

基于土壤背景估计算法和SpectraNet联合的XRF土壤重金属Zn超标分析方法研究

于腾越*, 李 野, 赵 鹏

长春理工大学物理学院, 吉林 长春

收稿日期: 2026年1月12日; 录用日期: 2026年2月6日; 发布日期: 2026年2月14日

摘 要

随着社会现代化进程的迈进, 频繁的人类活动加剧了土壤的重金属污染。当土壤中的重金属元素含量超过风险筛选值, 会经过食物链摄入人体, 过量的重金属累积对人体健康造成损害。因此, 识别含有重金属污染风险的土壤样本, 是开展污染防治工作的关键步骤。本研究采用X射线荧光(XRF)光谱仪获取了22份国家标准土壤样品的光谱数据, 并结合airPLS与高斯卷积方法对光谱进行本底扣除等预处理操作; 最后, 利用本研究提出的深度学习模型SpectraNet对土壤样本是否存在重金属污染风险进行预测。实验结果显示, 该模型在多个土壤元素数据集上展现出较好的识别性能和稳定性, 本研究所提出的网络在多种土壤的定性分析任务中展现了更为全面的判别性能, 在资源受限的情况下能够实现快速推理。

关键词

X射线荧光光谱, 背景扣除, 自适应迭代重加权惩罚最小二乘法(airPLS), 锌元素, SpectraNet

Research on the Analysis Method of Excessive Zn in Soil Based on Soil Background Estimation Algorithm and SpectraNet Combined with XRF

Tengyue Yu*, Ye Li, Peng Zhao

School of Physics, Changchun University of Science and Technology, Changchun Jilin

Received: January 12, 2026; accepted: February 6, 2026; published: February 14, 2026

*通讯作者。

文章引用: 于腾越, 李野, 赵鹏. 基于土壤背景估计算法和 SpectraNet 联合的 XRF 土壤重金属 Zn 超标分析方法研究[J]. 物理化学进展, 2026, 15(1): 39-50. DOI: 10.12677/japc.2026.151005

Abstract

With the advancement of social modernization, frequent human activities have intensified the heavy metal pollution of soil. When the content of heavy metal elements in the soil exceeds the risk screening value, they will be ingested into the human body through the food chain, and excessive accumulation of heavy metals can cause damage to human health. Screening out the soil with heavy metal pollution is an important step in soil pollution control. The spectral data of 22 national standard soil samples were obtained by using X-ray fluorescence (XRF) spectrometer, and then the background subtraction was carried out by combining airPLS and Gaussian convolution for pretreatment. Finally, the SpectraNet algorithm was used to predict whether the soil samples had the risk of heavy metal pollution. The experimental results show that the model has good recognition performance and stability on multiple soil element datasets. The network proposed in this study has shown more comprehensive discrimination performance in the qualitative analysis tasks of various soils and can achieve rapid inference under resource-limited conditions.

Keywords

X-Ray Fluorescence Spectroscopy, Background Subtraction, Adaptive Iteratively Reweighted Penalized Least Squares (airPLS), Zinc, SpectraNet

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

土壤中的重金属污染是世界上突出的环境污染问题之一。随着社会城市化的快速发展,燃料燃烧、矿产开采、农药使用等人类活动[1][2]加剧着土壤中重金属的污染程度。镍、铜、砷、铅、锌等重金属在人类生活和工业排放中通过大气沉降和农业灌溉等方式进入土壤,并且长时间滞留难以降解。这些累积在土壤中的重金属对生态环境影响大,同时也对人类的健康构成了严重威胁。锌是植物必需的微量元素之一。锌参与色氨酸代谢,是生长素合成的关键调节因子,缺锌会间接导致植物生长素含量下降,表现为植株矮小、节间短、生长停滞;同时锌也是许多酶的活化剂,通过对植物碳、氮代谢产生广泛的影响,因此有助于光合作用;同时锌还可增强植物的抗逆性;提高籽粒重量,改变籽实与茎秆的比率[3]。锌在土壤中富集,必然导致在植物体内的富集,这种富集不仅对植物,而且对食用这种植物的人和动物都有危害。用含锌污水灌溉农田会造成小麦出苗不齐,分蘖少,植株矮小,叶片发生萎黄。过量的锌还会使土壤酶失去活性,细菌数目减少,土壤中的微生物作用减弱[4]。过量锌也会对人体健康产生危害。能抑制吞噬细胞的活性和杀菌力,从而降低人体的免疫功能,使抗病能力减弱,而对疾病易感性增强。过量的锌能抑制铁の利用,从而影响铁的吸收与利用,导致人体发生顽固性缺铁性贫血,并且在体内高锌情况下,即使服用铁制剂,也很难使贫血治愈。人体内含有的元素硒具有抗癌功能,但由于锌与硒有拮抗性,所以过量锌能减弱硒的生理作用,如果体内含锌量高,将会导致人体抗癌能力降低,甚至刺激肿瘤生长[5]。鉴于我国南方红壤、黄壤等多表现为酸性,pH 值在 5.0~6.5 [6][7],土壤中重金属的迁移转化特性更为复杂。根据 GB15618-2018《中华人民共和国国家标准土壤环境质量农用地土壤污染风险管控标准(试行)》,当土壤中重金属元素含量超过其风险筛选值时,可能会对人体健康造成威胁,应该开展进一步的详细调查以评估该区域的土壤污染风险程度。然而,土壤重金属污染具有隐蔽性[8],人们很难凭借肉

眼分辨出土壤中是否存在重金属及其含量是否超标, 需要依靠相关检测技术对土壤中的元素含量进行测定及风险评估。能量色散型 X 射线荧光光谱(ED-XRF)分析是一种常用的无损分析、快速检测元素的方法, 因其制样简单、快捷、环保、经济、准确有效等优点被广泛用于土壤、合金等的重金属检验[9][10]。在土壤 EDXRF 光谱分析中, 背景估计通常采用空白实验的方法获取: 以高纯二氧化硅粉末代替实际样品, 重复采集能谱并取均值作为本底基线。该方案虽然操作简易, 但二氧化硅与实际土壤差异显著, 难以复现多组分基体吸收和多元素特征峰叠加等效应, 因此得到的基线与实际土壤光谱中的本底基线存在系统性偏差, 尤其在元素含量跨度大或颗粒组成差异明显的样品上更为突出。由于通常需要检测的重金属含量很低, 元素特征峰会出现与其对应干扰元素特征峰发生谱线重叠, 从而导致元素检测分析结果的误差显著增加[11], 因此在采用 ED-XRF 分析时, 对光谱背景基线进行准确的估计是重要的。吴和喜等[12]使用统计敏感的迭代剥峰算法(SNIP)对 γ 能谱仪测得的谱线进行本底扣除, Tan 等人[13]将迭代小波变换法(IWT)应用于多变量光谱数据的基线校正, 有效消除了非恒定、变化的背景干扰; Zhang 等[14]通过自适应迭代重加权惩罚最小二乘法(airPLS)应用到拉曼光谱中进行背景基线估计中, 但尽管上述方法在各自的应用场景中已取得一定的效果, 但面对不同类型的土壤中多种重金属的 XRF 光谱时, 其通用性仍显不足。同时随着人工智能算法进入成分分析领域, 解决土壤重金属污染风险测定出现了新思路。何迎一[15]实现了基于支持向量机建立土壤重金属污染评价模型, Hu [16]等利用随机森林算法来预测土壤重金属含量。本工作以土壤中重金属 Zn 元素为研究对象, 采集国家标准土壤样品的 XRF 光谱数据, 通过小波阈值去噪、airPLS 结合高斯卷积对谱线进行预处理, 然后利用 SpectraNet 模型, 将经过本底扣除后的一维光谱数据作为模型的输入, 对土壤中的重金属污染进行风险筛选。

2. 实验部分

2.1. 样品与仪器

实验样品采用 22 份国家标准土壤样品, 主要包含 GSS-5、GSS17、GSS-20 和 GSS23 等 GSS 土壤成分分析标准物质, 还包含了地方土壤成分分析标准物质和 GSD 水系沉积物成分分析标准物质共三个系列土壤样品。实验使用了 EDXRF 光谱仪, 工作电压为 40 kV, 工作电流为 100 μ A。激发源为 Mo 靶 X 射线光管, 光子激发特定样品后, 通过 SDD 探测器接收特定样品被激发后的能量, 并通过采集板进行解析, 从而生成能谱。

2.2. 数据分析方法和评价指标

2.2.1. 自适应迭代重加权惩罚最小二乘法原理

自适应迭代重加权惩罚最小二乘法(adaptive iterative re-weighted penalized least squares, airPLS)常用于一维光谱的基线校正[14]。算法将最小二乘平滑与非平滑惩罚项相结合, 并引入了可迭代更新的权重向量, 在无需预先标注峰区的情况下自适应区分噪声或背景, 即使在峰形复杂的数据中也能得到贴近真实本底的平滑曲线。在迭代过程中, 算法会根据残差符号调节权重: 位于暂定基线之上的点(正残差)权重指数衰减, 减弱其对下一轮估计的影响; 位于基线之下的点(负残差)保留较大权重, 加强对基线的下拉约束。使估计结果避开峰形并更贴近真实背景, 直至收敛。

假设原始信号 $x = [x_1, \dots, x_N]^T$ 的长度为 N , 为等间隔采样得到的信号序列, z 为待拟合基线, 拟合的基线应大致跟随 x 具有相同的趋势, 通过 F 代表误差平方和:

$$F = \sum_{i=1}^N (x_i - z_i)^2 \quad (2.1)$$

同时可以利用基线相邻两项的平方和用来表示粗糙度 R :

$$R = \sum_{i=2}^N (z_i - z_{i-1})^2 \quad (2.2)$$

通过误差平方和加上粗糙度的惩罚，可以构成：

$$Q = F + \lambda R \quad (2.3)$$

这里的 λ 作为超参数调整，较大的 λ 可以带来更加平滑的基线，相反较小的 λ 使得基线更加贴切于原始信号。通过引入自适应迭代重加权向量 w_i^t ，其中 t 是迭代次数，加权向量 w 初始值为 0，在每一次迭代中，我们均计算加权惩罚最小二乘式 Q^t 。

$$Q^t = \sum_{i=1}^N w_i^t |x_i - z_i^t|^2 + \lambda \sum_{j=2}^N w_j^t |z_j^t - z_{j-1}^t|^2 \quad (2.4)$$

对于自适应迭代重加权的过程是在迭代过程中计算权重控制惩罚，从而控制基线的平滑性。权重向量 w_i^t 迭代公式为：

$$w_i^t = \begin{cases} 0, & x_i \geq z_i^{t-1} \\ \frac{t(x_i - z_i^{t-1})}{|d^t|}, & x_i < z_i^{t-1} \end{cases} \quad (2.5)$$

其中，向量 d^t 是在第 t 次迭代过程中，原始信号 x 与 z^{t-1} 之差的小于 0 部分所组成。在上一次迭代中，我们令 z^{t-1} 作为基线的参考值，此时如果原始信号 x_i 大于基线的参考，则认为它是峰值的一部分，令其权值设为 0，在下一次迭代中此点则被忽略。反之根据式(2.5)调整权值，迭代将在最大迭代次数或满足下式时终止：

$$|d^t| < \varepsilon \times |x_i| \quad (2.6)$$

式中， ε 为阈值系数，经验值通常取 0.001 [17]。

2.2.2. 基于改进 airPLS 结合高斯卷积的土壤背景估计算法

为了提升背景拟合的精度，本研究提出了一种基于改进 airPLS 结合高斯卷积的土壤背景估计算法，该方法在保留 airPLS 方法局部平滑优势的基础上，通过划分光谱窗口进而克服了其在特征峰附近容易过拟合的问题，并有效消除了全局特征峰对部分区域的干扰。此外，所采用的高斯卷积可避免相邻窗口背景曲线之间出现断裂现象而造成的误判。算法流程图如图 1 所示。

(1) 输入原始土壤 XRF 光谱 $x = [x_1, \dots, x_N]^T$ 。

(2) 通过小波变换分离出信号中的噪声部分，方便后续分析。

(3) 将经过去噪处理后的光谱划分为 m 个等宽光谱窗口 X ，并为每个窗口设置 airPLS 算法最大迭代次数 n 。

(4) 分别对每个光谱窗口进行 airPLS 基线估计算法，得到第 i 个窗口在第 t 次迭代时的背景基线 B_i^t 。根据式(2.6)，若 B_i^{t-1} 与 B_i^t 之间的误差小于设定阈值 ε 则将 B_i^t 视为该窗口的最优背景基线，否则继续迭代至设定的最大迭代次数 n ，算法停止。其中判别基线误差的公式为：

$$\sum_{i=1}^N \left| \frac{B_i^t - B_i^{t-1}}{B_i^t} \right| < \varphi \quad (2.7)$$

(5) 最后，通过高斯卷积滤波平滑各个窗口的背景基线，得到全局背景基线 B 。使用宽度为 5 的高斯卷积核对各窗口的背景基线进行迭代平滑，以消除断裂现象[18]。高斯卷积核构造公式如下：

$$Kernel = \left[\frac{G(-2)}{\sum_{x=-2}^2 G(x)} \dots \frac{G(0)}{\sum_{x=-2}^2 G(x)} \dots \frac{G(2)}{\sum_{x=-2}^2 G(x)} \right] \in R^{1 \times 5} \quad (2.8)$$

令经过高斯卷积后的第 q 个平滑本底为 B^q ，第 $q+1$ 个卷积的计算公式如下：

$$\hat{B}^{q+1}(k) = \sum_{z=-2}^2 B^q(z) * Kernel(k-z), k=1, 2, \dots, N \quad (2.9)$$

其中， k 表示 XRF 光谱中的通道数。为了防止高斯卷积后的平滑背景基线被抬高，令第 $q+1$ 次的高斯卷积滤波得到的背景基线为：

$$B^{q+1}(k) = \min \{ B^q(k), \hat{B}^{q+1}(k) \}, k=1, 2, \dots, N \quad (2.10)$$

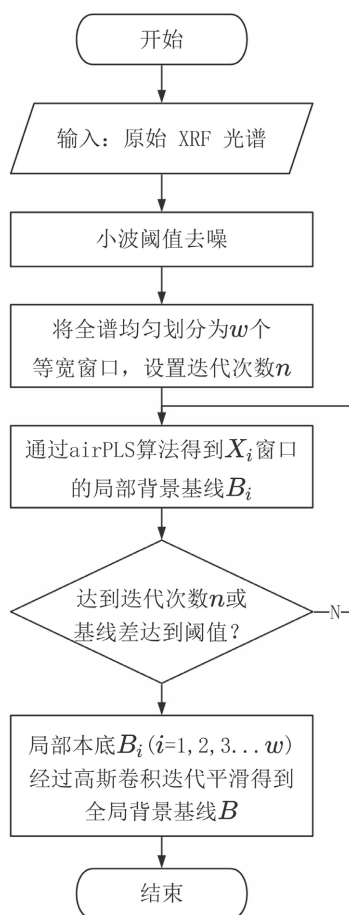


Figure 1. Flowchart of the soil background estimation algorithm based on airPLS and gaussian convolution

图 1. 基于 airPLS 结合高斯卷积的土壤背景估计算法流程图

2.2.3. 基于 SpectraNet 的土壤重金属元素超标判别研究

为了识别土壤中的重金属元素是否超标，本研究提出了 SpectraNet 模型，网络输入为背景扣除处理后的土壤光谱数据的 2048 个通道计数值，整体仅由输入层、卷积模块、残差注意模块层，以及分类模块层所组成，该网络同时兼顾了轻量化设计和良好的光谱数据特征提取性能，能够在计算资源受限的嵌入式设备上部署，具备实时推理的能力。网络结构如图 2 所示。

通过基于改进 airPLS 结合高斯卷积的土壤背景估计算法得到原始土壤光谱数据的背景基线，原始光谱数据与背景基线相减即为背景扣除后的光谱，并将其作为网络整体的输入。首先通过卷积模块提取初步的局部频域特征，该模块由一维卷积核和最大池化组合构成，兼具平滑与降维作用。接下来通过的残差注意模块层则为模型的核心特征提取阶段，分别叠加了标准残差结构与 Squeeze-and-Excitation (SE) 注意机制模块，该结构能够有效捕捉不同尺度下的谱图细节特征，同时借助 SE 注意力机制对通道信息进行自适应重加权，增强模型对高贡献通道特征的关注能力，抑制冗余无效通道，提升特征表达的鲁棒性。网络末端通过自适应平均池化压缩特征序列长度，并通过 Dropout 计算剔除冗余信息，最后使用全连接层输出类别概率，用于判断元素是否超标。相对于依赖手工特征或全局统计量的传统方法不同，SpectraNet 通过一维残差卷积与 SE 注意力在端到端训练中能够捕捉光谱数据的局部形状与多尺度邻域关系，从而更加适合 XRF 光谱的物理特性。

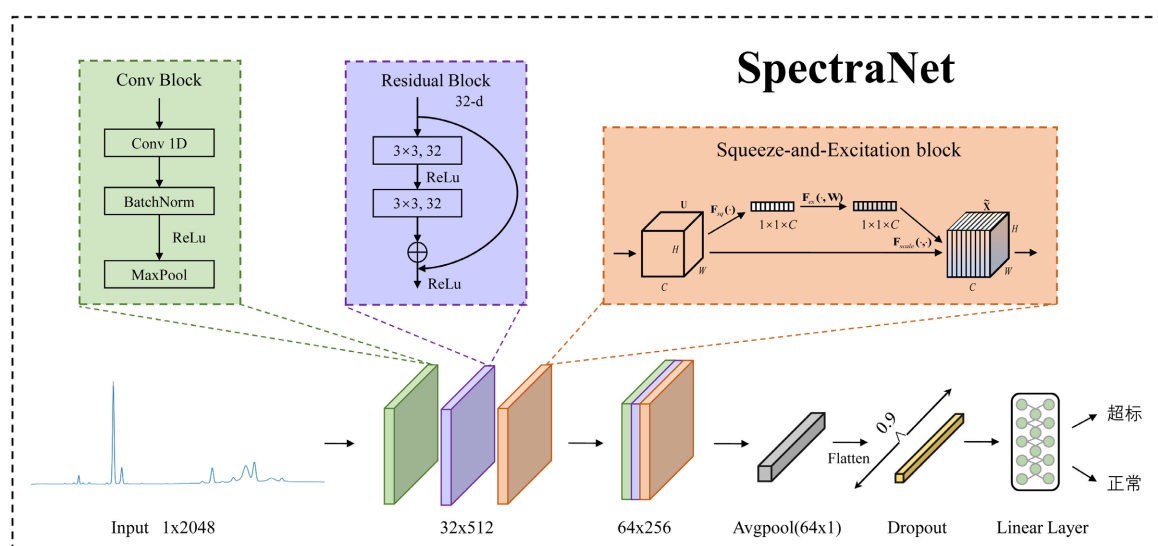


Figure 2. Schematic diagram of SpectraNet structure

图 2. SpectraNet 结构示意图

3. 结果与讨论

3.1. 土壤背景估计算法对光谱数据的处理

3.1.1. 背景估计算法的有效性

如图 3(A)~(D)所示，实验选取了 GSS-5、GSS-17、GSS20、GSS23 四种典型土壤样品的光谱为研究对象，我们采用原始 airPLS 算法作为基准方法，对比验证基于改进 airPLS 结合高斯卷积的背景估计算法的有效性。从整体趋势来看，原 airPLS 算法应用于土壤背景估计中更偏向全局光滑，对慢变化的背景能够保持一致性，但在重叠峰、强弱峰交替附近或峰变化较快的区段往往呈现欠拟合；相对而言，本研究提出的“先分窗、后拼接”策略在不改变 airPLS 基本思想的前提下补足了这一短板，即能保证总体基线相对平滑，又注重了局部基线细节不丢失。

为了进一步验证提出背景估计方法的有效性，本研究将 IWT、SNIP 和基于改进 airPLS 结合高斯卷积的背景估计算法在四种土壤上进行应用，如图 4(A)~(D)所示，SNIP 方法在高峰值附近易收到峰形牵引，基线向上鼓起表现出过拟合，经过背景扣除后会出现峰型被过度的扣除的结果；而 IWT 方法在整体趋势虽然表现光滑，但其对强峰区域不敏感，表现出了背景低估的现象。

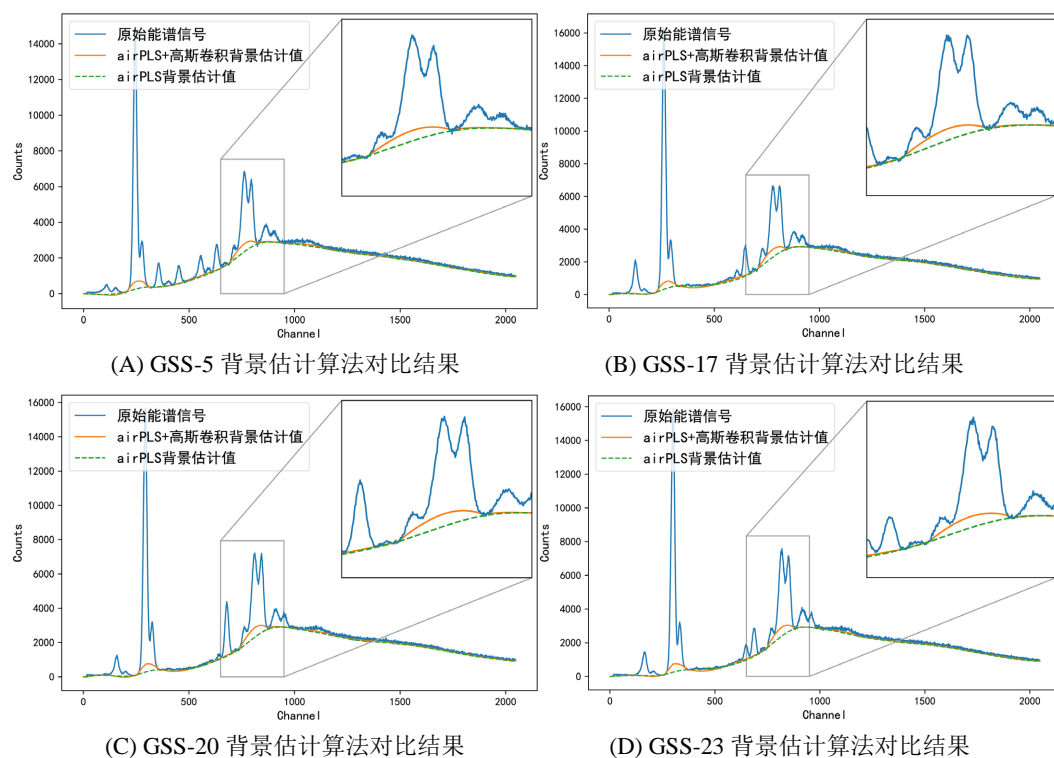


Figure 3. airPLS vs. improved airPLS-gaussian convolution background estimation

图 3. airPLS 与基于改进 airPLS 结合高斯卷积的背景估计算法对比

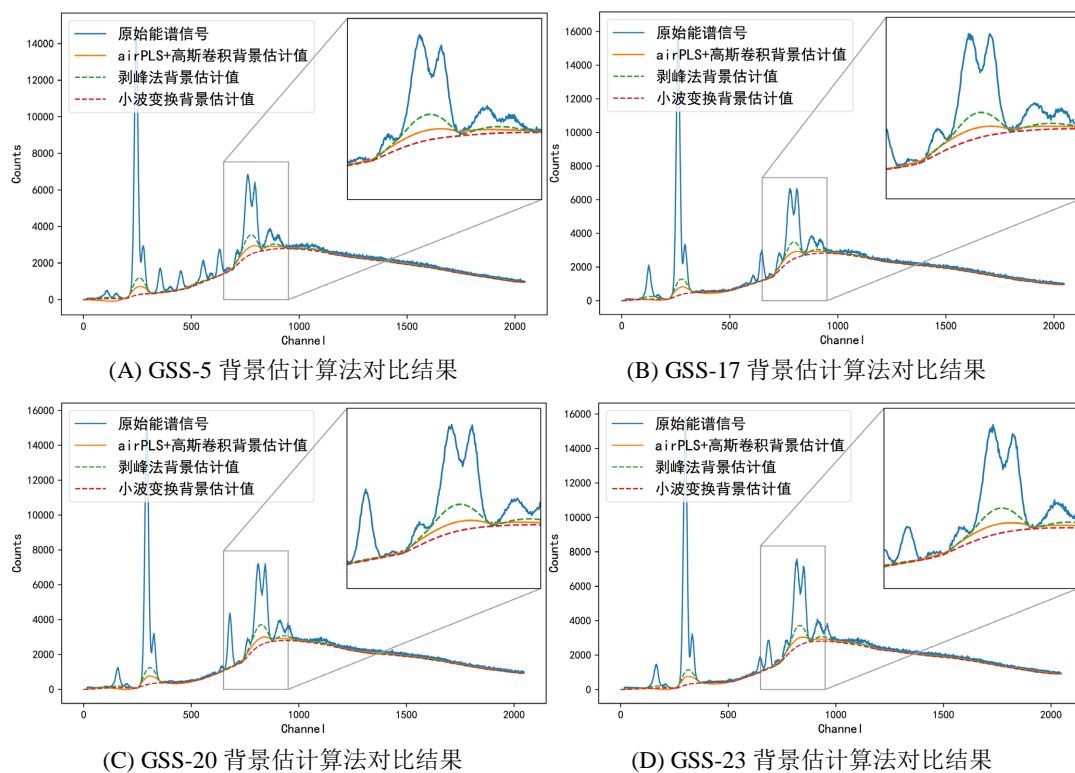


Figure 4. Comparison of background curves derived from different background estimation methods

图 4. 基于不同背景估计方法获得的背景谱线对比

3.1.2. 算法定量分析准确性

由于本底的存在会在一定程度上干扰净峰面积(特征峰计数)与元素浓度的正相关关系,干扰程度会因元素差异而不同,为了证明本底信号能够被精准扣除、同时保留完整的峰信息,可以利用这种线性关系来表现背景估计算法的准确性。本研究以重金属 Zn ($K_{\alpha} = 8.63 \text{ keV}$)为研究对象,对 22 份国家标准土壤样品应用迭代小波变换法(IWT)、剥峰法(SNIP)和基于改进 airPLS 结合高斯卷积的背景估计算法进行背景扣除。如图 5 所示,红色圆点代表原始光谱数据,蓝色圆点代表背景扣除后的光谱数据,通过对二者线性回归,描述出两条不同的直线,并使用 R^2 值来表示背景扣除前后计数值与元素浓度线性关系的变化,从而体现不同背景扣除算法的准确性。

通过对比不同的背景估计算法,可以看出 SNIP 在对 22 种土壤样品进行背景估计时,其本底扣除后的拟合线较远偏离了原始数据的拟合直线,同时 R^2 值下降了 0.08,实验结果表明 SNIP 在 Zn 元素的能道处的背景估计效果不具备适用性,并具有过拟合的表现;迭代小波变换法虽然在原始数据拟合线的基础上 R^2 值提升了 0.005,但其与原始数据拟合线相距较近,分析出 IWT 在土壤背景估计应用中表现为欠拟合。相对而言,基于改进 airPLS 结合高斯卷积的背景估计算法不仅在背景扣除后的拟合线 R^2 值提升了 0.018,同时两条拟合线有平行的趋势,这表现了本研究提出的方法在背景估计时的准确性和稳定性,证明了本底扣除准确的将背景信号过滤,使得光谱信号更“干净”。

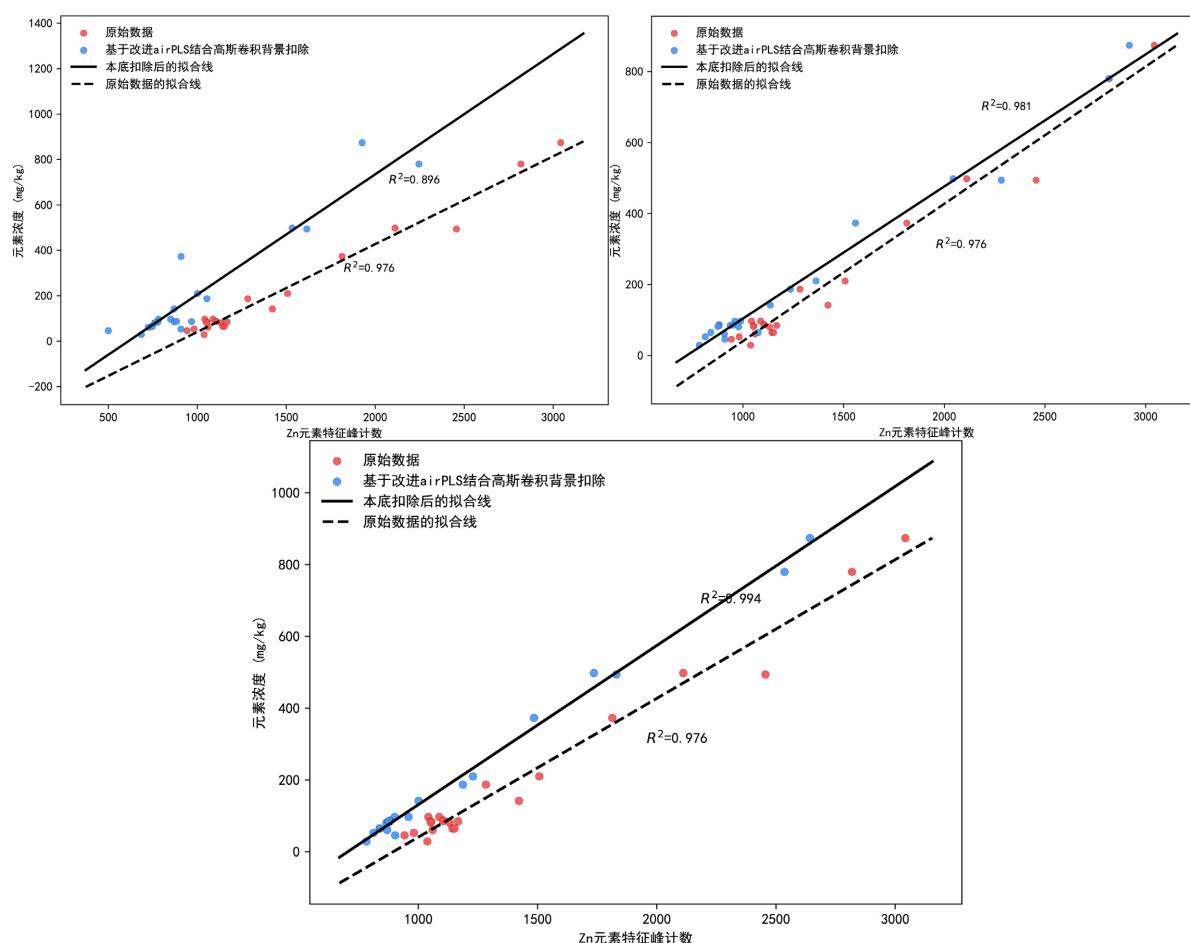


Figure 5. Comparison of linear regression results under different background estimation methods

图 5. 基于不同背景估计算法的线性回归对比

3.2. 基于 SpectraNet 的土壤重金属元素超标判别实验验证

本研究为了提高重金属元素超标判别的普适性，使用 EDXRF 光谱仪对土壤样品进行光谱测量，其中工作电压为 40 kV，工作电流为 100 μ A。考虑到测量仪器的随机误差影响，分别对 22 份国家标准土壤样品测量 3 次，通过基于改进 airPLS 结合高斯卷积的背景估计算法对其进行背景扣除，并将经过预处理后的数据输入到 SpectraNet 中训练，获得能够判别土壤样品中某重金属元素是否超标的分类模型，结构及参数量如表 1 所示。

Table 1. Structure table of SpectraNet model
表 1. SpectraNet 模型结构表

层类型	输入	输出	参数详情	参数量
输入	(1,2048)	(1,2048)	-	0
Conv1d	(1,2048)	(32,1024)	Conv (1 \rightarrow 32, $k = 7$, $s = 2$, $p = 3$)	224
BatchNorm1d	(32,1024)	(32,1024)	γ, β	64
ReLU	(32,1024)	(32,1024)	-	0
MaxPool1d	(32,1024)	(32,512)	$k = 3$, $s = 2$, $p = 1$	0
ResidualBlock	(32,512)	(32,512)	2xConv (32 \rightarrow 32, $k = 3$, $p = 1$)	6272
SEBlock	(32,512)	(32,512)	32 \rightarrow 4 \rightarrow 32 ($r = 8$)	256
Conv1d	(32,512)	(64,256)	Conv (32 \rightarrow 64, $k = 3$, $s = 2$, $p = 1$)	6144
BatchNorm1d	(64,256)	(64,256)	γ, β	128
ReLU	(64,256)	(64,256)	-	0
ResidualBlock	(64,256)	(64,256)	dilation = 2 2xConv (64 \rightarrow 64, $k = 3$)	24,832
SEBlock	(64,256)	(64,256)	64 \rightarrow 8 \rightarrow 64 ($r = 8$)	1024
AvgPool1d	(64,256)	(64,1)	output_size = 1	0
Flatten	(64,1)	(64,)	-	0
Dropout	(64,)	(64,)	$p = 0.1$	0
Linear	(64,)	(2,)	64x2 + 2 (bias)	130
参数合计				39,074

在基于 XRF 光谱的土壤重金属元素污染风险判别任务中，对于模型是一个“是”或“否”的二分类问题，依据 GSS5、GSS17 和 GSS20 等 22 份国家土壤中某元素含量的标准值，将其标记为：“超标”或“未超标”，当土壤样品中的元素超过风险筛选值我们则标记此土壤样品为某元素重金属污染超标的风险，即“超标”标签。本研究继续以重金属元素 Zn 为研究对象，根据 GB15618-2018《中华人民共和国国家标准土壤环境质量 - 农用地土壤污染风险管控标准(试行)》确认风险筛选值，通过不同 pH 值下农用地、草地土壤中重金属元素污染风险筛选值，筛查的 pH 值标准设定为 $5.5 < \text{pH} \leq 6.5$ ，确认 Zn 的筛选值为 200 (mg/kg)，然后将所有土壤样品背景扣除预处理后的数据进行标记，标记结果：在 66 份土壤样品中(22 份国家标准土壤样品测量 3 次)，有 42 份土壤样品为“未超标”，24 份土壤样品为“超标”。将这些数据输入到 SpectraNet 中，建立土壤重金属元素量超标检测模型。

对于 pH 值的筛选标准设定，土壤样品 pH 值绝大多数集中在 5.4 至 6.5 之间，这也是大多数植物最

适宜生长的 pH 值范围。当土壤 pH 值超出这个范围时,可能会影响植物对营养元素的吸收和利用,从而影响植物的生长和产量,因此根据相关研究[19],不同 pH 值下农用耕地、草地等土壤中重金属元素污染的风险筛选值如表 2 所示。

Table 2. Risk screening values of heavy metals (unit: mg/kg)

表 2. 重金属元素的风险筛选值(单位: mg/kg)

元素	pH ≤ 5.5	5.5 ≤ pH ≤ 6.5	6.5 ≤ pH ≤ 7.5	pH > 7.5
Cr	150	150	200	250
Ni	60	70	100	190
Cu	50	50	100	100
Zn	200	200	250	300
As	40	40	30	25
Pb	70	90	120	170

为了全面评估重金属元素浓度超标判别模型的性能,本研究采用了 3 折交叉验证评估模型的泛化能力,将整个数据集随机分成 3 个大小相等的子集,每轮随机选择 2 个子集作为训练集来训练模型,将剩余的 1 个子集作为测试集用于性能评估。该方法减少了由于随机性对评估效果的影响,并且能更好的反映模型在未见过的数据上的表现。本研究 Accuracy (准确率)、AUC、PR-AUC (Precision-Recall)、F1 分数和 Recall (正样本召回率)作为主要性能指标。其计算公式如下:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (3.1)$$

$$\text{TPR} = \frac{TP}{TP + FN}, \text{FPR} = \frac{FP}{FP + TN} \quad (3.2)$$

其中, TP (True Positive): 实际为正类且预测为正类的样本数; TN (True Negative): 实际为负类且预测为负类的样本数; FP (False Positive): 实际为负类但预测为正类的样本数; FN (False Negative): 实际为正类但预测为负类的样本数。AUC 是以 FPR (假阳性率)为横轴, TPR (真阳性率)为纵轴绘制出的 ROC 曲线下面积,用于衡量模型对正负样本分类的能力,其取值范围为[0, 1],该指标值越接近 1 表明模型分类能力越强。PR-AUC 同样是一个曲线下面积,但 PR 曲线是以 Recall (召回率)为横轴, Precision (精确率)为纵轴绘制曲线下的面积。公式如下:

$$\text{Recall} = \text{TPR} = \frac{TP}{TP + FN} \quad (3.3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3.4)$$

F1 分数是 Precision 和 Recall 的调和平均值,它能够有效的反映模型在分类过程中的综合性能。计算公式如下:

$$\text{F1} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.5)$$

对于 Zn 元素的元素超标判别实验混淆矩阵如图 6 所示。从交叉验证的混淆矩阵可以看到, SpectraNet 在各折中都能显著减少漏报样本,且误报水平保持可控,体现出了相对稳定的决策边界。

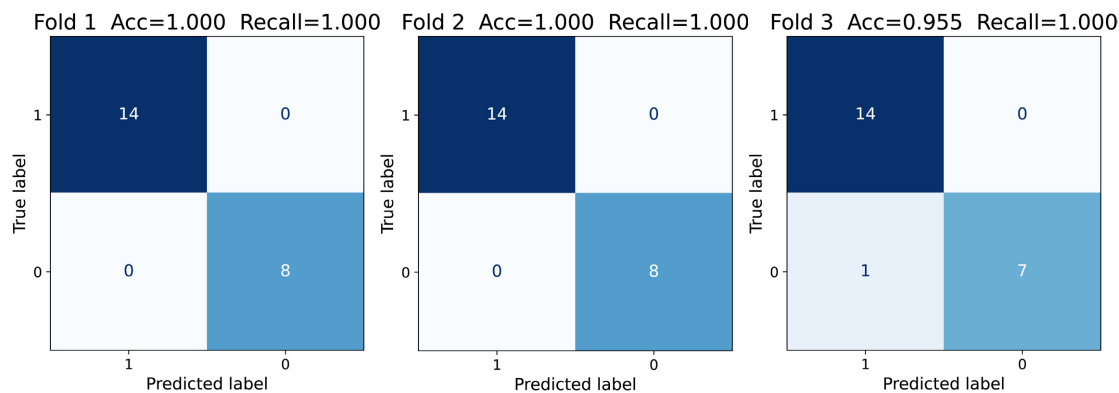


Figure 6. Confusion matrix of the accuracy of K-fold cross-verification of Zn element exceeding the standard (K = 3)
图 6. Zn 元素超标判别 K 折交叉验证准确率混淆矩阵(K = 3)

为了进一步系统评估本研究提出的 SpectraNet 模型的有效性，我们在完成背景扣除后，将同一批光谱数据输入到多种经典分类器和 SpectraNet 中，并采用 3 折交叉验证进行对比。通过表 3 所示，传统模型若无针对性的特征工程，在小样本场景下会受到一定影响，导致分类性能下降。相对而言，SpectraNet 不依赖复杂的特征工程或手工峰提取流程，通过残差卷积和 SE 注意力机制在训练过程中能够捕捉土壤光谱数据的局部特征与多尺度邻域关系，有效地提取了数据特征，从而在重金属元素超标判别任务中表现出色，证明了 SpectraNet 模型能够有效提取光谱特征。

Table 3. Cross-verification of Zn elements exceeding the standard (K = 3)
表 3. Zn 元素超标判别 K 折交叉验证(K = 3)

方法	Accuracy (%)	AUC (%)	PR-AUC (%)	F1 (%)	Recall (%)
SVM	95.24	90.47	92.38	91.86	87.5
随机森林	96.43	99.11	98.42	95.58	92.85
XGboost	86.9	92.26	89.53	87.88	83.33
Logistic 回归	93.75	87.5	91.65	95.15	87.5
SpectraNet (ours)	97.91	99.7	99.53	98.46	95.83

4. 结论

改进的 airPLS 方法通过优化权重函数减少了基线欠估计风险，有效提升了土壤 XRF 光谱的基线校正精度。该方法在校正后实现了最低的均方根误差和最优的相对误差，大幅增强了 XRF 谱的定量分析能力。结合高斯卷积滤波后，高频噪声被进一步平滑，有助于提高复杂土壤基质中 Zn 谱峰提取的准确性。SpectraNet 模型采用多尺度一维卷积网络对预处理光谱进行层次特征提取，实现了对谱线模式的鲁棒表示。该深度网络在训练和验证阶段表现出优秀的分类能力，能够从有限的光谱数据中提取有效信息，提高 Zn 超标判别的准确度。改进的基线校正与 SpectraNet 模型结合应用于土壤 Zn 污染识别，展现出良好的可行性与潜力。高质量的背景扣除光谱输入增强了对 Zn 信号的敏感性，使模型能够快速、稳定地识别样本中的超标风险。实验表明，该方法在 22 份国家标准土壤样本(共 66 组测量数据)上可稳定识别 Zn 污染趋势，为现场快速监测和污染风险评估提供了可靠手段。总之，本研究证明了改进的 airPLS + 高斯卷积和 SpectraNet 模型在土壤 Zn 污染检测中的有效性，为重金属污染的快速识别提供了可行的技术路径。

参考文献

- [1] 余涛, 蒋天宇, 刘旭, 等. 土壤重金属污染现状及检测分析技术研究进展[J]. 中国地质, 2021, 48(2): 460-476.
- [2] Zwolak, A., Sarzyńska, M., Szpyrka, E., *et al.* (2019) Sources of Soil Pollution by Heavy Metals and Their Accumulation in Vegetables: A Review. *Water, Air, & Soil Pollution*, **230**, Article No. 164. <https://doi.org/10.1007/s11270-019-4221-y>
- [3] 曲格平. 环境科学词典上[M]. 上海: 上海辞书出版社, 1994.
- [4] 赵睿新. 环境污染化学[M]. 北京: 化学工业出版社, 2004.
- [5] 丁文秀. 超声波活化风化煤对土壤中铜、锌形态和土壤酶活性影响研究[D]: [硕士学位论文]. 泰安: 山东农业大学, 2013.
- [6] 李晓平. 土壤酸碱度与肥效的关系[J]. 农家参谋, 2014(6): 10.
- [7] 戴万宏, 黄耀, 武丽, 等. 中国地带性土壤有机质含量与酸碱度的关系[J]. 土壤学报, 2009, 46(5): 851-860.
- [8] 吴施萌. 浅谈土壤重金属污染现状、原因、危害及修复[J]. 生态环境与保护前沿, 2020, 3(1): 7-8.
- [9] 杏艳, 田渭花, 刘锦华, 等. 单波长激发能量色散 X 射线荧光光谱法测定土壤和沉积物中 19 种元素[J]. 环境化学, 2022, 41(10): 3182-3195.
- [10] 豆卫全, 高明, 夏培民, 等. 粉末压片-能量色散 X 射线荧光光谱法分析硅铬合金中铬硅磷[J]. 冶金分析, 2019, 39(9): 54-58.
- [11] 陈颖, 刘峥莹, 肖春艳, 等. 基于麻雀搜索算法的土壤重金属 X 射线荧光光谱重叠峰解析[J]. 光谱学与光谱分析, 2021, 41(7): 2175-2180.
- [12] 吴和喜, 刘庆成, 杨波, 等. SNIP 法在天然放射性核素 γ 能谱分析中的应用[J]. 核技术, 2010, 33(7): 513-516.
- [13] Tan, H. and Brown, S.D. (2002) Wavelet Analysis Applied to Removing Non-Constant, Varying Spectroscopic Background in Multivariate Calibration. *Journal of Chemometrics*, **16**, 228-240. <https://doi.org/10.1002/cem.717>
- [14] Zhang, Z.M., Chen, S. and Liang, Y.Z. (2010) Baseline Correction Using Adaptive Iteratively Reweighted Penalized Least Squares. *The Analyst*, **135**, 1138-1146. <https://doi.org/10.1039/b922045c>
- [15] 何迎一. 基于支持向量机的土壤重金属污染评价研究[J]. 信息通信, 2020(12): 122-125.
- [16] Hu, B.F., Xue, J., Zhou, Y., *et al.* (2020) Modelling Bioaccumulation of Heavy Metals in Soil-Crop Ecosystems and Identifying Its Controlling Factors Using Machine Learning. *Environmental Pollution*, **262**, Article ID: 114308. <https://doi.org/10.1016/j.envpol.2020.114308>
- [17] Guo, Y., Jin, W., Wang, W., He, Y. and Qiu, S. (2023) Baseline Correction for Raman Spectra Using a Spectral Estimation-Based Asymmetrically Reweighted Penalized Least Squares Method. *Applied Optics*, **62**, 4766-4776. <https://doi.org/10.1364/ao.489478>
- [18] Birch, P., Mitra, B., Bangalore, N.M., Rehman, S., Young, R. and Chatwin, C. (2010) Approximate Bandpass and Frequency Response Models of the Difference of Gaussian Filter. *Optics Communications*, **283**, 4942-4948. <https://doi.org/10.1016/j.optcom.2010.07.047>
- [19] 杨婉琪. 基于 XRF 的土壤重金属元素智能分析算法研究及应用[D]: [博士学位论文]. 成都: 电子科技大学, 2024.