

文章编号: 1001-1749(2023)02-0275-06

# 基于 DS 算法下耕种区重金属 Cd 污染的反演研究

吴林黛, 曾 涛

(成都理工大学 地球科学学院, 成都 610059)

**摘要:** 传统的重金属元素测定需要耗费大量的人力物力而且效率低下, 因此需要采用更便捷与快速的手段获取土壤的重金属含量值。目前在遥感中, 可以利用光谱的特征波段反演出重金属含量值, 能够做到以点带面, 并且实现大面积的重金属反演, 这是传统土壤重金属含量测定手段无法实现的。以保定市安新县某耕作区为研究区, 采取了 97 组样本点。利用 DS 算法对野外原位光谱进行校正, 再经过微分、包络线去除、倒数的对数等预处理, 让野外原位光谱直接参与重金属反演模型的构建。结果表明: 在研究区内, 基于 DS 算法下的提纯光谱与重金属 Cd 含量之间有较好的相关性。DS 算法能够在一定程度上减少外界环境对野外原位光谱的影响, 提高光谱的信号质量。在四种光谱预处理中, 一阶微分参与建模的反演精度最高 ( $R^2$  可达到 0.704 4, RMSE 为 0.098 19), 为该研究区内重金属 Cd 元素的最佳反演模型。该模型方法能够有效估测该研究区内的重金属 Cd 含量, 实现了野外原位光谱直接参与模型构建并且有效反演重金属 Cd 含量值。

**关键词:** 土壤重金属; DS 算法; 高光谱; 遥感反演

**中图分类号:** P 631.4 **文献标志码:** A **DOI:** 10.3969/j.issn.1001-1749.2023.02.16

## 0 引言

人类活动及工业化进程的不断推进, 是土壤受到重金属污染的主要原因之一。在土壤中, 重金属污染物具有极强的隐蔽性、积累性、难以迁移性和危害性等特点<sup>[1]</sup>。耕作区中的重金属含量超标不仅会对农业生产和发展产生严重的影响, 食用重金属含量超标的粮食, 也会引发人的身体健康问题<sup>[2]</sup>, 其中镉元素(Cd)是人体非必需且有害的元素, 人体摄入 Cd 会导致易致癌, 致突变等危害<sup>[3]</sup>。通过光谱数据对耕作区的重金属 Cd 元素进行有效反演, 能够快速、动态、大范围地获取耕作区土壤中的重金属含量<sup>[4]</sup>。

传统的土壤重金属测定一般采用土壤样本点与化学分析的方法。这种方式花费昂贵且效率低下。而高光谱反演重金属具有便捷、快速等优势, 逐步成为国内、外学者的研究重点<sup>[5]</sup>。在野外光谱的测定中, 我们发现野外光谱受气候、光照、土壤湿度、土壤表面状况等外界环境影响较大, 其中土壤水分掩盖了土壤的主要光谱吸收, 扭曲了总体的光谱形状<sup>[6]</sup>。因此消除野外环境因素对光谱带来的干扰成为研究者关注的重点, 在前人的报道中, 可以使用直接校正法 (direct standardization DS) 和分段直接校正法 (Piecewise direct standardization PDS) 对野外光谱进行提纯与校正<sup>[7]</sup>。Ji W, Rossel RAV 等<sup>[8]</sup>用 DS 算法对野外光谱数据提纯。结果表明, DS 算法可以有效地解释土壤中水含量和环境因素对于野外光谱

收稿日期: 2021-12-24

第一作者: 吴林黛(1996—), 女, 硕士, 主要研究方向为摄影测量与遥感, E-mail: 769081411@qq.com。

通信作者: 曾涛(1975—), 女, 博士, 副教授, 主要研究方向为摄影测量与遥感, E-mail: 649218129@qq.com。

的影响。野外原位光谱在 DS 的转换基础上,结合实验室光谱波段,进行光谱联合反演的效果最好<sup>[9]</sup>,然后将光谱进行预处理后,直接对进行 DS 转换后的原位光谱进行单一建模的研究较少,因此笔者基于经过 DS 转换后的原位光谱,对最优模型的构建以及重金属 Cd 元素的反演展开讨论与研究。

笔者基于保定市安新县研究区内的 60 组土壤样本点,对经过 DS 算法提纯后的野外原位光谱进行光谱预处理,分析特征光谱波段与重金属含量直接的相关性,筛选出光谱特征波段,利用多元回归分析等方法对研究区内的重金属含量值进行反演,根据决定系数与均方根误差综合精度,来评价不同光谱预处理方法下的重金属 Cd 反演精度。以期实现对通过 DS 转换以及光谱预处理后的野外原位光谱,能够直接参与模型的反演。

## 1 数据与方法

### 1.1 研究区

研究区位于河北省保定市辖县—安新县的西南方向,地理位置为东经  $38^{\circ}44'E \sim 38^{\circ}50'E$ ,北纬  $115^{\circ}33'N \sim 115^{\circ}44'N$ (图 1)。该区地处温暖带半湿润大陆季风区,适合农业与渔牧业的发展。该区域内为耕地种植区与建筑群区,面积大约为  $90 \text{ km}^2$ 。

### 1.2 数据采集

土壤光谱采集使用 SVC HR-1024i 便携式高光谱地物波谱仪,波长范围为  $350 \text{ nm} \sim 2500 \text{ nm}$ 。研究区内对每个采样点的地表原始土壤面进行光谱的收集,每个采样点再分别采取两份约  $1 \text{ kg}$  耕地土壤等量样本,去除较大石块及草木渣滓后密封装袋。将土壤样本在实验室风干研磨后分别过 20 目及 100 目筛制成标准样本。分别用于实验室内土壤光谱测定以及电感耦合等离子体制谱仪下重金属含量 Cd 值的测定。

### 1.3 野外原位光谱 DS 转换

DS 算法是一种针对野外光谱的直接校正法<sup>[10]</sup>,其核心就是利用野外实测光谱与实验室实测光谱之间的矩阵函数关系来转换野外光谱,目的是为了消除外界环境对于野外光谱的影响。

设定野外原位光谱与实验室实测光谱的  $n \times p$  阶矩阵分别为  $\mathbf{X}_{\text{lab}}$  和  $\mathbf{X}_{\text{field}}$ 。其中矩阵中的  $n$  代表样本数, $p$  为光谱波段数。实验室与野外原位光谱矩阵之间的关系式如式(1)所示。

$$\mathbf{X}_{\text{lab}} = \mathbf{X}_{\text{field}} * \mathbf{B} + \mathbf{E} \quad (1)$$

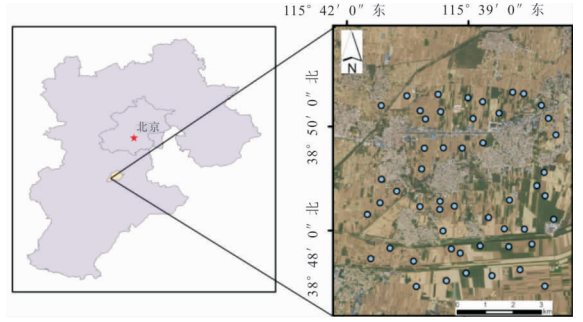


图 1 研究区位置与采样点分布

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

式中: $\mathbf{B}$  是  $p \times p$  个位置参数构成的转换矩阵,由  $\mathbf{X}_{\text{lab}}$  和  $\mathbf{X}_{\text{field}}$  决定; $\mathbf{E}$  为背景残差,可以表示为:

$$\mathbf{E} = \boldsymbol{\lambda} \times \mathbf{d}_s^T \quad (2)$$

式中: $\mathbf{d}_s^T$  为  $p \times p$  阶矩阵,代表两种测量环境下光谱间的基线差异; $\boldsymbol{\lambda}$  为  $n \times 1$  的列向量,并且向量中元素值都为“1”。由式(1)和式(2)可得式(3)。

$$\mathbf{X}_{\text{lab}} = \mathbf{X}_{\text{field}} * \mathbf{B} + \boldsymbol{\lambda} \mathbf{d}_s^T \quad (3)$$

为计算转换矩阵  $\mathbf{B}$ ,引入  $m \times m$  阶中心化矩阵  $\mathbf{C}_m$

$$\mathbf{C}_m = \mathbf{I}_m - (1/m)\boldsymbol{\lambda} * \boldsymbol{\lambda}^T \quad (4)$$

为了方便计算矩阵  $\mathbf{B}$  将式(3)两边同时左乘矩阵  $\mathbf{C}_m$ ,因  $\boldsymbol{\lambda} \mathbf{d}_s^T$  所有元素都一样,中心化后结果为 0,可得

$$\mathbf{X}_{\text{lab}} = \bar{\mathbf{X}}_{\text{field}} * \mathbf{B} \quad (5)$$

$$\mathbf{B} = \bar{\mathbf{X}}_{\text{field}}^+ + \mathbf{X}_{\text{lab}} \quad (6)$$

式中, $\mathbf{X}_{\text{lab}}$  和  $\bar{\mathbf{X}}_{\text{field}}$  表示中心化后实验室与野外原位光谱矩阵,+代表 Moore-Penrose 广义逆矩阵,基线差异矩阵则为:

$$\mathbf{d}_s = \bar{\mathbf{X}} \bar{\mathbf{X}}_{\text{lab}}^T - \mathbf{B}^T \bar{\mathbf{X}}_{\text{lab}}^T \quad (7)$$

式中: $\bar{\mathbf{X}}$  是光谱矩阵, $\mathbf{X}$  每列元素均值构成的  $n \times p$  阶矩阵。最终经过上述 DS 转换后的土壤样本野外光谱可以表示为式(8)。

$$\bar{\mathbf{X}}_{\text{field}} = \mathbf{X}_{\text{field}} \mathbf{B} + \mathbf{E} \quad (8)$$

## 2 实验结果与分析

### 2.1 数据预处理与相关性分析

为了避免数据的冗余,对经过 DS 转换后的实测点光谱数据进行单位为  $5 \text{ nm}$  的重采样。在确保数据量压缩的情况下也能保留光谱的特征波段<sup>[11]</sup>。

在光谱预处理中,微分变化可以有效地消除背景噪声值<sup>[12]</sup>。其中一阶微分变换(First Derivative

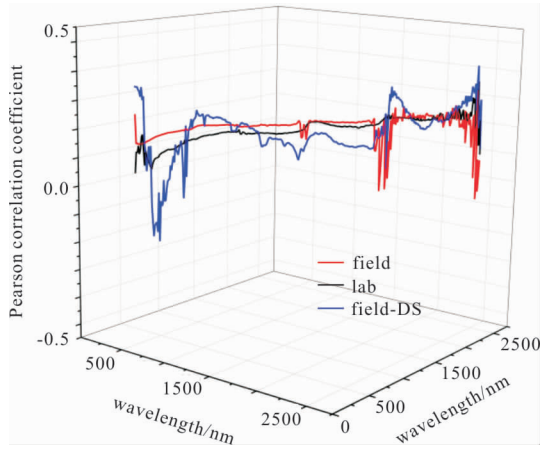


图 2 波谱反射率与 Cd 含量之间的相关性系数  
Fig. 2 Correlation coefficient between spectral reflectance and heavy metal Cd content

FD)后的光谱可以去除部分线性噪声,增强局部的敏感光谱响应值。二阶微分变换(Second Derivative SD)可以消除基线漂移以及仪器背景的影响<sup>[13]</sup>。利用倒数的对数(Alog)进行预处理,可以在调整由于光线的亮度变化以及土壤表明凹凸不平给实验结果带来的影响<sup>[14]</sup>。包络线去除(Continue Remove CR)可以将每个波段对应的反射率值归一到同一背景下,从而突出光谱的吸收和反射特征<sup>[15]</sup>。在进行建模之前,分别对野外原位光谱、实验室实测光谱、野外 DS 以及预处理结果后的数据与重金属 Cd 进行皮尔逊相关性分析,相关性结果如图 2 所示。

由图 2 可以看出,野外原位光谱受外界环境影响较大,与重金属 Cd 含量这二者之间的变量之间皮尔逊相关系数小于 0.3,可以判断出两变量基本不相关。结合实验室实测光谱的先验知识,再经过 DS 算法校正后,在波长为 2 477 等波段的相关系数可以在 0.01 显著水平上达到 0.42。在统计学意义上,可以认为两变量中度相关。这反映出通过 DS 算法,可以提高野外原位光谱与重金属含量值之间的相关系数值。

为了进一步提高波谱响应函数的敏感性,将通过 DS 算法转换后的野外原位光谱进行微分、倒数的对数以及连续统去除等预处理,实测光谱以及光谱预处理如图 3 所示。预处理分别包括微分预处理、包络线预处理与倒数的对数预处理。经过预处理之后,相关性系数均有所提升,其中微分处理后波段与 Cd 含量在 0.01 显著水平上的皮尔逊相关系数可以达到 0.566(图 4)。

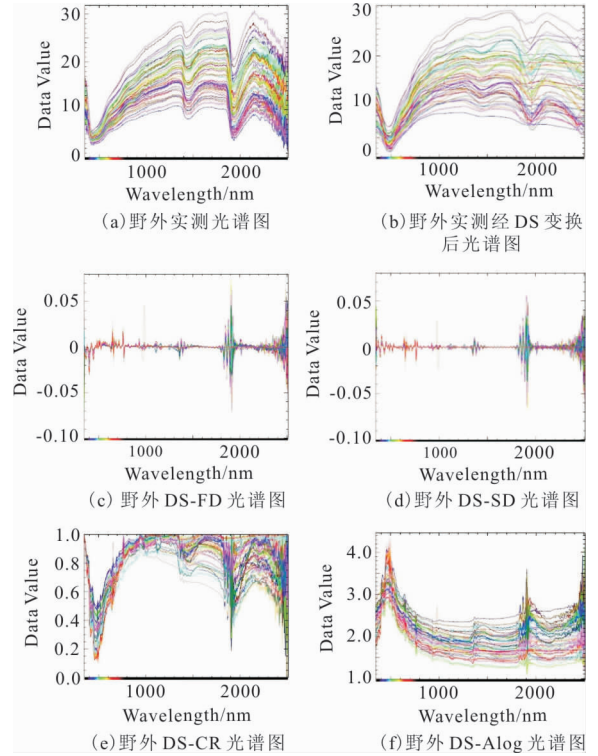


图 3 各波谱曲线示意图

Fig. 3 Schematic diagram of each spectral curve

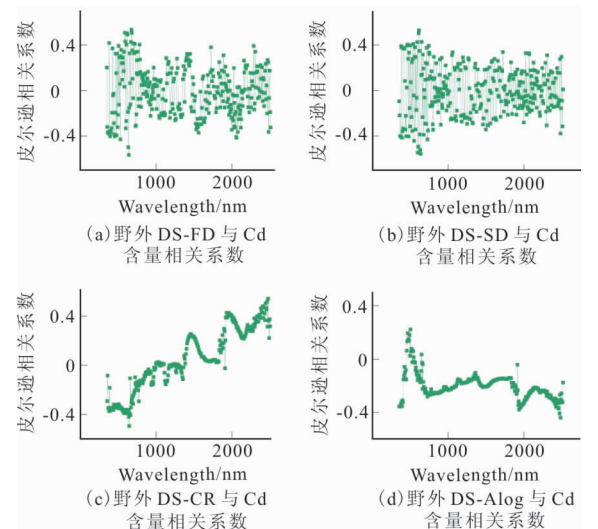


图 4 光谱预处理与重金属 Cd 含量值之间的相关性系数

Fig. 4 Correlation coefficient between spectral pretreatment and heavy metal Cd value

## 2.2 特征因子选择

为了保证建模精度,在建模之前,筛选经过野外 DS 算法后相关系数较高的波段作为建模的备选波段。具体特征波段选择表 1。

## 2.3 多元逐步回归建模

根据实测土壤光谱反射率与重金属 Cd 含量之

表 1 特征波段选择

Tab. 1 Characteristic spectral band selection

Heavy Metal	Spectral Index	Characteristic Spectral band
Cd	Der 1st	376.3、388.9、451.2、524.1、583.3— 629.6、635.3、663.3—697.1、713.6、 777.8、2059.5、2054.3
	Der 2nd	357.4—382.6、438.8、4457.3、469. 6、530.1、542、583.3—595、600.8、 606.6—641、719.1
	CR	635.3—658、1929、1944.9—1992.2、 2356—2397.1、2403.9、2417.4、 2424.1、2430.8、2444.2、2450.9— 2457.5、2464.1、2477.3
	Alog	344.7、351—357.4、363.7—382.6、 1918.3、1929、1934.3—1939.6、 1944.9、1950.2—1992.2、2403.9、 2417.4、2430.8、2444.2—2464.1、 2477.3

Der 1st 表示光谱一阶微分, Der 2nd 表示光谱二阶微分, CR 表示光谱包络线去除, Alog 表示光谱倒数的对数

间的皮尔逊相关系数结果, 摘取相关性较高的特征因子参与建模分析。本次实验建模采用多元逐步回归法(MSR), 先由一个自变量开始, 每次引入一个偏回归平方和最大具有统计学意义的自变量, 由少至多, 直到没有具有统计学意义的因素可以带入方程为止。在多元逐步回归分析中, 将重金属 Cd 含量作为因变量, 其余相关性特征波段作为自变量。建立多个变量之间的数学模型数量关系, 经过 DS 校正算法后, 符合参与建模的数据点为 59 组。根据 RANK 法, 将重金属含量值按照升序排列, 每隔一组进行数据取值, 选取三分之二的的数据为建模集, 剩下三分之一的数据为模型验证集。

最后根据建模结果的决定系数  $R^2$  以及均方根误差 RMSE 两种因素综合评定建模结果的精度。其中  $R^2$  的取值范围在 0 到 1 之间, 数值越高, 则相对应表明模型精度越稳定。均方根误差则是数学统计中衡量一组数据中观测值与真值之间的偏差。结合决定系数以及均方根误差结果, 选取出最优重金属反演模型。多元逐步回归模型建模结果实测值与预测值如图 4 所示。

由图 5 可以看出, 多元逐步回归分析模型精度中最高的是经过一阶微分预处理的实测光谱,  $R^2$  为 0.7044, 预测精度最高, 验证点都在  $Y=X$  的趋势

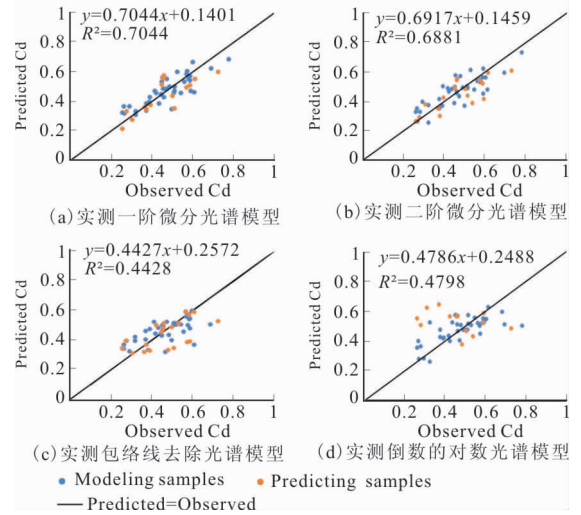


图 5 多元逐步回归模型实测值和预测值散点  
Fig. 5 Scatter plot of the measured and predicted values of the MSR model

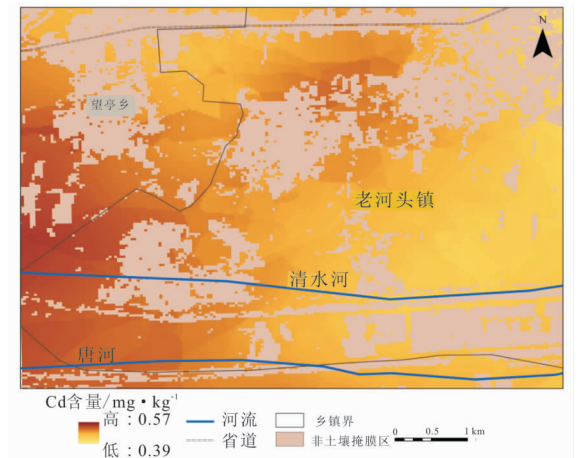


图 6 研究区重金属 Cd 反演图  
Fig. 6 Inversion map of heavy metal Cd in the study area

线附近。次之高精度为经过二阶微分预处理的实测光谱,  $R^2$  为 0.6881。而经过包络线去除与倒数的对数预处理之后参与建模的结果不太理想。验证点的重金属含量相对于实测值来说都大幅度偏高或者偏低。预测值与实测值的点位分布都偏离  $Y=X$  的趋势线。最后将多元逐步回归建模结果得出的回归方程进行模型验证, 验证结果(表 2)。

由表 2 的数据结果可以看出, 光谱变换形式为一阶微分的建模精度与验证精度是最高的, 分别为 0.7044 与 0.6512。光谱变换形式为二阶微分的建模精度与验证精度次之, 分别为 0.6881 与 0.6383。而光谱变化形式为包络线去除以及倒数

表 2 建模结果评价

Tab. 2 Modeling results evaluation

Model index	Modeling		Prediction	
	$R^2$	RMSE	$R^2$	RMSE
FD	0.7044	0.09819	0.6512	0.11438
SD	0.6881	0.09973	0.6383	0.10346
CR	0.4428	0.07323	0.3877	0.09342
Alog	0.4798	0.08464	0.1369	0.07301

的对数建模精度与验证精度并不理想,多元逐步回归模型不够稳定,不能将此模型直接应用于重金属 Cd 的反演。

根据建模结果,选择综合精度指标最高的光谱变换形式对研究区内的 Cd 重金属含量值进行克里金插值与反演(图 6)。由图 6 可以看出,研究区内的 Cd 含量高值位于西部以及中北部,都位于非土壤掩膜区周围,即房屋建筑周围。而东部以及中部的耕地区中重金属含量较低。

### 3 结论

笔者通过对野外原位光谱进行 DS 转换以及波谱预处理后,筛选出光谱特征波段参与到多元逐步回归模型中的构建,为野外原位光谱直接参与到高光谱反演重金属模型中,提供了例证以及验证了实验的可行性,并得出以下结论:

1)野外 DS 转换法可以有效去除野外原位光谱中的外界环境噪声,增强波谱敏感性,提高波段反射率与重金属含量值之间的皮尔逊相关系数,提升高光谱反演重金属值含量中的物理解释意义。

2)经过 DS 转换后的野外原位光谱再进行光谱预处理,可以发现四种预处理手段都可以增强波谱的响应值,使得波谱与重金属含量之间的相关性系数进一步提高,能帮助在后面的研究中识别出更多的特征波段。

3)通过对野外 DS 光谱进行多元逐步回归模型的建立,发现一阶微分处理后的特征波段参与模型精度的指标最高,能为反演重金属 Cd 含量提供思路。

4)研究区内的西部以及北部重金属含量都偏高,这可能是由于人类活动以及交通运输等产生的气体和粉尘以气溶胶的形态进入大气,再通过自然沉降以及降水等方式进入土壤中,引起土壤内的重

金属元素值较高。

### 参考文献:

- [1] 王鹤亭. 土壤重金属污染现状与修复技术应用[J]. 南方农机, 2021, 52(13): 64—66.  
WANG H T. Current status of soil heavy metal pollution and application of remediation technology[J]. South China Agricultural Machinery, 2021, 52(13): 64—66. (In Chinese)
- [2] 王炫凯,曲宝成,艾孜买提·阿合麦提,等. 我国农田重金属污染状况及修复技术研究进展[J]. 清洗世界, 2021, 37(8): 55—58, 61.  
WANG X K, QU B C, AIZIMATI A, et al. Research progress of heavy metal pollution in farmland and remediation technology in my country[J]. Clean the World, 2021, 37(8): 55—58, 61. (In Chinese)
- [3] 王豫. 浅谈食品中重金属对人体的危害及预防[J]. 青海农技推广, 2010(4): 8—9, 16.  
WANG Y. Talking about the harm of heavy metals in food to human body and its prevention[J]. Qinghai Agricultural Technology Extension, 2010(4): 8—9, 16. (In Chinese)
- [4] 宋练,简季,谭德军,等. 万盛采矿区土壤 As, Cd, Zn 重金属含量光谱测量与分析[J]. 光谱学与光谱分析, 2014, 34(3): 812—817.  
SONG L, JIAN J, TAN D J, et al. Spectral measurement and analysis of heavy metal content of As, Cd, and Zn in soil in Wansheng mining area[J]. Spectroscopy and Spectral Analysis, 2014, 34(3): 812—817. (In Chinese)
- [5] 钱佳,郭云开,蒋明,等. 不同类型土壤 Cu 含量高光谱联合反演建模[J]. 测绘科学, 2020, 45(8): 138—144.  
QIAN J, GUO Y K, JIANG M, et al. Hyperspectral joint inversion modeling of Cu content in different types of soils[J]. Science of Surveying and Mapping, 2020, 45(8): 138—144. (In Chinese)
- [6] OGEN Y, FAIGENBAUM—GOLOVIN S, GRANOT A, et al. Removing moisture effect on soil reflectance properties: a case study of clay content prediction[J]. Pedosphere, 2019, 29(4): 421—431.
- [7] ACKERSON J P, MORGAN C, Ge Y. Penetrometer mounted visir spectroscopy—application of epoplsgemode[J]. Geoderma, 2017(286): 131—138.
- [8] JI W, ROSSEL RAV, SHI Z. Accounting for the effects of water and the environment on proximally sensed Vis—nir Soil spectra and their calibrations[J]. European Journal of Soil Science, 2015, 66(3): 555—565.

- [9] 邹滨,涂宇龙,姜晓璐,等. 土壤 Cd 含量实验室与野外 DS 光谱联合反演[J]. 光谱学与光谱分析, 2019, 39(10): 3223—3231.  
ZOU B, TU Y L, JIANG X L, et al. Joint inversion of soil Cd content in laboratory and field DS spectra [J]. Spectroscopy and Spectral Analysis, 2019, 39(10): 3223—3231. (In Chinese)
- [10] 张霞,丁松滔,岑奕,等. 结合野外光谱与实验室光谱的土壤 Pb 含量反演方法研究[J]. 武汉大学学报(信息科学版), 2022, 47(09): 1479—1485.  
ZHANG X, DING S T, CEN Y, et al. Research on the inversion method of soil Pb content combining field spectroscopy and laboratory spectroscopy [J]. Journal of Wuhan University (Information Science Edition), 2022, 47(09): 1479—1485. (In Chinese)
- [11] THOMAS K, STEFAN S. Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy [J]. Environmental Science & Technology, 2002, 36(12): 2742—7.
- [12] 雷宇斌,刘宁,郭云开,等. 高光谱组合变换下土壤 Cd 含量 GWR 模型反演研究[J]. 测绘工程, 2018, 27(11): 71—76.  
LEI Y B, LIU N, GUO Y K, et al. Research on GWR model inversion of soil Cd content under combined hy-
- perspectral transformation [J]. Geomatics Engineering, 2018, 27(11): 71—76. (In Chinese)
- [13] 周伟,谢利娟,杨晗,等. 基于高光谱的三江源区土壤有机质含量反演[J]. 土壤通报, 2021, 52(3): 564—574.  
ZHOU W, XIE L J, YANG H, et al. Inversion of soil organic matter content in the source region of three rivers based on hyperspectral [J]. Chinese Journal of Soil Science, 2021, 52(3): 564—574. (In Chinese)
- [14] 沈强,张世文,葛畅,等. 矿业废弃地重构土壤重金属含量高光谱反演[J]. 光谱学与光谱分析, 2019, 39(4): 1214—1220.  
SHEN Q, ZHANG S W, GE C, et al. Hyperspectral inversion of heavy metal content in reconstructed soil from mining wasteland [J]. Spectroscopy and Spectral Analysis, 2019, 39(4): 1214—1220. (In Chinese)
- [15] 陶超,王亚晋,邹滨,等. 土壤重金属铅、锌高光谱反演模型可迁移能力分析研究[J]. 光谱学与光谱分析, 2018, 38(6): 1850—1855.  
TAO C, WANG Y J, ZOU B, et al. Analysis of the transferability of soil heavy metal lead and zinc hyperspectral inversion models [J]. Spectroscopy and Spectral Analysis, 2018, 38(6): 1850—1855. (In Chinese)

## Research on inversion of heavy metal Cd pollution in cultivated area based on DS algorithm

WU Lindai, ZENG Tao

(Chengdu University of Technology College of Earth Science Chengdu, Chengdu 610059, China)

**Abstract:** The traditional determination of heavy metal elements requires a lot of workforce and material resources and is inefficient. Therefore, it is necessary to adopt a more convenient and rapid method to obtain the value of the heavy metal content of the soil. In remote sensing, it is possible to use the spectrum's characteristic bands to invert heavy metals' content. It is possible to achieve a large area of heavy metal inversion. This is not possible with traditional soil heavy metal content determination methods. Using a farming area in Anxin County, Baoding City as the research area, 97 sample points were taken. The DS algorithm is used to correct the field spectra and then pre-processing such as FD, SD, CR, and Alog, so that the field spectra can directly participate in constructing the heavy metal inversion model. The results show that: in the study area, there is a good correlation between the purification spectrum based on the DS algorithm and the heavy metal Cd content. Therefore, the DS algorithm can reduce the influence of the external environment on the field spectrum to a certain extent and improve the signal quality of the spectrum. Among the four kinds of spectral preprocessing, the first derivative involved in the modeling has the highest inversion accuracy ( $R^2$  can reach 0.7044, RMSE is 0.09819), the best inversion model for heavy metal Cd in the study area. The model method can effectively estimate the heavy metal Cd content in the study area. Realize that the field spectroscopy directly participates in the model construction and effectively retrieves the heavy metal Cd content value.

**Keywords:** soil heavy metals; DS algorithm; hyperspectral; remote sensing inversion