

土壤重金属砷的高光谱估算模型

李志远^{1,2}, 邓帆^{1*}, 贺军亮², 魏薇¹

1. 长江大学地球科学学院, 湖北 武汉 430100

2. 石家庄学院资源与环境科学学院, 河北 石家庄 050035

摘要 砷是严重危害人体的重金属之一。利用高光谱技术进行土壤重金属砷含量的估测具有很大的应用潜力, 但受区域和土壤背景的影响, 估算模型适用性和精度都会有很大的差异。针对石家庄市地表水源地保护区土壤砷含量的高光谱估算, 在水源地保护区的主要采矿地和冶炼企业进行了土壤实地采样和实验室重金属分析, 对土壤样本的原始光谱反射率采用 Savitzky-Golay 7 点平滑处理, 进行一阶微分(FD)、二阶微分(SD)、倒数(RT)、倒数一阶微分(RTFD)、倒数二阶微分(RTSD)、倒数对数(AT)、倒数对数一阶微分(ATFD)、倒数对数二阶微分(ATSD)、连续统去除(CR)9种光谱变换后, 再对重金属砷实测含量与经光谱变换后的光谱指标进行相关分析, 并提取各光谱指标的最大敏感波段。运用多元线性逐步回归(MLSR)、单光谱变换指标偏最小二乘回归(U-PLSR)和多光谱变换指标偏最小二乘回归(M-PLSR)方法构建土壤重金属砷含量估算模型, 最后通过相关系数 r 、均方根误差(root mean square error, RMSE)和统计值 F 来对比建模效果。结果表明: 研究区部分土壤样本重金属砷含量已经出现了轻度污染, 大部分样本处于污染的临界状态; 经连续统去除变换后的光谱特征与砷的相关性最大, 一阶微分与砷含量存在最大负相关性; 相较于多元线性逐步回归和单光谱变换指标偏最小二乘回归, 采取多光谱变换指标偏最小二乘回归方法土壤重金属砷含量模型估算值与实测值最为接近, 建模 R^2 达到0.852, RMSE和 F 值分别达到0.147和32.384, 多光谱变换指标建模集成效果显著。因此研究结果可以为石家庄水源地保护区主要采矿地和冶炼企业重金属砷污染高光谱快速监测提供科学依据。

关键词 重金属砷; 高光谱; 多元线性逐步回归; 偏最小二乘回归

中图分类号: TP79 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)09-2872-07

引言

土壤重金属监测在农业生产中至关重要, 尤其工业和农业密集地区存在重金属扩散到食物链的风险^[1]。砷(As)是人体所需微量元素之一, 但若吸入出现超标, 会对人体健康造成极大危害。含砷三废的排放和农药化肥的过度使用人为加剧了重金属砷污染^[2]。

传统的土壤特性检测方法能够准确测得重金属含量, 但是花费高、步骤繁琐、耗时长, 无法做到实时高效^[3]。高光谱技术具有成本低、效率高、用时短等特点, 被广泛用于重金属含量估算研究中^[4]。目前, 用于土壤重金属As高光谱反演的线性建模方法主要有单变量回归、多元线性逐步回

归、主成分回归、偏最小二乘回归、地理加权回归等, 这些方法在对应的研究区域取得了较好的效果。

多元线性逐步回归(multiple linear stepwise regression, MLSR)和偏最小二乘回归(partial least squares regression, PLSR)在估算重金属As含量中运用非常普遍, 它尽可能选出对重金属As含量影响较大的自变量, 但是容易出现过度拟合情况, 导致模型稳定性降低^[5]; 偏最小二乘回归综合了主成分回归和多元线性逐步回归方法, 能够在波段个数较多且自相关严重时建立回归模型, 避免共线性问题^[5]。

以采矿地和冶炼企业较多的石家庄市地表水源地保护区褐土为研究对象, 主要根据重金属As含量与光谱指标进行相关分析选出特征波段, 运用多元线性逐步回归和偏最小二乘回归方法来估算土壤重金属As含量, 找出反演重金属As

收稿日期: 2020-06-11, 修订日期: 2020-10-25

基金项目: 国家自然科学基金青年科学基金项目(41201215), 国家重点研发项目(2017YFB0503903-04), 河北省自然科学基金项目(D2016106013), 河北省高等学校科学技术研究项目(Z2019009), 石家庄市高等教育科学研究项目(20191004)资助

作者简介: 李志远, 1996年生, 长江大学地球科学学院硕士研究生 e-mail: oliverblues@outlook.com

* 通讯作者 e-mail: dengfan@yangtzeu.edu.cn

的优选方法,对研究该区域土壤 As 污染的快速监测提供借鉴。

1 实验部分

1.1 研究区概况

研究区位于石家庄市滹沱河中游的平山县和井陘县境内,主体(一级保护区)包括滹沱河干流、黄壁庄水库和岗南水库及其支流,主体外设有二级和三级保护区(如图1)。总体呈西北向东南逐渐降低,地势差距较大。由于样本采集区域开发较早,人地矛盾突出,矿产的开采造成了土壤重金属污染问题,对研究区现有生态安全造成一定的负面影响^[6]。

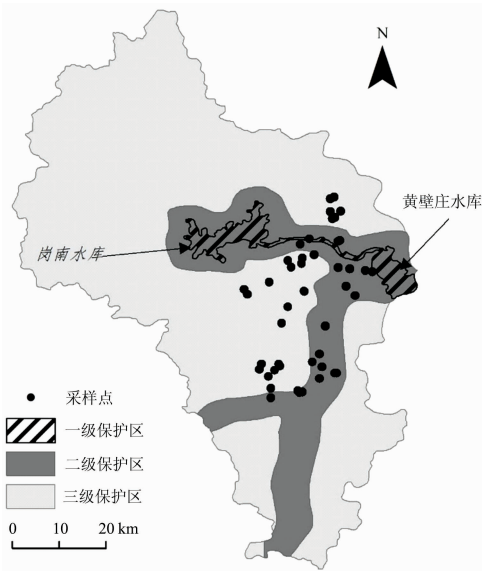


图1 研究区和土壤采样点分布图
Fig. 1 Study area and the distribution of the soil sampling points

1.2 样本采集

基于研究区地形和土壤空间分布不均匀的特性,依据研究区内主要采矿点和冶炼企业的分布,按照地表径流路径设置三条样带,采集土壤样本48个,采集深度为0~20 cm,同时使用GPS定位,采样点位置如图1所示。采集的所有土壤样本带回室内,经风干、研磨、过100目筛,变成粉状土,然后将每份样本一分为二,分别进行土壤基本理化性质和高光谱数据的测定。采用微波消解仪(CEM-Mars)消解,利用电感耦合等离子体质谱仪(PE-Nexion 300 ICP-MS)测定土壤中的砷含量。表1为48个土壤样本中砷含量的描述性统计值。

研究区土壤中重金属 As 含量分布在 10.844~15.125 mg·kg⁻¹之间,平均值为 13.230 mg·kg⁻¹,接近河北省背景值^[7],说明采集的部分土壤样本中重金属 As 含量已经富集到了一定程度。土壤重金属污染程度采用单因子评价法^[8]来分析,计算公式为

$$P_i = C_i/S_i \quad (1)$$

其中 P_i 表示土壤中重金属 As 的污染指数, C_i 表示土壤中

重金属 As 的实测值, S_i 表示土壤中重金属 As 的河北省背景值。 $P_i \leq 0.7$ 为安全范围, $0.7 < P_i \leq 1$ 为警戒范围, $1 < P_i \leq 2$ 为轻度污染范围。

表1 土壤中砷含量的描述性统计值(mg·kg⁻¹)

重金属	最小值	平均值	最大值	标准差	河北省背景值
As	10.844	13.230	15.125	1.005	13.6

表2 土壤中砷的单因子污染指数统计结果

元素	最小值	平均值	最大值	样点数	
				0.7 < P _i ≤ 1	1 < P _i ≤ 2
As/(mg·kg ⁻¹)	0.797	0.973	1.112	29	19

由表2统计得到,污染指数的最小值为0.797,平均值为0.973,最大值为1.112。有29个样点处于警戒范围;有19个样点受到重金属 As 的轻度污染^[9]。综合看来,约40%土壤样本受到重金属 As 的轻度污染,受到了采矿点和冶炼企业的影响,有必要对重金属 As 含量进行长期动态监测。

1.3 光谱测定

采用 FieldSpec Pro 便携式光谱仪测定土壤光谱反射率,波段范围为 350~2 500 nm,采样间隔为 1.4 nm(350~1 000 nm)和 2 nm(1 000~2 500 nm),重采样间隔为 1 nm。光谱测量在可调节光照的暗室内进行,为尽可能减少外来光线的影响,选用卤素灯作为唯一光源。将处理好的土壤样本放入玻璃培养皿(直径 10 cm、深 2 cm)内,置于黑色绒布上,用工具将土壤表面刮平。光源天顶角为 45°,距土样表面 30 cm,探头置于距土壤样本表面垂直上方的 15 cm 处,光纤探头视场角为 5°。每次测试前进行白板标定,每个土壤样本测试 10 次光谱曲线,取算数平均值作为该样本实际光谱数据。

1.4 光谱数据预处理

为消除仪器信噪比、样本风干和杂散光等因素的影响,采用 Savitzky-Golay 7 点平滑,去除噪声较大的异常波段,保留 1 640 个波段用于后续建模研究。为了提高光谱信噪比,去除背景噪声,突出光谱特征,通过原始光谱反射率(SR)分别计算一阶微分(FD)、二阶微分(SD)、倒数(RT)、倒数一阶微分(RTFD)、倒数二阶微分(RTSD)、倒数对数(AT)、倒数对数一阶微分(ATFD)、倒数对数二阶微分(ATSD)、连续统去除(CR)。将重金属实测值与以上经光谱变换后的数据进行相关性检验,得到相关性最高的波段。光谱变换由 Origin2020b 和 ENVI5.1 软件进行,结果见图2。相关性分析及后续模型建立在 Minitab18 软件中完成。

1.5 多变量反演方法

(1) MLSR

多元线性逐步回归是通过一个因变量和多个自变量组合表达它们之间的线性关系。回归方程为

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon \quad (2)$$

其中, y 表示因变量, x 表示自变量, $\beta_i (i=0, 1, 2, \dots, k)$ 表示回归系数, ϵ 表示随机误差, k 表示自变量的数量^[10]。将土壤重金属 As 实测含量作为建模样本的因变量, 将光谱指标作为自变量, 根据贡献率高低来建立 MLSR 模型。

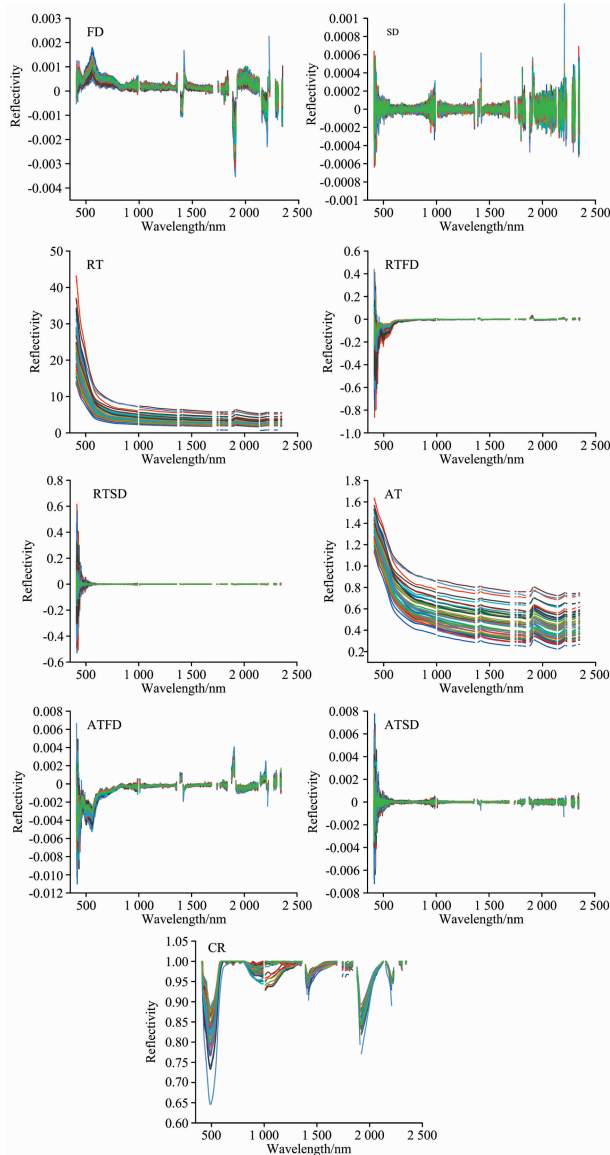


图 2 光谱变换后的土壤样本光谱反射率曲线

Fig. 2 Transformed soil reflectance spectra after

(2) PLSR

偏最小二乘回归是一种结合主成分回归和 MLSR 的分析方法, 由 Herman Wold 于 1966 年提出, 目前已被广泛应用于光谱数据处理。

PLSR 的主要研究内容是在大量的两组高线性相关变量的情况下, 建立自变量的线性模型, 以解决样本数小于变量数的问题, 避免出现过拟合, 因此 PLSR 是对 MLSR 的改进。PLSR 的原理是: 首先从自变量 (x_1, x_2, \dots, x_m) 中提取相互独立的分量 $T_h (h=1, 2, \dots)$, 并且提取的主成分携带尽可能多的原始成分; 然后从因变量 (y_1, y_2, \dots, y_m) 中提

取独立分量 $U_h (h=1, 2, \dots)$, 要求 T_h 和 U_h 之间的协方差最大化, 并利用 MLSR 得出的因变量建立回归模型。偏最小二乘回归的基本模型是

$$X = T_h P^T + E \quad (3)$$

$$Y = U_h Q^T + F \quad (4)$$

其中, P 和 Q 分别是 $m \times h$ 的正交载荷矩阵, E 和 F 是误差项(服从正态分布的随机变量)^[11]。以土壤重金属 As 实测含量为因变量, 选用单一光谱指标的特征波段作为自变量建立单光谱变换指标偏最小二乘回归 (univariate partial least squares regression, U-PLSR) 模型, 用 MLSR 筛选出的多种光谱指标作为自变量建立多光谱变换指标偏最小二乘回归 (multivariate partial least squares regression, M-PLSR) 模型。

1.6 模型建立与检验

随机将 48 个样本分为两部分, 32 个用作建立模型, 16 个用来验证模型。采用相关系数 r 、均方根误差 (root mean square error, RMSE) 和统计值 F 来衡量建模效果。 F 值和 r 越大, RMSE 越小, 建模样本和验证样本的散点分布越靠近 1:1 线, 建模效果越好, 由此确定最佳回归模型。

2 结果与讨论

2.1 研究区土壤光谱特性

原始光谱曲线经过平滑之后得到的曲线如图 3 所示, 48 个土壤样本的光谱曲线形态基本相同。在可见光波段, 光谱反射率随波长增长而增大, 在近红外波段, 整体趋于平缓。光谱曲线在 600 和 800 nm 附近有有机质的反射峰; 1 400, 1 900 和 2 200 nm 波段附近存在三处明显的水分吸收带, 这是受到土壤中硅酸盐矿物和粘土矿物有关^[12]。FD, SD 和 ATFD 变换光谱曲线起伏较大; RT 和 AT 变换呈现从起始波段迅速下降, 到 600 nm 附近降幅变缓趋势; RTFD, RTSD 和 ATSD 变换后反射率变化主要出现在 500 nm 之前; 经 CR 变换, 光谱曲线归一化在 0~1 之间, 吸收谷出现在 500, 1 010, 1 400, 1 900 和 2 200 nm 附近(图 3)。

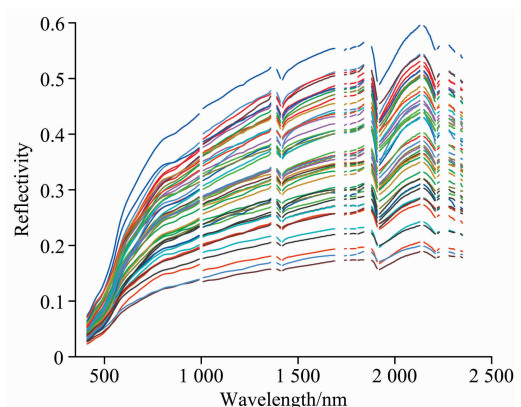


图 3 平滑后的土壤样本光谱反射率曲线

Fig. 3 Smoothed reflectance spectra of soil samples

2.2 特征波段提取

为了探究不同光谱指标与重金属砷含量的关系, 将 As

含量与各种光谱变换进行 Pearson 相关性分析, 表 3 表示最大相关系数及其对应的敏感波段。

表 3 土壤重金属砷含量与光谱指标的最大相关系数

Table 3 Maximum correlation coefficients of heavy metal As content and spectral indexes

光谱指标	最大相关波段/nm	相关系数
SR	2 342	-0.289*
FD	1 211	-0.527**
SD	2 208	-0.470**
RT	1 740	0.357*
RTFD	1 359	-0.249
RTSD	1 065	0.485**
AT	1 740	0.353*
ATFD	464	-0.370**
ATSD	2 208	0.372**
CR	682	0.603**

注: ** 表示通过了 0.01(双尾)显著相关检验, * 表示通过了 0.05(双尾)显著相关检验

Note: ** significant at 0.01 level (2-tailed) and * at 0.05 level (2-tailed)

可以发现, 经 CR 光谱变换后 682 nm 与重金属 As 含量的相关系数出现明显峰值, 达到 0.603(如图 4), 这可能与重金属 As 受到土壤有机质的吸附有关。FD 的 1 211 nm 与土壤重金属 As 含量存在最大负相关性。其中通过 0.01(双尾)显著性检验的波段还有 SD 和 ATSD 的 2 208 nm、RTSD 的 1 065 nm、ATFD 的 464 nm, SR 的 2 342 nm, 相关系数均超过 0.370, 表示相关性显著, 而 RT 和 AT 的 1 740 nm 相关系数大于 0.28, 通过了 0.05(双尾)显著性检验, 表示相关性比较显著。RTFD 未能通过任何相关性检验, 因此在最后建立模型时不予考虑。

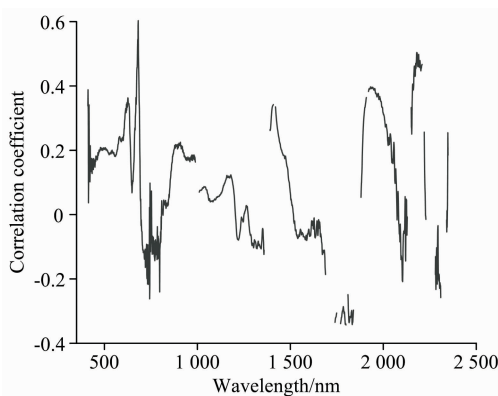


图 4 连续统去除相关系数曲线图

Fig. 4 Continuum removal correlation coefficient curve

2.3 MLSR

MLSR 是筛选对因变量影响较大的自变量建立回归模型的方法。根据自变量对因变量的作用大小, 剔除不显著的自变量, 由大到小逐个引入回归方程, 由于考虑了多个变量因素, 建模结果具有一定的可靠性。

根据土壤重金属砷含量与光谱指标的相关分析, 将表 3 中各光谱指标对应的敏感波段作为 As 含量估算模型的自变量, 以 As 实测值为因变量, 建立 MLSR 模型。

通过建模分析, 将 FD_{1211} , $RTSD_{1065}$, $ATFD_{464}$ 和 CR_{682} 4 个对模型拟合度影响较大的光谱指标保留。具体模型为

$$Y = -638.304 - 6\,629.954X_{FD} + 1\,110.784X_{RTSD} - 391.284X_{ATFD} + 651.907X_{CR} \quad (5)$$

MLSR 模型的散点图和验证结果如表 4 和图 5 所示, 建模样本 r 超过 0.8, 基本分布在 1:1 线两侧, 说明该模型对建模数据具有良好的解释能力。而验证样本数据分布相较于建模样本距离 1:1 线存在较大偏离, 说明模型预测效果不如建模效果, 模型稳定度需要进一步探究。

表 4 MLSR 模型结果

Table 4 Results of MLSR model

参数	建模样本	验证样本
r	0.814	0.776
RMSE	0.159	0.318
F	29.437	8.482

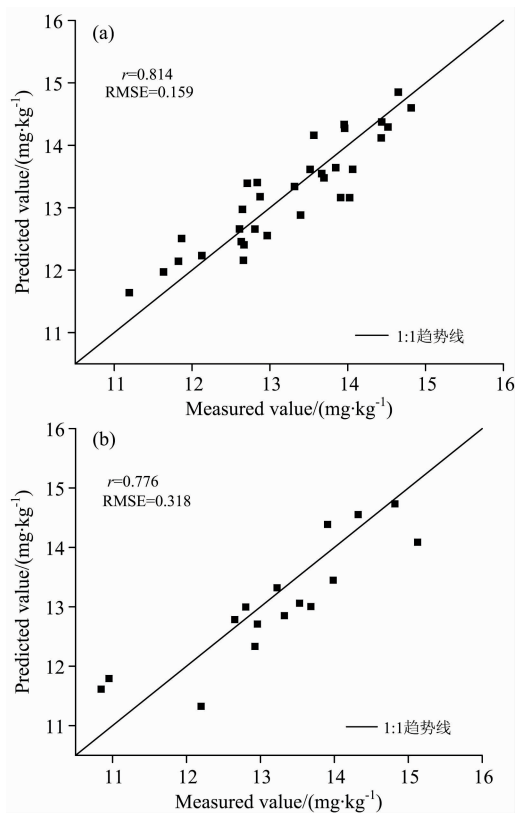


图 5 MLSR 模型建模样本和验证样本散点图

(a): 建模样本; (b): 验证样本

Fig. 5 Scatter plots of fitting and testing sets for MLSR

(a): Fitting set; (b): Testing set

2.4 U-PLSR

依照 2.3 分析, 将筛选出来模型精度影响较大的光谱指标(FD_{1211} , $RTSD_{1065}$, $ATFD_{464}$ 和 CR_{682}) 分别作为自变量,

构建 U-PLSR 模型。为避免发生过度拟合现象,采用逐一剔除法进行交叉验证(下同)。U-PLSR 模型结果见表 5。

表 5 U-PLSR 模型结果

Table 5 Results of U-PLSR model

参数	建模样本			验证样本		
	r	RMSE	F	r	RMSE	F
FD	0.260	0.753	2.374	0.320	0.687	1.194
ATFD	0.200	0.814	1.689	0.025	0.994	0.239
RTSD	0.213	0.801	1.826	0.296	0.722	0.939
CR	0.365	0.647	3.872	0.381	0.649	2.465

依照模型的 r 和 F 值最大, RMSE 最小原则,经 CR 光谱变换的建模和验证效果最好,其次是 FD, ATFD 效果最差。相比 MLSR 建模结果,4 种 U-PLSR 模型的 r 均较低,最高仅达到 0.381,这表明 U-PLSR 模型稳定性较差,不能做到准确有效的估算重金属 As 含量。

2.5 M-PLSR

基于 MLSR 和 U-PLSR 模型,为了寻找更为稳定的模型,进一步建立 M-PLSR 模型,所用模型光谱指标和 MLSR 模型相同。

M-PLSR 模型的散点图和验证结果如表 6 和图 6 所示,具体模型为

$$Y = -690.390 - 6011.343X_{FD} + 1174.766X_{RTSD} - 391.829X_{ATFD} + 643.903X_{CR} \quad (6)$$

建模样本和预测样本均超过 0.8,分布更为均匀,相较于 MLSR 模型更靠近 1:1 线。

表 6 M-PLSR 模型结果

Table 6 Results of M-PLSR model

参数	建模样本	验证样本
r	0.852	0.812
RMSE	0.147	0.312
F	32.384	8.598

M-PLSR 模型的建模样本 r , F 值和 RMSE 分别达到 0.852, 32.384 和 0.147,验证样本 r 为 0.812, F 值为 8.598, RMSE 为 0.312,比 MLSR 模型相比,建模 r 增加 0.038, F 值增加 2.947, RMSE 降低 0.012, M-PLSR 建模效果好于 MLSR,预测模型也有所提升。综合来看, M-PLSR 模型拟合度更高,建模效果更佳,模型稳定度也更强,更适合用于估算研究区重金属 As 含量。

3 结论

通过实地采样分析,研究区重金属 As 含量平均值为 $13.230 \text{ mg} \cdot \text{kg}^{-1}$,接近河北省背景值。采矿和冶炼企业对土壤重金属 As 含量有一定影响,使得部分土壤样本 As 含量已经存在轻度污染情况,大部分样本略低于背景值,处于污染的警戒水平。

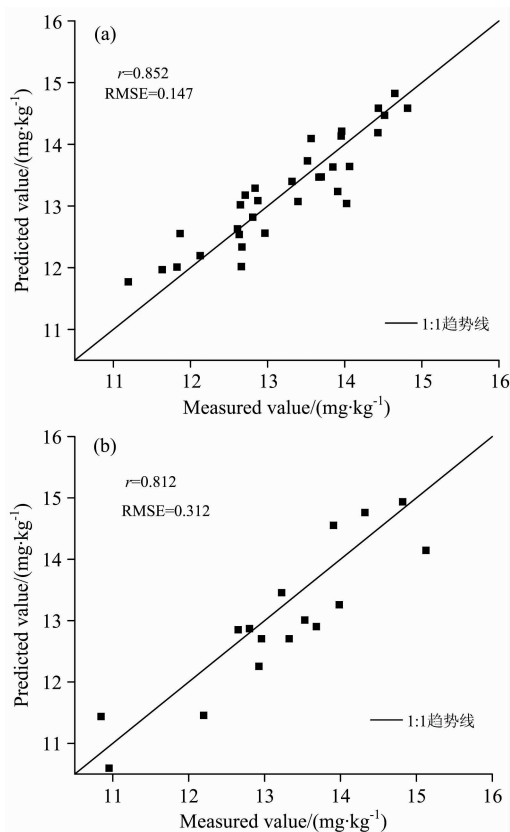


图 6 M-PLSR 模型建模样本和验证样本散点图

(a): 建模样本; (b): 验证样本

Fig. 6 Scatter plots of the two set of samples of fitting and testing for M-PLSR

(a): Fitting samples; (b): Testing samples

研究区域和土壤背景的不同使得重金属 As 的特征波段存在差别。某矿区农田潜育性水稻土重金属 As 的特征波段有 450, 700, 1 200 和 2 200 nm 等^[13];三江源地区以高山草甸土和高山草原土为主的特征波段为 400, 429, 470, 550, 600, 815, 900, 1 415, 1 915, 2 205, 2 355 和 2 455 nm 等^[12];商洛黄褐土的特征波段包括 450, 470, 640~660, 1 030~1 060, 1 080~1 110, 1 230 和 2 190~2 220 nm 等^[11]。这些特征波段与本研究成果基本一致(464, 682, 1 065, 1 211 和 2 208 nm),但并不完全相同,其中本研究与商洛黄褐土重合率最高,可能因为两种土壤类型相似性较高。

土壤重金属 As 含量的研究多采用单光谱变换指标建模,突出单一光谱指标的特征,模型可能存在误差。本工作选用了 10 种光谱指标,CR 与土壤重金属 As 含量相关性最高,且在 U-PLSR 模型中 CR 建模效果最好,这主要是因为 CR 能够有效的消除研究区潜在因素对光谱的影响,可以增强光谱对 As 的敏感性。多光谱变换指标集成建模结合多种最佳特征波段,一定程度上避免了单一光谱指标的不足,达到提升建模效果和模型稳定性的目的。

重金属估算模型受到区域位置和土壤类型甚至污染水平的影响,构建的最佳模型是否适用其他土壤类型和其他地区

土壤还有待深入研究。

针对石家庄市水源地保护区褐土重金属 As 含量的高光谱估测,通过相关分析,建立了土壤重金属砷含量 MLSR, U-PLSR 和 M-PLSR 估测模型,具体结论如下:

(1)重金属 As 与经连续统去除变换后的光谱特征的相关性最大,达到 0.603,与一阶微分含量存在最大负相关性,达到-0.527。

(2)相较于 MLSR 和 U-PLSR, M-PLSR 方法估算土壤重金属 As 含量建模和验证样本的 r 分别达到 0.852 和 0.812,模型预测值和重金属 As 实测值整体拟合度最高。

(3)多光谱变换指标建模效果显著,相比单光谱变换指标模型,它的建模结果和预测能力表现更好,可以对估算重金属 As 含量起到较好的作用。

References

- [1] Araújo Suzana R, Demattê J A M, Vicente S. International Journal of Remote Sensing, 2014, 35(10): 3579.
- [2] WEN Qi-qian, YAN Xiu-lan, SHEN Jun-feng, et al(温其谦, 阎秀兰, 申俊峰, 等). Environmental Science(环境科学), 2019, 40(11): 5090.
- [3] Yu Xiang, Liu Qing, Wang Yebao, et al. Catena, 2016, 137: 340.
- [4] Nocita M, Kooistra L, Bachmann M, et al. Geoderma, 2011, 167-168: 295.
- [5] HE Jun-liang, ZHANG Shu-yuan, ZHA Yong, et al(贺军亮, 张淑媛, 查勇, 等). Remote Sensing Technology and Application(遥感技术与应用), 2015, 30(3): 407.
- [6] ZHAO Qing, LIU Zheng, ZHAO Xu-yang(赵晴, 刘征, 赵旭阳). Research of Soil and Water Conservation(水土保持研究), 2013, 20(2): 121.
- [7] YAO Na, PENG Kun-guo, LIU Zu-gen, et al(姚娜, 彭昆国, 刘足根, 等). Journal of Agro-Environment Science(农业环境科学学报), 2014, 33(2): 313.
- [8] SONG Jing-yi, FU Kai-dao, SU Bin, et al(宋静宜, 傅开道, 苏斌, 等). Acta Geographica Sinica(地理学报), 2013, 68(3): 389.
- [9] FANG Xiao-bo, SHI Jian, LIAO Xin-feng, et al(方晓波, 史坚, 廖欣峰, 等). Chinese Journal of Applied Ecology(应用生态学报), 2015, 26(6): 1883.
- [10] Zhao Li, Hu Yueming, Zhou Wu, et al. Sustainability-Basel, 2018, 10(7): 2474.
- [11] Han Lei, Chen Rui, Zhu Huili, et al. Sustainability-Basel, 2020, 12(4): 1476.
- [12] ZHANG Wei, GAO Xiao-hong, YANG Yang, et al(张威, 高小红, 杨扬, 等). Soil(土壤), 2014, 46(6): 1052.
- [13] REN Hong-yan, ZHUANG Da-fang, QIU Dong-sheng, et al(任红艳, 庄大方, 邱冬生, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2009, 29(1): 114.

Hyperspectral Estimation Model of Heavy Metal Arsenic in Soil

LI Zhi-yuan^{1,2}, DENG Fan^{1*}, HE Jun-liang², WEI Wei¹

1. School of Geosciences, Yangtze University, Wuhan 430100, China

2. College of Resources and Environment Sciences, Shijiazhuang University, Shijiazhuang 050035, China

Abstract Arsenic is one of the heavy metals that seriously harm the human body. The estimation of arsenic content of heavy metals in soil by using hyperspectral technology has great potential for application. However, the applicability and accuracy of the estimation model will vary greatly due to the influence of the region and soil background. The hyperspectral estimation was done in the protection zone of surface water zone of Shijiazhuang. Soil field samples were conducted in the main mining sites and smelting enterprises in the study area and analyzed the heavy metal in the laboratory. After 7-point Savitzky-Golay smoothing of the original spectral reflectance of soil samples, 9 spectral transformations were performed: first derivative (FD), second derivative (SD), reciprocal transformation (RT), reciprocal first derivative (RTFD), reciprocal second derivative (RTSD), absorbance transformation (AT), absorbance transformation and first derivative (ATFD), absorbance transformation and second derivative (ATSD), and continuum removal (CR). Then, the measured content of heavy metal arsenic was correlated with the spectral indexes after spectral transformations, and the maximum sensitive band of each spectral index was extracted. In order to compare the estimation effect of each model, and find the optimal model, the multivariate linear stepwise regression (MLSR), univariate partial least squares regression (U-PLSR) and multivariate partial least squares regression (M-PLSR) method were used to construct the soil heavy metal arsenic content estimation model. The model of arsenic content estimation finally compared the modeling effect through correlation coefficient R^2 , root means square error (RMSE) and statistical value F.

The results showed that: most of the samples in the study area were in the critical state of contamination, although some of the soil samples were already mildly contaminated with heavy metal arsenic. The maximum positive and negative correlations existed between CR and ATFD and arsenic content, respectively. Compared with MLSR and U-PLSR, the estimated value of M-PLSR was closest to the measured value, and the fitting samples' R^2 , RMSE and F of M-PLSR were 0.852, 0.147 and 32.384, respectively, which indicated that the integration effect of multivariate transformation modeling was effective. Therefore, the research results can provide a scientific basis for rapid monitoring of heavy metal arsenic pollution in the region.

Keywords Heavy metal arsenic; Hyperspectrum; Multiple linear stepwise regression; Partial least squares regression

(Received Jun. 11, 2020; accepted Oct. 25, 2020)

* Corresponding author