

基于惩罚最小二乘算法的土壤重金属检测光谱基线校正

江晓宇^{1,2} 李福生^{1,2} 王清亚^{1,2} 郝 军^{1,2} 徐木强^{1,2} 罗 杰³

(1. 东华理工大学核资源与环境国家重点实验室, 南昌 330013;
2. 东华理工大学核技术应用教育部工程研究中心, 南昌 330013;
3. 长江大学资源与环境学院, 武汉 430100)

摘要: 针对 X 射线荧光光谱分析技术在检测土壤重金属过程中由于土壤背景复杂、包含大量噪声和干扰信息而易受基体效应影响的问题,为了提高定量分析模型的精度,利用惩罚最小二乘算法拟合基线与真实基线之间的保真度和平滑度,对 X 射线荧光光谱进行基线校正,从而减小基线漂移的影响。选用无基线扣除、非对称最小二乘 (ASLS)、自适应迭代重加权惩罚最小二乘 (AIRPLS)、非对称重加权惩罚最小二乘 (ARPLS)、局部对称重加权惩罚最小二乘 (LSRPLS) 和多约束重加权惩罚最小二乘 (DRPLS) 等 6 种处理方法对土壤重金属元素铅和砷的测量光谱进行基线校正;结合偏最小二乘 (PLS) 算法建立相应的校正模型,以选择最优基线校正算法;与神经网络 (BP) 和支持向量机 (SVR) 建立的校正模型进行比较,对模型进行评价。结果显示,铅元素的最佳模型为 DRPLS-PLS,模型的 R^2 达到 0.982,预测均方根误差 (RMSEP) 为 0.056 mg/kg;砷元素的最佳模型为 DRPLS-PLS 模型,模型的 R^2 达到 0.985,RMSEP 为 0.796 mg/kg。与 PLS 和 BP 模型相比,铅、砷两种元素的 SVR 模型建模均最优,模型的 R^2 分别达到 0.998 和 0.993,RMSEP 分别为 0.015、0.596 mg/kg。实验表明,通过基线校正后模型的预测精度、检出限和稳定性均有所提高,该方法可有效提高 X 射线荧光光谱在土壤中的定量分析能力。

关键词: X 射线荧光光谱; 基线校正; 惩罚最小二乘算法; 重金属

中图分类号: O657.3; X53 文献标识码: A 文章编号: 1000-1298(2021)08-0205-08 OSID: 

Spectrum Baseline Correction for Soil Heavy Metal Detection Based on Penalized Least Squares Algorithm

JIANG Xiaoyu^{1,2} LI Fusheng^{1,2} WANG Qingya^{1,2} HAO Jun^{1,2} XU Muqiang^{1,2} LUO Jie³

(1. State Key Laboratory of Nuclear Resources and Environment, East China University of Technology, Nanchang 330013, China
2. Engineering Research Center of Nuclear Technology Application, Ministry of Education, East China University of Technology, Nanchang 330013, China
3. College of Resource and Environmental Science, Yangtze University, Wuhan 430100, China)

Abstract: X-ray fluorescence spectrometry has the advantages of nondestructive and rapid detection of heavy metals in soil. However, in the process of practical application, because the soil background is complex and contains a lot of noise and interference information, it is easy to be affected by the matrix effect. In order to improve the accuracy of the quantitative analysis model, it is necessary to carry out baseline correction of X-ray fluorescence spectrum and reduce the effect of baseline drift. Penalty least squares algorithm, as a common baseline algorithm, was used to further optimize the fitting baseline based on least squares by fitting the fidelity and smoothness between the baseline and the real baseline. No baseline deduction, asymmetric least squares (ASLS), adaptive iterative reweighted penalty least squares (AIRPLS), asymmetric reweighted penalty least squares (ARPLS), local symmetric reweighted penalty least squares (LSRPLS) and multi-constrained reweighted penalty least squares (DRPLS) were selected for baseline correction of the measured spectrum of heavy metal elements lead and arsenic in soil, and then the corresponding correction models were established with partial least squares (PLS)

收稿日期: 2021-03-25 修回日期: 2021-05-19
基金项目: 2019 年江西省“双千计划”引进项目(2120800003)和国家自然科学基金项目(21876014)
作者简介: 江晓宇(1988—),男,博士生,主要从事核仪器检测研究,E-mail: jjxy1988@126.com
通信作者: 李福生(1976—),男,教授,博士生导师,主要从事核技术应用及智能仪器开发研究,E-mail: lifusheng@ecit.cn

algorithm to select the optimal baseline correction algorithm. At last, the partial least square (PLS) model was compared with the correction model established by neural network (BP) and support vector machine (SVR) to evaluate the advantages and disadvantages of different models. The results showed that the optimal baseline correction algorithm of the two elements was DRPLS, which the R^2 of the lead corresponding PLS model was 0.982, the prediction root mean square error (RMSEP) was 0.056 mg/kg, and the R^2 of the arsenic corresponding PLS model was 0.985, the RMSEP was 0.796 mg/kg. Besides, the SVR models of lead and arsenic were optimal compared with PLS and BP models. And the R^2 of the model reached 0.998 and 0.993, respectively. The RMSEP was 0.015 mg/kg and 0.596 mg/kg, respectively. Experiments showed that the prediction accuracy, detection limit and stability of the model established after baseline correction can effectively improve the quantitative analysis ability of X-ray fluorescence spectroscopy in soil.

Key words: X-ray fluorescence spectroscopy; baseline correction; penalty least squares algorithm; heavy metal

0 引言

能量色散 X 射线荧光 (Energy dispersive X-ray fluorescence, EDXRF) 光谱仪是一种间接定量分析仪器,可对多种重金属元素进行快速同步分析^[1],除了能无损检测外,EDXRF 因其体积小、质量轻、分析速度快及准确度高而广泛应用于野外现场分析中^[2-3]。近几年,EDXRF 已经成为地质、钢铁、石油、环境和医学等领域的首选仪器^[4]。在获取 X 射线荧光光谱的过程中,仪器本身(如 SSD 探测器和采集板)产生的统计涨落所带来的噪声干扰以及 X 射线管出射谱中连续韧致辐射造成的背景干扰噪声是高频信号,基线变化是低频信号。基线问题是一个相对复杂的问题,理论上很难研究出一种完美的方法来处理基线问题。通常直线用于连接信号峰的两端,以直线为基线,在此基础上进一步计算峰面积或峰高。如果直线不符合实际基线,则计算就会产生误差。

近年来,研究人员提出了各种基线校正方法,例如小波变换^[5-7]、多项式拟合^[8-9]、傅里叶变换^[10-13]、形态学^[14]和惩罚最小二乘^[15]。其中,惩罚最小二乘算法因其在光谱背景扣除中的高效率和简便性而被广泛采用。1922 年,惩罚最小二乘由 WHITTAKER^[16]首次提出,并用作数据平滑,随后研究人员在惩罚最小二乘算法的基础上进行了很多改进^[17-21]。虽然改进的基于惩罚最小二乘算法的基线校正方法在近红外光谱领域取得了一些良好的实验效果,但在 X 射线光谱分析领域的具体应用效果还有待进一步研究和探讨。

本文选择 5 种惩罚最小二乘(或其改进)算法,结合 PLS 技术分别对土壤样品中的铅、砷元素建立 X 射线荧光光谱定量分析模型,并对比分析模型的相关指标,选取最适合 X 射线荧光光谱基线校正的算法,然后采用选好的 PLS 模型分别与神经网络

(BP)^[22-24]和支持向量机(SVR)^[25-27]所建立的模型进行对比分析,以选择较优的土壤中 X 射线荧光光谱定量分析模型,以期为土壤污染调查分析提供技术支持,也为惩罚最小二乘算法在 X 射线光谱领域的研究应用提供理论基础。

1 材料与方法

1.1 实验仪器

主要仪器:TS-XH4000 型便携式 X 射线荧光光谱仪,浙江泰克松德能源科技有限公司;SDD 探测器,能量分辨率为 125 eV,美国 Amptek 公司;球磨机,江苏宜兴丁蜀浩强机械设备有限公司;样品杯(聚乙烯),尺寸为 $\Phi 3\text{ cm} \times 1\text{ cm}$,单开口,带固定麦拉膜的颈圈;麦拉膜,厚 3.6 μm ,宽 7.6 cm,美国 Chemplex 公司。

1.2 土壤样品制备与光谱采集

本实验中,共计样品 89 个,其中野外采集土壤样品 30 个(江西省鄱阳湖地区),另外 59 个为国家土壤标准样品,其中 20 个水系沉积物标准物质(GSD)和 39 个土壤标准物质(GSS 系列),样品来源多样,且基本覆盖了全国大部分土壤质地。样品采集和制备方法严格按照土壤环境质量标准(GB 15618—2018)中的技术规范执行。将采集的土壤样品自然风干,除去土壤样品中沙石、草屑、动物碎片等明显的异物,使用四分法取 2 份,1 份用于实验分析,1 份备用。将国家标准认证样品和实验分析的土壤样品均匀填入玛瑙钵体中,在球磨仪上研磨 5 min 后,取出粉末过 200 目筛子。将处理后的土壤样品使用便携式 XRF 分析仪在管压 35 keV、电流 40 μA 和时间 90 s 下,采集土壤 X 射线荧光光谱数据,测量时每个样本测量 3 次,3 次移动不同位置,取平均值作为光谱数据,共获取样品在 0~45 keV 范围内共 2 048 个通道的光谱信息。

Kennard-Stone(K-S)算法^[28]作为样本划分

方法将光谱原始数据的 69 个样品作为训练集,20 个样品作为预测集。

1.3 光谱数据处理方法

1.3.1 惩罚最小二乘算法

假设 y 为原始光谱信号,长度为 N 。用 z 表示拟合基线的向量,长度也为 N 。则惩罚最小二乘方法中 z 对 y 的保真度 F 为

$$F = \sum_{i=1}^N \omega_i (y_i - z_i) \tag{1}$$

式中 y_i —— y 在通道 i 时的信号强度
 z_i —— z 在通道 i 时的信号强度
 ω_i ——权重

ω_i 可以表示为对角线矩阵 W

$$W = \begin{bmatrix} \omega_1 & & & & \\ & \omega_2 & & & \\ & & \ddots & & \\ & & & \omega_{N-1} & \\ & & & & \omega_N \end{bmatrix} \tag{2}$$

拟合的基线 z 的粗糙度 R 表示为

$$R = \sum_{i=2}^N (z_i - z_{i-1})^2 = \sum_{i=1}^{N-1} \Delta z_i^2 \tag{3}$$

对于原始数据保真度和拟合基线数据粗糙度之间的平衡,采用保真度权重和平滑参数来调节,可以描述为

$$Q = F + \lambda R = W \|y - z\|^2 + \lambda \|Dz\|^2 \tag{4}$$

式中 D ——2 阶差分矩阵 λ ——平滑参数
 Q ——综合变量,表示原始基线与拟合基线保真度与粗糙度间的平衡

为求解式(4)的极值问题,通过找到偏导数的向量并将其导数设置为零 $\left(\frac{\partial Q}{\partial z} = 0\right)$,得到方程

$$z = (W + \lambda D^T D)^{-1} W y \tag{5}$$

因此,式(2)、(5)是惩罚最小二乘算法的基本形式。该算法的核心思想是如何选择合适的平滑参数 λ 和权重向量 W 。其中,加权方案一直受到大家

的关注和重视。一般来说,非峰值区域的权重应接近于 1,峰值区域的权重应接近于零。

1.3.2 算法实现

根据实验,所有基于惩罚最小二乘(或者改进)算法的程序都在计算机上使用 Matlab 2016b 编写。惩罚最小二乘(或者改进)算法的基本实现步骤为:

- (1)输入原始 X 射线荧光光谱数据 y 。
- (2)加载最优参数:最大迭代次数 Itermax、平滑参数 λ 和初始化权重 ω 。
- (3)迭代次数从 $t = 1$ 开始,对权重 ω 进行初始化 $\omega^1 = [1, 1, \dots, 1]$,则权重矩阵 W 为对角矩阵, $W^{t=1} = \text{diag}(1, 1, \dots, 1)$ 。
- (4)将初始权重 ω^1 代入式(5),通过惩罚最小二乘拟合基线数据 z' 。
- (5)计算 X 射线荧光光谱与拟合基线之间的残差信号 $d = y - z$,标准差 σ 和估计权重 ω 。
- (6)判断是否满足迭代终止条件或者达到最大迭代次数,若不满足转步骤(5),否则转至步骤(7)。
- (7)迭代终止,输出最优基线 z 并计算校正后的光谱 y 。

2 结果与讨论

2.1 光谱预处理

采集到的土壤原始 X 射线荧光光谱数据包含仪器操作、土壤背景和散射效应等引起的噪声和无关信息,因此在对光谱数据进行操作之前,需要对光谱进行平滑去噪处理,有助于消除干扰因素。以国家土壤标准物质 GSS-7 为例,采用信噪比(SNR)作为去噪效果评价指标,去噪处理采用以下方法:样品颗粒分布不均匀及颗粒粒径产生的散射影响,使用多元散射校正法(MSC)消除。光谱平滑去噪采用小波变换进行处理,小波基函数选取 sym4,计算不同分解层数下的 SNR,结果见表 1。可以看出,光谱信号在分解层数为 7 层时,去噪效果最好,效果见图 1,其信噪比为 19.712 7 dB。

表 1 sym4 小波函数在不同分解层数时的去噪结果

Tab.1 sym4 wavelet functions at different decomposition layers denoising results								
小波函数	sym4-2	sym4-3	sym4-4	sym4-5	sym4-6	sym4-7	sym4-8	sym4-9
SNR/dB	12.824 5	13.321 4	15.421 2	18.445 2	17.012 2	19.712 7	18.323 5	15.111 4

2.2 惩罚最小二乘算法基线校正

以国家土壤标准物质 GSS-7 为例,5 种基线校正方法拟合得到基线如图 2 所示。从图 2 中可以看出,DRPLS 算法对基线有较好的拟合效果,所得的估计基线能够更好地拟合本底,估计基线与 X 射线荧光光谱的真实基线最为接近,可以精确拟合土壤

实际光谱目标元素特征峰附近的基线,且基线校正后的本底扣除最为明显。而 ASLS、ARPLS 和 LRSPLS 算法所得估计基线存在着不同程度的位置偏低现象,从而低估了土壤样品实际光谱的基线,而这一现象也间接证明了 ASLS、ARPLS 和 LRSPLS 的加权方案总是受到实际光谱噪声的影响。另外,

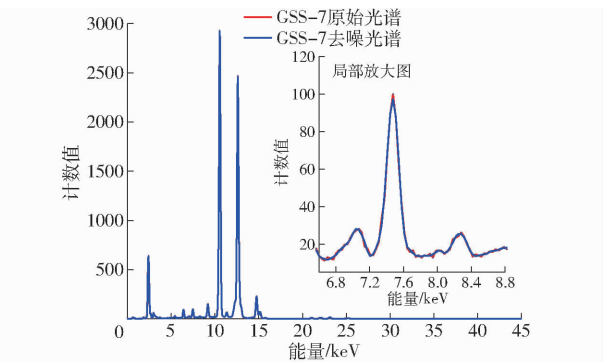
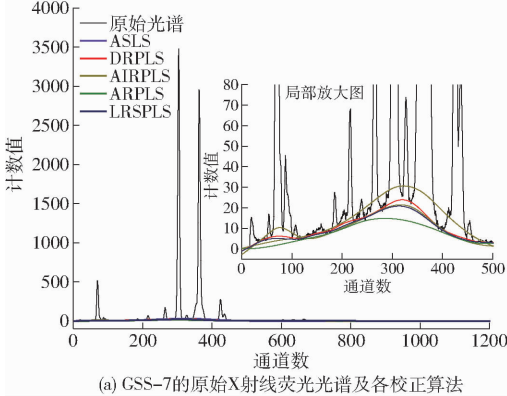


图 1 土壤标准物质 GSS-7 的 X 射线荧光原始光谱与去噪结果

Fig. 1 GSS-7 original X-ray fluorescence spectrum with denoised results

AIRPLS 算法的估计基线沿着底部信号中间穿过,而不是沿着 X 射线荧光光谱真实基线的底部,这将造



成校正后光谱向下凹起,进而导致谱强度偏小,小峰无法识别,从而影响对 X 射线荧光光谱的后续分析,特别是定量分析结果的准确性。

为了更具体地比较不同方法对实测光谱信号拟合的效果,删去校正后光谱特征峰区域的数据,然后求取信号方差。方差越小,说明校正后信号的基线越平缓,拟合效果越好,去除特征峰后信号的方差如表 2 所示。

由表 2 可以看出,与其他 4 种方法相比,DRPLS 算法在土壤样品实测数据中得到的方差最小,为 1.52×10^{-3} ,说明 DRPLS 算法在样品非特征峰区的基线不会太高或太低,校正后的基线最平缓,波动最小,拟合效果最好。综上所述,该方法可以有效地去除 XRF 谱的基线,为后续定量分析模型的精度提高

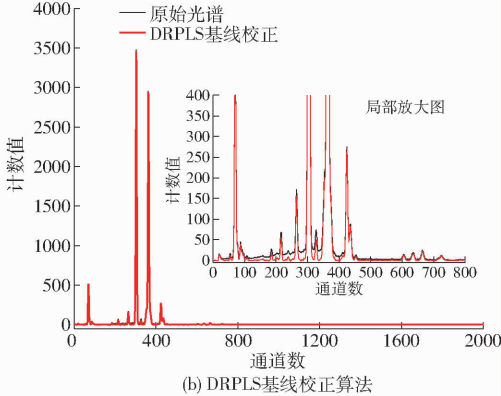


图 2 GSS-7 的 X 射线荧光光谱估计基线与校正结果

Fig. 2 GSS-7 X-ray fluorescence spectra estimate baseline and correction results

表 2 去除特征峰后信号的方差

Tab.2 Variance of signal after removing characteristic peak

方法	方差
ASLS	5.750×10^{-3}
AIRPLS	1.558×10^{-2}
ARPLS	2.730×10^{-3}
LRSPLS	2.490×10^{-3}
DRPLS	1.520×10^{-3}

提供了保障。

利用原始数据和使用 5 种处理方法对土壤样品 X 射线荧光光谱进行基线校正处理后的光谱数据绘制校准曲线,分别基于 PLS 建模,进行铅和砷元素的定量分析,结果见图 2 和表 3。由表 3 可以看出:不同算法基线校正处理对 PLS 模型精度的影响差异明显,6 种光谱基线校正方法中,铅的预测集 R^2 最大为 0.982,最小为 0.912;RMSECV 和 RMSEP 最大分别为 0.233 mg/kg 和 0.239 mg/kg,最小分别为 0.113 mg/kg 和 0.056 mg/kg。图 2 表示 2 种元素的 5 种基线校正方法的校正效果,从图 2 和表 2 中均

可看出,AIRPLS 的效果最差,经基线扣除后的 RMSECV 和 RMSEP 最大,且低于无基线处理的 PLS 模型。图 2b 的 DRPLS 基线校正的结果表明,经基线校正处理后,原始光谱的基线有较明显改善。根据表 2 和表 3 可以看出,DRPLS 基线校正后的模型训练集和预测集 R^2 、RMSECV 和 RMSEP 均有提高,且主成分数也减少,改善了模型的精度。结合表 2、3 和图 2 并从多个指标综合考虑,相比较其他几种基线校正方法,可以得出 DRPLS 基线校正处理方法对 PLS 模型精度改善最好,但也可以看出不是所有的惩罚最小二乘基线校正算法均对 PLS 模型精度有改善,有可能对模型产生负影响。

2.3 不同模型建模对比分析

为了验证该最优算法在不同模型中对定量分析的影响,将铅和砷元素作为定量分析对象,用本实验所用便携式 X 射线荧光光谱仪对土壤样品(共 89 个)进行检测,每个样品测量 3 次,对应的计数率取平均值,与其相应的标准值(CRMs)进行 PLS、BP 和 SVR 建模,以模型决定系数(R^2)和预测均方根误差

表 3 基线校正后的结果
Tab. 3 Results of baseline corrected

元素	基线校正	训练集		预测集		主成分分数
		R^2	RMSECV/ ($\text{mg}\cdot\text{kg}^{-1}$)	R^2	RMSEP/ ($\text{mg}\cdot\text{kg}^{-1}$)	
铅	无基线校正	0.973	0.233	0.978	0.166	3
	AIRPLS	0.891	0.451	0.912	0.239	3
	ASLS	0.975	0.210	0.979	0.096	3
	ARPLS	0.981	0.152	0.980	0.074	2
	LRSPLS	0.973	0.173	0.981	0.099	3
	DRPLS	0.981	0.113	0.982	0.056	2
砷	无基线校正	0.963	1.621	0.966	1.144	4
	AIRPLS	0.855	1.451	0.892	1.239	4
	ASLS	0.955	1.210	0.963	1.162	3
	ARPLS	0.974	0.952	0.977	0.774	3
	LRSPLS	0.971	1.273	0.972	1.120	3
	DRPLS	0.984	0.913	0.985	0.796	3

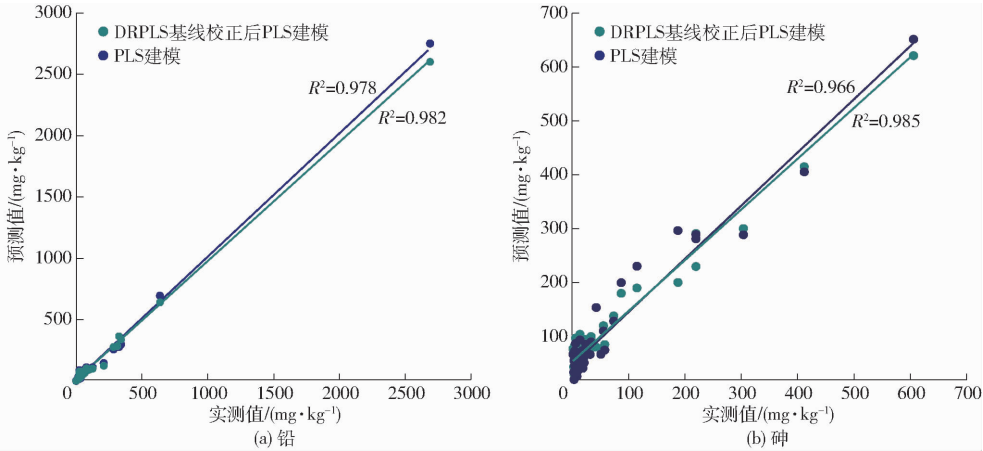


图 3 铅和砷元素的 DRPLS - PLS 模型的建模和预测结果
Fig. 3 Calibration and prediction results of DRPLS - PLS model for Pb and As

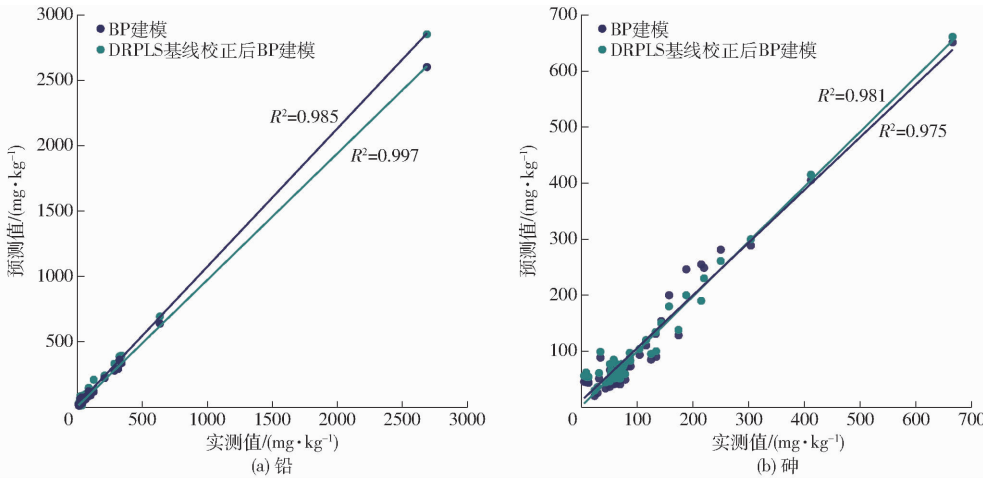


图 4 铅和砷元素的 DRPLS - BP 模型的建模和预测结果
Fig. 4 Calibration and prediction results of DRPLS - BP model for Pb and As

综上所述，SVR 比 PLS 模型和 BP 模型更适用于土壤重金属浓度的测定，因为 SVR 模型具有较高的准确性。

经 DRPLS 基线校正处理后结合 SVR 建模的预

(RMSEP)作为评价标准,对不同模型建模的预测结果进行评价。

评估最佳 PLS、BP 和 SVR 模型的实际预测能力是利用预测集中的 20 个未知样本进行的。采用 DRPLS 对土壤重金属铅和砷的光谱基线校正后进行 PLS、BP 和 SVR 建模,20 个预测集预测值与实测值的拟合结果见图 3 ~ 5。铅和砷的预测值和实测值的 R^2 和 RMSEP 见表 4,通过对比分析 R^2 和 RMSEP,可以得出 DRPLS - SVR 具有最佳回归效果,铅和砷的 R^2 分别为 0.998 和 0.993, RMSEP 分别为 0.015 mg/kg 和 0.596 mg/kg,相比较于无任何基线校正处理,SVR 不仅在精度方面有较大的提升,比 DRPLS - BP 在时间性能方面也有所提高,且与 DRPLS - PLS 相比具有更好的泛化性能和预测效果。

测值与实测值(均值, $n = 3$)的结果如表 5 所示,从表 5 可以看出,DRPLS 基线校正后铅和砷的预测值和实测值的相对标准偏差(RSD)范围分别为 0.56% ~ 4.35%、0.61% ~ 4.85%,未经处理相对标

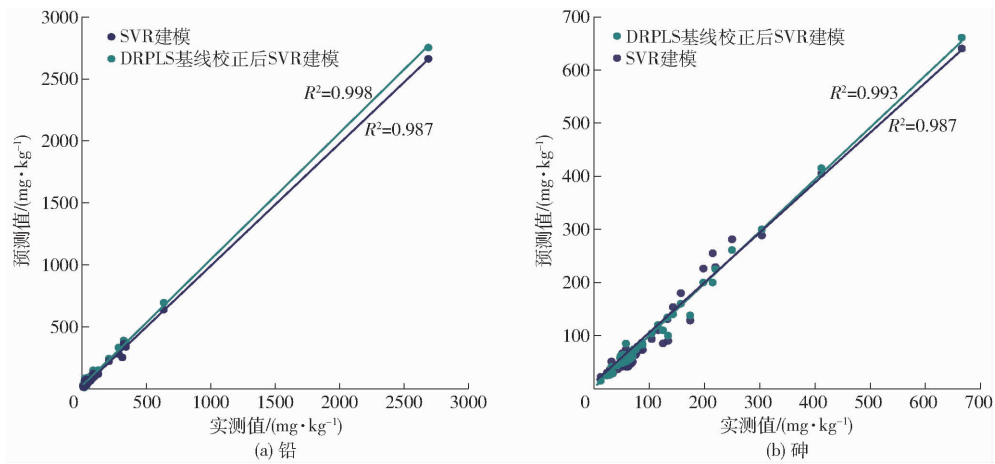


图 5 铅和砷元素的 DRPLS-SVR 模型的建模和预测结果

Fig. 5 Calibration and prediction results of DRPLS-SVR model for Pb and As

表 4 2 种重金属的预测值和实测值的拟合结果
Tab. 4 Fitting results of predicted and actual values of two heavy metals

元素	校正模型	R^2	RMSEP/($\text{mg}\cdot\text{kg}^{-1}$)
铅	DRPLS-PLS	0.982	0.056
	PLS	0.978	0.166
	DRPLS-BP	0.997	0.044
	BP	0.985	0.056
	DRPLS-SVR	0.998	0.015
	SVR	0.987	0.016
砷	DRPLS-PLS	0.985	0.796
	PLS	0.966	1.144
	DRPLS-BP	0.981	0.669
	BP	0.975	0.955
	DRPLS-SVR	0.993	0.596
	SVR	0.987	0.774

表 5 DRPLS 背景扣除前后 SVR 建模的预测结果与实际值

Tab. 5 Prediction results and actual values of SVR modeling before and after DRPLS background deduction

元素	样品 序号	标准值/ ($\text{mg}\cdot\text{kg}^{-1}$)	背景扣除前			背景扣除后		
			实测值/ ($\text{mg}\cdot\text{kg}^{-1}$)	RSD/ %	RE/ %	实测值/ ($\text{mg}\cdot\text{kg}^{-1}$)	RSD/ %	RE/ %
铅	1	98.1	98.2	0.67	0.20	97.6	0.56	0.41
	2	20.2	19.3	2.24	3.50	20.2	0.95	1.00
	3	58.7	61.0	2.36	5.17	60.4	1.91	4.14
	4	14.1	13.4	4.55	4.29	13.6	4.03	2.86
	5	25.5	26.4	5.69	5.60	25.6	1.95	2.40
	6	247.2	251.4	2.61	1.78	245.4	1.74	0.65
	7	19.1	19.6	6.85	3.16	18.8	2.38	1.05
	8	31.2	32.6	4.41	5.16	30.8	2.33	0.65
	9	34.2	33.0	2.14	2.94	33.2	2.52	2.35
	10	137.4	138.9	2.77	1.39	137.6	1.76	1.02
	11	158.1	159.8	4.02	3.10	157.2	3.19	2.28
	12	38.2	37.6	5.06	0.83	38.2	1.67	4.58
	13	84.3	87.6	2.46	4.29	85.2	0.61	1.43
	14	74.4	75.4	5.11	1.89	74.6	0.75	0.81
	15	12.2	12.4	15.10	1.64	12.6	4.35	3.28
	16	36.8	40.4	6.55	6.67	38.6	4.03	1.67
	17	36.1	39.4	4.51	2.67	36.6	2.03	0.97
	18	31.9	33.0	2.74	3.34	31.2	2.62	2.25
	19	57.6	58.9	2.77	1.39	57.2	1.76	1.02
	20	228.0	229.8	5.12	3.13	227.2	3.91	1.68
砷	1	59.0	59.9	5.16	0.93	59.1	1.97	4.98
	2	53.6	53.8	2.16	4.09	53.4	0.61	1.53
	3	57.1	56.2	5.21	1.99	57.2	0.77	0.91
	4	110.5	92.5	13.50	1.54	110.2	4.85	3.18
	5	53.0	51.3	6.55	6.67	53.3	4.03	1.67
	6	53.8	62.6	6.14	6.14	52.6	2.72	2.95
	7	28.7	27.7	2.71	1.69	28.7	1.56	1.12
	8	54.6	53.2	4.12	3.71	54.2	1.89	1.58
	9	51.0	60.7	10.06	6.83	50.7	4.67	3.58
	10	65.8	63.1	4.46	3.29	65.2	0.71	0.93
	11	50.4	52.3	5.16	2.89	50.3	1.75	1.81
	12	65.7	63.1	13.50	4.64	66.1	3.51	3.58
	13	60.2	62.2	6.15	6.17	60.2	1.03	0.67
	14	64.2	63.9	2.14	2.94	64.1	1.12	0.45
	15	43.3	42.2	5.77	3.39	43.6	1.86	1.42
	16	33.6	40.6	4.02	3.10	33.4	1.09	1.68
	17	49.8	51.5	4.06	2.83	50.5	2.67	3.58
	18	44.7	44.0	2.46	4.29	44.5	0.69	1.13
	19	26.5	25.9	2.14	2.81	26.9	1.75	0.91
	20	68.6	66.6	1.16	2.64	68.6	1.35	0.88

3 结论

(1) 研究选取了 89 个土壤样本,采用手持式 XRF 仪获取土壤重金属原始光谱,并对原始光谱采用 sym4 小波基进行小波变换平滑去噪处理。结果表明,在分解层数为 7 层时,去噪效果最好,信噪比最高,达到了 19.712 7 dB。

(2) 对比了土壤重金属光谱的无基线校正处理以及 ASLS、LRSPLS、ARPLS、AIRPLS、DRPLS 等基线

校正方法的处理效果。结果表明,与其他 5 种方法相比,DRPLS 算法在土壤光谱中的方差最小,为 0.001 52,说明 DRPLS 算法校正后的基线波动最小,拟合效果最好。

(3) 采用 DRPLS 算法基线校正后的光谱分别采用 PLS、BP 和 SVR 进行定量建模分析,结果显示 SVR 回归得到的建模效果最好,铅和砷预测值和实测值的 R^2 分别达到 0.998 和 0.993。

参 考 文 献

[1] 卓尚军,李国会,吉昂.能量色散 X 射线荧光光谱[M].北京:科学出版社,2011.

[2] 邓赛文,李松,王毅民,等. X 射线荧光光谱在现场分析中的应用评介[J]. 冶金分析, 2020,40(10):76-85.
DENG Saiwen, LI Song, WANG Yimin, et al. Review on the application of X-ray fluorescence spectrometry in field analysis [J]. Metallurgical Analysis, 2020, 40(10):76-85. (in Chinese)

[3] 张锐,马龙. 手持 X 射线荧光光谱法在矿山生产中的应用[J]. 矿产勘查, 2018,9(2):341-345.
ZHANG Rui, MA Long. Application of hand-held X-ray fluorescence spectrometry analysis method in minerals exploration[J]. Mineral Exploration, 2018, 9(2):341-345. (in Chinese)

[4] 赵奉奎,王爱民. 能量色散 X 射线荧光光谱仪研究现状[J]. 核技术, 2013, 36(10):100402.
ZHAO Fengkui, WANG Aimin. Research status of energy-dispersive X-ray fluorescence spectrometer[J]. Nuclear Technology, 2013, 36(10):100402. (in Chinese)

[5] XI Yang, LI Yuee, DUAN Zhizhen, et al. A novel pre-processing algorithm based on the wavelet transform for Raman spectrum [J]. Appl. Spectrosc, 2018,72(1):1752-1763.

[6] 郭媛,赵学珏,张锐,等. 小波变换应用于谐波谱线的噪声滤除与基线校正[J]. 光谱学与光谱分析, 2013,33(8):2172-2176.
GUO Yuan, ZHAO Xuegong, ZHANG Rui, et al. The noise filtering and baseline correction for harmonic spectrum based on wavelet transform[J]. Spectroscopy and Spectral Analysis, 2013, 33(8):2172-2176. (in Chinese)

[7] 李玲. 近红外光谱基线漂移校正的方法研究[J]. 激光杂志, 2019,40(2):153-156.
LI Ling. Research on method of baseline drift correction in near infrared spectroscopy [J]. Laser Journal, 2019, 40(2):153-156. (in Chinese)

[8] HU Haibing, BAI Jing, XIA Guo, et al. Improved baseline correction method based on polynomial fitting for Raman spectroscopy[J]. Photon. Sens., 2018, 18(4):332-340.

[9] 宁志强,刘家祥,吴越,等. 基于改进迭代多项式拟合的红外光谱基线校正方法[J]. 激光与光电子学进展, 2020,57(3):247-253.
NING Zhiqiang, LIU Jiexiang, WU Yue, et al. Infrared spectrum baseline correction method based on improved iterative polynomial fitting[J]. Laser Optoelectronics Progress, 2020, 57(3):247-253. (in Chinese)

[10] WANG Tuo, DAI Liankui. Background subtraction of Raman spectra based on iterative polynomial smoothing[J]. Appl. Spectrosc, 2017,71(6):1169-1179.

[11] 王丽,邵利民. 基线对傅里叶变换红外光谱的目标因子分析结果的影响研究[J]. 分析化学, 2021,49(3):474-481.
WANG Li, SHAO Limin. Influence of baseline on target factor analysis results of fourier transform infrared spectroscopy[J]. Chinese Journal of Analytical Chemistry, 2021, 49(3):474-481. (in Chinese)

[12] TANG Xiaojun, WANG Jin, ZHANG Lei, et al. Spectral baseline correction by piecewise dividing in Fourier transform infrared gas analysis[J]. Spectroscopy and Spectral Analysis, 2013,33(2):334-339.

[13] 赵安新,汤晓君,张钟华,等. Rubberband 方法中分段数量的选择对基线校正效果的影响分析[J]. 红外与激光工程, 2015(4):1172-1177.
ZHAO Anxin, TANG Xiaojun, ZHANG Zhonghua, et al. Influence analysis of Rubberband's piecewise quantity on baseline correction effect[J]. Infrared and Laser Engineering, 2015(4):1172-1177. (in Chinese)

[14] VIDAL J, PUEYO R, SOSANNA M. Automatic morphology based cubic p-spline fitting methodology for smoothing and baseline-removal of Raman spectra[J]. J. Raman Spectrosc, 2017, 48:878-883.

[15] CHEN Yunliang, DAI Liankui. An automated baseline correction method based on iterative morphological operations[J]. Appl. Spectrosc, 2018,72(5):731-739.

[16] WHITTAKER E T. On a new method of graduation[J]. Proc. Edinb. Math. Soc., 1922,41:63-75.

[17] EILERS P H C. A perfect smoother[J]. Anal. Chem., 2003,75(14):3631-3636.

[18] ZHANG Zhimin, CHEN Shan, Liang Yizeng, et al. Baseline correction using adaptive iteratively reweighted penalized least squares[J]. Analyst, 2010,135(5):1138-1146.

[19] BAEK S J, PARK A, AHN Y J, et al. Baseline correction using asymmetrically reweighted penalized least squares smoothing [J]. Analyst, 2015, 140(1):250-257.

[20] 赵恒,陈娱欣,续小丁,等. 基于局部对称加权惩罚最小二乘的拉曼基线校正[J]. 中国激光, 2018, 45(12):274-285.

- ZHAO Heng, CHEN Yuxin, XU Xiaoding, et al. Baseline correction for raman spectra based on locally symmetric reweighted penalized least squares[J]. Chinese Journal of Lasers, 2018, 45(12): 274–285. (in Chinese)
- [21] XU Degang, LIU Song, CAI Yaoyi, et al. Baseline correction method based on doubly reweighted penalized least squares[J]. Appl. Opt., 2019, 58(14): 3913–3920.
- [22] WU Wei, WANG Jian, CHENG Mingsong, et al. Convergence analysis of online gradient method for BP neural networks[J]. Neural Netw., 2011, 24(1): 91–98.
- [23] 冯帅, 许童羽, 周云成, 等. 基于深度卷积神经网络的水稻知识文本分类方法[J/OL]. 农业机械学报, 2021, 52(3): 257–264.
FENG Shuai, XU Tongyu, ZHOU Yuncheng, et al. Rice knowledge text classification based on deep convolution neural network [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(3): 257–264. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20210329&flag=1. DOI: 10.6041/j.issn.1000-1298.2021.03.029. (in Chinese)
- [24] TANG Yongsheng, CHEN Zhengguang. Soil pH prediction based on convolution neural network and near infrared spectroscopy [J]. Spectroscopy and Spectral Analysis, 2021, 41(3): 892–897.
- [25] YANG Xun, CHEN Ying, LIANG Hui. Square root based activation function in neural networks[C]//Proc. IEEE ICALIP, 2018: 84–89.
- [26] 周华茂, 陈添兵, 刘木华, 等. 基于粒子群算法-支持向量机-激光诱导击穿光谱技术对稻壳中铬元素的定量分析模型[J]. 分析化学, 2020, 48(6): 811–816.
ZHOU Huamao, CHEN Tianbing, LIU Muhua, et al. Quantitative analysis of chromium in rice husks by laser induced breakdown spectroscopy based on particle swarm optimization – support vector machine[J]. Chinese Journal of Analytical Chemistry, 2020, 48(6): 811–816. (in Chinese)
- [27] FU Xinghu, ZHAO Fei, WANG Zhenxing, et al. Quantitative analysis of goat serum protein content by Raman spectroscopy based on IABC – SVR[J]. Spectroscopy and Spectral Analysis, 2021, 41(2): 540–545.
- [28] KENNAND R W, STONE L A. Computer-aided design of experiments[J]. Technometrics, 1969, 11(1): 137–148.

(上接第 204 页)

- [27] 马友忠, 孟小峰. 云数据管理索引技术研究[J]. 软件学报, 2015, 26(1): 145–166.
MA Youzhong, MENG Xiaofeng. Research on cloud data management index technology[J]. Journal of Software, 2015, 26(1): 145–166. (in Chinese)
- [28] ZHANG X Y, AI J, WANG Z Y, et al. An efficient multidimensional index for cloud data management[C]//Proc. of the 1st Int'l Workshop on Cloud Data Management. New York: ACM, 2009: 17–24.
- [29] 孟必平, 王腾蛟, 李红燕, 等. 分片位图索引: 一种适用于云数据管理的辅助索引机制[J]. 计算机学报, 2012, 35(11): 2306–2316.
MENG Biping, WANG Tengjiao, LI Hongyan, et al. Regional bitmap index: a secondary index for data management in cloud computing environment[J]. Chinese Journal of Computers, 2012, 35(11): 2306–2316. (in Chinese)
- [30] RAJESHBABU C. Hindex[EB/OL]. [2016–06–15]. <https://github.com/Huawei-Hadoop/hindex>.
- [31] The Apache Software Foundation. Apache Phoenix overview[EB/OL]. 2019. <https://phoenix.apache.org>.
- [32] 苑严伟, 冀福华, 赵博, 等. 基于 Solr 的农田数据索引方法与大数据平台构建[J/OL]. 农业机械学报, 2019, 50(11): 186–192.
YUAN Yanwei, JI Fuhua, ZHAO Bo, et al. Farmland data indexing method based on Solr and construction of big data platform [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(11): 186–192. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20191121&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2019.11.021. (in Chinese)
- [33] ZOU Y Q, LIU J, WANG S C, et al. CCIndex: a complementary clustering index on distributed ordered tables for multi-dimensional range queries[C]//Proc. of the 7th IFIP Int'l Conf. on Network and Parallel Computing. Berlin: Springer-Verlag, 2010: 247–261.
- [34] NISHIMURA S, DAS S, AGRAWAL D, et al. MD-HBase: a scalable multi-dimensional data infrastructure for location aware services[C]//Proc. of the 11th Int'l Conf. on Mobile Data Management. Washington: IEEE, 2011: 7–16.
- [35] 钱志彬, 赵卫东. 轻量级 WebGIS 中间件的设计与实现[J]. 计算机工程, 2006(23): 265–267, 277.
QIAN Zhibin, ZHAO Weidong. Design and implementation of lightweight WebGIS middleware[J]. Computer Engineering, 2006(23): 265–267, 277. (in Chinese)
- [36] GORMLEY C, TONG Z. Elastic search the definitive guide: a distributed real-time search and analytics engine[J]. Oreilly Media, 2015, 1(1): 1–4.
- [37] 马智勤, 廖雪花, 邓威, 等. 基于分布式 Elasticsearch 相似内容比对算法研究[J]. 计算机与数字工程, 2020, 48(12): 2843–2849.
MA Zhiqin, LIAO Xuehua, DENG Wei, et al. Research on similarity content allocation algorithm based on distributed Elasticsearch [J]. Computer and Digital Engineering, 2020, 48(12): 2843–2849. (in Chinese)