

光学学报

基于多尺度连续小波分解的土壤氧化铁反演

赵海龙¹, 甘淑^{1,2*}, 袁希平^{2,3}, 胡琳¹, 刘帅¹, 王俊杰¹

¹昆明理工大学国土资源工程学院, 云南 昆明 650093;

²云南省高校高原山区空间信息测绘技术应用工程研究中心, 云南 昆明 650093;

³滇西应用技术大学, 云南 大理 671000

摘要 为准确快速地预测土壤氧化铁含量信息,以禄丰恐龙谷南缘地表的土壤为研究对象,共采集 135 个样品,并在室内测得土壤光谱数据和氧化铁含量信息。首先,在对原始光谱进行 Savitzky-Golay 滤波平滑后,进行常规光谱变换和连续小波变换,并利用相关系数(CC)法对变换光谱与氧化铁含量进行相关性分析,筛选出每个尺度中通过 0.01 显著性检验的波长作为粗选的波长。然后,进一步利用竞争性自适应重加权(CARS)算法选择的波长作为特征波长。最后,通过遗传算法优化的支持向量机(SVR)进行建模。结果表明:连续小波变换可以提高土壤光谱反射率与氧化铁含量的相关性;通过 CC-CARS 波长选择方法可以有效地减少建模的自变量数目;第 4 尺度连续小波分解构建的模型(L₄-CC-CARS-SVR)效果最好,其建模集的决定系数 R^2 为 0.760,均方根误差 E_{RMSE} 为 $5.236 \text{ g}\cdot\text{kg}^{-1}$,验证集的 R^2 为 0.663, E_{RMSE} 为 $7.798 \text{ g}\cdot\text{kg}^{-1}$,性能与四分位数间距比 R_{PIQ} 达到了 2.598,即模型具有很好的稳定性和预测能力。

关键词 光谱学; 土壤; 氧化铁; 高光谱; 连续小波变换; 遗传算法; 支持向量机

中图分类号 TP79

文献标志码 A

DOI: 10.3788/AOS202242.2230002

Inversion of Soil Iron Oxide Based on Multi-Scale Continuous Wavelet Decomposition

Zhao Hailong¹, Gan Shu^{1,2*}, Yuan Xiping^{2,3}, Hu Lin¹, Liu Shuai¹, Wang Junjie¹

¹Faculty of Land and Resources Engineering, Kunming University of Science and Technology, Kunming 650093, Yunnan, China;

²Yunnan Institute of Engineering Research and Application of Plateau Mountain Spatial Information Surveying and Mapping Technology, Kunming 650093, Yunnan, China;

³West Yunnan University of Applied Sciences, Dali 671000, Yunnan, China

Abstract In order to predict the content of iron oxide in soil accurately and quickly, 135 soil samples are collected from the southern edge of Lufeng Dinosaur Valley, and soil spectral data and iron oxide content are measured in the laboratory. The original spectrum is smoothed by the Savitzky-Golay filter, and then conventional spectral transform and continuous wavelet transform are performed. The correlation coefficient (CC) method is used to analyze the correlation between the transform spectrum and iron oxide content. The wavelengths that pass the 0.01 significance test in each scale are selected as the coarse wavelengths, and the wavelengths selected by the competitive adaptive reweighted sampling (CARS) are further used as the characteristic wavelengths. Finally, the support vector regression (SVR) optimized by the genetic algorithm is used for modeling. The results reveal that continuous wavelet transform can improve the correlation between the soil spectral reflectance and iron oxide content. The number of independent variables for modeling can be effectively reduced by the CC-CARS wavelength selection method. The model constructed by the fourth-scale continuous wavelet decomposition (L₄-CC-CARS-SVR) has the best effect. The coefficient of determination R^2 and root-mean-square error E_{RMSE} of its calibration set is 0.760 and $5.236 \text{ g}\cdot\text{kg}^{-1}$, respectively. The R^2 , E_{RMSE} and performance to interquartile range ratio R_{PIQ} of its validation set is 0.663, $7.798 \text{ g}\cdot\text{kg}^{-1}$ and 2.598, respectively, which indicates that the model has good stability and predictive ability.

Key words spectroscopy; soil; iron oxide; high-spectrum; continuous wavelet transform; genetic algorithm; support

收稿日期: 2022-04-08; 修回日期: 2022-05-18; 录用日期: 2022-05-27

基金项目: 国家自然科学基金(41561083,41861054)、云南省自然科学基金(2015FA016)

通信作者: *1193887560@qq.com

1 引言

土壤的环境质量通常以有机质、全氮和重金属元素作为重要的指标^[1]。中国云南省楚雄彝族自治州禄丰县恐龙谷南缘地区(下文简称禄丰恐龙谷南缘地区)地形地貌复杂,农业生产活动频繁,掌握土壤中氧化铁含量信息对于揭示云南高原土壤环境状况,指导农作物生产具有重要意义。因此,本文选择禄丰恐龙谷南缘地区作为研究区进行土壤采样,并尝试进行该地区土壤氧化铁含量的定量分析研究。

常规的氧化铁含量测定技术成本高、周期长,利用高光谱技术可以快速且高效地进行土壤元素的测定。高光谱以其光谱分辨率高、波段连续性强的特点^[2],被广泛用来进行土壤元素含量的反演^[3-5],从而为精准农业提供技术支撑^[6]。目前国内外学者对土壤中氧化铁含量的高光谱反演研究均取得了较好的成果。何挺等^[7]指出土壤中氧化铁含量与反射率呈负相关,在400、450、490、700、870 nm等处存在吸收峰,且氧化铁含量反演的精度可能与土壤类型分类相关。谭洁等^[8]指出不同森林土壤氧化铁含量和不同土壤类型光谱曲线在可见光-近红外波段内变化趋势基本一致,均形似陡坎。在光谱变换方面的研究成果有:谭洋等^[9]以东北黑土的总铁含量进行反演,先对原始光谱进行标准正态处理,再进一步进行连续小波变换,此时验证集决定系数达到了0.67,比常规光谱变换方法要高;郭颖等^[10]以亚热带水稻土为研究对象,借助一阶微分(FD)、二阶微分和倒数的对数(RL)等光谱变换预处理方法发现,微分光谱能够显著提升光谱与氧化铁含量之间的相关性;于雷等^[11]通过对三种土壤类型的有机质进行反演,利用连续小波变换对原始光谱和连续统去除光谱进行预处理,研究结果表明连续小波处理可以抑制背景噪声的干扰,挖掘土壤的有效信息,并提高土壤光谱与有机质的相关性。基于特征波长筛选方面的研究有:丁海宁等^[12]利用相关系数峰值-逐步分析方法筛选出了氧化铁的特征波长,构建的估算模型均较稳定;李双权等^[13]利用相关性达到0.05显著性水平的波长构建了氧化铁的偏最小二乘回归估算模型,模型的决定系数达到了0.7906;阳洋等^[14]提取的敏感波长是利用多元逐步线性手段剔除通过0.01显著性检验的波长中的共线性波长获得的,最终构建的模型的决定系数可达到0.79。李冠稳等^[15]利用稳定性自适应重加权采样、连续投影算法、遗传算法和迭代保留有效信息变量来筛选特征波长,并与回归方法结合建模,发现不仅模型运算效率得到了提高,模型的预测能力较全波段还有一定的提高。

综合来看,前人研究在对具体的土壤类型进行氧化铁含量的反演时均取得了很好的效果,但在面对不同土壤类型和不同土地利用类型时,反演精度难以保证。尤其是,在有机质含量丰富的地方,有机质对氧化

铁的光谱特征具有掩盖作用。因此,本文先以不同土壤类型和不同土地利用类型的土壤为研究对象,对原始光谱进行常规光谱变换和连续小波变换(CWT),并利用特征波长选择的方法选出特征波长和敏感小波系数。然后,以基于遗传算法的支持向量机(SVR)回归建立土壤氧化铁的估算模型。最后,通过对比各模型的精度,选出最佳的模型,进而实现土壤氧化铁含量的快速、准确预测。

2 材料与方 法

2.1 研究区概况

禄丰恐龙谷南缘地区位于云南省中部滇中高原的禄丰县境内,地理坐标为:24°55′25″N~25°22′05″N,102°00′00″E~102°9′00″E。研究区所在的禄丰县属亚热带低纬度高原季风气候。禄丰内陆高原盆地是一个小型的中生代红色沉积盆地,根据禄丰县1982年和1985年两次土壤调查资料,县内有5个土类(棕壤、黄棕壤、红壤、紫色土和水稻土)、10个亚类、20个土属和40个土种,其中:紫色土占土地面积的56.9%,为辖区内最主要的土壤类型;其次为红壤,占土地面积的22.8%;黄棕壤占7.8%;水稻土占6.3%^[16]。

2.2 数据采集与数据获取

土壤样品于2021年7月底采于禄丰恐龙谷南缘地区山地,根据地势地貌差异进行采样点布设。每个样本点采样范围为5 m×5 m。在采样范围内,根据5点采样法,采集5~20 cm深度范围内的表层土壤,取约1 kg重的土壤装袋保存,并用手持全球定位系统(GPS)记录该点的经纬度坐标,共135份样品,采样点分布如图1所示。根据实地调查,土壤类型有紫壤、红壤和黄棕壤,土地利用类型有林地和耕地。将采集回来的土壤首先去除杂草、石块等杂质,再进行自然风干处理,最后用玛瑙球研磨机研磨并过100目筛。将处理后的每份样品分成两份,一份用于测定土壤氧化铁的含量,另一份用于土壤的高光谱数据测量。结合样品分析质量要求等技术规范,在满足样品的检出限、准确度和精密度的前提下,参照《土壤农业化学分析方法》,选用X射线荧光光谱法进行土壤氧化铁含量的测定。

土壤光谱测量在暗室中进行,光谱仪为ASD Field Spec 3地物光谱分析仪,光源为探头内置卤素光源,探头内径为21 mm,前视场角为25°,波段范围为350~2500 nm。在350~1000 nm波长范围内,光谱采样间隔和分辨率分别为1.4 nm和3 nm。在1000~2500 nm波长范围内,光谱采样间隔和分辨率分别为2 nm和10 nm。将光谱采样间隔重采样为1 nm后,得到的波长数为2151^[17]。在进行光谱测量时,把土壤样品放在宽10 cm、高2 cm的容器内,并把土壤样品刮平,以减少土壤样品粗糙度对光谱测量的影响。探头与土壤样品保持2 cm的高度,并垂直对准样品。每个

样品采集 5 条光谱曲线,对 5 条光谱曲线取平均后作为样本的实际光谱反射率。

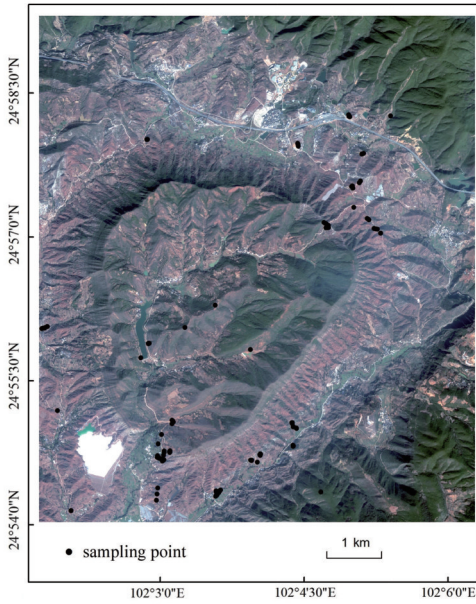


图 1 土壤采样点分布图

Fig. 1 Distribution map of soil sampling points

2.3 数据预处理

受仪器自身的影响,350~399 nm、2451~2500 nm 波段处信噪比较低,因此对这两段光谱数据进行去除处理,共得到 2051 个波长。为了消除仪器噪声、土壤颗粒分布不均匀和随机因素的干扰,将窗口数为 9、多项式阶数为 2 的 Savitzky-Golay^[18]平滑之后的曲线作为原始光谱(OS)。此外,进一步对 OS 进行 FD 和 RL 变换处理。

小波变换包括离散小波变换和 CWT。利用离散小波变换进行尺度分解时,光谱维数会随着尺度分解而减半,故不能对土壤光谱数据展开较好的分析。由于土壤光谱曲线特征与 Gaussian 函数相似^[19],因此本研究选取 Gaussian4 函数作为小波基函数,对 OS 进行 2^1 、 2^2 、 2^3 、 \dots 、 2^{10} 的 CWT 分解,即 1~10 尺度的 CWT 分解,并生成一系列的小波系数。研究中使用的小波基函数为

$$\psi_{a,b}(\lambda) = \frac{1}{\sqrt{a}} \psi\left(\frac{\lambda - b}{a}\right), \quad (1)$$

式中: λ 为波谱曲线的波段数; a 为伸缩因子; b 为平移因子; $\psi(\cdot)$ 为小波基函数。小波分解后可以得到不同波长和分解尺度下的小波系数,即

$$W_f(a,b) = \int_{-\infty}^{+\infty} f(\lambda) \psi_{a,b}(\lambda) d\lambda, \quad (2)$$

式中: $f(\lambda)$ 为光谱的反射率。小波系数 $W_f(a,b)$ 是一个行为尺度数,其列为波长数的矩阵^[20]。

2.4 特征波长选择

由于高光谱数据冗余度较大,故在进行回归分析时,并不是所有的波长都对建模精度的提高有益。若对所有的波长都进行建模分析,则不仅运算量巨大,还

会降低建模精度。因此,在建模之前进行特征波长选取是有必要的。

在相关系数(CC)法中,将氧化铁含量逐波长与变换光谱进行 Pearson 相关性分析,并计算每个波长与土壤样本氧化铁含量的相关系数,把通过 $P=0.01$ 显著性检验的波长作为特征波长。相关系数的表达式为

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}, \quad (3)$$

式中: r 为土壤氧化铁含量与光谱值的相关系数; n 为样本个数; X_i 是第 i 个样本的光谱值; Y_i 是第 i 个样本的实测值; \bar{X} 是 X_i 的平均值; \bar{Y} 是 Y_i 的平均值。

竞争性自适应重加权(CARS)算法是一种基于蒙特卡罗采样和偏最小二乘(PLS)回归系数的特征波长选择方法,把每个变量当成一个个体,并选择适应能力较强的个体。具体步骤为:随机选取固定比率的样本作为校正集,并建立 PLS 模型;计算模型的回归系数的绝对值和每个波长对应的权重,并利用指数衰减函数和自适应重加权采样法对变量进行选择;计算交叉验证均方根误差,在 n 次采样后,选择均方根误差最小的子集作为最优变量子集^[21]。

2.5 模型建立与验证

支持向量机模型是由 Vapni 等提出的一种基于统计学习理论的非线性建模方法。支持向量机模型利用训练样本中的支持向量来设计一个最优的决策边界,来处理线性和非线性问题,特别是在处理高维和小样本数据时表现良好。本文选择支持向量机的核函数为高斯核函数,为了让模型更稳定和结果更可靠,模型训练过程中利用基于差分进化的遗传算法对支持向量机的参数(c 和 γ)进行调优。与传统的遗传算法相比,基于差分进化的遗传算法具有鲁棒性强、收敛速度快和全局优化搜索能力强的优点。基于差分进化的遗传算法的参数设置为: c 和 γ 的取值范围为 2^{-8} ~ 2^8 ,种群数量为 50,编码方式为实数编码,基向量的选择方式为精英复制选择,变异算子 F 为 0.5,交叉算子为 $F_{CR}=0.5$,最大进化代数 1000,适应度函数为交叉验证的平均均方根误差。算法利用第三方库 Geatpy 实现。

采用 Kennard-Stone(K-S)算法进行建模集和验证集的划分,选用 70% 的样本为建模集,剩下的 30% 为验证集。将根据 CC 和 CC-CARS 分别筛选出的特征波长和敏感小波系数作为自变量,并将土壤氧化铁含量作为因变量来分别构建模型。由于土壤样本呈现非正态分布,故采用性能与四分位数间距比(R_{PIQ})可给出模型更真实的评价^[22]。反演模型的拟合效果和预测精度均采用决定系数(R^2)、均方根误差(E_{RMSE})、 R_{PIQ} 和 1:1 线 4 个参数来衡量。其中, R^2 、 E_{RMSE} 和 R_{PIQ} 的表达式为

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y_i^*)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (4)$$

$$E_{\text{RMSE}} = \sqrt{\frac{\sum_{i=1}^n (y_i - y_i^*)^2}{n}}, \quad (5)$$

$$R_{\text{PIQ}} = \frac{D_{\text{IQ}}}{E_{\text{RMSE}}}, \quad (6)$$

式中: y_i^* 为第 i 个样本的观测值; y_i 为第 i 个样本的真实值; \bar{y} 为样本的平均值; D_{IQ} 为样本第三个四分位数和第一个四分位数之差。 R^2 越大, 表示模型越稳定。 E_{RMSE}

越小, 表示模型精度越高。 R_{PIQ} 越大, 表示模型的预测能力越好^[23]。

3 结果与分析

3.1 土壤氧化铁含量分析

利用 K-S 法选出来的建模集共有 95 个样本, 验证集共有 40 个样本。利用 Origin 软件对获取的土壤氧化铁含量进行统计, 得到的相关统计参数如表 1 所示。

表 1 氧化铁含量统计特征

Table 1 Statistical characteristics of iron oxide content

Sample classification	Sample number	Maximum / ($\text{g} \cdot \text{kg}^{-1}$)	Minimum / ($\text{g} \cdot \text{kg}^{-1}$)	Mean / ($\text{g} \cdot \text{kg}^{-1}$)	Standard deviation / ($\text{g} \cdot \text{kg}^{-1}$)	Variable coefficient / %
Total set	135	66.978	18.293	41.201	11.698	28.393
Calibration set	95	64.808	23.311	42.141	10.736	25.476
Validation set	40	66.978	18.293	38.969	13.605	34.912
Red soil	68	66.978	23.297	46.838	11.804	25.016
Purple soil	48	62.365	23.311	35.834	7.840	21.880
Yellow brown soil	19	54.356	18.293	34.583	9.559	28.400

可以看出: 研究区中土壤氧化铁含量的最小值为 $18.293 \text{ g} \cdot \text{kg}^{-1}$, 最大值为 $66.978 \text{ g} \cdot \text{kg}^{-1}$, 均值为 $41.201 \text{ g} \cdot \text{kg}^{-1}$, 变异系数为 28.393%, 属于中等变异; 建模集和验证集的平均值和标准差差距不大, 故可以认为建模集和验证集属于同一分布; 红壤的平均氧化铁含量最高, 其次为紫色土, 最低的为黄棕壤。

3.2 土壤原始高光谱响应特征

全部土壤样本的原始光谱曲线如图 2 所示。可以发现: 研究区的土壤样本的反射率范围在 0.02~0.49 之间, 总体反射率不高; 在 400~550 nm 波段处斜率上升缓慢; 随着波长的增加, 在 550~600 nm 波段处斜率剧增, 形似陡坎; 在 600~1000 nm 波段处, 斜率变化不均匀, 形成几个波折; 在 500 nm 和 900 nm 波长处有微弱氧化铁吸收峰; 1400 nm 波长处是 OH 和 H_2O 二者的合成峰; 1900 nm 波长处的峰是 H_2O 的吸收峰; 2210 nm 波长处的峰是由黏土矿物中金属-OH 振动的合频产生的。

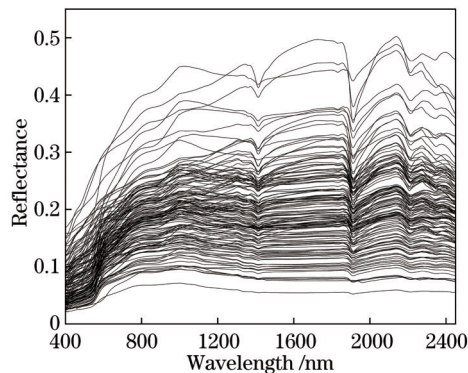


图 2 全部土壤样本原始光谱反射率曲线

Fig. 2 Original spectral reflectance curves of all soil samples

3.3 变换光谱与土壤氧化铁含量的相关性

对光谱进行数学变换可以降低背景干扰的影响, 扩大光谱差异, 提供更加明显的光谱特征。因此, 本研究分别对 OS、FD 和 RL 与氧化铁含量进行 Pearson 相关性分析, 得到的变换光谱反射率与氧化铁含量的相关性如图 3 所示。

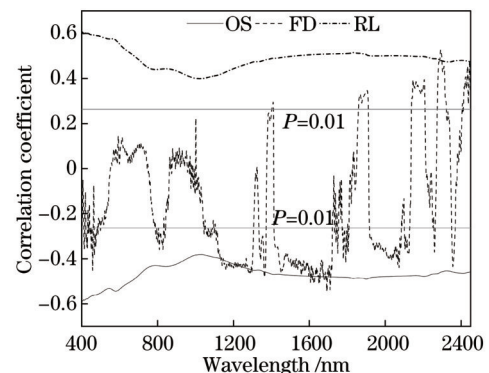


图 3 光谱与氧化铁含量的相关系数

Fig. 3 Correlation coefficient between spectrum and iron oxide content

由分析结果可知: OS 与氧化铁含量呈负相关, 在 400~600 nm 波长范围内相关系数较大, 其相关系数在 402 nm 波长处达到最大值 ($r = -0.589$), 全波段均通过 0.01 显著性检验; FD 与氧化铁含量的相关系数在 1693 nm 波长处达到最大值 ($r = -0.548$), 通过 0.01 显著性检验的波长数为 1451; RL 与氧化铁含量的相关系数在 403 nm 波长处达到最大值 ($r = 0.606$), 全波段均通过 0.01 显著性检验。FD 与 OS 相比, 最大相关系数绝对值降低了 0.041, 这说明微分变换虽然能够消除背景噪声的干扰且能对重叠混合光谱进行分

解,但是会增加高频噪声,进而导致相关性降低;RL的最大相关系数绝对值比OS提高了0.017,说明RL不仅可以增强可见光的光谱差异,还可以减少光照、粗糙度等外界条件的干扰,从而提升与氧化铁含量的相关性。

3.4 小波系数与土壤氧化铁含量的相关性

将OS进行1~10尺度的CWT分解,生成一系列的小波系数,并将各尺度小波系数与土壤氧化铁含量进行相关性分析。将第一尺度记为 L_1 ,第二尺度记为 L_2 ,第 m 尺度记为 L_m 。实验中OS经过CWT分解后,其与氧化铁含量的决定系数热力图如图4所示。

由图4可知:小波系数与氧化铁含量的决定系数较高的主要分布在第3~6尺度的可见光波段中,最高决定系数可达到0.367,说明有效信息主要集中在第3~5尺度;在第1尺度和第2尺度下,决定系数较低,说明部分光谱特征消失,有效信息较少。各尺度通过 $P=0.01$ 显著性检验的波长数和最大相关系数如表2

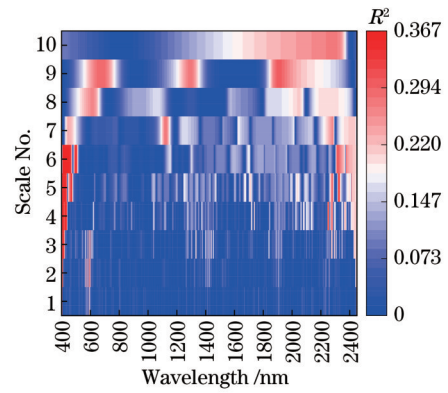


图4 小波系数与氧化铁含量的决定系数热力图
Fig. 4 Thermodynamic diagram of wavelet coefficient and coefficient of determination of iron oxide content 所示。

表2 变换光谱与氧化铁含量相关性分析

Table 2 Correlation analysis of transform spectrum and iron oxide content

Transform spectrum	Wavelet decomposition scale	Significant band number	Maximum correlation coefficient
CWT	L_1	92	-0.590
	L_2	171	-0.593
	L_3	395	0.606
	L_4	663	-0.602
	L_5	1136	0.603
	L_6	1056	-0.604
	L_7	1320	0.527
	L_8	1357	-0.511
	L_9	1273	-0.548
	L_{10}	1447	-0.523
OS		2051	-0.589
FD		1252	-0.548
RL		2051	0.606

从表2可以看出:CWT的最大相关系数为0.606,其最大相关系数绝对值分别比OS和FD光谱大0.017和0.058,与RL持平。同时,各尺度小波系数与氧化铁含量的相关系数绝对值呈现先增大后减小的趋势,显著波长数呈现逐步增加的趋势。

3.5 特征波长选择

如果直接利用变量优选算法对全波段进行筛选,则不仅会造成效率低下,还可能会造成对建模有用的信息丢失。因此,本文先利用相关性分析进行变量的粗选,再利用CARS算法进行变量的精选。具体步骤为:将训练集样本OS、FD、RL和CWT与氧化铁含量进行相关性分析,把通过0.01显著性检验的波长作为粗选的特征波长;利用CARS对光谱数据进行进一步的降维。由于蒙特卡罗采样方法是不稳定的,故多次运行的结果有所不同。因此,本文在相关性分析的基础上,利用CARS循环进行了50次实验,将得到的结果中出现了20次以上的波长作为特征波长。根据

CARS算法选择的特征波长结果如图5所示。

由图5可知:第1~3尺度选择的特征波长在可见光和近红外波段中均有分布;第4~7尺度选择的特征

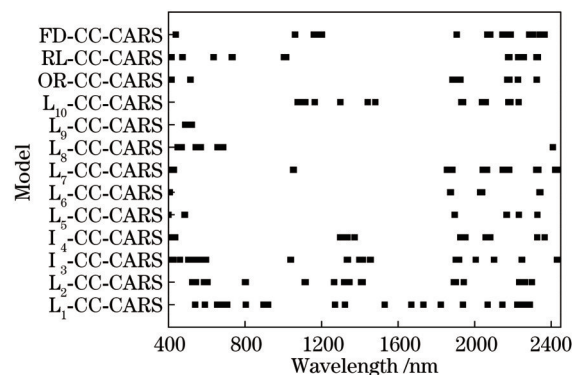


图5 CC-CARS算法选择的特征波长
Fig. 5 Characteristic wavelengths selected by CC-CARS algorithm

波长主要分布在 400~500 nm 和 1900~2450 nm 波段内;第 8 尺度和第 9 尺度选择的特征波长主要分布在可见光波段中;第 10 尺度选择的特征波长主要分布在近红外波段中;OS、FD 和 RL 选择的特征波长在可见光和近红外波段中均有分布。

表 3 土壤氧化铁反演模型结果

Table 3 Results of soil iron oxide inversion model

Transform spectrum	Calibration set		Validation set		
	R^2	$E_{\text{RMSE}} / (\text{g} \cdot \text{kg}^{-1})$	R^2	$E_{\text{RMSE}} / (\text{g} \cdot \text{kg}^{-1})$	R_{PIQ}
L_1	0.783	4.969	0.168	12.251	1.654
L_2	0.691	5.932	0.370	10.657	1.901
L_3	0.825	4.464	0.590	8.600	2.356
L_4	0.760	5.236	0.663	7.798	2.598
L_5	0.622	6.565	0.637	8.095	2.503
L_6	0.548	7.183	0.450	9.967	2.034
L_7	0.769	5.133	0.618	8.298	2.442
L_8	0.511	7.463	0.498	9.517	2.129
L_9	0.321	8.801	0.102	12.726	1.592
L_{10}	0.340	9.593	0.342	11.548	1.755
OS	0.678	6.062	0.616	8.328	2.433
FD	0.647	6.343	0.638	8.076	2.509
RL	0.744	5.402	0.637	8.092	2.504

由表 3 可知:OS 的验证集的 R^2 和 E_{RMSE} 分别为 0.616 和 $8.328 \text{ g} \cdot \text{kg}^{-1}$, 其 R_{PIQ} 值为 2.433;FD 的验证集的 R^2 和 E_{RMSE} 分别为 0.638 和 $8.076 \text{ g} \cdot \text{kg}^{-1}$, 其 R_{PIQ} 值为 2.509;RL 的验证集的 R^2 和 E_{RMSE} 分别为 0.637 和 $8.092 \text{ g} \cdot \text{kg}^{-1}$, 其 R_{PIQ} 值为 2.504。三种模型的预测能力均一般,可以粗略估计土壤中氧化铁的含量。

在 CWT 中,各尺度的验证集的 R^2 从 L_1 到 L_4 逐渐增大,随后逐渐减小。其中, L_4 的验证集的 R^2 最高,达到了 0.663,相应的 E_{RMSE} 和 R_{PIQ} 分别为 $7.798 \text{ g} \cdot \text{kg}^{-1}$ 和 2.598,均比 OS、FD 和 RL 的精度要高。通过对比, L_4 的验证集的 R^2 比 OS、FD 和 RL 提高了 0.047, 0.025 和 0.026。

同时,对 L_4 -CC-CARS-SVR、OS-CC-CARS-SVR、FD-CC-CARS-SVR、RL-CC-CARS-SVR 4 个模型进行 1:1 线分析,如图 6 所示。

可以发现,这 4 种模型中 L_4 -CC-CARS-SVR 的验证效果较好,其验证样本的实测值和预测值样点分布都更接近 1:1 线,能更好地实现土壤氧化铁含量的估算,说明 CWT 的最佳尺度为第 4 尺度。因此, L_4 -CC-CARS-SVR 可以作为该地区土壤氧化铁含量的模型。

4 结 论

以禄丰恐龙谷南缘地区的土壤为研究对象,获取该地区土壤的高光谱数据和氧化铁含量值,发现红壤的平均氧化铁含量最高,紫壤其次,而黄棕壤最低。这

3.6 土壤氧化铁高光谱模型构建及验证

支持向量机回归将获得的特征波长和敏感小波系数作为输入自变量,并将氧化铁含量作为因变量进行回归建模。建模结果如表 3 所示。

与前人的研究基本一致,对红壤土色起主导作用的是赤铁矿和针铁矿。对紫壤土色起主导作用的是结晶的氧化铁和锰化合物。对黄棕壤土色起主导作用的是针铁矿和水铁矿。

对 OS 进行 RL 变换可以提高最大相关性,相关系数的绝对值提高了 0.017,但进行 FD 变换却会降低最大相关性,相关系数的绝对值降低了 0.041。对 OS 进行 CWT 分解后发现: $L_3 \sim L_6$ 的最大相关性比 OS 和 FD 要强; L_3 的最大相关系数与 RL 相当。

利用 CC-CARS 算法选出特征波长作为支持向量机回归建模的自变量,并将土壤中氧化铁的含量作为建模的因变量。利用 CC-CARS 算法得到的变量相比全波段大大减少,并充分保留了对模型性能有效提升的变量。对 OS、FD 和 RL 构建的反演模型估测结果显示,RL 模型精度最好(R^2 达到了 0.637),其次为 FD,最差的为 OS。该现象说明:经过常规的光谱变换也可以在一定程度上提高模型的精度;经过 CWT 分解后,以不同尺度上的敏感小波系数作为自变量建立的模型,在 L_4 、 L_5 和 L_7 尺度建立的模型精度较高。其中, L_4 尺度建立的模型精度最高,其建模集和验证集的 R^2 分别为 0.760 和 0.663,相应的 E_{RMSE} 分别为 $5.236 \text{ g} \cdot \text{kg}^{-1}$ 和 $7.798 \text{ g} \cdot \text{kg}^{-1}$,验证集的 R_{PIQ} 为 2.598,优于常规光谱建立的模型。该现象说明,在研究区中,不同的土壤类型和不同的土地利用类型下 CWT 与常规光谱相比,更有利于挖掘土壤的有效信息。

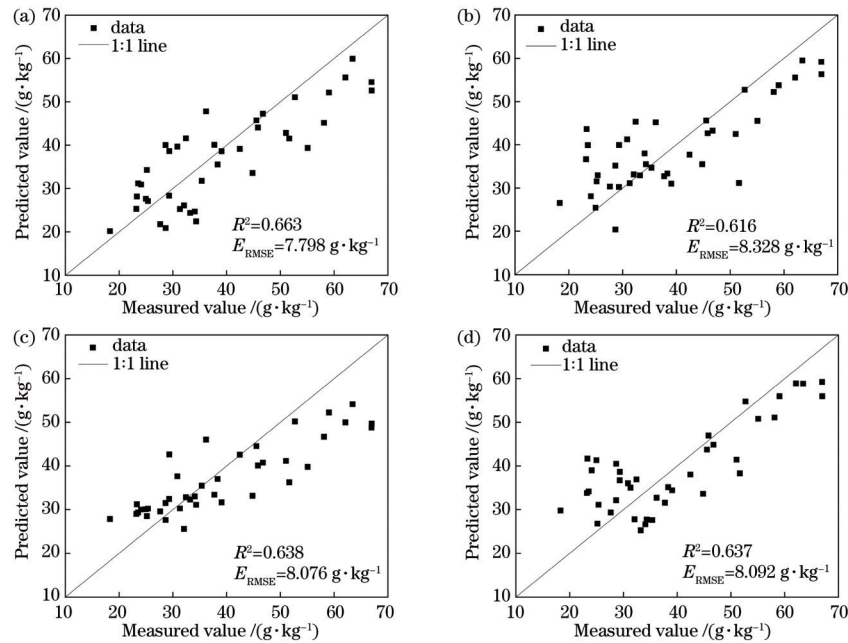


图 6 各模型下土壤氧化铁含量实测值与预测值散点图。(a) L_4 -CC-CARS-SVR; (b) OS-CC-CARS-SVR; (c) FD-CC-CARS-SVR; (d) RL-CC-CARS-SVR

Fig. 6 Scatter plot of measured and predicted values of soil iron oxide content under different models. (a) L_4 -CC-CARS-SVR; (b) OS-CC-CARS-SVR; (c) FD-CC-CARS-SVR; (d) RL-CC-CARS-SVR

参 考 文 献

- 傅佳丽. 平潭海岛土壤元素的高光谱遥感反演模型[D]. 福州: 福建师范大学, 2016.
Fu J L. Modeling island soil elements with visible near infrared reflectance spectroscopy in Pingtan[D]. Fuzhou: Fujian Normal University, 2016.
- Hu L, Gan S, Yuan X P, et al. Spatial differentiation analysis of water quality in Dianchi Lake based on GF-5 NDVI characteristic optimization[J]. Journal of Spectroscopy, 2021, 2021: 5542126.
- Wei L F, Pu H C, Wang Z X, et al. Estimation of soil arsenic content with hyperspectral remote sensing[J]. Sensors, 2020, 20(14): 4056.
- 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 Q X, Zhang H B, Liu W K, 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, 2017, 33(12): 230-239.
- 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.
- 史舟. 土壤地面高光谱遥感原理与方法[M]. 北京: 科学出版社, 2014.
Shi Z. Soil ground hyperspectral remote sensing principles and methods[M]. Beijing: Science Press, 2014.
- 何挺, 王静, 程焯, 等. 土壤氧化铁光谱特征研究[J]. 地理与地理信息科学, 2006, 22(2): 30-34.
He T, Wang J, Cheng Y, et al. Study on spectral features of soil Fe_2O_3 [J]. Geography and Geo-Information Science, 2006, 22(2): 30-34.
- 谭洁, 陈严, 周卫军, 等. 基于实验室高光谱数据的大围山森林土壤氧化铁含量反演[J]. 土壤, 2021, 53(4): 858-864.
Tan J, Chen Y, Zhou W J, et al. Inversion of iron oxide contents in forest soils of Dawei Mountains using laboratory hyperspectral data[J]. Soils, 2021, 53(4): 858-864.
- 谭洋, 姜琦刚, 刘骅欣, 等. 基于多尺度 SNV-CWT 特征的黑土有机质、水分、总铁及 pH 值估测[J]. 光谱学与光谱分析, 2021, 41(11): 3424-3430.
Tan Y, Jiang Q G, Liu H X, et al. Estimation of organic matter, moisture, total iron and pH from back soil based on multi scales SNV-CWT transformation[J]. Spectroscopy and Spectral Analysis, 2021, 41(11): 3424-3430.
- 郭颖, 郭治兴, 刘佳, 等. 亚热带典型区域水稻土氧化铁高光谱反演: 以珠江三角洲为例[J]. 应用生态学报, 2017, 28(11): 3675-3683.
Guo Y, Guo Z X, Liu J, et al. Hyperspectral inversion of paddy soil iron oxide in typical subtropical area with Pearl River Delta, China as illustration[J]. Chinese Journal of Applied Ecology, 2017, 28(11): 3675-3683.
- 于雷, 洪永胜, 周勇, 等. 连续小波变换高光谱数据的土壤有机质含量反演模型构建[J]. 光谱学与光谱分析, 2016, 36(5): 1428-1433.
Yu L, Hong Y S, Zhou Y, et al. Inversion of soil organic matter content using hyperspectral data based on continuous wavelet transformation[J]. Spectroscopy and Spectral Analysis, 2016, 36(5): 1428-1433.
- 丁海宁, 陈玉, 陈芸芝. 黄土高原土壤铁元素含量遥感

- 反演方法[J]. 遥感技术与应用, 2019, 34(2): 275-283.
- Ding H N, Chen Y, Chen Y Z. Remote sensing inversion method of soil iron content in the Loess Plateau [J]. Remote Sensing Technology and Application, 2019, 34(2): 275-283.
- [13] 李双权, 马玉凤, 刘勋, 等. 郑州邙山枣树沟黄土剖面常量元素含量的高光谱反演[J]. 自然资源遥感, 2021, 33(3): 121-129.
- Li S Q, Ma Y F, Liu X, et al. Hyperspectral inversion of macro element content in loess based on the profile of Zaoshugou Village, Mangshan Mountain, Zhengzhou City[J]. Remote Sensing for Natural Resources, 2021, 33(3): 121-129.
- [14] 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 Y, Huang W H, Lu Y, 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.
- [15] 李冠稳, 高小红, 肖能文, 等. 特征变量选择和回归方法相结合的土壤有机质含量估算[J]. 光学学报, 2019, 39(9): 0930002.
- Li G W, Gao X H, Xiao N W, et al. Estimation of soil organic matter content based on characteristic variable selection and regression methods[J]. Acta Optica Sinica, 2019, 39(9): 0930002.
- [16] 苑志娟. 云南禄丰恐龙国家地质公园地质遗迹资源特征及保护[D]. 北京: 中国地质大学(北京), 2015.
- Yuan Z J. Study on the characteristics of the geoheritages and protection in Lufeng dinosaur national geopark, Yunnan[D]. Beijing: China University of Geosciences, 2015.
- [17] 张焱, 丁建丽, 张子鹏, 等. 光谱配置对最优波段组合算法预测土壤有机质和电导率的影响[J]. 激光与光电子学进展, 2021, 58(21): 2128001.
- Zhang Y, Ding J L, Zhang Z P, et al. Effect of spectral configuration on soil organic matter and electrical conductivity predicted by optimal band combination algorithm[J]. Laser & Optoelectronics Progress, 2021, 58(21): 2128001.
- [18] 赵安新, 汤晓君, 张钟华, 等. 优化 Savitzky-Golay 滤波器的参数及其在傅里叶变换红外气体光谱数据平滑预处理中的应用[J]. 光谱学与光谱分析, 2016, 36(5): 1340-1344.
- Zhao A X, Tang X J, Zhang Z H, et al. Optimizing Savitzky-Golay parameters and its smoothing pretreatment for FTIR gas spectra[J]. Spectroscopy and Spectral Analysis, 2016, 36(5): 1340-1344.
- [19] Zhang S W, Shen Q, Nie C J, et al. Hyperspectral inversion of heavy metal content in reclaimed soil from a mining wasteland based on different spectral transformation and modeling methods[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019, 211: 393-400.
- [20] 彭咏石, 陈水森, 陈金月, 等. 基于连续小波系数的叶绿素 a 浓度估测模型[J]. 激光与光电子学进展, 2021, 58(8): 0828002.
- Peng Y S, Chen S S, Chen J Y, et al. Estimation model of chlorophyll-a concentration based on continuous wavelet coefficient[J]. Laser & Optoelectronics Progress, 2021, 58(8): 0828002.
- [21] 李江波, 郭志明, 黄文倩, 等. 应用 CARS 和 SPA 算法对草莓 SSC 含量 NIR 光谱预测模型中变量及样本筛选[J]. 光谱学与光谱分析, 2015, 35(2): 372-378.
- Li J B, Guo Z M, Huang W Q, et al. Near-infrared spectra combining with CARS and SPA algorithms to screen the variables and samples for quantitatively determining the soluble solids content in strawberry[J]. Spectroscopy and Spectral Analysis, 2015, 35(2): 372-378.
- [22] Bellon-Maurel V, Fernandez-Ahumada E, Palagos B, et al. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy[J]. TrAC Trends in Analytical Chemistry, 2010, 29(9): 1073-1081.
- [23] 唐海涛, 孟祥添, 苏循新, 等. 基于 CARS 算法的不同类型土壤有机质高光谱预测[J]. 农业工程学报, 2021, 37(2): 105-113.
- Tang H T, Meng X T, Su X X, et al. Hyperspectral prediction on soil organic matter of different types using CARS algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(2): 105-113.