与因变量  $Y$  的主成分分别为  $t_1$  和  $u_1$ , 要求主成分包含原数据含尽量多的信息, 同时  $t_1$  对  $u_1$  具有最大解释能力, 即: ①  $t_1$  与  $u_1$  方差最大; ②  $t_1$  与  $u_1$  相关程度最大。由

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}} \quad (5)$$

可知, 当协方差 ( $\text{Cov}(X, Y)$ ) 最大时, 即可满足两个条件。式中  $\text{Cov}(X, Y)$  表示  $X$  与  $Y$  的相关系数,  $\text{Var}(X)$  表示  $X$  的方差。

(3) 建立  $X_0, Y_0$  对  $t_1$  和  $u_1$  的三个回归方程, 如式(6)

$$\begin{cases} X_0 = t_1 p'_1 + Y_1 \\ Y_0 = u_1 q'_1 + Y'_1 \\ Y_0 = t_1 r'_1 + Y_1 \end{cases} \quad (6)$$

式(6)中回归系数向量是

$$\begin{cases} p_1 = \frac{E_0' t_1}{t_1^2} \\ q_1 = \frac{Y_0' u_1}{u_1^2} \\ r_1 = \frac{Y_0' t_1}{t_1^2} \end{cases} \quad (7)$$

式(7)中,  $X_1, Y'_1, Y_1$  分别是三个回归方程的残差矩阵。

(4) 继续求主成分, 直到满足设定条件

将残差矩阵  $X_1$  和  $Y_1$  替代  $X_0$  和  $Y_0$ , 求更多主成分  $t_i$  以及  $u_i$ , 如式(8)

$$\begin{cases} t_i = X_{i-1} w_i \\ u_i = Y_{i-1} c_i \end{cases} \quad (8)$$

(5) 推导回归表达式

若  $X_0$  的秩为  $A$ , 可以得出

$$\begin{cases} X_0 = t_1 p'_1 + \dots + t_A p'_A \\ Y_0 = t_1 r'_1 + \dots + t_A r'_A + Y_A \end{cases} \quad (9)$$

即

$$y_k^* = a_{k1} x_1^* + \dots + a_{kp} x_p^* + Y_{Ak} \quad (10)$$

式(10),  $Y_{Ak}$  为残差矩阵  $Y_A$  的第  $k$  列。回归方程得以求出。

表 4 潜在因子解释的方差比例

Table 4 Explained variance ratio of potential factors

潜在因子	统计				
	X 方差	累计的 X 方差	Y 方差	累计的 Y 方差 ( $R^2$ )	调整后的 $R^2$
$t_1$	0.419	0.419	0.471	0.471	0.436
$t_2$	0.378	0.797	0.467	0.938	0.929
$t_3$	0.120	0.917	0.017	0.955	0.944
$t_4$	0.047	0.964	0.025	0.980	0.973
$t_5$	0.029	0.993	0.004	0.983	0.975
$t_6$	0.004	0.997	0.012	0.996	0.993

(6) 检验有效性

从 17 组气体样本中选取 12 组作为预测样本, 其余 5 组作为检验样本, 建立计算气体含量的偏最小二乘法模型。实验发现, 如表 4 所示, 取到 6 个潜在因子时调整后的拟合优度  $R^2$  已经达到 0.993, 即使再增加潜在因子对拟合效果影响微小, 同时会让模型更加复杂, 因此  $t=6$  时较为适合。

## 4 结果与讨论

取拉曼光谱图中谱峰高度、半高宽和谱峰面积作为自变量, 建立偏最小二乘数学回归模型, 得到的回归方程参数如表 5。

表 5 回归方程参数

Table 5 Regression equation parameters

自变量	因变量	
	CH <sub>4</sub>	C <sub>2</sub> H <sub>6</sub>
常数	-0.004	-0.001
S <sub>1</sub>	-9.956 × 10 <sup>-8</sup>	-3.233 × 10 <sup>-7</sup>
S <sub>2</sub>	2.366 × 10 <sup>-7</sup>	4.498 × 10 <sup>-8</sup>
S <sub>3</sub>	3.827 × 10 <sup>-8</sup>	5.829 × 10 <sup>-7</sup>
S <sub>4</sub>	8.628 × 10 <sup>-7</sup>	-5.934 × 10 <sup>-7</sup>
FWHM <sub>1</sub>	0.002	-0.006
FWHM <sub>2</sub>	0.007	0.000
FWHM <sub>3</sub>	-0.002	0.008
FWHM <sub>4</sub>	-0.007	-0.001
A <sub>1</sub>	-1.923 × 10 <sup>-7</sup>	2.363 × 10 <sup>-6</sup>
A <sub>2</sub>	9.060 × 10 <sup>-7</sup>	2.185 × 10 <sup>-7</sup>
A <sub>3</sub>	6.280 × 10 <sup>-7</sup>	4.171 × 10 <sup>-6</sup>
A <sub>4</sub>	6.083 × 10 <sup>-6</sup>	-1.153 × 10 <sup>-6</sup>

由表 5 分析可知, 在此算法中, 谱峰半高宽参数在回归方程中是其他参数的 10<sup>4</sup> ~ 10<sup>6</sup> 倍, 表明半高宽对两种气体的含量预测结果有明显影响, 谱峰面积以及谱峰高度对气体含量的预测有影响, 精确度不及谱峰半高宽。混合气体实验中, 激光拉曼光谱图的基线去除会影响谱峰面积以及谱峰高度, 难以得到理论上正确的谱峰高度以及谱峰面积, 所以使用理论公式计算气体含量误差较大, 偏最小二乘法刚好回避这一问题。

回归方程参数中四个谱峰的半高宽均贡献明显, 表明甲烷含量预测取决于四个谱峰信息, 这与纯甲烷拉曼测量时表现出了不同的特性, 说明混合气体测量时不是简单的重叠相加, 也侧面论证了混合气体定量计算时需要算法的需求; 乙烷的主要影响因素是峰 1、峰 2 和峰 4 的半高宽, 同样说明对乙烷预测含量时不再是仅仅考虑理论上两个谱峰。

得到预测结果如表 6。表 6 的测试结果及误差显示, 误差大体上随气体含量增大而减小, 平均相对误差 (average relative error, ARE) 为 2.25%。两种气体的真实值和实测值拟合如图 7 所示, 模型对不同气体量的实测值没有交叉及重叠, 实测值与真实值的误差均保持在 3% 以内, 表明基于偏最小二乘法 (PLS) 回归分析对变压器油中此两种特征气体量的定量分析模型具有极高的准确率。

表 6 测试结果对比  
Table 6 Comparison of test results

气体	真实值/MPa	实测值/MPa	测试误差/MP
甲烷	0.05	0.052	0.002
	0.10	0.095	-0.005
	0.15	0.147	-0.003
	0.20	0.199	-0.001
	0.25	0.253	0.003
乙烷	0.05	0.048	-0.002
	0.08	0.081	0.001
	0.10	0.097	-0.003
	0.15	0.152	0.002
	0.20	0.200	0

该方法仅采用 12 组混合气体数据作为训练量，能够得到误差在 3% 的测试结果，表明了选取拉曼光谱图中谱峰高度、半高宽以及谱峰面积为参量，建立偏最小二乘回归模型对拉曼光谱中变压器油中特征气体的定量分析检测具有明显优势。取样数据少，模型计算不复杂，检测精度高，效果良好。

### 5 结 论

(1)对变压器油中特征气体拉曼光谱图预处理时，采用“移动平滑去噪和中位线去基线”效果良好，得到的拉曼光谱图特征谱峰明显，便于气体定量计算。

(2)通过高斯拟合对拉曼光谱特征峰的谱峰高度、半高宽以及谱峰面积进行提取，建立偏最小二乘法回归模型，分析气体含量，调整后的拟合优度  $R^2$  为 0.993，拟合效果良好，方法可行。

(3)在室温 25 °C，积分时间 15 s，积分次数 2，狭缝 100  $\mu\text{m}$  条件下，选取半高宽、谱峰面积和谱峰高度，使用偏最小二乘法对变压器特征气体甲烷乙烷进行定量，平均相对误差

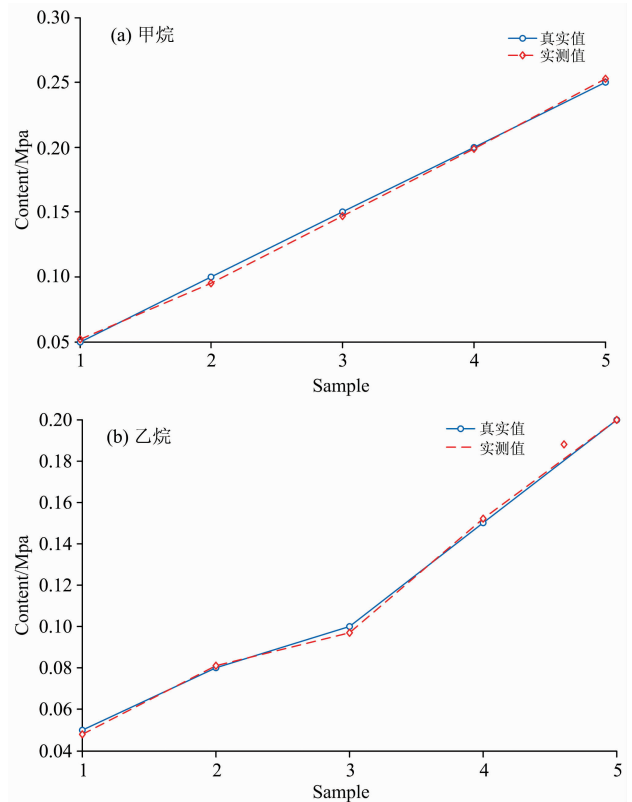


图 7 甲烷和乙烷含量真实值与实测值对比  
(a): 甲烷; (b): 乙烷

Fig. 7 Comparison of real and measured methane and ethane contents  
(a): Methane; (b): Ethane

为 2.25%。

基于半高宽、谱峰面积和谱峰高度的多参量激光拉曼光谱研究，为变压器油中特征气体的快速在线检测提供了思路。

### References

[ 1 ] WANG Wen-bin, BAI Wen-guang, SHI Lei-lei, et al(王文宾, 白文广, 石磊磊, 等). Power System Protection and Control(电力系统保护与控制), 2019, (4): 91.

[ 2 ] LI Zhen-jie, BIAN Chao-hui, CHEN Xue-min, et al(李振杰, 卞朝辉, 陈学民, 等). High Voltage Apparatus(高压电器), 2018, 54(6): 158.

[ 3 ] CHEN Xin-gang, LI Song, MA Zhi-peng, et al(陈新岗, 李松, 马志鹏, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2016, 36(8): 2492.

[ 4 ] XIA Jia-bin, ZHU Lian-qing, YAO Qi-feng, et al(夏嘉斌, 祝连庆, 姚齐峰, 等). Chinese Journal of Scientific Instrument(仪器仪表学报), 2018, 39(3): 141.

[ 5 ] Hou Kaiyuan, Shao Guanghui, Wang Haiming, et al. Protection and Control of Modern Power Systems, 2018, 3(3): 119.

[ 6 ] CHEN Wei-gen, ZHAO Li-zhi, PENG Shang-yi, et al(陈伟根, 赵立志, 彭尚怡, 等). Proceedings of the CSEE(中国电机工程学报), 2014, 34 (15): 2485.

[ 7 ] WAN Fu, YANG Man-lin, HE Peng, et al(万福, 杨曼琳, 贺鹏, 等). Chinese Journal of Scientific Instrument(仪器仪表学报), 2016, 37(11): 2482.

[ 8 ] Gu Zhaoliang, Chen Weigen, Xiang Yingzhu, et al. 2016 IEEE International Conference on High Voltage Engineering and Application (ICHVE), 2016.

[ 9 ] TIAN Zhi-xian, ZHANG Xin, LIU Chang-ling, et al(田陟贤, 张鑫, 刘昌岭, 等). China Environmental Science(中国环境科学),

- 2014, 34(10): 2631.
- [10] Toshihiro Somekawa, Masayuki Fujita, Yasukazu Izawa, et al. IEEE Transactions on Dielectrics and Electrical Insulation, 2015, 22(1): 229.
- [11] HU Li, ZHAO Nan-jing, LIU Wen-qing, et al(胡 丽, 赵南京, 刘文清, 等). Chinese Journal of Lasers(中国激光), 2014, 41(7): 0715003.
- [12] GAO Ying, DAI Lian-kui, ZHU Hua-dong, et al(高 颖, 戴连奎, 朱华东, 等). Chinese Journal of Analytical Chemistry(分析化学), 2019, 47(1): 67.
- [13] Chilaka Ranga, Ashwani Kumar Chandel. 2017 3rd International Conference on Condition Assessment Techniques in Electrical Systems (CATCON), 2017.
- [14] LI Ling(李 玲). Laser Journal(激光杂志), 2019, 40(2): 153.
- [15] CHEN Xin-gang, YANG Ding-kun, TAN Hao, et al(陈新岗, 杨定坤, 谭 昊, 等). High Voltage Engineering(高电压技术), 2017, 43(7): 2256.
- [16] Tadesse Ghirmai. 2014 48th Annual Conference on Information Sciences and Systems (CISS). 2014.

## Quantitative Analysis of Dissolved Gases in Transformer Oil Based on Multi-Parameter

CHEN Xin-gang<sup>1,2</sup>, FENG Yu-xuan<sup>1\*</sup>, LI Chang-xin<sup>1</sup>, CHEN Shu-ting<sup>1</sup>, CHEN Xiao-qing<sup>1</sup>, LONG Yao<sup>1</sup>, CHEN Lin-chi<sup>3</sup>

1. Chongqing University of Technology, Chongqing 400054, China

2. Chongqing Energy Internet Engineering Technology Research Center, Chongqing 400054, China

3. Chongqing University of Post and Telecommunications, Chongqing 400065, China

**Abstract** The rapid and accurate detection of characteristic gases ( $H_2$ ,  $CO$ ,  $CO_2$ ,  $CH_4$ ,  $C_2H_4$ ,  $C_2H_6$ ,  $C_2H_2$ ) in oil is an important part of transformer on-line monitoring. Laser Raman spectroscopy is suitable for the detection of characteristic gases and can overcome many shortcomings of traditional on-line monitoring. When the characteristic gas in transformer oil is detected, the characteristic peaks of methane ( $CH_4$ ) and ethane ( $C_2H_6$ ) gather in the Raman spectrum from 2 900 to 3 300  $cm^{-1}$ . It is of great significance to study the mixed gases samples with different content ratios in this spectrum for the quantitative analysis of mixed gases in transformer oil. Based on the research of Raman spectroscopy detection of single characteristic gas, the spectral peak height, full-width at half-maximum and spectral peak area parameters of the characteristic peaks in the pretreated spectrum are selected as the characteristic factors to quantitatively analyze the mixed gas in the transformer oil. According to the second-order perturbation theory, there are four characteristic peaks in the methane Raman spectrum, and the selected spectrum contains two peaks with 3 111 and 3 284  $cm^{-1}$  as the Raman shift center. Six characteristic peaks exist in the Raman spectra of ethane, and there are two peaks of 3 111 and 3 187  $cm^{-1}$  in the selected spectral bands. In theory, the amount of both gases can be calculated by the information of the characteristic peak carried in the spectrum bands. Through the detection of Raman spectroscopy platform, the characteristic peaks of the mixed gas spectrum will produce translation and polymerization. In practice, four peaks with shifts center of 1 902, 2 918, 2 956 and 3 022  $cm^{-1}$  were found in the spectrum. A Gaussian function model was established for the four mixed peaks, and the spectral peak height, full-width at half-maximum and spectral peak area of the characteristic peak was obtained. A partial least squares regression model (PLS) was established. The spectral peak height, full-width at half-maximum and spectral peak area were taken as independent variables, and the two gas contents were taken as dependent variables for calculation and analysis. When the potential factor of the model is taken to  $t_6$ , the adjusted R-square is 0.993, indicating that the independent variable has a definite relationship with the dependent variable, and the regression model is reliable. The analysis of regression equation parameters shows that the full-width at half-maximum of spectral peaks contribute significantly to the area and height of spectral peaks, which is in line with the expected target. Four characteristic spectral peaks in the spectrum of mixed gases have an effect on both gases. It can be concluded from the experiments that for the methane-ethane mixed gas, at room temperature 25  $^{\circ}C$ , integration time 15 s, integral number 2, slit 100  $\mu m$ , by obtaining the peak height, peak area and full-width at half-maximum three parameters, the gas content can be accurately measured, which lays a foundation for the simultaneous detection of various characteristic gases in transformer oil.

**Keywords** Raman spectra; Transformer; Characteristic gases; Quantitative analysis; Multi-parameter

\* Corresponding author

(Received Jul. 24, 2019; accepted Nov. 8, 2019)