

基于有机质特征谱段的土壤 Cd 含量高光谱遥感反演

张霞¹ 孙友鑫^{1,2} 尚坤³ 丁松滔^{1,2} 孙伟超¹

(1. 中国科学院空天信息创新研究院, 北京 100101; 2. 中国科学院大学资源与环境学院, 北京 100049;

3. 自然资源部国土卫星遥感应用中心, 北京 100048)

摘要: 针对土壤 Cd 高光谱遥感定量反演中的机理性不足及数据冗余问题, 提出一种基于有机质特征谱段的反演方法。该方法首先提取土壤光谱中对重金属 Cd 具有吸附作用的有机质特征谱段, 进而通过竞争性自适应重加权采样法 (Competitive adaptive reweighted sampling, CARS) 优选特征谱段, 采用偏最小二乘回归法 (Partial least squares regression, PLSR) 建立重金属 Cd 的反演模型, 并利用郴州矿区土壤实验室光谱数据和哈密黄山南矿区野外光谱数据进行方法验证。研究表明: 有机质特征谱段提取在降低数据冗余的同时提高了重金属 Cd 的反演精度, CARS 算法相对于相关系数法 (Correlation coefficient, CC) 和遗传算法 (Genetic algorithm, GA) 特征选择具有更高的反演精度, 基于有机质特征谱段的 CARS-PLSR 算法在土壤实验室光谱和野外实测光谱所得验证精度 R^2 分别为 0.94 和 0.80, 表明该算法对于实验室和野外光谱均具有一定适用性。研究可为土壤重金属含量高光谱反演的特征波段选择和算法优选提供参考。

关键词: 高光谱遥感; 土壤重金属; 土壤光谱活性物质; 特征选择; 反演

中图分类号: TP79

文献标识码: A

文章编号: 1000-1298(2024)01-0186-10

OSID:



Soil Cd Content Retrieval from Hyperspectral Remote Sensing Data Based on Organic Matter Characteristic Spectral Bands

ZHANG Xia¹ SUN Youxin^{1,2} SHANG Kun³ DING Songtao^{1,2} SUN Weichao¹

(1. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China

2. College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

3. Land Satellite Remote Sensing Application Center, Ministry of Natural Resources, Beijing 100048, China)

Abstract: To address the mechanistic limitations and data redundancy issues in the quantitative retrieval of soil Cd using hyperspectral remote sensing, an inversion method was proposed based on organic matter characteristic spectral bands. The method involved the extraction of characteristic spectral bands of organic matter with adsorption effects on heavy metal Cd in soil spectra. Subsequently, competitive adaptive reweighted sampling (CARS) was employed to optimize the selected spectral bands, and a partial least squares regression (PLSR) model was developed for the inversion of heavy metal Cd. The proposed method was validated by using laboratory spectral data from the Chenzhou mine and field spectral data from the Hami Huangshan South mine. The results demonstrated that the extraction of organic matter characteristic spectral bands not only reduced data redundancy but also significantly improved the accuracy of Cd inversion. In comparison to the correlation coefficient (CC) and genetic algorithm (GA) methods, the CARS algorithm exhibited superior performance in feature selection and inversion accuracy. The validation accuracies, expressed as R^2 , were 0.94 for the Chenzhou laboratory spectral data and 0.80 for the Hami field spectral data, indicating the robustness of the CARS-PLSR algorithm for both laboratory and field spectra. The findings can provide valuable references for feature band selection and algorithm optimization in the hyperspectral estimation of soil heavy metal content. The proposed method effectively addressed the limitations of existing approaches by leveraging the unique spectral characteristics of organic matter in soil.

Key words: hyperspectral remote sensing; soil heavy metal; soil spectrally active substance; feature selection; retrieval

收稿日期: 2023-06-05 修回日期: 2023-07-03

基金项目: 国家自然科学基金项目(42371360)和中国科学院战略性先导科技专项(XDA28080500)

作者简介: 张霞(1972—), 女, 研究员, 主要从事高光谱遥感信息提取与应用研究, E-mail: zhangxia@radi.ac.cn

0 引言

近年来,土壤遭受重金属污染的状况日益严重,造成严重的环境和社会问题,尤其是在工业化发展迅速的地区,土壤重金属污染问题更加严重^[1]。据《全国土壤污染状况调查公报》统计,我国土壤污染超标率高达 16.1%,其中,重金属 Cd 污染点位超标率最高^[2]。重金属 Cd 具有较大的毒性,并且很难自然降解,会在土壤中不断积累,通过食物链对人体健康造成威胁^[3]。传统土壤重金属检测多采用实地调查与实验室化学分析相结合的方法,但该方法成本高、耗时长^[4],难以满足空间大尺度快速监测土壤重金属含量及空间分布的需求^[5]。随着遥感技术的发展,特别是高光谱遥感通过获取连续而精细的土壤反射光谱,使得一次性实现多种重金属元素的快速探测成为可能^[6]。

在利用高光谱数据反演土壤重金属含量时,国内外学者大多将整个可见近红外-短波红外反射光谱(VNIR-SWIR)区间用于反演建模,会造成数据冗余问题^[7],也未充分考虑土壤重金属的光谱反演机理^[8]。土壤反射光谱产生机理研究表明,土壤在 VNIR-SWIR 区间的光谱特征主要由铁氧化物、有机质、黏土矿物引起^[9],这三者又称为土壤光谱活性物质。重金属与土壤光谱活性物质的内在联系是土壤反射光谱反演重金属含量的机理^[10]。SUN 等^[11]通过提取对土壤重金属 Zn 起吸附作用的有机质和黏土矿物的特征谱段,实现了 Zn 含量的反演。提取对土壤重金属起主要吸附作用的土壤光谱活性物质特征谱段,作为光谱子集用于建模是解决高光谱数据冗余问题的有效途径^[12]。

目前,土壤重金属含量估算研究多采用实验室光谱,较少使用土壤野外实测光谱数据。实验室光谱是土壤样本经过干燥、研磨、过筛等一系列预处理后,在实验室可控条件下测量的光谱。实验室光谱不受大气、土壤水分和表面粗糙度等环境因素的影响,而野外光谱测量的是自然状态下的原位土壤,不可避免地受到这些环境因素的影响^[13]。土壤重金属含量反演方法对两种光谱数据的适用性有待检验。

特征选择算法是解决高光谱数据冗余问题的另一种有效途径,可以优选土壤光谱中特征性强的波段,降低数据冗余,提高重金属反演精度^[14],目前土壤重金属反演中常用的特征选取算法有相关系数法(CC)、遗传算法(GA)和竞争性自适应重加权采样法(CARS)等。阳洋等^[15]在 CC 优选的波段基础上,进一步采用多元逐步回归剔除共线性波长,构建

的氧化铁模型决定系数达 0.790。ZHAO 等^[16]采用 GA 算法进行特征选取,相较于直接利用多元线性回归和反向传播神经网络反演,土壤汞含量的反演精度得到了明显提高。通常直接利用 CC 算法进行特征选择时,建模输入变量较多且未考虑变量间的共线性导致预测精度不高^[17],GA 和 CARS 算法均属于以“适者生存”为原则的特征优化算法,GA 以精度最大化为目标,通过随机寻优的方式筛选波段,降低数据间相关性^[18],但 GA 算法会过早收敛,陷入局部最优解^[19],而 CARS 算法以各波段重要性为依据,通过逐步去除冗余和不重要特征^[20],可以实现最优特征段组合的高效提取^[21]。

本文以湖南省郴州市矿区土壤实验室光谱和新疆维吾尔自治区哈密市黄山南矿区土壤野外光谱为数据源,提出一种基于有机质特征谱段的 CARS-PLSR 算法的土壤重金属 Cd 含量反演方法,该方法通过提取有机质特征谱段增强反演的机理性,进而采用 CARS 进行特征优选,实现对两种光谱数据的土壤重金属 Cd 含量建模反演,并将 CARS 与 CC、GA 算法在全谱段和有机质特征谱段的建模结果进行对比,探究基于有机质特征谱段的 CARS-PLSR 算法在降低数据冗余、提高土壤重金属 Cd 反演精度方面的有效性,以及在野外实测光谱反演重金属 Cd 含量中的适用性。

1 研究数据与预处理

1.1 研究区概况

本文采用的土壤数据源于两个研究区(图 1):湖南省郴州市某铅锌矿区(郴州矿区),位于湖南省境内资江水系上游,属东江湖库区六大采矿区之一。属中亚热带季风性湿润气候区,冬春两季多冷空气活动,夏秋两季盛行热带湿润季风,年平均降水量 1 452.1 mm。矿区土壤以黄壤和红壤为主,几十年矿产开采产生的矿渣以及废液渗入土壤,使得当地土壤存在严重的重金属污染。新疆维吾尔自治区哈密市黄山南矿区(哈密矿区),位于新疆维吾尔自治区东部,属于温带大陆性气候,干燥少雨,年降水量 33.8 mm,年蒸发量 3 300 mm,夏季酷热、蒸发强,矿区土壤类型主要为棕漠土、灰棕漠土、盐土等。该矿区是重要的铜、镍、铁、铅、锌等大型矿床集中区,于 2007 年开始勘探,至今仍在开采中。矿区开采过程中,矿石矿渣中的重金属经雨水冲刷、淋溶渗入周边土壤中,易造成周围土壤的重金属污染^[22]。两个研究区在气候分区和土壤类型等方面具有典型性。

1.2 研究数据

2015 年 11 月沿郴州市清江镇东河西岸在

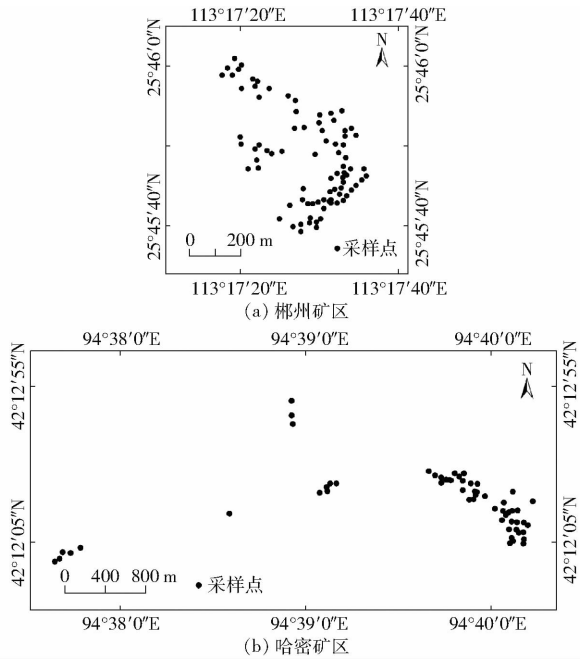


图 1 土壤采样点分布
Fig. 1 Distribution of soil sampling points

113°17'18" ~ 113°17'36"E, 25°45'39" ~ 25°46'01"N 范围采集土壤表层(0 ~ 20 cm)样本。采集的土壤样本于实验室内通风处晾干,土壤干燥后用有机玻璃棒将干燥的样本压散,去除植物残体、砾石等杂质,然后使用玛瑙研钵对土壤样本进行研磨,并用 100 目筛对土壤样本进行分选。对处理后的 72 个土壤样本进行实验室光谱测量,利用波长范围为 350 ~ 2 500 nm 的 PSR - 3500 型便携式光谱仪和一个视场角为 25°的光纤在暗室进行土壤样本光谱测量,以一个 50 W 的卤素灯作为光源,使用 BaSO₄板作为定标板,共采集 72 个不同环境土质的土壤样本及其实验室反射率光谱。

2021 年 7 月 22—30 日,在哈密矿区开展土壤样本采集与光谱测量,为避免周围地物的临近反射光和地物阴影的影响,土壤野外光谱测量采用 PSR - 3500 型便携式光谱仪,选择在开阔、平坦且具有代表性土壤的地区,清除土壤表面杂物进行采样。土壤野外光谱测量时对每一个样本单独进行反射率定标以消除太阳辐射随时间变化的影响,光谱测定重复 10 次取平均,通过野外光谱测量共采集 57 个不同环境土质的土壤样本及其反射率光谱。

郴州和哈密矿区每个样点均采集 1 kg 的土壤样本,对 2 个研究区的土壤样本采用酸消解法和火焰原子吸收光谱法进行重金属 Cd 含量测定。首先在微波消解仪中利用 HNO₃ - HF - HClO₄ 对土壤样本进行消解,然后采用火焰原子吸收光谱法测定样本重金属 Cd 含量,每个土壤样本平行测定 3 次,取均值记为样本重金属含量。为了避免土壤样本出现

偏分布,影响反演精度,对郴州和哈密矿区土壤样本的 Cd 含量进行对数变换,使其满足正态分布。

1.3 数据预处理

2 组光谱数据在波长两端信噪比较低,为了降低噪声影响,将 350 ~ 400 nm、2 400 ~ 2 500 nm 两个光谱区间剔除,利用剩余 400 ~ 2 400 nm 土壤反射光谱进行重金属含量反演。对于哈密矿区野外实测土壤反射光谱数据,因受到 1 900 nm 水汽强吸收影响,而 1 400 nm 附近水汽吸收影响较小,因此,保留 1 400 nm 附近的光谱数据,仅去除 1 900 nm 附近受水汽影响严重的区域^[23]。

为进一步去除土壤反射光谱数据受环境、采集设备等影响而产生的光谱噪声,利用 Savitsky - Golay(SG)平滑去噪,SG 可以有效减少原始光谱曲线毛刺,消除随机噪声^[24]。400 ~ 2 400 nm 范围内的两组光谱均存在不同程度的噪声,根据光谱噪声严重程度采用不同的窗口尺寸进行分段 SG 滤波^[25]。郴州矿区土壤实验室光谱可分为 3 个噪声区间:900 ~ 1 000 nm 和 1 800 ~ 2 400 nm 为严重噪声区间,400 ~ 500 nm 和 1 000 ~ 1 800 nm 为中度噪声区间,500 ~ 900 nm 为轻度噪声区间。哈密矿区土壤野外实测光谱的严重噪声区间为 900 ~ 1 000 nm 和 1 800 ~ 2 400 nm,中度噪声区间为 500 ~ 900 nm 和 1 000 ~ 1 800 nm,轻度噪声区间为 400 ~ 500 nm。对于严重噪声、中度噪声和轻度噪声区间,分别使用窗口尺寸为 21、15、7 的二次多项式进行滤波处理。滤波前后的郴州矿区土壤实验室光谱和哈密矿区土壤野外实测光谱如图 2、3 所示。对比原始光谱数据,滤波后的光谱质量明显提升。微分变换可以去除部分线性或接近线性的噪声对目标光谱的影响,被广泛应用于光谱分析,二阶微分在一阶微分消去基线漂移的基础上,还可以消去线性趋势,因此对滤波处理后的数据进行二阶微分处理^[26]。

2 研究方法

图 4 为本文方法技术路线图,首先对光谱数据进行 SG 滤波去噪,通过包络线去除增强光谱特征;根据对 Cd 起主要吸附作用的土壤光谱活性物质(即有机质),提取其特征谱段,作为光谱特征子集;采用 CASR 算法对光谱特征子集进行特征优选;将优选的特征组合利用 PLSR 建立反演模型,实现对土壤重金属 Cd 含量反演。

2.1 特征光谱子集提取方法

重金属与土壤光谱活性物质的内在联系是土壤反射光谱反演重金属含量的机理,对于不同重金属,铁氧化物、锰氧化物、黏土矿物和有机质对其吸附强

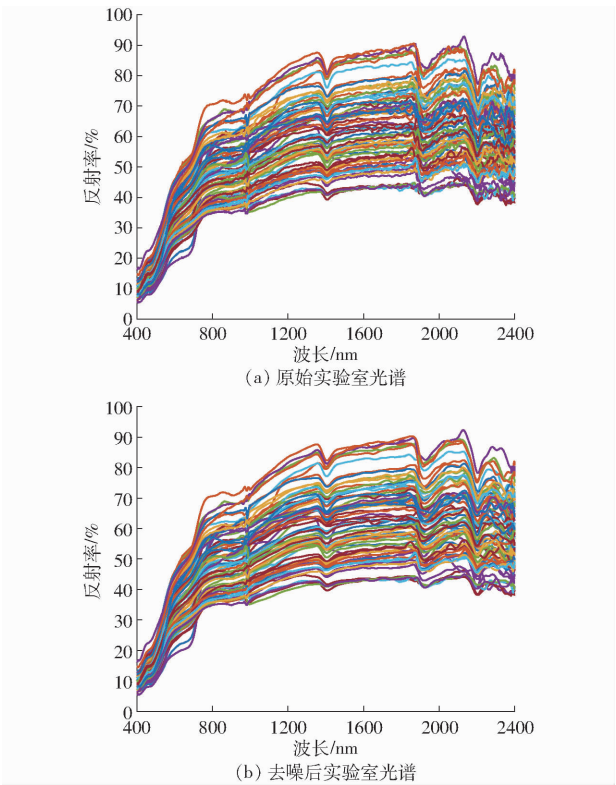


图 2 郴州矿区土壤原始实验室光谱与滤波后结果
Fig.2 Original and filtered laboratory spectra of soils from Chenzhou mining area

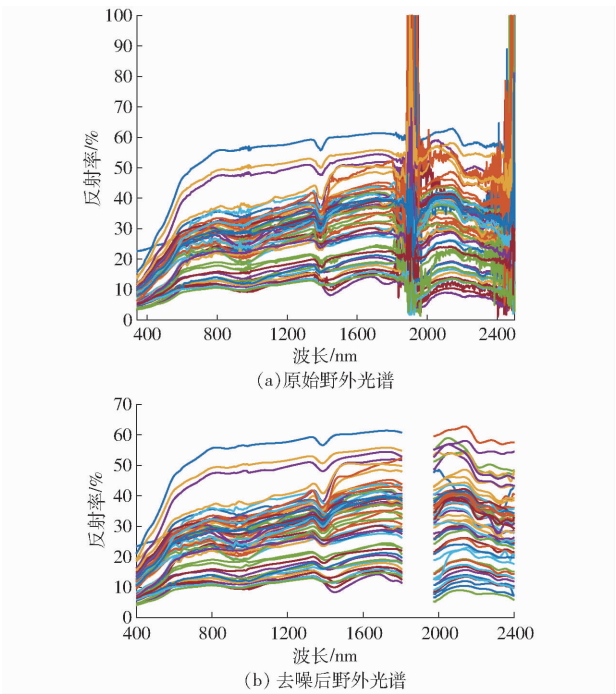


图 3 哈密矿区土壤原始野外光谱与滤波后结果
Fig.3 Original and filtered field spectra of soils from Hami mining area

度不同,起主导作用的土壤组分不同^[27],土壤组分对重金属 Cd 吸附和解析作用的影响研究表明:土壤对 Cd 的固定作用与土壤有机质含量显著相关,在 Cd 的吸附作用中占主导作用的是土壤中的有机

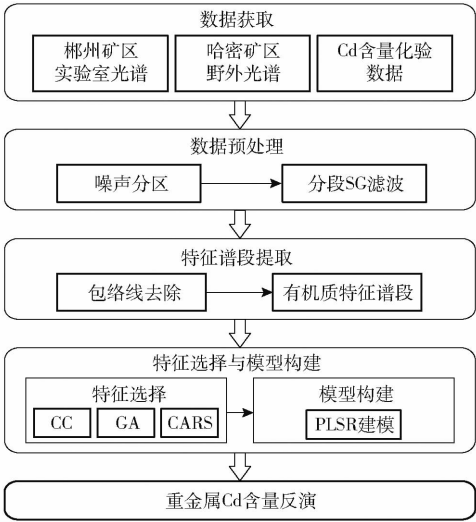


图 4 本研究方法技术路线图
Fig.4 Flowchart of proposed method

质^[28]。有机质对土壤反射光谱影响最大的范围为 600 ~ 800 nm 吸收峰^[29],因此,本文提取 600 ~ 800 nm 内的吸收峰作为特征光谱子集,用于土壤重金属 Cd 含量反演。特征光谱提取采用包络线去除法增强光谱后,将光谱转换至相同基准上,有效提取光谱吸收特征^[30],由于土壤类型及土壤样本处理方法、测量条件存在差异,土壤反射光谱吸收峰位置也有所不同,在提取有机质特征谱段时,以 600 ~ 800 nm 内的吸收峰最大吸收深度为中心,提取左右两侧吸收宽度为 1/2 的光谱波段用于土壤重金属 Cd 含量反演,得到郴州矿区实验室光谱数据和哈密矿区野外光谱数据的有机质特征光谱子集。

2.2 竞争自适应重加权采样法

CARS 是一种将蒙特卡罗 (MC) 和 PLS 回归系数相结合从而完成特征选取的方法^[6]。该方法结合自适应重加权采样算法,从光谱中提取回归系数绝对值最大的光谱波段,去除冗余和不重要的光谱信息,再利用 PLS 方法对各波长变量子集进行蒙特卡罗交叉验证建模,从数据集中选出均方根误差最小的子集,作为最优数据变量^[31]。算法具体分为 4 步:

(1) 利用蒙特卡洛法对数据进行采样,在采样过程中,通过随机选择的方式选择样本。

(2) 利用衰减指数对所选择的样本进行评价,保留权重绝对值较大的样本,而当权重绝对值小于阈值时,进行舍弃,定义权重 W_i 为

$$W_i = \frac{|b_i|}{\sum_{i=1}^n |b_i|}$$

式中 $|b_i|$ ——第 i 个波段对估算 Cd 含量的贡献
 n ——光谱波段总数

通过 CARS 算法去除的变量,将其 W_i 设为 0。

(3)将 CARS 算法应用到特征波段选择中,从反射波谱中选择对于建模精度影响较大的波段,进一步降低数据冗余。

(4)利用交叉验证来选择最优变量子集,并将其作为 CARS 最终选择的特征波长。

2.3 反演建模方法与模型评价指标

偏最小二乘法 (PLSR) 能够处理高度自相关的数据和变量数远大于样本数的数据,被证明是一种非常有效的多变量数据分析方法,该方法通过求解误差的平方和,寻找最小值来找到一组数据的最佳函数匹配,尤其可以解决自变量存在严重相关性的情况,可以很好地应对高光谱数据变量之间相关性的问题^[32]。因而,本文采用 PLSR 进行建模反演。

采用决定系数 (Coefficients of determination, R^2)、预测均方根误差 (Root mean square error of prediction, RMSEP) 和相对偏差 (Relative percent difference, RPD) 在验证样本集上评价模型精度。模

型优劣参考现有的土壤属性含量高光谱反演的评价标准^[33]: $R^2 \geq 0.9$, RPD 大于 3.0 时为出色模型; $0.82 \leq R^2 < 0.9$, RPD 为 2.5 ~ 3.0 时为良好模型; $0.65 \leq R^2 < 0.82$, RPD 为 2.0 ~ 2.5 时为近似模型; $0.5 \leq R^2 < 0.65$, RPD 为 1.5 ~ 2.0 时模型具有一定估算能力; $R^2 < 0.5$, RPD 小于 1.5 时该模型不具备估算能力。

3 结果与分析

3.1 样本集划分与模型构建

按照土壤采样顺序,将数据以比例 2:1 划分训练样本和验证样本,自第 3 个样本开始,每隔 2 个样本选择 1 个作为验证样本,其余样本作为训练样本。郴州矿区数据划分后得到 48 个训练样本和 24 个验证样本,哈密矿区数据划分后得到 38 个训练样本和 19 个验证样本。郴州和哈密矿区样本全集、训练样本和验证样本的土壤重金属 Cd 含量 (质量比) 的统计值分别如表 1、2 所示。

表 1 郴州矿区样本集土壤 Cd 含量统计值
Tab.1 Statistics of soil Cd content of Chenzhou mining area

数据集	最小值/(mg·kg ⁻¹)	最大值/(mg·kg ⁻¹)	平均值/(mg·kg ⁻¹)	中值/(mg·kg ⁻¹)	标准差/(mg·kg ⁻¹)	变异系数
全部样本	0.48	31.17	5.35	2.61	6.08	1.13
训练样本	0.48	31.17	6.26	2.55	6.99	1.12
验证样本	0.52	11.49	3.58	2.82	3.07	0.86

表 2 哈密矿区样本集土壤 Cd 含量统计值
Tab.2 Statistics of soil Cd content of Hami mining area

数据集	最小值/(mg·kg ⁻¹)	最大值/(mg·kg ⁻¹)	平均值/(mg·kg ⁻¹)	中值/(mg·kg ⁻¹)	标准差/(mg·kg ⁻¹)	变异系数
全部样本	0.01	0.21	0.07	0.06	0.04	0.59
训练样本	0.01	0.21	0.07	0.06	0.04	0.62
验证样本	0.01	0.14	0.06	0.06	0.02	0.38

为了验证本文方法的有效性,同时利用 CC 和 GA 进行特征选择,结合 PLSR 进行建模,并将 CC-PLSR 和 GA-PLSR 算法作为对照,与本文 CARS-PLSR 建模结果做比较。设置 CC 显著性水平 $p < 0.01$, GA 算法的染色体数设置为 20,代际间隙为 90%,基因变异概率为 10%,CARS 算法设置 MC 采样次数为 40、每次迭代抽取 90% 的样本用于运算。同时,为扩大变量选择的区间和降低不同初始条件对重金属含量反演建模的影响,运行 5 次算法并选择最优模型进行含量反演。

3.2 土壤重金属 Cd 含量反演结果与分析

3.2.1 郴州矿区土壤实验室光谱特征选择与土壤重金属 Cd 含量反演

郴州矿区土壤实验室光谱数据提取得到的有机

质特征谱段为 638.2 ~ 727 nm。基于有机质特征谱段数据和全谱段数据分别应用 CARS-PLSR 算法、CC-PLSR 算法、GA-PLSR 算法进行重金属 Cd 含量反演,5 次运行所得验证集精度如表 3 所示,最优验证结果散点图如图 5 所示。

CC-PLSR 算法基于郴州矿区土壤实验室光谱全谱段的重金属 Cd 反演精度 R^2 为 0.42,而基于有机质特征谱段的反演精度 R^2 为 0.76;GA-PLSR 算法基于全谱段的 R^2 为 0.58,基于有机质特征谱段 R^2 为 0.85;CARS-PLSR 算法基于全谱段的 R^2 为 0.67,基于有机质特征谱段的 R^2 达到 0.94,RPD 为 5.16。根据模型评价标准,基于有机质特征谱段的 CARS-PLSR 所建模型精度最佳,为出色模型,基于全谱段的 CARS-PLSR 建模精度为 3 种特征选择和建模方法中最佳。

表 3 郴州矿区土壤实验室光谱全谱段与特征谱段反演土壤 Cd 含量精度

Tab.3 Retrieval accuracies of soil Cd content using entire spectral bands and characteristic spectral bands associated with soil organic matter of Chenzhou mining area laboratory spectra							
特征选择方法	R^2	全谱段			R^2	有机质特征谱段	
		RMSEP/($\text{mg}\cdot\text{kg}^{-1}$)	RPD			RMSEP/($\text{mg}\cdot\text{kg}^{-1}$)	RPD
CC	1	0.42	2.01	1.30	0.76	1.58	2.10
	2	0.42	2.01	1.30	0.76	1.58	2.10
	3	0.42	2.01	1.30	0.76	1.58	2.10
	4	0.42	2.01	1.30	0.76	1.58	2.10
	5	0.42	2.01	1.30	0.76	1.58	2.10
GA	1	0.58	1.81	1.57	0.83	1.44	2.51
	2	0.57	1.85	1.55	0.85	1.41	2.82
	3	0.52	1.92	1.50	0.77	1.54	2.15
	4	0.51	1.95	1.47	0.81	1.48	2.38
	5	0.56	1.87	1.53	0.79	1.51	2.27
CARS	1	0.62	1.78	1.68	0.91	1.31	3.45
	2	0.67	1.73	1.81	0.89	1.36	3.10
	3	0.60	1.80	1.61	0.94	1.20	5.16
	4	0.64	1.75	1.74	0.92	1.30	3.54
	5	0.63	1.76	1.69	0.93	1.23	3.85

3.2.2 哈密矿区土壤野外实测光谱特征选择与土壤重金属 Cd 含量反演

哈密矿区野外实测数据提取的有机质特征谱段为 610 ~ 746.8 nm。基于全谱段和有机质特征谱段的 CARS - PLSR、CC - PLSR 和 GA - PLSR 重金属 Cd 含量验证集精度如表 4 所示,最优验证结果散点

图如图 6 所示。
CC - PLSR 算法基于哈密矿区土壤野外实测光谱全谱段数据所得的 Cd 反演精度 R^2 为 0.46,基于有机质特征谱段的 R^2 为 0.66;GA - PLSR 算法基于全谱段的 R^2 为 0.55,基于有机质特征谱的 R^2 为 0.73;CARS - PLSR 算法基于全谱段和有机质特征谱段的 Cd 反演精度 R^2 分别为 0.62 和 0.80。在 3 种特征选择与建模方法中,CARS - PLSR 算法在有机质特征谱段和全谱段的反演精度均为最佳,有机质特征谱段所建模型达到近似模型。

3.3 两组光谱数据反演结果对比

土壤有机质具有较多的含氧功能团,对重金属具有很强的亲和性^[34],对土壤中 Cd、Cr、Cu、Ni、Pb、Zn 的同时吸附和解吸实验表明,土壤有机质对 Cd 有很强的吸附作用^[35],这使得利用土壤有机质和 Cd 的相关性来反演土壤 Cd 含量成为可能。由郴州矿区土壤实验室光谱和哈密矿区土壤野外实测光谱的土壤 Cd 含量反演结果可以看出,有机质特征谱段提取对于提高重金属 Cd 反演精度具有较好的效果,3 种特征选择算法所得的反演结果中,基于有机质特征谱段所得精度均高于利用土壤 400 ~ 2 400 nm 全谱段的反演精度。因此,有机质特征谱段的提取在降低波段冗余的同时提高了重金属 Cd 含量反演精度。

在 3 种特征选择算法中,利用 CARS 算法进行特征选择所得反演精度均高于 CC 和 GA 算法,应用于有机质特征谱段的对比分析表明,对于郴州矿区

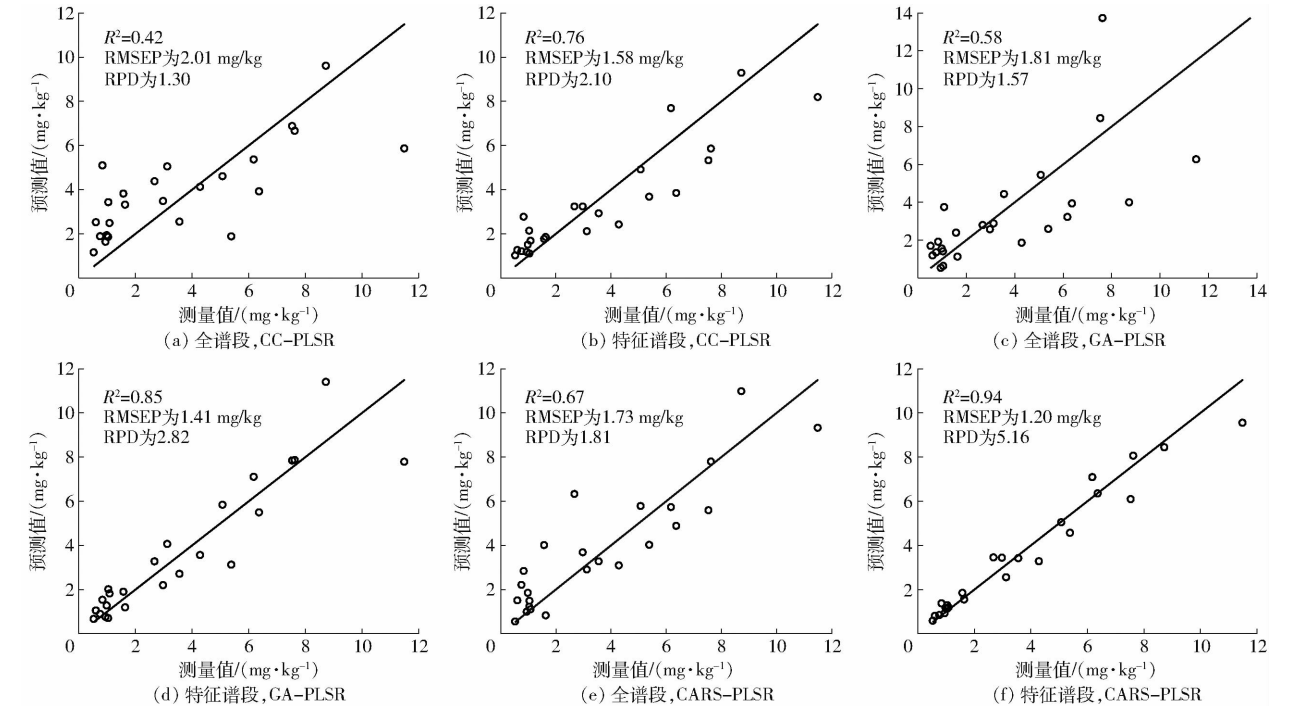


图 5 郴州矿区土壤实验室光谱重金属 Cd 含量反演最优验证结果散点图
Fig.5 Retrieval results of soil Cd content using Chenzhou mining area laboratory spectra

表 4 哈密矿区土壤野外实测光谱全谱段与特征谱段的 Cd 含量反演精度

Tab.4 Retrieval accuracies of soil Cd content using entire spectral bands and characteristic spectral bands associated with soil organic matter of Hami mining area field spectra

特征选择方法		全谱段			有机质特征谱段		
		R^2	RMSEP/ ($\text{mg}\cdot\text{kg}^{-1}$)	RPD	R^2	RMSEP/ ($\text{mg}\cdot\text{kg}^{-1}$)	RPD
CC	1	0.46	0.028	1.40	0.66	0.020	1.76
	2	0.46	0.028	1.40	0.66	0.020	1.76
	3	0.46	0.028	1.40	0.66	0.020	1.76
	4	0.46	0.028	1.40	0.66	0.020	1.76
	5	0.46	0.028	1.40	0.66	0.020	1.76
GA	1	0.51	0.026	1.47	0.73	0.019	1.96
	2	0.55	0.025	1.51	0.71	0.019	1.92
	3	0.52	0.026	1.49	0.72	0.019	1.94
	4	0.49	0.027	1.42	0.68	0.020	1.84
	5	0.50	0.026	1.43	0.67	0.020	1.81
CARS	1	0.62	0.024	1.67	0.80	0.017	2.32
	2	0.58	0.025	1.60	0.77	0.018	2.14
	3	0.57	0.025	1.59	0.79	0.017	2.22
	4	0.60	0.024	1.63	0.75	0.019	2.04
	5	0.59	0.024	1.62	0.78	0.018	2.16

土壤实验室光谱,利用 CARS-PLSR 算法所得 R^2 比 CC-PLSR 和 GA-PLSR 算法分别高 0.18 和 0.09;对于哈密矿区土壤野外实测光谱,利用 CARS-PLSR 算法所得 R^2 比 CC-PLSR 和 GA-PLSR 算法高 0.14 和 0.07。CC 算法在 3 种特征选择算法中所得的反演精度最低,在郴州矿区土壤实验室光谱数

据和哈密矿区土壤野外实测光谱数据全谱段反演结果 R^2 均低于 0.5,模型不具备估算能力,这是由于 CC 算法提取的波段数仍有较多冗余,使得建模变量较多,降低了反演精度。GA 算法作为一种常用的特征选择算法,将 GA 与 PLSR 算法相结合,所得的土壤重金属 Cd 反演精度较好,高于 CC 算法,但是低于 CARS 算法,并且 GA 算法容易出现早熟收敛,使种群多样性降低,丧失进化能力,陷入局部最优解而无法找到全局最优变量。而 CARS 算法能避免上述现象,它通过逐步去除冗余和不重要信息,实现最优波段特征组合的高效提取,达到建模精度最佳,因而利用 CARS 算法进行特征选择的土壤重金属 Cd 反演精度最高,基于有机质特征谱段的 CARS-PLSR 建模精度在郴州矿区土壤实验室光谱数据土壤重金属 Cd 的反演精度 R^2 达到 0.94,为出色模型,在哈密矿区土壤野外实测光谱数据中的反演精度也达到 0.80,为近似模型。

基于有机质特征谱段的 CARS-PLSR 算法在哈密野外实测光谱反演土壤 Cd 含量中的精度 R^2 为 0.80,表明该算法在野外原位测量光谱反演重金属 Cd 含量时仍适用,并可以达到较高的精度。但对比图 5 和图 6,除基于全谱段的 CC-PLSR 算法外,其余结果的估算精度郴州矿区土壤实验室数据均高于哈密矿区土壤野外实测光谱数据。究其原因,郴州矿区土壤实验室光谱数据通过晾干和研磨、过筛等操作,去除了土壤水分和表面粗糙度对光谱测量的影响,光谱曲线更为理想,而哈密矿区野外光谱数据

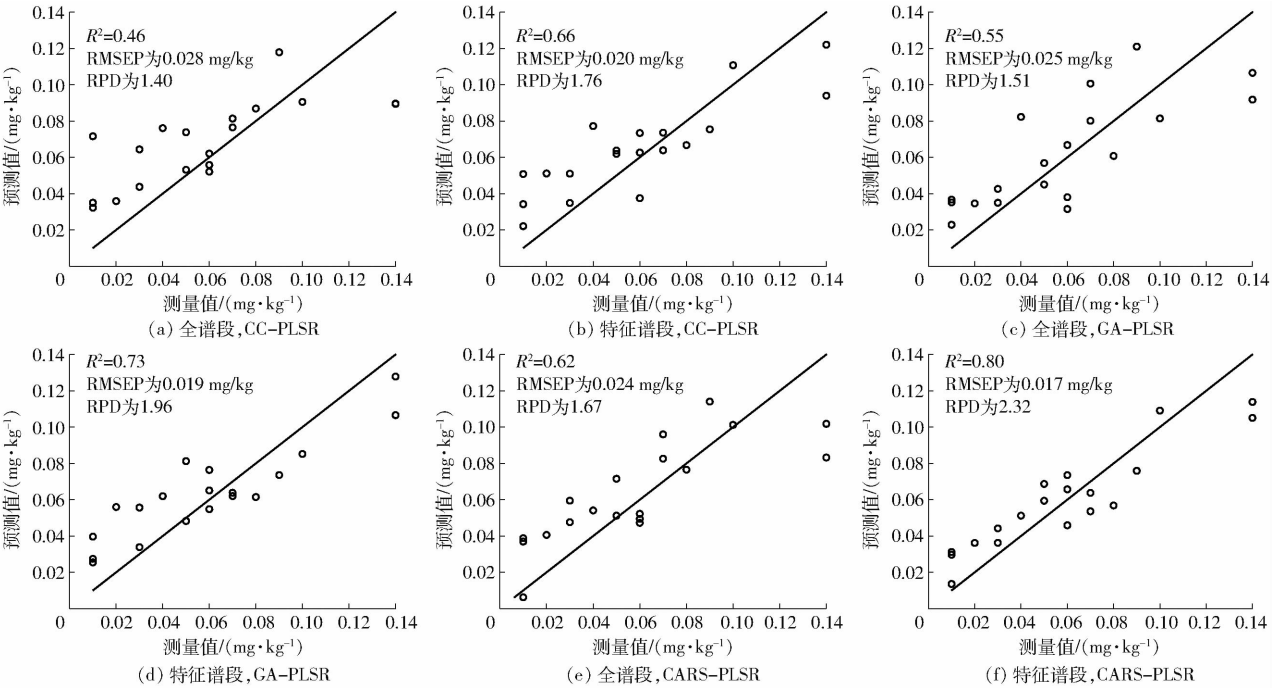


图 6 哈密矿区土壤野外实测光谱重金属 Cd 含量反演最优验证结果散点图

Fig.6 Retrieval results of soil Cd content using Hami mining area field spectra

不可避免地受到这些因素的干扰。对比图 2 和图 3 可以发现,野外环境对哈密矿区野外实测光谱造成了显著的影响,光谱噪声更为严重,对反演精度造成影响,这是导致哈密矿区数据土壤重金属 Cd 反演精度低于郴州矿区数据的主要原因之一。

在提取有机质特征谱段用于土壤 Cd 含量反演时,土壤有机质含量也可能对反演精度有一定影响。本研究所采用的郴州和哈密矿区土壤样本有机质平均含量分别为 28.44 g/kg 和 6.75 g/kg,根据我国第二次土壤普查有机质含量分级表,土壤有机质含量分为 6 级,分别为极高、高、中上、中、低和极低。郴州矿区土壤有机质含量为第 3 级,处于中上水平,而哈密矿区土壤有机质含量为第 5 级,为低等水平,哈密矿区土壤有机质含量低于郴州矿区,这可能也是导致哈密矿区土壤 Cd 反演精度低于郴州矿区的原因。但基于哈密矿区野外光谱数据有机质特征谱段的 CARS-PLSR 算法在重金属 Cd 含量反演时,仍得到 $R^2 = 0.80$ 的较高精度,这说明,在较低的土壤有机质含量下,基于有机质特征谱段的 CARS-PLSR 算法在重金属 Cd 含量反演中仍具有一定适用性。

4 结论

(1)在土壤重金属 Cd 反演中,提取对 Cd 起主要吸附作用的土壤有机质的特征光谱用于含量反演,相较于 400 ~ 2 400 nm 全谱段建模,不仅能够增强反演机理,降低数据冗余,而且还能够显著提高反演精度。对于郴州矿区土壤实验室光谱和哈密矿区土壤野外实测光谱,相较于全谱段建模,基于有机质特征谱段建模的反演精度均得到提高。

(2)CARS 算法通过对光谱数据进行特征寻优,从反射光谱中提取对建模精度影响最大的波段,用于重金属含量反演,在郴州矿区土壤实验室光谱和哈密矿区土壤野外实测光谱的土壤 Cd 含量反演中,有机质特征谱段和全谱段建模中均取得了优于 CC 算法和 GA 算法的反演精度。

(3)本文方法在消除表面粗糙度、水分等环境因素影响的土壤实验室光谱取得 R^2 为 0.94 的高精度,在哈密矿区野外实测光谱受环境因素影响情况下,也获得了较高精度($R^2 = 0.80$),说明本文基于有机质特征谱段的特征选择建模方法在利用野外光谱数据反演重金属含量时也具有一定的适用性。

参 考 文 献

[1] GMOCHOWSKA W, PIETRANIK A, TYSZKA 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.

[2] 环境保护部,国土资源部.全国土壤污染状况调查公报[J]. *中国环保产业*, 2014, 36(5):10–11.

[3] WANG X B, LI X Y, YAN X, et al. Environmental risks for application of iron and steel slags in soils in China: a review[J]. *Pedosphere*, 2021, 31(1):28–42.

[4] 于海洋,谢赛飞,郭灵辉,等.融合光谱和空间特征的土壤重金属含量极端随机树估算[J]. *农业机械学报*, 2022, 53(8):231–239.

YU Haiyang, XIE Saifei, GUO Linghui, et al. Extremely randomized trees estimation of soil heavy metal content by fusing spectra and spatial features[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(8):231–239. (in Chinese)

[5] KHOSRAVI V, GHOLIZADEH A, AGYEMAN P C, et al. Further to quantification of content, can reflectance spectroscopy determine the speciation of cobalt and nickel on a mine waste dump surface[J]. *Science of the Total Environment*, 2023, 872:161996.

[6] TAN K, MA W, CHEN L, et al. Estimating the distribution trend of soil heavy metals in mining area from HyMap airborne hyperspectral imagery based on ensemble learning[J]. *Hazard Mater*, 2021, 401:123288.

[7] ZOU B, JIANG X L, FENG H H, 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.

[8] WANG X, AN S, XU Y Q. 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*, 2020, 10(1):51.

[9] 成永生,周瑶.土壤重金属高光谱遥感定量监测研究进展与趋势[J]. *中国有色金属学报*, 2021, 31(11):3450–3467.

CHENG Yongsheng, ZHOU Yao. Research progress and trend of quantitative monitoring of hyperspectral remote sensing for heavy metals in soil[J]. *The Chinese Journal of Nonferrous Metals*, 2021, 31(11):3450–3467. (in Chinese)

[10] 孙伟超.基于土壤光谱活性物质吸附作用的重金属高光谱定量反演[D].北京:中国科学院大学,2019.

SUN Weichao. Estimation of heavy metal concentration in soil using hyperspectral remote sensing based on the sorption on spectrally active soil constituents[D]. Beijing: University of Chinese Academy of Sciences, 2019. (in Chinese)

[11] SUN W C, LIU S, ZHANG X, et al. Performance of hyperspectral data in predicting and mapping zinc concentration in soil[J]. *Science of the Total Environment*, 2022, 824:153766.

[12] SUN W C, ZHANG X, SUN X J, et al. Predicting nickel concentration in soil using reflectance spectroscopy associated with

- organic matter and clay minerals[J]. *Geoderma*, 2018, 327:25 – 35.
- [13] YANG H F, XU H, ZHONG X N. Prediction of soil heavy metal concentrations in copper tailings area using hyperspectral reflectance[J]. *Environmental Earth Sciences*, 2022, 81(6):183.
- [14] TAN K, WANG H M, CHEN L H, et al. Estimation of the spatial distribution of heavy metal in agricultural soils using airborne hyperspectral imaging and random forest[J]. *Journal of Hazardous Materials*, 2020, 382:120987.
- [15] 阳洋, 黄伟濠, 卢瑛, 等. 土壤游离氧化铁高光谱特征与定量反演[J]. *华南农业大学学报*, 2020, 41(1):91 – 99.
YANG Yang, HUANG Weihao, LU Ying, et al. Spectral characteristics and quantitative retrieval of free iron content in soil[J]. *Journal of South China Agricultural University*, 2020, 41(1):91 – 99. (in Chinese)
- [16] ZHAO L, HU Y M, ZHOU W, et al. Estimation methods for soil mercury content using hyperspectral remote sensing[J]. *Sustainability*, 2018, 10(7):2474.
- [17] 周冰, 李秉璇, 贺宣, 等. 融合自适应稀疏表示和相关系数的高光谱伪装分类方法[J]. *光谱学与光谱分析*, 2021, 41(12):3851 – 3856.
ZHOU Bing, LI Bingxuan, HE Xuan, et al. Classification of camouflages using hyperspectral images combined with fusing adaptive sparse representation and correlation coefficient[J]. *Spectroscopy and Spectral Analysis*, 2021, 41(12):3851 – 3856. (in Chinese)
- [18] 江晓宇, 李福生, 王清亚, 等. X 射线荧光光谱结合 CARS 变量筛选选择方法用于土壤中铅砷含量的测定[J]. *光谱学与光谱分析*, 2022, 42(5):1535 – 1540.
JIANG Xiaoyu, LI Fusheng, WANG Qingya, et al. Determination of lead and arsenic in soil samples by X fluorescence spectrum combined with CARS variables screening method[J]. *Spectroscopy and Spectral Analysis*, 2022, 42(5):1535 – 1540. (in Chinese)
- [19] PAVEZ L B, SOTO C. A deterministic annular crossover genetic algorithm optimisation for the unit commitment problem[J]. *Expert Systems with Applications*, 2021, 38(6):6523 – 6529.
- [20] VOHLAND M, HARBICH M, LUDWIG M, et al. Quantification of soil variables in a heterogeneous soil region with VIS – NIR – SWIR data using different statistical sampling and modeling strategies[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2016, 9(9):4011 – 4021.
- [21] VOHLAND M, LUDWIG M, THIELE-BRUHN S, et al. Quantification of soil properties with hyperspectral data: selecting spectral variables with different methods to improve accuracies and analyze prediction mechanisms[J]. *Remote Sensing*, 2017, 9(11):1103.
- [22] 刘泓君, 牛腾, 于强, 等. 北京市平谷区桃林土壤重金属元素含量反演估测[J]. *光谱学与光谱分析*, 2022, 42(11):3552 – 3558.
LIU Hongjun, NIU Teng, YU Qiang, et al. Inversion and estimation of heavy metal element content in peach forest soil in Pinggu District of Beijing[J]. *Spectroscopy and Spectral Analysis*, 2022, 42(11):3552 – 3558. (in Chinese)
- [23] KOOISTRA L, SALAS E, 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.
- [24] 宁鸿章, 谭鑫, 李宇航, 等. 空-谱维联合 Savitzky – Golay 高光谱滤波算法及其应用[J]. *光谱学与光谱分析*, 2020, 40(12):3699 – 3704.
NING Hongzhang, TAN Xin, LI Yuhang, et al. Joint space – spectrum SG filtering algorithms for hyperspectral images and its application[J]. *Spectroscopy and Spectral Analysis*, 2020, 40(12):3699 – 3704. (in Chinese)
- [25] 张霞, 丁松滔, 岑奕, 等. 结合野外与实验室光谱的土壤 Pb 含量反演[J]. *武汉大学学报(信息科学版)*, 2022, 47(9):1479 – 1485.
ZHANG Xia, DING Songtao, CEN Yi, et al. Soil heavy metal Pb content estimation method by combining field spectra with laboratory spectra[J]. *Geomatics and Information Science of Wuhan University*, 2022, 47(9):1479 – 1485. (in Chinese)
- [26] 涂宇龙, 邹滨, 姜晓璐, 等. 矿区土壤 Cu 含量高光谱反演建模[J]. *光谱学与光谱分析*, 2018, 38(2):575 – 581.
TU Yulong, ZOU Bin, JIANG Xiaolu, et al. Hyperspectral remote sensing based modeling of Cu content in mining soil[J]. *Spectroscopy and Spectral Analysis*, 2018, 38(2):575 – 581. (in Chinese)
- [27] ZHANG S, SHEN Q, NIE C, et al. Hyperspectral inversion of heavy metal content in reclaimed soil from a mining wasteland based on different spectral transformation and modeling methods[J]. *Spectrochim Acta A Mol Biomol Spectrosc*, 2019, 211:393 – 400.
- [28] 贺军亮, 韩超山, 韦锐, 等. 基于偏最小二乘的土壤重金属镉间接反演模型[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)
- [29] LU Q, WANG S J, BAI X Y, et al. Rapid inversion of heavy metal concentration in Karst grain producing areas based on hyperspectral bands associated with soil components[J]. *Microchemical Journal*, 2019, 148:404 – 411.
- [30] 郭斌, 白昊睿, 张波, 等. 基于 RF 和连续小波变换的露天煤矿土壤锌含量高光谱遥感反演[J]. *农业工程学报*, 2022, 38(10):138 – 147.
GUO Bin, BAI Haorui, ZHANG Bo, et al. Inversion of soil zinc contents using hyperspectral remote sensing based on random

forest and continuous wavelet transform in an opencast coal mine[J]. Transactions of the CSAE, 2022, 38(10):138 – 147. (in Chinese)

[31] WEI L F, YUAN Z R, ZHONG Y F, et al. An improved gradient boosting regression tree estimation model for soil heavy metal (Arsenic) pollution monitoring using hyperspectral remote sensing[J]. Applied Sciences, 2019, 9(9):1943.

[32] 贺军亮, 崔军丽, 张淑媛, 等. 基于偏最小二乘的土壤重金属铜含量高光谱估算[J]. 遥感技术与应用, 2019, 34(5):998 – 1004.

HE Junliang, CUI Junli, ZHANG Shuyuan, et al. Hyperspectral estimation of heavy metal Cu content in soil based on partial least square method[J]. Remote Sensing Technology and Application, 2019, 34(5):998 – 1004. (in Chinese)

[33] 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.

[34] 章明奎, 郑顺安, 王丽平. 土壤中颗粒状有机质对重金属的吸附作用[J]. 土壤通报, 2007(6):1100 – 1104.

ZHANG Mingkui, ZHENG Shun'an, WANG Liping, et al. Adsorption of heavy metals by soil particulate organic matter[J]. Chinese Journal of Soil Science, 2007(6):1100 – 1104. (in Chinese)

[35] 马闯, 王雨阳, 周通, 等. 污染土壤颗粒态有机质镉锌富集特征及其解吸行为研究[J]. 生态环境学报, 2022, 31(9):1892 – 1900.

MA Chuang, WANG Yuyang, ZHOU Tong, et al. Enrichment characteristics and desorption behavior of cadmium and zinc in particulate organic matter of polluted soil[J]. Ecology and Environmental Sciences, 2022, 31(9):1892 – 1900. (in Chinese)

(上接第 46 页)

[12] 陈建能, 李杭, 任萍, 等. 螺旋管吸附式名优茶采收器设计与试验[J]. 农业机械学报, 2023, 54(5):131 – 139.

CHEN Jianneng, LI Hang, REN Ping, et al. Design and experiment of threaded tube adsorption type famous tea harvester[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(5):131 – 139. (in Chinese)

[13] 贾江鸣, 叶玉泽, 程培林, 等. 手持式名优茶嫩梢采摘机械手设计与试验优化[J]. 农业机械学报, 2022, 53(5): 86 – 92.

JIA Jiangming, YE Yuze, CHENG Peilin, et al. Design and experimental optimization of hand-held manipulator for picking famous tea shoot[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(5): 86 – 92. (in Chinese)

[14] 许丽佳, 刘琦, 代建武, 等. 茶叶嫩梢采摘末端执行器的设计研究[J]. 茶叶科学, 2021, 41(5): 705 – 716.

XU Lijia, LIU Qi, DAI Jianwu, et al. Design of end effector for picking tea shoots[J]. Journal of Tea Science, 2021, 41(5): 705 – 716. (in Chinese)

[15] MICKOVSKI S B, STOKES A, VAN BEEK R, et al. Simulation of direct shear tests on rooted and non-rooted soil using finite [J]. Ecological Engineering, 2011, 37: 1523 – 1532.

[16] KELLN C, SHARMA J, HUGHES D. A finite element solution scheme for an elastic-viscoplastic soil model[J]. Computers and Geotechnics, 2008, 35: 524 – 536.

[17] SOUZA V H S, SANTOS A A R, COSTA A L G, et al. Evaluation of the interaction between a harvester rod and a coffee branch based on finite element analysis[J]. Computers and Electronics in Agriculture, 2018, 150: 476 – 483.

[18] MENG Y, WEI J, WEI J, et al. An ANSYS LS-DYNA simulation and experimental study of circular saw blade cutting system of mulberry cutting machine[J]. Computers and Electronics in Agriculture, 2019, 157: 38 – 48.

[19] 施印炎, 陈满, 汪小岳, 等. 芦蒿有序收获机切割器动力学仿真与试验[J]. 农业机械学报, 2017, 48(2): 110 – 116.

SHI Yinyan, CHEN Man, WANG Xiaochan, et al. Dynamic simulation and experiments on *Artemisia selengensis* orderly harvester cutter[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(2): 110 – 116. (in Chinese)

[20] 黄汉东, 王玉兴, 唐艳芹, 等. 甘蔗切割过程的有限元仿真[J]. 农业工程学报, 2011, 27(2): 161 – 166.

HUANG Handong, WANG Yuxing, TANG Yanqin, et al. Finite element simulation of sugarcane cutting[J]. Transactions of the CSAE, 2011, 27(2): 161 – 166. (in Chinese)

[21] KOVACS A, KERENYI G. Physical characteristics and mechanical behaviour of maize stalks for machine development[J]. International Agrophysics, 2019, 33: 427 – 436.

[22] 吴昆, 宋月鹏. 农作物茎秆切割理论与方法研究进展分析[J]. 农业机械学报, 2022, 53(6): 1 – 20.

WU Kun, SONG Yuepeng. Research progress analysis of crop stalk cutting theory and method[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(6): 1 – 20. (in Chinese)

[23] 贾江鸣, 叶玉泽, 程培林, 等. 细长果蔬采摘软体气动抓手设计与参数优化[J]. 农业机械学报, 2021, 52(6): 26 – 34.

JIA Jiangming, YE Yuze, CHENG Peilin, et al. Design and parameter optimization of soft pneumatic gripper for slender fruits and vegetables picking[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(6): 26 – 34. (in Chinese)

[24] XIE L, WANG J, CHENG S, et al. Optimisation and finite element simulation of the chopping process for chopper sugarcane harvesting[J]. Biosystems Engineering, 2018, 175:16 – 26.