

# 土壤 Cd 含量实验室与野外 DS 光谱联合反演

邹 滨, 涂宇龙, 姜晓璐, 陶 超, 周 莱, 熊立伟

中南大学有色金属成矿预测与地质环境监测教育部重点实验室, 地球科学与信息物理学院, 湖南长沙 410083

**摘 要** 土壤重金属高光谱遥感建模理论上能够大大降低传统化学分析测定所需成本, 正逐步发展为有效探查土壤污染空间分布与开展污染土壤综合防治的关键技术。然而土壤重金属高光谱遥感调查技术目前多局限于稳定可控条件下的实验室光谱模型, 野外诸多因素(光照、湿度、土壤粗糙度等)影响下野外原位光谱模型的有效性已成为困扰该项技术大范围推广亟待突破的关键科学问题。以湖南衡阳市某矿区为例, 分别利用 ASD 地物光谱仪和等离子发射光谱法测定 46 个土壤样品 350~2 500 nm 的实验室光谱和 Cd 含量, 并在土壤取样时同步测量样品野外原位光谱。在运用 DS(direct standardization)转换算法处理野外光谱的基础上, 融合实验室光谱先验知识, 基于主成分逐步回归建模方法开展了土壤 Cd 含量实验室与野外原位 DS 光谱联合反演实验, 交叉验证了模型的稳定性。同时为深入探究实验室与野外原位 DS 光谱联合反演模型的有效性, 将其与基于实验室光谱、野外原位光谱、野外原位 DS 光谱、实验室与野外原位光谱联合建立的主成分逐步回归模型开展了对比分析。结果表明: 野外原位光谱反演模型精度( $R^2 = 0.56$ )明显低于实验室光谱反演模型( $R^2 = 0.64$ ), 野外原位 DS 光谱反演模型与之相比精度有所提升( $R^2 = 0.66$ ); 在野外原位光谱 DS 转换校正基础上, 联合实验室光谱先验知识的土壤 Cd 含量反演模型精度最高,  $R^2$  可达 0.72。与此同时, 实验室与野外原位 DS 光谱联合反演模型揭示 482, 565, 979 和 2 206 nm 波段对研究区土壤 Cd 含量有较好指示性, 此结果与实验室光谱反演模型所识别的特征波段一致, 两者物理意义相同。研究结果证实了实验室光谱先验知识以及 DS 转换算法能够提升野外原位光谱模型的可靠性, 可为发展土壤 Cd 含量野外原位高光谱遥感探测提供重要的提供理论与方法支撑。

**关键词** 土壤; 重金属; 高光谱; DS 转换; 遥感反演

**中图分类号:** TP79 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)10-3223-09

## 引 言

土壤 Cd 易通过植物富集进入食物链危害人体健康, 正受到广泛关注<sup>[1]</sup>。快速测定土壤中 Cd 的含量及空间分布已成为我国当下开展重金属污染土壤治理亟待突破的关键科学问题<sup>[2]</sup>。相比传统人工土壤取样与实验室化学测定花费大、效率低<sup>[5]</sup>, 高光谱遥感因高效、无破坏性、低成本特点, 正逐步发展成为探测与识别土壤物理、化学以及生物特性的新兴技术方法<sup>[2]</sup>。

而截止目前, 已报道的土壤特性高光谱探测研究多依赖环境可控条件下的实验室光谱<sup>[6]</sup>, 直接采用野外原位光谱的研究相对少见。但类比基于实验室光谱和野外原位光谱分别建立的土壤黏土、碳酸钙<sup>[7]</sup>、盐分含量<sup>[8]</sup>高光谱反演模型可

知, 基于实验室光谱建立的土壤成分含量反演模型常常难以直接运用于野外。有学者对此试图通过河流沉积物<sup>[9]</sup>与采矿区土壤野外原位光谱<sup>[10]</sup>模拟估计矿区及附近河流的重金属分布。但此种工作往往需要样品数量足够的重金属污染土壤野外光谱数据库支撑。不仅如此, 受湿度、温度、光照条件、土壤结构、表面粗糙度等因素影响<sup>[7]</sup>, 土壤野外原位光谱测定常会因取样点差异存在不确定性。为消除环境因素对取样点野外原位光谱的影响<sup>[6]</sup>, 系列光谱转换方法包括直接校正算法(direct standardization, DS)、分段直接校正算法(piecewise direct standardization, PDS)、外部参数正交化法(external parameter orthogonalization, EPO)相继用于野外原位光谱校正。基于转换后的野外原位光谱, 研究人员目前也已经成功估算了土壤有机质、黏土成分的含量<sup>[6]</sup>, 对比了光谱转换前后的模型效果差异<sup>[11]</sup>。但是, 光谱转换方法在野外原位

收稿日期: 2018-08-23, 修订日期: 2018-12-12

基金项目: 环保公益性行业科研专项经费项目(201509050), 中国博士后科学基金特别资助项目(2013T60780), 中南大学“创新驱动计划”项目(2015CXSS005), 湖南省自然科学基金项目(2017JJ3378)和教育部重点实验室开放基金项目(2017YJSJ07)资助

作者简介: 邹 滨, 1981 年生, 中南大学地球科学与信息物理学院教授 e-mail: 210010@csu.edu.cn

光谱模式下反演土壤重金属含量的效果却尚未见报道。因此,耦合实验室光谱先验知识与野外原位光谱在较大地理空间反演土壤中 Cd 含量仍然是一个悬而未决的问题。

## 1 实验部分

### 1.1 研究区与样品测定

选取湖南省衡阳市某矿区为例,矿区运营百余年来尾砂堆积与矿渣道路运输等开采活动已经对周边水田、菜地等土壤造成严重的 Cd 污染。顾及矿区尾砂坝、选矿点、冶炼厂、道路、河流水系等分布及地形起伏度,采集 46 个土壤样品(图 1),采样点土地利用类型主要为农业用地,少部分样本为建筑用地和矿区土壤,土壤类型以红壤为主。采样过程中,首先对每个采样点选取土壤表面约 20 cm<sup>2</sup> 平整区域,在去除石块、草木等杂质基础上运用 PSR-3500 野外便携式光谱仪(光谱波长范围 350~2 500 nm)测量该点野外原位光谱。测量前进行白板辐射定标,测量时,保持探头方向与太阳入射角方向相对稳定,每个采样点重复测量 10 次光谱曲线取算术平均值。野外原位光谱测量后,收集所测样点土壤样品、分装编号带回实验室用于重金属含量测定及实验室光谱测量。采样时期为 2017 年 3 月,整个野外采样过程在晴朗天气进行,时间从上午 9 点持续至下午 4 点。土壤样品实验室光谱测量过程在暗室进行,唯一光源小型卤素灯与仪器探头固定,测量时保持探头样品间距 5 cm。测量前同样开展白板辐射定标,对每个样品重复测量 10 次光谱曲线取算术平均值。土壤样品 Cd 含量测定包括研磨、过 100 目筛,采用盐酸、硝酸、高氯酸溶解,运用等离子发射光谱法测定等步骤。

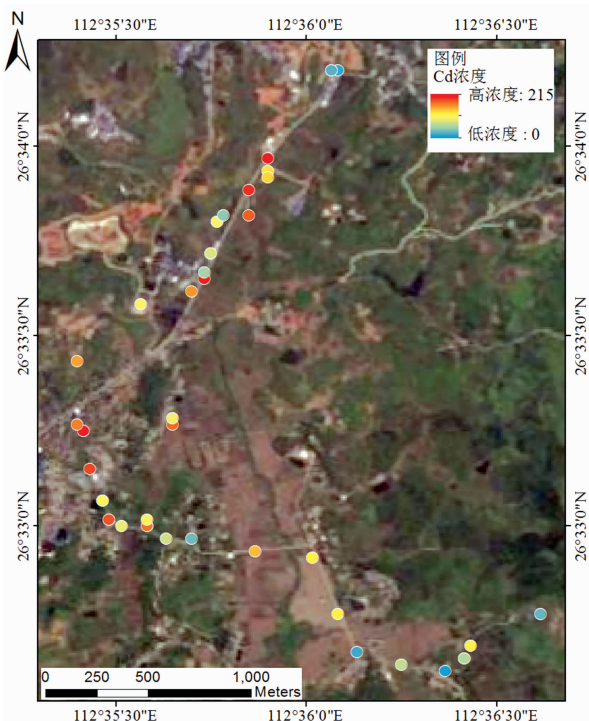


图 1 研究区土壤样品采样点分布示意图

Fig. 1 The distribution of soilsampling sites in study area

### 1.2 野外原位光谱 DS 转换

DS 转换是一种光谱直接校正算法,可通过光谱间的相互转换消除野外光谱测量中环境因素的影响<sup>[12]</sup>。假定土壤样品实验室光谱与样品采样点野外原位光谱  $n \times p$  阶矩阵分别为  $\mathbf{X}_{\text{lab}}$  和  $\mathbf{X}_{\text{field}}$ 。 $n$  为样本数, $p$  为光谱波段数。实验室与野外原位光谱矩阵间的关系如式(1)所示

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

式中, $\mathbf{B}$  是  $p \times p$  个未知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,可表示为

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

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

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

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

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

将式(3)左右同时左乘矩阵  $\mathbf{C}_m$ ,因  $\lambda\mathbf{d}_s^T$  所有元素都一样,中心化后结果为 0,可得

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

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

式中, $\bar{\mathbf{X}}_{\text{lab}}$  和  $\bar{\mathbf{X}}_{\text{field}}$  表示中心化后的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

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

式中, $\bar{\mathbf{X}}$  是光谱矩阵  $\mathbf{X}$  每列元素均值构成的  $1 \times p$  阶矩阵。最终经过上述 DS 转换后的土壤样品野外光谱可表示为

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

转换过程中,采用 Kennard-Stone 算法筛选具有代表性的土壤样品实验室光谱与采样点野外原位光谱,用于光谱 DS 转换矩阵计算。代表性光谱样品转换集筛选以样本光谱间距离最大为准则,最终筛选的转换集样本需要空间覆盖足够。具体步骤:(1)计算任意两个样本光谱间的欧式距离,选择间距最大的两个样本作为初始转换集;(2)计算剩余样本光谱与现有转换集样本间的距离(对于每个样本而言,到转换集各样本最短距离为该样本到转换集的距离);(3)选择距转换集距离最大的样本光谱作为新加入的转换集样本,重复(2)和(3)直至筛选的转换集样本数达到预设数量。预设数量合理性通过差异化转换集样本  $t(t=5, 10, 15, 20, 25, \dots)$  输入下的最终 Cd 含量光谱反演模型预报误差平方和(predicted error sum of squares, PRESS)加以判定。PRESS 值达到最小且变化趋向平稳状态是确定转换集样本数的重要准则。

### 1.3 实验室光谱与野外原位 DS 光谱联合建模

主成分分析作为一种重要的统计分析方法,在土壤成分光谱反演建模中相对于单一波段输入往往具有更强的物理指示作用。以此推断,假若土壤 Cd 含量变化具有主成分可解释特点,则土壤样品实验室与野外原位 DS 光谱理论上均具有对应的主成分特征。顾及实验室光谱主成分与野外原位 DS 光谱主成分之间的相关性,则可以将噪声较小的实验室光谱中的先验知识用于指导噪声较大的野外原位 DS 光谱建模,提升野外原位 DS 光谱土壤 Cd 含量反演模型的物理解释力。本研究相关性强弱采用 Pearson 相关系数评价。实验

室光谱主成分与野外原位 DS 光谱主成分间强相关则指示两者具有相同物理意义,反之实验室光谱对野外原位 DS 光谱土壤 Cd 含量反演建模作用有限。实验室与野外光谱联合建模过程中,先将实验室光谱主成分变量作为先验知识进行筛选,然后选择相关性最强野外光谱主成分参与建模。

鉴于主成分逐步回归(principle component stepwise regression, PCSR)在基于光谱主成分的土壤重金属含量反演建模中可有效选择合适主成分变量并同时规避哑变量<sup>[13-14]</sup>。在上述实验室光谱与野外原位 DS 光谱强相关主成分判定基础上,本研究采用 PCSR 方法分别构建了基于实验室光谱、野外原位光谱、野外原位 DS 光谱的研究区土壤 Cd 含量反演模型,并将实验室光谱分别与野外原位光谱和野外原位 DS 光谱联合建模,对比评价实验室与野外原位 DS 光谱联合下的土壤 Cd 含量反演效果。所建模型分别简称为:实验室-PCSR、野外-PCSR、野外 DS-PCSR、实验室-野外-PCSR、实验室-野外 DS-PCSR。模型稳健性采用 10 折交叉验证方法检验,精度评定指标包括可决系数  $R^2$ 、验证可决系数  $R_{cv}^2$ 、PRESS, RPD(residual predictive derivation)。PRESS 与 RPD 的计算公式如式(9)和式(10)所示, PRESS 值为样本预报误差的平方和, RPD 为预报样本标准差与均方根误差的比值,  $R^2$  与 RPD 越大、PRESS 越小,模型精度越高、偏差越小。

$$PRESS = \sum_{i=1}^n (\hat{y}_i - y_i) \quad (9)$$

$$RPD = \frac{SD}{RMSE} \quad (10)$$

## 2 结果与讨论

### 2.1 土壤样品重金属含量与光谱特性分析

46 个土壤样品的有机质含量、pH 值均值分别为 4.22% 和 5.49, Cd 含量描述性统计分析结果如表 1。数据表明,研究区土壤 Cd 污染整体较为严重,所有样品土壤 Cd 含量均显著高出国家土壤环境质量标准(GB15618—1995)一级污染标准( $0.3 \text{ mg} \cdot \text{kg}^{-1}$ )。与此同时,表 1 中的偏度、标准差、峰度等指标表明:研究区土壤 Cd 含量数据略微呈现右偏态,需通过 Box-Cox 变换使数据符合正态分布,其中参数  $\lambda = -0.113$ 。

表 1 土壤样本 Cd 含量统计特征( $\text{mg} \cdot \text{kg}^{-1}$ )

Table 1 Cd content statistic of soil samples

元素种类	极小值	极大值	均值	标准差	偏度	峰度
Cd	0.718	215.827	28.849	7.478	2.696	6.884

针对野外光谱在 400 nm 以下和 2 400 nm 以上波段信噪比过低的问题,以及 1 360~1 490 和 1 810~1 960 nm 波段的水汽吸收干扰问题,本研究在模型建立之前对此波段范围内的数据予以了剔除,同时考虑 DS 转换过程中野外与实验室光谱波长范围的一致性,本研究对实验室 1 360~1 490 和 1 810~1 960 nm 波段光谱数据也做了同样剔除处理<sup>[15]</sup>。问题波段剔除后所有样品实验室与野外原位原始和 SNV

(standard normal variate)变换光谱特征分别如图 2(a)和(b)所示。蓝色表示实验室光谱,红色表示野外原位光谱;实线表示样本光谱均值,虚线表示样本光谱的离散程度。图 2(c)显示了实验室与野外原位光谱差异,即两者反射率差值。蓝线与红线分别表示原始和 SNV 变换后野外原位光谱与相应实验室光谱的差异。对比分析图 2(a)和(c)可知,野外环境对 Cd 污染土壤光谱影响显著。在可见光区域,野外原位光谱与实验室光谱反射率几乎一致,但近红外区域的野外原位光谱反射率显著高于实验室光谱反射率,两者间差异随波长

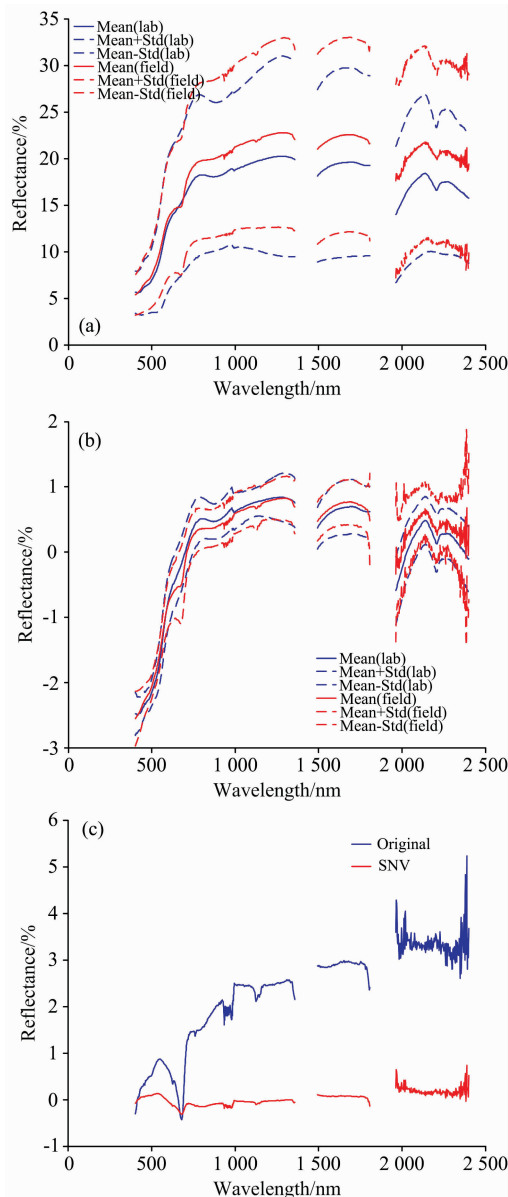


图 2 土壤样品实验室与野外原位光谱曲线

(a): 原始光谱; (b): SNV 光谱;  
(c): 实验室与野外光谱反射率差值

Fig. 2 Spectral of soil samples under laboratory and in-situ field conditions

(a): Original spectral; (b): SNV spectral;  
(c): Difference between laboratory and field spectral

呈线性趋势增加,且在 680, 960 和 1 120 nm 波段呈现明显的吸收特征,在 2 000 nm 之后光谱噪声明显增大。由图 2

表 2 实验室与野外原位光谱土壤 Cd 含量反演模型精度对比

Table 2 Precision comparison of soil Cd content estimating models based on laboratory and in-situ field spectral

	选入的主成分	$R^2$	$R_{cv}^2$	PRESS	RPD
实验室-PCSR	3 10 6 14 7	0.63	0.42	28.21	1.32
野外-PCSR	2 6 4 5	0.56	0.40	29.18	1.30
野外 DS-PCSR	3 6 14 15 9	0.66	0.50	25.16	1.40

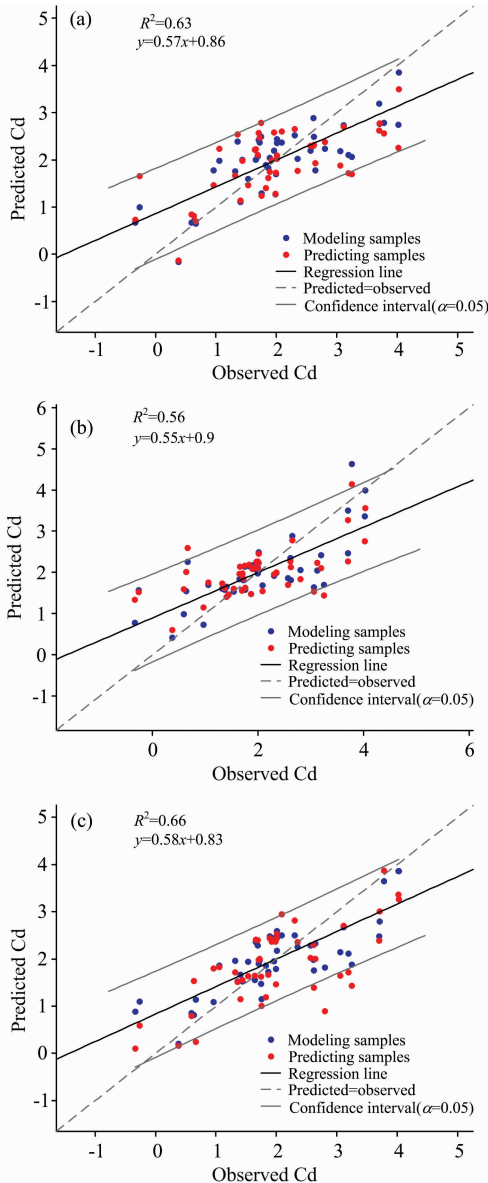


图 3 实验室与野外原位光谱独立反演土壤 Cd 含量散点图

(a): 实验室-PCSR; (b): 野外-PCSR; (c): 野外 DS-PCSR

Fig. 3 Scatterplots of predicted soil Cd content based on sole laboratory or in-situ field spectral

(a): Lab-PCSR; (b): Field-PCSR; (c): Field DS-PCSR

(b)和(c)可知,样品实验室与野外原位光谱均值与标准差在 SNV 变换后更为接近,两者间光谱差值和吸收特征差异显著减小,可有效提升后续参与建模的光谱数据质量。

## 2.2 独立反演建模

表 2 和图 3 是基于实验室光谱、野外原位光谱、野外原位 DS 光谱构建的 PCSR 模型精度验证结果。其中,实验室-PCSR 模型最终有效变量为第 3, 10, 6, 14 和 7 主成分。模型交叉验证预报误差平方和 PRESS 为 28.21,  $R^2$  和  $R_{cv}^2$  分别为 0.63 和 0.42, 验证集样本估计值基本落入置信区间范围 [图 3(a)], 模型总体可靠。但与此同时,对于较高 Cd 含量样本,图 3(a)中样本散点多位于 1:1 直线下方,模型对 Cd 含量较高样品总体低估。与实验室-PCSR 模型相比,野外-PCSR 模型最终有效变量为第 2, 6, 4 和 5 主成分,  $R^2$  和  $R_{cv}^2$  分别为 0.56 和 0.40, PRESS 为 29.18, 精度有所降低。对比

表 3 实验室野外原位光谱联合下的土壤 Cd 含量反演模型精度对比

Table 3 Precision comparison of soil Cd content estimating models based on joint laboratory and in-situ field spectral

	选入的主成分	$R^2$	$R_{cv}^2$	PRESS	RPD
实验室-野外-PCSR	2 28 5 26 8	0.36	0.05	53.37	0.96
实验室-野外 DS-PCSR	3 11 6 15 7	0.72	0.48	28.99	1.30

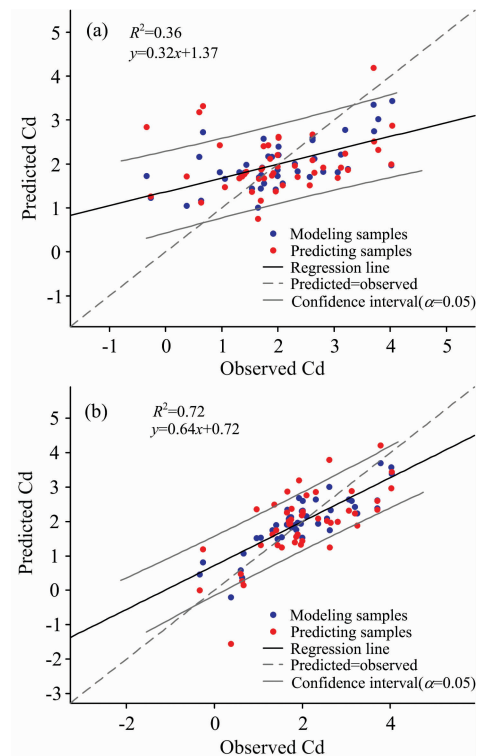


图 4 实验室与野外原位光谱土壤联合反演 Cd 含量散点图

(a): 实验室-野外-PCSR; (b): 实验室-野外 DS-PCSR

Fig. 4 Scatterplots of predicted soil Cd content based on joint laboratory and in-situ field spectral

(a): Lab-field-PCSR; (b): Lab-field DS-PCSR



图 3(b)和(a)可知,野外-PCSR 模型不仅类似实验室-PCSR 模型存在低估较高 Cd 含量样本的现象,而且还易高估较低 Cd 含量样本。而图 3(c)与图 3(a)和(b)的对比表明,基于野外原位 DS 光谱建立的野外 DS-PCSR 模型相比前两种模型精度有明显提升,  $R^2$ ,  $R_{cv}^2$  分别提升至 0.66 和 0.50, PRESS 值降低至 25.16。且相比野外-PCSR 模型,野外 DS-PCSR 模型与实验室-PCSR 模型在模型变量主成分方面更为相近(第 3, 6, 14, 15 和 9 主成分),两者光谱建模变量物理指示意义更相似(表 2),可靠性更强。

2.3 联合反演建模

表 3 和图 4 为基于野外原位光谱和野外原位 DS 光谱分别组合实验室光谱所建立的联合反演模型精度验证结果。表 4 为实验室光谱与野外原位光谱部分主成分间的相关性。表

4 表明,本研究中 Cd 污染土壤样品野外原位光谱第 2, 28, 5, 26 和 8 主成分分别与实验室 PCSR 模型变量主成分(第 3, 10, 6, 14 和 7 主成分)强相关。然而表 3 和图 4(a)中的精度指标指示,实验室-野外-PCSR 模型精度偏低,  $R^2$  和  $R_{cv}^2$  分别为 0.36 和 0.05, PRESS 值增至 53.37, 模型对土壤 Cd 含量的估计能力有限。仔细分析表 4 可知,实验室光谱第 10 主成分(Lab PC10)、第 14 主成分(Lab PC14)虽然分别与野外原位光谱第 28 主成分(Field PC28)、第 26 主成分(Field PC26)相关性较强,但两两间相关系数绝对值仍低于 0.5、难以具有相同的物理解释。分析原因,野外环境干扰使得野外原位光谱部分主成分丧失了其原有的物理解释力,导致真正有效主成分在建模过程中未能被模型识别从而最终影响建模精度。

表 4 野外原位光谱与实验室光谱部分主成分间的相关性

Table 4 Correlation between partial principal components of in-situ field spectral and laboratory spectral

	Field PC1	Field PC2	Field PC3	Field PC4	Field PC5	Field PC6	Field PC8	Field PC26	Field PC28
Lab PC1	<b>0.97*</b>	-0.11*	0.10*	0.13*	-0.04	0.00	0.01	0.00	-0.01
Lab PC2	-0.15*	-0.58*	<b>0.79*</b>	-0.05	-0.07	-0.07	0.05	-0.01	0.01
Lab PC3	0.00	<b>0.76*</b>	0.55*	-0.12*	-0.14*	-0.15*	-0.01	0.01	-0.02
Lab PC4	-0.15*	0.07	0.05	0.92*	0.02	-0.23*	0.03	0.01	0.02
Lab PC5	0.06	-0.08	-0.09	-0.12*	0.11*	<b>-0.77*</b>	-0.05	-0.02	-0.01
Lab PC6	0.02	0.00	0.09	-0.07	<b>0.69*</b>	-0.04	-0.42*	-0.07	-0.02
Lab PC7	0.05	0.02	-0.01	-0.15*	0.28*	-0.23*	<b>0.57*</b>	0.13	0.05
Lab PC10	0.03	-0.05	-0.02	-0.02	-0.06	-0.04	-0.03	-0.04	<b>(0.26)*</b>
Lab PC14	-0.01	-0.04	-0.09	-0.07	-0.16*	-0.19*	0.02	<b>(-0.19)*</b>	0.00

野外原位光谱主成分与实验室光谱相关性最强的用加粗数字表示,其中绝对值不超过 0.5 的用()表示; \* 表示  $p < 0.01$

表 5 不同转换集样本数下实验室-野外 DS-PCSR 模型精度

Table 5 Precision of lab-field DS-PCSR models under different number of transfer sets

$t$	实验室-PCSR 模型				实验室-野外 DS-PCSR 模型			
	选入主成分	PRESS	$R^2$	$R_{cv}^2$	选入主成分	PRESS	$R^2$	$R_{cv}^2$
10					3 10 6 14 7	50.68	0.18	0.02
15					3 10 6 14 7	60.60	0.32	0.03
20					4 10 5 14 7	50.44	0.28	0.11
21					3 10 6 14 7	35.16	0.44	0.28
22					3 11 6 15 7	48.38	0.52	0.27
23					3 11 6 15 7	38.76	0.50	0.29
24					3 11 6 16 7	34.16	0.59	0.35
25					3 11 6 15 7	36.97	0.53	0.28
26					3 11 6 16 7	39.43	0.50	0.29
27	3 10 6 14 7	28.21	0.63	0.42	3 11 6 14 7	46.38	0.61	0.29
28					3 11 6 15 7	40.09	0.72	0.38
29					3 11 6 15 7	28.99	0.72	0.48
30					3 11 6 13 7	31.74	0.69	0.40
31					3 11 6 13 7	35.45	0.65	0.31
32					3 11 6 13 7	35.41	0.64	0.32
33					3 11 6 13 7	33.96	0.63	0.36
34					3 10 6 13 7	33.35	0.67	0.36
35					3 10 6 14 7	30.64	0.60	0.39

野外原位光谱 DS 变换可有效消除野外环境对土壤样品原位光谱的干扰。表 5 是基于实验室光谱的野外原位光谱 DS 变换过程中转换集样本数量  $t$  变化下, 实验室-野外 DS-PCSR 模型变量及估计精度差异。表 5 表明, 第 3, 6, 11 和 7 主成分在  $t$  取值大于 20 时可稳定作为模型有效变量, 模型精度指标 PRESS 值在  $t$  取值 29 时趋于稳定。表 6 显示了最优转换集样本数 ( $t=29$ ) 时实验室光谱与野外原位 DS 光谱主成

分之间的相关性。此时, 实验室光谱主成分 Lab PC10, Lab PC14 与野外原位 DS 光谱主成分间的最大相关系数分别从 0.26 和 -0.19 增加至 0.86 和 -0.67, 野外原位 DS 光谱主成分的物理解释能力显著提升。与此同时, 表 3 中的  $R^2$  和  $R_c^2$  分别提高至 0.72 和 0.48, PRESS 值降低到 28.99; 图 4 (b) 中土壤样品校准集样本散点也大致沿 1:1 直线分布, 高含量 Cd 样本预测偏差明显降低, 模型精度可接受。

表 6 野外原位 DS 光谱 (29 个转换集) 与实验室光谱部分主成分间的相关性

Table 6 Correlation between partial principal components of in-situ field DS spectral and laboratory spectral

	Field PC1	Field PC2	Field PC3	Field PC4	Field PC5	Field PC6	Field PC7	Field PC11	Field PC15
Lab PC1	<b>1.00*</b>	0.06	-0.04	-0.05	0.03	0.00	0.00	0.01	0.00
Lab PC2	-0.05	<b>0.97*</b>	-0.15*	0.16*	0.00	-0.02	0.01	-0.01	-0.01
Lab PC3	0.01	0.14*	<b>0.98*</b>	0.14*	-0.02	-0.07	0.03	0.00	0.01
Lab PC4	0.05	-0.16*	-0.13*	<b>0.97*</b>	-0.07	0.10*	0.00	-0.02	0.00
Lab PC5	0.04	0.01	0.00	-0.06	<b>-0.99*</b>	0.03	0.01	-0.06	0.00
Lab PC6	0.00	0.04	0.06	-0.07	0.06	<b>0.90*</b>	0.30*	-0.08	-0.02
Lab PC7	0.00	0.02	0.05	-0.02	0.02	0.27*	<b>-0.95*</b>	0.00	-0.02
Lab PC10	0.00	0.00	-0.02	0.02	-0.01	-0.06	0.04	<b>0.86*</b>	0.08
Lab PC14	0.00	0.00	0.02	-0.01	0.01	-0.03	0.01	0.04	<b>-0.67*</b>

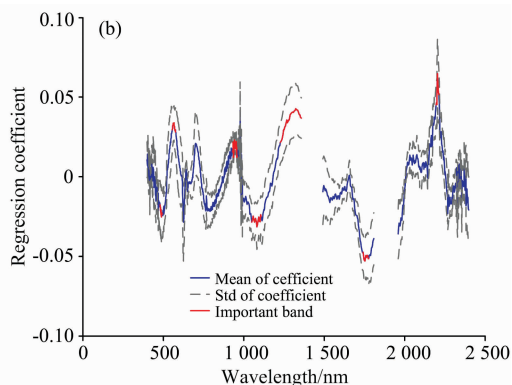
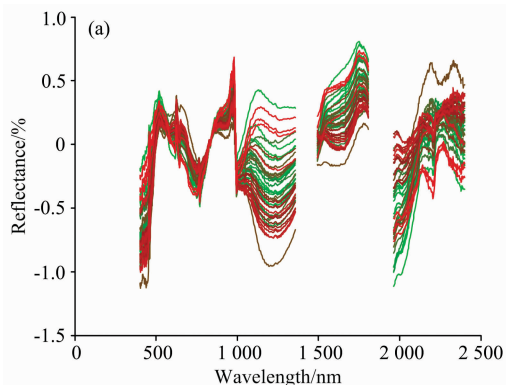
野外 DS 光谱主成分与实验室光谱相关性最强的用加粗数字表示; \* 表示在  $p < 0.01$

## 2.4 特征波段识别

模型变量与回归系数主成分逆变换是重建变量光谱、获取各波段贡献强度、进而识别特征波段的一种有效方式。图 5 左列为本研究中五类模型变量重建光谱, 其中红色、绿色线分别表示高、低浓度 Cd 样本重建光谱曲线; 右列为各模型回归系数, 蓝色曲线为 10 次交叉验证回归系数均值, 灰色虚线为回归系数标准差范围, 红色部分为识别的特征波段。

图 5(a) 表明, 实验室-PCSR 模型变量重建光谱在 500 nm 以及 1 000~2 200 nm 附近的反射率随样本变化呈现明显差异。图 5(b) 中回归系数在 488, 568, 935~955, 1 050~1 100, 1 225~1 358, 1 753 和 2 206 nm 位置达到峰值, 其中 1 753 和 2 206 nm 波段贡献最大; 而 703 以及 979 nm 波段因回归系数因标准差范围较大, 难以作为土壤 Cd 含量反演的稳定特征波段。图 5(c) 表明, 因模型变量为包含土壤光谱主要信息的第 2, 4, 5 和 6 主成分, 野外-PCSR 模型变量重建光谱的形态特征与原始光谱并无明显区别; 同时图 5(d) 中野

外 PCSR 模型仅揭示 539, 677, 1 000 和 2 198 nm 波段对土壤 Cd 含量变化具有较强指示作用, 与实验室-PCSR 模型中识别的特征波段明显不同。图 5(e) 和 (f) 可知, 野外 DS-PCSR 模型变量重建光谱在 493~732 以及 2 236 nm 处的吸收深度明显增强, 1 109~1 358 nm 位置的反射率曲线形态发生了改变; 但相对实验室-PCSR 模型, 568 和 2 206 nm 波段敏感性显著降低, 446, 650 以及 2 027~2 101 nm 附近波段对预测土壤 Cd 含量的贡献较大。图 5(g) 表明实验室-野外-PCSR 模型重建光谱形态与实验室-PCSR 模型有显著差异, 且图 5(h) 中大部分波段回归系数接近 0、标准差范围较大, 模型对于土壤 Cd 含量变化的特征波段识别均不敏感。图 5(i) 显示实验室-野外 DS-PCSR 模型变量重建光谱形态与实验室-PCSR 模型具有相似性, 两者物理意义较为一致, 图 5(j) 中回归系数表明 482, 565, 979 和 2 206 nm 波段对模型贡献最大, 与实验室-PCSR 模型识别的特征波段较为接近。



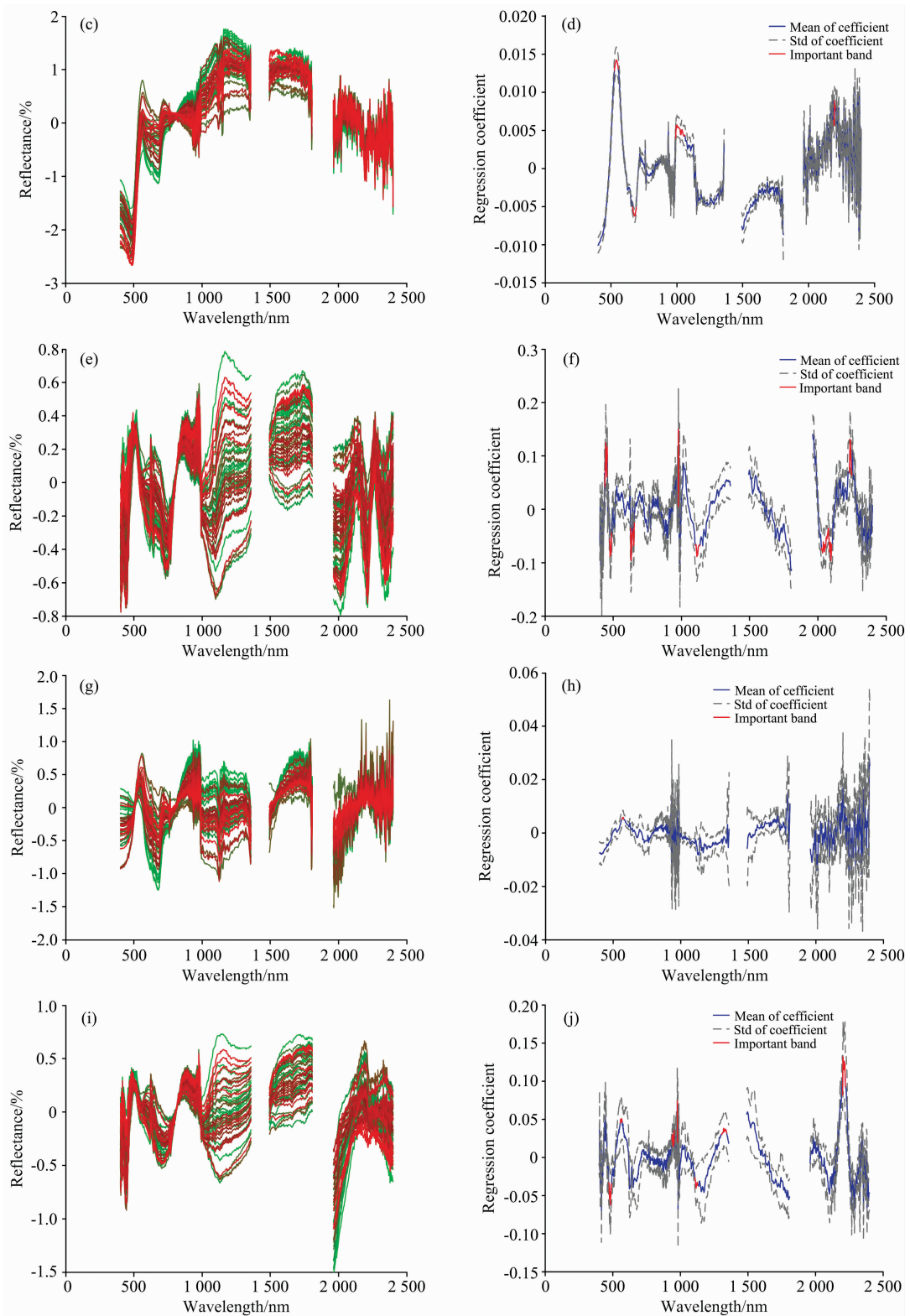


图 5 5 种基于实验室与野外原位光谱的土壤 Cd 含量反演模型变量重建光谱(左列)与模型回归系数(右列)

(a), (b): 实验室-PCSR; (c), (d): 野外-PCSR; (e), (f): 野外 DS-PCSR; (g), (h): 实验室-野外-PCSR; (i), (j): 实验室-野外 DS-PCSR

Fig. 5 Reconstruct spectral (left column) and regression coefficient (right column) of five models with different laboratory and in-situ field spectral

(a), (b): Lab-PCSR; (c), (d): Field-PCSR; (e), (f): Field DS-PCSR; (g), (h): Lab-field-PCSR; (i), (j): Lab-field DS-PCSR

对比前人研究, Siebielec 等发现 490~580 nm 光谱与土壤 Cd 含量高度相关<sup>[16]</sup>, 宋练等认为 2 260 与 2 210 nm 光谱反射率比值能够反演土壤 Cd 含量<sup>[17]</sup>, 本研究进一步证实了土壤 Cd 含量与土壤高光谱间的联系。研究发现实验室与野外原位光谱具有相同物理解释意义的波段主要集中在 482, 565, 979 以及 2 206 nm 波段附近, 其中 979 和 2 206 nm 光谱与土壤中氧化物和有机质的特征波段十分相近<sup>[9]</sup>。这一结果实际上进一步确认了程航等人发现的土壤 Cd 含量与有机质光谱特征波段高度相关的结论<sup>[18]</sup>。当然, 受文献报道有限的局限性, 本研究暂且仅能基于土壤 Cd 含量与有机质光谱特征波段高度相关的发现对特征波段予以一定的物理解释, 对土壤类型与土壤其他理化参数的顾及或许是进一步提升野外原位光谱反演土壤 Cd 含量精度的另一突破点。

### 3 结 论

基于重金属污染土壤样品的实验室、野外原位光谱测

量, 以及 Cd 含量测定, 本研究联合实验室和 DS 算法转换后的野外原位光谱开展了土壤 Cd 含量反演研究, 探究了运用野外原位光谱反演土壤重金属 Cd 的可行性, 为大范围普适性土壤重金属含量遥感高光谱反演理论与方法的突破提供了新的例证。主要结论是: (1) 由于野外环境因素的影响, 土壤原位高光谱反演重金属精度相比实验室光谱反演有所降低; (2) DS 方法能够大大降低野外环境噪声对土壤光谱的影响, 提升野外原位光谱在土壤重金属含量反演中的物理解释意义; (3) 联合实验室光谱先验知识, 以与实验室反演模型变量相关性最强为准则, 选择野外 DS 光谱主成分建立的模型能有效反演土壤重金属 Cd 含量, 且识别出贡献度较大的特征波段。当然, 因样本数量以及采样条件的限制, 本研究所得土壤重金属 Cd 含量高光谱反演模型的空间迁移性与普适性有待在其他区域或其他土壤类型条件下进一步验证。

### References

- [1] Gu Y W, Li S, Gao W, et al. *Acta Ecologica Sinica*, 2015, 35(13): 4445.
- [2] Shi T Z, Chen Y Y, Liu Y L, et al. *Journal of Hazardous Materials*, 2014, 265(2): 166.
- [3] Jiang Xiaolu, ZOU Bin, Feng Huihui, et al. *Journal of Geochemical Exploration*, 2019, 196: 1.
- [4] TAO Chao, WANG Ya-jin, ZOU Bin, et al(陶超, 王亚晋, 邹滨, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2018, 38(6): 1850.
- [5] Liu Y L, Li W, Wu G F, et al. *Geo-Spatial Information Science*, 2011, 14(1): 10.
- [6] Ackerson J P, Morgan C L S, Ge Y. *Geoderma*, 2017, 286: 131.
- [7] Lagacherie P, Baret F, Feret J B, et al. *Remote Sensing of Environment*, 2008, 112(3): 825.
- [8] Xu C, Zeng W, Huang J, et al. *Remote Sensing*, 2016, 8(1): 42.
- [9] Choe E, Meer F V D, Ruitenbeek F V, et al. *Remote Sensing of Environment*, 2008, 112(7): 3222.
- [10] Song L, Jian J, Tan D J, et al. *International Journal of Applied Earth Observation & Geoinformation*, 2015, 34(1): 1.
- [11] Ji W, Viscarra Rossel R A, Shi Z. *European Journal of Soil Science*, 2015, 66(4): 670.
- [12] CHEN Yi-yun, QI Kun, LIU Yao-lin, et al(陈奕云, 漆 锟, 刘耀林, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2015, 35(6): 1705.
- [13] TU Yu-long, ZOU Bin, JIANG Xiao-lu, et al(涂宇龙, 邹 滨, 姜晓璐, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2018, 38(2): 575.
- [14] Zhai L, Li S, Zou B, et al. *Atmospheric Environment*, 2018, 181: 145.
- [15] Piekarczyk J, Ka? mierzowski C, Królewicz S, et al. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 2016, 9(2): 827.
- [16] Siebielec G, Mccarty G W, Stuczynski T I. *Journal of Environmental Quality*, 2004, 33(6): 2056.
- [17] SONG Lian, JIAN Ji, TAN De-jun, et al(宋 练, 简 季, 谭德军, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2014, 34(3): 812.
- [18] Cheng Hang, Shen Ruili, Chen Yiyun, et al. *Geoderma*, 2019, 336: 559.



# Estimation of Cd Content in Soil Using Combined Laboratory and Field DS Spectroscopy

ZOU Bin, TU Yu-long, JIANG Xiao-lu, TAO Chao, ZHOU Mo, XIONG Li-wei

The Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring (Central South University), Ministry of Education, School of Geoscience and Info-Physics, Changsha 410083, China

**Abstract** Theoretically, hyperspectral remote sensing aided content estimation of soil heavy metal can greatly reduce the cost of conventional chemistries. As a result, hyperspectral remote sensing is gradually becoming a key technology to effectively explore the spatial distribution of soil heavy metal and consequently guide the prevention and remediation of heavy metal polluted soil. However, currently reported hyperspectral retrieval models for soil heavy metal estimation are mostly with laboratory spectra under specifically controlled conditions. Due to the impacts of environmental factors, such as illumination, soil moisture content, and roughness on in-situ field spectra, the wide implementation of in-situ field spectra based remote sensing detection of soil heavy metal is still experiencing the difficulty of reliability. For this, 46 soil samples were firstly collected from a mining area in Hengyang of Hunan Province, China. Then the spectra (ranged 350~2 500 nm) and Cd content of these soil samples were measured using ASD field spectrometer and ICP-atomic emission spectrometry under in-situ field and laboratory conditions, respectively. Then, considering the prior knowledge of laboratory spectra, the principal stepwise regression method was used to develop the Cd content estimation model based on combined laboratory and direct standardization (DS) algorithm transformed in situ field DS spectra with the model robust test by cross-validation. In order to further prove the effectiveness of the model with combined laboratory and DS transformed in-situ field DS spectra, the performance of this model was then compared with four types of hyperspectral remote sensing models including those with spectra from the laboratory, in-situ field, DS transformed in-situ field only, as well as with combined laboratory and in-situ field, one by one. The result shows that while the precision of the hyperspectral remote sensing model with in-situ field spectra ( $R^2=0.56$ ) is lower than the one with laboratory spectra ( $R^2=0.64$ ), the precision of the model with DS transformed in-situ field spectra is improved ( $R^2=0.66$ ). The model with combined laboratory and DS transformed in-situ field spectra is the one with the highest accuracy ( $R^2, 0.72$ ). Meanwhile, this highest robust model discloses that the wavebands located at 482, 565, 979, and 2 206 nm have significantly strong correlations with the soil Cd content. And this result is physically consistent with the model with laboratory spectra. In summary, results in this study confirm the role of the prior knowledge of laboratory spectra and DS algorithm in enhancing the reliability of in-situ field spectra based hyperspectral remote sensing model for soil Cd content estimation. It could provide new theoretical and methodological evidence for the development of soil Cd content estimation by using hyperspectral remote sensing.

**Keywords** Soil; heavy metal; Hyperspectral; DS method; Remote sensing retrieval

(Received Aug. 23, 2018; accepted Dec. 12, 2018)