

鞍山式铁尾矿区土壤重金属含量高光谱反演

贾玉娜, 董元, 白洋, 刘暑明, 李孟倩

(华北理工大学, 河北 唐山 063210)

摘要: 土壤中的铜浓度占比过高会对人体和环境造成一定的危害, 因此探讨重金属铜含量反演具有重要意义。本研究以唐山典型铁尾矿区 43 个土壤样品为例, 同时测定土壤中铜的反射光谱和含量信息, 经多种光谱变换后, 通过相关性分析法 (CA) 和连续投影法 (SPA) 进行土壤铜含量的特征波段选取, 然后利用多元线性回归 (MLR) 和偏最小二乘回归 (PLSR) 算法建立了铜含量的反演模型, 得到多种光谱数据的反演成效。结果显示: 二阶微分处理后的光谱数据, 其反演效果较好, CA-PLSR 和 SPA-MLR 两种反演模型中, SPA-MLR 的反演精度相对较准确; 二阶微分光谱变换后的 SPA-MLR 模型在估算土壤铜含量方面更有优势。

关键词: 高光谱; 铜; 特征波段; 反演

doi:10.3969/j.issn.1000-6532.2023.02.032

中图分类号: TD951; P237 文献标志码: A 文章编号: 1000-6532 (2023) 02-0213-06

我国铁矿资源丰富, 铁尾矿成分各异, 铁矿山排出的尾矿成分差别很大, 排除少量金属组成成分外, 其他基本矿物组成涵盖绿泥石、方解石、石英、赤铁矿和黄铁矿等。河北省是尾矿库大省, 总数量在全国第一, 河北省尾矿库数量在 2400 座左右。唐山铁尾矿多属于高硅鞍山型铁尾矿, 其特点是 SiO_2 含量高达 70% 左右, 高硅鞍山式铁尾矿是全国数量最大的一类尾矿类型。

有关国家土壤的污染统计情况显示, 金属矿区及矿业废弃地周边土壤铅、砷等重金属污染严重^[1]。尾砂浆中含有多种重金属, 其在环境中不易降解并能在动植物体内富集, 这些金属一旦通过外排废液或者扬尘进入大气、水和周围土壤中, 就会导致严重的污染和危害。近年来, 许多研究者通过地球化学检测^[2]、电化学方法^[3]、激光诱导击穿光谱技术^[4]、电镜扫描技术^[5]等方法进行了重金属污染方面的研究, 并在特定的研究环境下得到了一定的效果。近年来, 高光谱以其高效、快捷、节约成本、非破坏性的新技术优势广泛用于土壤重金属含量测定中, 为更加精准、普适的进行土壤重金属研究提供了一条新路径^[6]。

“十二五”以来, 国家对受关注度极高的尾矿库重金属污染的治理给出了更明确的要求, 使其成为目前解决环境污染问题的一大热点。如何准确快速地获取尾矿中重金属污染分布情况是进行污染风险评价及预测研究的首要问题。唐山作为矿产资源丰富的矿业城市, 每年尾矿的产出量很大, 但关于唐山铁尾矿的重金属污染研究目前尚未报道, 部分学者只是针对鞍山式铁尾矿 Fe、Si 等化合物的含量研究^[7-8]。因此, 本文在对唐山典型尾矿区土壤进行 As Cr Cd Ni Zn Cu Pb Hg 8 种典型重金属元素测定后, 以《中国土壤元素背景值》为参照, 发现铜的含量高于土壤背景值, 可能存在铜元素重金属污染, 由此对铜元素进行含量与光谱的特征波段选择, 筛选出铜的特征波段, 利用偏最小二乘法和多元线性回归法建立土壤重金属反演模型, 以期高硅鞍山型铁尾矿土壤重金属污染提供理论支撑。

1 材料与方 法

1.1 研究区概况

本研究区域主要分布在司家营铁尾矿库及其

收稿日期: 2021-02-16

基金项目: 河北省教育厅青年基金 (QN2020149); 河北省省属高等学校基本科研项目 (JQN2019008); 华北理工大学博士研究生创新项目 (CXZZBS2020142); 华北理工大学青年基金 (ZZ01618)

作者简介: 贾玉娜 (1986-), 女, 讲师, 博士生, 主要研究方向为遥感与地理信息应用。

周边土壤地区，即位于唐山市滦县下康各庄。此尾矿库为正在使用中的尾矿库，库内尾砂浆主要由司家营铁矿场经高压泵输送至此，目前尾矿库的最高坝为 19 期坝，平均 3 年进行 2 期筑坝，本次采样点分布以主坝的坝顶为中心向四周发散取样，尾矿库内采样点分别选取坝顶、坝坡、坝底，尾矿坝周边土壤以尾矿库为中心选取东、西、南、北四条采样路线，对周边的村庄、农田、林地、梯田、果园等不同地类进行采样。

1.2 样本处理与理化分析

1.2.1 样本采集

司家营尾矿库坝及其周边为采样区，依据发散线布点法开展采样工作，于坝顶、坝坡、坝底以及尾矿库四周采集土壤样本。其中尾矿库共取样 4 个，沿东、西、南、北四条发散线采用 S 型取样 43 个，除尾矿坝样本外，其余均为表层 (0~30 cm) 的土样，采集量大约 1 kg 左右，装入黑色自封袋，通过手持 GPS 定位每个采样点，编辑相应编号贴于自封袋上，对每个采样点进行拍照，并记录周围的土地利用类型、土壤性状、植被覆盖情况等信息。每个采样点土壤样品分为 2 部分，其中，一部分用来进行化学元素测定，另一部分用来进行光谱研究。

1.2.2 样本理化成分测定

将测量光谱后的土壤样本使用分析天平依次准确称取 4 mg 样品，样品转移至聚四氟乙烯消解管中，向样品消解管中滴入 3 mL 硝酸、3 mL 盐酸、0.5 mL 氢氟酸，经电加热板进行加热消解。消解时需要在消解液小于 2 mL 时按 1:1 的比例补加硝酸和盐酸。消解完成后，将消解管放入通风橱内进行自然冷却至室温，将消解好的样品过滤并使用去离子水稀释定容至 10 mL 的容量瓶中混合均匀，每批样品加做双份全程试剂空白对照实验，采用电感耦合等离子体质谱仪 (ICP-MS) 测定重金属铜元素含量，其中每个样本平行测量两次，选取平均值作为最终的含量。

1.3 土壤光谱处理

在实验室内使用 Field Spec 光谱仪在 350~2500 nm 光谱范围内进行土壤光谱信息的获取。具体操作为：将野外采集的土壤样本剔除其中的根系、枝叶、石块等，经风干处理，于 85 °C 烘箱内烘干 3 h，用球磨机磨细并过 0.15 mm 筛子。将土

样放置在 5×8×3 cm 的黑色纸盒内，探头垂直对准被测土壤的中部，光源用光谱仪自带光源。为了准确测量样本，需先进行白板定标，再经按压土壤表面后开始测量，对各样本重复测量 5 次后，求其算数平均值为该样本的实测光谱数据。

为了削弱测定过程中仪器、光照等条件的影响，对原始光谱曲线进行断点修正和 Savitzky-Golay 平滑处理，同时为了凸显土壤光谱反射率特征，提高反演精度，对土壤光谱反射率进行了光谱一阶微分 (FD)、二阶微分 (SD) 和光谱连续去除光谱变换 (CR)，增强有效波谱信息。

1.4 特征波段选择方法

1.4.1 相关性分析法

相关性分析法 (Correlation Analysis, CA) 是研究两个或两个以上处于同等地位的随机变量间的相关关系的统计分析方法。相关系数是衡量变量之间的相关程度，一般用字母 r 表示，用来度量两个变量间的线性关系。

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

其中， $Cov(X, Y)$ 为 X 与 Y 的协方差， $Var[X]$ 为 X 的方差， $Var[Y]$ 为 Y 的方差。经多次实验后选择相关性最高的前 70 个波段进行反演效果较佳。

1.4.2 连续投影法

连续投影法 (Successive Projections Algorithm, SPA) 是由 Araujo 等提出的一种消除变量间共线性的波段选择算法，该方法是提取全部光谱的一些特征波段来消除原始光谱中的冗余信息。其主要原理算法如下。

将光谱矩阵 $X^{n \times p}$ 进行分组，分为 p 组，设为集合 $sl = [s_1, s_2, \dots, s_m] \in R^{p \times m}$ 。每组选择 m 个波长 [$m \leq M = \min(n, p)$]。令 $i = 1, k = 1, 2, \dots, p, z_i = x_k; s_1^k = x_k; sl(k, 1) = k, u = 1, 2, \dots, m$ 。基于 z_i 构建正交投影算子。其中 I 为 $n \times n$ 的单位矩阵；

$$P_i = I - \frac{z_i(z_i)^T}{(z_i)^T z_i} \quad (2)$$

将剩余还未被选入的波长矢量的集合记作 v ，即为 $v \in [1, p] \&\& v \notin sl; s_v^k = p_i x_v$ 。计算各 s_v^k 的正交投影矢量，并得到波长位置，即

$$sl(k, u) = \arg \max_{v \in [1, p] \&\& v \notin sl} \|s_v^k\|; z_i = x_{sl(u)} \quad (3)$$

令 $i = i + 1$ ，若 $i < m$ ，重新建立正交投影矢量选

择下一波长，最后得到光谱矩阵 $sl = [s_1, s_2, \dots, s_m] \in R^{p \times m}$ ，较优波长的确定由多元定量校正模型决定。

1.5 模型评价

根据土壤重金属铜含量与光谱反射率两者的关联度，利用2种方法筛选出重金属铜的光谱特征波段，利用偏最小二乘法和多元线性回归法，开展铜含量与特征波段光谱信息的回归分析，将43个样本按照33个用来建模、10个用来验证的分组方式建立回归模型并对模型进行检验。本文以土壤重金属铜含量参数与不同光谱变换展开回归分析，并以光谱变换的特征波段的反射率为自变量 X 、土壤铜含量为因变量 Y 进行拟合，模型的效果好坏以判定系数 R^2 、均方根误差 $RMSE$ 指标为依据。 R^2 反映模型的拟合优化程度， R_C 、 R_P 表示建模样本和验证样本的 R^2 ， $RMSE$ 表示模型预算结果的精密度， $RMSEC$ 、 $RMSEP$ 表示建模样本和验证样本的 $RMSE$ 。 RPD 表示实测样本重金属含量和预测值之间的相关性， R^2 越高， $RMSE$ 值越小，预测模型就越准确。通过预测样本验证回归模型，得到铜的较佳反演模型及其相关参数。

$$R^2 = 1 - \frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2} \quad (4)$$

$$RMSE = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2} \quad (5)$$

$$RPD = \frac{SDs}{RMSEP} \quad (6)$$

式中， n 是样本容量为土壤样品的重金属铜含量的实测值； Y_i 是土壤铜含量的估值； \hat{Y}_i 为土壤铜含量的实测平均值； SDs 为土壤实测值标准差； $RMSEP$ 为模型预测均方根误差，单位为 $mg \cdot kg^{-1}$ 。

当 $RPD > 2.5$ 时，模型具有超高的估算准确度；当 $2 < RPD < 2.5$ 时，模型具有高度的反演准确性；当 $1.8 < RPD < 2$ 时，模型具有较高程度的反演准确性；当 $1.4 < RPD < 1.8$ 时，模型具有一般高度的反演准确性，但仍在可以接受的范围内；当 $1.0 < RPD < 1.4$ 时，模型具有较低高度的反演准确性；当 $RPD < 1.0$ 时，模型的反演精度极低已经不能满足研究需求^[9]。综合模型情况来看， R^2 和 RPD 值最大、 $RMSE$ 值最小时为较优估算模型。

2 结果与分析

2.1 土壤样本重金属含量分析

本次采集的尾矿库及周边43个土壤样本铜含量的结果见表1。

表1 铁尾矿区尾砂土壤重金属铜含量分析/ ($mg \cdot kg^{-1}$)
Table 1 Analysis of heavy metal copper content in tailing soil of iron tailings area

样本	均值	最大值	最小值	标准差	变异系数
Cu	11.80	25.05	4.49	4.07	34%

在土壤的相关研究中，变异系数可作为反映土壤样本变异程度的等级划定指标，如表2所示，由变异系数来看，铜元素处于16%~35%的中等变异区间；如果变异系数太大的话，说明有差别大的样本值，会对后面的模型建立的精度产生影响，这时候就要根据实际情况进行剔除，本次采集样本不需要剔除。

表2 土壤变异系数等级分类
Table 2 Classification of soil coefficient of variation

系数区间	等级
0~15%	小
16%~35%	中
>36%	高

2.2 土壤实测光谱曲线分析

矿区土壤样本(43个)在可见光-近红外波段区间的波谱曲线见图1。各土壤样品的光谱曲线形态特征呈现相同的特点，但因土壤组分存在差异，致使土壤光谱反射率高低不一。在可见光波段，土壤反射率较低，但增速较大，在760 nm之后近红外波段，反射率变化趋势总体较为平稳。在1400、1900和2200 nm处出现明显的吸收波

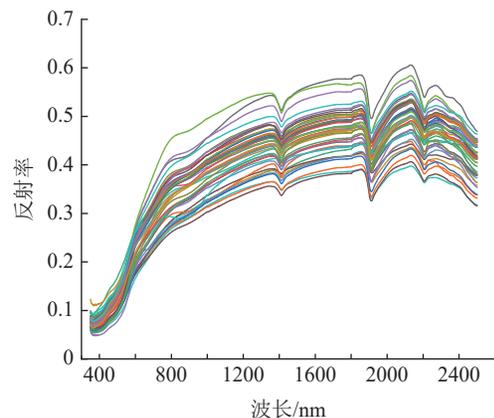


图1 土壤样本反射率光谱曲线
Fig.1 Reflectance spectrum curve of soil samples

谷，主要是土壤粘土矿物和硅酸盐类矿物中所含的水分子所致。

2.3 土壤重金属含量反演分析

偏最小二乘回归 (partial least squares regression, PLSR) 建模方法，通过考虑光谱数据与性质之间的内在联系，模型更加稳健，但模型建立过程复杂、抽象，计算速度较慢且繁琐。当变量之间相关性非常高时，PLSR 效果极佳。而多元线性回归 (multiple linear regression, MLR) 常用于线性较好的数据，产生的模型比 PLSR 更简单，更容易解释。故本文基于相关性分析法所选取的特征波段数量较多、相关性较好，通过建立 PLSR 模型分析其预测能力，基于连续投影法得到的特征波长线性较好，通过构建 MLR 模型分析其预测能力。

2.3.1 CA-PLSR 反演

根据 CA 筛选的敏感特征波段，经偏最小二乘回归分析建立土壤铜含量的反演模型，其建模和

验证结果见表 3，其中 SD 建模的 R_C 达到了 0.94，RMSEC 为 0.75 mg/kg，其验证模型的 R_p 、RMSEP 和 RPD 值分别为 0.68、3.41 mg/kg 和 1.78，此模型拟合的稳定性相比较适合。其他模型中 R 、FD 和 CR 的验证误差均大于 2，RPD 值分别为 1.07、1.22、1.07，这 3 种模型精度不够理想。

表 3 土壤铜含量 CA-PLSR 模型结果

反射率类型	R_C	RMSEC	R_p	RMSEP	RPD
R	0.17	2.94	0.13	5.66	1.07
FD	0.61	2.02	0.33	4.98	1.22
SD	0.94	0.75	0.68	3.41	1.78
CR	0.45	2.39	0.13	5.65	1.07

图 2 为土壤铜含量 CA-PLSR 模型的散点，所有模型验证数据都散落在 1:1 线周边区域，说明其估算准确率并不高。而从模型的拟合程度来看，二阶微分 SD 的拟合程度最好，相比其他三种可以利用此模型来反映土壤铜含量。

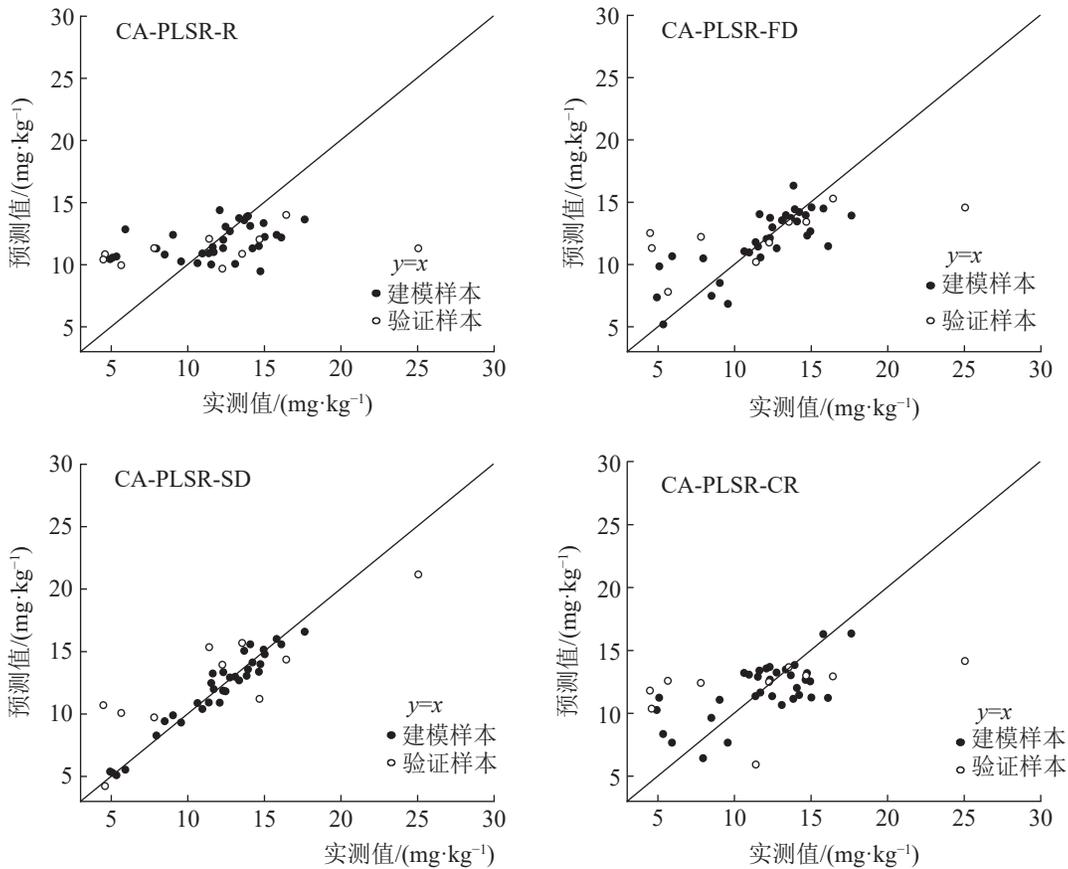


图 2 CA-PLSR 模型实测值预测值对比

Fig.2 Comparison of measured and predicted values of CA-PLSR model

2.3.2 SPA-MLR 反演结果

根据 SPA 选好的特征波段，经多元线性回归

分析后得到土壤铜含量反演模型，其建模和验证结果如表 4 所示。其中二阶微分 SD 模型的 R_C 达

到了 0.99, RMSEC 为 0.56 mg/kg, 其模型验证的 R_p 、RMSEP 和 RPD 值分别为 0.85、2.30 mg/kg 和 2.63。R、FD 两种模型的误差较大, 且 RPD 小于 2, 预测精度不理想, CR 不及 SD 的估算准确, 故此可利用二阶微分 SD 模型进行土壤铜含量估算。

图 3 为 SPA-MLR 模型的散点图, 图 3 中显示 SPA-MLR 模型的拟合程度均紧凑的分布在 1:1 线上, 因此, SPA-MLR 土壤铜含量估算模型

表 4 土壤铜含量 SPA-MLR 模型结果
Table 4 Results of SPA-MLR model for soil copper content

反射率类型	R_C	RMSEC	R_p	RMSEP	RPD
R	0.09	3.18	0.01	5.98	1.01
FD	0.98	1.04	0.68	3.32	1.83
SD	0.99	0.56	0.85	2.30	2.63
CR	0.99	0.93	0.75	2.93	2.07

精度要优于 CA-PLSR 构建的估算模型精度。

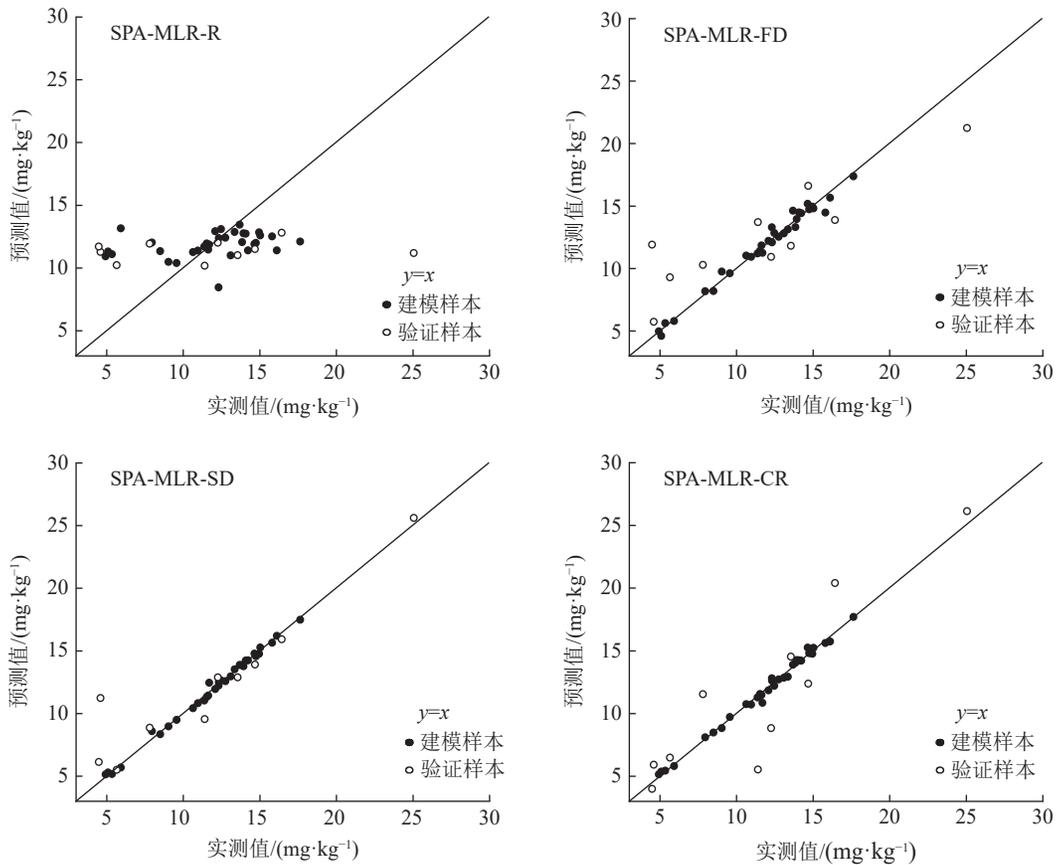


图 3 SPA-MLR 模型实测值预测值对比
Fig.3 Comparison of measured and predicted values of PA-MLR model

3 结论

(1) 分析不同光谱变换方法对土壤铜含量估算精度的影响, 其中二阶微分光谱变换优于其他方法, 为土壤铜含量的估算的准确度。

(2) SPA-MLR 模型估算效果优于 CA-PLSR 模型, 其较佳 SPA-MLR 模型校正和验证的 R_C 、 R_p 分别达到了 0.99 和 0.85 远大于 CA-PLSR 全部模型, 其校正和验证的 RMSEC、RMSEP 值分别为 0.56 mg/kg 和 2.30 mg/kg 优于 CA-PLSR 较佳模型的 0.75 mg/kg 和 3.41 mg/kg, 以及 RPD 值为 2.63 优于 CA-PLSR 较佳模型的 1.78, 得出 SPA-MLR

较佳模型可以更好的反演土壤铜含量。

参考文献:

[1] 宋文, 成少平, 迟晓杰, 等. 重金属污染土壤修复遥感监测研究进展[J]. 矿产综合利用, 2021(4):21-28.
SONG W, CHENG S P, CHI X J, et al. Research progress on remediation of heavy metal contaminated soil monitored by remote sensing[J]. Multipurpose Utilization of Mineral Resources, 2021(4):21-28.

[2] 杨帆, 郝志红, 张舜尧, 等. 长江武汉段冲积土壤中重金属的环境地球化学特征[J]. 地质学报, 2016, 90(8):1955-1964.
YANG F, HAO Z H, ZHANG S Y, et al. Environmental

geochemistry of heavy metals in alluvial soils of Yangtze River in Wuhan City[J]. *Acta Geologica Sinica*, 2016, 90(8):1955-1964.

[3] 黄善洛, 金妍, 郭慧, 等. 基于 Android 平台的便携式痕量重金属离子检测仪的研制[J]. *分析化学*, 2015, 43(7):1098-1103.

HUANG S L, JIN Y, GUO H, et al. Development of portable trace heavy metal ion detector based on Android platform[J]. *Chinese Journal of Analytical Chemistry*, 2015, 43(7):1098-1103.

[4] 余克强, 赵艳茹, 刘飞, 等. 激光诱导击穿光谱技术检测土壤中的铅和镉含量[J]. *农业工程学报*, 2016, 32(15):197-203.

YU K Q, ZHAO Y R, LIU F, et al. Determination of lead and cadmium in soil by laser induced breakdown spectroscopy[J]. *Journal of Agricultural Engineering*, 2016, 32(15):197-203.

[5] 徐国栋, 葛建华, 杜谷, 等. 成都市中心城区地表沉积物中重金属分布及矿物学特征[J]. *岩矿测试*, 2019, 38(4):418-428.

XU G D, GE J H, DU G, et al. The heavy metal distribution and mineralogical characteristics of surface sediments in Chengdu[J]. *Rock and Mineral Analysis*, 2019, 38(4):418-428.

[6] 汪金花, 曹兰杰, 白洋, 等. 铁尾矿粒径和湿度因子对高光

谱特征参量影响[J]. *矿产综合利用*, 2019(2):128-133.

WANG J H, CAO L J, BAI Y, et al. Influence of iron tailings' particle size and humidity factor on hyperspectral characteristic parameters[J]. *Multipurpose Utilization of Mineral Resources*, 2019(2):128-133.

[7] 杨晓峰, 苏兴强, 张廷东. 鞍山铁尾矿特性及综合利用前景[J]. *矿业工程*, 2008(5):47-49.

YANG X F, SU X Q, ZHANG T D. Features of Anshan iron ore tailings and future of their comprehensive utilization[J]. *Modern Mining*, 2008(5):47-49.

[8] ROSSEL R A V, MCGLYNN R N, MCBRATNEY A B. Determining the composition of mineral-organic mixes using UV-vis-NIR diffuse reflectance spectroscopy[J]. *Geoderma*, 2006, 137:70-82.

[9] Eunyoung Choe, Freek van der Meer, Frank van Ruitenbeek, Harald van der Werff, Boudewijn de Smeth, Kyoung-Woong Kim. Mapping of heavy metal pollution in stream sediments using combined geochemistry, field spectroscopy, and hyperspectral remote sensing: A case study of the Rodalquilar mining area, SE Spain[J]. *Remote Sensing of Environment*, 2008, 112(7):3222-3233.

Hyperspectral Inversion of Soil Heavy Metal Content in Anshan-style Iron Tailings Area

Jia Yuna, Dong Yuan, Bai Yang, Liu Shuming, Li Mengqian

(North China University of Science and Technology, Tangshan, Hebei, China)

Abstract: Excessive copper concentration in the soil will cause certain harm to human body and the environment, so it is of great significance to explore the inversion of heavy metal copper content. In this study, 43 soil samples in the typical iron tailings area of Tangshan were taken as examples, and the reflectance spectrum and content information of copper in the soil were measured at the same time. After a variety of spectral transformations, the correlation analysis method (CA) and the continuous projection method (SPA) were carried out. The characteristic wavebands of soil copper content were selected, and then the inversion model of copper content was established using multiple linear regression (MLR) and partial least square regression (PLSR) algorithms, and the inversion results of various spectral data were obtained. The results show that the spectrum data after the second-order differential processing has the best inversion effect. Among the two inversion models of CA-PLSR and SPA-MLR, the inversion accuracy of SPA-MLR is relatively accurate; after the second-order differential spectrum transformation te SPA-MLR model has more advantages in estimating soil copper content.

Keywords: Hyperspectral; Copper; Characteristic band; Inversion