第16卷第11期
2026年11月

中国无机分析化学
Chinese Journal of Inorganic Analytical Chemistry

Vol. 16, No. 11
1969～1979

doi: 10. 20236/j. CJIAC. 2026. 11. 001

郑培超，张文硕，王金梅，等. 特征谱线矩阵与样品聚类辅助的钢铁中铬元素激光诱导击穿光谱(LIBS)定量分析[J]. 中国无机分析化学，2026，16(11)：1969-1979.

Zheng Peichao, Zhang Wenshuo, Wang Jinmei, et al. Quantitative LIBS Analysis of Chromium in Steels Assisted by a Characteristic Spectral-line Matrix and Sample Clustering[J]. Chinese Journal of Inorganic Analytical Chemistry, 2026, 16(11): 1969-1979.

特征谱线矩阵与样品聚类辅助的钢铁中铬元素激光诱导击穿光谱(LIBS)定量分析

郑培超　张文硕　王金梅*　代攀杨　李思琪　张　艺　黄浩轩

（重庆邮电大学 电子科学与工程学院，重庆 400065）

摘　要　成分跨度较大的钢铁样品存在明显的基体依赖响应，全谱建模又包含大量冗余变量。在保持谱线物理归属和低维表达的前提下，评价样品聚类对钢铁中Cr定量的作用。将27条样品平均光谱作为独立建模单元，经稀疏性基线估计与去噪(BEADS)处理后提取Fe、Cr、Ni、Mn、Mo和Ti共23条特征谱线的峰面积，以Fe Ⅰ 271.902 nm谱线构建内标比值；采用主成分分析(PCA)与K-means++形成K=3分组，并在外层留一法(逐样品留一)过程中，将积分窗口选择、标准化、PCA、聚类、留出样品归类及偏最小二乘(PLS)潜变量选择全部限制在训练折内。K-means++与高斯混合模型(GMM)均得到11、9和7个样品的相同分组；27个外层折的训练样品分组与全样本描述性分组的调整兰德指数(ARI)均为1.000。Fe Ⅰ 271.902 nm在4条候选Fe谱线中具有最低的峰面积相对标准偏差(RSD)中位数，但内标校正对不同分析线的重复性影响具有谱线依赖性。各模型的负预测值在汇总评价前按质量分数非负边界置为0。非负后处理后，K=3聚类辅助偏最小二乘回归(Cluster-PLSR)模型获得决定系数(R^2)、均方根误差(RMSE)、平均绝对误差(MAE)和性能偏差比(RPD)分别为0.953 1、1.886%、1.314%和4.70；相对于相同23变量的全局Fe内标PLSR，RMSE降低43.24%。与16 355个变量的全谱模型相比，特征变量数减少99.86%。物理可解释的特征谱线矩阵结合成分感知分组，可在所研究钢样范围内显著改善低维全局模型性能；全谱模型仍保持最高绝对精度，本文方法的主要价值在于变量轻量化和结果可解释性。

关键词　激光诱导击穿光谱；钢铁分析；铬；特征谱线；聚类分析；偏最小二乘回归

要点

1）23维特征矩阵使光谱变量减少99.86%。

2）K=3分组对应三类可解释的钢样成分组成。

3）同维度聚类模型RMSE降低43.24%。

中图分类号：O657.38　　**文献标志码**：A　　**文章编号**：2095-1035(2026)11-1969-11

收稿日期：2026-07-31　　**修回日期**：2026-08-07　　**录用日期**：2026-08-08　　**在线出版日期**：2026-08-17

Received：2026-07-31　　**Revised**：2026-08-07　　**Accepted**：2026-08-08　　**Online published**：2026-08-17

基金项目：国家自然科学基金资助项目(62575044)；重庆市自然科学基金资助项目(CSTB2024NSCQ-LZX0078)；重庆市教委科技研究项目重大项目(KJZD-M202600608)

Fund：Supported by the National Natural Science Foundation of China (62575044); Natural Science Foundation of Chongqing Municipality (CSTB2024NSCQ-LZX0078); Major Science and Technology Research Program of Chongqing Municipal Education Commission (KJZD-M202600608)

作者简介：郑培超，男，教授，主要从事新型光谱检测方面的研究。E-mail：zhengpc@cqupt. edu. cn

Author：Zheng Peichao, male, professor, mainly engaged in novel spectroscopic detection. E-mail: zhengpc@cqupt. edu. cn

*通信作者：王金梅，女，教授，主要从事光电子和微电子技术方面的研究。E-mail：wangjm@cqupt. edu. cn

*Corresponding author: Wang Jinmei, female, professor, mainly engaged in optoelectronic and microelectronic technologies. E-mail: wangjm@cqupt. edu. cn

©Chinese Journal of Inorganic Analytical Chemistry Editorial Department 2026. This is an open access article under the CC-BY-NC-ND 4.0 License.

钢铁中Cr、Ni、Mn、Mo和Ti等合金元素的含量直接影响材料的强度、韧性、耐蚀性及服役稳定性，快速获取其组成信息是冶炼过程控制、牌号判定和质量检验的重要基础。激光诱导击穿光谱(Laser-induced Breakdown Spectroscopy，LIBS)通过脉冲激光烧蚀样品并采集瞬态等离子体发射光谱，具有样品前处理少、多元素同步检测、分析速度快以及便于原位和远程测量等特点[1-4]。已有研究将LIBS用于钢水、钢渣和成品钢的离线或在线检测，面向工业过程的仪器稳定性、自动化和定量评价体系也在持续完善[5-6]，表明该技术适用于钢铁成分的快速筛查与过程监测。

围绕钢铁中合金元素的定量分析，研究者已从实验条件、定标策略和仪器形式等方面开展探索。通过优化激光聚焦与光谱采集条件，可改善低合金钢中C、Si、Cr和Ni的定量响应[7]；对于高合金钢，元素间校正有助于减弱由组成和烧蚀响应差异引起的系统偏差[8]。无标样定量、便携式光纤激光LIBS、遗传算法优化神经网络和手持式LIBS等方法进一步拓展了钢铁分析场景[9-12]。在谱线拥挤或仪器分辨率受限时，导数预处理结合(Partial Least Squares Regression，PLSR)仍可提取Mn、Cr、Ni和Ti等元素的定量信息[13]；图像特征辅助建模也为高合金钢基体效应补偿提供了补充途径[14]。

随着化学计量学方法的发展，LIBS还被用于钢种识别、异常光谱筛除和特征变量压缩。随机森林能够利用全谱信息区分不同钢牌号[15]，主成分分析(Principal Component Analysis，PCA)、相关性和总强度等准则可用于识别异常光谱并提高后续分析稳定性[16]。针对全谱维数高和冗余变量多的问题，特征选择可在减少变量的同时保持不锈钢中Ni、Ti和Cr的定量能力[17]；卷积神经网络协同优化和PCA结合机器学习算法则分别用于痕量Cr分析和同基合金识别[18-19]。这些结果说明，围绕元素谱线和样品组成进行信息筛选，有望获得比直接堆叠全谱变量更轻量、可解释的数据表达[20]。

然而，复杂钢种的统一定量模型仍受到基体差异和高维冗余的共同制约。不同钢种的合金组成、组织和表面烧蚀行为会改变烧蚀质量、等离子体状态、自吸收及谱线干扰，使分析线强度与元素含量之间呈现基体依赖关系[6-7,21-22]。内标、总强度和标准正态变量变换等方法能够补偿部分整体波动，但效果取决于分析线和样品条件，不能预设某一种归一化方式对全部谱线均有利[23-24]。具有明确元素归属的峰面积、适合共线变量的PLSR以及稀疏性基线估计与去噪(Baseline Estimation and Denoising with Sparsity，BEADS)背景分离，为构建低维定量表达提供了基础[25-26]；无监督聚类结合局部回归则可针对不同基体建立校准关系[27]。同时，若标准化、聚类或超参数选择预先使用全部样品，交叉验证结果会因信息泄漏而偏乐观[28-30]，因此局部模型的性能必须在训练折内重建完整流程后评价。

基于上述认识，本研究以27条钢铁样品平均光谱为独立建模单元，经BEADS校正后构建Fe、Cr、Ni、Mn、Mo和Ti共23条特征谱线的峰面积及Fe内标比值矩阵，并以预先设定、具有组成解释的$K=3$分组建立聚类辅助PLSR。积分窗口选择、标准化、PCA、聚类、留出样品归类和偏最小二乘(Partial Least Squares，PLS)潜变量选择均限制在样品级嵌套留一交叉验证(Leave-One-Out Cross-Validation，LOOCV)的训练折内。本文同时比较全谱、原始峰面积和Fe内标特征的全局模型，并以相同23维输入检验聚类辅助建模的实际增益。研究目的不是以局部模型取代绝对精度最高的全谱模型，而是评价在大幅压缩变量的条件下，组成感知分组能否改善低维全局关系，并为钢铁LIBS定量提供轻量、可解释的建模方案。

1 实验部分

1.1 钢铁样品与组成信息

研究采用27个钢铁样品，统一编号为样品1#—27#。Cr和Ni质量分数范围分别为0.023%～24.36%和0.011%～18.500%，同时包含Mn、Mo和Ti的质量分数差异，可覆盖低Cr低Ni钢、Cr-Ni合金钢以及含Ti或Mo的特殊合金钢。用于特征选择和分组解释的5种元素的质量分数列于表1。该组成跨度为后续特征谱线筛选及不同基体条件下的模型比较提供了样品基础。

每个样品选取10个位置，每个位置连续采集50次激光脉冲光谱；舍弃前20次后，将后30次平均为1条位置平均光谱，共得到270条位置平均光谱。再将同一样品的10条位置平均光谱平均为1条样品平均光谱。聚类和定量均以27条样品平均光谱为独立建模单元，270条位置平均光谱仅用于计算重复性和形成样品均值，不作为彼此独立的建模样品。

表1　27个钢样品中Cr、Ni、Mn、Mo和Ti的质量分数

Table 1　Mass fractions of Cr, Ni, Mn, Mo and Ti in the 27 steel samples

Samples	Cr/%	Ni/%	Mn/%	Mo/%	Ti/%
1	0.023	0.011	0.018	0.005 3	0.000 2
2	0.042	0.045	0.104	0.304	0.02
3	0.236	0.563	0.303	0.034	0.049
4	0.102	0.40	0.669	0.144	0.202
5	0.106	0.393	1.04	0.105	0.178
6	0.409	0.206	1.47	0.071	0.124
7	0.612	0.107	2.1	0.196	0.088
8	22.39	7.96	2.07	0.08	-
9	19.78	14.19	0.425	0.955	0.025
10	16.4	10.32	1.15	2.04	0.008
11	17.13	9.05	1.03	0.051	0.482
12	13.3	13.51	0.349	0.458	-
13	23.04	12.44	1.05	0.033	-
14	24.36	18.5	1.54	0.182	0.017
15	18.08	8.17	1.27	0.114	0.005
16	21.8	4.67	1.07	2.78	-
17	16.18	6.19	0.972	0.092	-
18	1.89	1.06	0.292	0.258	0.000 8
19	1.68	0.03	0.582	0.306	0.001 9
20	19.4	6.48	0.49	0.021	0.206
21	14.75	13.68	0.321	0.466	0.001 1
22	17.41	9.94	1.04	0.104	-
23	20.12	7.33	0.39	0.055	0.19
24	15.89	9.06	0.8	0.16	0.34
25	12.98	10.97	1.19	0.055	0.55
26	9.95	12.93	1.62	0.2	0.77
27	18.84	8.44	0.14	-	-

1.2　LIBS实验系统与光谱采集

实验使用调Q Nd：YAG激光器(法国Quantel公司，Ultra 100型)作为激发源，工作波长1 064 nm、单脉冲能量40 mJ、重复频率20 Hz。激光经能量调节、扩束和反射后，由焦距50 mm的平凸透镜垂直聚焦于样品表面。等离子体发射由与样品法线约45°布置的收光系统采集，并耦合至4台AvaSpec-Mini4096CL光谱仪；系统光谱范围为186～797 nm，分辨率约0.05 nm、积分时间1 ms。激光器采用外触发模式并由MaxLIBS软件控制。样品经机械打磨和无水乙醇清洗后置于三维移动平台，以保证各次烧蚀作用于新的位置。实验系统见图1，主要参数列于表2。在一致的采集条件下获得样品光谱后，进一步通过背景校正和谱线积分构建可比较的特征变量。

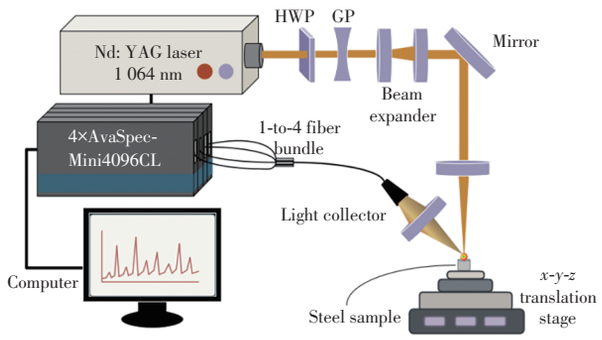


图1　LIBS实验系统的光路与信号采集原理

Figure 1　Optical path and signal-acquisition principle of the LIBS system.

表2　LIBS光谱采集参数

Table 2　LIBS acquisition parameters

Parameter	Setting	Description
Laser	Quantel Ultra 100	Q-switched Nd: YAG
Laser wavelength/nm	1 064	Actual single-pulse setting
Laser energy/mJ	40	
Repetition rate/Hz	20	External trigger
Focal length/mm	50	Normal focusing; collection at ~45°
Spectrometer	Four AvaSpec-Mini4096CL spectroneters	Resolution ~0.05 nm
Integration time/ms	1	MaxLIBS control
Spectral range/nm	186－798	16 355 wavelength variables
Measurement positions per sample	10	Uniformly distributed
Laser shots per position	50	First 20 shots discarded; last 30 averaged

1.3　光谱预处理与特征谱线矩阵构建

采用BEADS算法估计并扣除缓慢变化的连续背景[26]。单条光谱表示为真实发射信号、背景与随机噪声之和：

$$y=x+b+\varepsilon \tag{1}$$

式中，y、x、b和ε分别为原始光谱、真实发射信号、缓慢变化背景和随机噪声向量。BEADS参数设为freq_cutoff=0.04，$\lambda_0=\lambda_1=\lambda_2=1$，asymmetry=6，最大迭代次数50，收敛阈值0.01；基线校正后不对负值作强制截断。全谱对照模型在BEADS校正后进一步进行标准正态变量变换(Standard Normal Variate，SNV)：

$$x_{ij}^{\mathrm{SNV}}=\frac{x_{ij}-\bar{x}_i}{s_i},\ j=1,\cdots,p \tag{2}$$

式中，x_{ij}为第i条样品光谱在第j个波长变量处的强度，x_i和s_i分别为该条光谱全部p个变量的均值和样本标准差。对中心波长λ_j附近的峰面积及非Fe特征的内标比值分别计算为：

$$A_{ij}(\Delta\lambda)=\int_{\lambda_j-\Delta\lambda}^{\lambda_j+\Delta\lambda} I_i(\lambda)\mathrm{d}\lambda \tag{3}$$

$$R_{ij}=\frac{A_{ij}}{A_{i,\mathrm{Fe\ I\ 271.902}}},\ j\notin \mathrm{Fe} \tag{4}$$

式中，$\Delta\lambda$为积分半宽，在±0.07、±0.10、±0.12和±0.18 nm中由每个外层训练折的内层误差选择。4条Fe谱线保留峰面积，19条非Fe谱线以Fe Ⅰ 271.902 nm峰面积归一化，形成23维FeRatio矩阵；未归一化峰面积矩阵LineArea作为低维对照。

候选谱线依据样品主要组成及NIST Atomic Spectra Database (version 5.12)逐条核对原子/离子归属[31]。23条谱线覆盖Fe基体、Cr目标元素以及Ni、Mn、Mo和Ti等主要合金元素，具体构成见表3。由此得到的23维FeRatio矩阵既保留了目标元素响应信息，也提供了描述样品组成差异的基体与合金元素信息，作为后续聚类与定量建模的统一输入。

表3 23维特征谱线矩阵构成

Table 3 Composition of the 23-dimensional characteristic spectral-line matrix

Elements	Central wavelength and assignment/nm	Variable form	Purpose
Fe	248.327, 271.902, 438.354, 527.039 (Fe Ⅰ)	Peak area	Matrix response; Fe Ⅰ 271.902 nm as common internal standard
Cr	357.869, 359.349, 360.533, 425.433, 427.480, 428.972 (Cr Ⅰ)	Peak area/Fe Ⅰ 271.902	Target-element information
Ni	231.604 (Ni Ⅱ); 341.476, 344.626, 352.454 (Ni Ⅰ)	Peak area/Fe Ⅰ 271.902	Major alloying element
Mn	257.610, 259.373 (Mn Ⅱ); 403.307 (Mn Ⅰ)	Peak area/Fe Ⅰ 271.902	Major alloying element
Mo	550.650, 553.305 (Mo Ⅰ)	Peak area/Fe Ⅰ 271.902	Special alloying element
Ti	334.941, 337.280 (Ti Ⅱ); 498.173, 499.107 (Ti Ⅰ)	Peak area/Fe Ⅰ 271.902	Special alloying element

1.4 样品聚类辅助PLSR与样品级嵌套验证

聚类前，对23维FeRatio特征进行训练折标准化，并以PCA提取得分，随后采用K-means++形成3个样品组，同时以Ward连接的层次聚类分析(Hierarchical Cluster Analysis, HCA)和高斯混合模型(Gaussian Mixture Model, GMM)检验分组的算法一致性。K=3在定量建模前依据样品组成结构预先设定：3组分别对应低Cr低Ni的Fe基钢、Cr-Ni合金钢和含Ti特殊合金钢，且每组均保留足以用于局部PLSR的样品量。该设定不使用Cr预测误差，因此不以K=2～6的监督性能反向选择聚类数。标准化和PCA均采用常规定义，K-means目标函数及PLSR预测关系表示如式(5)、(6)所示：

$$\min_{C_1,C_2,C_3} J=\sum_{k=1}^{3}\sum_{i\in C_k}\left\|t_i-\mu_k\right\|^2 \tag{5}$$

式中，t_i为第i个样品的PCA得分向量，C_k和μ_k分别为第k个样品簇及其中心，K固定为3。标准化参数、PCA载荷和聚类中心均在每个外层训练折内重新计算。

$$\hat{y}_i=b_0+z_i^{\mathrm{T}}\beta \tag{6}$$

式中，$\hat{y}_i$为第$_i$个留出样品的Cr质量分数预测值，其数值以%表示，z_i为应用训练折参数后的特征向量，b_0和β分别为PLSR截距与回归系数；聚类辅助模型使用该样品所分配簇的局部参数。

$$R^2=1-\frac{\sum_{i=1}^{N}\left(y_i-\hat{y}_i\right)^2}{\sum_{i=1}^{N}\left(y_i-\bar{y}\right)^2} \tag{7}$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i-\hat{y}_i\right)^2} \tag{8}$$

模型同时报告决定系数(Coefficient of Determination, R^2)、均方根误差(Root Mean Square Error, RMSE)、平均绝对误差(Mean Absolute Error, MAE)、对称平均绝对百分比误差(symmetric Mean Absolute Percentage Error, sMAPE)和性能偏差比(Ratio of Performance to Deviation, RPD)，其中RPD=s_y/RMSE，s_y为Cr参考值的样本标准差。由于低Cr样品会放大相对误差，模型比较主要依据R^2、RMSE、MAE、RPD以及预测值与残差分布。

分别建立统一PLSR(Global PLSR)与K=3聚类辅助PLSR(Cluster-PLSR)。每个外层折留出1个完整样品，仅以其余26个样品重新选择积分窗口、计算标准化参数、拟合PCA并完成K-means++；留出样品应用训练折参数后分配至最近的聚类中心。对应局部训练簇再通过内层LOOCV选择PLS潜变

量数，重新拟合并预测留出样品。全局模型也在每个外层训练折内独立选择窗口和潜变量。27个外层预测汇总后，依据元素质量分数的非负物理边界，采用$\hat{y}^{+}=\max(0,\hat{y})$将负预测值统一置0，再计算$R^2$、RMSE、MAE、sMAPE和RPD；原始预测值同时保留用于追溯。上述设置保证留出样品不参与任何数据驱动步骤[28-30]。据此，结果部分依次检验特征表达、聚类结构及其对同维度定量模型的影响。

2　结果与讨论

2.1　样品组成与典型LIBS光谱特征

表1显示，27个样品既包含Cr和Ni含量均接近零的Fe基钢，也包含Cr质量分数高于20%的高合金钢，且在Mo和Ti含量上进一步分化。由于样品组成差异较大，共存元素与烧蚀条件可能影响Cr含量与光谱响应之间的对应关系。因此，选取Fe、Cr、Ni、Mn、Mo和Ti的特征谱线，以兼顾目标元素信息与基体差异的表征，并进一步评价其低维定量能力。样品平均光谱包含16 355个实测波长变量；图2展示了包含全部23条特征谱线的255～560 nm实测光谱区段，表明特征矩阵仅提取全谱中元素归属明确的少量区域。为使这些谱线可稳定积分，首先考察背景校正与Fe内标表达。

2.2　BEADS校正与Fe内标特征表达

图3a在各自真实强度坐标下给出原始光谱、BEADS基线和校正光谱。缓慢变化的背景被削弱，而离散发射峰的位置与峰形得以保留；图3b展示Fe Ⅰ 271.902 nm和Cr Ⅰ 425.433 nm附近的实际积分窗口。±0.07、±0.10、±0.12和±0.18 nm 4个候选积分半宽在27个外层折中的选择次数依次为16、2、8、1，表明±0.07 nm在多数情况下能兼顾峰信号与邻近谱线干扰，但仍保留逐折选择，以避免固定窗口的全样本偏倚。

基于270条位置平均光谱重新计算重复性，Fe Ⅰ 271.902、248.327、438.354和527.039 nm峰面积的相对标准偏差(Relative Standard Deviation，RSD)中位数分别为3.97%、4.40%、4.50%和6.12%，因此选取271.902 nm作为统一内标。需要强调的是，内标效果具有谱线依赖性：19条非Fe谱线中既有RSD降低的分析线，也有RSD上升的分析线，总体并不存在“所有谱线稳定性均提高”的结论，见图3c和3d。因此，本文选取该Fe线的依据是其自身在候选谱线中的重复性优势及Fe基体含义，内标矩阵的有效性最终仍由严格的外层预测评价，而不是由单一RSD指标预设。完成背景校正与内标特征构建后，进一步利用多元素响应检验样品是否形成具有组成意义的稳定分组。

2.3　*K*=3样品聚类及组成解释

23维FeRatio矩阵经标准化与PCA降维后，前3个主成分累计解释87.8%的特征方差。图4a以第一主成分(PC1)和第二主成分(PC2)展示样品的二维分布，实际聚类采用前3个主成分。主成分保留数敏感性分析表明，采用2～8个主成分时，K=3聚类均得到相同的11/9/7分组，与3个主成分结果相比，调整兰德指数(Adjusted Rand Index，ARI)均为1.000，说明该分组结构不依赖于特定的主成分保留数。K-means++聚类结果中，Cluster 1包括样品8～10、12～17、22和27；Cluster 2包括样品1～7、18和19；Cluster 3包括样品11、20、21和23～26。各簇的主要元素组成差异见图4b和表4。

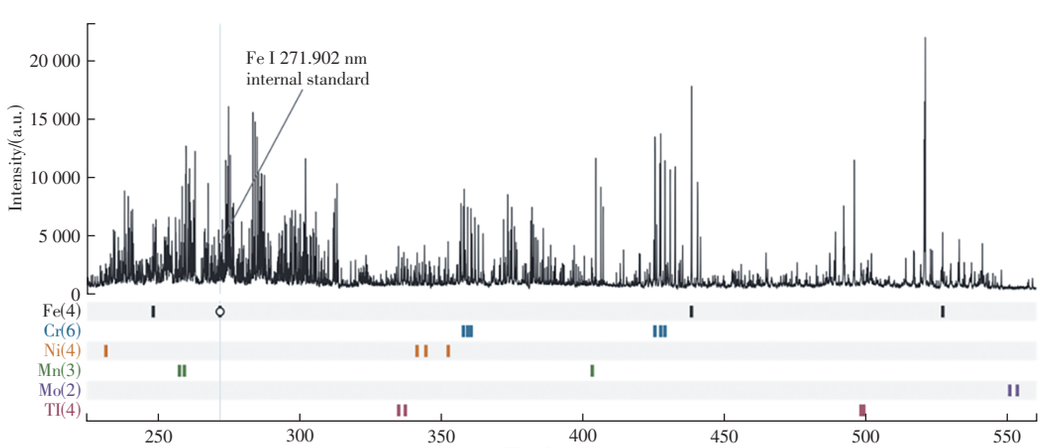


图2　典型钢铁样品实测LIBS光谱及23条特征谱线位置

Figure 2　Measured LIBS spectrum of a representative steel sample and positions of the 23 characteristic lines.

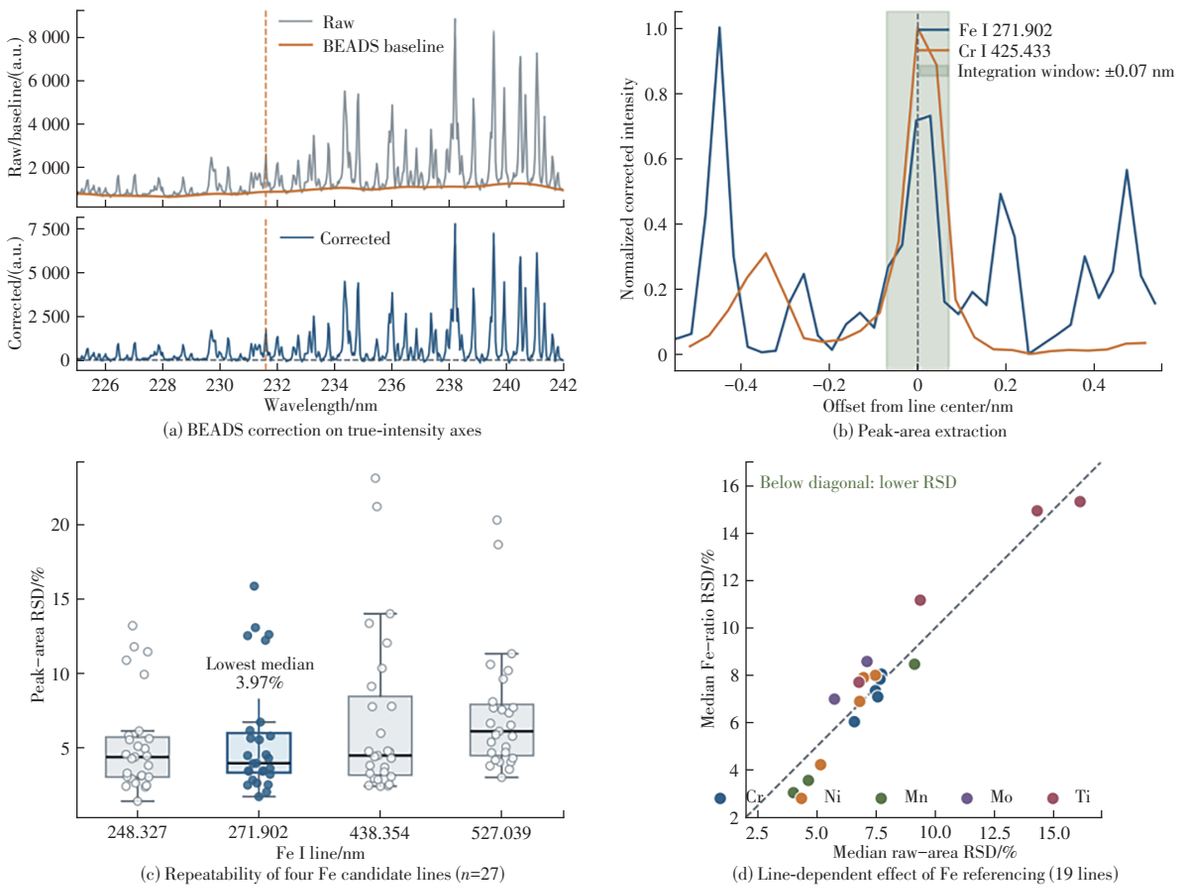


图3 基于实测光谱的BEADS基线校正、特征峰积分与Fe内标重复性诊断

Figure 3 BEADS baseline correction, peak-area integration and repeatability assessment for Fe referencing using measured spectra.

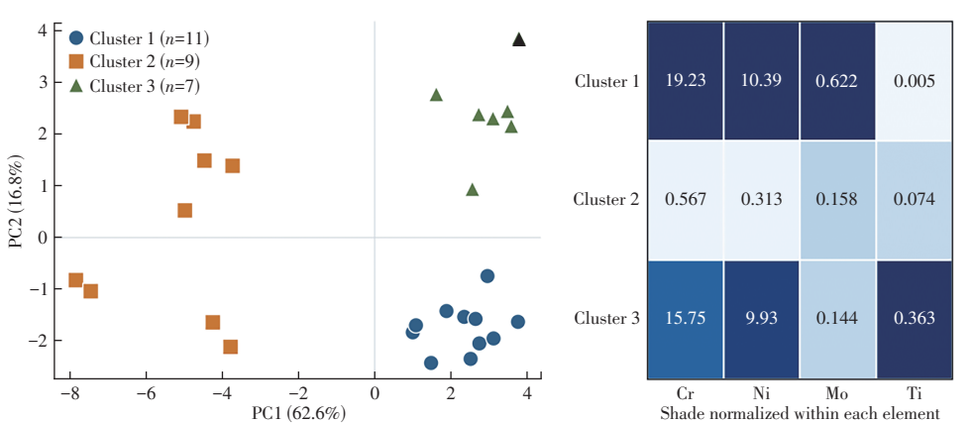


图4 *K*=3样品PCA分组及三类样品的主要元素组成

Figure 4 PCA grouping at *K*=3 and major-element compositions of the three sample groups.

表4 三类样品的主要合金元素平均质量分数

Table 4 Mean mass fractions of major alloying elements in the three clusters /%

Cluster	*n*	Cr	Ni	Mo	Ti
Cluster 1	11	19.235	10.394	0.622	0.005
Cluster 2	9	0.567	0.313	0.158	0.074
Cluster 3	7	15.746	9.929	0.144	0.363

在相同的3个主成分输入下，K-means++与GMM得到完全一致的样品分组，二者的轮廓系数(Silhouette Coefficient)、Calinski-Harabasz指数(CH)和Davies-Bouldin指数(DB)相同，ARI为1.000；HCA-Ward仅在少量边界样品的归属上存在差异，相对于K-means++的ARI为0.890，具体

结果见表5。进一步在27个外层留一折中，均使用26个训练样品重新完成标准化、PCA与聚类；与全样本描述性分组匹配后，各折训练样品的ARI均为1.000，留出样品类别匹配率为100%。上述结果表明，K=3分组能由不同聚类算法及不同留一训练集稳定复现，并非由单次初始化偶然形成。

表5　K=3条件下不同聚类算法的评价

Table 5　Evaluation of the clustering algorithms at K=3

Method	Silhouette	CH	DB	Cluster size	ARI
K-means++	0.513 1	47.22	0.732 2	11/9/7	1.000 (*vs.* GMM)
HCA-Ward	0.506 8	45.81	0.733 5	12/9/6	0.890 (*vs.* K-means++)
GMM	0.513 1	47.22	0.732 2	11/9/7	1.000 (*vs.* K-means++)

图4b及表4进一步给出三类样品的组成依据。Cluster 1中Cr和Ni的平均质量分数分别为19.235%和10.394%，表现为Cr-Ni高合金钢；Cluster 2的Cr和Ni均值分别仅为0.567%和0.313%，对应低Cr低Ni Fe基钢；Cluster 3仍具有较高的Cr和Ni含量，同时Ti均值为0.363%，明显高于Cluster 1中的相应值。由此，K=3并非为了将Cr浓度机械地划分为3段，而是以多元素谱线共同表征3种组成环境，为后续局部定量分析提供化学解释。在引入局部模型之前，仍需明确相同特征在统一全局关系下的性能基线。

2.4　不同光谱表达方式的全局定量性能

为区分变量轻量化与聚类分组的作用，首先比较不使用局部模型的3种全局表达。对所有模型采用相同的非负后处理后，FullSpectrum Global保留16 355个变量，在样品级LOOCV下取得R^2=0.9846、RMSE=1.082%和RPD=8.20，说明全谱保留了更充分的Cr相关综合信息。将输入压缩至23条谱线后，LineArea Global和FeRatio Global的RMSE分别为3.342%和3.322%，Fe内标仅带来0.57%的有限改善，但仍无法以单一低维全局关系覆盖跨钢种响应。表6同时显示：99.86%的变量压缩伴随一定的绝对精度损失，因此本文不将轻量化等同于精度提高。基于这一全局基线，下一步在完全相同的23维FeRatio输入下评价K=3局部建模能否补偿部分精度损失。

表6　不同全局PLSR方案经非负后处理后的样品级LOOCV性能

Table 6　Sample-wise LOOCV performance of the global PLSR strategies after nonnegative post-processing

Model	No. of variables	R^2	RMSE/%	MAE/%	sMAPE/%	RPD
FullSpectrum Global	16 355	0.984 6	1.082	0.743	36.23	8.20
LineArea Global	23	0.852 6	3.342	2.548	63.09	2.65
FeRatio Global	23	0.854 3	3.322	2.427	56.37	2.67

2.5　聚类辅助PLSR性能与结果分析

在完全相同的23维FeRatio输入下，经统一非负后处理后，K=3 Cluster-PLSR取得R^2=0.9531、RMSE=1.886%、MAE=1.314%和RPD=4.70，相对于FeRatio Global，RMSE降低43.24%，见表7。图5显示，局部模型明显缩小了全局低维模型在中高Cr区间的预测离散度。Cluster-PLSR各外层折的潜变量数(Latent Variable，LV)数为1～8，中位数为4；FeRatio Global为2～8，中位数为8，说明局部关系在多数折中可由更少的潜变量表达。该比较表明，组成感知分组能够在本文样品范围内改善低维全局关系。

表7　相同23维Fe内标特征下全局与聚类辅助PLSR经非负后处理后的性能

Table 7　Performance of global and cluster-assisted PLSR after nonnegative post-processing using the same 23 Fe-referenced features

Model	No. of variables	LV range (median)	R^2	RMSE/%	MAE/%	RPD
Global PLSR	23	2−8(8)	0.854 3	3.322	2.427	2.67
Cluster-PLSR (K=3)	23	1−8(4)	0.953 1	1.886	1.314	4.70

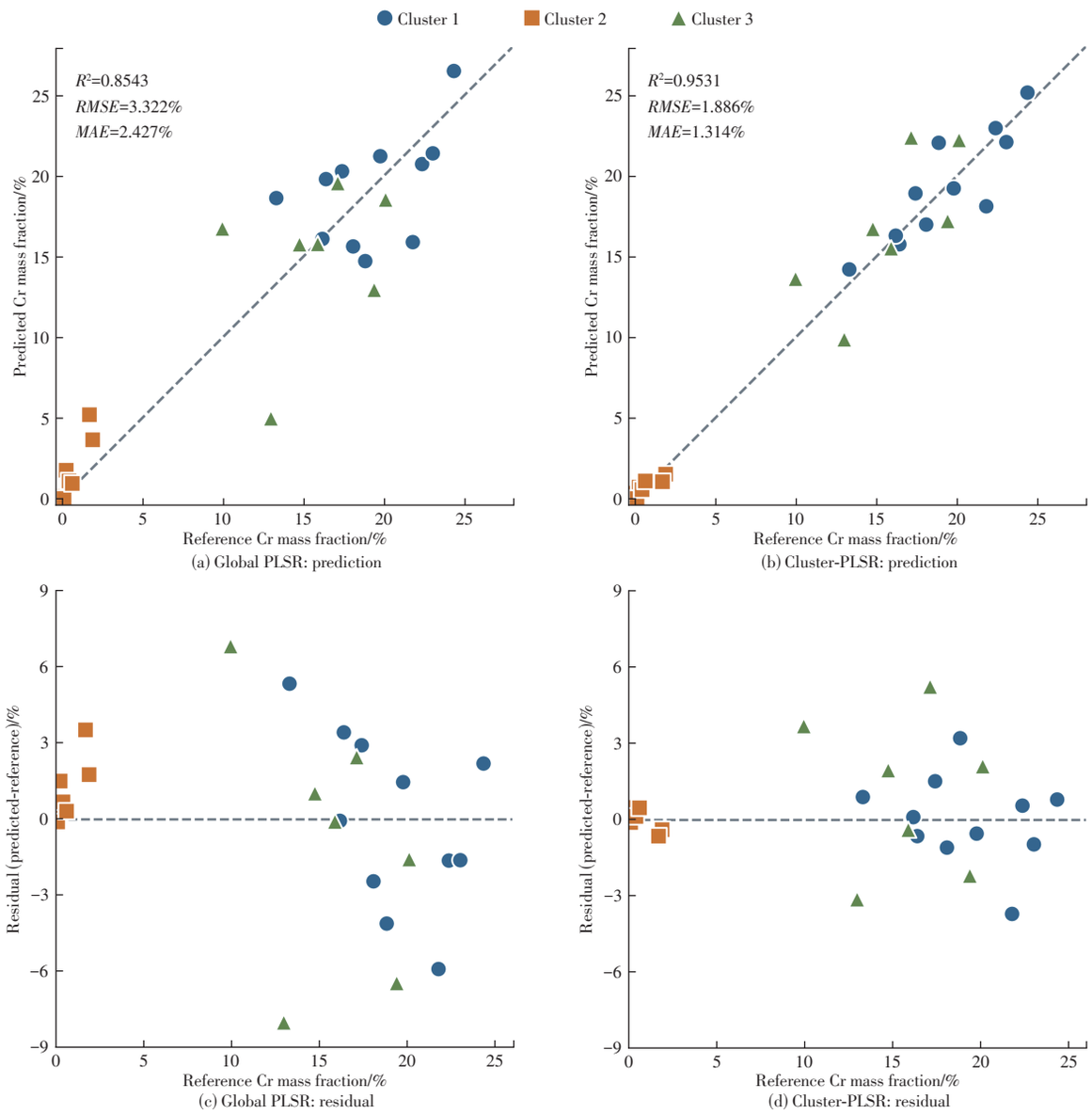


图5　相同23维Fe内标特征下全局PLSR与K=3聚类辅助PLSR的非负后处理预测及相应残差

Figure 5　Nonnegative post-processed predictions and residuals of global PLSR and K=3 cluster-assisted PLSR using the same 23 Fe-referenced features.

从方法作用机制看，本方法并不假定所有钢种共享同一条严格线性校准关系。BEADS校正用于削弱连续背景，FeRatio用于降低共同的烧蚀和采集波动，多元素特征同时表征Cr响应及共存元素的差异；随后，K=3聚类将组成相近的样品限制在较一致的局部空间内，再由PLSR处理局部共线性响应。该流程能够减弱跨基体统一建模造成的系统偏差，但并不等同于从物理上消除自吸收或谱线干扰。

为检验分组是否由目标元素直接驱动，进一步从聚类门控中移除6条Cr I谱线，但在PLSR预测器中保留Cr特征。全样本分组仍为11/9/7，ARI=1.000；27个外层折的类别分配、窗口选择和预测结果均保持不变。这说明Fe、Ni、Mn、Mo和Ti特征已足以维持三类组成结构，聚类增益并非简单的Cr浓度分段。另一方面，Cluster-PLSR的绝对性能仍低于FullSpectrum Global，且sMAPE受接近零的Cr含量放大。因此，本方法适合被表述为可解释、低维的组成感知校准，而非全谱模型的无条件替代；其性能结论适用于本文27个钢铁样品所覆盖的组成范围。表8对本文方法与4项代表性国外LIBS定量研究进行定性比较。由于各研究的样品基体、目标元素、仪器条件及验证方式不同，本文不直接比较R^2或RMSE等数值，而重点比较特征表达、基体效应处理方式及方法边界。

表8　本文方法与代表性国外LIBS定量研究的方法学比较

Table 8　Methodological comparison of the present method with representative international LIBS quantitative studies

Ref.	Input and model	Treatment of matrix effects	Advantages	Limitations
[8]	Selected analyte lines; calibration curves with iterative interelement correction	Corrects composition-dependent interelement interference	Direct and physically interpretable; residual deviations reduced	Requires certified calibration samples and element-specific correction
[13]	Low-resolution spectral regions/ first derivatives; PLSR	PLSR extracts unresolved features; derivatives reduce baseline difficulty	Compatible with compact low-resolution instruments	Focused on low-alloy steels; no explicit composition grouping
[14]	Spectral-image (HOG) features; PLSR	Multidimensional image features describe composition-dependent spectral patterns	Strong feature representation for stainless steels	Image conversion and feature extraction reduce direct line interpretability
[27]	LIBS spectra; UMAP+ K-means+local linear/ridge/ PLS models	Similar sample matrices are clustered before local calibration	Blind tests across distinct drill cores; improved low-concentration calibration	Developed for Li-bearing geological samples; workflow is relatively complex
This work	23 Fe-referenced characteristic lines; PCA+ K-means++local PLSR	Composition-aware clustering followed by local calibration	99.86% variable reduction and explicit elemental interpretation	Conclusions are limited to the composition range of the 27 steel samples

3　结论

本文面向复杂钢种中全谱变量冗余及统一低维关系难以覆盖基体差异的问题，以27条样品平均光谱为独立单元构建了Fe、Cr、Ni、Mn、Mo和Ti共23条特征谱线矩阵。Fe Ⅰ 271.902 nm在4条候选Fe线中具有最低峰面积RSD中位数，但真实位置光谱同时表明内标作用具有谱线依赖性，因此其价值由后续严格外层预测而非普遍降RSD假设来评价。

K-means++与GMM均得到11、9和7个样品的相同K=3分组，外层留一训练折的分组ARI均为1.000；3类样品分别对应Cr-Ni高合金钢、低Cr低Ni Fe基钢和含Ti特殊合金钢。在每个外层折内重新完成窗口选择、标准化、PCA、聚类、留出样品归类和PLS潜变量选择后，外层预测经统一非负后处理后，Cluster-PLSR获得R^2=0.9531、RMSE=1.886%、MAE=1.314%和RPD=4.70，相对于相同23维FeRatio Global模型的RMSE降低43.24%。

全谱统一模型仍具有最高绝对预测性能，但需要16 355个变量；23维方案使变量数减少99.86%，并以元素归属明确的特征保留了有效的Cr定量信息。全局与局部低维模型的对比进一步表明，组成感知分组能够补偿部分特征压缩造成的预测能力损失。因此，本文方法的主要价值在于兼顾变量轻量化、谱线物理可解释性和所研究组成范围内的定量能力，可为复杂钢种的快速LIBS分析提供一种清晰、可实施的建模思路。

参考文献

[1] Hahn D W, Omenetto N. Laser-induced breakdown spectroscopy (LIBS), part Ⅰ: review of basic diagnostics and plasma-particle interactions: still-challenging issues within the analytical plasma community[J]. Applied Spectroscopy, 2010, 64(12): 335A-366A.

[2] Hahn D W, Omenetto N. Laser-induced breakdown spectroscopy (LIBS), part Ⅱ: review of instrumental and methodological approaches to material analysis and applications to different fields[J]. Applied Spectroscopy, 2012, 66(4): 347-419.

[3] El Haddad J, Canioni L, Bousquet B. Good practices in LIBS analysis: review and advices[J]. Spectrochimica Acta Part B: Atomic Spectroscopy, 2014, 101: 171-182.

[4] Bol′shakov A A, Yoo J H, Liu Chunyi, et al. Laser-induced breakdown spectroscopy in industrial and security applications[J]. Applied Optics, 2010, 49(13): C132-C142.

[5] Noll R, Bette H, Brysch A, et al. Laser-induced breakdown spectrometry: applications for production control and quality assurance in the steel industry[J]. Spectrochimica Acta Part B: Atomic Spectroscopy, 2001, 56(6): 637-649.

[6] Noll R, Fricke-Begemann C, Brunk M, et al. Laser-induced breakdown spectroscopy expands into industrial applications[J]. Spectrochimica Acta Part B: Atomic Spectroscopy, 2014, 93: 41-51.

[7] Aragón C, Aguilera J A, Peñalba F. Improvements in quantitative analysis of steel composition by laser-induced breakdown spectroscopy at atmospheric pressure using an infrared Nd: YAG laser[J]. Applied Spectroscopy, 1999,

53(10): 1259-1267.

[8] Vrenegor J, Noll R, Sturm V. Investigation of matrix effects in laser-induced breakdown spectroscopy plasmas of high-alloy steel for matrix and minor elements[J]. Spectrochimica Acta Part B: Atomic Spectroscopy, 2005, 60(7/8): 1083-1091.

[9] Shah M L, Pulhani A K, Gupta G P, et al. Quantitative elemental analysis of steel using calibration-free laser-induced breakdown spectroscopy[J]. Applied Optics, 2012, 51(20): 4612-4621.

[10] Zeng Qingdong, Guo Lianbo, Li Xiangyou, et al. Quantitative analyses of Mn, V, and Si elements in steels using a portable laser-induced breakdown spectroscopy system based on a fiber laser[J]. Journal of Analytical Atomic Spectrometry, 2016, 31(3): 767-772.

[11] Li Kuohu, Guo Lianbo, Li Jiaming, et al. Quantitative analysis of steel samples using laser-induced breakdown spectroscopy with an artificial neural network incorporating a genetic algorithm[J]. Applied Optics, 2017, 56(4): 935-941.

[12] Afgan M S, Hou Zongyu, Wang Zhe. Quantitative analysis of common elements in steel using a handheld μ-LIBS instrument[J]. Journal of Analytical Atomic Spectrometry, 2017, 32(10): 1905-1915.

[13] Kim H, Nam S H, Han S H, et al. Laser-induced breakdown spectroscopy analysis of alloying elements in steel: partial least squares modeling based on the low-resolution spectra and their first derivatives[J]. Optics & Laser Technology, 2019, 112: 117-125.

[14] Yan Jiujiang, Hao Zhongqi, Zhou Ran, et al. A quantitative analysis method assisted by image features in laser-induced breakdown spectroscopy[J]. Analytica Chimica Acta, 2019, 1082: 30-36.

[15] Zhang Tianlong, Xia Donghui, Tang Hongsheng, et al. Classification of steel samples by laser-induced breakdown spectroscopy and random forest[J]. Chemometrics and Intelligent Laboratory Systems, 2016, 157: 196-201.

[16] Pořízka P, Klus J, Prochazka D, et al. Laser-Induced Breakdown Spectroscopy coupled with chemometrics for the analysis of steel: the issue of spectral outliers filtering[J]. Spectrochimica Acta Part B: Atomic Spectroscopy, 2016, 123: 114-120.

[17] 吴卓, 苏晓慧, 范博文, 等. 不同特征选择方法结合激光诱导击穿光谱量化不锈钢镍、钛和铬元素含量[J]. 光谱学与光谱分析, 2024, 44(12): 3339-3346.
Wu Zhuo, Su Xiaohui, Fan Bowen, et al. Different feature selection methods combined with laser-induced breakdown spectroscopy were used to quantify the contents of nickel, titanium and chromium in stainless steel[J]. Spectroscopy and Spectral Analysis, 2024, 44(12): 3339-3346.

[18] 汪倩, 钟绪, 李国文, 等. 基于LIBS-CNN协同优化的结构金属材料中痕量Cr元素的定量分析[J]. 中国激光, 2025, 52(22): 2211003. DOI: 10.3788/CJ250598.
Wang Qian, Zhong Xu, Li Guowen, et al. Quantitative analysis of trace chromium element in structural metal materials based on LIBS-CNN collaborative optimization[J]. Chinese Journal of Lasers, 2025, 52(22): 2211003. DOI: 10.3788/CJ250598.

[19] 李婉雪, 何亚雄, 李杨, 等. 激光诱导击穿光谱结合机器学习算法的同基合金快速识别研究[J]. 激光与光电子学进展, 2024, 61(17): 1730001. DOI: 10.3788/LOP232417.
Li Wanxue, He Yaxiong, Li Yang, et al. Rapid identification of homogeneous alloys based on laser-induced breakdown spectroscopy combined with machine-learning algorithms[J]. Laser & Optoelectronics Progress, 2024, 61(17): 1730001. DOI: 10.3788/LOP232417.

[20] Brunnbauer L, Gajarska Z, Lohninger H, et al. A critical review of recent trends in sample classification using Laser-Induced Breakdown Spectroscopy (LIBS)[J]. TrAC Trends in Analytical Chemistry, 2023, 159: 116859. DOI: 10.1016/j.trac.2022.116859.

[21] Zheng Peichao, Chen Guanghui, Gu Weilun, et al. Upgrade from signal enhancement to plasma modulation for LIBS signal improvement[J]. TrAC Trends in Analytical Chemistry, 2026, 197: 118709. DOI: 10.1016/j.trac.2026.118709.

[22] Tognoni E, Cristoforetti G, Legnaioli S, et al. Calibration-free laser-induced breakdown spectroscopy: state of the art[J]. Spectrochimica Acta Part B: Atomic Spectroscopy, 2010, 65(1): 1-14.

[23] Zorov N B, Gorbatenko A A, Labutin T A, et al. A review of normalization techniques in analytical atomic spectrometry with laser sampling: from single to multivariate correction[J]. Spectrochimica Acta Part B: Atomic Spectroscopy, 2010, 65(8): 642-657.

[24] Guezenoc J, Gallet-Budynek A, Bousquet B. Critical review and advices on spectral-based normalization methods for LIBS quantitative analysis[J]. Spectrochimica Acta Part B: Atomic Spectroscopy, 2019, 160: 105688. DOI: 10.1016/j.sab.2019.105688.

[25] Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics[J]. Chemometrics and Intelligent Laboratory Systems, 2001, 58(2): 109-130.

[26] Ning Xiaoran, Selesnick I W, Duval L. Chromatogram baseline estimation and denoising using sparsity (BEADS)[J]. Chemometrics and Intelligent Laboratory

Systems, 2014, 139: 156-167.
[27] Silva N A, Capela D, Ferreira M, et al. Towards robust calibration models for laser-induced breakdown spectroscopy using unsupervised clustered regression techniques[J]. Results in Optics, 2022, 9: 100245. DOI: 10.1016/j.rio.2022.100245.
[28] Reunanen J. Overfitting in making comparisons between variable selection methods[J]. Journal of Machine Learning Research, 2003, 3: 1371-1382.
[29] Varma S, Simon R. Bias in error estimation when using cross-validation for model selection[J]. BMC Bioinformatics, 2006, 7(1): 91. DOI: 10.1186/1471-2105-7-91.
[30] Cawley G C, Talbot N L C. On over-fitting in model selection and subsequent selection bias in performance evaluation[J]. Journal of Machine Learning Research, 2010, 11: 2079-2107.
[31] Kramida A, Ralchenko Y, Reader J, et al. NIST Atomic Spectra Database (version 5.12)[DB/OL]. Gaithersburg: National Institute of Standards and Technology, 2024 [2026-07-28]. https://physics.nist.gov/asd.

Quantitative LIBS Analysis of Chromium in Steels Assisted by a Characteristic Spectral-line Matrix and Sample Clustering

Zheng Peichao, Zhang Wenshuo, Wang Jinmei*, Dai Panyang, Li Siqi, Zhang Yi, Huang Haoxuan
(*School of Electronic Science and Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China*)

Abstract Matrix-dependent responses and high-dimensional redundancy hinder quantitative laser-induced breakdown spectroscopy (LIBS) analysis of compositionally diverse steels. This study aimed to evaluate whether composition-aware clustering can improve a physically interpretable low-dimensional model for chromium determination. Twenty-seven sample-mean spectra were treated as independent modeling units. After correction using baseline estimation and denoising with sparsity (BEADS), peak areas of 23 Fe, Cr, Ni, Mn, Mo and Ti lines were extracted and referenced to Fe Ⅰ 271.902 nm. Integration-window selection, standardization, principal component analysis (PCA), K=3 clustering, held-out-sample assignment and partial least squares (PLS) latent-variable selection were reconstructed inside every outer sample-wise leave-one-out cross-validation (LOOCV) training fold. Negative outer-fold predictions were mapped to zero according to the nonnegative physical boundary before metric calculation. K-means++ and Gaussian mixture model (GMM) produced the same 11/9/7 grouping, and the adjusted Rand index of the training-sample partition was 1.000 in all 27 outer folds. Fe I 271.902 nm had the lowest median peak-area relative standard deviation (RSD) among four Fe candidates, whereas the effect of Fe referencing on repeatability was line-dependent. K=3 cluster-assisted partial least squares regression (PLSR) achieved a coefficient of determination (R^2) of 0.953 1, a root mean square error (RMSE) of 1.886%, a mean absolute error (MAE) of 1.314% and a ratio of performance to deviation (RPD) of 4.70, reducing RMSE by 43.24% relative to global PLSR using the same 23 variables. The variable count was reduced by 99.86% compared with the 16 355-variable full-spectrum model. Composition-aware local modeling recovered much of the predictive capability lost by low-dimensional global modeling while retaining explicit elemental assignments. The full-spectrum model remained the most accurate in absolute terms; therefore, the principal advantage of the proposed strategy is lightweight and interpretable calibration within the studied composition range.

Keywords laser-induced breakdown spectroscopy; steel analysis; chromium; characteristic spectral lines; clustering; partial least squares regression

HIGHLIGHTS

1) The variable count was reduced by 99.86% using a 23-dimensional matrix.
2) Three interpretable steel compositions were represented by the K=3 grouping result.
3) The RMSE was reduced by 43.24% via cluster-assisted PLSR under the identical dimensionality.