

多光谱影像的陕西大西沟矿区土壤重金属含量反演

王腾军^{1,2}, 赵明海³, 杨 耘^{1*}, 张 扬^{2,4}, 崔琴芳¹, 李陇同¹

1. 长安大学地质工程与测绘学院, 陕西 西安 710054
2. 国土资源部退化及未利用土地整治工程重点实验室, 陕西 西安 710016
3. 陕西铁路工程职业技术学院, 陕西 渭南 714000
4. 陕西省土地工程建设集团, 陕西 西安 710075

摘要 传统的以“点采样+实验室分析”为主的土壤重金属含量分析技术成本高、效率低下,而基于多光谱遥感的土壤重金属高精度定量反演中存在重金属含量影响因子的优化这一难题,以陕西大西沟矿区这类山区地形条件下的金属矿区为例,利用 Landsat8/OLI 多光谱卫星影像、DEM 数据以及外业土壤采样分析数据,开展了矿区土壤重金属含量指示因子分析及定量反演研究。首先,考虑研究区地形地貌特点,设计了沿研究区地形特征线及其两侧坡面均匀分布的样点分布方案,采集了 45 个样本。并对 45 个样本的混合样中的 8 种重金属含量进行了兴趣度分析,根据含量超标程度及矿的类型选取了铜、铅、砷 3 种元素作为分析对象。其次,根据研究区土地利用现状及地形特点,提出了以 Landsat8/OLI 影像 B2 至 B7 波段光谱反射率、粘土矿物比(CMR)、改进归一化水体指数(MNDWI)、差异植被指数(DVI)等八种光谱指数、以及反映研究区地形坡度和坡向三类因子作为反映土壤重金属含量空间分布特征的候选因子。进而,对上述三类候选因子与样本中 3 种金属含量进行了最小二乘相关性分析。根据分析结果,引入了基于估算误差最小准则的金属含量估算模型——基于规则的 M5 模型树的分段线性估算模型。以上述三大类共 17 个指示因子作为模型的输入,利用 80% 的土壤样本分析数据作为模型的训练数据,经过 M5 模型树的构建、平滑和树枝修剪过程,建立了 3 种金属的反演模型实现了研究区土壤中 3 种金属含量的估算。同时,基于均方根误差(RMSE)最小准则确定了以光谱因子为主的最利于反演的最佳指示因子集。最后,用随机选取的 20% 的检验样本对模型进行了反演精度分析,验证了该模型对铜、铅、砷 3 种金属含量的反演精度比普通的线性模型分别提高了 27.3%, 24.6%, 20.9%, 同时,铜、铅元素的可信度也有所提高。利用上述模型的反演结果实现了 3 种金属含量的空间分布制图,并将反演结果与 1990 年公布的国家土壤元素背景值进行了对比。此外,分析了研究区铜、铅、砷 3 种金属的空间分布规律,并利用野外调查结果进行了验证。

关键词 土壤重金属; 多光谱遥感影像; 反演; 空间分布; M5 模型树

中图分类号: X53 **文献标识码**: A **DOI**: 10.3964/j.issn.1000-0593(2019)12-3880-08

引言

全国土壤污染状况调查结果表明:金属矿区周边土壤镉、砷、铅等重金属污染严重^[1]。人们长期生活在重金属污染超标的环境中会导致严重的健康问题^[2]。因此,快速、高精度地获取土壤被重金属污染的信息,并为污染防治提供技术支持是中国当前急需解决的问题。

传统的采用“点采样+实验室分析+空间插值”的土壤重金属调查方法,其采样点位精度高,但非采样点位精度不均匀,需要利用更多的样本及多变量辅助的自适应插值算法^[3]。近年来,基于地物光谱仪的光谱分析技术^[4]以及诱导击穿光谱(LIBS)技术^[5]也被用于土壤重金属含量估算研究。但以上基于点采样的分析技术对于大范围的污染调查来说样点采集成本高且耗时。而矿山企业大多处于山区,交通不便,特别是近年来,企业管理人员警惕性高,配合意愿低,

收稿日期: 2018-10-26, **修订日期**: 2019-02-20

基金项目: 国土资源部退化及未利用土地整治工程重点实验室开放基金项目(SXDJ2017-3), 中央高校基本科研业务费(300102269205, 300102269201), NSFC-新疆联合基金项目(U1703121)和国家自然科学基金项目(41301386)资助

作者简介: 王腾军, 1967 年生, 国土资源部退化及未利用土地整治工程重点实验室副教授 e-mail: wangtj@chd.edu.cn

* 通讯联系人 e-mail: yangyunbox@163.com

给土壤采样及调查等工作带来一定的困难。因此,开展了利用卫星遥感技术的土壤成分、重金属含量间接估算研究。杨灵玉等^[6]利用野外采集的土壤样本数据以及 Hyperion 高光谱影像,利用偏最小二乘回归方法建立土壤重金属含量反演模型,以间接估算土壤锌等元素含量。有学者^[7-8]利用高光谱影像的多种光谱信息,利用回归模型实现土壤重金属含量估计。文献^[9]总结了利用高光谱影像估算土壤成分的优势及挑战。

目前,大多数高光谱卫星影像空间分辨率不高但获取成本高,常用于大区域性土壤污染调查与监测。因此,有学者采用多光谱遥感数据开展了重点区域的土壤重金属污染调查。Peng 等^[10]组合 Landsat8 影像光谱指数和多种环境指示因子建立了土壤重金属含量估算模型,反演了砷(As)、铬(Cr)等元素在土壤中的空间分布,更多研究参见文献^[11-12]。但通常土壤和重金属混杂在一起,难以直接获取重金属的光谱信息,且多光谱卫星影像光谱波段有限,因此,需要综合土壤中各元素含量受地形、植被覆盖等多种自然环境和人为因素作为间接指示因子来提高土壤重金属含量的估算精度。如,文献^[10]将土壤类型等 17 种辅助数据与 Landsat8 光谱特征组合反演土壤重金属含量。更多研究见文献^[11-13]。

前人研究中仍存在一个难题:对于具体的研究对象及研究区地形、地貌特点,如何确定土壤重金属含量的指示因子来实现高精度的定量反演?考虑到陕西大西沟矿生态环境的退化及其特殊的地形地貌及环境特点,开展了该矿区土壤中重点关注的三种重金属含量的指示因子分析及反演模型研究。

1 数据准备

1.1 研究区遥感数据准备

陕西大西沟矿是中国特大型菱铁矿之一,占陕西省铁矿石总储量的 47.6%。矿区位于商洛市柞水县小岭镇,年平均气温 11.1~12.1℃。年均降水量 722.2~811.7 mm,植被覆盖率高达 80% 以上。研究区山高沟深坡陡,最低、最高海拔为 805 和 1 800 m。自 2005 年来主要是露天开采,采矿活动造成了矿区耕地重金属污染、生态及地质环境的破坏。

从美国地质调查局(<https://earthexplorer.usgs.gov/>)获取了研究区 2017 年 1 月 1 日至 12 月 31 日共 24 景 Landsat8/OLI 影像。研究区主要土地利用类型:采矿区、林地、工矿设施、居民生活区及耕地,其中,采矿区、林地、居民生活区占大部分土地面积。利用 ENVI 工具对 Landsat8/OLI 影像进行辐射定标及大气校正,并对影像进行拼接、裁剪处理。研究区内多山,地形复杂,因此,又收集了 30 m 的 ASTER GDEM 数据。

1.2 土壤样本采集、分析与处理

1.2.1 土壤样本采集

采样方案的设计是根据研究区的地形、地貌特点及土地利用现状,并充分考虑样点的代表性,采取随机取点、多点取样、等量混合的原则。按照统计学的抽样要求,需考虑土

壤母质、土壤类型等信息布设样点。研究区土壤冲刷使沟谷的重金属埋深深,采样深度在 30~40 cm 左右,而坡面上重金属埋深浅,采样深度在 10~20 cm 左右。依据国家土壤样本采集、处理和贮存方法,以及研究区地域地貌及流域特征,沿研究区三条主要山脊线方向采样,所有土样均在裸地及稀疏植被区内采集,具体情况视现场而定。

根据上述设计方案,在设计样点位置处的 30 m×30 m 范围内,采取梅花状布点,多点取样混合的方法采集土样。采样时,根据现场实际位置,先记录采样区域中心点的 WGS84 坐标,最后,用取样铲除去土壤取样点上的明显杂质及散落的树枝后,在每个土壤取样点等量采集并均匀混合在一起,保留 1 kg 土壤混合样封装。最后,研究区内布设了 45 个采样点,样点分布如图 1 所示。从图 1 中可以看出,土壤采样点(黑点表示)分布均匀,主要分布在接近沟谷的位置以及人能到达的山坡下、中、上部,这些采样点位代表了重金属聚集程度不同的区域。采样现场图以及所采集的 45 个土壤样本如图 2 所示。

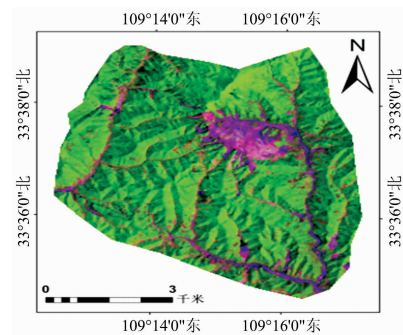


图 1 研究区土壤采样点分布

Fig. 1 Soil sampling points distribution in the study area



图 2 土壤采集现场照片(a)及 45 个样本(b)

Fig. 2 The photos for sampling fieldwork (a) and 45 samples (b)

1.2.2 土壤样本理化分析

实验室内,将采集的原始土壤样本碾碎,去除动植物残留体等处理后,风干、敲碎、过目尼龙筛后用专用设备化验分析。为了确定研究区最受关注的几种重金属,先由 45 个样本形成一个全区域混合样本,利用火焰原子吸收分光光度法、原子荧光法等方法检测 8 种常见的重金属元素含量(单位: $\text{mg} \cdot \text{kg}^{-1}$),如表 1。其次,将分析出的 8 种元素含量与该地区 1990 年公布的国家土壤元素背景值对比,根据该金属矿的类型以及各金属含量超标严重程度,确定以铜(Cu)、

铅(Pb)、砷(As)3种金属为调查对象(下文称目标金属)。

1.2.3 样本统计分析处理

对研究区 45 个有效土壤样本中的三种目标金属化验分析发现:有部分样本铜、铅含量存在异常值。对 45 个土壤样本的三种元素含量统计直方图分布进行了分析,剔除了异常值,并与同区域国家土壤元素背景值进行对比,见表 2。

表 2 表明:研究区这三种金属元素含量明显超出同区域国家土壤元素含量的背景值。

表 1 混合样本中 8 种常见重金属含量分析结果

Table 1 The results of content analysis of 8 popular metals in a mixed soil sample

名称	含量	名称	含量
镉(Cd)	0.200	镍(Ni)	44.4
汞(Hg)	0.111 2	锌(Zn)	130.238
铜(Cu)	68.650	铅(Pb)	163.52
砷(As)	34.8	铬(Cr)	35.2

表 2 剔除异常值后研究区三种金属含量与国家土壤统计数据对比

Table 2 A statistical contrast of the contents of three metals with national data when abnormal values are rejected

类型	样本数	最小值		中位数		最大值		算术平均值		标准差	
		本文	背景值	本文	背景值	本文	背景值	本文	背景值	本文	背景值
铜	44	42.78	6.8	101.20	19.5	338.2	43.6	129.8649	21.4	70.560 6	7.74
铅	40	40.89	13.7	58.20	20.5	143.98	34.5	68.759 5	21.4	23.343 1	5.04
砷	43	17.9	6.3	60.30	10.8	118.4	21	65.688 4	11.1	23.878 6	2.62

2 目标金属含量指示因子及其相关性分析

2.1 土壤金属含量的指示因子

2.1.1 光谱指示因子

为了提高土壤重金属含量反演精度,引入了反映土壤属性、土壤湿度、植被覆盖等环境因素敏感的光谱指数,以及坡度、坡向等地形特征来增强土壤重金属分布规律。选取了 Landsat8/OLI 影像 B2~B7 波段的光谱反射率数据以及其 8

种光谱指数,包括:反映土壤中粘土矿物信息的粘土矿物比(CMR),反映土壤含水状况的改进归一化水体指数(MNDWI),对土壤背景变化敏感的差异植被指数(DVI),反映土壤表层植被覆盖情况的增强植被指数(EVI)和归一化植被指数(NDVI),以及经穗帽变换(K-T)生成的反映地表植被覆盖及土壤湿度的绿度(Greenness)、亮度(Brightness)、以及湿度(Wetness)分量^[14]。各光谱指数定义如表 3;8 种光谱指数图像见图 3 所示。

表 3 三种目标金属含量估计的候选光谱因子

Table 3 The candidates of spectral indicators for three target metals content estimation

类型	因子名称	定义
光谱指数	MNDWI	$(B3 - B6) / (B3 + B6)$
	DVI	$B5 - B4$
	CMR	$B6 / B7$
	EVI	$2.5 \times (B5 - B4) / (B5 + 6 \times B4 - 7.5 \times B2 + 1)$
	NDVI	$(B5 - B4) / (B5 + B4)$
	Greenness	$-0.294 \times B2 - 0.243 \times B3 - 0.542 4 \times B4 + 0.727 6 \times B5 + 0.071 3 \times B6 - 0.160 8 \times B7$
	Brightness	$0.302 9 \times B2 + 0.278 6 \times B3 + 0.473 3 \times B4 + 0.559 9 \times B5 + 0.508 \times B6 + 0.187 2 \times B7$
	Wetness	$0.151 1 \times B2 + 0.197 3 \times B3 + 0.328 3 \times B4 + 0.340 7 \times B5 - 0.711 7 \times B6 - 0.455 9 \times B7$

图 3 表明:选取的上述光谱因子增强了采矿区及道路等重金属污染严重区域的特征。

2.1.2 地形因子

研究区高差较大,坡度较陡,使得矿区顶部土壤携带重金属易向坡下迁移。而坡向会影响植被类型及其覆盖度,会阻碍雨水对土壤重金属的冲刷,以及植物对金属元素的吸收,这两个因素都使得土壤金属的空间分布。考虑到本矿区地形对土壤重金属空间分布的影响,引入了反映地形特征的海拔、坡度与坡向属性来分析这三种重金属含量的空间分布。

2.2 目标金属含量与指示因子相关性分析

2.2.1 三种目标金属之间的相关性分析

为了便于土壤金属反演,基于最小二乘回归分析法,采

用相关系数 r [其定义式(1)]分析上述三种目标金属之间的相关性,结果见表 4 所示

$$r(X, Y) = \text{Cov}(X, Y) / \sqrt{\text{Var}[X]\text{Var}[Y]} \quad (1)$$

式(1)中, $\text{Cov}(X, Y)$ 为金属 X 与金属 Y 的协方差, $\text{Var}[X]$ 为 X 的方差, $\text{Var}[Y]$ 为 Y 的方差。

从表 4 可以得出,三种目标金属之间的相关性不大,故对三种元素单独分析与反演。

2.2.2 目标金属含量与遥感影像季节因素的相关性分析

从研究区 24 景 Landsat8/OLI 影像中排除云层干扰影像后,选取了 3 月、5 月、12 月 3 个月份(日期分别为 2017-03-09, 2017-05-28 和 2017-12-06 代表春、夏、冬 3 个季节)进行相关性分析。对上述 3 个月 Landsat8/OLI 影像的 B2~B7 波

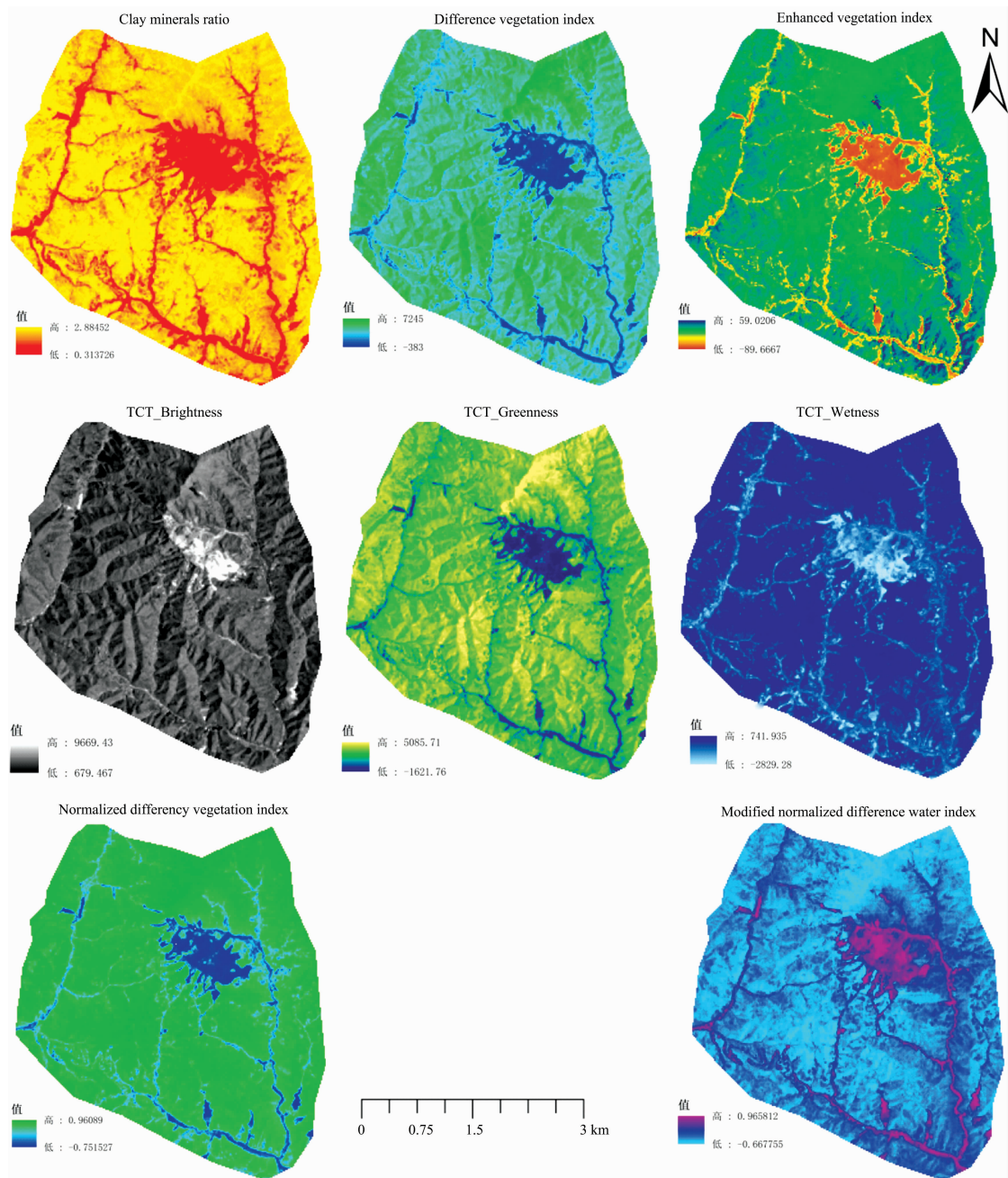


图 3 8 种光谱指数指示因子图像

Fig. 3 Eight images of spectral indices as indicators

表 4 目标金属元素间的相关性矩阵
Table 4 Correlation coefficients matrix among the three target metals

元素	铜	砷	铅
铜	1	0.439 9*	0.587 8**
砷	0.439 9*	1	0.557 4**
铅	0.587 8**	0.557 4**	1

注: * 表示 $p < 0.05$, ** 表示 $p < 0.01$ (下同)。
Note: * represents $p < 0.05$; ** represents $p < 0.01$ (the same below)

段反射率与三种金属含量进行了相关性分析, 结果如表 5 所示。

从表 5 中数据可以看出, 相关系数值都不超过 0.7, 其中, 铜元素对 5 月份的数据较为敏感, 6 个波段的数据, 有五个波段都与其含量达到了极显著相关。但是考虑红、绿波段, 则 12 月份的数据与铜元素相关性最显著; 而整体上看, 5 月、12 月份的影像数据与铅元素含量相关性较好; 对于砷元素, 12 月份的整体相关性较高。以上表明, 春季(如 2、3 月初)、冬季(如 12 月)温度偏低, 研究区遥感影像植被覆盖度低, 利于对土壤金属含量的观测。

表 5 三种目标金属含量与 Landsat8/OLI 影像季节性因素的相关性分析
Table 5 Correlation analysis of three target metals with seasonal Landsat8/OLI imagery

	月份	B2	B3	B4	B5	B6	B7
铜	3	0.441 1**	0.456 3**	0.394 4**	0.124 9	0.027 1	0.150 3
	5	0.403 4**	0.438 8**	0.423 4**	-0.441 0**	0.129 5	0.391 3**
	12	0.517 6**	0.465 6**	0.363 4**	0.005 4	-0.087 5	0.022 9
铅	3	-0.003 2	-0.003 0	-0.017 2	0.085 3	0.020 3	-0.015 0
	5	-0.055 2	-0.240 0	-0.041 3	0.004 9	0.023 8	-0.019 6
	12	0.025 2	0.019 7	-0.009 2	0.069 5	-0.027 7	-0.045 9
砷	3	0.141 8	0.116 3	0.086 8	-0.034 9	-0.059 0	-0.016 7
	5	0.082 3	0.061 1	0.065 5	-0.096 5	-0.107 0	-0.012 3
	12	0.128 4	0.102 4	0.081 3	-0.062 8	-0.066 0	-0.043 3

2.2.3 三种目标金属含量与影像光谱因子相关性分析

对三种金属与 12 月 Landsat8/OLI 影像的 14 种光谱因子进行了相关性分析, 见表 6 和表 7。

从表 6、表 7 得出, 铜元素与红、绿、蓝三个波段以及 CMR, MNDWI, Greenness 表现为极显著相关 ($p < 0.01$), 与 EVI 和 NDVI 显著相关 ($p < 0.05$); 铅元素与 EVI 表现为显著相关 ($p < 0.05$); 砷元素与 MNDWI 表现为显著相关 ($p < 0.05$)。表明这些光谱指数可作为土壤金属含量反演模型建立的指示因子。

表 6 三种目标金属与 6 种光谱反射率因子相关系数

Table 6 Correlation coefficients between three target metals and 6 spectrum reflectivity indicators

名称	B2	B3	B4	B5	B6	B7
铜	0.517 6**	0.465 6**	0.363 4**	0.005 4	-0.087 5	0.022 9
铅	0.025 2	0.019 7	-0.009 0	0.069 5	-0.027 7	-0.045 9
砷	0.128 4	0.102 4	0.081 3	-0.062 8	-0.066 0	-0.043 3

表 7 三种目标金属与 8 种光谱指示因子的相关系数

Table 7 Correlation coefficients between three target metals and 8 spectrum indicators

名称	DVI	EVI	CMR	NDVI	MNDWI	Brightness	Greenness	Wetness
铜	-0.251 0	-0.363 9*	-0.453 3**	-0.370 8*	0.396 4**	0.074 0	-0.386 1**	0.207 2
铅	-0.011 6	-0.313 3*	-0.163 6	-0.127 2	-0.004 0	0.234 8	-0.119 3	-0.117 8
砷	-0.156 5	-0.095 0	-0.220 7	-0.275 6	0.297 5*	-0.021 7	-0.215 0	0.120 5

2.2.4 三种目标金属含量与地形因子的相关性分析

基于最小二乘回归分析法进行了三种金属样本数据与地形因子的相关性分析, 结果表明: 不同的金属与三种地形因子的相关性不同。其中, 坡向 (Aspect) 与铜含量的相关性最大。而海拔 (Altitude)、坡度 (Slope) 分别与铅和砷元素含量的相关性最大。

上述相关性分析结果表明: 铜、铅、砷这三种金属含量与各个指示因子相关性不同, 但实验表明: 仅根据相关性大小来确定各金属的最佳指示因子集并不一定能实现最低的估算误差。为了解决反演模型中最优指示因子集确定的难题, 引入了基于估算误差最小原则的 M5 模型树来实现金属含量的高精度反演, 同时可以确定各金属的最佳指示因子集。

3 基于 M5 模型树的金属含量反演及空间分布分析

3.1 反演模型的建立

不同于普通线性模型和 BP 神经网络模型, 基于 M5 模型树的回归模型是一种基于规则并利用树结构实现回归分析及反演的模型。该模型在建模方面更灵活, 反演精度更高^[15]。因此, 以上述的 17 种指示因子集作为 M5 模型树

的输入, 随机选取 80% 的土壤样本参与建模 (其中, 分别有 35, 32 和 34 个样本参与铜、铅、砷元素含量的反演建模), 经过模型树的构建、平滑, 以及树的修剪三个步骤, 分别建立了三种金属的分段线性模型 (以下简称本模型) 来实现对铜、铅、砷三种元素含量的估算。

铜元素反演模型如式 (2)、式 (3) 和式 (4)。

当 $V_{B2} \leq 474$ 时,

$$V_{Cu} = 160.183 5 + 115V_{MNDWI} \quad (2)$$

当 $V_{B2} > 474$ 且 $V_{Brightness} \leq 2 467.812$ 时,

$$V_{Cu} = -573.327 5 + 0.270 3V_{Brightness} + 0.395V_{B2} \quad (3)$$

当 $V_{B2} > 474$ 且 $V_{Brightness} > 2 467.812$ 时,

$$V_{Cu} = 127.33 + 70V_{MNDWI} \quad (4)$$

铅元素反演模型如式 (5)

$$V_{Pb} = 13.453 82 + 0.182V_{B2} + 0.164V_{Wetness} - 0.096 9V_{B5} \quad (5)$$

对于砷元素, 其反演模型形式如式 (6)

$$V_{As} = 166.56 + 155V_{MNDWI} - 0.050 6V_{B2} \quad (6)$$

其中, V_{Cu} , V_{Pb} , V_{As} 表示铜、铅、砷含量的估算值; V_{B2} , V_{B5} 表示 Landsat8/OLI 影像 B2 和 B5 波段反射率; V_{MNDWI} , $V_{Brightness}$ 和 $V_{Wetness}$ 分别表示 MNDWI 指数值、K-T 变换亮度和湿度值。

上述三种金属的反演模型是基于均方根误差最小准则实现了三种金属含量的估算,同时确定了适合本研究区三种金属含量反演的指示因子集。其中,铜元素选取了 3 月份 Landsat8/OLI 影像 B2 波段、K-T 变换的亮度值、3 月份的改进归一化水体指数作为其最优指示因子集。对于铅元素,选取了 12 月份 Landsat8/OLI 影像 B5 波段、K-T 变换的湿度值以及 Landsat8/OLI 影像 B6 波段作为其最佳指示因子集。对于砷元素,选取了 2 月 Landsat8/OLI 影像 B2 波段、改进归一化水体指数作为其最佳指示因子集。

3.2 反演误差分析

用剩余的 20% 的土壤样本(其中,铜、铅、砷的检验样本个数分别是 9, 8, 9)作为检验样本对上述三类金属反演模型的结果进行误差分析。平均误差(ME)、均方根误差(RMSE)和平均相对误差(MRE)为评价指标。其中, RMSE 能反映模型反演结果的精确度。ME 反映反演值高于 (ME < 0) 或低于 (ME > 0) 实测值的趋势。MRE 反映模型反演结果的平均可信度,三者的值越接近 0 表示模型反演效果越好。ME、RMSE 和 MRE 表达式如式(7)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (M_i - P_i)^2}{n-1}}, ME = \frac{\sum_{i=1}^n (M_i - P_i)}{n},$$

$$MRE = \frac{\sum_{i=1}^n \left(\frac{M_i - P_i}{M_i}\right)}{n} \quad (7)$$

其中, M_i 和 P_i 分别表示第 i 个检验样本金属含量的实测值和估算值; n 为土壤样本数。上述误差评价指标对本模型与普通线性模型(简称普通模型)反演精度对比见表 8。

表 8 三种目标金属含量的估算误差统计
Table 8 The statistics of estimation error of three target metals content

	铜		铅		砷	
	普通模型	本模型	普通模型	本模型	普通模型	本模型
ME	-14.022 1	5.822	5.400 0	5.011 1	5.529 9	-6.937 3
RMSE	29.872 0	21.711	24.514 8	18.496 4	18.480 3	14.609 5
MRE	-0.151 2	0.072 7	0.046 1	0.037 9	0.143 9	-0.211 8

从表 8 可以看出,就均方根误差(RMSE)这一指标而言,对研究区内铜、铅、砷三种元素来说,本文模型比普通的线性模型反演误差分别降低了 27.3%, 24.6% 和 20.9%。表明:本模型对这三种元素的反演精度整体上优于普通的线性回归模型,改善了三种金属含量的反演精度,但反演误差均大于 14.0。而砷元素反演结果的可信度有待改善。

为了更直观的反映估算误差总体趋势,图 4 列出了两种模型估算三种金属含量的残差分布。从图中看出:利用本模型及选取的 17 种候选因子对铜与铅两种元素估算残差明显小于普通线性模型,而本模型的估算残差值更趋近于 0。

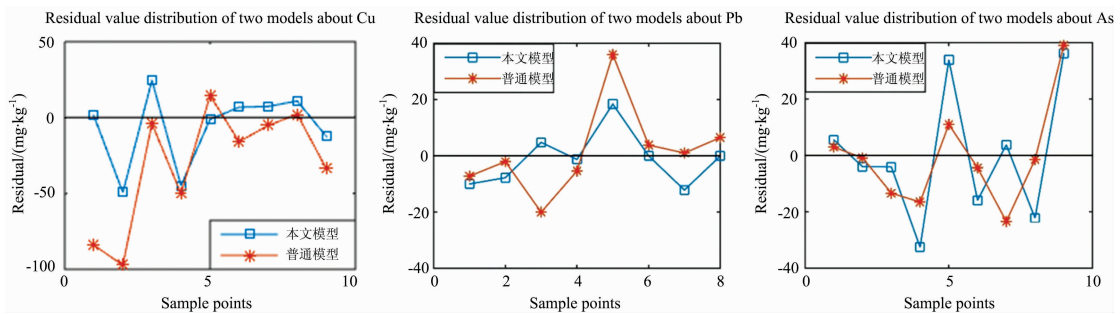


图 4 三种目标金属含量估算误差分析的两种模型残差分布对比

Fig. 4 Residuals distribution of estimation error analysis obtained from two models

3.3 土壤金属含量空间分布规律分析

本模型对三种金属含量预测值的空间分布见图 5,其反演结果统计见表 9。

从图 5 和表 9 中总结出以下规律和结论:

(1)研究区内铜含量变化范围为 0~350 mg · kg⁻¹,其中,铜元素含量在 40~165 mg · kg⁻¹ 区间的土壤面积占研究区总面积的 95.2%; 铅元素含量在 13~198 mg · kg⁻¹ 范围内变化,其含量在 13~86 mg · kg⁻¹ 区间内的面积占总面积的 95%; 对砷元素来说,估算的含量值在 18~222 mg · kg⁻¹ 范围内变化,其中,在 40~100 mg · kg⁻¹ 区间内变化的面积占总面积的 93.6%。以上表明:自 1990 年以来该地区土壤受三种目标金属的污染面积增加。

(2)对比本估算的三种元素含量和表 2 国家公布的陕西省土壤元素背景值最大值,发现:研究区内土壤中铜、锌、

砷三种金属含量超出国家统计数据中最大值的面积分别占整个研究区面积的 98% 以上,表明自 1990 年来该地区土壤受三种目标金属的污染程度加剧。

(3)受三种金属污染严重的区域主要分布在开采区及道路两侧。经实地调查,矿石从坡顶的采矿区向下运输,并经过山谷中居民生活区的道路,由于掉落矿石的积累,使经过的道路金属含量偏高。表明本模型反演结果与实际情况一致。

4 结 论

以陕西大西沟金属矿区为例,利用 Landsat8/OLI 多光谱卫星影像及 DEM 数据,以及 45 个实测土壤样本,提出了沿研究区地形特征线及其两侧坡面均匀分布的样点分布方案。其次,提取了 Landsat8/OLI 影像的光谱反射率、光谱指数和地形三大类指示因子。进而,对铜、锌、砷三种金属含

量与影像光谱、季节及地形等因子进行了相关性分析。最

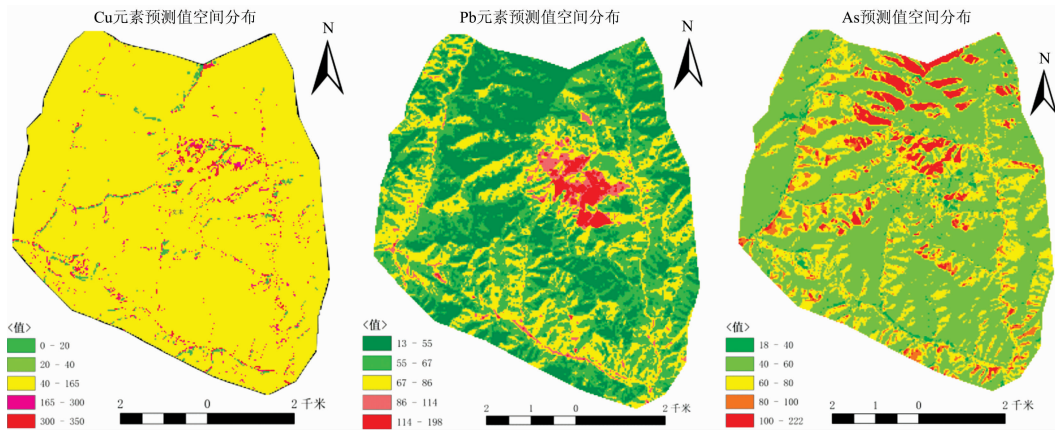


图 5 三种目标金属元素含量空间分布图

Fig. 5 Spatial distributions of the estimated contents of three target metals

表 9 三种目标金属含量反演结果统计

Table 9 The statistics of the estimation result of three target metals content

铜		铅		砷	
含量/(mg · kg ⁻¹)	百分比/%	含量/(mg · kg ⁻¹)	百分比/%	含量/(mg · kg ⁻¹)	百分比/%
0~20	1.1	13~55	34.8	18~40	2.3
20~40	0.1	55~67	35.2	40~60	66.1
40~165	95.2	67~86	25.0	60~80	22.0
165~300	2.4	86~114	3.4	80~100	5.5
300~350	1.2	114~198	1.6	100~222	4.1

后, 以上述三大类影响因素共 17 种因子作为候选因子, 利用基于规则的 M5 模型树作为重金属含量的反演模型, 实现了三种金属元素含量的反演及空间分布成图。同时, 基于误差最小准则确定了三种金属含量估算的最佳指示因子集。实验结果表明: (1) 对普通线性回归模型等反演模型来说, 仅根据相关性大小来确定各金属的最佳指示因子集并不一定能实现高精度的金属含量估算。而引入的 M5 模型树是基于估算误差最小准则, 通过对 17 种候选因子的优化实现了金属含量的高精度反演, 同时可以确定各金属的最佳指示因子集,

解决了反演模型中最佳指示因子集选择的难题; (2) 从反演误差来看, 三种金属元素反演结果 RMSE 值整体偏大, 今后将从指示因子选择及优化等方面提高其估算精度; (3) 研究区土壤中三种金属元素含量的估算值均远超出国家土壤元素背景值, 表明自 1990 年以来研究区土壤受这三种金属污染的面积和污染程度加剧; (4) 三种金属含量空间分布规律的分析表明: 这三种金属含量的最大值通常分布在研究区中部的开采区, 道路两侧以及坡底, 该反演结果与野外实地调查结果一致。

References

[1] Ministry of Environmental Protection & Ministry of Land and Resources(环境保护部, 国土资源部). Report on the National General Survey of Soil Contamination(全国土壤染公报). China Environmental Protection Industry(中国环保产业), 2014, 5: 10.

[2] Yang J, Ma S, Zhou J, et al. Journal of International Medical Research, 2018, 1.

[3] WANG Sheng-li, LIU Wei, ZHANG Lian-peng, et al(王胜利, 刘伟, 张连蓬, 等). Research of Soil and Water Conservation(水土保持研究), 2018, 25(1): 132.

[4] MA Wei-bo, TAN Kun, LI Hai-dong, et al(马伟波, 谭琨, 李海东, 等). Journal of Ecology and Rural Environment(生态与农村环境学报), 2016, 32(2): 213.

[5] YU Yang, ZHAO Nan-jing, MENG De-shuo, et al(余洋, 赵南京, 孟德硕, 等). Chinese Journal of Laser(中国激光), 2018, 45(8): 811001.

[6] YANG Ling-yu, GAO Xiao-hong, ZHANG Wei, et al(杨灵玉, 高小红, 张威, 等). Chinese Journal of Applied Ecology(应用生态学报), 2016, 27(6): 1775.

[7] YUAN Zhong-qiang, CAO Chun-xiang, BAO Da-min, et al(袁中强, 曹春香, 鲍达明, 等). Wetland Science(湿地科学), 2016, 14(1): 113.

[8] QU Yong-hua, JIAO Si-hong, LIU Su-hong, et al(屈永华, 焦思红, 刘素红, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析)

析), 2015, 35(11): 3176.

- [9] Wang F, Gao J, Zha Y. *ISPRS Journal of Photogrammetry & Remote Sensing*, 2018, 136: 73.
- [10] Peng Y, Kheir R, Adhikari K, et al. *Remote Sensing*, 2016, 8(12).
- [11] Liu M, Wang T, Skidmore A K, et al. *Science of the Total Environment*, 2018, 637-638: 18.
- [12] Naderi A, Delavar M A, Kaboudin B, et al. *Environmental Monitoring & Assessment*, 2017, 189(5): 214.
- [13] Liu M, Liu X, Zhang B, et al. *International Journal of Applied Earth Observation and Geoinformation*, 2016, 51: 91.
- [14] Baig M H A, Zhang L, Tong S, et al. *Remote Sensing Letters*, 2014, 5(5): 423.
- [15] Rahimikhoob A. *Water Resources Management*, 2016, 30(9): 3063.

Inversion of Heavy Metals Content in Soil Using Multispectral Remote Sensing Imagery in Daxigou Mining Area of Shaanxi

WANG Teng-jun^{1, 2}, ZHAO Ming-hai³, YANG Yun^{1*}, ZHANG Yang^{2, 4}, CUI Qin-fang¹, LI Long-tong¹

1. College of Geology Engineering and Surveying, Chang'an University, Xi'an 710054, China

2. Key Laboratory of Degraded and Unused Land Consolidation Engineering, the Ministry of Land and Resources, Xi'an 710016, China

3. Shaanxi Railway Institute, Weinan 714000, China

4. Land Construction Group of Shaanxi, Xi'an 710075, China

Abstract The problem of low efficiency and higher cost exists in the traditional method mainly on “field-work point sampling then indoor experimental analysis”. Also the problem that how to choose the optimal factors indicating the content of heavy metals in soils is difficult to solve for the quantitative inversion of high precision using multispectral remote sensing technology. Using Landsat8/OLI satellite imagery, DEM data and soil samples data, the paper performed the analysis indicators of heavy metals in soil and the quantitative inversion of the content of heavy metals in soil in order to achieve an improved accuracy, taking a study case of a mountainous and forestry mining area called Daxigou mineral of Shaanxi in China. The work was as follows: A soil sampling scheme considering terrain and geomorphology characteristics was designed and evenly sampled in both sides along main topographic feature lines in the study area and 45 soil samples were acquired. Furthermore, a mixed samples from 45 samples were analyzed in laboratory so as to choose the most interested metals (i. e. Cu, Zn, As) as our focus according to both the degree of metals content bigger than that of national authoritative statistics and the type of mineral. Secondly, the paper suggested three types of factors including six spectral reflectivity from band two to seven of Landsat8/OLI imagery, and several spectral indices such as CMR, MNDWI, DVI, EVI etc. , derived from Landsat8 image and also slope and aspect factors derived from DEM data were adopted to indicate the characteristics of the spatial distribution of the content of the three metals candidates considering land use and terrain circumstances in the study area. Subsequently, a correlation analysis of the content of three interested metals individually with six spectral reflectivity data, eight spectral indices and three terrain indicators was done using Least Squares principle. According to the consequence of the correlation analysis, the paper introduced the rule-based M5 model tree in the form of piecewise linear model which was used to estimate the content of Cu, Zn, As three metals in the principle of minimizing error rate. And an inversion model for the content of the three metals was constructed through the simulation, smoothing and pruning of the model tree with an input of all three types and 17 indicators mentioned above and 80% training samples. Also, a set of optimal indicators focusing on spectrum for the inversion were determined according to the principle of minimizing RMSE. Finally, the inversion results using 20% random samples were verified, showing that our suggested method achieved a decrease of RMSE value by 27.3%, 24.6%, 20.9%, and an improvement in confidence level for Cu and As, compared to that of the three interested metals using ordinary linear regression model. Also the thematic images showing the spatial distribution were mapped using the model. Then, the comparisons between the estimated value of the content of three metals and the background value published by Chinese government in 1990 were made. Furthermore, the statistical distribution rules of the three metals were concluded and verified using field survey results.

Keywords Soil heavy metal; Multispectral remote sensing imagery; Inversion; Spatial distribution; M5 model tree

* Corresponding author

(Received Oct. 26, 2018; accepted Feb. 20, 2019)