

DOI: 10.13382/j.jemi.B2407934

# 改进的自适应惩罚最小二乘法基线校正 用于钻井岩屑激光诱导击穿光谱分析\*

闫 碧<sup>1</sup> 李忠兵<sup>1</sup> 朱迅果<sup>1</sup> 梁海波<sup>2</sup> 倪朋勃<sup>3</sup>

(1. 西南石油大学电气信息学院 成都 610500; 2. 西南石油大学机电工程学院 成都 610500;

3. 中法渤海地质服务有限公司 天津 300457)

**摘要:** 利用激光诱导击穿光谱(laser-induced breakdown spectroscopy, LIBS)技术对钻井岩屑元素进行分析不仅能够提供地下地层信息,优化钻井过程,还能提高钻井的安全性和经济性。受钻井复杂环境的影响,钻井岩屑 LIBS 普遍呈现较严重的基线漂移现象,而现有基线校正方法容易造成基线低估或过度估计等问题。为此,提出一种改进的自适应惩罚最小二乘法基线校正方法,在非对称惩罚最小二乘算法基础上,引入 tanh 函数根据光谱信号的峰高自动调整权重矩阵,并利用光谱数据和估计基线的差值和标准差设计了一种平滑参数自动调整策略,以平衡基线校正时光谱数据的平滑度与保真度之间的冲突。在模拟光谱和实际测量的钻井岩屑 LIBS 中进行了验证,结果表明,该方法在不同噪声程度的仿真光谱上具有更低的均方根误差值,在实测岩屑 LIBS 上提高了元素的分析精度, Si、Ca、Mg、Al、Fe 5 种元素浓度预测的决定系数值分别达到了 0.992 6、0.993 0、0.968 4、0.969 1、0.977 4,均超过了 0.96。该方法可有效促进激光诱导击穿光谱技术在复杂油气环境下的钻井岩屑元素分析。

**关键词:** 钻井岩屑; 激光诱导击穿光谱; 基线校正; 自适应惩罚最小二乘法

**中图分类号:** TH744.16 **文献标识码:** A **国家标准学科分类代码:** 510.40

## Improved adaptive penalty least squares method for baseline correction of laser-induced breakdown spectroscopy analysis of drilling cuttings

Yan Bi<sup>1</sup> Li Zhongbing<sup>1</sup> Zhu Xunguo<sup>1</sup> Liang Haibo<sup>2</sup> Ni Pengbo<sup>3</sup>

(1. School of Electrical Engineering and Information, Southwest Petroleum University, Chengdu 610500,

China; 2. School of Mechatronic Engineering, Southwest Petroleum University, Chengdu 610500, China;

3. China-France Bohai Geo services Co., Ltd., Tianjin 300457, China)

**Abstract:** The analysis of elements in drilling cuttings using laser-induced breakdown spectroscopy (LIBS) technology can not only provide information about underground formations, optimize the drilling process, but also enhance the safety and economy of drilling. Affected by the complex drilling environment, the LIBS of drilling cuttings generally exhibits a relatively severe baseline drift phenomenon, while the existing baseline correction methods are prone to issues such as baseline underestimation or overestimation. Therefore, an improved adaptive penalized least-squares baseline correction method is proposed. Based on the asymmetric penalized least-squares algorithm, the tanh function is introduced to automatically adjust the weight matrix according to the peak height of the spectral signal, and a smooth parameter automatic adjustment strategy is designed by utilizing the difference and standard deviation between the spectral data and the estimated baseline to balance the conflict between the smoothness and fidelity of the spectral data during baseline correction. Verification was conducted on both simulated spectra and the measured LIBS of drilling cuttings. The results indicate that the proposed method has lower root mean square error (RMSE) values on simulated spectra with different noise levels, and improves the quantitative analysis accuracy of elements on the measured LIBS of cuttings with the  $R^2$  values of 0.992 6, 0.993 0, 0.968 4, 0.969 1, and 0.977 4 for five elements, namely Si, Ca, Mg, Al, and Fe, respectively, all exceeding 0.96. It can effectively promote the element analysis of drilling cuttings in complex oil and gas environments using laser-induced breakdown spectroscopy

收稿日期: 2024-11-03 Received Date: 2024-11-03

\* 基金项目: 四川省自然科学基金(2024NSFSC0202)、国家自然科学基金(52074233)、油气藏地质及开发工程国家重点实验室开放基金(PLN2022-42)项目资助

technology.

**Keywords:** drilling cuttings; LIBS; baseline correction; adaptive penalized least squares method

## 0 引言

岩屑元素定量分析在钻井过程中起着至关重要的作用。钻井岩屑是从地层中钻取的碎屑,分析其元素成分可以提供地下地层的重要信息<sup>[1]</sup>,帮助工程师优化钻井策略和提高效率。不同地层的矿物组成具有独特的元素特征,分析这些特征有助于判断地层类型<sup>[2]</sup>、选择合适的钻井液和钻头,从而减少井壁塌方和设备磨损<sup>[3]</sup>。此外,岩屑分析还可以揭示地层的孔隙度和渗透率,优化油气勘探和开采计划,提高经济效益。异常元素浓度变化还能用于监测地质不稳定或污染问题<sup>[4]</sup>,增强钻井的安全性和经济性<sup>[5]</sup>。

激光诱导击穿光谱(laser-induced breakdown spectroscopy, LIBS)是一种元素检测技术,以其非接触、快速、高灵敏度和多元素分析的特点,成为强大的分析工具,在钻井工程中逐渐引起重视<sup>[6]</sup>。LIBS通过高能激光脉冲激发岩屑样本,使其表面产生等离子体,通过分析等离子体发出的光谱来确定钻井岩屑的元素组,提供地下地层的详细信息。然而,LIBS的分析结果受多种因素影响,包括样品的物理状态和化学成分、激光脉冲的能量和频率、环境条件的稳定性,以及光谱仪的灵敏度和分辨率等<sup>[7-8]</sup>。这些因素直接影响到等离子体的形成和稳定性,进而影响光谱的信噪比和峰形特征,从而影响到元素的检测限、定量分析的精度<sup>[9]</sup>,最终影响到钻井岩屑样品中元素种类和浓度的准确性和可靠性评估。

通过有效的基线校正方法<sup>[10-14]</sup>,可以消除光谱中由于仪器本身或环境干扰引起的背景信号,如连续背景光、大气吸收和其他光谱仪系统的噪声<sup>[15]</sup>,较大程度地提高LIBS技术在不同应用领域中的实用性和应用价值,提供更加精确、可靠的元素分析结果。偏最小二乘法(partial least squares, PLS)因其在处理高维数据时的高效性,以及对异常值和噪声的出色鲁棒性而备受研究者的青睐。PLS通过采用Whittaker平滑策略,初步确定基线,并为位于基线之上的数据点赋予较低的权重,通过迭代不断优化,直至达到所需的精确度,从而有效地消除光谱数据中的背景干扰。但不同的权重函数和权重策略极大地影响了基线校正的性能,因此,近年来在这一领域出现了许多研究。Boelens等<sup>[16]</sup>最先提出了非对称(惩罚)最小二乘法(asymmetric (penalised) least squares, AsLS)并将其应用于光谱信号基线校正,通过将平滑参数与非对称加权相结合,实现了频谱拟合基线的灵活调整。Zhang等<sup>[17]</sup>提出了自适应迭代重加权惩罚最小二乘

法(adaptive iterative reweighted penalised least squares, airPLS),自适应迭代重加权过程与AsLS方法类似,通过基于信号和修正基线之间的差异分配权重,使得算法的收敛速度显著提高,具有很强的实用性,成为目前最流行的惩罚最小二乘基线校正算法。同时,He等<sup>[18]</sup>通过向AsLS目标函数添加一阶导数约束项,提出了一种改进的非对称最小二乘法(improved asymmetric least squares, IASLS)来校正基线,使其可预测性性能进一步提高。Baek等<sup>[19]</sup>提出了一种非对称重加权惩罚最小二乘平滑方法(asymmetric reweighted penalised least squares, arPLS),通过引入广义逻辑函数自适应获得权重,可适用于不同的噪声环境。Zhang等<sup>[20]</sup>提出一种自适应平滑参数惩罚最小二乘法(adaptive smoothing parameter penalised least squares, asPLS)的基线校正方法,通过在峰区设置较大的平滑参数,在无峰区设置较小的平滑参数,实现了平滑参数的自动调整。虽然该方法能够获得较好的效果,但是asPLS方法引入了新的参数,增加了复杂度。Li等<sup>[21]</sup>提出了一种自适应扩展高斯峰值导数重新加权惩罚最小二乘法(adaptive extended Gaussian peak derivative reweighted penalised least squares, agdPLS)来消除光谱中的基线漂移,该方法将扩展高斯峰添加到光谱中,在迭代过程中添加光谱和基线差异的导数项,并自适应调整惩罚系数。Dong等<sup>[22]</sup>设计了一种改进的非对称最小二乘平滑(optimized asymmetric least squares, O-ALS)算法,根据被测光谱的二阶凹点个数选择合适的平滑因子,基于估计基线和测量频谱之间的差异,在每次迭代中自适应地优化非对称惩罚因子。这些基线校正方法通过优化和改进偏最小二乘算法的权重参数,能够在一定程度上提升基线校正算法的性能,但却忽略了平衡参数 $\lambda$ 对于校正性能的影响。

在钻井现场,由于地质条件和操作环境的复杂性<sup>[23]</sup>,钻井岩屑的激光诱导击穿光谱常常会出现较大的基线漂移。而现有基线校正方法大多应用于拉曼光谱,极少数应用于激光诱导击穿光谱,特别是钻井岩屑方面。因此,本文提出一种针对钻井岩屑激光诱导击穿光谱的改进自适应惩罚最小二乘法基线校正方法,其核心思想是利用已知的光谱数据和估计基线的差值和标准差,通过迭代改变tanh函数构建新的权重更新策略,并设计了新的自动调整平滑参数的方法,来自适应调整平滑参数和权重,以提升激光诱导击穿光谱基线校正的性能。为了验证该方法在不同基线上的有效性和通用性,在模拟光谱和实际测量的激光诱导击穿光谱中,将所提出的方法与airPLS、arPLS和asPLS进行了比较,预测了钻井岩

屑中的 Ca、Si、Mg、Al 和 Fe 5 种元素的浓度,并比较了不同方法处理后特征谱线强度与元素浓度之间的相关系数,通过模拟实验和定量实验验证了该方法的有效性和实际适用性。

## 1 自适应惩罚最小二乘法基线校正算法

### 1.1 惩罚最小二乘法

设原始信号是长度为  $N$  的数据序列  $y = (y_1, y_2, \dots, y_N)$ , 用一平滑序列  $z = (z_1, z_2, \dots, z_N)$  表示待估计的基线。该算法通过最小化以下目标函数将原始序列  $y$  平滑为平滑序列  $z$  :

$$S(z) = (y - z)^T(y - z) + \lambda z^T D^T D z \quad (1)$$

式中:  $D$  为二阶差分矩阵。

$$D = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 \end{pmatrix} \quad (2)$$

其中,  $(y - z)^T(y - z)$  表示  $z$  对数据  $y$  的保真度,  $\lambda z^T D^T D z$  表示  $z$  的光滑度。保真度和光滑度之间的平衡由平滑参数  $\lambda$  控制。 $\lambda$  越大,  $z$  越平滑, 对于原始信号  $y$  的保真度就越低。为了平衡这二者间的冲突, 为每个点引入一个权重  $w_i$ , 构造权重矩阵  $W = \text{diag}(w_1, w_2, \dots, w_N)$  来获得平滑的基线。因此, 式(1)中的目标函数变为如下形式:

$$S(z) = (y - z)^T W (y - z) + \lambda z^T D^T D z \quad (3)$$

令偏导数为 0, 得到最小化方程的解为:

$$z = (W + \lambda D^T D)^{-1} W y \quad (4)$$

在 AsLS 方法中, 引入不对称参数  $p$  来分配权重  $w_i$ , 如式(5)所示, 并推荐设置  $p$  在 0.001~0.01 之间。而平滑参数  $\lambda$  一般取很大的值, 其范围为  $10^2 \sim 10^9$ 。由于优化目标是凸函数, 迭代过程会很快收敛。

$$w_i = \begin{cases} p, & y_i > z_i \\ 1 - p, & y_i \leq z_i \end{cases} \quad (5)$$

在该算法中, 权重  $w_i$  根据  $y_i$  与  $z_i$  之间的关系进行赋值: 当光谱信号高于估计基线时, 认为其属于光谱信号部分, 给较小的权重  $p$ ; 当光谱信号低于估计基线时, 认为其属于基线部分, 给较大的权重  $1 - p$ 。由于  $p$  值为事先给定的, 对于长度为  $N$  的光谱数据, 尤其是包含不同信号峰高度的数据, 如重叠峰区域, 这种固定的权重分配可能不太合理。

### 1.2 改进算法原理

针对上述存在的问题, 提出了一种改进的自适应惩罚最小二乘法基线校正算法 (improved adaptive penalty least squares, IAPLS), 设计新的自动调整平滑参数的方

法且不增加新的参数, 并基于  $\tanh$  函数重新给出了权重更新策略, 避免实际应用于激光诱导击穿光谱是产生基线高估或低估现象。

本文算法的核心思想是利用已知的光谱数据和估计基线的差值来自适应的调整平滑参数  $\lambda$  和权重。将权重矩阵  $W$  初始设置为全 1 矩阵, 然后给平滑参数  $\lambda$  一个初始值, 求解式(4)得到初始基线  $z$ 。而在迭代过程中, 权重  $w_i$  可以通过式(6)自动更新。

$$w_i = \begin{cases} (\tanh(\sigma_{d_-}) + 1)/2, & y_i \leq z_i \\ (\tanh(-d_i/\sigma_{d_-}) + 1)/2, & y_i > z_i \end{cases} \quad (6)$$

式中:  $d$  是原始信号  $y$  和拟合基线信号  $z$  之间的差值;  $d_-$  由  $d$  的负元素组成;  $\sigma_{d_-}$  是  $d_-$  的标准差。给定标准差, 在光谱数据与估计基线的差值  $d < 0$  的部分, 即  $y_i < z_i$  时, 此时基线可能处于高估状态, 随着标准差的增大, 权重也会增大, 使得在下一次迭代时拉低基线; 当标准差为 0 时, 权重处在 0.5 左右, 因此具有去噪效果, 估计基线不会被噪声拉低; 而当  $d > 0$  时, 此时可能处于有峰的地方, 随着  $d$  的增加, 权重逐渐变小, 故有峰处基线不会被抬高。

同时, 平滑参数  $\lambda$  根据式(7)进行自动调整。

$$\lambda = \max(\text{abs}(y_i - z_i)) \times \sigma_{d_-} \quad (7)$$

式中:  $\sigma_{d_-}$  是  $d_-$  的标准差。将式(7)代入式(4)就可以实现平滑参数的自动更新调整且不增加额外的未知参数, 同时也能使得拟合的基线  $z$  更接近所有区域的基线。迭代达到最大迭代次数或满足以下终止条件时停止。

$$|d| < 0.001 \times y \quad (8)$$

### 1.3 算法实现

IAPLS 基线校正算法的步骤如下:

步骤 1) 载入需要进行基线拟合估计的原始光谱数据  $y$ , 设置初始平滑参数  $\lambda$ , 范围为  $10^2 \sim 10^9$ ;

步骤 2) 对权重进行初始化  $w_0 = [1, 1, \dots, 1]$ , 则权重矩阵  $W$  是一个稀疏对角矩阵,  $W = \text{diag}(w_1, w_2, \dots, w_N)$ ,  $N$  为光谱数据点个数;

步骤 3) 将初始权重矩阵  $W_0$  代入式(4), 计算初始估计基线,  $z_0 = (W_0 + \lambda D^T D)^{-1} W_0 y$ ;

步骤 4) 判断是否满足迭代终止条件, 若不满足转步骤 5), 否则转步骤 8);

步骤 5) 计算 LIBS 数据和估计基线之间的差值信号  $d = y - z$ , 计算每个点的差值信号的绝对值, 得到最大值, 取差值信号的负值部分  $d_-$ , 并计算标准差  $\sigma_{d_-}$ ;

步骤 6) 将最大值和标准差  $\sigma_{d_-}$  代入式(7), 得到更新后的平滑参数  $\lambda$ ;

步骤 7) 将差值信号  $d$  和标准差  $\sigma_{d_-}$  代入式(6), 使用  $\tanh$  函数对权重进行迭代得到更新后的  $W$ , 将更新后的  $W$  和更新后平滑参数  $\lambda$  的重新代入式(4), 得到新的

估计基线；

步骤 8) 迭代终止,用原始光谱减去最终的估计基线即可实现 LIBS 数据的基线校正。

综上所述,可以得到本文基线校正算法流程如图 1 所示。

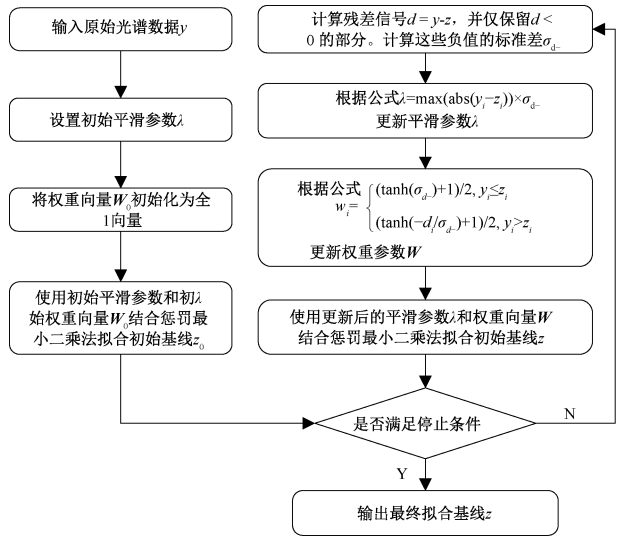


图 1 IAPLS 基线校正算法流程

Fig. 1 Flow chart of the IAPLS baseline correction algorithm

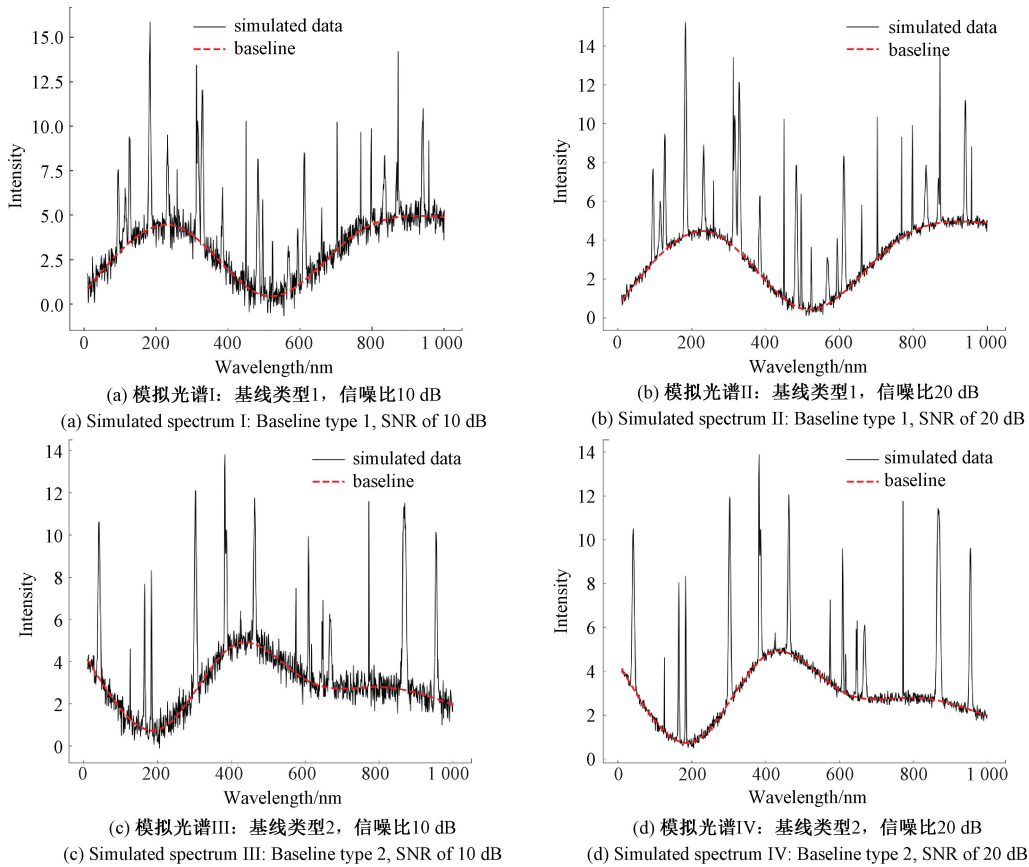


图 2 两种不同基线类型两种不同信噪比的模拟光谱

Fig. 2 Simulated spectra with two different baseline types and two different signal-to-noise ratios

2 实验结果及分析

2.1 模拟 LIBS 光谱仿真实验

1) 模拟数据

本文模拟光谱数据由模拟谱峰信号、模拟基线和随机噪声 3 部分组成。模拟光谱的产生过程如下:首先,随机确定峰的数量和位置,以便可以表达各种峰,峰的数量设置在 5~50 之间;其次,在一定范围内随机选择和宽度,在本文实验中,峰的宽度设置在 0.025~1 nm 之间,峰的高度设置在 5~21 之间,使用汉宁窗对光谱峰进行建模;最后,由于频谱基线容易出现线性和非线性漂移,使用锚点和插值法对基线进行建模,能够更好的模拟出漂移的基线,将锚点的数量设置在 2~8 之间,通过三次样条插值法生成长度为 1 000 的基线。

通过基线、光谱峰和附加的高斯白噪声叠加形成最终的模拟光谱。使用这种方式能够生成更加复杂的基线,更适用于测试基线校正算法的有效性。图 2 所示分别为随机生成的两种不同基线分别在 10、20 dB 两种信噪比(signal-to-noise ratio, SNR)情况下的模拟光谱。



2) 方法对比与结果讨论

为了更好的评估所提出的方法,利用 airPLS、arPLS 和 asPLS 方法和本文提出的算法 IAPLS 进行比较。采用真实基线和估计基线的均方根误差 (RMSE) 来比较方法的性能。RMSE 计算公式如下:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (b_i - \hat{b}_i)^2}{N}}$$

(9)

式中:  $b$  是真实基线;  $\hat{b}$  是估计基线;  $N$  表示光谱数据点数。均方根误差的值越小,说明基线校正的准确性越高。图 3 所示为 4 种方法在 4 个模拟光谱上的均方根误差与平滑参数  $\lambda$  的变化关系。可以看出,在 4 个模拟光谱上,

在平滑参数  $\lambda$  遵循的选取原则范围内,本文提出的算法 IAPLS 对不同基线、不同噪声均具有较小的 RMSE 值,且变化较小,表明本文算法提出的方法具有很强的自适应能力。

选取 RMSE 最小值对应的  $\lambda$  作为最优平滑参数,用作 airPLS、arPLS 和 asPLS 方法的最优参数,用作后续实验验证。选取 4 种方法的 RMSE 变化最小范围内对应的  $\lambda$  值作为最优  $\lambda$  值范围,如表 1 所示,可见,airPLS、arPLS 和 asPLS 方法对不同基线、不同噪声所对应的最优  $\lambda$  值范围不同,应用于不同类型的光谱,所需的参数难以确定,而本文方法对不同基线、不同噪声所对应的最优  $\lambda$  值范围相同,说明本文方法具有较强的鲁棒性。

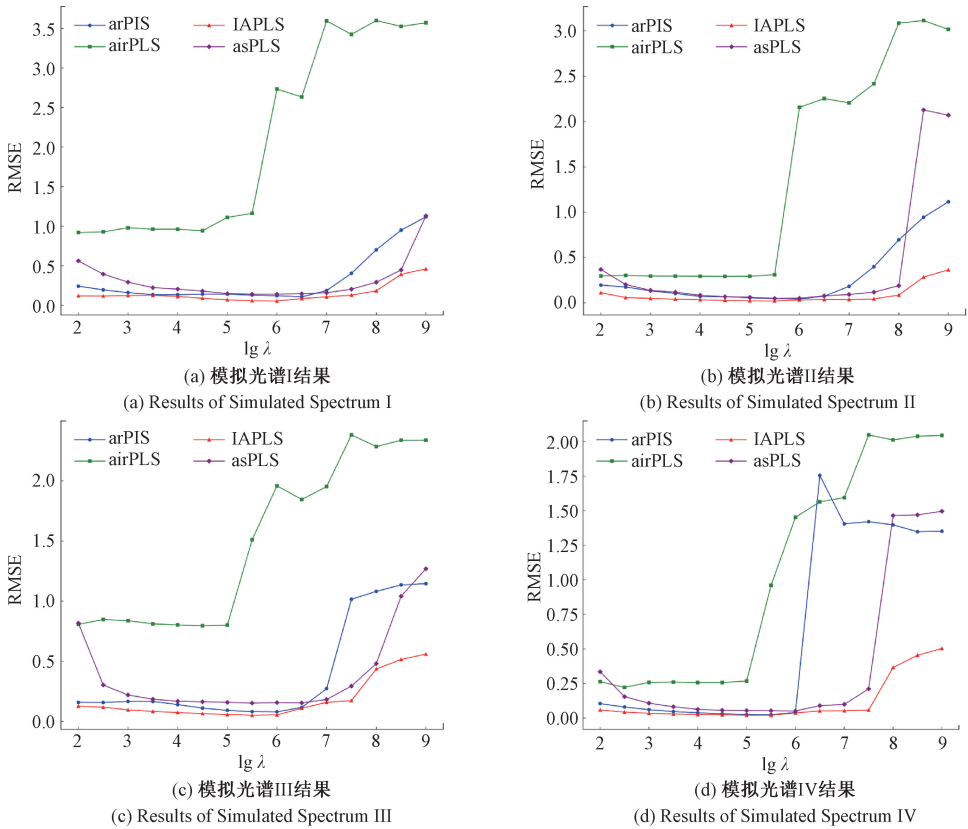


图 3 4 种方法在 4 个模拟光谱上的 RMSE 与平滑参数  $\lambda$  的变化关系

Fig. 3 The variation relationships between RMSE and smoothing parameter  $\lambda$  of the four methods on the four simulated spectra

表 1 4 种方法在 4 个模拟光谱的平滑参数  $\lambda$  最优值范围

方法	模拟光谱 I	模拟光谱 II	模拟光谱 III	模拟光谱 IV
airPLS	$10^2 \sim 10^3$	$10^3 \sim 10^4$	$10^2 \sim 10^3$	$10^4 \sim 10^5$
arPLS	$10^5 \sim 10^6$	$10^4 \sim 10^5$	$10^4 \sim 10^5$	$10^5 \sim 10^6$
asPLS	$10^6 \sim 10^7$	$10^6 \sim 10^7$	$10^6 \sim 10^7$	$10^5 \sim 10^6$
IAPLS	$10^6 \sim 10^7$	$10^6 \sim 10^7$	$10^6 \sim 10^7$	$10^6 \sim 10^7$

图 4 所示为 4 种方法在模拟光谱 III、模拟光谱 IV 下的估计基线及 200~600 nm 范围内的局部放大图。对比

图 4(a) 和 (c) 可以看出,当噪声较大时,airPLS 方法得到的估计基线偏离真实基线的程度最大,这与其权重函数

设置相关,没有考虑到噪声的影响;从局部放大图可以观察到,arPLS 在有峰处会产生抬升的基线,容易导致校正后峰的强度变弱,而 asPLS 的估计基线低于真实基线,均受到噪声的明显影响。与 airPLS、arPLS 和 asPLS 方法相比,本文提出的方法在两种信噪比下均具有较好的效果,

产生的估计基线几乎与真实基线重合。4 种基线校正方法在 4 个模拟光谱上的 RMSE 值如表 2 所示,本文提出的 IAPLS 算法均具有最小的 RMSE 值,表明本文算法具有最好的基线拟合效果。

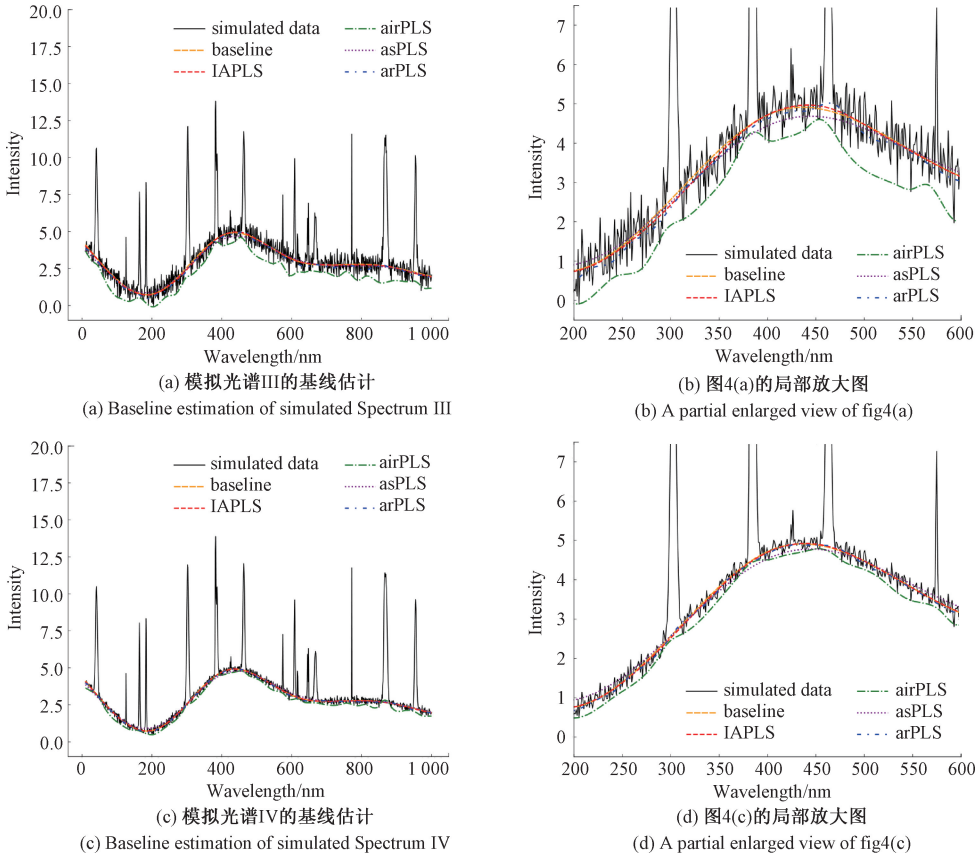


图 4 4 种方法对模拟光谱 III、模拟光谱 IV 的估计基线

Fig. 4 The estimated baselines of simulated Spectrum III and simulated Spectrum IV by the four methods

表 2 模拟光谱上 4 种方法的 RMSE 值

Table 2 RMSE values of the simulated spectra by the four methods

方法	模拟光谱 I	模拟光谱 II	模拟光谱 III	模拟光谱 IV
airPLS	1.110 6	0.294 3	0.801 3	0.267 3
asPLS	0.120 6	0.043 1	0.156 7	0.051 1
arPLS	0.089 5	0.072 1	0.054 3	0.040 0
IAPLS	0.055 0	0.033 3	0.046 8	0.023 3

为了进一步测试本文算法对不同噪声情况下光谱的适应性,图 5 所示为本文算法与 airPLS、arPLS 和 asPLS 算法在不同信噪比下的 RMSE 变化情况。可以看出,airPLS 算法在低信噪比下效果较差,具有较高的误差,asPLS、arPLS 及本文算法效果依次提高。本文提出的 IAPLS 算法,具有更低的 RMSE 变化,均方根误差平稳的保持在较小范围内(0.015 3,0.163 9)。因此,本文算法对不同噪声下光谱的适应性更强,具有较好的噪声消除

效果。

将本文提出的自适应调整平滑参数方法和权重分配方案应用于 4 个模拟光谱,得到迭代过程中的  $\lambda$  变化情况如表 3 所示。可以看出,权重只需迭代两次就达到迭代停止条件,每一次迭代所产生的  $\lambda$  值都在表 1 中最优  $\lambda$  值范围内,且平滑参数值在迭代过程中具有较大的变化,并不是单一的值,说明本文方法能够根据光谱数据自适应调整平滑参数。

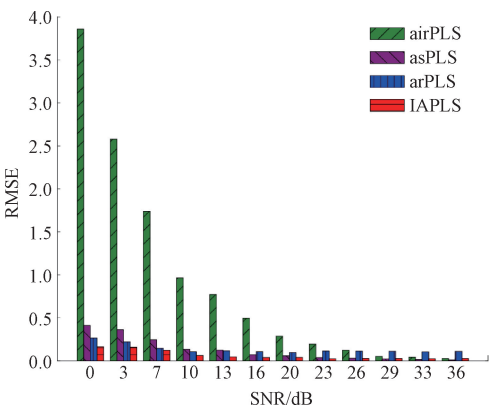


图 5 4 种方法不同噪声下模拟光谱的 RMSE 变化情况  
Fig. 5 The variation of RMSE values of the four methods on the simulated spectra under different noises

表 3 算法迭代过程中平滑参数  $\lambda$  变化情况  
Table 3 The variation of the smoothing parameter  $\lambda$  during algorithm iteration process

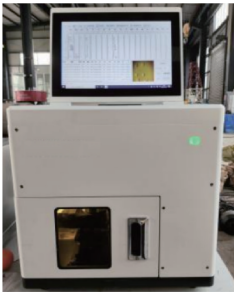
	模拟光谱 I	模拟光谱 II	模拟光谱 III	模拟光谱 IV
第 1 次迭代	4 694 755.8	2 784 533.8	2 014 967.8	3 347 124.1
第 2 次迭代	3 623 815.7	1 319 599.9	963 582.4	2 410 570.4



(a) 电磁粉碎机  
(a) The electromagnetic crusher



(b) 手动压片机  
(b) The manual tablet-pressing machine



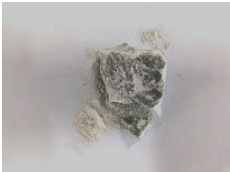
(c) LIBS光谱仪  
(c) The LIBS spectrometer

图 6 实验系统

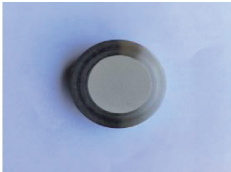
Fig. 6 The experiment system

采样时,在常温常压环境下,首先取适量干燥岩屑物质放入粉碎机,电磁研磨 2 min,将其研磨为 200  $\mu\text{m}$  粒径的岩屑粉末,然后取约 3 g 岩屑粉末放入直径为 30 mm 的模具中,利用手动压片机在 10 MPa 下压制成药饼,使

样品表面平整,再将制成的圆饼放入 LIBS 光谱仪中,取 30 个检测点的光谱数据平均值作为一个样本的原始光谱数据,以进一步减少样品不均匀性对实验的影响。图 7 所示为经过不同处理后得到的岩屑样本图片。



(a) 原始岩屑样本 I  
(a) Initial cuttings sample



(b) 经过研磨和压片后得到的圆饼  
(b) The round cake obtained after grinding and pressing



(c) 经过 LIBS 光谱仪进行 30 个检测点采样后的样本  
(c) The sample after 30-point measurements by the LIBS spectrometer

图 7 不同处理过程的岩屑样本图片

Fig. 7 The images of the sample with different processes

国家标准岩屑样品及钻井岩屑样品均通过上述流程采集获得光谱数据,其中,99 个国标岩屑样品的光谱图如图 8 所示,图 8(b) 为图 8(a) 中国标样品 GBW07114

的光谱图,可以观察到明显的基线漂移现象。表 4 为部分岩屑标准样品的国标编号和 Ca、Si、Mg、Al、Fe 5 种元素的真实浓度含量。

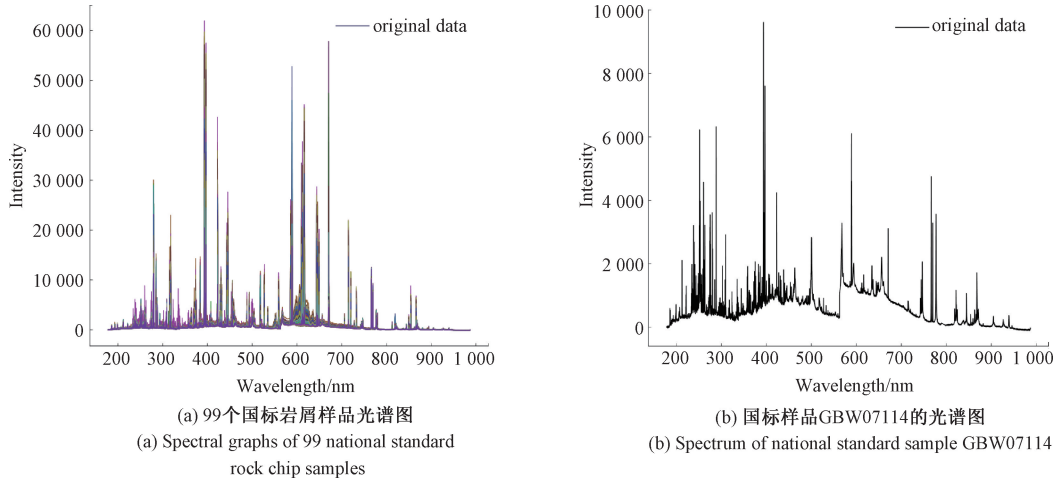


图 8 岩屑样品光谱图  
Fig. 8 Spectra of cuttings samples

表 4 部分岩屑标准样品的国标编号和 Ca、Si、Mg、Al、Fe 5 种元素的真实浓度含量

Table 4 GB numbers (National Standard codes) and certified concentrations of Ca, Si, Mg, Al, and Fe in selected rock chip reference materials (%)					
样品编号	Si	Ca	Mg	Al	Fe
GBW07152	34.765 4	0.239 4	0.032 5	7.811 3	0.323 7
GBW07153	30.216 9	0.054 3	0.021 7	10.118 8	0.226 0
GBW07733	30.011 2	0.100 0	0.036 1	13.415 9	0.322 4
GBW07734	30.086 0	0.050 0	0.021 1	13.378 8	0.309 2
GBW07735	34.718 6	0.192 9	0.028 9	8.404 1	0.142 9
GBW07114	0.289 3	21.442 8	13.080 0	0.052 9	0.028 0
GBW07136	3.850 0	23.621 4	10.800 0	0.052 9	0.039 9
GBWE070157	3.929 3	20.521 4	11.856 0	0.635 2	0.332 5
GBWE070158	0.872 6	21.585 7	12.510 0	0.108 5	0.170 8
GBWE070159	1.008 0	21.535 7	12.546 0	0.132 3	0.173 6
GBWE070160	2.436 0	21.071 4	12.258 0	0.373 7	0.249 9
GBW07103	33.987 3	1.107 1	0.252 0	7.094 1	1.498 0
GBW07104	28.289 3	3.714 2	1.032 0	8.560 5	3.430 0
GBW07105	20.832 0	6.292 8	4.662 0	7.321 7	9.380 0
GBW07109	25.424 0	0.992 8	0.390 0	9.381 1	5.184 6
GBW07110	29.428 0	1.764 2	0.504 0	8.523 5	3.304 7

2) 钻井岩屑 LIBS 光谱基线校正

应用上述 4 种方法对采集的所有岩屑 LIBS 光谱进行了基线校正,图 9 所示为图 8(b) 所示光谱基线校正情况。可见,4 种方法均成功实现 LIBS 光谱的基线校正。从局部放大图看,airPLS 方法生成的基线明显低于原始光谱数据且偏离幅度较大,存在低估基线的现象,这与模拟实验结果一致;arPLS 和 asPLS 方法在有峰处容易产生类似谱峰的基线,存在基线过度估计的现象,容易导致有效信息的丢失;本文方法产生了较为稳定的基线,在无峰

处可有效去除噪声的影响,在有峰处紧密跟随光谱的下部轮廓,保证了拟合曲线的稳定和平滑。

为验证本文算法对钻井岩屑 LIBS 光谱定量分析的影响,在 99 个钻井岩屑标准物质的 LIBS 光谱数据集上建立了偏最小二乘回归分析模型<sup>[24]</sup>,以预测不同岩屑样品中 Ca、Si、Mg、Al、Fe 等元素的含量。光谱数据集分为训练集和测试集两组,训练集:测试集=4:1。为了评价模型的性能,采用决定系数  $R^2$  和 RMSE 两个指标评估定量分析模型性能,其中决定系数表示的是实际值与预



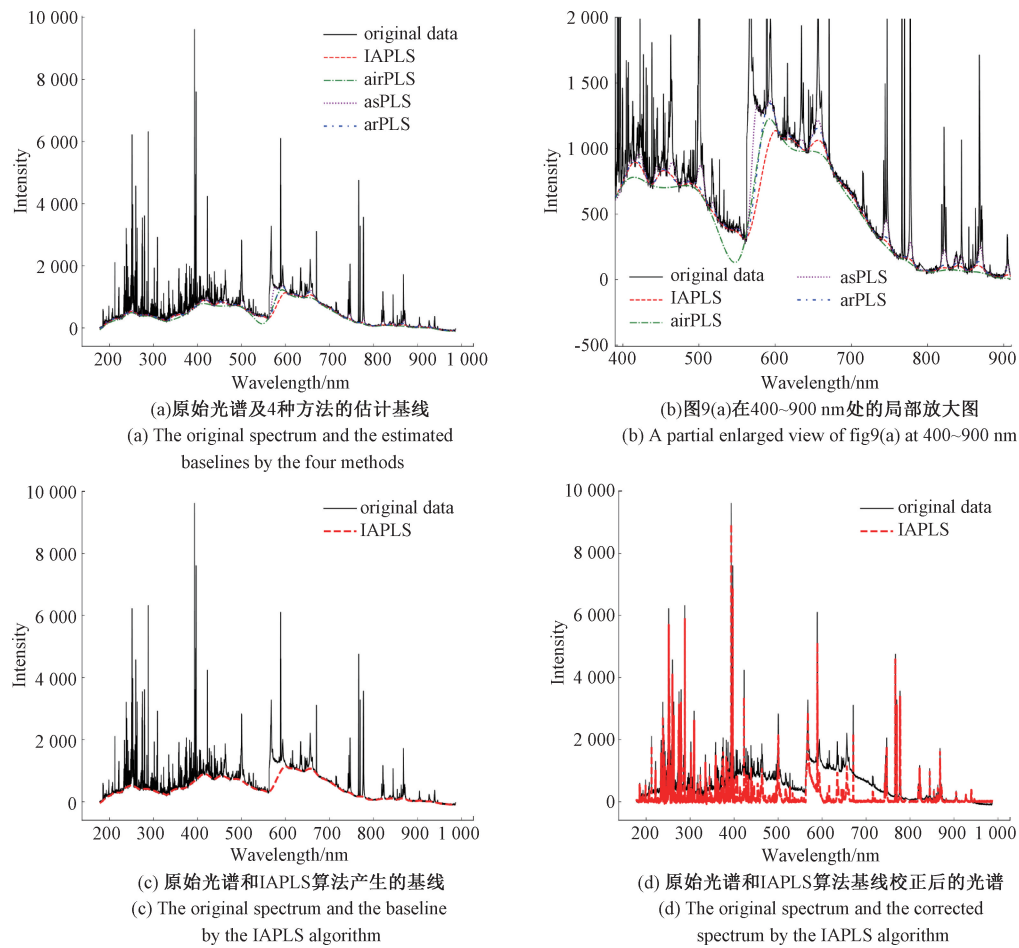


图 9 钻井岩屑实测 LIBS 光谱数据估计基线与校正结果

Fig. 9 Estimated baselines and correction results of the measured LIBS spectral data of drilling cuttings

测值之间的相关性,该值参考范围为(0,1),越接近于 1 意味着该模型的拟合优度越高。误差评价指标反映了实际值与预测值的差异,均方根误差数值越小代表模型预测精度越高,决定系数公式如式(10)所示。

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \tag{10}$$

式中:  $i$  为测量样本序号;  $N$  为测量样本数量;  $y_i$  为第  $i$  个测量样本实际浓度值;  $\hat{y}_i$  为第  $i$  个测量样本预测浓度;  $\bar{y}$  为  $N$  个样本真实浓度的均值。

分别采用 4 种基线校正方法后,利用偏最小二乘回归方法对 Ca、Si、Mg、Al、Fe 等 5 种元素浓度进行了定量分析,其相关系数和均方根误差如表 5 所示。与未进行基线校正的 PLS 模型相比,基线校正后的 PLS 模型性能均有所提高,说明基线校正可以提高模型的精度。IAPLS 方法得到的  $R^2$  和 RMSE 值明显优于 airPLS、asLS 和 arPLS,能够有效去除实际岩屑光谱基线,提高光谱强度与浓度之间的线性相关性,提高元素浓度预测精度,Si、

Ca、Mg、Al、Fe 5 种元素浓度预测的  $R^2$  值分别达到了 0.992 6、0.993 0、0.968 4、0.969 1、0.977 4。

3 结 论

本文提出了一种改进的自适应惩罚最小二乘法基线校正方法,在非对称惩罚最小二乘法基础上,引入 tanh 函数根据光谱信号的峰高自动调整权重矩阵,并利用光谱数据和估计基线的差值和标准差设计了一种平滑参数自动调整策略,以平衡基线校正时光谱数据的平滑度与保真度之间的冲突。模拟光谱的实验结果表明,本文方法在不同噪声及不同基线情况下,能有效去除使光谱抬高的基线和部分噪声,而不会出现较为严重的基线低估或高估,比 airPLS、arPLS 和 asPLS 方法获得了更好的性能。在应用于钻井岩屑激光诱导击穿光谱实验中,利用该方法及 airPLS、arPLS 和 asPLS 方法对实测 LIBS 光谱进行基线校正后,采用偏最小二乘法预测钻井岩屑中 Ca、Si、Mg、Al 和 Fe 元素含量。实验结果表明,采用本文

表 5 偏最小二乘回归方法结合不同基线校正方法的定量分析效果

Table 5 Quantitative analysis performance of partial least squares regression combined with different baseline correction methods

PLS	Si		Ca		Mg		Al		Fe	
Method	$R^2$	RMSE	$R^2$	RMSE	$R^2$	RMSE	$R^2$	RMSE	$R^2$	RMSE
original	0.932 4	3.318 9	0.957 4	2.957 4	0.911 6	0.724 7	0.767 7	2.033 3	0.758 2	0.940 6
airPLS	0.975 6	2.542 7	0.988 8	2.506 5	0.962 8	0.561 1	0.903 2	1.696 5	0.889 7	0.796 0
arPLS	0.975 3	2.581 4	0.980 4	2.770 6	0.947 1	0.611 6	0.908 2	1.661 4	0.885 2	0.809 9
asPLS	0.952 1	3.099 5	0.979 6	2.855 8	0.925 9	0.642 3	0.933 3	1.632 0	0.964 9	0.780 7
IAPLS	0.992 6	2.238 2	0.993 0	2.268 8	0.968 4	0.527 3	0.969 1	1.518 3	0.977 4	0.743 1

方法进行钻井岩屑 LIBS 光谱基线校正后,5 种元素浓度预测的相关系数均高于 0.96。本文方法显著提高了钻井岩屑元素预测浓度的准确性,可有效促进激光诱导击穿光谱技术在复杂油气环境下的钻井岩屑元素分析。

参考文献

[ 1 ] 问晓勇,张敏,牟春国,等. 基于钻井岩屑评价水平井地层力学性质——以鄂尔多斯盆地太原组灰岩为例[J]. 钻采工艺,2024,47(1):87-93.

WEN X Y, ZHANG M, MU CH G, et al. Evaluation of formation mechanical properties of horizontal wells based on drilling cuttings: Taking limestone of Taiyuan Formation in Ordos Basin as an example[J]. Drilling & Production Technology, 2024, 47(1):87-93.

[ 2 ] ZHANG H, CHEN Q, NI P, et al. Study on the intelligent identification method of formation lithology by element and gamma spectrum[J]. Neural Computing & Applications, 2022, 34(5):3375-3383.

[ 3 ] WU J, ZHANG S, DUAN X, et al. Rotary thermal desorption technology for treatment of oil-based drilling cuttings in shale gas industry [ J ]. Separation and Purification Technology, 2024, 337:126319.

[ 4 ] XU Q, ZHANG D, MA L, et al. Treatment of oil-based drilling cuttings by floatation-advanced oxidation two-step process[J]. RSC advances, 2024, 14(5):3000-3009.

[ 5 ] 张雪菲,张万栋,方胜杰,等. 含油岩屑处理和资源利用的研究进展[J]. 化工技术与开发, 2024,53 ( Z1 ): 99-104,114.

ZHANG X F, ZHANG W D, FANG SH J, et al. Research progress on treatment and resource utilization of oil-bearing cuttings [ J ]. Technology & Development of Chemical Industry, 2024, 53 ( Z1 ): 99-104,114.

[ 6 ] WANG J, XU J, SHI L, et al. A computer-controlled in situ analysis instrument based on laser induced breakdown spectroscopy [ J ]. Applied Mechanics and Materials, 2014, 556:561-566.

[ 7 ] ZhANG Q, ZHAO T, LI X, et al. Correction method for the influence of environmental factors on laser-induced

breakdown spectroscopy [ J ]. Optics letters, 2021, 46(20):5244-5247.

[ 8 ] ERIK K, PAVEL P, JAKUB K, et al. Influence of baseline subtraction on laser-induced breakdown spectroscopic data [ J ]. Journal of Analytical Atomic Spectrometry, 2018, 33(12):2107-2115.

[ 9 ] HERMANN J, GROJO D, AXENTE E, et al. Local thermodynamic equilibrium in a laser-induced plasma evidenced by blackbody radiation [ J ]. Spectrochimica Acta Part B: Atomic Spectroscopy, 2018, 144:82-86.

[ 10 ] LONG L, GUANG X, MING Z, et al. Baseline correction algorithm for Raman spectroscopy based on adaptive window spline fitting [ J ]. Spectroscopy and Spectral Analysis, 2021, 41(1):111-115.

[ 11 ] CHEN H, SHI X, HE H, et al. Automatic background correction method for laser-induced breakdown spectroscopy [ J ]. Spectrochimica Acta Part B: Atomic Spectroscopy, 2023, 208:106763.

[ 12 ] YANG Z, HISAYUKI A. A double sliding-window method for baseline correction and noise estimation for Raman spectra of microplastics [ J ]. Marine pollution bulletin, 2023, 190:114887-114887.

[ 13 ] XI Y, LI Y, DUAN Z, et al. A novel pre-processing algorithm based on the wavelet transform for Raman spectrum [ J ]. Applied Spectroscopy, 2018, 72 ( 12 ): 1752-1763.

[ 14 ] MA S, XU S, CHEN Y, et al. A LIBS spectrum baseline correction method based on the non-parametric prior penalized least squares algorithm [ J ]. Analytical methods: Advancing Methods and Applications, 2024, 16(26):4360-4372.

[ 15 ] FENG C, YANG R, LI Q, et al. Quantitative measurement of hydrogen isotopes in titanium using laser-induced breakdown spectroscopy [ J ]. Applied optics, 2020, 59(9):2866-2873.

[ 16 ] BOELENS H F M, EILERS P H C, THOMS H. Sign constraints improve the detection of differences between complex spectral data sets: LC-IR as an example [ J ].

- Analytical chemistry, 2005, 77(24):7998-8007.
- [17] ZHANG Z, CHEN S, LIANG Y. Baseline correction using adaptive iteratively reweighted penalized least squares[J]. The Analyst, 2010, 135(5):1138-46.
- [18] HE S, ZHANG W, LIU L. Baseline correction for Raman spectra using an improved asymmetric least squares method[J]. Analytical Methods, 2014, 6(12): 4402-4407.
- [19] BAEK S J, PAR K A, AHN Y J, et al. Baseline correction using asymmetrically reweighted penalized least squares smoothing[J]. The Analyst, 2015, 140(1): 250-257.
- [20] ZHANG F, TANG X, TONG A, et al. Baseline correction for infrared spectra using adaptive smoothness parameter penalized least squares method [J]. Spectroscopy Letters, 2020, 53(3):222-233.
- [21] LI X, TANG X, WANG B, et al. An adaptive extended Gaussian peak derivative reweighted penalised least squares method for baseline correction[J]. Analytical Methods: Advancing Methods and Applications, 2023, 15(44):6048-6060.
- [22] DONG Y, XU L. Baseline estimation using optimized asymmetric least squares (O-ALS) [J]. Measurement, 2024, 233:114731.
- [23] 唐正. 石油钻井中安全影响因素与控制途径分析[J]. 石化技术, 2023, 30(11):184-186.
- TANG ZH. Analysis of safety-Influencing factors and

control approaches in petroleum drilling [J]. Petrochemical Industry Technology, 2023, 30(11): 184-186.

- [24] DUMA S, SUSILUOTO J, LAMMINPAA O, et al. KF-PLS: Optimizing kernel partial least-squares (K-PLS) with kernel flows [J]. Chemometrics and Intelligent Laboratory Systems, 2024, 254:105238-105238.

## 作者简介



闫碧, 2023 年于西南石油大学获得学士学位, 现为西南石油大学硕士研究生, 主要研究方向为数据处理、光谱分析。

E-mail: 2582190151@qq.com

**Yan Bi** received her B. Sc. degree from Southwest Petroleum University in 2023. Now she is a M. Sc. candidate at Southwest Petroleum University. Her main research interests include data processing and spectral analysis.



李忠兵(通信作者), 2009 年于武汉大学获得学士学位, 2014 年于武汉大学获得博士学位, 现为西南石油大学副教授, 主要研究方向为图像处理、机器学习和光谱分析。

E-mail: lzb@swpu.edu.cn

**Li Zhongbing** (Corresponding author) received his B. Sc. degree from Wuhan University in 2009 and his Ph. D. degree from Wuhan University in 2014. Now he is an associate professor at Southwest Petroleum University. His main research interests include image processing technology, machine learning, and spectral analysis.