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
2068～2078

doi: 10. 20236/j. CJIAC. 2026. 11. 010

郑一博，杨慧珂，王震宇，等. 基于激光诱导击穿光谱(LIBS)预处理优化的高鲁棒性煤质定量分析研究[J]. 中国无机分析化学，2026，16(11)：2068-2078.

Zheng Yibo, Yang Huike, Wang Zhenyu, et al. High-robustness Quantitative Analysis of Coal Quality Based on LIBS Spectral Preprocessing Optimization[J]. Chinese Journal of Inorganic Analytical Chemistry, 2026, 16(11): 2068-2078.

基于激光诱导击穿光谱(LIBS)预处理优化的高鲁棒性煤质定量分析研究

郑一博[1]　杨慧珂[1]　王震宇[1]　魏文涛[1]　李　安[1*]
张新宇[2]　刘瑞斌[3]　王占荥[4]

(1. 河北地质大学，河北省光电信息与地球探测技术重点实验室，石家庄 050031；
2. 齐鲁工业大学(山东科学院) 海洋仪器仪表研究所，山东 青岛 266100；
3. 北京理工大学 物理学院，北京 100081；
4. 邢台劳特斯新材料科技有限公司，河北 邢台 054000)

摘　要　激光诱导击穿光谱(LIBS)技术定量分析精度和准确度依赖于对原始光谱数据的有效预处理。光谱数据的稳定性、数据对样本的代表性是决定物质定量分析模型性能的基础。研究量化分析中隐含光谱的统计特征和潜在的物理机理更需要高质量和高重复性的光谱数据。本文提出基于自适应光谱特征的可变窗口平移平滑去背景方法以及细分通道内两步归一化的光谱预处理优化方法，基于此建立了煤质灰分、全硫和热值的定量分析模型，并对模型的鲁棒性和泛化能力进行了评估检验。盲测结果表明，该方法显著提升了模型的预测精度与鲁棒性。尤其对于热值，盲测集预测均方根误差(RMSE)从1 283. 4 J/g大幅降至407. 6 J/g，降幅超过68%，决定系数(R^2)从0. 818 3提升至0. 981 7。同时，灰分与全硫的预测性能也得到有效优化，对应盲测集的RMSE分别从1. 2215%和0. 1741%降低至0. 8924%和0. 1648%。本研究从光谱预处理算法上提升了数据的稳定性和代表性，定量分析模型可以有效学习到光谱数据和回归量之间稳定的和固有的物理关联，增强了模型的鲁棒性和预测精度。该方法可为LIBS技术在不同场景下的应用提供算法支撑，也为定量分析的深入研究奠定可靠的数据基础。

关键词　激光诱导击穿光谱；光谱预处理；定量分析；煤；偏最小二乘法

收稿日期：2025-12-16　**修回日期**：2026-01-13　**录用日期**：2026-01-17　**在线出版日期**：2026-06-22

Received：2025-12-16　**Revised**：2026-01-13　**Accepted**：2026-01-17　**Online published**：2026-06-22

基金项目：国家自然科学基金资助项目(12502420)；河北省教育厅科学研究项目(BJ2025077)；河北省科技计划项目(25347618D)；河北地质大学研究生教育项目(CXZZDD202626)

Fund：Supported by the National Natural Science Foundation of China (12502420)；Science Research Project of Hebei Education Department (BJ2025077)；S & T Program of Hebei, China (25347618D)；Graduate Education Programs of Hebei GEO University (CXZZDD202626)

作者简介：郑一博，男，教授，主要从事激光光谱与智能感测技术方面的研究。E-mail：yibo_zheng@hgu. edu. cn

Author：Zheng Yibo, male, professor, mainly engaged in laser induced breakdown spectroscopy and intelligent sensing technology. E-mail：yibo_zheng@hgu. edu. cn

*通信作者：李安，男，助理研究员，主要从事激光诱导击穿光谱方面的研究。E-mail：anli@hgu. edu. cn

*Corresponding author：Li An, male, research assistant, mainly engaged in laser induced breakdown spectroscopy. E-mail：anli@hgu. 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.

要点

1）提出了可变窗口平移平滑光谱背景拟合及去除方法。

2）提出了分通道两步归一化光谱信号稳定性提升算法。

3）研究了基于光谱预处理优化的定量分析算法。

中图分类号：O657.38　　**文献标志码：**A　　**文章编号：**2095-1035(2026)11-2068-11

激光诱导击穿光谱(LIBS)是一种基于原子分析光谱的化学计量学分析方法，从20世纪80年代开始发展至今已有40余年[1]，发展迅速、成果丰富。因LIBS分析技术具有快速、实时、无污染等优点，对样品的物理状态(固、液、气)无特殊要求[2-4]，可以实现全元素分析等优点，目前已经成为一种诊断物质元素含量及其物化性质的高效分析方法，并且成功应用于煤质检测[5-6]、生物医疗[7-8]、珠宝分析[9-10]以及环境监测[11-12]等领域。但是LIBS光谱的噪声和信号不稳定性一直制约着该技术的工业应用和分析精度的进一步提升。

光谱的稳定性决定了数据对样本的代表性，提升光谱信号稳定性的预处理方法主要是基于光谱背景的去除及光谱强度的归一化。LIBS光谱背景去除的目的是消除有效谱峰下方的连续背景而不引起谱峰信息的丢失(峰形、峰位不变)。Motto-Ros等[13]评估了LIBS光谱在背景去除前后对量化分析的影响，经过背景去除后量化分析的性能和预测精度得到一定的提升；Képeš等[14]提出了一种通过计算某条谱线在背景去除前后元素检出限(Limit of Detection，LOD)的变化来评估背景基线拟合优劣的方法，作者综合对比了四种背景去除方法对特征谱线的LOD影响。1)Yaroshchyk方法[15]，即无模型背景去除算法，该方法基于小波变换，无需阈值选择可实现LIBS光谱的背景拟合，且对含噪声的光谱信号也具有较好的背景拟合效果，利用偏最小二乘回归(Partial Least Squares Regression，PLSR)和留一交叉验证的方法评估了该背景去除算法对定量分析的影响，他们发现经过该方法处理的光谱对铁矿石中铁成分的预测精度提升了1%；2)Galloway方法[16]，利用小波函数迭代的方式，不断对重建光谱与输入光谱进行比较，直到两个光谱的差异值达到设置的阈值，即可实现对光谱背景的拟合；3)Gornushkin方法[17]，该方法将光谱分割为若干个窗口，并在窗口内找到最小值，并利用高阶多项式拟合的方法来实现光谱背景的拟合，该方法也称为窗口平移平滑去背景方法；4)手动背景选择，该方法并不适用大批量光谱数据的处理。目前，光谱去背景的算法主要集中在利用小波变换、窗口平移平滑实现对背景连续信号和光谱特征信号的分离，此两类方法对背景识别的效果较好，可以避免背景拟合对一些分子峰的有效带状宽光谱产生影响，是目前LIBS光谱信号背景去除中最多采用的两种方法。

光谱信号的波动对LIBS分析结果有显著影响，而光谱信号经过强度归一化后，可以进一步降低光谱强度的波动，有大量的研究将归一化前后的光谱信号分别用作定量分析，并对比精度提升效果。Feng等[18]对比了分段光谱归一化的方法和全谱面积归一化方法，相对标准偏差RSD由30%降至10%。熊凯等[19]使用最大最小归一化方法将谱线组间RSD由15%降至5%。Li等[20]提出自适应加权归一化方法，结合线性加权网络提高混合基质条件下Cr检测精度。Guezenoc等[21]系统梳理了背景值归一化、总面积归一化、内标法和标准正态变量法(Standard Normal Variate，SNV)的原理。背景值归一化是将目标发射线的强度除以背景辐射强度。Dell′Aglio等[22]采用背景归一化处理方法，在铬、铜、铅等元素检测中得到了成功验证；总面积归一化是将分析物对应发射线的辐射强度除以全谱的积分面积。Oh等[23]对浆料中Fe的原子发射线进行全谱归一化，平均RSD强度由34.74%降至5.92%；内标法是将待测物的发射线强度与内标物对应的发射谱线的峰强度进行比对。Cao等[24]提出了一种改进的内标归一化方法，选择Fe原子谱线作为内标来计算其他元素(如Cr)的相对浓度分布，消除了光谱强度波动对分析结果的影响；SNV是将光谱以平均值为中心进行校正，即使用波长原始强度减去光谱平均强度，然后将矫正后的光谱除以原始光谱的标准差[25]。Ismaël等[26]将SNV归一化方法成功应用于土壤的LIBS数据中，显著降低了标准差。以上方法均为利用光谱内部信息进行归一化，近年来，基于外部信号的归一化策略逐渐受到关注，例如，Wang等[27]提出监测激光输出能量和等离子体的关键参数与反向传播的神经网络(KPBP)方法，使Si、Al的RSD分别降至3.20%和3.40%。Wang等[28]利用收集到的声波信号和等离子体图像组合归一化应用到水下LIBS分析，结果显示，Mn的

RSD从6.12%降至3.53%，更有效地降低了信号波动。总之，多项研究结果证实合理的归一化能有效降低光谱信号间的波动性，显著改善光谱一致性并提升模型的鲁棒性。

综上，LIBS信号的预处理是环环相扣、彼此影响，光谱的归一化效果一定程度上也取决于光谱背景拟合的效果，对LIBS光谱信号进行去背景和归一化处理可以一定程度上提升光谱信号稳定性，而且光谱信号稳定性提升对定量分析模型的性能提升具有积极影响。然而，现有的预处理方法并没有针对光谱特征进行优化，导致同样参数的预处理对分子带状峰和原子离子线状峰产生不同程度影响[23,29]，有些预处理可能会导致谱峰变形，极大地削弱了谱线携带的物理信息[16,30]。因此，在建立量化分析模型前有必要对光谱进行细致的预处理，并对预处理效果进行评估。本文对LIBS光谱预处理中的去背景和归一化方法进行了优化，并将其和定量分析有机地结合起来，以煤的灰分、全硫及热值为对象，系统研究了光谱预处理方法优化后的信号质量提升对定量分析结果的影响，为LIBS在复杂场景中应用提供了一种算法支撑。

1 实验部分

1.1 实验样品

本文研究中以煤为研究对象，所用样本均为标准煤样粉末，粒径均小于0.2 mm，共121个样本，样品的相关参数和统计情况如表1所示。利用200 kN压力机将4 g粉末状样本压制成直径30 mm、厚度4 mm的饼状样本作为待测样。

表1 煤样品中各参数的真值

Table 1 The reference data of coal samples

No.	Ash/%	Sulfur/%	Calorific values/(J·g^{-1})
1	12.56	0.64	27 905
2	12.07	2.20	29 084
3	11.06	2.26	29 406
4	25.36	0.98	23 151
5	8.82	1.21	28 462
6	8.62	0.61	30 376
7	7.30	0.78	29 145
8	4.94	0.63	30 687
⋮	⋮	⋮	⋮
120	7.51	1.01	31 171
121	14.99	0.62	27 064
[min, max]	[3.47, 61.70]	[0.23, 4.76]	[14 364, 31 632]
Mean	24.56	1.21	25 418
Standard deviation	16.55	0.79	4 952

1.2 LIBS实验装置

实验所用LIBS测试装置如图1所示，该装置为典型的LIBS装置，光学系统采用收发光路异轴的形式，其中，R为1 064 nm介质膜反射镜，L_1为熔石英聚焦透镜(焦距f=100 mm)，L_2(f=60 mm)、L_3(f=60 mm)组成光信号采集系统。将等离子体光信号耦合至光纤。激光器(Nd: YAG，1 064 nm)输出能量100 mJ、脉宽5 ns；光谱仪(Avantes，190～655 nm)为四通道光纤光谱仪，设置积分时间为30 μs，光谱仪数据采集与激光器Q信号之间的延迟时间为1 μs。实验中激光器的频率设置为5 Hz，样本放置在三维移动样品台(3D-MS)上，每个烧蚀点采集5个光谱，随后控制载物台步进1 mm，进行下一点的数据采集，最终在样本上形成6×7的激光烧蚀阵列，共采集光谱210组。

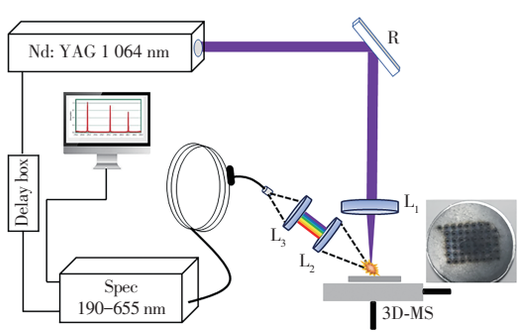


图1 收发光路异轴LIBS测试系统：反射镜R和聚焦透镜L_1组成激发光路，透镜L_2和L_3组成信号收集光路，3D-MS为三维移动样品台

Figure 1 The LIBS measurement system based on the off-axis optical path: The plasma is stimulated by the optical components R and L_1, and the lenses L_2 and L_3 constitute the light collection part, the 3D-MS is three-dimensional motion stage.

1.3 模型评价

本文采用的定量分析模型建立方法主要基于主成分分析的偏最小二乘法(Principal Components Analysis Based Partial Least Squares Regression，PCA-PLS)，对模型预测性能的评价依靠训练集(Training Set)、验证集(Validation Set)的拟合优度R^2、均方根误差(Root Mean Square Error，RMSE)和平均绝对误差(Mean Absolute Error，MAE)，对模型鲁棒性和泛化能力的评价主要由模型在盲测数据集(Blind Set)上的均方根误差RMSE和MAE决定。各评价指标的计算如式(1)～(3)所示。

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

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

$$\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right| \tag{3}$$

其中，y_i是煤中某参数的真值或参考值，$\hat{y}_i$是煤中某参数的预测值，$\overline{y}_i$是参考值的平均值，n为样品总个数，该值在训练集、验证集和盲测集中各异。

2　光谱预处理算法优化

2.1　可变窗口平移平滑去背景

本文中对光谱数据的预处理方法主要集中在两个方面，连续背景拟合去除与光谱归一化。其中连续背景的拟合最为重要，拟合结果的优劣直接影响光谱背景面积归一化的效果。通过研究发现，通过窗口平移平滑方法可快速精准实现光谱连续背景拟合，利用通道内拟合的光谱背景面积对光谱数据归一化，光谱数据的稳定性可得到一定提升[31]。然而，传统窗口平移平滑方法通常采用固定窗口宽度和步进参数，在处理不同谱峰宽度特征的光谱时存在局限性。具体而言，在光谱背景拟合和去除方面，受限于固定窗口宽度和滑动步长，在寻找窗口内光谱极值时易遗漏局部最小值，导致拟合背景过大或过小，进而致使去除背景后的光谱特征峰变形和失真。此外，在处理较宽分子峰或原子峰时，普通窗口平移平滑方法因窗口宽度不变，相对较窄窗口易导致最小值频繁出现在特征峰上，背景拟合削弱了这些光谱特征的强度。如图 2a、2b 所示，分子光谱特征 CN($\Delta\nu$=0) 的背景强度被过多估计导致去背景后其形状失真。上述异常处理均在后续归一化步骤中传导，对光谱稳定性和相对强度产生不同程度影响，降低建模数据质量，从而限制定量分析模型精度的提升。

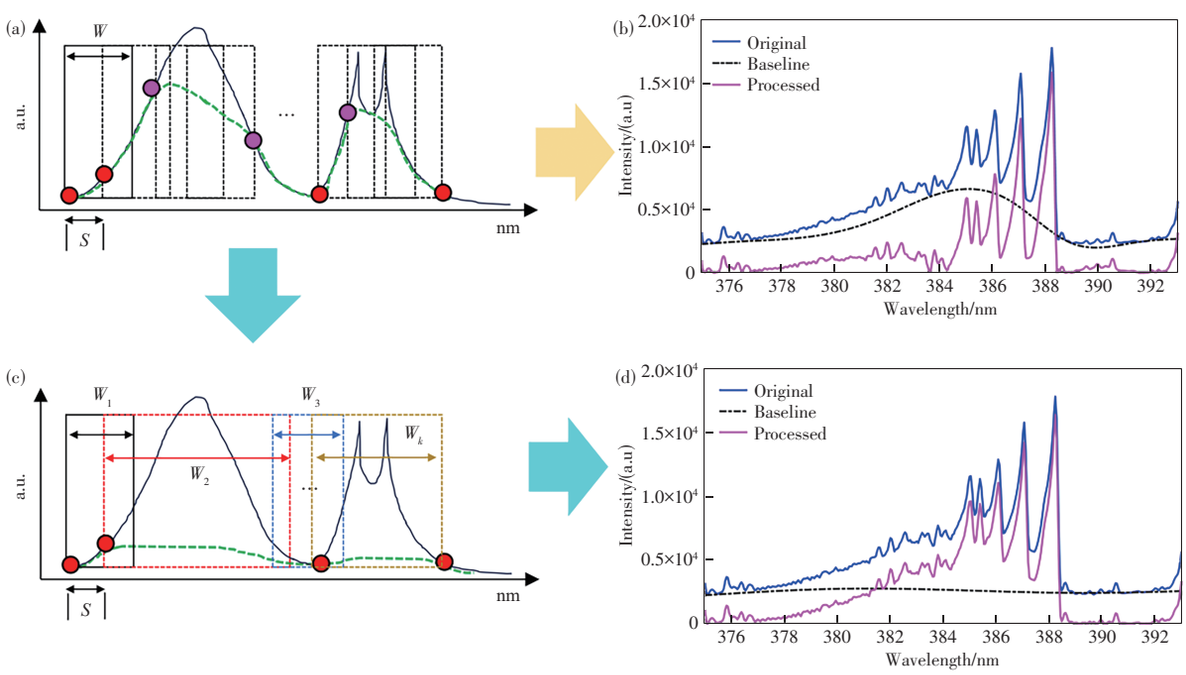


图 2　光谱去背景方法优化：(a)(b) 窗口平移平滑方法；(c)(d) 可变窗口平移平滑方法

Figure 2　Optimization of spectral background removal methods: (a)(b) Window translation smoothing method; (c)(d) Variable window translation smoothing method.

为进一步提升 LIBS 光谱信号的稳定性，本文提出一种可变窗口平移平滑的 LIBS 光谱背景拟合方法，该方法核心思想是在保持滑动窗口平移方式和中心位置不变的前提下，引入窗口宽度自适应调节机制，以提高窗口平移平滑法对不同谱峰宽度特征的适应能力。其具体操作方法：

1）将光谱强度数据视为 N 个数据点群，即所有光谱强度是 CCD 像素点上光强度的表现，标记为 $I(X)$，其中 X 为像素点索引；

2）分割数据点群：将 N 个数据点分为 n 个小点群，确定平滑窗口宽度为 W_n，n 个数据点群中的 W_n 可根据峰特征改变大小；

3）确定可变窗口大小：标记 CN、C_2 和 H_α 光谱特征，确定每个光谱特征的波长范围（像素点范围）为

$\left[P_{\min}^{Z}, P_{\max}^{Z}\right]$，满足$\Delta P^{Z}=P_{\max}^{Z}-P_{\min}^{Z}$；

4）寻找极值点并自适应改变滑动窗口大小：将窗口步进宽度固定为S，找出n组数据中每组中的最小值L_i，当滑动窗口W_k（k=1，2，3，…，n）遇到特征Z时，自动将窗口宽度变为$W_k+\Delta P^{Z}$，其中，窗口中心位置由基础窗口宽度W_k决定（即窗口左右边界像素索引的算术平均值），并在窗口宽度自适应扩展为$W_k+\Delta P^{Z}$的过程中保持不变；可变窗口始终以该基础窗口中心为对称轴进行双侧扩展。随后确定该窗口内的极值点为$I\left(P_{\min}^{Z}\right)$、$I\left(P_{\max}^{Z}\right)$、$L_k$；

5）基线拟合：对窗口内的极值点进行三次样条插值拟合，由于极值点确定时考虑了特征峰形状，这样更容易处理较宽的背景光谱和一些带状分子光谱特征，并不会改变光谱特征的形状；

6）背景扣除：将h_i组中n个数据点减去极小值L_i作为最终光谱强度；

7）链接窗口：将得到的n组消除背景的数据点按照步骤2）的逆向操作链接为新的光谱。

经过上述1）～7）步骤的处理后，光谱背景拟合避免了较宽谱峰的背景基线被过多估计的情况，如图2c、2d所示，其中分子光谱特征CN（$\Delta\nu$=0）被完整的保留了下来，其强度未被削弱，光谱峰形状未出现较大变形，为研究光谱强度归一化方法及定量分析中光谱特征与回归量之间的内在物理关联提供了稳定的光谱特征数据。

2.2 光谱归一化

光谱信号的归一化可进一步提升光谱信号的稳定性，增强数据对样本的代表性。本文对光谱信号的归一化不仅依赖于背景积分面积，而是综合考虑光谱积分强度与光谱背景面积，对去除背景信号后的光谱数据分两步进行归一化处理。首先，光谱信号的归一化依托于通道内光谱背景拟合基线面积A完成，用通道内光谱背景面积A归一化去除背景后的光谱信号S'，得到第一次归一化光谱S_1^N，$S_1^N=\dfrac{S'}{A}$。然后，将光谱通道内的数据再次细分，每通道内包含的2 048个数据点进一步划分为4个光谱细分范围，用每细分光谱范围内总强度I_S再次对第一次归一化光谱S_1^N进行第二次归一化，得到第二次归一化光谱S_2^N，$S_2^N=\dfrac{S_1^N}{I_s}$。其中，强度I_S是光谱细分范围内去除背景后的光谱信号积分总强度，即$I_S=\int I_{n'}$，$I_{n'}\in S'$。对光谱仪每个通道重复上述背景信号去除和归一化步骤。最终光谱信号由各通道的第二次归一化光谱拼接组成。为对比经过两次归一化和未经过归一化处理的光谱波动性大小，计算每个像素点光谱强度波动，其中，第j个像素点光谱强度波动STD_j计算如式（4）所示：

$$\mathrm{STD}_j=\frac{\sqrt{\dfrac{\sum_{i=1}^{n}\left(I_{ij}-\bar{I}_j\right)^2}{m}}}{\dfrac{1}{n}\sum_{i=1}^{n}I_{ij}} \tag{4}$$

其中，$\bar{I}$表示第j个像素点上所有光谱强度的平均值，I_{ij}表示第j个像素点上光谱强度值，n表示光谱总个数。为直观观察两步归一化预处理方法对光谱信号波动的抑制效果，对比了原始光谱、一步归一化（即去背景后仅用背景面积归一化）和本研究提出的两步归一化方法处理后的光谱信号波动，如图3所示。最终，经过上述两步归一化处理后，光谱信号的平均波动从10.07%降至4.91%。

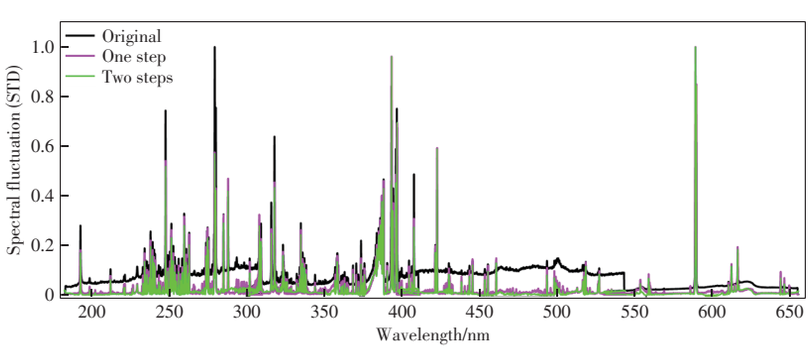


图3 原始光谱与不同归一化方法处理后的信号波动对比

Figure 3 Comparison of signal fluctuations between raw spectra and those processed by different normalization methods.

为进一步定量评价光谱预处理方法对数据结构稳定性的提升效果，采用主成分分析（Principal Component Analysis，PCA）对预处理前后的同一样本的光谱数据进行对比分析。从主成分分析的几何意义来看，一致性较差的数据点在特征空间中的分布为细长椭球形，投影点在第一个主成分分量上的

分布与在其他分量方向上的分布明显不同，这意味着数据协方差矩阵的第一特征值的大小明显大于其他特征值[32]。而一致性较高的数据点分布更接近球形，各主成分上的本征值大小接近，即各主成分捕捉到的信息变异程度相似。

图4a为同一样本重复性测试光谱数据预处理前后在得分空间中的分布。可以看出，预处理前的蓝色数据点在主成分空间中呈现明显的离散性，数据间差异较大，表明原始光谱中存在较强的噪声干扰和非相关变化。而经可变窗口平移平滑去背景及两步归一化处理后的紫色数据点，在得分空间中呈现更紧凑的聚集分布，整体结构更为稳定。这说明所使用的预处理算法有效降低了光谱信号的波动性，且提高了数据的一致性。图4b显示前五个主成分的方差贡献率，预处理前的数据中PC1的方差贡献率较高，反映整体光谱特征受噪声和强度波动的影响较大；经预处理后，各主成分的方差贡献率分布更加均衡，PC1与PC2的贡献下降而后续主成分的贡献相对提升，光谱数据中的有效信息更加均匀地分布在多个维度中，噪声和非相关信号主导的变化被显著抑制。上述分析表明，本文提出的光谱预处理方法不仅提升了光谱信号的稳定性，还改善了数据在低维空间中的结构特征，为后续建立PCA-PLS定量分析模型提供稳定的且具有样本代表性的光谱数据。

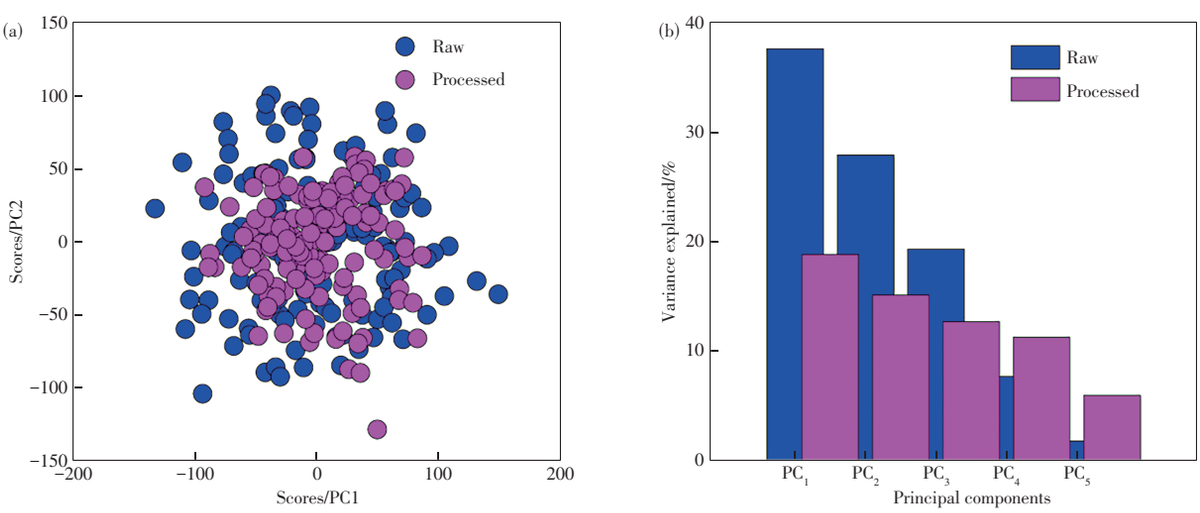


图4　主成分分析（PCA）结果：（a）前两个主成分得分图；（b）前5个主成分的方差贡献率

Figure 4　Principal component analysis (PCA) results: (a) Scores of first two PCs; (b) PCA variance explained of first five PCs.

3　结果与讨论

3.1　模型建立过程

在定量分析之前，通常要对采集的光谱数据进行数据集的划分，一般会划分为训练集和验证集来训练模型，测试集的确定可以通过数据集的划分得到，或通过对一些样本进行重复性测试得到，也可以使用独立于建模样本外的其他样本作为测试集（也称盲测数据集）。为确保模型评估的客观性与泛化能力，本研究采用以下数据集划分策略，首先将数据集分为建模数据集和盲测数据集。建模数据集共包含121个样本，按8∶2的比例划分为训练集和验证集，其中验证集主要用于模型超参数的选择和优化。数据集的划分方法采用分层随机抽样的方法，确保各数据集灰分含量分布的一致性与数据集之间的相互代表性。盲测数据集由独立于模型训练样本的另外17个样本组成，该样本集完全不参与模型训练过程，其光谱数据是利用同一装置在同样的实验条件下间隔一周时间后采集的，用于评估模型的最终泛化性与系统鲁棒性。

研究中采用偏最小二乘回归（Partial Least Squares Regression，PLSR）建立定量分析模型，PLSR通过同时分解自变量矩阵$\boldsymbol{X}$和因变量矩阵$\boldsymbol{Y}$，提取具有最大协方差的潜变量，有效解决光谱数据中常见的多重共线性问题。其基本形式为如式（5）所示：

$$\boldsymbol{Y}=\beta\boldsymbol{X}+\varepsilon \tag{5}$$

其中，$\boldsymbol{Y}$为目标变量矩阵，$\boldsymbol{X}$为光谱数据矩阵，β为回归系数矩阵，ε为残差矩阵。具体建模步骤为：

1）模型训练与超参数优化：使用训练集数据建立PLSR模型，并在验证集上评估模型性能。为确定最优超参数即主成分个数，在1～20内进行搜索。

对每个候选超参数，训练模型并计算验证集的R^2和RMSE。

2）最优超参数选择：选择验证集RMSE最小时对应的主成分数作为最优超参数，例如图5为灰分训练集和验证集在不同超参数下的RMSE，超参数为18时对应的验证集RMSE最小，故18为灰分预测模型的最优超参数。

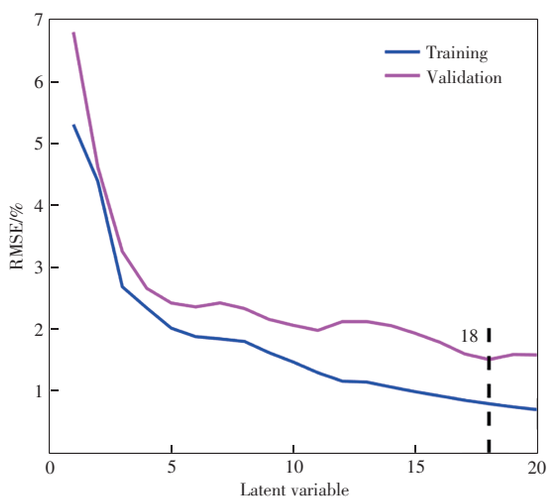


图5 训练集和验证集在不同超参数下的RMSE(灰分)

Figure 5 RMSE (Ash) of the training set and validation set under different hyperparameters.

3）最终模型训练：使用最优超参数，基于全部建模数据(训练集与验证集)重新训练，得到最终PLSR模型。

4）模型性能评估：使用训练好的模型对盲测集进行预测，并计算R^2、RMSE和MAE指标，全面评估模型的预测精度与泛化能力。模型在验证集和盲测数据集上的表现将用于进一步分析光谱预处理对模型稳健性的影响。

3.2 模型性能评价

为系统量化评估预处理方法对定量分析模型性能的影响，对比分析了预处理前后煤质灰分、全硫、热值的预测结果，如图6所示。图6a、6b、6c为预处理前的数据建立的模型，图6d、6e、6f则为采用可变窗口平移平滑去背景与两步归一化后建立的模型结果。可以直观地看到，预处理后验证集$RMSE_{cv}$结果由1.259%、0.124%及266.929 J/g降低至0.875%、0.097%及246.515 J/g。在盲测集上的煤质各参数的预测精度提升

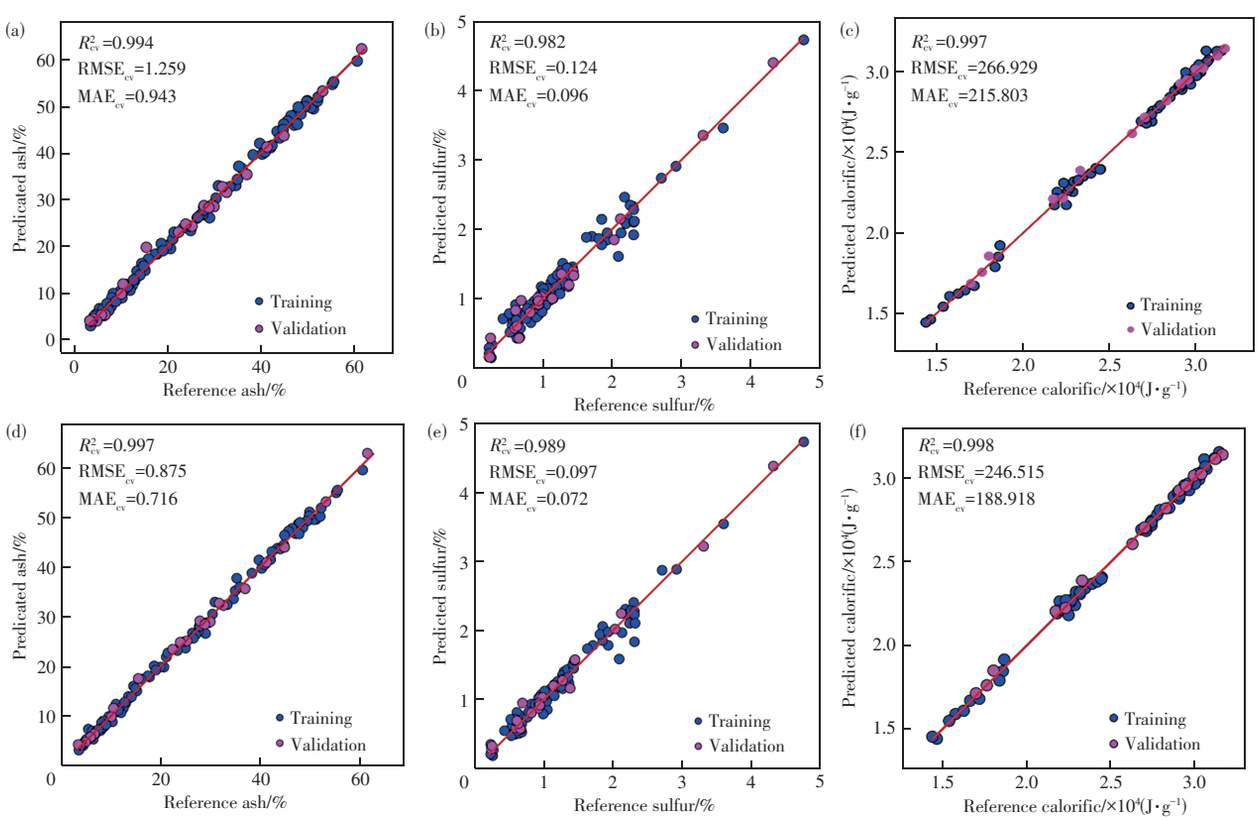


图6 预处理前后训练集、验证集预测结果：(a)(d)灰分；(b)(e)全硫；(c)(f)热值

Figure 6 Prediction results of the training and validation sets using raw and preprocessed data: (a)(d) Ash; (b)(e) Sulfur; (c)(f) Calorific values.

更明显，其中灰分、全硫和热值的$RMSE_b$(Root Mean Square Error of Blind Set)从1.2215%、0.1741%和1 283.4 J/g分别降至0.8924%、0.1648%和407.567 7 J/g，其余各参数的量化结果见表2。已有研究表明，不同预处理策略在煤质参数的LIBS定量分析中对模型性能具有显著影响。例如，Li等[33]采用11点平滑结合二阶导数的预处理方法，对煤热值进行定量建模，其验证集RMSE达到0.47 MJ/kg；李盛冬等[34]通过异常值剔除并引入一种基于稀疏矩阵技术的基线估计与降噪算法进行基线校正，将煤灰分预测均方根误差$RMSE_P$由2.2731%降至2.0017%，平均绝对误差(MAE)由1.9747%降至1.5572%。上述研究一致表明，合理的预处理策略对提升模型稳定性具有重要作用。在此基础上，本文提出的预处理方法通过有效增强数据代表性、改善数据稳定性，为模型提取稳健光谱特征提供了更加有利的数据基础，进而增强了模型在实际应用中的准确性与鲁棒性。

表2　各参数预处理前后的R^2、RMSE、MAE

Table 2　R^2, RMSE, and MAE of the models using raw and preprocessed data

Method	Items	Ash/%		Sulfur/%		Calorific values/(J·g^{-1})	
		Before preprocessing	After preprocessing	Before preprocessing	After preprocessing	Before preprocessing	After preprocessing
Validation set	R^2_{cv}	0.994	0.997	0.982	0.989	0.997	0.998
	$RMSE_{cv}$	1.259	0.875	0.124	0.097	266.929	246.515
	MAE_{cv}	0.943	0.716	0.096	0.072	215.803	188.918
Blind set	R^2_b		0.987 4	0.893 9	0.905 2	0.818 3	0.981 7
	$RMSE_b$	1.221 5	0.892 4	0.174 1	0.164 8	1 283.4	407.567 7
	MAE_b	1.096 4	0.763 7	0.144 7	0.129 0	820.794 9	304.909 7

4　结论

本研究提出了基于可变窗口平移平滑去背景与两步归一化的LIBS光谱预处理方法，提升了光谱信号稳定性和数据代表性，并结合基于PCA-PLS算法实现对煤样灰分、全硫与热值的高精度定量分析。主要结论为：

1）光谱预处理方法的细节优化。可变窗口平移平滑通过调整窗口宽度，避免了传统固定窗口因对宽谱峰背景的过多估计而导致峰形失真，保留了光谱特征峰的完整性；两步归一化通过背景面积与细分强度共同校正，降低了光谱强度的波动，提升了数据稳定性。PCA分析表明，预处理后数据在低维空间的分布更紧凑，方差贡献率更均衡，有效抑制了噪声主导的变化。

2）定量分析模型预测性能的提升。预处理后，模型在盲测集上的预测精度显著改善，灰分、全硫及热值的RMSE分别降低27%、5%和68%以上，R^2均提升至0.98以上。这种提升源于预处理后数据质量的提高，使模型能更有效地提取光谱与煤质参数之间固有的物理关联，增强了模型的泛化能力。需要指出的是，相比灰分和热值，全硫预测的提升幅度相对有限，这一现象主要源于硫元素在LIBS分析中受多方面因素制约。一方面，硫的发射谱线主要分布在紫外波段，特征谱线强度较弱，并且在空气环境下，等离子体中的激发态硫与氧元素结合，导致硫相关谱线的辐射极易发生猝灭；另一方面，硫谱线易与其他元素的谱线发生重叠，引入更复杂的谱线干扰，最终影响煤质中微量硫元素光谱的强度与稳定性，进而限制了定量分析模型的鲁棒性。在此条件下，光谱预处理对全硫预测的作用主要体现在抑制随机波动、提高预测稳定性。对于大幅度提升硫等微量元素的定量预测精度，除光谱预处理外，未来将在基体效应校正方面开展更深入的研究。

本研究将可变窗口平移平滑与两步归一化有机结合，解决了传统预处理方法中光谱失真和归一化信息单一的局限性，为LIBS光谱预处理提供了新思路。所提出的光谱预处理优化方法能够有效降低光谱噪声，提升数据稳定性和代表性，增强了煤质定量预测模型的准确性与泛化能力。该方法可推广至其他具有复杂基体物质的LIBS定量分析中，为LIBS技术在复杂工业场景中的应用提供算法支撑。

参考文献

[1] Radziemski L J, Loree T R, Cremers D A, et al. Time-resolved laser-induced breakdown spectrometry of aerosols[J]. Analytical Chemistry, 1983, 55(8): 1246-1252.

[2] Anabitarte F, Cobo A, Lopez-Higuera J M. Laser-induced breakdown spectroscopy: fundamentals, applications, and challenges[J]. ISRN Spectroscopy, 2012, 2012: 285240. DOI: 10.5402/2012/285240.

[3] Noll R. Laser-Induced Breakdown Spectroscopy: Fundamentals and Applications[M]. Heidelberg: Springer, 2012.

[4] Cremers D A, Radziemski L J. Handbook of laser-induced breakdown spectroscopy[M]. New York, USA: John Wiley & Sons, 2013.

[5] Li An, Zhang Xinyu, Wang Xianshuang, et al. High-accuracy quantitative analysis of coal by small sample modelling algorithm based laser induced breakdown spectroscopy[J]. Journal of Analytical Atomic Spectrometry, 2022, 37(10): 2022-2032.

[6] Peng Tong, Feng Junrong, Yi Wen, et al. Coal classification and analysis based on shadowgraphy and deep learning methods[J]. Optics Letters, 2025, 50(13): 4294-4297.

[7] Lin Jingjun, Li Yao, Lin Xiaomei, et al. Decision-level data fusion based on laser-induced breakdown and Raman spectroscopy: a study of bimodal spectroscopy for diagnosis of lung cancer at different stages[J]. Talanta, 2024, 275: 126194. DOI: 10.1016/j.talanta.2024.126194.

[8] 马越, 邓美慧, 陈建军. 激光诱导击穿光谱结合主成分分析在生物医学领域的应用研究[J]. 现代仪器与医疗, 2025, 31(3): 89-96.
Ma Yue, Deng Meihui, Chen Jianjun. Research on the application of laser-induced breakdown spectroscopy combined with principal component analysis in the field of biomedicine[J]. Modern Instruments & Medical Treatment, 2025, 31(3): 89-96.

[9] 王亚军. 激光诱导击穿光谱技术(LIBS)定量分析缅甸翡翠中的镁与钙[J]. 中国无机分析化学, 2024, 14(11): 1569-1575.
Wang Yajun. Quantitative analysis of Mg and Ca in Myanmar jadeite by laser-induced breakdown spectroscopy[J]. Chinese Journal of Inorganic Analytical Chemistry, 2024, 14(11): 1569-1575.

[10] Shih M, Yuan Ye, Shi Guanghai. Comparative analysis of LDA, PLS-DA, SVM, RF, and voting ensemble for discrimination origin in greenish-white to white nephrites using LIBS[J]. Journal of Analytical Atomic Spectrometry, 2024, 39(6): 1560-1570.

[11] Tavares T R, Mouazen A M, Nunes L C, et al. Laser-induced breakdown spectroscopy (LIBS) for tropical soil fertility analysis[J]. Soil and Tillage Research, 2022, 216: 105250. DOI: 10.1016/j.still.2021.105250.

[12] 杨新艳, 王鑫, 李东东, 等. 基于激光诱导击穿光谱技术的水中氮元素检测[J]. 量子电子学报, 2024, 41(3): 506-513.
Yang Xinyan, Wang Xin, Li Dongdong, et al. Determination of nitrogen in water based on laser-induced breakdown spectroscopy[J]. Chinese Journal of Quantum Electronics, 2024, 41(3): 506-513.

[13] Motto-Ros V, Syvilay D, Bassel L, et al. Critical aspects of data analysis for quantification in laser-induced breakdown spectroscopy[J]. Spectrochimica Acta Part B: Atomic Spectroscopy, 2018, 140: 54-64.

[14] Képeš E, Pořízka P, Klus J, et al. Influence of baseline subtraction on laser-induced breakdown spectroscopic data[J]. Journal of Analytical Atomic Spectrometry, 2018, 33(12): 2107-2115.

[15] Yaroshchyk P, Eberhardt J E. Automatic correction of continuum background in laser-induced breakdown spectroscopy using a model-free algorithm[J]. Spectrochimica Acta Part B: Atomic Spectroscopy, 2014, 99: 138-149.

[16] Galloway C M, Le Ru E C, Etchegoin P G. An iterative algorithm for background removal in spectroscopy by wavelet transforms[J]. Applied Spectroscopy, 2009, 63(12): 1370-1376.

[17] Gornushkin I B, Eagan P E, Novikov A B, et al. Automatic correction of continuum background in laser-induced breakdown and Raman spectrometry[J]. Applied Spectroscopy, 2003, 57(2): 197-207.

[18] Feng Jie, Wang Zhe, West L, et al. A PLS model based on dominant factor for coal analysis using laser-induced breakdown spectroscopy[J]. Analytical and Bioanalytical Chemistry, 2011, 400(10): 3261-3271.

[19] 熊凯, 邹祥波, 陈创庭, 等. 激光诱导击穿光谱结合机器学习的煤质多指标同步快速检测[J]. 热力发电, 2025, 54(4): 129-139.
Xiong Kai, Zou Xiangbo, Chen Chuangting, et al. Simultaneous rapid determination of multiple coal quality indicators using laser-induced breakdown spectroscopy combined with machine learning[J]. Thermal Power Generation, 2025, 54(4): 129-139.

[20] Li Xiaolong, Huang Jing, Chen Rongqin, et al. Chromium in soil detection using adaptive weighted normalization and linear weighted network framework for LIBS matrix effect reduction[J]. Journal of Hazardous Materials, 2023, 448: 130885. DOI: 10.1016/j.jhazmat.2023.130885.

[21] 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.

[22] Dell′Aglio M, Gaudiuso R, Senesi G S, et al. Monitoring of Cr, Cu, Pb, V and Zn in polluted soils by laser induced breakdown spectroscopy (LIBS)[J]. Journal of Environmental Monitoring, 2011, 13(5): 1422. DOI: 10. 1039/c0em00780c.

[23] Oh S Y, Yueh F Y, Singh J P, et al. Preliminary evaluation of laser induced breakdown spectroscopy for slurry samples[J]. Spectrochimica Acta Part B: Atomic Spectroscopy, 2009, 64(1): 113-118.

[24] Cao Zhi, An Yongtao, Wang Zeji, et al. Improved internal standard LIBS method used in CLF-1 exposure to liquid lithium[J]. Nuclear Materials and Energy, 2020, 24: 100786. DOI: 10. 1016/j. nme. 2020. 100786.

[25] Syvilay D, Wilkie-Chancellier N, Trichereau B, et al. Evaluation of the standard normal variate method for laser-induced breakdown spectroscopy data treatment applied to the discrimination of painting layers[J]. Spectrochimica Acta Part B: Atomic Spectroscopy, 2015, 114: 38-45.

[26] Ismaël A, Bousquet B, Pierrès K M, et al. In situ semi-quantitative analysis of polluted soils by laser-induced breakdown spectroscopy (LIBS)[J]. Applied Spectroscopy, 2011, 65(5): 467-473.

[27] Wang Rui, Ma Xiaohong. Study on LIBS standard method via key parameter monitoring and backpropagation neural network[J]. Chemosensors, 2022, 10(8): 312. DOI: 10. 3390/chemosensors10080312.

[28] Wang Beibei, Song Wenhua, Tian Ye, et al. Applying plasma acoustic and image information for underwater LIBS normalization[J]. Journal of Analytical Atomic Spectrometry, 2023, 38(2): 281-292.

[29] Zhang Lei, Dong Lei, Dou Haipeng, et al. Laser-induced breakdown spectroscopy for determination of the organic oxygen content in anthracite coal under atmospheric conditions[J]. Applied Spectroscopy, 2008, 62(4): 458-463.

[30] Wang Zhe, Afgan M S, Gu Weilun, et al. Recent advances in laser-induced breakdown spectroscopy quantification: from fundamental understanding to data processing[J]. TrAC Trends in Analytical Chemistry, 2021, 143: 116385. DOI: 10. 1016/j. trac. 2021. 116385.

[31] Xu Xiangjun, Li Angze, Wang Xianshuang, et al. The high-accuracy prediction of carbon content in semi-coke by laser-induced breakdown spectroscopy[J]. Journal of Analytical Atomic Spectrometry, 2020, 35(5): 984-992.

[32] Esbensen K H, Geladi P. Principal component analysis: concept, geometrical interpretation, mathematical background, algorithms, history, practice[M]// Comprehensive Chemometrics. Amsterdam: Elsevier, 2009: 211-226.

[33] Li Wenbing, Lu Jidong, Dong Meirong, et al. Quantitative analysis of calorific value of coal based on spectral preprocessing by laser-induced breakdown spectroscopy (LIBS)[J]. Energy & Fuels, 2018, 32(1): 24-32.

[34] 李盛冬，倪明辉，许斐，等. 基于马氏距离和稀疏矩阵技术的激光诱导击穿光谱(LIBS)煤质灰分分析[J]. 中国无机分析化学, 2022, 12(4): 97-102.
Li Shengdong, Ni Minghui, Xu Fei, et al. LIBS coal ash analysis based on mahalanobis distance and baseline estimation and denoising using sparsity data preprocessing[J]. Chinese Journal of Inorganic Analytical Chemistry, 2022, 12(4): 97-102.

High-robustness Quantitative Analysis of Coal Quality Based on LIBS Spectral Preprocessing Optimization

Zheng Yibo[1], Yang Huike[1], Wang Zhenyu[1], Wei Wentao[1], Li An[1*], Zhang Xinyu[2], Liu Ruibin[3], Wang Zhanying[4]

(*1. Hebei Key Laboratory of Optoelectronic Information and Geo-detection Technology, Hebei GEO University, Shijiazhuang, Hebei 050031, China;*
2. Institute of Oceanographic Instrumentation, Qilu University of Technology (Shandong Academy of Sciences), Qingdao, Shandong 266100, China;
3. School of Physics, Beijing Institute of Technology, Beijing 100081, China;
4. Xingtai Laotesi New Material Technology Co., Ltd., Xingtai, Hebei 054000, China)

Abstract The accuracy and precision of laser-induced breakdown spectroscopy (LIBS) quantitative analysis depend on effective preprocessing of the raw spectral data. The stability of the spectral data and its representativeness of the samples constitute the foundation determining the performance of quantitative analysis models for materials. Investigating the statistical characteristics and underlying physical mechanisms implicit in quantitative analysis further demands high-quality and highly reproducible spectral data. This paper proposed an

optimized spectral preprocessing method that integrated adaptive spectral-feature-based variable-window translation smoothing with background removal and two-step normalization within subdivided spectral channels. Specifically, the variable-window translation smoothing method adaptively adjusted the window size based on broad molecular bands, thereby eliminating distortion and deformation of characteristic peaks after background removal-a problem inherent in traditional translation smoothing methods due to their fixed window width and sliding step. The two-step normalization algorithm within subdivided channels operates as follows: the first normalization step was performed based on the spectral background area (i.e., the fitted baseline area) within a channel. Subsequently, the data within the channel were further segmented, and the total intensity within these subdivided spectral ranges was used to perform a second normalization on the initially normalized spectra. Based on this, quantitative analysis models for coal ash content, total sulfur, and calorific value were established, and the robustness and generalization capability of the models were evaluated and tested. Blind test results showed that this method significantly improves model prediction accuracy and robustness. Particularly for calorific value, the prediction root mean square error (RMSE) and mean absolute error (MAE) decreased from 1 283.4 J/g and 820.8 J/g to 407.6 J/g and 304.9 J/g, representing reductions of over 68% and 63%, respectively. Simultaneously, the coefficient of determination (R^2) increased from 0.818 3 to 0.981 7. Furthermore, the prediction performance for ash content and total sulfur was also effectively optimized. For ash content, the RMSE and MAE decreased from 1.22% and 1.10% to 0.89% and 0.76%, for total sulfur, they decreased from 0.17% and 0.14% to 0.16% and 0.13%. This study enhanced spectral stability and data representativeness through data processing algorithms, enabling the quantitative analysis models to effectively learn the implicit physical relationships between spectral data and regression variables, thereby improving model robustness and prediction accuracy. The proposed method provides algorithmic support for the application of LIBS technology in various scenarios and establishes a reliable data foundation for in-depth research into quantitative analysis.

Keywords laser-induced breakdown spectroscopy; spectral preprocessing; quantitative analysis; coal; PCA-PLS

HIGHLIGHTS

1) A variable-window translation smoothing method was proposed for spectral background fitting and removal.
2) A two-step normalization algorithm for spectral signals within subdivided channels was proposed to enhance stability.
3) Quantitative analysis algorithms based on optimized spectral preprocessing were investigated.