

基于铁氧化物特征光谱和改进遗传算法反演土壤 Pb 含量

张霞¹, 王一博^{1,2}, 孙伟超¹, 黄长平¹, 张茂^{1,2}

(1. 中国科学院空天信息创新研究院, 北京 100101; 2. 中国科学院大学, 北京 100049)

摘要: 近年来, 高光谱的快速发展使野外实时监测土壤重金属含量成为可能。然而高光谱分辨率数据在提高信息量的同时也造成了信息冗余, 该研究针对光谱冗余问题, 提出一种基于铁氧化物特征光谱和改进遗传 (Improved Genetic Algorithm, IGA) 特征优选算法的反演方法: 依据 Pb 在土壤中的吸附机理, 提取土壤光谱中的铁氧化物特征波段用于 Pb 含量反演, 减少数据冗余的同时提高方法的机理性。改进遗传算法解决传统遗传算法 (Genetic Algorithm, GA) “过早收敛”的问题, 增强算法的有效光谱的提取能力。使用雄安新区农田野外土壤样本构建偏最小二乘回归模型 (Partial Least Square Regression, PLSR) 反演土壤 Pb 含量, 研究表明: 相对于全谱段建模, 基于铁氧化物特征波段的 IGA-PLSR 模型的 R^2 和 RPD 分别提升了 0.397、1.037, RMSE 下降了 1.958 mg/kg; 改进后的 IGA-PLSR 在运行初期能够跳出局部解区域寻找更加有效的光谱波段组合, 平均的 R^2 、RPD 分别为 0.822、2.377, RMSE 为 2.221 mg/kg, 相对于传统 GA-PLSR 算法的精度 (平均 R^2 、RPD 为 0.782、2.117, RMSE 为 2.487 mg/kg) 有显著提升。该研究表明, 从反演机理和波段选择算法两方面提出的反演方法有利于提高土壤 Pb 含量的估算精度。该研究为雄安地区农田土壤 Pb 含量的高光谱估算提供了参考。

关键词: 土壤; 反演; 高光谱; 重金属; 特征选择; 光谱分析

doi: 10.11975/j.issn.1002-6819.2020.16.013

中图分类号: TP97

文献标志码: A

文章编号: 1002-6819(2020)-16-0103-07

张霞, 王一博, 孙伟超, 等. 基于铁氧化物特征光谱和改进遗传算法反演土壤 Pb 含量[J]. 农业工程学报, 2020, 36(16): 103-109. doi: 10.11975/j.issn.1002-6819.2020.16.013 http://www.tcsae.org

Zhang Xia, Wang Yibo, Sun Weichao, et al. Inversion of Pb content in soil based on iron oxide characteristic spectrum and improved genetic algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(16): 103-109. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2020.16.013 http://www.tcsae.org

0 引言

土壤为自然生态系统的运转和人类活动提供了基本保障。随着工业化的不断发展, 通过污水排放、大气降尘等途径进入土壤的重金属元素在土壤中持续累积, 造成土壤重金属污染^[1]。Pb 是土壤中常见的一种重金属污染物, 渗透进土壤后移动性较差, 残留时间长, 且不易被微生物分解, 其治理和修复难度大, 不仅影响农作物的生长, 还会通过食物链威胁人类健康^[2]。因此, 广泛开展土壤 Pb 含量调查和监测, 对保障农业生产、保护人类健康和生态系统安全具有重要意义。近年来高光谱遥感技术^[3-4]不断发展, 其具有光谱分辨率高、波段多且连续性强等特点^[5], 能够快速高效地反演土壤重金属含量^[6-8]。另外, 随着高光谱图像性能的不提高, 土壤重金属含量的大范围制图以及周期性监测成为可能。

对土壤光谱进行分析时, 国内外学者常常使用全谱段来对土壤重金属的含量进行预测^[9-13]。一般情况下, 自然环境中的土壤重金属含量较低, 其本身微弱的光谱响应信号又受到多种成像因素的干扰, 难以直接通过分析土壤重金属元素的光谱特征来估算其含量^[14]。但土壤光谱活性物质 (主要指黏土矿物、铁氧化物和有机质等)

对重金属元素具有吸附或赋存作用^[15], 因而土壤重金属元素对土壤光谱的影响间接反映在这些光谱活性物质的特征光谱波段上。Sun 等^[16-17]使用有机质和黏土矿物的特征波段组合建立 Zn 和 Ni 的反演模型, 与全谱段建模相比精度有显著提升。贺军亮等^[18]基于土壤有机质敏感波段对应的多种光谱变换指标, 建立了 Cd 的高光谱间接反演模型。虽然现有研究在间接反演土壤重金属元素含量方面取得了一定进展, 但关于提取土壤 Pb 相关光谱活性物质的特征光谱进行 Pb 含量反演的可行性还有待验证。

在建模算法的选择方面, 遗传算法-偏最小二乘回归 (Genetic Algorithm-Partial Least Squares Regression, GA-PLSR)^[16-17,19-20]具有特征优选的能力, 且能对具有强相关性的自变量进行建模, 常被用于高光谱数据模型构建。由于高光谱数据波段众多、相关性强且冗余性高, 使用随机寻优的遗传算法对参与建模的波段进行筛选能够减少自变量的个数, 降低数据间相关性, 提取出光谱的有效信息, 最终提高模型估算精度^[21-22]。但是遗传算法仍然存在“过早收敛”的问题^[23], 算法在迭代时只接受适应度更高的解, 所以在开始迭代后便立即向初始种群的局部解收敛, 从而容易陷入局部解区域, 难以做到全局寻优, 对最终模型的精度产生了一定的影响。

以往研究多以矿区作为研究对象, 而矿区以外的一般农作区也应是监测的重要方面。本文以河北雄安一般农作区农田采集的土壤样本数据为研究对象, 提取土壤中对 Pb 起主导吸附作用的土壤光谱活性物质的特征谱段

收稿日期: 2020-04-27 修订日期: 2020-07-28

基金项目: 中国地质调查局地质项目 (202012000000180102)

作者简介: 张霞, 研究员, 主要从事高光谱遥感信息提取与应用研究。

Email: zhangxia@radi.ac.cn

进行 Pb 含量反演, 以避免全谱段建模的数据冗余问题。为解决 GA 算法的“过早收敛”问题, 提出改进遗传算法 (Improved Genetic Algorithm, IGA), 使算法在运行前期能够跳出局部解区域, 寻找更优解, 并结合偏最小二乘法^[24]进行回归建模反演土壤 Pb 含量。

1 研究数据与预处理

1.1 研究数据

雄安新区是中国新规划的数字经济创新发展试验区, 为关注该地区的环境建设, 将其作为研究区。2018 年 9 月 20 日至 9 月 23 日, 于雄县和安新县的农田开展土壤样本采集与野外原位光谱测量试验, 采集农田土壤表层(0~20 cm)作为土壤样本。野外光谱采集使用 SVC HR1024i 地物光谱仪, 该光谱仪波长范围为 350~2 500 nm, 测量时探头离地面距离大约 5 cm。采集土壤光谱时每个样点测 3~5 条光谱取平均作为样点光谱, 共采集了土壤样本及光谱 73 组。

土壤样本送实验室进行 Pb 含量测试。测量方法为 HNO₃-HF-HClO₄ 消煮法, 将样本在微波消解仪进行消解, 然后用电感耦合等离子体质谱 ICP-MS 测定 Pb 元素含量, 每个土壤样本测量 3 次后取平均值。

1.2 数据预处理

光谱数据处理主要包括光谱维去噪、异常光谱剔除、光谱特征增强及铁氧化物特征谱段提取等。

野外光谱测量时, 大气中的水汽在 1 400 和 1 900 nm 处存在强吸收, 且 1 900~2 500 nm 区间的光谱存在较大噪声^[25]。为去除大气水汽吸收波段和受噪声影响的波段, 并保留尽可能多的光谱区间, 将 1 800 nm 之后的光谱波段从土壤野外光谱中剔除。为进一步去除野外光谱在 350~1 800 nm 的随机噪声, 使用包络线去除及分段 Savitzky-Golay (SG) 滤波对光谱曲线进行平滑处理。SG 滤波中 890~1 020 和 1 330~1 520 nm 采用窗口大小为 15 的二次多项式, 其余区间采用窗口大小为 7 的二次多项式。为排除光谱异常和重金属含量异常对试验结果的影响, 通过光谱曲线筛选和 z-score 异常点检测方法, 去除 3 个异常样本, 最终保留 70 个有效样本用于反演建模。

高光谱数据相邻两波段具有较强的相关性, 使得整个波段的反射率具有强多重共线性, 给反演带来极大困难。因此采用一阶差分的方式, 减弱各波段之间的共线性, 增强光谱的特征^[4]。差分公式见式 (1)

$$B'_i = B_{i+1} - B_i, \quad i = 1, 2, \dots, n-1 \quad (1)$$

式中 B'_i 为第 i 段光谱的一阶差分, B_i 为第 i 段光谱反射率, n 为光谱总数。

2 研究方法

2.1 铁氧化物光谱特征提取

Covelo 等^[26]的重金属吸附和解析试验表明: 土壤对重金属 Pb 的吸附作用最强, 且土壤有机质、铁氧化物、锰氧化物均对重金属 Pb 具有很强的吸附作用, 其中铁氧化物在重金属 Pb 的吸附作用中占主导地位, 而且随着土壤中不同重金属之间对吸附位置的竞争加剧, 铁氧化物在重金属 Pb 的吸附作用中的重要性也随之上升。Wu

等^[27]的研究表明, 铁氧化物与重金属含量相关性在所有土壤光谱活性物质中占主导地位, 并试验得出铁氧化物的吸收特征在以 500 和 950 nm 为中心的吸收峰。因此, 本文提取 500 和 950 nm 处的吸收特征谱段预测土壤 Pb 含量, 提取波段区间为 $[B_m - W/4, B_m + W/4]$, 其中 B_m 是最大的吸收波段, W 是吸收区域的宽度。

综上, 从 70 条有效土壤光谱中提取 500 和 950 nm 为中心的铁氧化物吸收峰, 提取出的光谱波段范围为 450.7~523.2 以及 914.2~1 027.9 nm。

2.2 改进的遗传波段选择算法 (IGA)

遗传算法^[28-30]仿照物种演化机制, 模仿了遗传进化中发生的基因突变和自然选择现象, 通过随机选择、交叉和变异等遗传操作, 产生适应度高的个体作为最终解。传统遗传算法中存在最严重的问题是“过早收敛”: 运行初期会直接向局部解收敛而难以跳出。为解决该问题引入模拟退火算法中的 Metropolis 准则^[31-32]: 若新解优于当前解, 则接受新解, 否则以某一概率接受新解, 且此概率随适应度增大而降低。Metropolis 准则使得算法在运行初期有一定的概率接受较差解而跳出当前局部解区域, 从而做到全局寻优。IGA 算法具体流程如图 1。

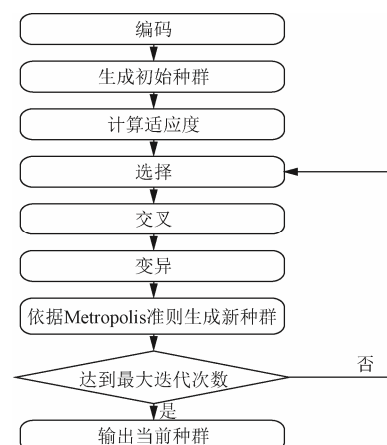


图 1 改进的遗传算法流程图

Fig.1 Process of improved genetic algorithm

算法的改进核心为生成新种群, 其中具体的 Metropolis 准则为: 当新个体目标函数值小于父个体目标函数值时, 用新个体取代父个体; 否则以 $\exp(-k \cdot \Delta f / E)$ 的概率接受新个体为父个体。其中 k 为大于 0 的实数, E 为种群目标函数值的方差。令 n 为算法总迭代次数, 当 E 大于 $0.01/n$ 时, $k = E + 0.01/n$; 否则 $k = 0.01/n$ 。 k 保证了 E 大于 $0.01/N$ 时算法有足够概率能够跳出局部解区域, E 小于 $0.01/N$ 时算法能够保持收敛状态。 E 定义如式 (2)

$$E = \frac{1}{N} \sum_{i=1}^N (f_i - \bar{f})^2 \quad (2)$$

式中 N 为种群个体数, f_i 为个体 i 的目标函数, \bar{f} 为种群目标函数的平均值。

2.3 偏最小二乘建模与模型评价指标

偏最小二乘法 (Partial Least Squares Regression, PLSR) 能较好地解决高光谱反演中自变量之间的强多重共线性问题, 并且能够实现多对多线性回归建模, 特别是在自变量存

在多重相关性而且个数较多，而样本数目又远小于自变量的情况下 PLSR 仍适用^[33-35]。因此，本研究采用 PLSR 建模。

使用拟合优度 R^2 ，均方根误差 (Root Mean Square Error, RMSE) 和相对偏差 (Relative Percent Difference, RPD) 评价模型的优劣。较好的模型通常拥有较高的 RPD 和 R^2 及较低的 RMSE。模型的评价参考现有的土壤属性含量高光谱估算的评价标准^[19,35]：出色模型， $R^2 > 0.9$ ， $RPD > 3.0$ ；良好模型， $0.9 > R^2 > 0.82$ ， $3.0 > RPD > 2.5$ ；近似模型， $0.82 > R^2 > 0.65$ ， $2.5 > RPD > 2.0$ ；具有一定估算能力， $0.65 > R^2 > 0.50$ ， $2.0 > RPD > 1.5$ ；不具备估算能力， $0.50 > R^2$ ， $RPD < 1.5$ 。

3 结果与分析

3.1 样本集划分

将 70 个样本分为训练集和测试集。依据重金属 Pb 的含量从小到大进行排序，对样本集按照 2：1 的比例进行分层抽样，获得 47 个样本的训练集和 23 个样本的测试集，样本集的数据统计见表 1。

表 1 样本集中 Pb 含量数据统计
Table 1 Data statistics of Pb content in the sample set

样本集 Sample set	样本数 Numbers of sample	最小值 Min (mg·kg ⁻¹)	最大值 Max (mg·kg ⁻¹)	平均值 Mean (mg·kg ⁻¹)	标准差 Standard deviation (mg·kg ⁻¹)	变异系数 Variable coefficient
训练集 Train set	47	3.57	28.39	15.34	5.71	0.37
测试集 Test set	23	5.52	26.06	15.54	5.43	0.35
总集 Total set	70	3.57	28.39	15.40	5.58	0.36

3.2 模型构建

为验证 IGA 波段选择算法改进的有效性，本文对 GA-PLSR 和 IGA-PLSR 这 2 种算法进行建模。2 种算法的参数设置相同，个体数为 20，个体编码方式为二进制编码，代际差为 90%。将 RMSE 作为波段选择算法的目标函数，即算法目标是最小化 RMSE。将选择算法运行结果中 RMSE 最小的个体作为当前最优解输出，并将其作为 PLSR 的自变量建立最终的回归预测模型。最后使用测试集数据计算模型评价指标。

3.3 土壤 Pb 含量估算模型评价及分析

根据 2.1 节的理论与方法，提取出铁氧化物特征谱段 (450.7~523.2 nm 和 914.2~1 027.9 nm) 作为输入，使用 GA-PLSR 及其改进算法 IGA-PLSR 对土壤 Pb 含量进行反演建模。为对比两算法的运行效果，选取迭代次数为 100、250、500 和 1 000，每种迭代次数下运行 5 次算法，评价参数及运行时间取 5 次运算的平均值。

试验结果如表 2 所示，可以看到随着迭代次数的增加，2 种算法的平均耗时增加，同时平均精度也在不断提高，IGA-PLSR 因加入了 Metropolis 准则的筛选过程，相同迭代次数下耗时略大于 GA-PLSR 算法。由模型精度随迭代次数的变化可得，2 种算法的最高精度对应的迭代次数均为 1 000 次，这与 Wang 等^[36]应用 GA 反演土壤重金属的迭代次数设置一致。在迭代次数为 1 000 次时，GA-PLSR 算法的反演精度 R^2 和 RPD 平均值分别为

0.782、2.117，RMSE 的平均值为 2.487 mg/kg，满足了近似模型的条件，能够近似估计 Pb 含量；而 IGA-PLSR 的 3 种评价参数均要优于改进前的算法，模型的 R^2 和 RPD 为 0.822、2.377，RMSE 为 2.221 mg/kg，其中 R^2 已经达到了良好模型的标准，表明模型对土壤 Pb 含量具有良好的估算性能。因此，IGA 算法相对于传统的 GA 算法具有更优的波段选择能力，显著提升了反演模型的精度。

表 2 基于铁氧化物特征谱段建模的反演精度
Table 2 Inversion accuracy based on iron oxide characteristic spectrum modeling

算法 Algorithm	迭代次数 Iterations	RMSE (mg·kg ⁻¹)	RPD	R^2	运行时间 Time consuming/s
GA-PLSR	100	2.861	1.860	0.711	104.28
	250	2.590	2.063	0.765	248.67
	500	2.503	2.106	0.778	490.03
	1 000	2.487	2.117	0.782	966.36
IGA-PLSR	100	2.775	1.953	0.726	124.75
	250	2.676	2.041	0.750	291.92
	500	2.341	2.282	0.802	576.23
	1 000	2.221	2.377	0.822	1 128.47

注：GA 为遗传算法；IGA 为改进遗传算法；PLSR 为偏最小二乘回归。铁氧化物波段范围为 450.7~523.2 和 914.2~1 027.9 nm。RPD 为相对偏差。
Note: GA: Genetic Algorithm; IGA: Improved Genetic Algorithm; PLSR: Partial Least Squares Regression. The iron oxide band ranges from 450.7-523.2 and 914.2-1 027.9 nm. RPD: Relative Percent Difference.

图 2 分别展示了 2 种算法在 1 000 次迭代时的 5 次运算中的最优结果。可以看到 IGA-PLSR 的 R^2 、RPD 为 0.837、2.403，RMSE 是 2.063 mg/kg，显著优于传统算法 GA-PLSR (R^2 、RPD 分别为 0.788、2.140，RMSE 为 2.447 mg/kg)，且通过散点的分布可以看到 IGA-PLSR 模型中的样本点更加集中于 1：1 线附近。因此，IGA-PLSR 算法的反演能力显著优于改进前的 GA-PLSR 算法。

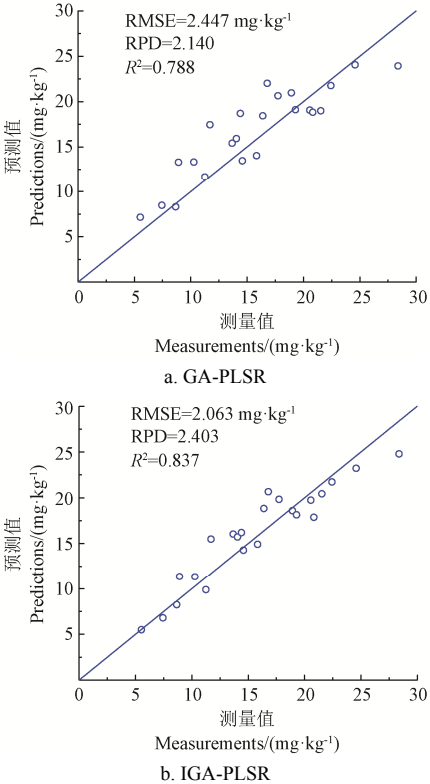


图 2 最优 Pb 含量估算结果
Fig.2 Optimal results of Pb content pridiction

3.4 铁氧化物特征谱段的有效性分析

为对比分析铁氧化物特征谱段的有效性, 基于 350~1 800 nm 的全部光谱波段对土壤重金 Pb 含量进行建模反演, 根据 3.3 节模型分析结果, 设置迭代次数为 1 000 次, 表 3 展示了试验结果。

表 3 基于全谱段建模的反演精度(迭代次数设为 1 000 次)

Table 3 Inversion accuracy based on full spectrum modeling (1 000 iterations)

算法 Algorithm	RMSE/(mg·kg ⁻¹)	RPD	R ²
GA-PLSR	4.511	1.216	0.292
IGA-PLSR	4.012	1.366	0.440

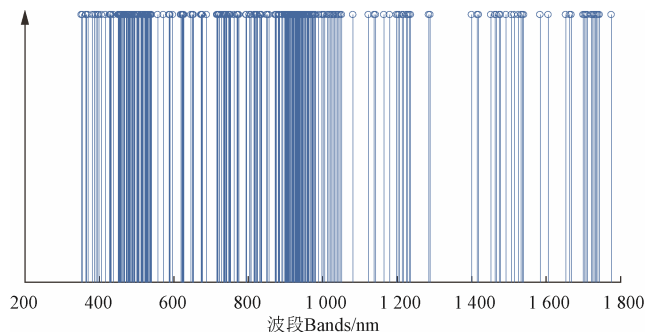
注: 全谱段范围为 350~1 800 nm。

Note: The full spectrum ranges from 350-1 800 nm.

对比表 2 (铁氧化物特征谱段) 和表 3 (全谱段) 的结果可以看出, 2 种建模方法在采用全谱段建模时模型精度较低。GA-PLSR 模型和 IGA-PLSR 模型的平均 R² 分别为 0.292、0.440, 平均 RPD 分别为 1.216、1.366。IGA-PLSR 模型有较高的精度和反演能力, 表明 IGA 波段选择算法在光谱波段较多、冗余度较高的情况下依然能选择出相对有效的光谱波段。但此时 2 种模型均未达到估算的精度要求, 而表 2 中使用铁氧化物特征谱段建模的结果要显著优于表 3 全谱段的结果, 说明铁氧化物特征波段的提取能切实有效地提高 Pb 含量反演的精度。

为直观显示铁氧化物特征波段对土壤 Pb 含量预测的贡献度, 图 3 展示出全谱段 IGA-PLSR 模型中 IGA 算法的波段筛选结果。波段选择算法的作用是筛选出能有效反演土壤重金属含量的波段, 可以看到筛选的波段在铁氧化物的特征谱段 (500 和 950 nm 吸收峰) 分布最为密集, 表明对研究区土壤 Pb 含量进行建模反演时, 铁氧化物特征谱段相比其他波段更加有效。

以上试验均说明了提取相关土壤活性物质的光谱特征能够在一定程度上减少光谱信息的冗余, 从而提高 Pb 含量反演的精度。本试验为基于铁氧化物特征波段反演土壤 Pb 含量提供了试验依据, 同时为反演土壤中其他重金属含量提供方法参考。



注: 竖直线表示对应横坐标波段被选中。

Note: Vertical line indicates that the corresponding abscissa band is selected.

图 3 全谱段 IGA-PLSR 模型波段筛选结果

Fig.3 Band selecting results of the IGA-PLSR model with full spectrum

3.5 算法收敛性及不确定性分析

为进一步分析 IGA-PLSR 算法对“过早收敛”的改

进, 选择 GA-PLSR 和 IGA-PLSR 迭代 1 000 次时的模型, 绘制代价函数 RMSE 随迭代次数变化的曲线, 如图 4。该曲线图将算法寻优过程可视化, 以便对比 2 种算法运算过程中的异同。向局部最优解收敛的特点为缓慢逼近, 由图 4 可看到 GA-PLSR 模型由于仅接受更优解, 整体呈单调下降趋势, 在 250 次迭代后 RMSE 基本稳定, 随后 400、600、720 及 900 次迭代邻域中都仅仅在某一次迭代后 RMSE 值有极小的降低, 表明已陷入局部最优, 后续仅是逼近局部最优解; IGA-PLSR 算法在迭代初期有较高的概率接受较差解, RMSE 波动较大, 在 300 次后开始收敛, 每次迭代都有一定的概率接受较差解, 从而跳出当前的局部解区域, 如在 380 和 640 次迭代附近接受较差解后 RMSE 都会有大幅降低出现, 说明跳出了当前局部解区域从而进一步寻找更优解。RMSE 曲线图的对比表明了 IGA-PLSR 算法提高反演精度的原因, 凸显了其在全局寻优方面的优势。

改进前后的 GA 和 IGA 都是随机寻优算法, 故存在一定的不确定性。主要体现在随机产生初始解、随机选择、交叉和变异等一系列遗传操作中, 同时引入的 Metropolis 准则中依概率接受较差解也增加了模型的不确定性。因此在实际应用时, 应多次运行算法后选择最优的模型进行重金属含量的反演。

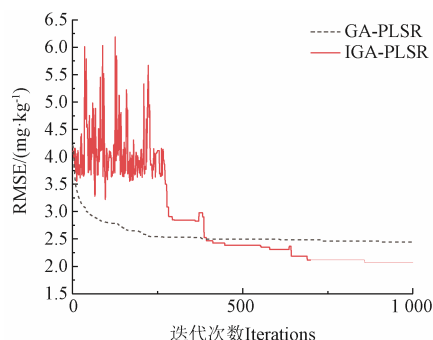


图 4 不同算法 RMSE 随迭代次数的变化

Fig.4 Changes of RMSE with iterations of different algorithms

4 结论与展望

本文基于河北雄安一般农作区采集的 70 个土壤样本的野外光谱数据, 研究具有机理性的特征波段的选取方法对土壤 Pb 含量进行反演。首先, 为规避全谱段建模的数据冗余和机理性不足问题, 提取出土壤中对 Pb 具有主导吸附性作用的铁氧化物的特征波段用于 Pb 含量反演, 在此基础上, 通过引入 Metropolis 准则对 GA 波段选择算法改进并提出改进遗传算法 (Improved Genetic Algorithm, IGA), 解决遗传算法的“过早收敛”的问题, 利用 IGA 算法筛选铁氧化物特征谱段中更加有效的光谱波段组合, 采用偏最小二乘法 (Partial Least Squares Regression, PLSR) 构建反演模型。结论如下:

1) IGA-PLSR 模型相对 GA-PLSR 拥有更高的反演精度, 其最优模型的 R²、RPD 分别为 0.837、2.403, RMSE 为 2.063 mg/kg, 其中 R² 达到了良好模型的标准, 可用于雄安新区农田土壤 Pb 含量的定量估算。

2) 铁氧化物特征谱段的反演精度要显著优于全谱段的精度, 说明了依据土壤 Pb 的吸附机理而提取相关土壤光谱活性物质的特征谱段能够有效减少全谱段建模的冗余信息, 从而提高 Pb 含量反演精度。

3) 通过对 2 种算法运行过程中 RMSE 的变化曲线分析, 表明 IGA-PLSR 能够有效地解决 GA 算法的“过早收敛”问题, 能够跳出局部解区域寻找更加有效的光谱波段组合, 提高 PLSR 的建模精度。

综上所述, 在应用于实际区域土壤 Pb 含量快速检测时, 可以从两方面提高土壤元素含量反演模型的精度和适用性: 提取出与土壤元素相关的光谱活性物质的特征光谱代替全谱段进行建模反演; 改进现有 GA 算法使用 IGA-PLSR 算法进行反演建模。但是, 在野外光谱及航空/卫星高光谱图像获取过程中, 不可避免地会受到土壤粒径、含水率等环境因素影响, 增加土壤重金属高光谱反演的不确定性, 因此, 下一步将研究野外光谱乃至像元光谱的环境因素去除方法, 以期提高土壤重金属高光谱遥感反演方法的稳定性。

[参 考 文 献]

- [1] Wei Binggan, Yang Linsheng. A review of heavy metal contaminations in urban soils, urban road dusts and agricultural soils from China[J]. *Microchemical Journal*, 2010, 94(2): 99-107.
- [2] Gmochowska W, Pietranik A, Tyska R, et al. Sources of pollution and distribution of Pb, Cd and Hg in Wrocław soils: Insight from chemical and Pb isotope composition[J]. *Geochemistry*, 2019, 79(3): 434-445.
- [3] Soriano-Disla J M, Janik L J, Rossel R A V, et al. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties[J]. *Applied Spectroscopy Reviews*, 2014, 49(2): 139-186.
- [4] 张贤龙, 张飞, 张海威, 等. 基于光谱变换的高光谱指数土壤盐分反演模型优选[J]. *农业工程学报*, 2018, 34(1): 110-117.
Zhang Xianlong, Zhang Fei, Zhang Haiwei, et al. Optimization of soil salt inversion model based on spectral transformation from hyperspectral index[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2018, 34(1): 110-117. (in Chinese with English abstract)
- [5] Shi Tiezhu, Chen Yiyun, Liu Yaolin, et al. Visible and near-infrared reflectance spectroscopy: An alternative for monitoring soil contamination by heavy metals[J]. *Journal of Hazardous Materials*, 2014, 265(2): 166-176.
- [6] Wei Lifei, Pu Haochen, Wang Zhengxiang, et al. Estimation of soil arsenic content with hyperspectral remote sensing[J]. *Sensors (Basel, Switzerland)*, 2020, 20(14): 4056.
- [7] 张秋霞, 张合兵, 刘文锴, 等. 高标准基本农田建设区域土壤重金属含量的高光谱反演[J]. *农业工程学报*, 2017, 33(12): 230-239.
Zhang Qiuxia, Zhang Hebing, Liu Wenkai, et al. Inversion of heavy metals content with hyperspectral reflectance in soil of well-facilitated capital farmland construction areas[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2017, 33(12): 230-239. (in Chinese with English abstract)
- [8] Zhao Li, Hu Yueming, Zhou Wu. Estimation methods for soil mercury content using hyperspectral remote sensing[J]. *Sustainability*, 2018, 10(7): 2474.
- [9] Wang Xi, An Shi, Xu Yaqing. A back propagation neural network model optimized by mind evolutionary algorithm for estimating Cd, Cr, and Pb concentrations in soils using Vis-NIR diffuse reflectance spectroscopy[J]. *Applied Sciences-basel*, 2020, 10(1): 51.
- [10] Zou Bin, Jiang Xiaolu, Feng Huihui, et al. Multisource spectral-integrated estimation of cadmium concentrations in soil using a direct standardization and Spiking algorithm[J]. *Science of the Total Environment*, 2020, 701: 134890.
- [11] 陈元鹏, 张世文, 罗 明, 等. 基于高光谱反演的复垦区土壤重金属含量经验模型优选[J]. *农业机械学报*, 2019, 50(1): 170-179.
Chen Yuanpeng, Zhang Shiwen, Luo Ming, et al. Empirical model optimization of hyperspectral inversion of heavy metal content in reclamation area[J]. *Transactions of the Chinese Society for Agricultural Machinery*. 2019, 50(1): 170-179. (in Chinese with English abstract)
- [12] Hou Le, Li Xinju, Li Fang. Hyperspectral-based inversion of heavy metal content in the soil of coal mining areas[J]. *Journal of Environmental Quality*. 2019, 48(1): 57-63.
- [13] 郭云开, 钱佳, 蒋明, 等. 基于核偏最小二乘的矿区土壤 Cu 含量高光谱反演[J]. *土壤通报*, 2019, 50(1): 52-56.
Guo Yunkai, Qian Jia, Jiang Ming, et al. Hyper-spectral inversion of Cu content in mining soil based on kernels partial least-squares[J]. *Chinese Journal of Soil Science*, 2019, 50(1): 52-56. (in Chinese with English abstract)
- [14] Rathod P H, Rossiter D G, Noomen M F, et al. Proximal spectral sensing to monitor phytoremediation of metal-contaminated soils[J]. *International Journal of Phytoremediation*, 2013, 15(5): 405-426.
- [15] Malley D F, Williams P C. Use of near-infrared reflectance spectroscopy in prediction of heavy metals in freshwater sediment by their association with organic matter[J]. *Environmental Science and Technology*, 1997, 31(12): 3461-3467.
- [16] Sun Weichao, Zhang Xia. Estimating soil zinc concentrations using reflectance spectroscopy[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2017, 58: 126-133.
- [17] Sun Weichao, Zhang Xia, Sun Xuejian, et al. Predicting nickel concentration in soil using reflectance spectroscopy associated with organic matter and clay minerals[J]. *Geoderma*, 2018, 327: 25-35.
- [18] 贺军亮, 韩超山, 韦锐, 等. 基于偏最小二乘的土壤重金

- 属间接反演模型 [J]. 国土资源遥感, 2019, 31(4): 96-103.
- He Junliang, Han Chaoshan, Wei Rui, et al. Research on indirect hyperspectral estimating model of heavy metal Cd based on partial least squares regression[J]. Remote Sensing for Land and Resources, 2019, 31(4): 96-103. (in Chinese with English abstract)
- [19] Vohland M, Besold J, Hill J, et al. Comparing different multivariate calibration methods for the determination of soil organic carbon pools with visible to near infrared spectroscopy[J]. Geoderma, 2011, 166(1): 198-205.
- [20] Wang Junjie, Cui Lijuan, Gao Wenxiu, et al. Prediction of low heavy metal concentrations in agricultural soils using visible and near-infrared reflectance spectroscopy[J]. Geoderma, 2014, 216(4): 1-9.
- [21] Elliott G N, Worgan H, Broadhurst D, et al. Soil differentiation using fingerprint Fourier transform infrared spectroscopy, chemometrics and genetic algorithm-based feature selection[J]. Soil Biology and Biochemistry, 2007, 39(11): 2888-2896.
- [22] 解宏图, 宣然然, 彭义, 等. 基于遗传算法的波长选择方法对土壤有机碳预测模型影响[J]. 土壤通报, 2014, 45(4): 795-800.
- Xie Hongtu, Xuan Ranran, Peng Yi, et al. Effects of prediction model on soil organic carbon using genetic algorithm[J]. Chinese Journal of Soil Science. 2014, 45(4): 795-800. (in Chinese with English abstract)
- [23] 杨从锐, 钱谦, 王锋, 等. 改进的自适应遗传算法在函数优化中的应用[J]. 计算机应用研究, 2018, 35(4): 1042-1045.
- Yang Congrui, Qian qian, Wang Feng, et al. Application of improved adaptive genetic algorithm in function optimization[J]. Application Research of Computers. 2018, 35(4): 1042-1045. (in Chinese with English abstract)
- [24] 尼加提·卡斯木, 师庆东, 王敬哲, 等. 基于高光谱特征和偏最小二乘法的春小麦叶绿素含量估算[J]. 农业工程学报, 2017, 33(22): 208-216.
- Nijat Kasim, Shi Qingdong, Wang Jingzhe, et al. Estimation of spring wheat chlorophyll content based on hyperspectral features and PLSR model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2017, 33(22): 208-216. (in Chinese with English abstract)
- [25] Kooistra L, Salas E A L, Clevers J, et al. Exploring field vegetation reflectance as an indicator of soil contamination in river floodplains[J]. Environmental Pollution, 2004, 127(2): 281-290.
- [26] Covelo E F, Vega F A, Andrade M L. Simultaneous sorption and desorption of Cd, Cr, Cu, Ni, Pb, and Zn in acid soils I. Selectivity sequences[J]. Journal of Hazardous Materials, 2007, 147(3): 852-861.
- [27] Wu Yunzhao, Chen Jun, Ji Junfeng, et al. A mechanism study of reflectance spectroscopy for investigating heavy metals in soils[J]. Soil Science Society of America Journal, 2007, 71(3): 918-926.
- [28] Ghoshal S P. Application of GA/GA-SA based fuzzy automatic generation control of a multi-area thermal generating system[J]. Electric Power Systems Research, 2003, 70(2): 115-127.
- [29] 毛博慧, 李民赞, 孙红, 等. 冬小麦苗期叶绿素含量检测光谱学参数寻优[J]. 农业工程学报, 2017, 33(增刊 1): 164-169.
- Mao Bohui, Li Minzan, Sun Hong, et al. Optimization of spectroscopy parameters and prediction of chlorophyll content at seeding stage of winter wheat[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2017, 33(Supp.1): 164-169. (in Chinese with English abstract)
- [30] 王春雷, 卢彩云, 陈婉芝, 等. 基于遗传算法和阈值滤波的玉米根茬行图像分割[J]. 农业工程学报, 2019, 35(16): 198-205.
- Wang Chunlei, Lu Caiyun, Chen Wanzhi, et al. Image segmentation of maize stubble row based on genetic algorithm and threshold filtering noise[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(16): 198-205. (in Chinese with English abstract)
- [31] 宾俊, 范伟, 周冀衡, 等. 智能优化算法应用于近红外光谱波长选择的比较研究[J]. 光谱学与光谱分析, 2017, 37(1): 95-102.
- Bin Jun, Fan Wei, Zhou Jiheng, et al. Application of intelligent optimization algorithms to wavelength selection of near-infrared spectroscopy[J]. Spectroscopy and Spectral Analysis. 2017, 37(1): 95-102. (in Chinese with English abstract)
- [32] 祝诗平, 王一鸣, 张小超, 等. 基于遗传算法的近红外光谱谱区选择方法[J]. 农业机械学报, 2004, 35(5): 152-156.
- Zhu Shiping, Wang Yiming, Zhang Xiaochao, et al. Region selection method of near infrared spectrum based on genetic algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery. 2004, 35(5): 152-156. (in Chinese with English abstract)
- [33] Vasques G M, Grunwald S, Sickman J O. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra[J]. Geoderma, 2008, 146(1): 14-25.
- [34] Rossel R A V, Behrens T. Using data mining to model and interpret soil diffuse reflectance spectra[J]. Geoderma, 2010, 158(1): 46-54.
- [35] Saeys W, Mouazen A M, Ramon H. Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy[J]. Biosystems Engineering, 2005, 91(4): 393-402.
- [36] Wang Junjie, Cui Lijuan, Gao Wenxiu, et al. Prediction of low heavy metal concentrations in agricultural soils using visible and near-infrared reflectance spectroscopy[J]. Geoderma, 2014, 216: 1-9.

Inversion of Pb content in soil based on iron oxide characteristic spectrum and improved genetic algorithm

Zhang Xia¹, Wang Yibo^{1,2}, Sun Weichao¹, Huang Changping¹, Zhang Mao^{1,2}

(1. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: A hyperspectral reflectance can be expected as an alternative method to quickly predict the concentration of heavy metals in soil at present. However, excessive band counting of hyperspectral data can lead to information redundancy during the quantitative inversion of surface parameters, and thereby to make it difficult to extract valid data from the spectrum. In this study, two aspects were addressed to solve spectral redundancy during the inversion of Pb content in soil: One is to extract the characteristics of iron oxides from the soil spectrum, according to the adsorption mechanism of Pb element in soil, in order to reduce data redundancy, while enhance the mechanism of inversion. Another is to propose an Improved Genetic Algorithm (IGA), in order to deal with the *premature convergence* of traditional Genetic Algorithm (GA), further to improve the extracting ability of algorithm for the required spectrum. Taking the Xiong'an New Area near Beijing, a national new-district in China, as the research region, the field experiment was carried out to quickly collect the Pb content of soil in the farmland, resulting from the rapid construction and developing facilities in recent years. 70 soil samples from the Xiong'an farmland were selected to establish the inversion model of Pb content. SVC HR1024I type ground object spectrometer was used to measure the raw hyperspectral reflectance of soil samples according to the national standard procedure in field farmland. After that, the collected soil samples were transferred to the laboratory for chemical analysis, particularly for the Pb content. Before modeling, the methods, including Savitzky-Golay filter, envelope removal and first-order difference, were used to pretreat with segmented the curves of soil reflectance. Partial Least Squares Regression (PLSR) was selected to serve as inversion model, in order to well deal with the strong correlation between hyperspectral bands. The models were also constructed in terms of full spectrum and characteristic bands of iron oxides, to verify the effectiveness of extracting method from characteristic bands of iron oxides during the inversion process. In full spectrum, the average values of R^2 , RPD, RMSE in the IGA-PLSR models were 0.440, 1.366, and 4.012 mg/kg, respectively. In the characteristic bands of iron oxides, the average values of R^2 , RPD, RMSE in the IGA-PLSR models were 0.822, 2.377, and 2.221 mg/kg, respectively. It infers that the models using the characteristic bands of iron oxides were much better than those using the full spectrum, indicating that the extraction for the characteristic bands of iron oxides can be contributed to reduce the redundancy of spectral information, and thereby to significantly improve the accuracy of the inversion model. Consequently, a final inversion model was selected from the optimal model for the improvement of inversion algorithm, based on the utilization of the characteristic bands of iron oxides to construct the GA-PLSR and IGA-PLSR models with different iterations. In the GA-PLSR models, the average values of R^2 , RPD, RMSE in the optimal model were 0.788, 2.140, and 2.447 mg/kg, respectively, whereas, the IGA-PLSR model achieved the highest accuracy, where the average values of R^2 , RPD, RMSE were 0.837, 2.403, and 2.063 mg/kg, respectively. Therefore, the optimal model of IGA-PLSR was chosen to estimate the Pb content in soil for the Xiong'an farmland area. The IGA-PLSR method has demonstrated better performance in the inversion of Pb content in soil, indicating that the proposed method was valid to improve genetic algorithm. At last, a visualized profile can be conducted for the changes of RMSE during the 1000 iterations of two models. The visualization revealed that the RMSE in the GA-PLSR model decreased monotonously during the operation process, whereas, the RMSE in the IGA-PLSR model fluctuated greatly at the beginning of the iteration, showing a certain probability to jump out from the local solution area to a better solution area at the later stage, while meaning that the problem of "premature convergence" was solved properly. This study revealed that two aspects, including the extraction of iron oxides spectral features associated with Pb absorption, and the improved algorithm of bands selection, both can be used to enhance the accuracy of Pb content estimation in soil. The proposed algorithm can provide a sound reference to real-time monitor the variation of heavy metal content in soil for Xiong'an area, China.

Keywords: soils; inversion; hyperspectral; heavy metals; feature selection; spectrum analysis