

## 优化光谱指数的露天煤矿区土壤重金属含量估算

亚森江·喀哈尔<sup>1,2</sup>, 茹克亚·萨吾提<sup>1,2</sup>, 尼加提·卡斯木<sup>1,2</sup>, 尼格拉·塔什甫拉提<sup>1,2\*</sup>,  
张 飞<sup>1,2\*</sup>, 阿不都艾尼·阿不里<sup>2,3</sup>, 师庆东<sup>1,2,3</sup>

1. 新疆大学资源与环境科学学院, 新疆 乌鲁木齐 830046
2. 新疆大学绿洲生态教育部重点实验室, 新疆 乌鲁木齐 830046
3. 新疆大学干旱生态环境研究所, 新疆 乌鲁木齐 830046

**摘 要** 光谱学提供了对土壤中许多元素进行定量分析和快速无损检测的方法。可见光和近红外反射光谱(Vis-NIR)为研究土壤重金属污染提供了一个有用的工具。于新疆准东露天煤矿区采集 51 个 0~10 cm 深度的土壤样品, 在实验室中分别测定样品的有机质(SOM)含量、重金属砷(As)含量与高光谱; 使用基于JAVA语言自主开发的两波段组合软件 V1.0(No: 2018R11S177501)计算不同高光谱数据变换形式(原始反射率( $R$ ), 倒数( $1/R$ ), 对数( $\lg R$ )和平方根( $\sqrt{R}$ )下 Vis-NIR 区域(400~2 400 nm)所有两波段组合得到的优化光谱指数(NPDI)与 As 的相关性, 在最优光谱指数( $|r| \geq 0.73$  和  $p=0.001$ )中通过变量重要性准则(VIP)进一步筛选  $VIP \geq 1$  的指数作为模型自变量, 基于地理加权回归(GWR)模型估算 As 含量并使用四个交叉验证度量标准: 相对分析误差(RPD), 决定系数( $R^2$ ), 均方根误差(RMSE)和最小信息准则(ACD)评价模型精度, 从而探讨优化光谱指数方法应用于高光谱检测露天煤矿区土壤重金属砷含量的可行性。结果表明: (1)研究区 As 含量离散度较高, 所有样品中 SOM 含量均小于 2%, 且 As 含量与 SOM 含量在 0.01 的显著性水平上无显著相关性( $|r|=0.113$ )。 (2)As 含量与单波段光谱反射率的相关性很低( $|r| \leq 0.228$ ), 而通过  $R$ ,  $1/R$ ,  $\lg R$ ,  $\sqrt{R}$  计算的 NPDI 与 As 含量的相关性在近红外(NIR, 780~1 100 nm)和短波红外(SWIR, 1 100~1 935 nm)光谱中发现最高的相关系数和最低的  $p$  值( $|r| \geq 0.73$  和  $p=0.001$ ), 在长波近红外(LW-NIR)区域基于  $R$  形成的 NPDI 与 As 含量相关性最高( $|r|=0.74$ )。 (3)VIP 方法分别筛选  $NPDI_{R(1\ 417/1\ 246)}$ ,  $NPDI_{1/R(799/953, 825/947)}$ 、 $NPDI_{\lg R(1\ 023/1\ 257, 1\ 008/1\ 249, 1\ 021/1\ 250, 1\ 020/1\ 247)}$  和  $NPDI_{\sqrt{R}(801/953, 811/953, 817/951, 825/947, 828/945)}$  为 GWR 模型自变量。 (4)从 4 个预测模型的表现可以看出, Model-a( $R$ )与其他三个模型(Model-b( $1/R$ ), Model-c( $\sqrt{R}$ )和 Model-d( $\lg R$ ))相比, 它具有最高的验证系数( $R^2=0.831$ ,  $RMSE=4.912 \mu\text{g} \cdot \text{g}^{-1}$ ,  $RPD=2.321$ )和最低的最小信息准则值( $AIC=179.96$ )。优化光谱指数  $NPDI_{R(1\ 417/1\ 246)}$  有助于快速准确地估算 As 含量, 为进一步获取地表土壤重金属污染分布信息提供理论支持和应用参考, 促进露天煤矿区环境污染快速有效调查和生态可持续发展。

**关键词** 土壤重金属; 优化光谱指数; 地理加权回归模型; 露天煤矿区

**中图分类号:** X53; X87 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)08-2486-09

## 引 言

随着中国工业化和城市化的快速发展, 一系列环境问题和健康风险逐渐增加<sup>[1]</sup>。对土壤污染进行的相关研究表明,

当污染物量超过标准量时, 土壤污染成为最具危害的问题<sup>[2]</sup>。在污染物中, 重金属是主要的污染源。重金属不容易被土壤微生物分解, 导致重金属在土壤中不断积累, 当重金属含量超过一定的阈值时就可能转化成更有毒的化合物<sup>[3]</sup>。在中国, 农田和矿区的土壤正受到人为活动(如采矿,

收稿日期: 2018-06-07, 修订日期: 2018-10-25

基金项目: 国家自然科学基金项目(41761077)和新疆维吾尔自治区自然科学基金项目(2017D01C065)资助

作者简介: 亚森江·喀哈尔, 1993年生, 新疆大学资源与环境科学学院硕士研究生 e-mail: ysj\_0801@163.com

茹克亚·萨吾提, 1987年生, 新疆大学资源与环境科学学院博士研究生 e-mail: 13079931238@163.com

\* 通讯联系人 e-mail: ngr.t@hotmail.com; zhangfei3s@163.com

施肥和农药施用)引起的重金属污染<sup>[4]</sup>。土壤中过量的重金属严重影响作物的生长和质量并通过食物链积累在动物和人体内。例如,As 通过饮用水和其他来源进入人体而引起的慢性病是世界范围内影响数亿人的重要公共卫生问题<sup>[5]</sup>。目前的研究表明,重金属在整个矿区的分布和扩散是一个重要的问题,废弃的尾矿可能是环境污染的主要来源<sup>[6]</sup>。因此,土壤重金属污染问题已引起世界各国的广泛关注。

在重视环境问题和生态保护的同时,矿区的实时监测已成为迫切需要解决的问题<sup>[7]</sup>。传统的主要采用系统抽样和化学分析重金属空间分布调查方法比较复杂。虽然该方法的准确性很高,但它耗时且结果高度依赖于样品数量和质量。因此,这种方法需要物质资源和人力资本的重大投资。遥感为估算土壤重金属含量提供了廉价、快速和环保的替代方法,许多研究已经证明了土壤反射光谱估算土壤重金属含量的可能性<sup>[8]</sup>。

目前,使用高光谱技术在各种尺度上识别和绘制环境因子的应用越来越多<sup>[9]</sup>。高光谱技术的优势在于光谱信息的丰富性和连续性,使得土壤和植物的一些非光谱活性特性如重金属的估算成为可能。近红外光谱(NIRS)反射率覆盖可见光和近红外波段(400~2 400 nm)是另一种定量和定性技术,它是对初步研究有用的快速非破坏性方法<sup>[10]</sup>。NIRS 对于检测土壤物理、化学性质如水分,总碳、氮和磷含量,矿物氮、有机质质量以及土壤的生物参数(如呼吸和微生物)表现出良好的效果。NIRS 提供了对土壤进行快速分析而无需化学预处理的方法,通过使用合适的预测模型方法估算土壤中的重金属是可行的。

尽管 As 在 Vis-NIR 中是光谱中性的,但是与 As 结合的矿物的光谱特征可以用于利用光谱数据间接检测土壤中的金属离子<sup>[11]</sup>。高光谱估算 SOM 含量的经验模型主要基于 Vis-NIR 区域(450, 500, 550, 600, 700, 900, 1 000, 1 400, 1 900, 2 050, 2 200, 2 250, 2 400 和 2 470 nm)<sup>[12]</sup>,估算 SOM 的波段理论上也可以估算 As 含量,因为 SOM 和 As 之间存在显著的负相关关系。Shi<sup>[13]</sup>发现,几种重要波长(480, 600, 670, 810, 1 980, 2 050 和 2 290 nm)是 As 的敏感波长,其中大部分对应于 SOM 预测的特征波长,因此得出结论:利用土壤光谱预测 As 含量的机制可能是基于光谱与农田光谱活性有机物的关系。Chakraborty<sup>[14]</sup>探讨了漫反射光谱(DRS)的能力,利用 Vis-NIR 检测 As 污染,并建议 As 的预测取决于 As 与光谱活性 SOM 和铁/铝氧化物之间的密切关联。Zheng 利用高光谱数据对 As 和 SOM 进行了研究,结果表明 SOM 丰富的预测模型精度低于 SOM 含量较低的预测模型。Singh<sup>[15]</sup>在低 SOM 条件下研究了 As 含量的敏感波段,并提出 450, 500, 600, 650, 700, 750, 800, 900, 1 000, 1 200, 1 400, 1 900, 2 050, 2 200, 2 250, 2 350, 2 400 和 2 470 nm 是估算 As 含量的重要波长。然而,一些研究也证实,仅仅选择与土壤中 As 相关的波段对于应用来说是不够的,需要尽更多的努力来提高解释能力,并减少使用反射光谱法估算 As 含量的计算成本。使用多种敏感波段估算重金属含量可能在一定程度上受到限制,或许多种高光谱指数的组合可以提高土壤重金属的估算精度。

因此,旨在利用 Vis-NIR 区域构建优化光谱指数,并挑选出通过原始反射率和变换反射率计算出的敏感优化光谱指数,基于最敏感的指数构建估算模型,用于预测研究区土壤 As 含量,探讨通过优化光谱指数提高土壤 As 估算精度的可行性。研究结果有望帮助改进遥感方法对土壤重金属污染的评估。

## 1 研究区概况

准东煤田位于天山北侧,吉木萨尔县,奇台县,木垒县境内。位于北纬 44°30′—45°00′,东经 88°40′—91°20′之间,海拔 300~600 m,煤炭总储量达 3 900 亿 t。研究区位于欧亚大陆腹地,地势北高南低,远离海洋,属于典型的极端干旱大陆性气候,夏季适度温暖,冬季寒冷干燥,年平均气温 7℃,年平均降雨量和蒸发量分别为 183.5 和 2 042.3 mm<sup>[16]</sup>。土质差,主要为灰棕色和棕色碳质土,石膏土(含石膏)和粗砾土,属于碱性土壤。表层土壤有机质含量在 2% 以下,自然植被类型少,结构单一,植被稀疏,多为沙漠植被。本工作主要讨论五彩湾准东煤田矿区及其周边区域。

## 1 实验部分

### 1.1 样品采集

2015 年 7 月在准东露天煤矿区进行了详细的土壤采样调查。共采集 51 个 0~10 cm 深度的土壤样品。用 GPS 记录所有采样点的确切位置。为了减少含水量的影响,所有样品充分风干(20~24℃)后,粉碎并通过 2 mm 筛网筛分以除去结石,植物残渣和其他杂质。重铬酸钾法用于测定 SOM。每个样品被分为两个子样品,一个用于光谱测量,另一个用于重金属含量的物理和化学分析。

通过以下步骤制备用于重金属化学分析的土壤样品:加入混酸(氢氟酸和氯酸组合物),将样品置于电热板上加热,然后干燥并溶于 10% 稀硝酸中直至土壤溶液变成灰白色,随后将土壤溶液澄清并在 20 mL 蒸馏水中冷却。As 含量用日立 Z-2000 原子分光光度计测定。为确保准确性,每个样品进行了三组平行实验,其平均值用作样品 As 含量的最终结果。GSS-8 用于质量控制,各种金属的回收率均在国家标准参考物质设定的检测结果质量控制要求的允许范围内。考虑到环境背景,与当地地质和采矿活动的关系,光谱吸收属性以及数据的完整性和可变性,As 的存在被认为是环境污染的指示。为了尽量减少实验误差,国家标准土壤样品被用于重金属含量的化学分析。

### 2.2 光谱测量与预处理

光谱测量在暗室中进行,以减少杂散光的影响。所有土壤样品分别放入直径为 20 cm,深度为 2 cm 的宽圆形容容器中并进行表面平滑处理。为了降低测量误差,这些容器被涂成黑色。使用 ASDFieldSpec® 3 便携式光谱仪(Analytical Spectral Device, Boulder, CO, USA)测量所有土壤样品的反射光谱,该光谱仪覆盖的光谱范围为 350~2 500 nm 并提供每秒 10 次扫描的数据采集,采样间隔为 1.4 nm(350~1 000 nm)

和 2 nm(1 000~2 500 nm), 而重采样间隔为 1 nm。一个 50 W 的卤素灯作为实验室反射率测量的光源, 垂直放置在距离每个土壤样品表面 50 cm 处。探头安装的天顶角为 30°并且距离每个土壤样本 15 cm。每个反射率测量都通过 100% 反射率的标准化平板进行校准以确保准确性。每个土壤样品, 采集 20 条光谱曲线, 并将 20 个光谱反射率的平均值作为最终反射率, 使用 ViewSpecProV6.0 软件计算平均光谱反射率。由于仪器噪声的存在, 从土壤样品的初始反射光谱中去除了光谱带(350~399 和 2 451~2 500 nm), 以改善信噪比。然后对光谱进行 Savitzky-Golay 平滑法处理, 这种方法可以减少随机噪声对校准模型鲁棒性的影响<sup>[17]</sup>。然后, 分别用  $1/R$ ,  $\lg R$  和  $\sqrt{R}$  等 3 种数学方法对高光谱反射率进行变换, 所有的变换都是使用 Origin-pro 2016 软件和 ENVI 5.3 实现。

### 2.3 研究方法

#### 2.3.1 优化光谱指数的计算

NPDI 是在全光谱区域(400~2 400 nm)使用可用波段( $i$  nm 和  $j$  nm)的所有可能组合来计算。使用两波段组合软件 V1.0(No: 2018R11S177501, 基于 JAVA 开发)计算 NPDI 和 As 含量之间的相关性, 并在 Matlab 2012a 软件中绘制二维等势图, 其允许评估不同波段组合并选择敏感光谱指数区域。使用 0.01 水平的显著性检验确定 As 的最有效和最敏感指数。光谱指数定义为

$$NPDI = (R_i + R_j) \times R_j \quad (1)$$

其中  $R$  是光谱反射率值, 下标( $i$  nm 和  $j$  nm)是以纳米(nm)为单位的波长。

#### 2.3.2 建模方法

GWR4 是用于校准 GWR 模型的基于 Microsoft Windows 的应用软件的新版本, 可用于探索因变量与自变量之间不同地理上的关系。GWR 模型可以被认为是一种具有不同地理参数的回归模型, 常规的 GWR 由方程描述<sup>[18]</sup>。

#### 2.3.3 统计分析

Wold 等首次发明了 VIP 选择方法。VIP 分数总结了个体变量对 PLS 模型的影响, VIP 分数为选择贡献最大的自变量提供了一个有效的办法。对于给定的模型和数据集, 总是只有一个 VIP 分数向量总结所有分量和因变量<sup>[19]</sup>。

#### 2.3.4 模型验证

为了评估土壤重金属含量预测模型的性能, 使用了 4 个交叉验证度量标准: RPD,  $R^2$ , RMSE 和 ACI。通常, 一个稳健的模型具有高  $R^2$ , RPD 和低 RMSE。 $R^2$  和 RPD 通常用于评估预测精度, 因为 RMSE 取决于测量值的范围。Chong 将 RPD 分为三类, 如果  $RPD < 1.4$  或  $R^2 < 0.50$  表示模型不可用;  $1.4 \leq RPD < 2.0$  或  $0.50 \leq R^2 < 0.80$  显示中等质量的模型, 可以进行近似定量预测;  $2.0 \leq RPD$  或  $0.8 \leq R^2$  表明优秀的定量模型<sup>[20]</sup>。在地理加权回归建模的统计预测方面, ACI 最为合适, ACI 越小, 预测模型精度就越高。

## 3 结果与讨论

### 3.1 土壤重金属含量变化特征

采用描述性统计, 单因素方差分析(表 1)对重金属 As 和 SOM 的空间分布状况进行了总结。As 含量平均值为  $31.5 \mu\text{g} \cdot \text{g}^{-1}$ , 是新疆土壤背景值的 3 倍, 在土壤中有显著的积累趋势, 变异系数为 39%, 属强变异性。根据国家土壤质量标准(GB15618—1995), 51 个样品中有 11 个土样(22%)超标( $40 \mu\text{g} \cdot \text{g}^{-1}$ )。As 含量与 SOM 含量在 0.01 的显著性水平上无显著相关性( $|r| = 0.113$ ), SOM 含量平均值为  $6.35 \text{g} \cdot \text{kg}^{-1}$ , 所有样品中 SOM 含量均小于 2%, 变异系数(CV)值为 160%, 属于高度可变。As 与 SOM 含量没有很强的相关性, 因此, SOM 对 As 含量的预测没有明显影响。

建模时, 根据 As 的含量将样本从高到低排序, 并且基于等距离方法选择建模集和验证集。从描述性统计特征(表 1)可以看出, 建模集和验证集的 As 含量最大值分别为 55.4 和  $49.4 \mu\text{g} \cdot \text{g}^{-1}$ , 最小值分别为 1.5 和  $2.2 \mu\text{g} \cdot \text{g}^{-1}$ , 平均值分别为 30.4 和  $31.4 \mu\text{g} \cdot \text{g}^{-1}$ , CV 分别为 37% 和 40%。CV 反映了重金属含量的变异性, 表明人类活动对重金属含量的影响程度, 更大的系数值表示土壤受人类活动影响更大<sup>[21]</sup>, 变异系数范围:  $CV < 15\%$  为轻微变化,  $15\% < CV < 36\%$  是中等变化,  $CV > 36\%$  是高度可变的。在该研究中, As 具有大于 36% 的变异系数, 这表明, 研究区土壤 As 含量离散度较高, 且煤矿区周边土壤受到的影响最大。

表 1 研究区土壤 As 含量统计特征

Table 1 Statistical characteristics of As concentration in soil of sampling area

Element	Sampling points /n	Maximum	Minimum	Mean	Standard deviation	Coefficient of variation (CV/%)
Whole set/ $(\mu\text{g} \cdot \text{g}^{-1})$	51	55.4	1.5	31.5	12.2	39
SOM/ $(\text{g} \cdot \text{kg}^{-1})$	51	95.90	0.26	6.35	10.15	160
Calibration set/ $(\mu\text{g} \cdot \text{g}^{-1})$	26	55.4	1.5	30.4	11.1	37
Validation set/ $(\mu\text{g} \cdot \text{g}^{-1})$	25	49.4	2.2	31.4	12.6	40

### 3.2 优化光谱指数与土壤重金属含量相关性分析

As 含量与单波段光谱反射率的相关性如图 1 所示, 可以看出, 51 个土壤样品反射率的原始反射率和 3 个数学变换反射率与 As 含量呈低度相关, 低于 0.228。相关顺序为:  $R$

$> \sqrt{R} > \lg R > 1/R$ 。在可见光范围(400~500 nm)内, 原始反射率与 As 含量之间相关性最高达 0.16。根据相关性结果, 可以得出结论: 直接使用高光谱数据估算研究区土壤 As 含量是毫无意义的。因此, 决定计算高光谱指数来研究指数

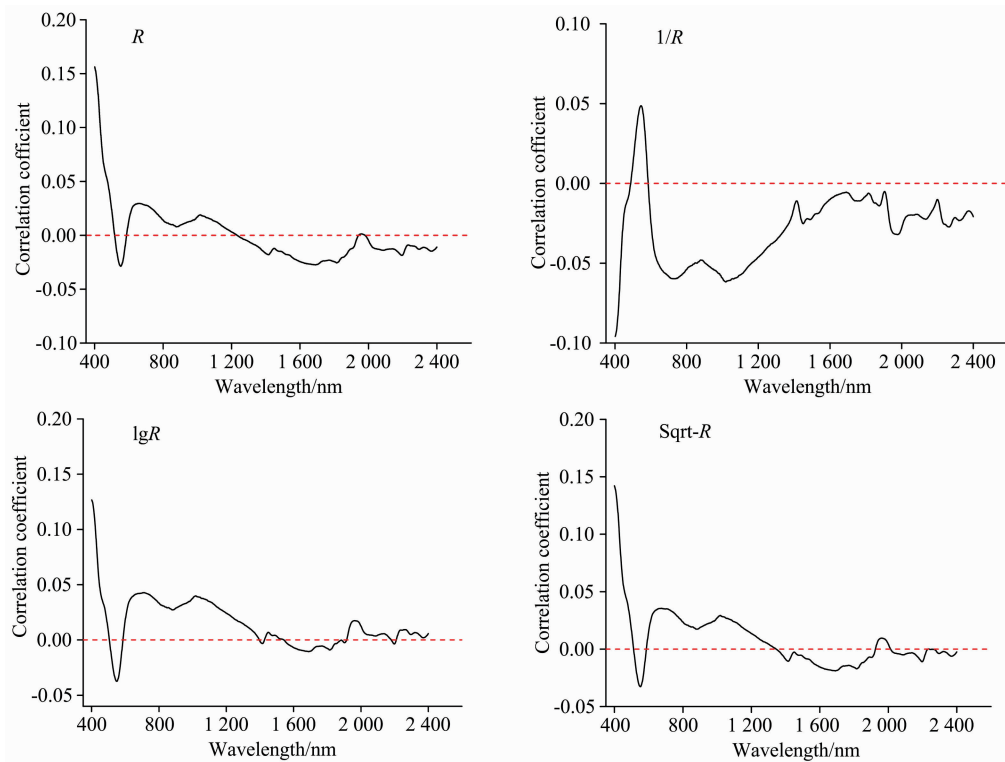


图 1 土壤光谱反射率与 As 含量的关系

Fig. 1 Correlation of soil spectral reflectance and As contents

在估算土壤 As 含量方面的可行性。

正如先前的研究所示, Vis-NIR 区域(450, 480, 500, 550, 600, 670, 700, 810, 900, 1 000, 1 400, 1 900, 1 980, 2 050, 2 200, 2 250, 2 290, 2 400 和 2 470 nm)是估算土壤中 As 含量的重要波长, 并且当土壤中富含有机质时预测模

型精度会低于 SOM 含量较少时的模型。为了验证这一结果, 分析了研究区 As 和 SOM 之间的相关性( $|r| = -0.143$ ), 并且 SOM 含量低于 2%。因此, 在忽略 SOM 干扰的条件下, 基于 Vis-NIR 区域光谱的原始反射率和转换反射率计算优化光谱指数对土壤 As 的敏感性(图 2)。

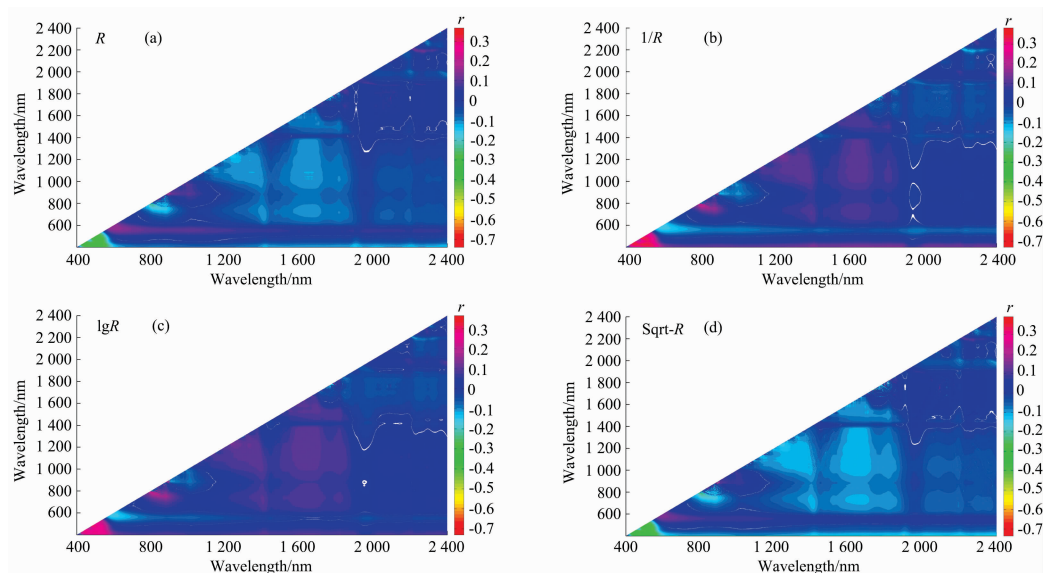


图 2 As 含量与 NPDIs 之间的相关系数二维等势图

Fig. 2 The two dimensional maps of correlation coefficient between As contents and NPDIs

表 2 表示使用 400~2 400 nm 计算的相关系数( $r$ ), 在 780~1 100 和 1 100~1 935 nm 发现最高的  $r$  值和最低的  $p$

值( $|r| \geq 0.73$  和  $p = 0.001$ )。通过原始和变换反射率( $\sqrt{R}$ ,  $\lg R$ )计算的 NPDIs, 在 SWIR 区(1 408/1 240, 1 417/1 246,

1 404/1 251, 1 935/2 391, 1 935/2 372 nm) 波段组合和 SWIR 与 NIR 区(1 242/1 018, 1 241/1 011, 1 243/1 010, 1 247/1 020, 1 250/1 021, 1 249/1 008 和 1 257/1 023 nm) 组合时与 As 含量具有较好的相关性( $r \geq 0.73$ )。通过转换反射率(1/R, lgR)计算的 NPDIs 在 NIR 区(825/947, 817/951, 799/952, 811/953, 828/945, 825/947 和 817/951 nm) 波段相互组合时与 As 含量也表现出较高的相关性( $|r| \geq 0.73$ )。在 LW-NIR 区域基于原始光谱反射率形成的 NPDIs 与 As 含量相关性最高( $|r|=0.74$ )。在 NIR 和 SWIR 区域也发现高  $r$  值[0.73, 图 2(b), (c)和(d)]。

### 3.3 地理加权回归分析

挑选的优化指数再经 VIP 过滤。如果  $VIP \geq 1$ , 则该指数将更适合作为自变量使用; 如果不是, 则指数不会被选作模型自变量。从图 3 中可以看出, 在通过  $R$  计算的优化指数 NPDIs 中,  $VIP$  (NPD<sub>I</sub>(1417/1 246)) 的值大于 1。1/R 中 NPDI<sub>I</sub>(799/953) 与 NPDI<sub>I</sub>(825/947) 的  $VIP$  值均  $\geq 1$ , 因此选择这两个 NPDI 为自变量。根据  $VIP$  结果可知, 尽管在  $\sqrt{R}$  和 lgR 形式下分别有 7 个 NPDI 与 As 含量具有高度相关性, 但其中只有 4

个来自转换的 NPDIs (NPDI<sub>I</sub>(1023/1 257, NPDI<sub>I</sub>(1008/1 249, NPDI<sub>I</sub>(1021/1 250, NPDI<sub>I</sub>(1020/1 247)) 更适合构建模型; 通过 lgR 获得的 5 个 NPDIs (NPDI<sub>I</sub>(801/953, NPDI<sub>I</sub>(811/953, NPDI<sub>I</sub>(817/951, NPDI<sub>I</sub>(825/947, NPDI<sub>I</sub>(828/945)) 的  $VIP$  值高于 1, 通过  $VIP$  检验被筛选为模型自变量。

表 2 高光谱指数与 As 含量之间的相关性  
Table 2 Correlation between hyperspectral indices and As contents

Index	$R$	1/R	lgR	$\sqrt{R}$
			$ r  \geq 0.73$	$ r  \geq 0.73$
NPDI	$ r  \geq 0.73$	$ r  \geq 0.73$	828/945	1 242/1 018
	1 408/1 240	825/947	825/947	1 241/1 011
	1 417/1 246	817/951	817/951	1 243/1 010
	1 404/1 251	799/953	811/953	1 247/1 020
		811/952	801/953	1 250/1 021
			1 935/2 372	1 249/1 008
			1 935/2 391	1 257/1 023

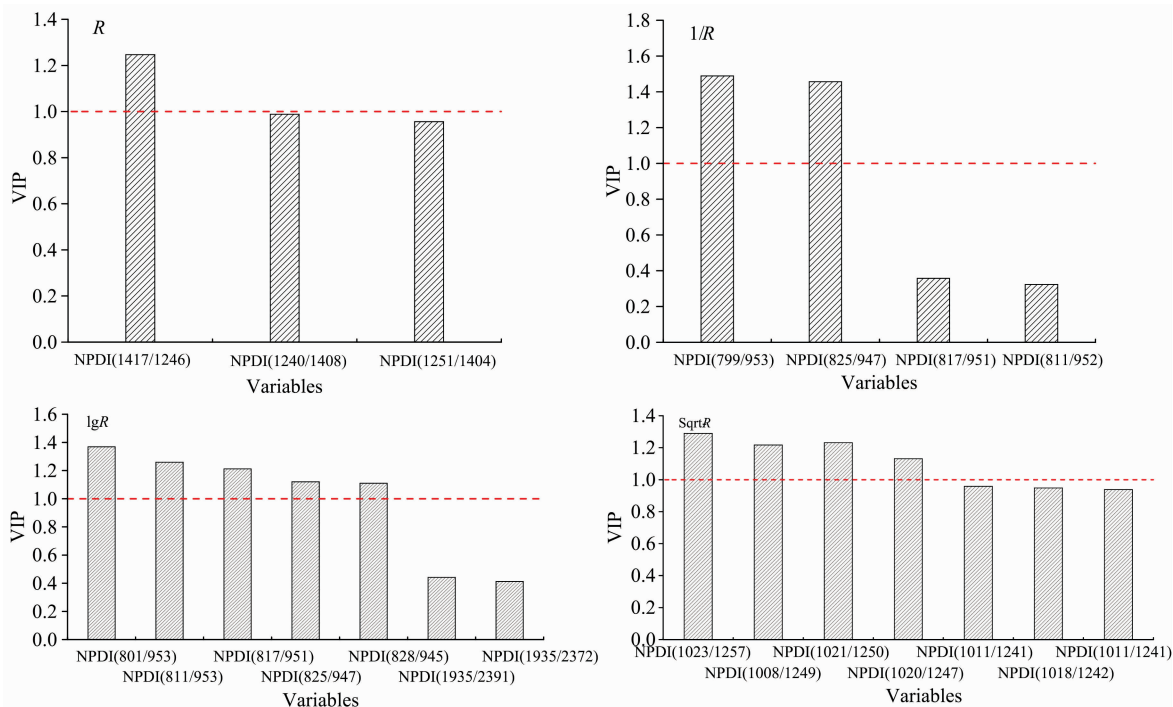


图 3 模型自变量 VIP 筛选图

Fig. 3 Map of the model parameters (VIPs)

26 个采样点  $VIP$  值较高的 NPDIs 用作重金属含量的预测因子, 25 个样品用于验证。表 3 是 GWR 预测模型的模型精度指标, 图 4 是模型预测和实测重金属含量的散点图。如表 3 和图 4 所示, 4 个模型的 RMSE 值范围为  $12.781 \sim 4.912 \mu\text{g} \cdot \text{g}^{-1}$ , RPD 值为  $0.543 \sim 2.321$ ,  $R^2$  为  $0.210 \sim 0.831$ 。a 模型 (NPDI<sub>I</sub>(1417/1 246)) 表现优于 b 模型 (NPDI<sub>I</sub>(799/953, 825/947)) 和 d 模型 (NPDI<sub>I</sub>(801/953, 811/953, 817/951, 825/947, 828/945)), 它具有最高验证系数 ( $R^2=0.831$ ,  $\text{RMSE}=4.912 \mu\text{g} \cdot \text{g}^{-1}$ ,  $\text{RPD}=2.321$ ) 和最

低 AIC 值 ( $\text{AIC}=179.96$ )。c 模型 (NPDI<sub>I</sub>(1023/1 257, 1 008/1 249, 1 021/1 250, 1 020/1 247)) 显示出较高的验证系数 ( $R^2=0.649$ ,  $\text{RMSE}=7.141 \mu\text{g} \cdot \text{g}^{-1}$ ,  $\text{RPD}=1.774$ ) 和较低的 AIC 值 ( $\text{AIC}=197.16$ )。d 模型 (lgR) 和 b 模型 (1/R) 产生的验证系数较低 ( $R^2=0.210$ ,  $\text{RPD}=0.811$ ,  $\text{RMSE}=12.781 \mu\text{g} \cdot \text{g}^{-1}$ ;  $R^2=0.371$ ,  $\text{RPD}=1.051$ ,  $\text{RMSE}=9.95 \mu\text{g} \cdot \text{g}^{-1}$ ) 并且 AIC 的值较高 ( $\text{AIC}=236.17$ ,  $\text{AIC}=218.26$ )。根据模型的预测结果, 从室内原始光谱数据得到的优化指数 (NPDI<sub>I</sub>(1417/1 246)) 证明了光谱指数定量方法检测土壤中重金属 As

含量的可行性。尽管 1 400 nm 附近的吸水特征可能会影响土壤反射率,但 Singh 认为 1 200 和 1 400 nm 是预测土壤 As 含量的重要波长。

表 3 GWR 预测模型验证系数

Table 3 Description of validation factors for every GWR prediction model

Cross-Validation metrics	Model-a (R)	Model-b (1/R)	Model-c ( $\sqrt{R}$ )	Model-d (lgR)
$R^2$	0.831	0.371	0.649	0.210
RMSE	4.912	9.951	7.141	12.781
RPD	2.321	1.051	1.774	0.543
AIC	179.96	218.26	197.16	236.17

由于土壤光谱是土壤中有机质,氧化铁,粘土矿物质和其他物质的全反射率,很难用普通的光谱方法估算土壤中的重金属含量。为了估算重金属含量,以前的研究通常是通过原始反射率或转换反射率(一阶微分,二阶微分,数学变换等)获得的敏感波段作为模型参数用于预测重金属含量。偏最小二乘回归(PLSR),逐步多元线性回归(MLR)和人工神

经网络(ANN)模型被广泛应用于估算各种环境中的重金属含量。Nawar 等<sup>[22]</sup>使用 ASTER 图像与 PLSR 模型估算重金属含量,得出结论:该方法能够很好地预测 Cr, Zn, Pb, Ni, Mn 等重金属含量,  $R^2$  分别为 0.70, 0.89, 0.89, 0.72 和 0.74, 但对 Fe, Cu 和 Cd 的预测效果欠佳。Stazi<sup>[23]</sup>采用控制实验的方式,使用高光谱仪检测土壤中重金属的含量,并比较两种多元统计回归分析模型——偏最小二乘和支持向量机。通过 VIP 方法筛选出反射光谱的 580, 660, 715 和 780 nm 波段为最敏感的模型自变量。Shi 等分析了高光谱植被指数在估算稻田土壤 As 含量方面的表现,并建议将三波段植被指数  $(R_{716} - R_{568}) / (R_{552} - R_{568})$  作为估算认为土壤 As 含量的指标。Sun 等<sup>[24]</sup>使用 2 200 nm 附近的组合波段估算土壤中重金属 Zn 的含量,达到较高的预测精度,  $R^2$  达到 0.640。江振蓝等证明使用 GWR 预测模型能提供更强大的重金属含量(Cr, Cu, Zn)预测能力,其  $R^2$  值分别为 0.64, 0.79 和 0.77。由于 GWR 预测模型是一种非线性的建模方法,因此它能在重金属含量预测时表现得更好。GWR 模型对土壤重金属的预测准确性取决于重金属与自变量之间相关性的空间非平稳程度。

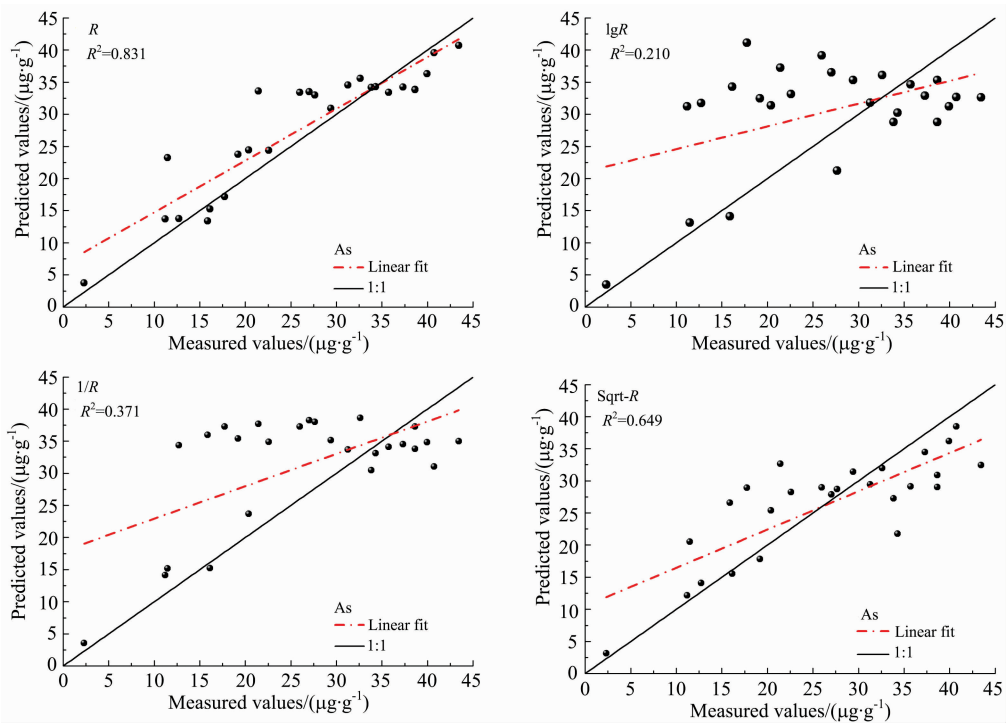


图 4 GWR 预测模型散点图

Fig. 4 Scatter plot map of GWR predicted model

在本工作中,计算优化光谱指数并讨论光谱指数在研究土壤重金属 As 含量时的可行性。利用各种光谱反射率计算指数,并用于预测土壤重金属 As 的含量。新构建的 NPDIs (表 2)与 As 含量显著相关,基于  $R, 1/R, \lg R, \sqrt{R}$  得出的 NPDIs 构建 GWR 模型用于估算 As 含量。结果证明, VIP 方法在样本较少且几个自变量间相关性较强的情形下可以有效筛选自变量,并增强模型的效率与应用性,在基于 NPDIs 构

建的模型中, Model-a(R) (NPD<sub>I1 417/1 246</sub>) (图 4) 表现出优异的效果和最高的验证系数 ( $R^2 = 0.898$ , RMSE = 4.912  $\mu\text{g} \cdot \text{g}^{-1}$ , RPD = 2.321) 和最低的 AIC 值 (AIC = 179.96)。空间分布图 [图 5(a), (b)] 表明 As 的预测状况 [图 5(b)] 与研究区内 As 的实测状况 [图 5(a)] 呈现相似分布状态,该结果进一步验证了预测模型的准确性。

研究结果表明,双波段优化指数 (NPD<sub>I1 417/1 246</sub>) 具有

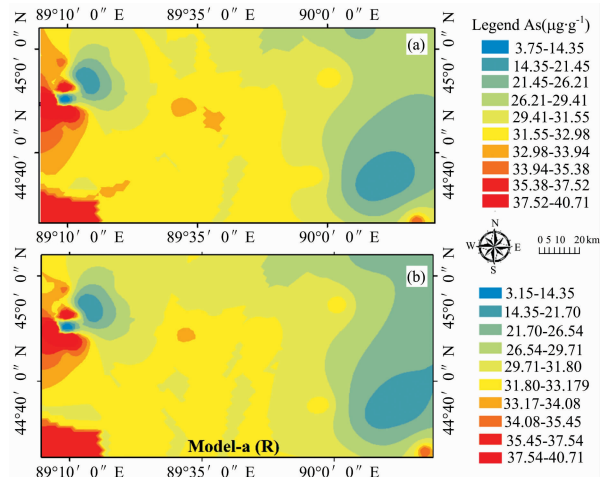


图 5 实测值和预测值的 As 分布等值线图

(a): 实测值; (b): Model-a 的预测值

Fig. 5 Arsenic distribution contour map of field measured value and predicted value

(a): Field measured value; (b): Predicted value by model-a

估算土壤中 As 含量的潜力。根据前人的研究, As 在 430 nm 附近波段会因为土壤中的铁氧化物(主要是赤铁矿和针铁矿)而出现反射特征,而在 1 290~1 310 nm 之间的反射特征可以归因于 SOM 或含铝氧化物粘土矿物中的 O—H 和 C—H 键,特别是 2 250 和 2 450 nm 之间的反射峰很可能与 SOM 中的 C—H 键有关。值得注意的是,这些光谱特征在 As 含量和优化光谱指数进行相关分析过程中被消除了。Singh 提出 1 200 和 1 400 nm 是预测土壤 As 含量的重要波长。这些结果表明,新构建的双波段优化指数可以被推荐为评估研究区其他土壤重金属含量的指标。

## References

- [1] Cai L M, Xu Z C, Qi J Y, et al. *Chemosphere*, 2015, 127: 127.
- [2] Zhuang W, Liu Y, Chen Q, et al. *Ecological Indicators*, 2016, 69: 252.
- [3] ZHOU Jian-jun, ZHOU Ju, FENG Ren-guo(周建军, 周 桔, 冯仁国). *Bulletin of Chinese Academy of Sciences(中国科学院院刊)*, 2014, 29(3): 315.
- [4] ZHANG Xiao-min, ZHANG Xiu-ying, ZHONG Tai-yang, et al(张小敏, 张秀英, 钟大洋, 等). *Environmental Science(环境科学)*, 2014, 35(2): 692.
- [5] Rahman M A, Rahman A, Khan M, et al. *Ecotoxicology & Environmental Safety*, 2018, 150: 335.
- [6] Qin H B, Takeichi Y, Nitani H, et al. *Environmental Science & Technology*, 2017, 51(11): 224.
- [7] LI Xin-long, CHEN Sheng-bo, CHEN Lei(李鑫龙, 陈圣波, 陈 磊). *Science Technology and Engineering(科学技术与工程)*, 2014, 14(7): 121.
- [8] FU Xin, ZHAO Yan-ling, LI Jian-hua, et al(付 馨, 赵艳玲, 李建华, 等). *China Mining Magazine(中国矿业)*, 2013, 22(1): 65.
- [9] 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.
- [10] Horta A, Malone B, Stockmann U, et al. *Geoderma*, 2015, 241: 180.
- [11] Tonmoy T H, Hanif M A, Rahman H A, et al. *International Conference on Computer and Information Technology, IEEE*, 2017, 111: 343.
- [12] Nawar S, Buddenbaum H, Hill J, et al. *Soil & Tillage Research*, 2016, 155: 510.

## 4 结 论

探讨了优化光谱指数在土壤 As 含量估算中的潜力。基于 GWR 模型用原始反射率及其三种变换反射率计算的 NPDI 估算重金属含量,得出以下结论:

(1)用  $R$ ,  $1/R$ ,  $\lg R$ ,  $\sqrt{R}$  计算的 NPDI 在 NIR 和 SWIR 中发现最高的  $r$  值和最低的  $p$  值( $|r| \geq 0.73$  和  $p = 0.001$ )。在 LW-NIR 区域基于  $R$  形成的 NPDI 与 As 含量相关性最高( $|r| = 0.74$ )。

(2)根据 VIP 值选择 NPDI 的,在通过  $R$ ,  $1/R$ ,  $\sqrt{R}$ ,  $\lg R$  计算的 NPDI 中,  $\text{NPDI}_{R(1.417/1.246)}$ ,  $\text{NPDI}_{1/R(799/953)}$ ,  $\text{NPDI}_{1/R(825/947)}$ ,  $\text{NPDI}_{\sqrt{R}(1.023/1.257)}$ ,  $\text{NPDI}_{\sqrt{R}(1.008/1.249)}$ ,  $\text{NPDI}_{\sqrt{R}(1.021/1.250)}$ ,  $\text{NPDI}_{\sqrt{R}(1.020/1.247)}$ ,  $\text{NPDI}_{\lg R(801/953)}$ ,  $\text{NPDI}_{\lg R(811/953)}$ ,  $\text{NPDI}_{\lg R(817/951)}$ ,  $\text{NPDI}_{\lg R(825/947)}$ ,  $\text{NPDI}_{\lg R(828/945)}$  的 VIP 值高于 1。因此,它们都被认为是合适的自变量。

(3)从 4 个预测模型的性能可以看出, Model-a( $R$ ) 相对于 Model-b( $1/R$ ) 和 Model-d( $\lg R$ ) 表现出优越的性能,它具有最高的验证系数( $R^2 = 0.831$ ,  $\text{RMSE} = 4.912 \mu\text{g} \cdot \text{g}^{-1}$ ,  $\text{RPD} = 2.321$ )和最低 AIC 值( $\text{AIC} = 179.96$ )。Model-c( $\sqrt{R}$ ) 也有较可靠的性能,更高的验证系数( $R^2 = 0.649$ ,  $\text{RMSE} = 7.141 \mu\text{g} \cdot \text{g}^{-1}$ ,  $\text{RPD} = 1.774$ )和较低的 AIC 值( $\text{AIC} = 197.16$ )。Model-d( $\lg R$ ) 和 Model-b( $1/R$ ) 的验证系数均较低。

优化光谱指数  $\text{NPDI}_{R(1.417/1.246)}$  可用于初步估算土壤重金属 As 的含量。优化光谱指数方法可以显著减少采样和分析过程中的时间和费用,光谱方法的潜在用途也表明需要根据金属结合反应理论进一步研究遥感光谱指数。由于数据采集的局限性,只涉及 SOM 和一种重金属,未考虑其他土壤性质的功能。其实,其他土壤特性(粘土和铁)也与重金属有密切的相互作用,因此需要进一步的研究。

- [13] Shi T, Liu H, Chen Y, et al. *Journal of Hazardous Materials*, 2016, 308: 243.
- [14] Chakraborty S, Li B, Deb S, et al. *Geoderma*, 2017, 296: 30.
- [15] Singh A N. *Pedosphere*, 2009, 19(6): 719.
- [16] LIU Wei, WANG Tao, WANG Jun, et al(刘巍, 王涛, 汪君, 等). *China Mining Magazine(中国矿业)*, 2017, 26(1): 60.
- [17] Ma Weibo, Tan K, Li Haidong, et al. *Journal of Ecology & Rural Environment*, 2016, 32(2): 213.
- [18] LI Ming-ze, GUO Hong-jun, FAN Wen-yi, et al(李明泽, 郭鸿郡, 范文义, 等). *Scientia Silvae Sinicae(林业科学)*, 2017, 53(6): 56.
- [19] Wold S, Trygg J, Berglund A, et al. *Chemometrics & Intelligent Laboratory Systems*, 2001, 58(2): 131.
- [20] JIANG Zhen-lan, YANG Yu-sheng, SHA Jin-ming(江振蓝, 杨玉盛, 沙晋明). *Acta Geographica Sinica(地理学报)*, 2017, 72(3): 533.
- [21] YANG Xiu-min, REN Guang-meng, LI Li-xin, et al(杨秀敏, 任广萌, 李立新, 等). *China Mining Magazine(中国矿业)*, 2017, 26(6): 79.
- [22] Nawar S, Buddenbaum H, Hill J. *Remote Sensing*, 2015, 7(2): 1181.
- [23] Stazi S R, Antonucci F, Pallottino F, et al. *Communications in Soil Science & Plant Analysis*, 2014, 45(22): 2911.
- [24] Sun W, Zhang X. *International Journal of Applied Earth Observation & Geoinformation*, 2017, 58: 126.

## Estimation of Heavy Metal Contents in Soil Around Open Pit Coal Mine Area Based on Optimized Spectral Index

Yasenjiang Kahaer<sup>1,2</sup>, Rukeya Sawut<sup>1,2</sup>, Nijat Kasim<sup>1,2</sup>, Nigara Tashpolat<sup>1,2\*</sup>, ZHANG Fei<sup>1,2\*</sup>, Abdugheni Abliz<sup>2,3</sup>, SHI Qing-dong<sup>1,2,3</sup>

1. College of Resources and Environmental Sciences, Xinjiang University, Urumqi 830046, China

2. Ministry of Education Key Laboratory of Oasis Ecology, Xinjiang University, Urumqi 830046, China

3. Institute of Arid Ecology and Environment, Xinjiang University, Urumqi 830046, China

**Abstract** Spectroscopy is regarded as a quick and nondestructive method to classify and analyze quantitatively many of elements of the soil. Visible and near-infrared reflectance spectroscopy offers a conducive tool for investigating soil heavy metal pollution. In this work, 51 soil samples with depths of 0~10 cm were collected, which were in the Eastern Junggar coal-field mining area, Xinjiang. The soil organic matter (SOM) content, Arsenic (As) content and indoor hyperspectra were measured in the laboratory. The significant relationship between As content and hyperspectral data was conducive analysis of NPDI, which were calculated from Vis-NIR region. For calculating the indices, on the basis of the raw spectral reflectance ( $R$ ), its three mathematical transformations were calculated, i. e., the reciprocal ( $1/R$ ), logarithm ( $\lg R$ ) and root mean square method ( $\sqrt{R}$ ), respectively. The two band combination of optimized indices software V1.0 (No: 2018R11S177501, independently developed based on the JAVA) was used during the calculation of the indices. NPDI were calculated using all possible combinations of available bands ( $i$  nm and  $j$  nm) in the full spectral region (400~2400 nm). In the optimal spectral indices ( $|r| \geq 0.73$  and  $p = 0.001$ ), an index of  $VIP \geq 1$  was further selected as a model independent variable by the Variable importance in projection (VIP) selection method. The main goal of this work is to obtain optimized spectral index (NPDI) related to soil heavy metal As, to estimate As concentration in soil based on geographically weighted regression (GWR) model, and to investigate the plausibility of using optimized spectral index for hyperspectral detection of heavy metal Arsenic in soil of coal mining areas. To assess the performance of the soil heavy metal contents prediction models, four cross-validation metrics were used; Residual Prediction Deviation (RPD), the Coefficient of Determination ( $R^2$ ), the Root Mean Square Error (RMSE) and Akaike Information Criterion (AIC). The results of this study are as follows: (1) As has the largest dispersion in the study area, SOM contents in all samples are less than 2%, and the As concentration has no significant correlation with the SOM content at a significance level of 0.01 ( $|r| = 0.113$ ). (2) Single-band reflectance shows low correlation with As contents, lower than 0.228. However, the highest correlation coefficient and lowest  $p$ -values ( $|r| \geq 0.73$  and  $p = 0.001$ ) between As and NPDI calculated by original and transformed reflectance ( $R$ ,  $1/R$ ,  $\lg R$ ,  $\sqrt{R}$ ) are found in the Near-infrared (NIR, 780~1100 nm) and Shortwave-infrared (SWIR, 1100~1935 nm) long wavelength infrared. The original spectral region formed with long wave length near-infrared (LW-NIR) regions show highest correlation with As contents ( $|r| = 0.74$ ). (3) VIP value of  $NPDI_{R(1417/1246)}$ ,  $NPDI_{1/R(799/953, 825/947)}$ ,  $NPDI_{\sqrt{R}(1023/1257, 1008/1249, 1021/1250, 1020/1247)}$  and  $NPDI_{\lg R(801/953, 811/953, 817/951, 825/947, 828/945)}$  higher than 1, thus these NPDI are chosen as independent variables. (4) From the four prediction model (GWR) performances it can be seen, the Model-a ( $R$ ) showed superior performance to other three models (Model-b ( $1/R$ ), Model-c ( $\sqrt{R}$ ) and Model-d ( $\lg R$ )), and it has the highest validation



coefficients ( $R^2=0.831$ ,  $RMSE=4.912 \mu\text{g} \cdot \text{g}^{-1}$ ,  $RPD=2.321$ ) and lowest AIC value ( $AIC=179.96$ ). The hyperspectral optimized index  $NPDI_{R(1.417/1.246)}$  may help to quickly and accurately evaluate Arsenic contents in soil, furthermore, the results provide theoretical and data support to access the distribution of heavy metal pollution in surface soil, promoting fast and efficient investigation of mining environment pollution and sustainable development of ecology.

**Keywords** Heavy metal; Optimized spectral indices; GWR model; Coal mine field

(Received Jun. 7, 2018; accepted Oct. 25, 2018)

\* Corresponding authors

欢迎投稿

欢迎订阅

欢迎刊登广告

## 《光谱学与光谱分析》2019年征订启事

国内邮发代码：82-68

国外发行代码：M905

《光谱学与光谱分析》1981年创刊，国内统一刊号：CN 11-2200/O4，国际标准刊号：ISSN 1000-0593，CODEN码：GYGFED，国内外公开发行，大16开本，332页，月刊；是中国科协主管，中国光学学会主办，钢铁研究总院、中国科学院物理研究所、北京大学、清华大学共同承办的学术性刊物。北京大学出版社出版，每期售价95元，全年1140元。刊登主要内容：激光光谱测量、红外、拉曼、紫外、可见光谱、发射光谱、吸收光谱、X射线荧光光谱、激光显微光谱、光谱化学分析、国内外光谱化学分析领域内的最新研究成果、开创性研究论文、学科发展前沿和最新进展、综合评述、研究简报、问题讨论、书刊评述。

《光谱学与光谱分析》适用于冶金、地质、机械、环境保护、国防、天文、医药、农林、化学化工、商检等各领域的科学研究单位、高等院校、制造厂家、从事光谱学与光谱分析的研究人员、高校有关专业的师生、管理干部。

《光谱学与光谱分析》为我国首批自然科学核心期刊，中国科协优秀科技期刊，中国科协择优支持基础性、高科技学术期刊，中国科技论文统计源刊，“中国科学引文数据库”，“中国物理文摘”，“中国学术期刊文摘”，同时被国内外的CSCD，SCI，AA，CA，Ei，AJ，MEDLINE，Scopus等文献机构收录。根据国家科技部信息研究所发布信息，中国科技期刊物理类影响因子及引文量《光谱学与光谱分析》都居前几位。欢迎国内外厂商在《光谱学与光谱分析》发布广告（广告经营许可证：京海市监广登字20170260号）。

《光谱学与光谱分析》的主编为高松院士。

欢迎新老客户到全国各地邮局订阅，若有漏订者可直接与《光谱学与光谱分析》期刊社联系。

联系地址：北京市海淀区学院南路76号（南院），

《光谱学与光谱分析》期刊社

邮政编码：100081

联系电话：010-62181070，62182998

电子信箱：chnghpxygpfx@vip.sina.com

修改稿专用邮箱：gp2008@vip.sina.com

网 址：<http://www.gpxygpfx.com>

