

## 基于高光谱反射率的喀斯特地区土壤重金属锌元素含量反演

王金凤<sup>1, 2, 5</sup>, 王世杰<sup>2, 3</sup>, 白晓永<sup>2, 3\*</sup>, 刘方<sup>1</sup>, 路茜<sup>1, 2</sup>, 田诗琪<sup>2, 4</sup>, 王明明<sup>2</sup>

1. 贵州大学资源与环境工程学院, 贵州 贵阳 550001
2. 中国科学院地球化学研究所环境地球化学国家重点实验室, 贵州 贵阳 550001
3. 中国科学院普定喀斯特生态系统观测研究站, 贵州 普定 562100
4. 贵州师范大学地理与环境科学学院, 贵州 贵阳 550001
5. 六盘水师范学院旅游与历史文化学院, 贵州 六盘水 553004

**摘要** 针对传统土壤重金属锌元素含量测定效率低下和喀斯特地区山高坡陡土壤样品采集难度大, 亟需先进手段获取土壤重金属锌元素含量的要求, 以喀斯特流域为研究区, 利用电感耦合等离子质谱测定土壤样品的锌元素含量和分光光度计采集土壤光谱数据。将所测定的原始光谱, 经过连续统去除、一阶、二阶微分、倒数、倒数对数、倒数对数一阶、倒数对数二阶微分 7 种数学变换, 基于高光谱吸收重金属元素的特征吸收带初步判断光谱特征变量, 利用相关分析进一步筛选特征变量, 运用逐步回归最终确定有效建模光谱变量。采用非线性和线性算法, 揭示光谱敏感波段反射率与重金属锌元素含量之间的映射关系, 进行土壤重金属含量估测。结果表明: 基于耦合的光谱特征变量甄选方式, 锌元素的特征波段 580, 810, 1 410, 1 910, 2 160, 2 260, 2 270, 2 350, 2 430 nm 与铁氧化物、有机质、粘土矿物吸收带关联, 表明一定程度上捕捉到喀斯特地区土壤重金属锌元素的光谱吸收特性; 运用随机森林、支持向量机、偏最小二乘 3 种算法进行元素含量与光谱变量建模后, 采用决定系数和均方根误差评价模型精度。从光谱变换形式和模型性能二个维度综合判断, 基于二阶微分变换的随机森林算法准确度最高, 为最佳估算模型。通过高光谱反射率估测重金属锌元素含量, 实现了喀斯特地区土壤重金属锌元素含量的高效快速反演, 为喀斯特地区重金属元素含量动态监测提供了可靠的技术支撑。

**关键词** 土壤重金属锌; 光谱反射率; 特征波段

中图分类号: O657.3 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2019)12-3873-07

## 引言

锌是普遍存在于土壤中的一种重金属元素, 随着工业化进程和人类活动的加剧, 土壤中重金属锌元素污染日益显著, 对人体健康和环境安全构成了重大威胁。传统土壤重金属通过大范围采样和长时间实验分析, 虽然元素含量测定精度高, 但是周期长、成本高, 且很难实现大面积土壤重金属含量动态监测。高光谱遥感的深入发展为快速获取土壤元素含量提供了新的途径, 为此国内外学者进行了大量研究。徐

良将等<sup>[1]</sup>利用高光谱反射率估算土壤水分、总氮、总磷、有机质等土壤物质的含量; Riedel Frank 等<sup>[2]</sup>利用室内土壤反射光谱和实验室测定的土壤 Al, As, Ca, Cu, Fe, K, Mn, Na, Ni, Pb 和 Zn 含量, 采用线性的偏最小二乘回归算法进行了元素含量估测; 江振蓝等<sup>[3]</sup>利用地理权重回归模型进行了光谱预测土壤铬含量; 许多学者<sup>[4-5]</sup>还采用了随机森林、支持向量机、极限学习机、神经网络等非线性智能算法, 建立了重金属元素含量与高光谱估算模型, 实现了重金属含量的反演。

尽管国内外在光谱反演重金属含量方面取得了大量的研

收稿日期: 2018-10-24, 修订日期: 2019-02-22

基金项目: 国家重点研发计划项目(2016YFC0502102, 2016YFC0502300), “西部之光”人才培养计划(A类), 中国科学院科技服务网络计划项目(KFJ-STIS-ZDTP-036), 国际合作局国际伙伴计划项目(132852KYSB20170029, 2014-3), 贵州省高层次创新型人才培养计划“十”层次人才项目(黔科合平台人才[2016]5648), 喀斯特科学研究中心联合基金项目(U1612441), 国家自然科学基金项目(41571130074, 41571130042), 贵州省科技计划项目(2017-2966)资助

作者简介: 王金凤, 女, 1987年生, 贵州大学资源与环境工程学院博士研究生 e-mail: jfwanggz@126.com

\* 通讯联系人 e-mail: baixiaoyong@126.com

究成果,但还存在几个问题需要解决,一是在土壤采样难度大的喀斯特地研究较少,实际上这些地区更需要先进的手段实现土壤重金属含量的快速监测。二是光谱全波段代入模型进行反演,不能辨识到不同元素独有的特征光谱,难以实现不同元素含量的有效反演。三是基于光谱反射率采用不同的方法对土壤重金属含量估算,结果差异巨大。鉴于此,本文以喀斯特地区贵州省后寨河流域为研究区,室内测定样品重金属锌元素含量和光谱反射率,预处理原始光谱,甄选重金属锌元素光谱特征波段,运用随机森林、支持向量机、偏最小二乘回归对重金属锌元素含量进行反演,通过比较确定最优估算模型,为喀斯特地区土壤重金属元素的反演预测提供技术手段和理论支撑。

## 1 研究区及实验分析

### 1.1 研究区概况

后寨流域位于长江流域和珠江流域的分水岭地区,地处 $105^{\circ}41'—105^{\circ}48'E$ , $26^{\circ}13'—26^{\circ}17'N$ ,总面积约 $75\text{ km}^2$ ,平均海拔高度为 $1\ 300\text{ m}$ ,流域范围内大于 $6^{\circ}$ 和小于 $6^{\circ}$ 的地势各占一半,具有典型的山坝耦合特征。气候属亚热带季风湿润气候,年平均气温 $15.0^{\circ}\text{C}$ ,年平均降水 $1\ 152.20\text{ mm}$ 。流域出露的地层为三叠系关岭组,地层产状平缓,岩性以白云

岩为主。土地利用方式以旱地、水田、林地、建设用地为主。

### 1.2 样品采集与实验测定

依据研究区自然地理条件和样点布设要求进行野外表层土壤样品采集(图1),共采集土壤样品100个。采样深度为 $0\sim 20\text{ cm}$ ,土壤样品重约 $1\text{ kg}$ ,将土壤样品带回室内自然风干,研磨并过 $200$ 目尼龙筛。每个样品分成两份,一份用于实验室化学分析,另一份用于室内土壤光谱测定。土样经盐酸-硝酸-氢氟酸-高氯酸微波消解后,采用电感耦合等离子质谱(ICP-MS)测定土壤中重金属锌元素的含量。为保证分析质量,用国家标准样本 GSB04-1767-2004 进行质量控制。土壤光谱测定采用紫外-可见-近红外分光光度计 Cary 5000 在室内进行,实验装置由计算机、光谱仪、积分球等组成。采用漫反射方式测定光谱,使用标准样采集基线,通过基线校正得到样品相对光谱反射率。光谱波段测定范围为 $500\sim 2\ 500\text{ nm}$ ,扫描间隔为 $1\text{ nm}$ ,从波长高端往低端扫描,记录不同波长对应的光谱反射率,光谱曲线经3次重复采集取平均值,数据以ASCII码形式导出。在光谱测定的过程中,由于仪器或环境因素影响,会产生不可预见的误差,为了减小估计误差,利用 The Unscrambler 软件进行光谱特性 PCA 分析,从样本中剔除3个光谱异常样本,剩下97个土壤反射光谱样本。

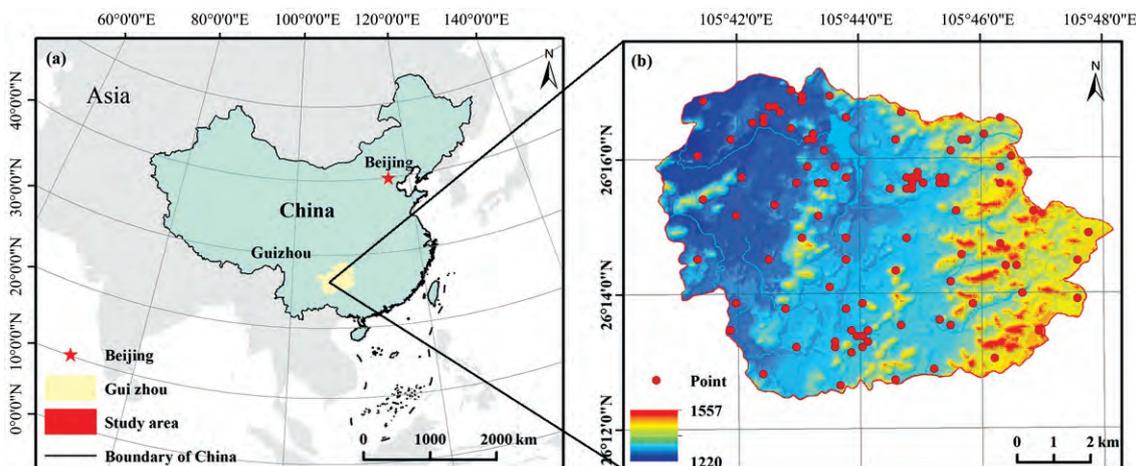


图1 研究区位置及采样点分布图

Fig 1 Location of the study area and distribution of soil sampling sites

## 2 研究方法

### 2.1 光谱数据处理

采用 Savitzky Golay 对原始光谱数据进行平滑处理(smoothness reflectivity, R)后,对其进行连续统去除(continuum removed, CR),一阶微分(first derivative, FD),二阶微分(second derivative, SD),倒数变换(reciprocal transformation, RT)、倒数对数变换(reciprocal logarithmic transformation, AT)、倒数对数一阶微分(first derivative of reciprocal logarithmic, AFD)、倒数对数二阶微分(second derivative of reciprocal logarithmic, ASD)等数学变换,增强有效光谱信息。

### 2.2 反演模型

随机森林(random forests, RF)是由 Leo Breiman 提出的一种基于统计学非线性组合智能算法。RF 实质是对决策树算法的一种改进,预测结果由投票或求平均的方法得出,其优势在于操作简便,对噪声有很高的容忍度,不容易产生过度拟合问题<sup>[6]</sup>。

支持向量机(support vector machines, SVM),是基于统计学习 VC 维理论和结构风险最小原理基础上发展出来的一种非线性新型智能算法。核心思想是将复杂的低维度非线性数据变为高维空间的线性问题,然后寻找最优分离超平面,再将超平面映射到低维空间<sup>[7]</sup>。

偏最小二乘回归(partial least squares regression,

PLSR), 作为一种数学优化技术, 主要基于主成分提取思想, 通过计算变量最小化误差的平方和, 使得拟合对象能够不断接近目标对象<sup>[8]</sup>。

2.3 模型检验

模型的稳定性用决定系数  $R^2$  的大小来检验, 模型的精度用均方根误差 RMSE 来检验。 $R^2$  越大、RMSE 越小, 模型的性能越好。

3 结果与分析

3.1 重金属锌元素含量基本特征分析

研究区 97 个样本数据分为训练样本和测试样本两组, 78 个训练样本用于建立拟合模型, 19 个测试样本用于模型验证。由表 1 可见, 研究区土壤重金属锌元素含量在 60.73 ~ 400  $\text{mg} \cdot \text{kg}^{-1}$  之间, 元素含量变化幅度较大。变异系数为 52%, 具有较强的空间异质性。训练和测试样本二者之间元素含量的平均值和标准差基本一致, 训练样本所建立的预测

模型能较好地适用于测试样本。97 个样本中有 58 超过国家一级标准自然标准值(GB 15618—1995), 表明后寨河流域土壤安全受到重金属锌元素的影响。

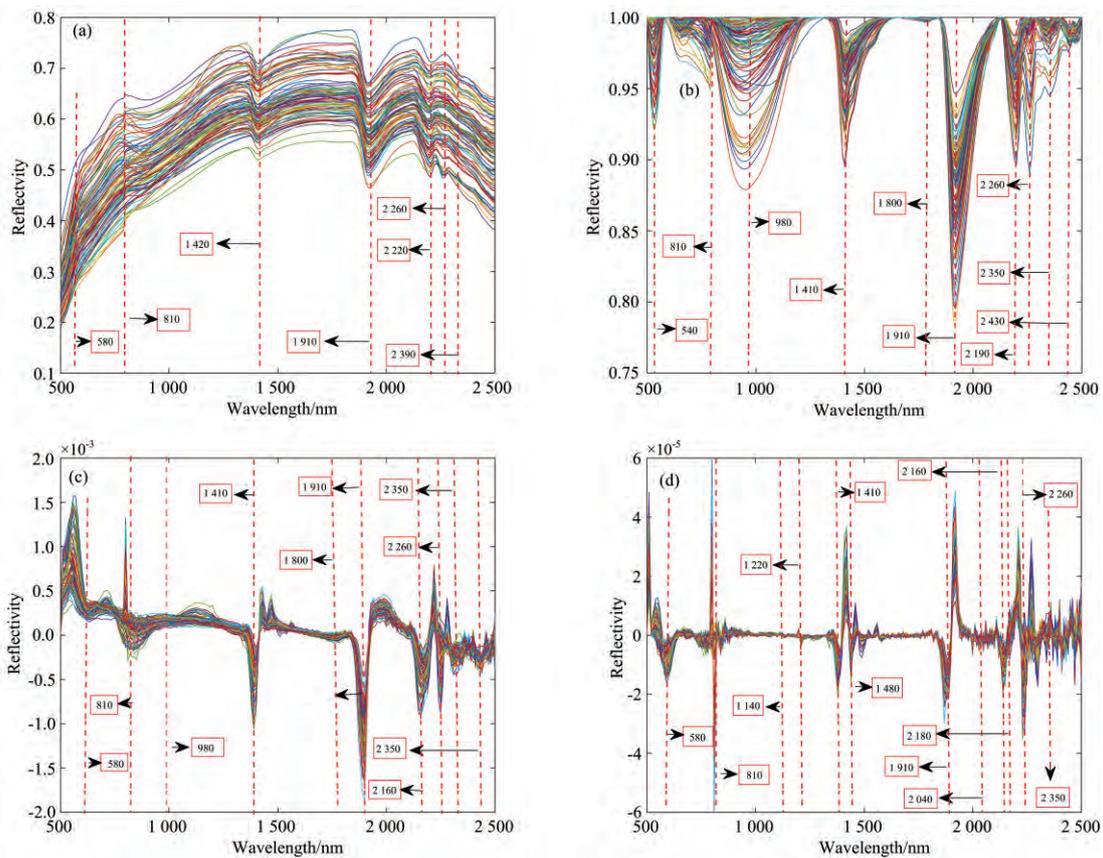
3.2 重金属锌元素光谱特征波谱甄选

从图 2(a)可以看出重金属锌元素原始光谱曲线的形状基本相同, 光谱曲线整体呈弧形, 在 500~1 000 nm 范围内光谱吸收带与有机质和氧化铁相关。1 000~2 200 nm 范围内存在 3 个明显的光谱吸收带, 1 400 和 1 900 nm 两处的吸收峰除了由羟基和层间水分引起, 还与有机物和粘土矿物吸附锌元素有关。2 200 nm 左右的吸收峰是由矿物晶格中水引起的<sup>[9]</sup>。2 200~2 500 nm 之间吸收峰与土壤碳酸盐矿物中离子基团振动关联<sup>[10]</sup>。

基于有机质、粘土矿物、铁锰氧化物等对土壤重金属锌元素的吸附机理<sup>[11-12]</sup>, 结合原始及 7 种变换形式光谱曲线中重金属锌元素在这些结合点上最大吸收带初步判断锌元素特征波段。从原始光谱曲线上可以筛选出 580, 810, 1 420, 1 910, 2 220, 2 260 和 2 390 nm 这几个光谱特征吸收波段。

表 1 土壤重金属锌含量统计  
Table 1 Statistics of soil heavy metal zinc content

样本	数量	最大值/ ( $\text{mg} \cdot \text{kg}^{-1}$ )	最小值/ ( $\text{mg} \cdot \text{kg}^{-1}$ )	平均值/ ( $\text{mg} \cdot \text{kg}^{-1}$ )	标准差	变异系数	峰度	偏度
训练样本	78	400	65.91	130.31	69.68	0.53	4.64	1.70
测试样本	19	223	80.3	136.88	66.38	0.48	4.12	1.55
总样本	97	400	60.73	139.84	69.09	0.52	4.16	1.98



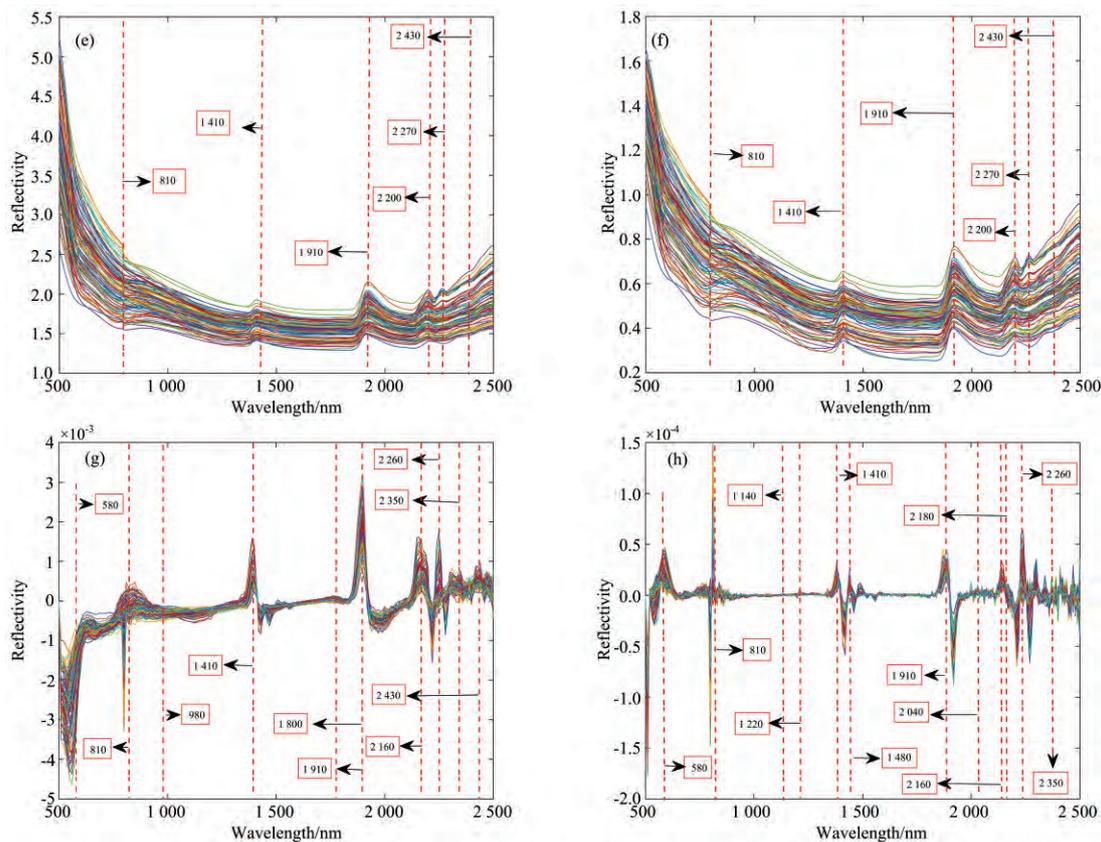


图 2 原始及 7 种变换形式的光谱数据

Fig 2 Spectra obtained after different transformations

表 2 重金属锌含量与各光谱形式初步特征吸收波段相关系数

Table 2 Correlation coefficients between absorption bands and heavy metal zinc content

变换形式	特征波段对应的相关系数
R	-0.08, -0.07, -0.27*, -0.30**, 0.35**, 0.32**, 0.31**
CR	-0.32**, 0.50**, -0.32*, -0.41**, -0.38**, -0.40**, -0.27, -0.183
FD	0.31**, 0.34**, 0.19, 0.34**, -0.28*, 0.33**, -0.38**, -0.46**, -0.33**, -0.27*
SD	-0.33**, 0.04, -0.15, -0.42**, 0.37**, -0.35**, 0.18, -0.14, 0.41**, 0.42**, 0.46**, -0.47**
RT	-0.12, 0.24, 0.38**, 0.32**, 0.32**, 0.34**
AT	-0.10, 0.29, 0.30*, 0.36**, 0.33**, 0.38**
AFD	-0.34**, 0.31**, 0.43**, 0.25, 0.27, 0.35**, 0.41**, 0.46**, 0.39**, 0.40**
ASD	0.34**, 0.01, 0.25, 0.42**, -0.40**, 0.36**, -0.15, 0.15, -0.31**, -0.26, -0.37**, 0.47**

注: \*\*表示 0.01 水平显著; \* 为 0.05 水平显著; 其余为不相关。

Note: \*\* and \* represent significant differences at 0.01 level and 0.05 level, respectively

经 CR 变换后的光谱曲线特征吸收波段集中在 540, 810, 980, 1410, 1800, 1910, 2190, 2260, 2350 和 2430 nm [图 2(b)]; 经 FD 变换后的光谱曲线吸收波段集中在 540, 810, 980, 1410, 1800, 1910, 2190, 2260, 2350 和 2430 nm [图 2(c)]; 经 SD 变换的光谱曲线特征吸收波段集中在 580, 810, 1140, 1220, 1410, 1480, 1910, 2040, 2160, 2180, 2260 和 2350 nm [图 2(d)]; RT 变换的光谱吸收波段集中在 810, 1410, 1910, 2200, 2270, 2430 nm [图 2(e)]. 经 AT [图 2(f)]、AFD [图 2(g)]、ASD [图 2(h)] 变换的光谱曲线特征波段与 FD, SD 和 RT 变换相同。

初步确定的土壤光谱特征吸收波段是根据土壤重金属光谱吸收特性的经验判断, 还需相关分析才能更加准确地揭示光谱变量与元素含量之间的关联。将原始光谱并结合 7 种变化形式初步确定的土壤光谱特征吸收波段与重金属含量进行相关分析, 如表 2 所示, 经过变换后的光谱数据的相关系数比原始反射率有所提高, 表明预处理一定程度上能更好去噪、增强信噪比, 突显有效光谱信息。为了提高模型精度, 达到更好的反演结果, 选择在 0.01 水平下极显著相关的波段作为土壤重金属锌特征波段。

相关系数筛选出的光谱特征波段虽与重金属含量有相关

性,但对重金属含量预测起重要作用的光谱特征波段还需拟合回归方程才能确定,将相关系数筛选出的光谱特征波段全部输入到逐步回归方程中,通过剔除回归建模过程中贡献率不显著的波谱,得到显著可以预测的变量(表 3)。

表 3 逐步回归确定的特征波段

Table 3 Characteristic bands selected by stepwise regression

变换形式	特征波段/nm
R	1 910, 2 220, 2 260
CR	810, 1 410, 1 910, 2 260
FD	580, 810, 1 410, 1 910, 2 160, 2 260, 2 350
SD	580, 1 410, 2 160, 2 180, 2 260, 2 350
RT	2 200, 2 270, 2 430
AT	1 910, 2 270, 2 430
AFD	580, 810, 1 410, 1 910, 2 160, 2 350, 2 430
ASD	580, 1 220, 1 410, 2 160, 2 260, 2 350

将逐步回归筛选的重复特征波段进行有效整合,最终确定重金属锌元素特征光谱变量为 580, 810, 1 410, 1 910, 2 160, 2 260, 2 270, 2 350 和 2 430 nm。

3.3 不同模型预测能力比较

将最终筛选出的重金属锌元素特征波谱信息,利用 RF,

SVM 和 PLSR 三种方法进行锌元素含量估算,并通过精度比较确定最优估算模型(表 3)。

从光谱不同变换形式 RF 算法检验结果来看。建模方面,通过 FD 光谱变换的建模精度最高,其  $R^2 = 0.94$ ,  $RMSE = 18.01$ 。建模精度最低的为 RT 变换后的模型( $R^2 = 0.85$ ,  $RMSE = 24.63$ );验证方面,基于 SD 变换的模型验证精度最高,最低的是未进行光谱变换的 R;虽然 FD 变换形式建模集有最大的  $R^2$  和最小的 RMSE,但其验证集的稳定性和精度却远不如 SD 变换形式建立的模型高。基于建模和验证结果评价参数的整体比较,RF 算法中,SD-RF 模型为最优估测模型。

SVM 算法中 R 和 CR 变换形式在 SVM 建模过程中具有较高的相关系数和较低的均方根误差,但是其验证结果并不理想,可能存在过度拟合。基于建模和验证结果的综合考虑,AT 变换形式和 R 所建立的模型精度较低,SD 变换形式建立的 SVM 模型明显优于其他变换形式所建立的模型。

在 PLSR 算法中,光谱经 SD 和 ASD 形式变换在建模和验证结果上优于其余 6 种光谱变换形式。根据模型精度评价指标,R 形式的 PLSR 模型精度最低,ASD 变换形式下的 PLSR 模型为不同变换形式下最优模型。

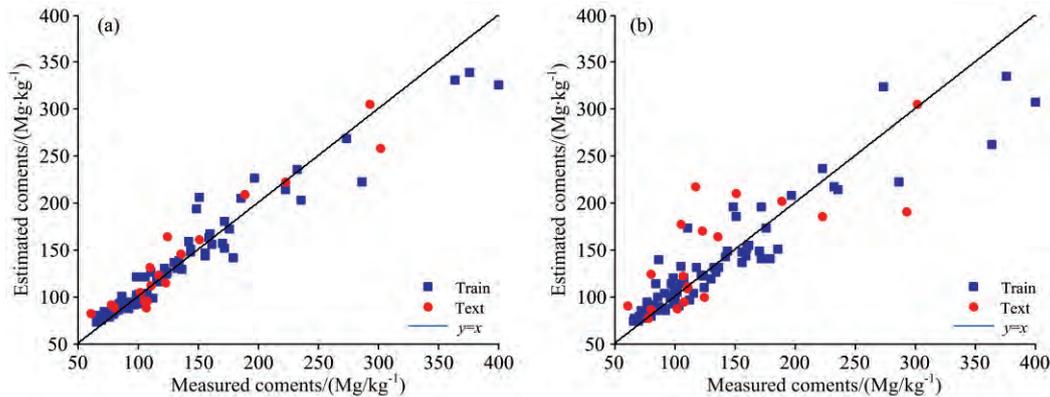
表 4 RF/SVM/PLSR 算法土壤锌含量建模和验证结果

Table 4 Modeling and verification results of soil zinc content by RF/SVM/PLSR algorithm

变换形式	RF 建模		RF 验证		SVM 建模		SVM 验证		PLSR 建模		PLSR 验证	
	$R^2$	RMSE	$R_p^2$	$RMSE_p$	$R^2$	RMSE	$R_p^2$	$RMSE_p$	$R^2$	RMSE	$R_p^2$	$RMSE_p$
R	0.88	22.02	0.59	38.93	0.84	15.87	0.49	52.4	0.61	43.21	0.49	52.53
CR	0.91	21.59	0.69	37.27	0.86	13.87	0.62	36.48	0.80	30.38	0.76	36.10
FD	0.94	18.01	0.75	29.24	0.70	39.82	0.73	31.04	0.80	30.05	0.75	31.64
SD	<b>0.94</b>	<b>18.02</b>	<b>0.93</b>	<b>18.49</b>	<b>0.83</b>	<b>20.87</b>	<b>0.80</b>	<b>25.48</b>	0.81	26.30	0.78	30.64
RT	0.85	24.63	0.59	36.85	0.70	37.16	0.71	36.67	0.65	36.28	0.68	40.64
AT	0.93	18.90	0.72	32.27	0.72	37.98	0.59	51.32	0.63	50.79	0.66	41.62
AFD	0.93	18.60	0.73	31.27	0.78	36.95	0.63	48.21	0.80	29.23	0.79	38.64
ASD	0.94	18.84	0.92	21.59	0.83	16.87	0.76	26.12	<b>0.81</b>	<b>27.38</b>	<b>0.80</b>	<b>27.10</b>

为了进一步比较 3 种算法估测重金属含量的准确性。从 3 种算法的最佳估算模型和最差估算模型的建模和验证评估指标可以看出,最佳估算模型整体性能  $SD-RF > SD-SVM > ASD-PLSR$ ,最差估算模型整体性能  $RT-RF > R-SVM > R-$

PLSR。因此,RF 是 3 种算法中预测能力最强和稳定性最好的算法,其次为 SVM,最差为 PLSR。同时,从 3 种最佳和最差估测模型的重金属锌元素回归散点图 3 可以看出,估测精度低的模型[图 3(b, d, f)]存在一个共同特点,即对于元



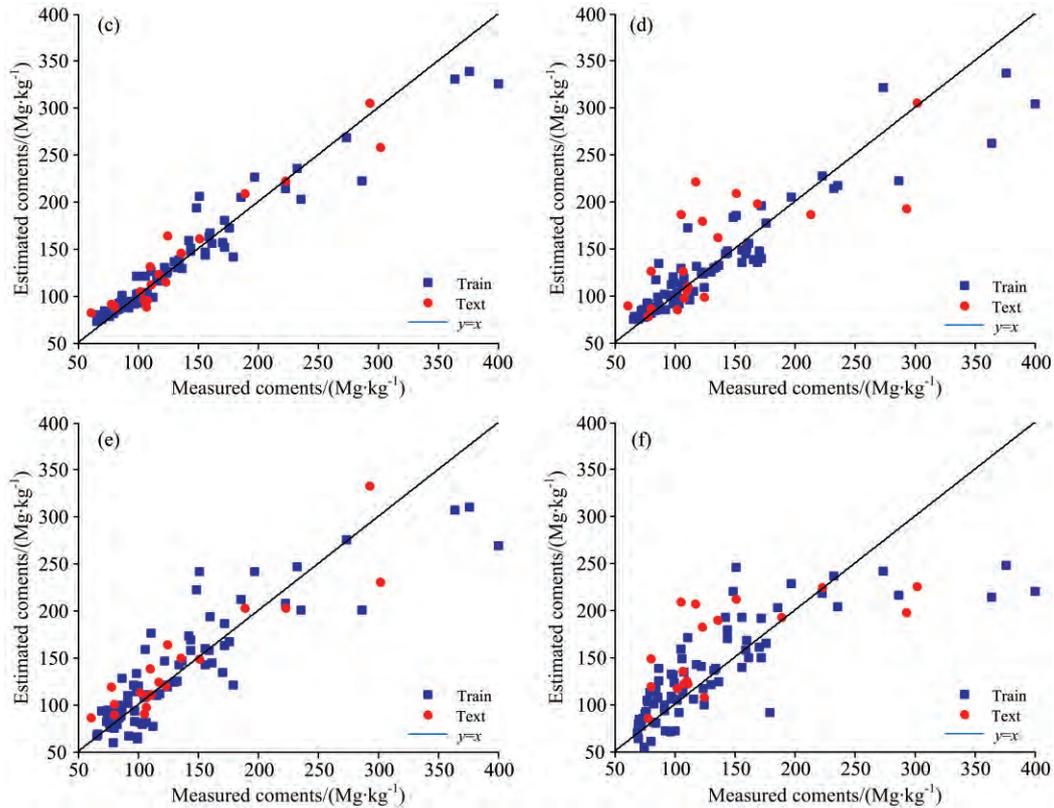


图 3 3 种最佳和最差估测模型的重金属锌元素回归散点图

Fig 3 The regression scatter plots of predictions for heavy metal zinc by the 3 best and the 3 worst models

素含量偏高的样本点, 预测值和实测值的偏差相对较大, 数据点偏离  $y=x$  直线。估测精度高的模型[图 3(a, c, e)], 提高了部分低浓度和中浓度重金属锌元素含量的预测能力, 数据点都较为靠近  $y=x$  直线两侧。SD-RF 模型样本点最接近  $y=x$  线, 能够更加准确地预测重金属锌元素的含量。因此, SD-RF 模型是研究区进行土壤重金属锌元素含量估测的最佳模型, 能更好地揭示光谱变量与土壤重金属锌元素含量之间的映射关系。

## 4 结 论

通过反射光谱与重金属特征吸收带初步判断光谱有效变量、利用相关分析筛选特征变量, 逐步回归剔除无效建模变量, 构建了耦合甄选光谱有效变量信息的方法体系。甄选出

的有效光谱变量 580, 810, 1 410, 1 910, 2 160, 2 260, 2 270, 2 350, 2 430 nm 与铁氧化物、有机质、粘土矿物等关联, 表明一定程度上捕捉到喀斯特地区土壤重金属锌元素的光谱吸收特性。

通过采用 RF, SVM 和 PLSR 三种算法, 分别建立的土壤重金属元素含量估算模型, 都能实现土壤重金属锌元素含量的反演, 其中 RF 是 3 种算法中预测能力最强的, 其次为 SVM, 最后为 PLSR。不同光谱变换形式与重金属含量拟合的模型精度, 原始光谱构建的模型性能差, 表明进行预处理以达到净化光谱噪声和突显光谱的重要性。7 种光谱变换形式中, 基于二阶微分变换的随机森林所建立的反演模型具有最强反演精度和预测能力, 为喀斯特地区高光谱反演土壤重金属锌元素的最好模型。

## References

- [ 1 ] XU Liang-jiang, HUANG Chang-chun, LI Yun-mei, et al(徐良将, 黄昌春, 李云梅, 等). Remote Sensing Technology and Application (遥感技术与应用), 2013, 28(4): 681.
- [ 2 ] Riedel F, Denk M, Müller L, et al. Geoderma, 2018, 315: 188.
- [ 3 ] JIANG Zhen-lan, YANG Yu-sheng, SHA Jin-ming(江振蓝, 杨玉盛, 沙晋明). Acta Ecologica Sinica(生态学报), 2017, 37(23): 8118.
- [ 4 ] Khosravi V, Ardejani F D, Yousefi S, et al. Geofisica Internacional, 2018, 31(8): 30.
- [ 5 ] Liu M, Liu X, Li J, et al. International Journal of Applied Earth Observation & Geoinformation, 2012, 19(19): 14.
- [ 6 ] Xia J, Chanussot J, Du P, et al. IEEE Transactions on Geoscience & Remote Sensing, 2015, 53(5): 2532.
- [ 7 ] Raczko E, Zagajewski B. European Journal of Remote Sensing, 2017, 50(1): 146.

- [ 8 ] Li S, Shao Q, Lu Z, et al. Spectrochim. Acta A: Mol. Biomol. Spectrosc. , 2018, 190: 284.
- [ 9 ] ZHANG Wei, GAO Xiao-hong, YANG Yang, et al(张 威, 高小红, 杨 扬, 等). Soil(土壤), 2014, (6): 1052.
- [10] CHENG Xian-feng, SONG Ting-ting, CHEN Yu, et al(程先锋, 宋婷婷, 陈 玉, 等). Acta Petrologica Et Mineralogica(岩石矿物学杂志), 2017, 36(1): 67.
- [11] Sun W, Zhang X. International Journal of Applied Earth Observation & Geoinformation, 2017, 58(1): 126.
- [12] YU Lei, HONG Yong-sheng, ZHOU Yong, et al(于 雷, 洪永胜, 周 勇, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2016, 32(13): 95.

## Prediction Soil Heavy Metal Zinc Based on Spectral Reflectance in Karst Area

WANG Jin-feng<sup>1, 2, 5</sup>, WANG Shi-jie<sup>2, 3</sup>, BAI Xiao-yong<sup>2, 3\*</sup>, LIU Fang<sup>1</sup>, LU Qian<sup>1, 2</sup>, TIAN Shi-qi<sup>2, 4</sup>, WANG Ming-ming<sup>2</sup>

1. College of Resource and Environment, Guizhou University, Guiyang 550001, China

2. State Key Laboratory of Environmental Geochemistry, Institute of Geochemistry, Chinese Academy of Sciences, Guiyang 550001, China

3. Puding Karst Ecosystem Observation and Research Station, Chinese Academy of Sciences, Puding 562100, China

4. School of Geographical and Environmental Sciences, Guizhou Normal University, Guiyang 550001, China

5. School of Tourism and Historical Culture, Liupanshui Normal University, Liupanshui 553004, China

**Abstract** In order to solve the problem of inefficiency in measuring heavy metal zinc content and soil samples collection difficulty with traditional way in karst area, it is greatly essential to get zinc content in soil by effective measures. The institutional area is a typical Karst region, soil zinc content as well as reflectance spectral of soil data were collected by inductively coupled plasma mass and Spectrophotometer. The reflectance spectra of measurement were handed by these steps. Firstly, 7 kinds of mathematical transformations were used including continuum removed, first differential, second differential, reciprocal transformation, absorbance transformation, first differential of absorbance, and second differential of absorbance. Secondly, spectral characteristic variables were determined based on the characteristic absorption band of spectral absorption of heavy metals. And then, further spectral characteristic variables were selected by correlation analysis. Finally, stepwise regression was used to determine the effective modeling spectral bands. Mapping relationships between Spectral bands and heavy metal zinc content were revealed by linear and nonlinear estimation algorithm, and the results aim to measure the heavy metal zinc in soil. It shows that the characteristic bands of zinc are associated with iron oxide, organic matter and clay mineral absorption band. It's focused on 580, 810, 1 410, 1 910, 2 160, 2 260, 2 270, 2 350, 2 430 nm, and these results reveal that the absorption characteristics of heavy metal zinc possible were captured in karst area. The models were funded by Random Forests, Support Vector Machines, Partial Least Squares Regression to precision evaluation by coefficient of determination and the root mean square error of prediction. The best estimation model was obtained from spectrum transformation and model performance. The algorithm of Random forests for second differential transformation has the highest accuracy and is chosen as the best model. The content of heavy metal zinc was estimated by spectral reflectance. It is a rapid, efficient method for indirect evaluation of zinc. It provides a technical support for the dynamic monitoring of heavy metal content in karst areas.

**Keywords** Soil heavy metal zinc; Spectral reflectance; Characteristic bands

(Received Oct. 24, 2018; accepted Feb. 22, 2019)

\* Corresponding author