

云南墨江矿区周边山地农田土壤重金属的高光谱反演

田安红^{1,2}, 李智缘¹, 付承彪^{2*}, 朱大明¹

(1. 昆明理工大学国土资源工程学院, 昆明 650093; 2. 昆明理工大学信息工程与自动化学院, 昆明 650500)

摘要: 为探究矿区周边山地农田土壤重金属的污染状况, 实现在复合污染情境下山地农田土壤中多种重金属含量的高效反演。以云南省墨江县某金矿附近的农田区域为例, 获取 121 个土壤样品实验室高光谱数据和重金属砷 (As)、铬 (Cr)、铜 (Cu)、镍 (Ni) 的含量数据, 构建高精度的高光谱反演模型, 实现对不同重金属含量的定量反演。结果表明: 1) 内梅罗污染指数法显示研究区土壤处于重度污染状态, 潜在生态风险指数法显示该区域处于中等生态风险水平。2) 一阶微分、二阶微分、标准正态变量以及倒数的对数能有效增强光谱响应, 竞争性自适应重加权算法 (competitive adaptive reweighted sampling, CARS) 的高效波段筛选能力结合迭代保留信息变量 (iteratively retains informative variables, IRIV) 算法的变量精炼优势, 能够实现在土壤重金属反演中的敏感波段选择, 该方法在特征波段数量、计算运行时间和模型反演精度方面都比单独的 CARS 和 IRIV 方法更有效。3) 对比发现反向传播神经网络 (back-propagation neural network, BPNN) 在 As 反演中取得最佳反演精度, 支持向量机 (support vector machine, SVM) 为 Cr、Cu 和 Ni 的最优反演模型, As、Cr、Cu、Ni 最优反演模型的 R^2 分别为 0.90、0.93、0.67、0.94, 均方根误差 (root mean squared error of external validation, RMSE) 分别为 87.33、142.63、2.63、70.31 mg/kg, 相对分析误差 (relative percent difference, RPD) 分别为 3.25、3.84、1.74、4.17。4) 重金属的空间分布结果显示, 高值区域主要集中在研究区的上下部分, 而低值区域则主要分布在边缘, 整体呈现从中心向四周逐渐降低的趋势。该研究可为监测矿区附近农田土壤重金属复合污染状况提供参考依据。

关键词: 高光谱; 反演; 机器学习; 土壤重金属; IRIV 算法; 特征波段选择

doi: 10.11975/j.issn.1002-6819.202407103

中图分类号: X53

文献标志码: A

文章编号: 1002-6819(2025)-05-0191-10

田安红, 李智缘, 付承彪, 等. 云南墨江矿区周边山地农田土壤重金属的高光谱反演[J]. 农业工程学报, 2025, 41(5): 191-200. doi: 10.11975/j.issn.1002-6819.202407103 <http://www.tcsae.org>

TIAN Anhong, LI Zhiyuan, FU Chengbiao, et al. Hyperspectral inversion of soil heavy metals in mountain farmland around Mojiang Mining Areas in Yunnan, China[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2025, 41(5): 191-200. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202407103 <http://www.tcsae.org>

0 引言

随着工农业进程的加快, 土壤环境中的重金属污染问题日益严峻^[1]。云南素有“有色金属王国”之称, 其矿产资源丰富, 矿业历史悠久, 但因早期的粗放式开采, 带来了一系列严重的环境问题。云南省墨江县内有一座集采矿、选矿和冶炼于一体的大型金矿, 目前针对金矿开发带来的土壤污染问题少有报道。墨江金矿附近农田大多分布在地形复杂、高差变化显著的山地区域。尽管高光谱在矿区周边土壤重金属反演方面取得了许多进展, 但云南墨江地区复杂的地形导致耕地破碎, 又因土壤类型的多样性和空间分布不均匀性, 给大范围复合污染监测带来了准确度不高与泛化能力不足的问题。因此,

准确地反演地域特性明显的墨江金矿附近山地农田土壤重金属含量, 对于治理和修复受污染的农田土壤具有重要意义。

近年来, 高光谱传感器的进步提高了其对地面监测数据的获取能力, 因此高光谱技术已成为土壤重金属定量反演的重要手段^[2-3]。而高光谱数据降维方法的选择对提高模型的准确性和计算效率至关重要, 现有降维方法主要分为两类: 一类是基于特征提取的方法, 另一类是基于波段选择的技术。光谱特征提取是一种将数据转换到新的特征空间的方法, 通过提取变换后的主要特征信息来实现光谱数据降维^[4]。光谱波段选择是一种根据特定评价标准从所有特征中筛选出最优子集的过程, 其目的是选择出一个光谱波段子集, 该子集尽可能保留原始光谱数据的关键波段信息, 从而实现降维并有效去除冗余波段。波段选择问题实际上是一个组合优化问题^[5], 特定评价标准即为目标函数, 目标函数直接影响选择的波段数目质量。常用的波段选择方法有竞争性自适应重加权算法 (competitive adaptive reweighted sampling, CARS) 以及迭代保留信息变量 (iteratively retains informative variables, IRIV) 等^[6]。虽然基于波段选择方

收稿日期: 2024-07-11 修订日期: 2024-12-06

基金项目: 国家自然科学基金项目 (42361007, 42067029); 云南省科技厅项目 (202205AC160005); 云南省“兴滇英才支持计划”青年人才项目 (KKXX202303001)

作者简介: 田安红, 教授, 研究方向为遥感信息处理与人工智能。

Email: tah@kust.edu.cn

*通信作者: 付承彪, 副教授, 研究方向为遥感图像处理。

Email: fcu@kust.edu.cn

法在预测土壤重金属含量反演方面已获得很多学者关注,但不同重金属元素的敏感波段机理揭示方面的探讨仍旧很少。

土壤是一种复杂综合体,不同区域土壤的成土母质、条件、过程等具有地域性,其反射率光谱是由其组成成分与结构等多种内在性质共同决定的结果。已有研究报告土壤重金属与光谱活性物质,如有机质、铁氧化物、黏土矿物之间存在内在联系^[7-8],土壤重金属的敏感波段与定量反演机理的研究一直处于探索阶段。迭代保留信息变量法是 2014 年提出的一种新的特征变量选择算法,该算法基于二进制矩阵重排过滤器的思想提出,在迭代运算过程中保留对模型有益的强信息波段变量和弱信息波段变量,删除对建模无益的无信息波段变量和干扰信息波段变量。最后,选取后向消除后的变量作为特征变量^[9]。IRIV 在农产品与食品等高光谱领域能表现出能够有效筛选特征波段的良好潜力^[10]。但目前 IRIV 用于提取土壤特征波段的报道较少,且缺乏强弱特征波段变量的机理性解释,需进一步探讨 IRIV 对土壤重金属的特征波段提取效果。

在土壤重金属高光谱反演领域中,研究需要重点关注与反演土壤参数强相关的光谱特征波段变量。因 IRIV 算法能有效提取出土壤重金属的强波段信息和弱波段信息,揭示出土壤重金属反演的机理,并有效精炼波段变量,提升模型的预测精度和稳定性。但已有研究发现 IRIV 算法的计算运行时间较慢^[11],无法满足未来大规模实时监测需求。而 CARS 能快速筛选出土壤参数的敏感波段信息,但其筛选出的特征波段通常较多,仍旧存在信息冗余的问题,这在一定程度上影响模型的效率与准确性。可见,结合 CARS 的高效筛选能力和 IRIV 的变量精炼优势,有可能在波段选择过程中获得更加稳健且具有较高预测能力的特征。由于墨江金矿附近污染物种复杂且空间分布差异较大,传统的单一波段选择方法难以全面反映污染的多样性和复杂性。因此,探讨 CARS 与 IRIV 结合使用的潜力,尤其是在土壤复杂污染状况下,仍需深入探究光谱波段选择与反演模型的组合对预测精度的影响。

本研究针对墨江金矿附近山地农田土壤的重金属复合污染状况,以云南省墨江县内某金矿周边农田为例,通过内梅罗污染指数和潜在生态风险指数评估土壤重金属污染,并结合光谱预处理和 CARS-IRIV 波段选择算法建立 4 种重金属含量的最佳反演模型,最后绘制土壤重金属含量的空间分布图。本研究目的如下:1) 对金矿附近土壤重金属 As、Cr、Cu、Ni 污染状况进行评估分析,确定研究区污染状况。2) 研究土壤光谱反射率在 5 种不同数学变换下对特征波段选择算法 CARS-IRIV 的影响,并探究 CARS-IRIV 方法筛选不同重金属特征波段的效果。3) 通过对比 3 种 PLSR、SVM、BPNN 模型分析得到研究区 4 种土壤重金属的最佳反演模型。4) 探究研究区内土壤重金属(As、Cr、Cu、Ni)含量的空间分布特征。本研究结合传统污染评价方法与高光谱反演技术,定量

评估山地农田土壤重金属污染,并构建土壤重金属空间分布图,为土壤污染的评估与治理提供技术支持。

1 材料与方法

1.1 研究区概况

墨江县位于云南省普洱市东部,地理位置处于 $101^{\circ}07' \sim 102^{\circ}04'E$, $22^{\circ}51' \sim 23^{\circ}59'N$ 之间。该地区雨水充沛,年均降水量为 1388 mm,光照充足,干湿季分明,年均气温为 $17.8^{\circ}C$,属于亚热带半湿润山地季风气候。海拔范围 478.5~2278 m,属于低纬度高海拔地区,土壤多为酸性红壤。墨江县金矿附近农田大多分布在地形复杂、高差变化显著的山地区域,呈现出北高南低的地形特征。由于地形的作用,在研究区内形成了自北向南的较小河流,沿河两侧分布着农田,使得研究区自然地划分为上、中、下游 3 个区域。这种特殊的地理环境使得土壤重金属污染分布具有明显的地域特征,长期的矿产开采不可避免地导致周边农田土壤中重金属含量有不同程度的超标。

1.2 土壤样品采集

样本采集时间为 2022 年 2 月中旬,采样前清除样点周围杂草和碎石,为确保所采集样本的代表性,采用五点采样法对表层土壤(0~20 cm)进行采集,在每个采样点中心位置,以该点为中心向 4 个方向等距采集样品,共计采集 5 个土壤样本。将这些样本充分混合,以获得样方区的混合土壤样本。此外,利用 GPS 测量仪记录样方区的中心位置坐标,对采样点周围环境进行记录。每份样本在去除石块、植物残留和根系后,采集约 1 kg 土壤样本并密封保存。共采集 121 个样点。采样点分布图如图 1 所示。样品带回实验室进行自然风干处理后,经过研磨过筛,将土壤样品分为两部分,一部分进行土壤重金属含量测定,另一部分进行土壤光谱反射率测定。

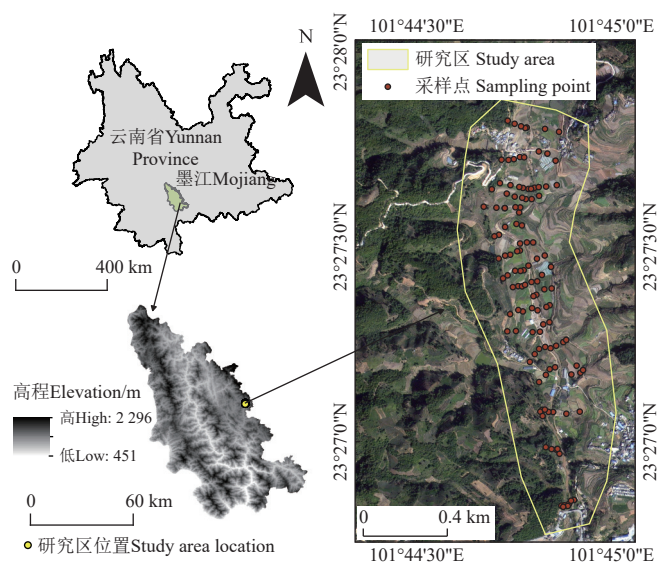


图 1 研究区位置及采样点分布

Fig.1 Location of the study area and distribution of sampling points

1.3 土壤光谱反射率测定与处理

土壤光谱反射率使用美国 ASD 公司的 FieldSpec3 便

携式地物光谱仪测定。为减少其他光源影响,测定在暗室内可控光照条件下进行。测量前,仪器需进行黑白校正。每个土壤样品被光谱仪重复测量10次,并将这10次测量的光谱平均值作为样本的代表性光谱。测定光谱波长范围为350~2500 nm,采样间隔为1 nm。在光谱采集过程中,由于仪器等原因350~399与2401~2500 nm区间内噪声较大,因此将这两部分区间内的光谱数据剔除。又因环境、仪器条件和人为操作等影响会产生噪声,这些噪声会对后续预测模型的精度造成影响,在使用高光谱构建反演模型之前应对光谱曲线进行预处理操作,本研究使用Savitzky-Golay (SG)方法对光谱曲线进行平滑去噪,处理得到原始光谱数据(original reflectance, OR)。同时,重金属在高光谱中的响应通常较微弱,有研究发现光谱变换能增强光谱特征、突出光谱隐含关键信息^[12]。但由于各种不同因素的影响,很难找到一种具体的变换方法同时适用于所有重金属并均表现出优异的性能,所以对于不同的重金属来说往往需要采用多种变换方法,探究得到每种重金属最适合的光谱变换方案,而常用的光谱变换方法主要为倒数的对数、导数以及散射校正等^[13]。因此,在经过SG处理的光谱曲线的基础上,再分别进行一阶微分(first-order derivative, FD)、二阶微分(second-order derivative, SD)、标准正态变换(standard normal variable, SNV)、倒数的对数(logarithm of reciprocal, LR)以及连续统去除(continuum removal, CR)变换,以探究不同变换方法对后续模型精度的影响。

1.4 重金属污染评价方法

结合单因子污染指数法、内梅罗污染指数法^[14]和潜在生态风险指数法,对研究区内的土壤重金属污染状况进行综合评价。单因子污染指数计算单一污染物的污染程度,内梅罗污染指数是基于单因子污染指数的综合计算,考虑了多个土壤重金属元素对土壤环境的综合污染情况。潜在生态风险指数法是由HAKANSON^[15]在1980年提出,用于评估重金属污染的生态风险,该指数不仅表示单一重金属元素的污染状况,同时还反映了所有参评的重金属元素的综合效应。计算时采用的标准值均使用云南省背景值^[16],在计算潜在生态风险指数时^[17],As、Cr、Cu、Ni所使用的毒性系数分别为10、2、5、5^[1]。

1.5 土壤重金属光谱特征选择方法

CARS算法参考达尔文进化论中的“适者生存”原则。通过随机抽取固定比率样本建立偏最小二乘(partial least squares, PLS)模型,计算回归系数的绝对值和对应该波长点的权重,结合指数衰减函数(exponentially decreasing function, EDP)和自适应重加权采样法(adaptive reweighted sampling, ARS)进行变量选择,并通过交叉验证计算均方根误差(root mean square error of cross-validation, RMSECV),最终通过N次蒙特卡洛采样选择RMSECV最小的波长子集作为最优变量组合^[18]。

IRIV算法通过随机组合变量并考虑变量间相互作用,基于二进制矩阵重排过滤器生成不同的随机组合,矩阵

每一行代表一个变量组合,基于该组合建立偏最小二乘模型(partial least squares, PLS),并通过交叉验证均方根误差(RMSECV)评估模型效果。然后计算每个波长变量包含与不包含时的RMSECV差异,并利用曼-惠特尼U检验计算P值,确定变量的重要性。剔除无信息变量与干扰变量,保留强信息与弱信息变量,通过多次迭代,直到无信息和干扰变量被移除,最终反向消除得到最优特征波长集合^[9]。

本文通过对光谱数据进行不同形式的光谱变换预处理,首先使用CARS进行初步特征筛选,再基于筛选出的波段进行二次筛选。CARS能够有效去除冗余特征,但由于蒙特卡洛(monte carlo, MC)采样过程的随机性,且无法完全消除与分析无关的波长,因此存在一定的局限性^[19]。而IRIV方法虽然需要构建大量模型,导致计算量较大,但其优势在于能够较好地保留变量之间的协同效应,从而提高模型的预测能力和稳定性。通过先使用CARS进行波段筛选,再应用IRIV进行选择,不仅可以增强变量选择的精确度,还能提高计算效率。因此,结合CARS和IRIV的波段选择协同方法具有重要的应用前景,值得进一步深入研究。

1.6 模型构建及精度评价

在建立土壤重金属浓度反演模型时,通常需要根据土壤性质和建模方法的原理选择合适的模型。目前,偏最小二乘回归(partial least squares regression, PLSR)广泛应用于土壤重金属反演建模^[20],通过分析因变量与自变量之间的统计关系,PLSR能够有效解决多重共线性问题。但是PLSR在处理非线性时精度受限,而机器学习模型在处理高维数据和非线性问题上具有显著优势,能够深入挖掘光谱特征与重金属浓度之间的潜在关系,从而提高预测精度。常用的机器学习模型包括支持向量机(support vector machine, SVM)和反向传播神经网络(back-propagation neural network, BPNN)。SVM在小样本和高维数据处理上表现出色^[21],而BPNN在非线性建模方面具有优势,但易陷入局部最优解^[22]。因此,选择合适的建模方法对于构建稳定且高精度的土壤重金属反演模型至关重要。研究采用PLSR、SVM与BPNN对重金属反演模型进行构建。并使用决定系数(R^2)、均方根误差(root mean squared error of external validation, RMSE)和相对分析误差(relative percent difference, RPD)对进行模型精度评价。 R^2 越接近1,说明模型拟合程度和稳定性越好。RMSE越低说明模型精度越高。其中,当 $RPD > 2.0$ 时,表明模型具有很好的预测能力;当在 $1.4 < RPD < 2.0$ 范围内时,模型预测能力一般;当 $RPD < 1.4$ 时,模型预测能力较差^[23]。并使用Kennard-Stone(KS)算法将121个土壤原始光谱按照3:1划分为建模集和验证集,该算法通过计算光谱曲线的欧式距离,确保训练集能代表整体样本,并按空间距离均匀分布。最终,91个样本作为训练集,30个样本作为验证集。

2 结果与分析

2.1 土壤重金属含量统计分析

表 1 为研究区内土壤重金属含量描述性统计结果，其中背景值采用云南省土壤元素背景值^[16]。由统计结果可知，土壤重金属 As、Cr、Ni 含量的平均值均大于云南省背景值，且分别为土壤背景值的 15.4、9.2、7.5 倍，而 Cu 的含量均值小于云南省背景值，表明该地区土壤重金属 As、Cr 和 Ni 污染严重，而重金属 Cu 在该地区土壤中积累含量较低。变异系数是重金属元素含量标准

差与均值的比值，可以用来衡量采样点之间重金属元素含量的离散程度与相对变异程度，表 1 结果表明 4 种重金属变异系数由大到小依次为 As、Cr、Ni、Cu，变化范围为 24.99%~123.57%。而变异系数小于 16% 为弱变异，16%~36% 属于中等变异，大于 36% 属于强变异^[24]。可见，研究区内重金属 As、Cr、Ni 的空间变异很强，重金属空间分布不均。而 Cu 的变异系数为 24.99% 属于中等变异，进一步说明 Cu 相对于其余 3 种重金属元素来说其空间分布差异不大。

表 1 研究区土壤重金属含量描述性统计

Table 1 Descriptive statistics of soil heavy metal content in the study area

重金属 Heavy metal	最大值 Maximum/(mg·kg ⁻¹)	最小值 Minimum/(mg·kg ⁻¹)	平均值 Average/(mg·kg ⁻¹)	标准差 Standard deviation/(mg·kg ⁻¹)	变异系数 Variation coefficient/%	背景值 Background value/(mg·kg ⁻¹)
As	1 780	5.79	283.33	350.12	123.57	18.40
Cr	2 543	62.80	600.49	581.90	96.90	65.20
Cu	41	11.6	22.18	5.54	24.99	46.30
Ni	1 051	20.1	318.33	305.90	96.10	42.50

2.2 土壤重金属污染评价

表 2 为土壤重金属污染评价结果，单因子污染指数的平均值从大到小依次为 As、Cr、Ni、Cu。研究结果表明，Cr 的污染率最高，显示大部分样点受到 Cr 元素的污染；而 As 和 Ni 的污染较为普遍，二者污染率接近 80%。Cu 则没有检测到污染。进一步分析发现，大部分土壤样本中 As、Cr、Cu 和 Ni 处于不同程度的污染状态，其中 As、Cr、Ni 在重度污染样本中的比例较高，而 Cu 未表现出污染迹象。单因子污染指数法能够识别造成污染的关键重金属元素，并评估土壤重金属的污染程度。然而，土壤重金属污染通常是多种元素共同作用的结果。内梅罗污染指数作为

一种多因子评价方法，可以全面体现多种污染物的综合污染水平，同时突显对污染贡献较大的重金属元素，从而更科学地反映污染特征。根据内梅罗污染指数评价分级标准^[25]，研究区大部分土壤样本处于重度污染状态，少部分样点处于警戒线或轻度污染状态，整体污染水平较为严重，污染率达到 71.07%。同时，表 2 显示潜在生态风险因子由大到小为 As、Ni、Cr、Cu，按照潜在生态风险指数分级标准^[26]，As 的单项生态风险因子等级为较强，生态风险较大，而 Cr、Cu、Ni 均为轻微级别，说明生态风险较小。通过综合计算研究区内的潜在生态风险指数为 212.25，表明该区域处于中等生态风险水平。

表 2 土壤重金属污染评价结果

Table 2 Evaluation results of soil heavy metal pollution

重金属 Heavy metal	单因子污染指数 single factor pollution index			内梅罗污染指数 Nemerow pollution index			潜在生态风险因子 Potential ecological risk factors	潜在生态风险指数 Potential ecological risk index
	范围 Range	平均值 Average	污染率/% Pollution rate/%	范围 Range	平均值 Average	污染率/% Pollution rate/%		
As	0.31~96.74	15.40	79.33	0.86~73.58	12.57	99	153.98	212.25
Cr	0.96~39.00	9.21	99.17				18.42	
Cu	0.25~0.88	0.48	0				2.40	
Ni	0.47~24.73	7.49	78.51				37.45	

2.3 光谱特征波段选择

有效提取出不同重金属元素的敏感波段，是进行高精度定量反演模型构建的关键。本研究采用 CARS-IRIV 方法提取重金属的敏感波段，以剔除冗余的无效波段信息，即先采用 CARS 得到初选的波段信息，再采用 IRIV 得到精选的波段信息。

2.3.1 基于 CARS 的初选波段

采用 CARS 算法进行特征筛选，主要的参数设置为：交叉验证数为 10，迭代次数为 100，蒙特卡罗采样时建模集与校正集的比例为 0.8。因为不同重金属元素的特征波段提取过程类似，本研究以重金属 Ni 元素在 SNV 变换下的光谱数据 CARS 采样过程为例来描述，如图 2 所示。图 2a 表明筛选出的特征波段变量数量出现快速下降趋势，并且速度由快到慢，之后逐渐减缓并趋于稳定，且 RMSECV 的变化趋势表现为先降低后升高。

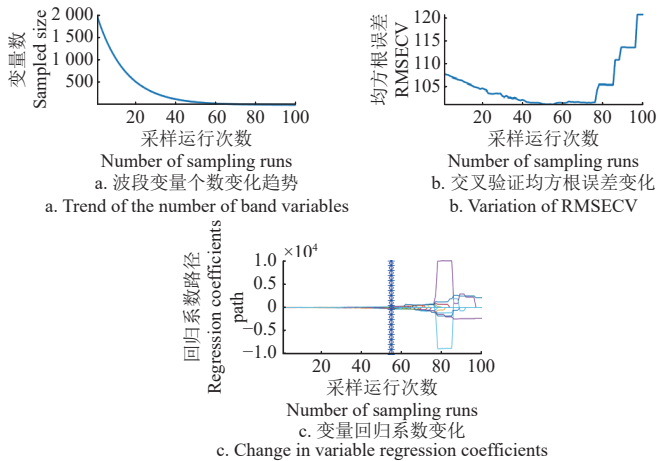


图 2 SNV 变换下重金属 Ni 的光谱数据 CARS 采样过程
Fig.2 CARS sampling process of spectral data of heavy metal Ni elements based on SNV (standard normal variable) transformation

当 RMSECV 值达到最低点时，采样运行次数为 55，最终 CARS 筛选出 46 个特征波段。

2.3.2 基于 CARS-IRIV 的精选波段

图 3 为基于 CARS-IRIV 筛选出的特征波段统计，从图 3 可看出在经过不同的预处理之后（OR、FD、SD、SNV、LR、CR），该方法挑选出来的最终特征波段数目差异较大，例如，重金属 Cr 在不同光谱变换中的特征波段数量差异较大，在 CR 光谱变换下的特征波段数量最多有 57 个特征波段，但原始光谱（OR）仅仅筛选出 7 个特征波段，特征波段相差 50 个，这种差异可能是因不同的光谱变换能突出的光谱特征信息不同造成的。

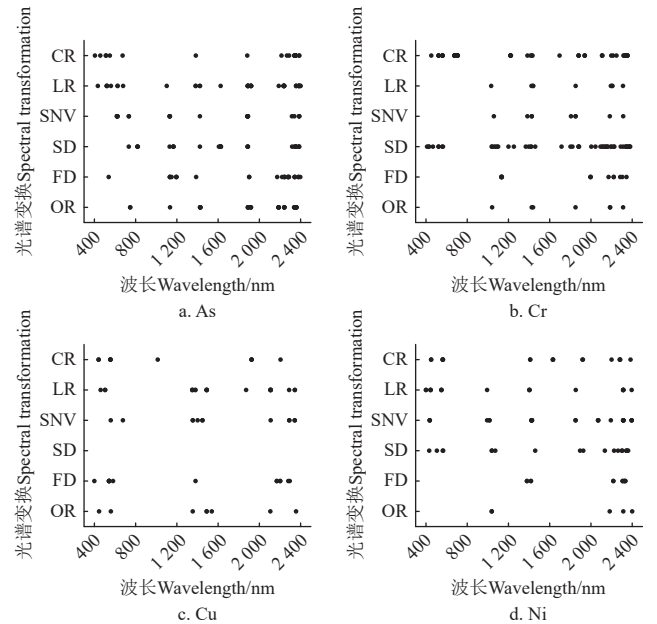
并且 CARS-IRIV 方法能够大幅度剔除与重金属不敏感的波段变量信息，以 CR 变换为例，与原始光谱波段相比，重金属 As、Cr、Cu、Ni 经过 CARS-IRIV 方法后提取的特征波段分别减少了 98.9%、97.15%、99.45%、99.1%。

图 3 结果也显示出 4 种重金属在不同数学变换下筛选出的特征波段具有较大的重叠，尤其是在 400~550、900~1 050、1 350~1 500、1 800~1900 和 2 150~2 400 nm 波段区间。这是因为土壤的光谱吸收特征通常与某些特定的土壤属性密切相关，重金属元素通常与黏土矿物、铁氧化物和有机质等土壤组分相互作用，这些组分具有显著的光谱特性^[27]，从而能够间接反演土壤重金属元素。

2.4 CARS-IRIV 方法的重金属含量反演模型精度评估

本研究使用 CARS-IRIV 方法筛选出的特征波段作为 PLSR、SVM 和 BPNN 模型的输入变量，构建了反演模型，其精度如表 3 所示。从 4 种重金属的预测精度来看，基于 5 种变换（FD、SD、SNV、LR 和 CR）处理后的大部分模型预测精度相对原始光谱 OR 有提升。其中，重金属 As 的最优反演模型为基于 LR 光谱变换的 BPNN 模型，其在验证集的 R^2 、RMSE 和 RPD 值分别为 0.90、87.33 mg/kg、3.25，该模型能很好地预测土壤

重金属 As 含量。对于重金属 Cr，其最优反演模型为基于 SD 光谱变换的 SVM 模型，该模型在验证集的 R^2 、RMSE 和 RPD 值分别为 0.93、142.63 mg/kg、3.84，模型的 RPD 值大于 2.0 说明能很好地预测土壤重金属 Cr 含量。对于重金属 Ni，其最优反演模型为基于 SNV 光谱变换的 SVM 模型，该模型在验证集的 R^2 、RMSE 和 RPD 值分别为 0.94、70.31 mg/kg、4.17，也能很好地预测土壤重金属 Ni 含量。



注：OR 为原始光谱；LR 为倒数的对数；SNV 为标准正态变换；SD 为二阶微分；FD 为一阶微分；CR 为连续统去除。
Note: OR: original reflectance; FD: first-order derivative; SD: second-order derivative; SNV: standard normal variable; LR: logarithm of reciprocal; CR: continuum removal.

图 3 CARS-IRIV 筛选的特征波段统计
Fig.3 Characteristic band statistics of CARS-IRIV screening

表 3 CARS-IRIV 选择方法在不同反演模型下重金属含量预测精度的比较
Table 3 Comparison of prediction accuracy for heavy metal content using different inversion models with CARS-IRIV selection method

重金属 Heavy metal	光谱变换 Spectral transformation	PLSR			SVM			BPNN		
		R^2	RMSE/ (mg·kg ⁻¹)	RPD	R^2	RMSE/ (mg·kg ⁻¹)	RPD	R^2	RMSE/ (mg·kg ⁻¹)	RPD
As	OR	0.84	125.84	2.49	0.81	121.51	2.33	0.85	109.81	2.58
	FD	0.89	104.67	3.07	0.87	101.35	2.81	0.75	141.73	2.01
	SD	0.87	114.69	2.79	0.88	97.17	2.92	0.76	140.39	2.02
	LR	0.89	104.91	3.04	0.82	121.35	2.34	0.90	87.33	3.25
	SNV	0.88	107.06	2.86	0.89	94.47	3.01	0.84	112.62	2.52
	CR	0.80	147.60	2.26	0.86	107.41	2.65	0.75	142.25	1.99
Cr	OR	0.87	206.17	2.81	0.89	180.36	3.03	0.90	171.44	3.19
	FD	0.91	178.56	3.29	0.92	156.96	3.49	0.89	181.53	3.01
	SD	0.91	175.65	3.39	0.93	142.63	3.84	0.81	234.89	2.33
	LR	0.88	189.38	2.98	0.89	176.06	3.11	0.79	247.85	2.21
	SNV	0.87	209.12	2.79	0.90	172.12	3.18	0.90	171.45	3.19
	CR	0.88	190.31	2.94	0.90	169.09	3.24	0.84	214.23	2.56
Cu	OR	0.64	2.80	1.67	0.59	2.94	1.56	0.52	3.18	1.44
	FD	0.65	2.77	1.71	0.67	2.63	1.74	0.66	2.69	1.70
	SD	—	—	—	—	—	—	—	—	—
	LR	0.57	3.53	1.53	0.55	3.07	1.49	0.57	2.99	1.53
	SNV	0.66	2.68	1.73	0.63	2.78	1.64	0.51	3.19	1.43
	CR	0.51	3.27	1.42	0.54	3.08	1.48	0.54	3.08	1.48
Ni	OR	0.74	158.09	1.99	0.87	102.64	2.85	0.82	120.98	2.42
	FD	0.85	116.94	2.62	0.90	91.71	3.19	0.86	107.22	2.73
	SD	0.92	91.25	3.54	0.93	72.59	4.03	0.92	81.41	3.59
	LR	0.92	84.37	3.54	0.90	92.37	3.17	0.91	87.43	3.35
	SNV	0.87	106.81	2.82	0.94	70.31	4.17	0.92	79.27	3.69
	CR	0.89	93.57	3.15	0.93	73.34	3.99	0.75	144.63	2.03

然而,相较于其他 3 种土壤重金属的反演结果,重金属 Cu 在 3 种模型中的预测精度均较差,其反演模型的 R^2 在 0.50~0.67 之间,而 RPD 在 1.5 左右,但 As、Cr、Ni 反演模型的 R^2 分别在 0.75~0.90、0.79~0.93、0.74~0.94 之间,模型预测能力大部分表现较好 (RPD>2)。这可能是土壤中 Cu 含量相对其他 3 种重金属较低,导致其光谱响应微弱,能捕捉的特征波段信息较少,进而对反演模型造成了较大影响。

通过综合分析可知,对于不同的重金属元素,As 和 Cr 分别在 LR 和 SD 构建的 BPNN 和 SVM 模型中表现最佳,Ni 则在 SNV 构建的 SVM 模型中预测最为准确,而 Cu 的最佳预测模型为基于 FD 构建的 SVM 模型。4 种最佳反演模型中的光谱变换并不一样,说明只有选取合适的光谱变换才能提高模型对土壤重金属含量的预测能力。

图 4 为 4 种重金属含量最佳反演模型的散点图,从图中可以看出,4 幅图中的样点大多数均集中在 1:1 线附近。其中反演 Ni、Cr 和 As 含量模型的样点最为集中,而 Cu 含量反演模型的样点相对前 3 种最为松散,说明其模型预测精度相对较差。

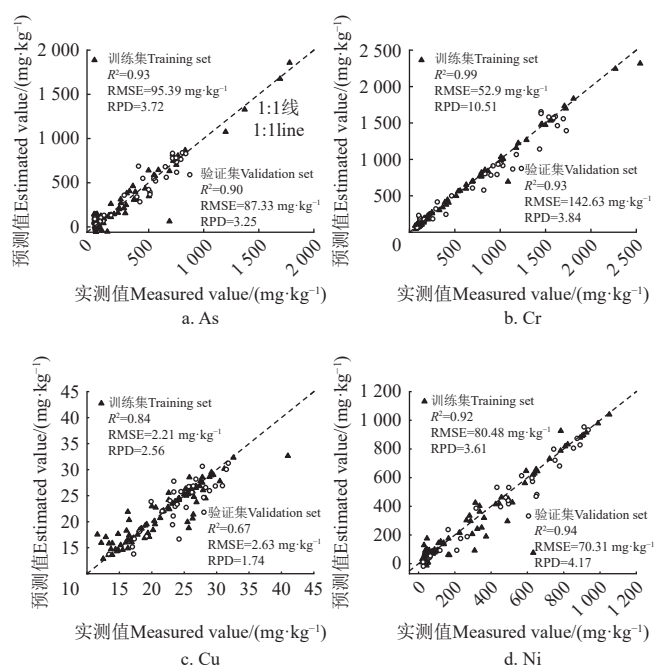


图 4 重金属含量最佳反演模型散点图

Fig.4 Scatter plot of the best inversion model for heavy metal content

2.5 最佳反演策略下不同重金属的敏感特征波段

因 IRIV 算法在迭代计算中能将光谱特征波段划分为强信息变量、弱信息变量、无信息变量和干扰信息变量,该算法能考虑到每个变量与其它变量之间的相互作用关系,通过不断迭代剔除无信息特征波段与干扰信息特征波段,使得变量数目减少,进行反向消除后可更好地提取特征变量,最后得到的这些变量应全部为“强信息变量波段”和“弱信息变量波段”。由于 IRIV 需要构建大量子模型导致计算时间长,研究先通过 CARS 方法对光

谱特征波段进行初步筛选,在此基础上进一步采用 IRIV 方法提取不同重金属的特征波段。

不同重金属的最佳反演模型在 CARS-IRIV 方法下筛选的特征波段位置如图 5 所示。重金属 As 的“强信息变量波段”有 5 个,分别为 434、1101、1103、1382 和 1383 nm; Cr 有 4 个,分别为 2192、2223、2289 和 2380 nm; Cu 有 5 个,分别为 1381、2169、2170、2202 和 2203 nm; Ni 有 4 个,分别为 438、2069、2071 和 2395 nm。“弱信息变量波段”在重金属 As、Cr、Cu 和 Ni 中分别为 39 个、46 个、9 个和 23 个波段。

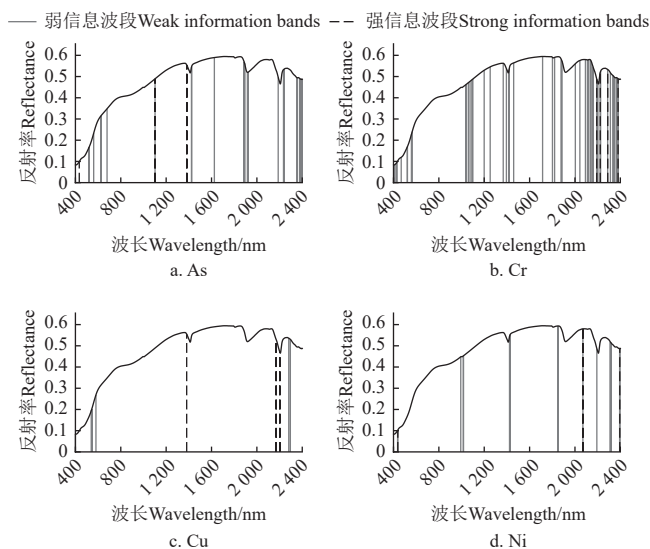


图 5 CARS-IRIV 方法对 4 种重金属最佳反演策略的光谱变换强信息波段筛选结果

Fig.5 Results of spectrally transformed strong information band screening of the CARS-IRIV method for the best inversion strategy for four heavy metals

2.6 不同光谱特征筛选方法下的土壤重金属反演精度

高光谱数据降维方法的选择对提高模型的准确性和计算效率至关重要。为了进一步验证 CARS-IRIV 方法的有效性,需对比分析单独使用 CARS 和 IRIV 方法的建模结果,并与最优反演模型进行比较,研究重点分析特征波段数量、计算运行时间和模型精度指标,其结果如表 4 所示。从表 4 的计算运行时间来看,基于 CARS-IRIV 方法不仅能提升土壤重金属反演精度,而且能大幅缩短计算运行时间。单独使用 IRIV 方法的计算时间较长,约为 3 h;而 CARS 方法运行效率极高,仅需 20 s 左右。然而,先通过 CARS 方法对高光谱波段进行初步筛选,再采用 IRIV 方法对初步筛选后的波段进行精选优化,计算运行时间可缩短至约 1 min,并仍保持较高的反演精度。从特征波段提取数量来看,在 4 种重金属 As、Cr、Cu 和 Ni 的最佳反演模型中,基于 CARS-IRIV 方法选取的特征波段分别为 44、50、14 和 27 个,占全波段的 2.19%、2.50%、0.70% 和 1.35%;而基于 CARS 方法提取出的特征波段分别为 70、87、15 和 46 个,占全波段的 3.49%、4.35%、0.75% 和 2.30%。从反演模型精度指标来看,对于 4 种重金属,基于 CARS-IRIV 方法的反演

模型精度均优于单独使用 CARS 或 IRIV 方法，建模精度依次为：CARS-IRIV > CARS > IRIV。

表 4 不同特征筛选方法对重金属反演精度的比较
Table 4 Comparison of inversion accuracy for heavy metals using different feature selection methods

重金属 Heavy metal	光谱变换 Spectral transformation	反演算法 Inversion algorithm	波段选择方法 Band selection method	特征波段数（占全波段 %） Number of selected bands (% of the whole band)	运行时间 Run time/s	R ²	RMSE/ (mg·kg ⁻¹)	RPD
As	LR	BPNN	CARS	70 (3.49%)	23.88	0.87	103.27	2.75
			IRIV	35 (1.75%)	10 662.14	0.75	140.56	2.02
			CARS-IRIV	44 (2.19%)	59.36	0.90	87.33	3.25
Cr	SD	SVM	CARS	87 (4.35%)	24.40	0.88	185.69	2.95
			IRIV	14 (0.70%)	11 319.99	0.72	291.71	1.88
			CARS-IRIV	50 (2.50%)	90.89	0.93	142.63	3.84
Cu	FD	SVM	CARS	15 (0.75%)	18.66	0.65	2.71	1.68
			IRIV	46 (2.30%)	7 089.66	0.61	2.87	1.59
			CARS-IRIV	14 (0.70%)	19.15	0.67	2.63	1.74
Ni	SNV	SVM	CARS	46 (2.30%)	23.78	0.85	111.78	2.62
			IRIV	9 (0.45%)	20 970.12	0.79	133.63	2.19
			CARS-IRIV	27 (1.35%)	30.53	0.94	70.31	4.17

2.7 土壤重金属空间分布

采用普通克里金插值方法，基于最佳估测值绘制了研究区 4 种土壤重金属含量的空间分布图（如图 6 所示），同时图 7 展示了基于实测值的空间分布情况。

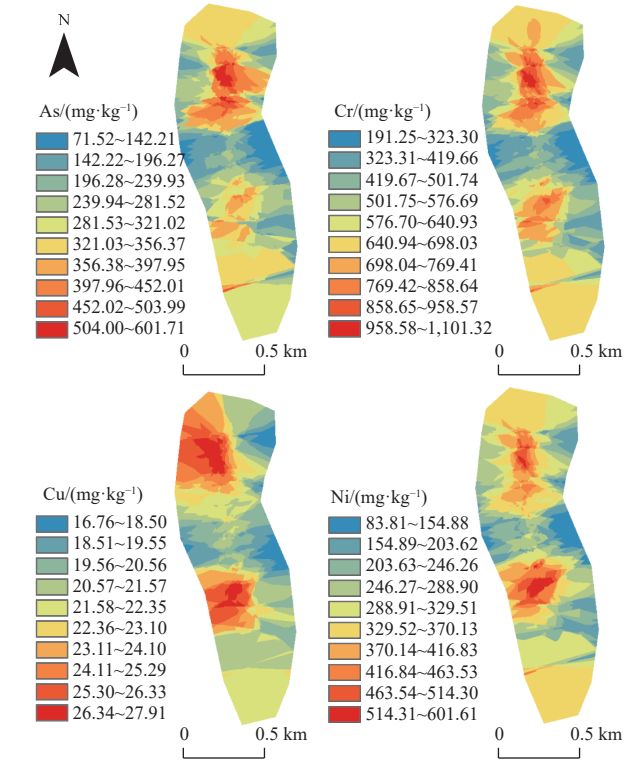


图 6 基于估测值的研究区土壤重金属空间分布图
Fig.6 Spatial distribution map of estimated values of soil heavy metals in the study area

图 6 和图 7 中的重金属空间分布图高度相似，表明插值结果与实测值的空间分布一致，说明所采用的 4 种重金属反演模型拟合效果较好，能够准确预测土壤重金属含量，从而有效揭示研究区土壤重金属的空间变化特征。并且 4 种土壤重金属浓度分布情况具有一定的相似性，但空间分布差异性较大。重金属高浓度区域主要集中在分布在上游与下游两部分，且下游重金属高浓度部分均向研究区西部延伸。与中部重金属浓度相比，上游与下游 4 种重金属浓度较高，并且整体呈现出一种由内向

外逐步降低的趋势。对于重金属 As、Cr 和 Ni 上游的高浓度区域均集中在中部，而重金属 Cu 上游高浓度区域则向西延伸。

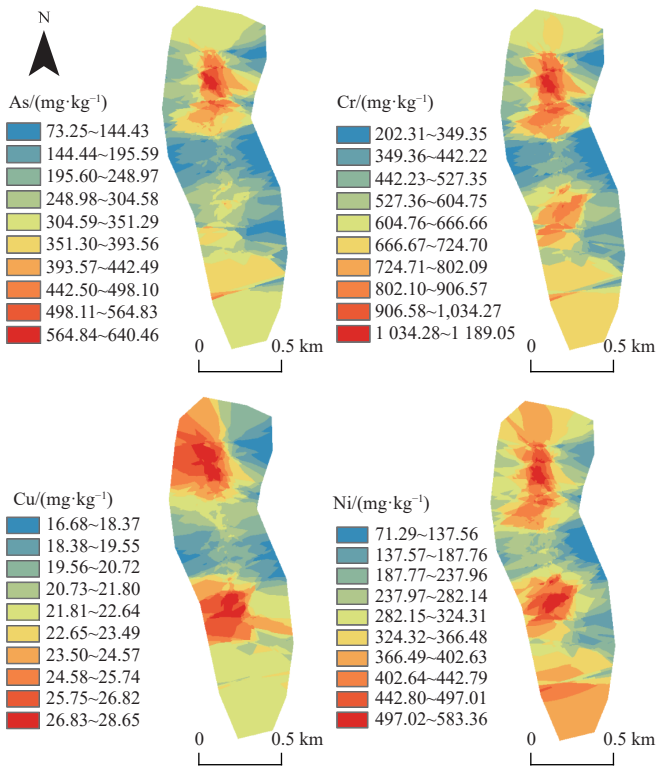


图 7 基于实测值研究区土壤重金属空间分布图
Fig.7 Spatial distribution map of measured values of soil heavy metals in the study area

3 讨 论

3.1 不同重金属的敏感特征波段分析

已有研究发现重金属 Cr、Cu、As 与土壤中光谱活性物质具有明显的相关性^[27]，图 5 显示尽管不同重金属元素的强波段响应的波谱位置有所差异，但它们大多数集中在可见光和近红外波段附近，且集中位于 1 200~1 500 和 2 000~2 400 nm，这与黏土矿物的吸收峰一致^[28]。而重金属 As、Cu、Ni 还有部分强波段出现在 400~600 nm，这主要是与有机质、铁氧化物的吸收峰一致^[29]。同时不

同重金属元素的“强信息变量波段”位置存在差异,主要是因为土壤组分信息对不同重金属元素的吸附作用不同导致。不同重金属在土壤中的吸附特性因组分的差异化作用而呈现显著不同,每种土壤组分对重金属的吸附强度具有独特的表现,如络合作用,离子交换机制等^[30]。在特定的吸附过程中,不同组分可能表现出主导作用,这种差异反映了土壤组分与重金属间复杂的物理化学相互作用机制。

3.2 不同波段筛选方法对重金属反演精度的影响

由表 4 可知, CARS-IRIV 方法提取出的特征波段数量相对较少,并且相较于单独的 CARS 方法其能进一步消除重金属的冗余波段信息,从而得到更高精度的反演模型。在特征波段选择中, CARS 方法能够在初步筛选阶段识别出土壤元素的敏感波段,特别是在处理高维数据时,能够聚焦于对模型性能影响较大的特征波段。但 IRIV 方法通过递归特征消除,在每一步计算中去除对模型贡献最小的特征,能够更好地保留变量间的协同效应^[31]。当这两种方法“粗-细”结合时,可以互补各自的不足,模型能够从不同的角度和方法中筛选出更为丰富、有效的特征波段,避免了单独使用任一方法时可能存在的信息丢失或冗余问题,相比单独使用 CARS 方法或 IRIV 方法构建的模型,该组合 CARS-IRIV 方法能够获得更高的精度,使其更好、更高效地适应数据。CARS-IRIV 方法充分发挥了 CARS 的高效性与 IRIV 的精度优势,为波段选择提供了一种高效可靠的解决方案,未来可望应用于嵌入式设备中,为实时、快速、高精度的土壤重金属监测提供技术支持。

3.3 研究区土壤污染源分析

由图 6、图 7 的重金属空间分布可知,4 种重金属高浓度区域主要集中分布在研究区上游与下游两部分,并且研究区东西两侧重金属浓度低,中间高。造成该现象的主要原因是在研究区域的上方存在一个大型的金矿,该金矿开发历史久远,金矿的开采以露天开采为主。该金矿是集采矿、选矿和冶炼为一体的大型企业,矿区内大规模的开采活动,导致重金属元素从地下深处被逐步释放出来,并逐渐迁移到表生环境,在外界风力、雨水、径流等因素的推动下,重金属元素被扩散沉降,长期积累对周围的生态环境造成了一定的污染^[32]。

土壤农产品的质量安全是全世界的研究热点,墨江金矿早期的粗放式开采,导致金矿的废水废气废渣直接排放,污染物被排入研究区域的河流内,当地居民长期采用河流的水源进行灌溉当地农作物,长期积累导致周边农田土壤中重金属含量有不同程度的超标。于海洋等^[33]研究发现污染源与地形特征对污染物的扩散和汇聚具有一定的影响。研究区上游金矿是重金属的来源之一,使得上游地区受到较高的重金属污染。而下游区域受到上游重金属污染物迁移的影响,也有一定的高浓度分布。然而,由于迁移过程中可能发生稀释或沉淀等现象,导致下游重金属浓度相对较低。

墨江县地处哀牢山脉中段,墨江金矿为云南省较大

型的著名金矿之一,该金矿中伴生有银、镍、铬、钴等稀有重金属。在云南省墨江和元江交界处,分布有一个大型的镍矿床,因此矿区附近农田土壤中重金属镍的污染也较为严重。

本研究在高光谱反演方面取得一定的效果,但由于受到高光谱遥感数据来源和天气环境等因素的限制,未能结合卫星影像进行重金属含量的空间分布反演。此外,研究主要依赖室内光谱数据,未能结合更多的野外实测光谱数据以提升模型的精确度。后续研究应在此基础上结合高质量的遥感影像与野外实测光谱数据,进一步提高模型的精度,并探索高光谱反演技术在复杂地形条件下的应用潜力。

4 结 论

本研究以云南墨江金矿附近山地农田为研究区域,针对重金属复合污染问题,构建出预测多种土壤重金属的反演模型,为当地未来土壤修复提供一定的科学参考价值。得出以下结论:

1) 研究区除了铜(Cu)未表现出污染外,As、Cr、Ni 3 种重金属均处于重度污染状态。研究区整体处于重度污染状态,污染率达到了 71.07%,面临中等生态风险。

2) 基于 CARS-IRIV 方法不仅能有效去除土壤高光谱的冗余波段信息,还能筛选出不同重金属元素的强波段变量信息和弱波段变量信息,As 和 Cu 有 5 个强波段,Cr 和 Ni 有 4 个强波段,进一步解释土壤重金属的光谱响应机理,该方法在复杂污染环境中展现了较强的应用潜力。

3) 通过对比分析,BPNN 和 SVM 方法均能有效反演 4 种重金属的含量。其中,As 的最佳反演模型为 BPNN, R^2 达到 0.90,均方根误差为 87.33 mg/kg; Cr、Cu 和 Ni 的最佳反演模型均为 SVM, R^2 分别为 0.93、0.67 和 0.94,均方根误差分别为 142.63、2.63、70.31 mg/kg。散点图显示,样本点紧密集中在 1:1 线附近,表明这两种方法在重金属浓度反演中具有很好效果。

4) 研究区内 4 种重金属的空间分布特征显示出一定的相似性。由于污染源和地势等因素的影响,研究区上游和下游存在重金属富集区,且上游的重金属浓度普遍高于下游。中游地区的重金属浓度较低,整体分布呈现由内向外逐步降低的趋势,表明重金属的空间分布存在明显的区域差异性。

[参 考 文 献]

- [1] 杨清可,王磊,李平星,等.长江干流沿江地区土壤重金属的分布、来源及潜在生态风险[J].农业工程学报,2024,40(9): 190-198.
- YANG Qingke, WANG Lei, LI Pingxing, et al. Distribution, sources and potential ecological risks of soil heavy metals in the main stream of the Yangtze River[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(9): 190-198. (in Chinese with English abstract)

- [2] TIAN A, ZHAO J, FU C, et al. Estimation of SO_4^{2-} ion in saline soil using VIS-NIR spectroscopy under different human activity stress[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2022, 282:121647.
- [3] ZOU Z, WANG Q, WU Q, et al. Inversion of heavy metal content in soil using hyperspectral characteristic bands-based machine learning method[J]. *Journal of Environmental Management*, 2024, 355: 120503.
- [4] 张兵. 高光谱图像处理与信息提取前沿[J]. *遥感学报*, 2016, 20(5): 1062-1090.
ZHANG Bing. Advancement of hyperspectral image processing and information extraction[J]. *Journal of Remote Sensing*, 2016, 20(5): 1062-1090. (in Chinese with English abstract)
- [5] TAN K, ZHU L, WANG X. A hyperspectral feature selection method for soil organic matter estimation based on an improved weighted marine predators algorithm[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 63: 5500411.
- [6] WANG X, ZHAO C, LI Z, et al. Modeling risk assessment of soil heavy metal pollution using partial least squares and fuzzy logic: A case study of a gully type coal-based solid waste dumpsite[J]. *Environmental Pollution*, 2024, 352: 124147.
- [7] FU C, FENG X, TIAN A. Whale optimization algorithm coupled with machine learning models for quantitative prediction of soil Ni content[J]. *Microchemical Journal*. 2025, 209: 112709.
- [8] SUN W C, ZHANG X, SUN X J, et al. Predicting nickel concentration in soil using reflectance spectroscopy associated with organic matter and clay minerals[J]. *Geoderma*, 2018, 327: 25-35.
- [9] YUN Y H, WANG W T, TAN M L, et al. A strategy that iteratively retains informative variables for selecting optimal variable subset in multivariate calibration[J]. *Analytica Chimica Acta*, 2014, 807: 36-43.
- [10] SUN Y, ZHANG H, LIU G, et al. Quantitative detection of myoglobin content in tan mutton during cold storage by near-infrared hyperspectral imaging[J]. *Food Analytical Methods*, 2022, 15(8): 2132-2144.
- [11] 郭云开, 张思爱, 王建军, 等. 特征变量选择结合 SVM 的耕地土壤 Hg 含量高光谱反演[J]. *测绘工程*, 2022, 31(1): 17-23.
GUO Yunkai, ZHANG Siai, WANG Jianjun, et al. Feature variable selection combined with SVM for hyperspectral inversion of cultivated soil Hg content[J]. *Engineering of Surveying and Mapping*, 2022, 31(1): 17-23.
- [12] ZHANG S, SHEN Q, NIE C, et al. Hyperspectral inversion of heavy metal content in reclaimed soil from a mining wasteland based on different spectral transformation and modeling methods[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2019, 211: 393-400.
- [13] 毛继华, 赵恒谦, 金倩, 等. 河北铅锌尾矿库区土壤重金属含量高光谱反演方法对比[J]. *农业工程学报*, 2023, 39(22): 144-156.
MAO Jihua, ZHAO Hengqian, JIN Qian, et al. Comparative study on the hyperspectral inversion methods for soil heavy metal contents in Hebei lead- zinc tailings reservoir areas[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2023, 39(22): 144-156. (in Chinese with English abstract)
- [14] ZHAO S, ZHANG H, ZHU Y, et al. Residual heavy metals and antibiotic pollution in abandoned breeding areas along the northeast coast of Hainan Island, China[J]. *Marine Pollution Bulletin*, 2025, 212: 117518.
- [15] HAKANSON L. An ecological risk index for aquatic pollution control. a sedimentological approach[J]. *Water Research*, 1980, 14(8): 975-1001.
- [16] 中国环境监测总站. 中国土壤元素背景值[M]. 北京: 中国环境科学出版社, 1990.
- [17] 吕留彦, 陈祥娥, 陈武, 等. 黔西北土法炼锌集聚区重金属污染评价及范围界定 [J]. *农业工程学报*, 2024, 40(18): 228-239.
LYU Liuyan, CHEN Xiang'e, CHEN Wu, et al. Assessment of heavy metal pollution and scope definition in the Indigenous Zinc Smelting agglomeration areas of northwest Guizhou Province of China[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2024, 40(18): 228-239. (in Chinese with English abstract)
- [18] 唐海涛, 孟祥添, 苏循新, 等. 基于 CARS 算法的不同类型土壤有机质高光谱预测[J]. *农业工程学报*, 2021, 37(2): 105-113.
TANG Haitao, MENG Xiangtian, SU Xunxin, et al. Hyperspectral prediction on soil organic matter of different types using CARS algorithm[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2021, 37(2): 105-113. (in Chinese with English abstract).
- [19] WANG S, SUN J, FU L, et al. Identification of red jujube varieties based on hyperspectral imaging technology combined with CARS - IRIV and SSA - SVM[J]. *Journal of Food Process Engineering*, 2022, 45(10): e14137.
- [20] SUN W, ZHANG X. Estimating soil zinc concentrations using reflectance spectroscopy[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2017, 58: 126-133.
- [21] DONG J, DAI W, XU J, et al. Spectral estimation model construction of heavy metals in mining reclamation areas[J]. *International Journal of Environmental Research and Public Health*, 2016, 13(7): 640.
- [22] SHEN Q, XIA K, ZHANG S, et al. Hyperspectral indirect inversion of heavy-metal copper in reclaimed soil of iron ore area[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2019, 222: 117191.
- [23] CHANG C W, LAIRD D A, MAUSBACH M J, et al. Near - infrared reflectance spectroscopy-principal components regression analyses of soil properties[J]. *Soil Science Society of America Journal*, 2001, 65(2): 480-490.
- [24] 姬丽, 马琨, 谢铁娜, 等. 宁夏供港蔬菜田土壤重金属分布特征及生态风险评价[J]. *环境科学*, 2024, 45(6): 3512-3522
JI Li, MA Kun, XIE Tiana, et al. Evaluation of heavy metal distribution characteristics and ecological risk of soil of vegetable land for Hong Kong in Ningxia[J]. *Environmental Science*, 2024, 45(6): 3512-3522. (in Chinese with English abstract)
- [25] LIU W, YUAN C, FAROOQ T, et al. Pollution index and distribution characteristics of soil heavy metals among four distinct land use patterns of Taojia River Basin in China[J]. *Gondwana Research*, 2024, 135: 198-207.
- [26] ZHU R, REN Z, PARAJULI M, et al. Assessment of potential ecological and health risk of potentially toxic elements in

- roadside green areas and urban parks[J]. *Journal of Environmental Chemical Engineering*, 2025, 13(1): 115045.
- [27] TIAN Y, QU G, XU R, et al. Iron-based materials for immobilization of heavy metals in contaminated soils: A critical review[J]. *Journal of Environmental Chemical Engineering*, 2024, 12(5): 113741.
- [28] 赵小敏, 杨梅花. 江西省红壤地区主要土壤类型的高光谱特性研究[J]. *土壤学报*, 2018, 55(1): 31-42.
- ZHAO Xiaomin, YANG Meihua. Hyper-spectral characteristics of major types of soils in red soil region of Jiangxi Province, China[J]. *Acta Pedologica Sinica*, 2018, 55(1): 31-42. (in Chinese with English abstract)
- [29] TAN K, CHEN L, WANG H, et al. Estimation of the distribution patterns of heavy metal in soil from airborne hyperspectral imagery based on spectral absorption characteristics[J]. *Journal of Environmental Management*, 2023, 347: 119196.
- [30] JIANG S X, ZHAI F J, ZHANG C, et al. Speciation distribution and risk assessment of heavy metals in sediments from the Yitong River City area[J]. *Environmental Science*, 2020, 41(6): 2653-2663.
- [31] BAI Z, CHEN S, HONG Y, et al. Estimation of soil inorganic carbon with visible near-infrared spectroscopy coupling of variable selection and deep learning in arid region of China[J]. *Geoderma*, 2023, 437: 116589.
- [32] WANG Y, ZOU B, LI S, et al. A hierarchical residual correction-based hyperspectral inversion method for soil heavy metals considering spatial heterogeneity[J]. *Journal of Hazardous Materials*, 2024, 479: 135699.
- [33] 于海洋, 谢赛飞, 郭灵辉, 等. 融合光谱和空间特征的土壤重金属含量极端随机树估算[J]. *农业机械学报*, 2022, 53(8): 231-239.
- YU Haiyang, XIE Saifei, GUO Linghui, et al. Extremely randomized trees estimation of soil heavy metal content by fusing spectra and spatial features[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(8): 231-239. (in Chinese with English abstract)

Hyperspectral inversion of soil heavy metals in mountain farmland around Mojiang Mining Areas in Yunnan, China

TIAN Anhong^{1,2}, LI Zhiyuan¹, FU Chengbiao^{2*}, ZHU Daming¹

(1. Faculty of Land Resource Engineering, Kunming University of Science and Technology, Kunming 650093, China;

2. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: An accurate and rapid identification of heavy metals (HMs) can be required to prevent environmental pollution and health risks near coal mining sites. However, the multiple contaminants have posed a great challenge to monitoring the heavy metals in agricultural soils. In this study, the hyperspectral inversion technique was developed for the soil heavy metals during pollution assessment and prevention. Spatial distribution maps were then constructed to quantitatively evaluate the heavy metal pollution in mountainous agricultural soils. An agricultural area was taken from the mountain farmland that was severely impacted by mining activities near a gold mine in Mojiang County, Yunnan Province, China. A total of 121 soil samples were collected from the study area. The laboratory hyperspectral data was obtained with the contents of heavy metals (arsenic (As), chromium (Cr), copper (Cu), and nickel (Ni)). The Nemerow pollution index and the potential ecological risk index were employed to assess the heavy metal pollution in the study area. Simultaneously, the CARS-IRIV (competitive adaptive reweighted sampling- iteratively retains informative variables) algorithm was applied to select the sensitive bands for the soil heavy metals. The hyperspectral inversion models were then constructed to estimate the contents of the various heavy metals in the soil. Additionally, the spatial distribution of soil heavy metals was obtained in the study area. The results indicated that: 1) The Nemerow pollution index revealed severe pollution in soil, while the potential ecological risk index suggested a moderate level of ecological risk. The first- and second-order derivatives, standard normal variate, and reciprocal transformations were found to significantly enhance the spectral responses. The efficient band selection of the CARS was combined with the variable refinement of the IRIV. The sensitive bands were successfully selected to explore the spectral response of heavy metals. The better performance of CARS-IRIV was achieved in the number of characteristic bands, computational runtime, and inversion accuracy, compared with the CARS and IRIV only. 3) The back-propagation neural network (BPNN) achieved the highest inversion accuracy for As, while the support vector machine (SVM) was identified as the optimal inversion model for Cr, Cu, and Ni. The R^2 values for the optimal inversion models of As, Cr, Cu, and Ni were 0.90, 0.93, 0.67, and 0.94, respectively. The root mean squared errors of external validation (RMSE) were 87.33, 142.63, 2.63, and 70.31 mg/kg, respectively, and the relative percent differences (RPD) were 3.25, 3.84, 1.74, and 4.17, respectively. 4) The spatial analysis indicated that the high-value areas of heavy metals in the study area were primarily concentrated in the upper and lower areas, while the low-value areas were along the edges. There was an overall trend of gradual decrease from the interior to the exterior. The distribution patterns of heavy metal concentration were attributed to the spatial heterogeneity of pollution under mining activities. The finding can provide a strong reference to monitor the heavy metal pollution in agricultural soils near mining areas. The data support can be offered for the monitoring accuracy of soil pollution for the subsequent assessment and remediation.

Keywords: hyperspectral; inversion; machine learning; soil heavy metals; IRIV algorithm; feature band selection