

基于 GWR 模型的土壤重金属含量高光谱预测

杨奇勇¹ 沈利娜² 李文军¹ 段小芳¹

(1. 邵阳学院农林生态学院, 邵阳 422000; 2. 中国地质科学院岩溶地质研究所, 桂林 541004)

摘要: 针对地理加权回归分析(GWR)自变量选取问题,本文通过相关分析确定表征土壤重金属 Cu 含量与光谱特征关系的最佳光谱变换方式,利用普通最小二乘法建立候选光谱变量子集,运用回归分析法构建 GWR 模型,并将预测结果与传统方法构建的 GWR 模型进行对比。结果表明:对数微分变换光谱与土壤重金属 Cu 含量相关系数为 $-0.626 \sim 0.618$,最能表征研究区域土壤光谱特征与土壤重金属 Cu 含量的关系,为研究区域最佳光谱变换方式。研究区土壤重金属 Cu 含量 GWR 预测最优模型的自变量组合为 2 308、1 400、2 197、2 138 nm 4 个波长,其构成 GWR 模型的赤池信息准则(AICc)最小,决定系数 R^2 和调整决定系数 R^2_{adj} 最大。与传统方法构建的 GWR 模型相比,本文方法构建的模型更简洁、稳健性更强、预测精度更高。

关键词: 土壤重金属; 地理加权回归; 变量选取; 回归分析; 土壤光谱

中图分类号: TP79; S151.3 文献标识码: A 文章编号: 1000-1298(2026)08-0367-08

OSID:



Hyperspectral Prediction of Soil Heavy Metal Content Based on GWR Model

YANG Qiyong¹ SHEN Li'na² LI Wenjun¹ DUAN Xiaofang¹

(1. Agriculture Forestry Ecology College, Shaoyang University, Shaoyang 422000, China

2. Institute of Karst Geology, Chinese Academy of Geological Sciences, Guilin 541004, China)

Abstract: Aiming at the selection of independent variables in geographically weighted regression (GWR) analysis, the construction of GWR model was explored by using soil heavy metal Cu content as the dependent variable and soil hyperspectral data as the independent variables. The best spectral transformation that characterized the relationship between the soil heavy metal Cu content and the spectral reflectance in the study area was determined, throughing the correlation analysis of the spectral reflectance of the original soil and its eight transformation methods. The candidate variable subsets was established by using the ordinary least squares method, and the GWR model was constructed by using the regression analysis method. The advantages of the GWR model, as well as its prediction accuracy, were compared with that of the GWR model constructed by traditional methods. The results showed that the correlation coefficient between the logarithmic differential transformation spectrum and the soil heavy metal Cu content was from -0.626 to 0.618 , which could best represent the relationship characteristics between the soil spectrum and the soil heavy metal Cu content in the study area, and it was the best spectral transformation method for the study area. The results of the regression analysis indicated that the optimal GWR prediction model for the soil heavy metal Cu content in the study area had four independent variables combined of 2 308 nm, 1 400 nm, 2 197 nm, and 2 138 nm wavelengths, which had the smallest Akaike information criterion (AICc), the largest determination coefficient (R^2) and adjust coefficient of determination (R^2_{adj}). Compared with the GWR model constructed by the traditional method, the GWR model constructed by the regression analysis method was more concise with stronger model robustness and higher prediction accuracy.

Key words: soil heavy metal; geographically weighted regression; variable selection; regression analysis; soil hyperspectral

收稿日期: 2025-10-28 修回日期: 2025-11-25

基金项目: 国家自然科学基金联合基金项目(U21A2041)和广西重点研发计划项目(桂科 AB25069160)

作者简介: 杨奇勇(1976—),男,研究员,博士,主要从事遥感监测与土壤数字制图研究,E-mail: yangqiyong0739@163.com

0 引言

铜(Cu)是植物必需微量元素,但土壤中过量的重金属 Cu 含量会抑制土壤微生物活性、降低养分矿化、破坏根系细胞膜结构、抑制植物光合作用酶活性,从而影响粮食^[1]、水果^[2]、蔬菜^[3]等的产量和质量,并通过食物链转移,对人体健康构成严重威胁^[4]。随着工矿业快速发展和农药、化肥长期不合理施用,越来越多的重金属污染物进入土壤环境,使得重金属污染已成为当今土壤污染中面积最广、危害最大的环境问题之一^[5]。因此,及时获取并掌握土壤重金属含量状况,对土壤污染的风险监控与治理及土壤修复都具有重要意义。传统的基于大量野外实地采样、实验室化学分析获取原位土壤重金属含量的方法,不但费时、费力,而且无法满足土壤重金属含量空间信息快速获取要求^[5]。高光谱遥感技术具有测定简单、省时、无损等优点,能够根据地物光谱特征对地物进行反演和识别,为土壤重金属含量快速预测提供了可能^[6-7]。

在土壤重金属含量光谱预测方面,前人已取得了大量研究进展。所采用方法主要有普通最小二乘法(Ordinary least square, OLS)^[8]、偏最小二乘回归(Partial least squares regression, PLSR)^[9-11]、极限学习机(Extreme learning machine, ELM)^[12-13]、人工神经网络(Artificial neural network, ANN)^[14-16]等,其中,OLS 是应用最多的参数估计方法,也是土壤重金属含量高光谱预测应用最常用模型^[8,17]。但这些方法都是假定研究区域内每个样点的环境因素对光谱反射率影响相同,即模型系数在研究区内是平稳、不变的。然而,土壤是一个时空连续的变体,受成土母质、气候、生物、地形、时间以及人类活动的影响,具有高度空间异质性。土壤光谱特征是土壤有关性质的综合反映,不同研究区域、同一研究区域的不同位置,土壤重金属含量与土壤光谱特征的关系都不尽相同^[6],即土壤重金属含量与土壤光谱特征关系具有空间异质性。而已有研究采用 OLS、PLSR、ELM、ANN 等方法预测土壤重金属含量时都忽视了这种空间异质性,因此其预测精度也受到影响。

地理加权回归模型(Geographically weighted regression, GWR)能有效揭示变量间关系的空间异质性^[18],被广泛应用于空间非平稳性关系探讨,并已在土壤重金属含量预测中取得较好效果^[19-21]。但目前还没有关于在 GWR 模型构建时,如何进行自变量选取的系统性研究^[21]。如学者直接用 OLS 模型^[22]或 PLS 模型^[23]中的自变量进行 GWR 建模,

或在逐步回归分析(Stepwise regression, SR)基础上,利用其自变量构建 GWR 模型^[24]。但这些模型系数在整个研究区域都是恒定不变的,而 GWR 模型系数在研究区域内随着位置的变化而变化。因此,需探讨直接利用这些模型的自变量构建 GWR 模型的合理性及该模型是否为最优模型。本文选择广西壮族自治区桂林市平乐县沙子镇为研究区域,以土壤重金属 Cu 含量为研究对象,通过确定最优光谱变换方式、遴选敏感波长构建候选变量子集;利用回归法分析确定土壤重金属 Cu 含量预测的最佳自变量组合和最优 GWR 模型,并将模型预测结果与传统方法进行比较,以期 GWR 模型推广应用提供理论和技术支持。

1 材料与方法

1.1 研究区域概况

沙子镇(24°42′29.51″~24°54′0.51″N, 110°37′41.82″~110°49′19.82″E)(图1)地处亚热带季风气候区,年均气温 20.3℃,年均降水量 1 355~1 865 mm。该镇总面积为 193 km²,海拔在 27~795 m 之间,属于典型的岩溶地貌区,成土母质以石灰岩为主。沙子镇主要农作物有平乐慈姑、葛根、淮山、水稻、玉米、油菜等。沙子镇也是广西的水果大镇,素有“水果之乡”的美誉^[25]。

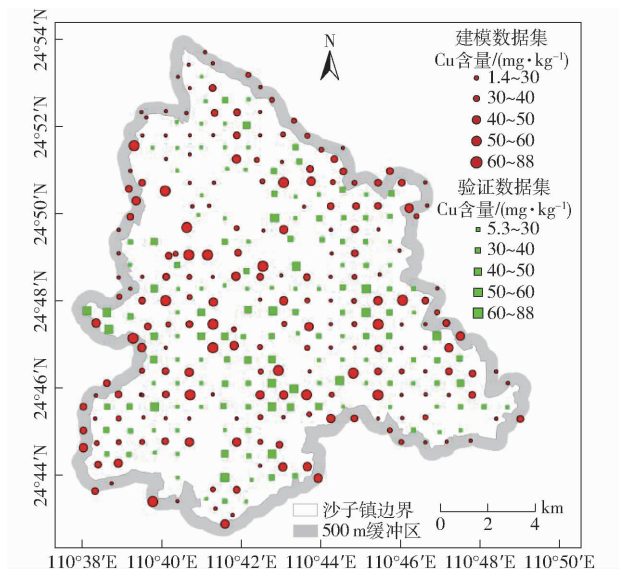


图1 研究区位置和采样点分布图

Fig. 1 Location of study area and sampling points

1.2 土壤样品采集与处理

1.2.1 土壤样品采集

以沙子镇边界建立 500 m 缓冲区为采样区,采用 2 个 1 km × 1 km 网格进行嵌套布点。采样时间为 2020 年 7—8 月,在 1 m² 范围采集 3 个 0~20 cm 土壤样品,同时用 GPS 记录采样点位置的经纬度、

海拔以及周围环境特征^[25]。共获得土壤重金属 Cu 含量有效样点 368 个(图 1)。

1.2.2 土壤样品实验与样品分组

土壤样品自然风干后分成 2 部分。一部分磨碎过 100 目筛,采用硝酸-双氧水电热板消解处理,并用石墨炉-原子吸收光谱法测定 Cu 含量。另一部分过 2 mm 孔筛,采用 ASD FieldSpec 4 型地物波谱仪采集土壤光谱数据。每个样品去除噪声较大边缘波段,只保留波段 400~2 400 nm 用于后续分析^[7]。

将沙子镇范围内的一个网格样点作为验证数据,其余所有样点作为建模数据(图 1、表 1)。

表 1 土壤重金属 Cu 含量统计特征

Tab.1 Statistical characteristics of heavy metal Cu content in soil

参数	全部数据集	建模数据集	验证数据集
样品数量/个	368	214	154
平均值/(mg·kg ⁻¹)	37.18	37.17	37.21
最小值/(mg·kg ⁻¹)	1.41	1.41	5.39
最大值/(mg·kg ⁻¹)	88.46	88.46	88.46
标准差/(mg·kg ⁻¹)	17.09	17.60	16.41
变异系数/%	45.97	47.35	44.10
偏度	0.56	0.41	0.83
峰度	3.21	3.03	3.48

1.2.3 土壤光谱变换与敏感波长确定

分别对土壤原始光谱反射率(R)进行一阶微分(R')、二阶微分(R'')、倒数($1/R$)、倒数微分($(1/R)'$)、对数($\lg R$)、对数微分($(\lg R)'$)、平方根($\sqrt{R}$)、平方根微分($(\sqrt{R})'$)等变换。利用土壤重金属 Cu 含量与这 9 种光谱反射率进行相关分析,确定表征土壤重金属 Cu 含量与土壤光谱反射率关系的最佳变换方式。绘制最佳变换光谱反射率曲线,利用局部极值法^[21]结合 t 检验确定土壤重金属 Cu 含量分析的敏感波长。

1.3 基于回归分析的 GWR 模型

1.3.1 候选变量子集确定

为从大量光谱变量中筛选出对土壤重金属 Cu 含量有显著解释力且避免多重共线性的候选变量子集,采用后向消除法(Backward elimination)和方差膨胀因子(Variance inflation factor, VIF)检验法实现,步骤为:①以土壤重金属 Cu 含量为因变量,所有敏感波长为自变量进行 OLS 分析,分别计算每个自变量的 p 值,移除 p 值最大的最不显著敏感波长。②再利用剩余的敏感波长作为自变量进行 OLS 分析,移除 p 值最大的最不显著敏感波长。③重复步骤②,直到所有剩余敏感波长都达到给定的显著性要求。④采用同样方法,移除 VIF 大

于等于 7.5 的敏感波长,直到所有敏感波长的 VIF 小于 7.5,将剩余所有敏感波长作为 GWR 分析的候选变量子集。

1.3.2 GWR 模型构建

GWR 法基本原理与算法详见文献[18]。本研究土壤重金属含量 GWR 建模过程自变量的确定采用前向选择(Forward selection)法。主要选取赤池信息准则(AICc)判断模型性能:AICc 越低,模型性能越好;当比较 2 个或多个 GWR 模型时,AICc 差 $\Delta AICc$ 大于等于 3 时,具有较低 AICc 的模型被认为更优。此外,模型决定系数(R^2)越大,对因变量空间变异的解释力越高;调整决定系数(R^2_{adj})越大,模型越稳健。在 GWR4 软件中建模的主要步骤为:①对候选自变量子集中所有变量分别进行单变量地理加权回归(Univariate GWR,UGWR)分析,通过比较确定最小 AICc 的模型为最优 UGWR 模型,命名为 GWR_1 。②在 GWR_1 模型基础上,依次添加自变量子集中的其它变量,进行双变量地理加权回归(Bivariate GWR,BGWR)分析,如果存在某个新模型的 AICc 最小,且与 GWR_1 模型的 $\Delta AICc$ 大于等于 3,则确定该模型为最优的 BGWR 分析模型,命名为 GWR_2 。③重复步骤②,直到 $\Delta AICc$ 小于 3 时迭代结束,得到 n 个最优自变量组合的最优 GWR 分析模型(GWR_n)。

1.4 模型预测精度验证

为更直观评判 GWR 模型性能,将预测结果与文献[22,24]中运用的 OLS 和 SR 法及其对应 GWR 法结果进行比较。选取预测数据与验证数据的 R^2 、均方根误差(RMSE)、平均绝对误差(MAE)^[25] 3 个指标,对模型预测精度进行评价。

2 结果与分析

2.1 土壤重金属 Cu 含量预测最佳光谱变换方式

研究区域土壤重金属 Cu 含量与土壤原始及其变换方式光谱反射率呈现显著相关性(图 2)。

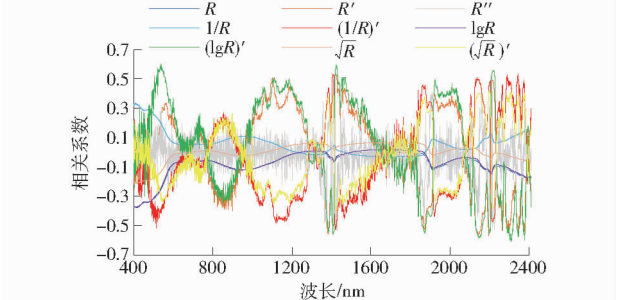


图 2 土壤重金属 Cu 含量与土壤原始及其变换方式光谱反射率相关性

Fig.2 Correlation coefficient between soil heavy metal Cu content and raw spectral reflectance and its transformation

从图 2 可看出,相关系数绝对值($|r|$)最大值处在 0.375 ~ 0.626 ($p < 0.01$) 之间。其中,只有 R' 、 R'' 、 $(1/R)'$ 和 $(\lg R)'$ 4 种变换方式光谱反射率中有 $|r| > 0.5$ 波长。在所有波长与土壤重金属 Cu 含量相关系数中,最大值($r = 0.618, p < 0.01$)和最小值($r = -0.626, p < 0.01$)均出现在 $(\lg R)'$ 变换光谱反射率中,且 $|r| > 0.5$ 波长数量最多(436 个),远高于 R' (198 个)、 R'' (28 个)和 $(1/R)'$ (52 个)等光谱反射率。因此, $(\lg R)'$ 变换最能表征土壤光谱特征与土壤重金属 Cu 含量的关系,为研究区域最佳光谱变换方式。

2.2 GWR 预测候选变量子集

建模数据为 214 个,在 99% 置信水平下,当土壤重金属 Cu 含量与光谱反射率 $|r| > 0.167$ 时, $p < 0.01$ 。利用局部极值法^[21],从 $(\lg R)'$ 光谱反射率中共选出 133 个敏感波长,其中 $|r| > 0.5$ 波长 31 个。以土壤重金属 Cu 含量为自变量,采用后向消除法进行 OLS 分析与 VIF 检验,得到 GWR 分析的候选

变量子集:578 nm ($r = 0.364$)、593 nm ($r = 0.346$)、876 nm ($r = -0.359$)、1 051 nm ($r = 0.460$)、1 086 nm ($r = 0.330$)、1 400 nm ($r = -0.379$)、2 138 nm ($r = -0.589$)、2 190 nm ($r = -0.365$)、2 197 nm ($r = 0.587$)、2 308 nm ($r = -0.617$)、2 375 nm ($r = -0.610$) nm 共 11 个敏感波长。

2.3 土壤重金属 Cu 含量预测 GWR 模型

由于本研究采用网格布点法,样点分布比较均匀。因此 GWR 建模时采用高斯函数核类型和黄金搜索法确定最优核带宽。通过 GWR 回归分析,发现当自变量数量增加到 5 时,GWR 模型中所有 ΔAIC_c 小于 3,迭代结束。这说明研究区域土壤重金属 Cu 含量 GWR 分析的最佳自变量组合数量为 4。因此,一共进行 38 次有效 GWR 建模,包括 11 次 UGWR 建模、10 次 BGWR 建模、9 次 3 变量地理加权回归(Trivariate GWR, TGWR)建模、8 次 4 变量地理加权回归(Four-variable GWR, FGWR)建模。最后一次添加到上述模型中变量 $|r|$ 与各模型参数如图 3 所示。

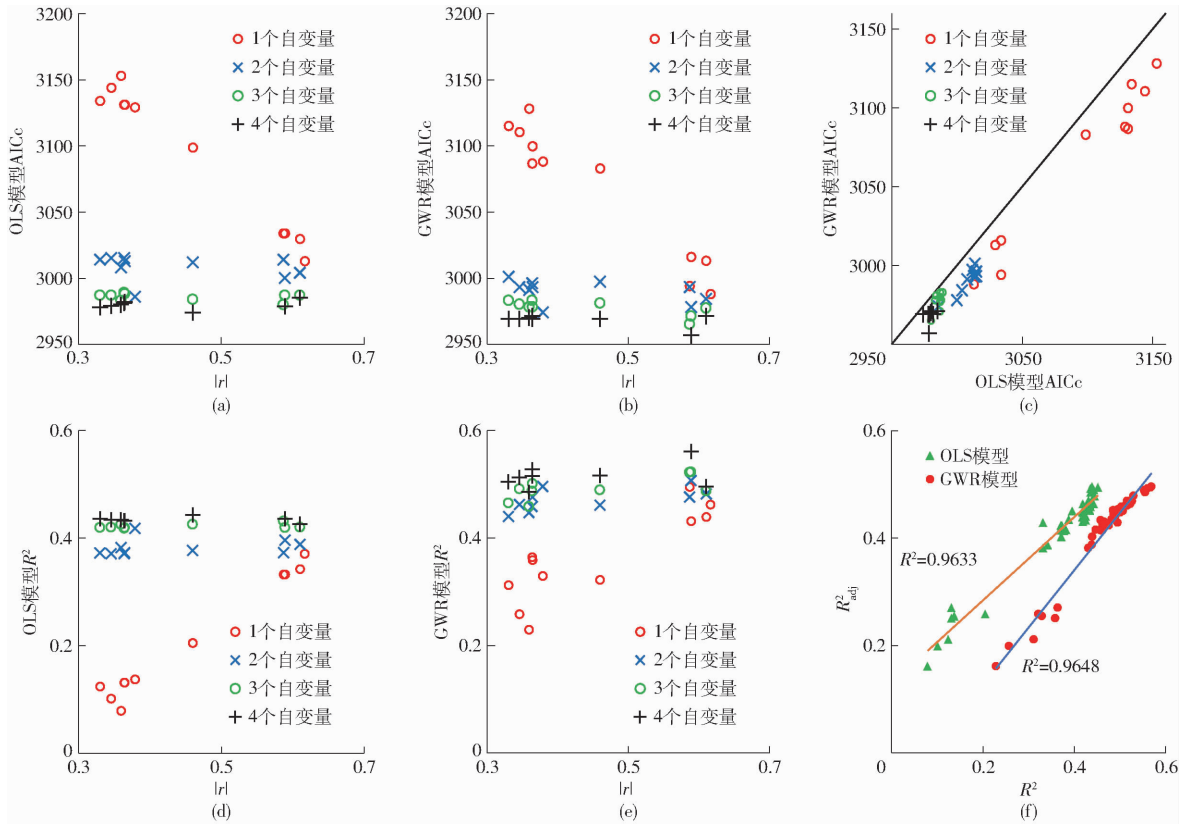


图 3 不同自变量下土壤重金属 Cu 含量预测的 OLS 和 GWR 模型参数

从图 3 可看出,在 11 次 UGWR 分析中,2 308 nm 与土壤重金属 Cu 含量相关性最强,作为自变量时构建的模型 AICc 最小,为 2 988(图 3b)、 R^2 和 R^2_{adj} 均最大(图 3e),分别为 0.462 和 0.423,说明 2 308 nm 作为自变量构建的模型为最优 UGWR 模型,该模型为 GWR_1 。在 10 次 BGWR 建模中有 3 个模型 AICc 比 GWR_1 小(图 3b),且 R^2 和 R^2_{adj} 均比 GWR_1 大(图 3e),其中 2 308、1 400 nm 为自变量时构建的模型 AICc 最小,且与 GWR_1 的 ΔAIC_c 为 14,因此该模型为 BGWR 模型中的最优模型(GWR_2);在 9 次 TGWR 建模中,有 2 个模型 AICc 比 GWR_2 小(图 3b),且 R^2 和 R^2_{adj} 比 GWR_2 大(图 3e),其中

2 308、1 400、2 197 nm 为自变量时构建 TGWR 模型最优($\Delta AICc$ 为 9),该模型为 GWR_3 ;在 8 次 FGWR 建模中,只有 2 308、1 400、2 197、2 138 nm 为自变量时构建的模型 $AICc$ 比 GWR_3 小($\Delta AICc$ 为 8),该模型为最优 FGWR 模型,即 GWR_4 。

综上分析,研究区土壤重金属 Cu 含量预测最佳自变量组合为 2 308、1 400、2 197、2 138 nm, GWR_4 为最优模型,该模型截距 (Intercept)、系数(β)和 R^2 空间分布见图 4(分辨率为 500 m)。

从图 4 可看出,截距、系数和模型 R^2 在空间上

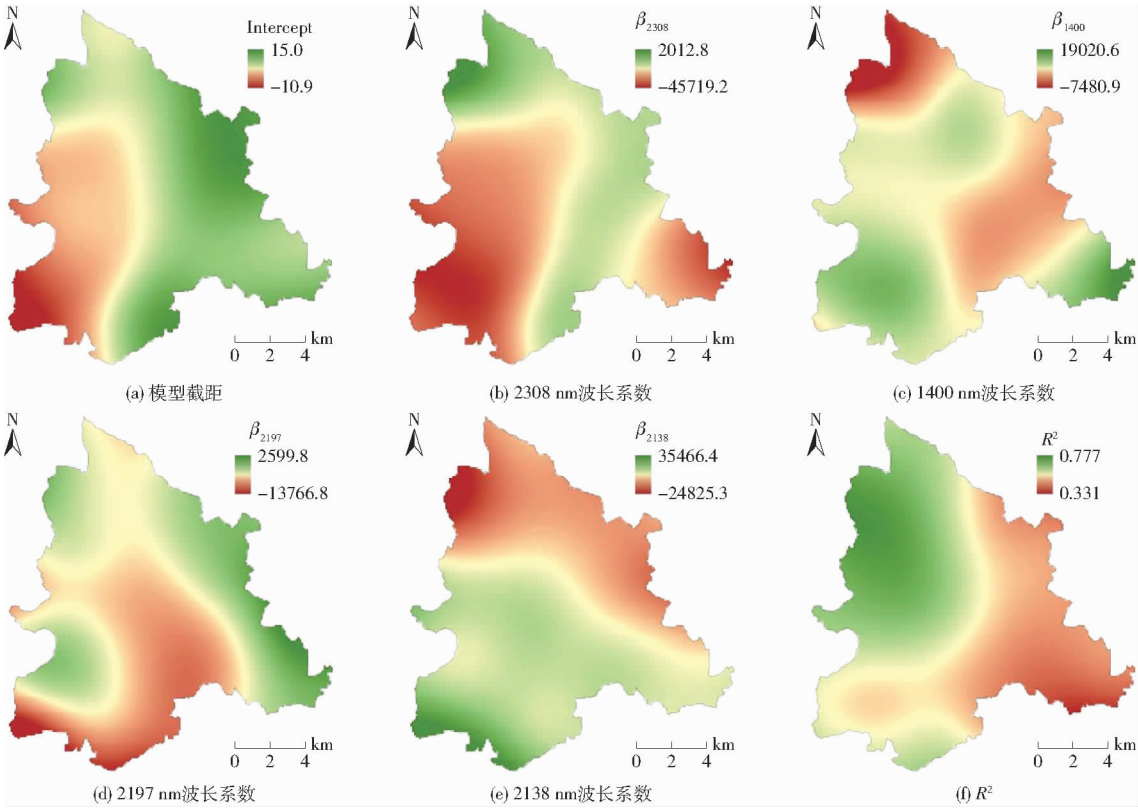


图 4 土壤重金属 Cu 含量预测的 GWR 模型参数

Fig. 4 Parameters of GWR model for predicting content of heavy metal Cu in soil

都具有较大变异性。研究区域土壤重金属 Cu 含量 (y_{Cu}) 预测的 GWR 模型为: $y_{Cu} = I_{Intercept} + \lambda_{2308} \beta_{2308} + \lambda_{1400} \beta_{1400} + \lambda_{2197} \beta_{2197} + \lambda_{2138} \beta_{2138}$ ($R^2 = 0.561, p < 0.01$),模型中 $I_{Intercept}$ 为截距, λ_{2308} 、 λ_{1400} 、 λ_{2197} 、 λ_{2138} 分别为波长 2 308、1 400、2 197、2 138 nm。模型能够解释研究区域土壤重金属 Cu 含量空间变异的 56.1%。

2.4 土壤重金属 Cu 含量预测对比

以土壤重金属 Cu 含量为因变量,候选变量子集为自变量进行 SR 分析,结果表明 11 个自变量子集中有 2 308、1 400、876、1 051、593、2 190、587 nm 等 7 个波长入选 SR 模型。因此将 SR 模型记作 SR_7 ,其对应的 GWR 模型记为 GWR_7 ;自变量子集对应的 OLS 模型记为 OLS_{11} ,相应的 GWR 模型记为 GWR_{11} 。通过其模型参数与不同自变量的最优 GWR 模型参数对比(表 2)可看出, $AICc$ 随自变量增加而减小,且所有 $\Delta AICc$ 大于 3。但从相同自变量组合的不同模型表现来看, GWR_4 模型 $AICc$ 、 R^2 和 R^2_{adj} 改善幅度比 GWR_7 和 GWR_{11} 模型更明显;从不同

表 2 不同自变量下 GWR 模型参数

Tab. 2 Parameters of GWR model under different auxiliary variables

模型	OLS/SR 模型			GWR 模型		
	AICc	R ²	R ² _{adj}	AICc	R ²	R ² _{adj}
GWR ₄	2 979	0.436	0.428	2 957	0.561	0.492
GWR ₇	2 955	0.480	0.468	2 953	0.494	0.473
GWR ₁₁	2 934	0.519	0.503	2 924	0.567	0.525

自变量数量组合的 GWR 模型来看, GWR_4 模型 R^2 和 R^2_{adj} 比 GWR_7 模型大,而与 GWR_{11} 模型 R^2 和 R^2_{adj} 差别不大。

将 GWR_4 、 GWR_7 、 GWR_{11} 以及对应回归模型的土壤重金属 Cu 含量预测值与实测值进行对比(图 5),发现在这 3 种模型中 GWR_4 预测效果最好,其预测值与实测值之间 R^2 最大,为 0.615、MAE 和 RMSE 最小,分别为 7.63、11.40 mg/kg。从图 5d ~ 5f 可看出, GWR_4 、 GWR_7 和 GWR_{11} 预测值表现有很大不同:从预测最小值来看, GWR_4 为 7.07 mg/kg,而 GWR_7 和 GWR_{11} 分别为 15.05、

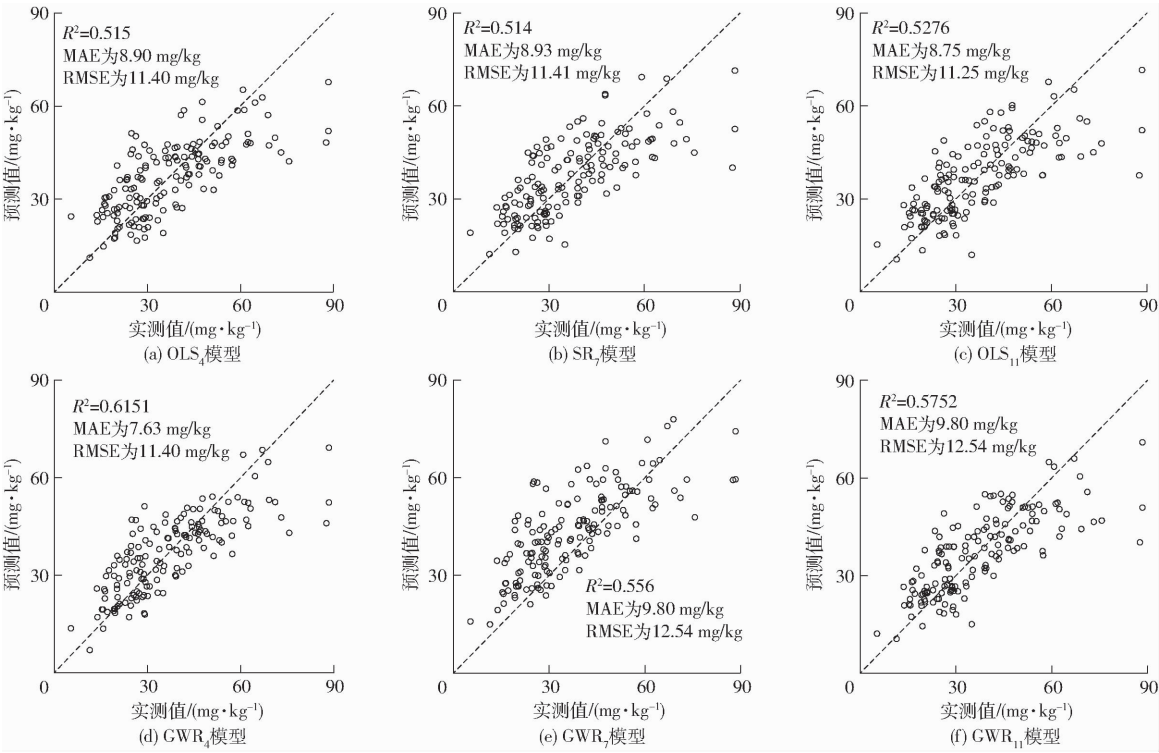


图5 不同模型 Cu 含量预测精度对比

Fig.5 Comparison of prediction accuracy for Cu content by different methods

10.67 mg/kg,很显然 GWR₄预测最小值与验证数据集最小值 5.39 mg/kg 最为接近(表 1);从预测最大值来看,GWR₄、GWR₇和 GWR₁₁分别为 69.29、77.95、71.00 mg/kg,GWR₇预测最大值与验证数据集中最大值 88.46 mg/kg 最为接近(表 1);从预测平均值来看,GWR₄、GWR₇和 GWR₁₁分别为 36.86、43.36、36.75 mg/kg,GWR₇预测平均值与验证数据集平均值 37.21 mg/kg 相差最大(表 1)。这说明随着自变量增加 GWR₇和 GWR₁₁模型虽然降低了 AICc,但模型可能存在过拟合现象^[26],预测精度并没有得到提高。因此,基于回归分析构建的 GWR₄模型为研究区域土壤重金属 Cu 含量预测的最优模型。

3 讨论

从 38 次有效 GWR 建模来看,GWR 与 OLS 模型的 AICc 呈显著正相关($r=0.966,p<0.01$),并且相同自变量情况下,GWR 的 AICc 比 OLS 的 AICc 小(图 3c),这与当前已有研究的 GWR 优于 OLS 结论吻合^[8]。

从自变量对模型的贡献来看,在 UGWR 建模中,自变量 $|r|$ 与模型 AICc 呈显著负相关($r=-0.966,p<0.01$)(图 3a、3b),与模型 R^2 呈显著正相关($r=0.859,p<0.01$)(图 3d、3e),并且这种相关关系在 $|r|$ 越大时越明显。这说明自变量对 UGWR 模型的贡献与 $|r|$ 直接相关。因此,在进行

UGWR 建模时,模型自变量应以 $|r|>0.5$ 为宜,且 $|r|$ 越大,AICc 越小、 R^2 越大,模型越优。但在 BGWR、TGWR 和 FGWR 建模中,最后一次添加到模型自变量 $|r|$ 与 AICc 并不存在这种关系(图 3a、3b、3d、3e)。可见,最后一次添加到模型的自变量对模型的贡献并不取决于 $|r|$,可能取决于因变量与自变量、自变量与自变量间的耦合关系。因此,在进行 GWR 建模时,首次添加进模型的自变量 $|r|$ 越大越好;后续添加进模型的自变量不完全取决于 $|r|$,而是由自变量组合与因变量的关系确定。

随着自变量数量增加,无论是 OLS 还是 GWR,模型 AICc 总体上都呈减小趋势(图 3a、3b),而 R^2 和 R^2_{adj} 呈增大趋势(图 3d ~ 3f)。但并不一定是自变量数量多,模型就越好。如图 3b、3c 中有 7 个 BGWR 模型的 AICc 大于 GWR₁,说明只有 3 个 BGWR 模型优于 GWR₁。在 TGWR 和 FGWR 建模中存在类似的自变量数量少的模型优于自变量多的模型情况:如在 10 次 BGWR 中,GWR₂并没有出现在 $|r|>0.5$ 的其它波长中,反而出现在 $|r|=0.379$ 的 1 400 nm;同样,GWR₃和 GWR₄分别出现在 $|r|=0.589$ 的 2 197 nm 和 $|r|=0.460$ 的 2 138 nm(图 3a、3b)。本研究中,利用回归分析构建的 GWR₄ 优于 OLS 分析基础上的 GWR₇和 GWR₁₁。这说明直接利用 OLS、RS 分析取得的自变量构建 GWR 模型不一定是最优模型。

4 结论

- (1)对数导数变换光谱与土壤重金属 Cu 含量具有显著相关性,相关系数为 $-0.626 \sim -0.618$, $|r| > 0.5$ 波长数量达 436 个,最能表征研究区域土壤光谱与土壤重金属 Cu 含量的关系特征,为最佳光谱变换方式。
- (2)研究区土壤重金属 Cu 含量 GWR 预测最优

模型自变量组合为 2 308、1 400、2 197、2 138 nm 4 个波长,其构成 GWR 模型的 AICc 最小, R^2 和 R^2_{adj} 最大。

(3)与传统方法构建的 GWR 模型相比,本文基于回归分析构建的 GWR 模型更加精简,自变量数量减少到 4 个,却提高了预测数据与实测数据的 R^2 、减小了 MAE 和 RMSE。

参 考 文 献

[1] MICHAUD A M, CHAPPELLAZ C, HINSINGER P. Copper phytotoxicity affects root elongation and iron nutrition in durum wheat[J]. Plant and Soil, 2008, 310(1-2): 151-165.

[2] ALVA A K, HUANG B, PRAKASH O, et al. Effects of copper rates and soil pH on growth and nutrient uptake by citrus seedlings[J]. Journal of Plant Nutrition, 1999, 22(11):1687-1699.

[1] BRUN L A, MAILLET J, HINSINGER P. Evaluation of copper availability to plants in copper-contaminated vineyard soils[J]. Environmental Pollution, 2001, 111(2):293-302.

[2] 周葆华, 胡睿鑫, 赵宽, 等. 皖江经济带土壤 Cd 空间分布特征与生态风险评价[J]. 自然资源学报, 2021, 36(12): 3261-3270.

ZHOU Baohua, HU Ruixin, ZHAO Kuan, et al. Spatial distribution characteristic of Cd in soils and its ecological risk assessment in the economic belt of Yangtze River in Anhui[J]. Journal of Natural Resources, 2021, 36(12): 3261-3270. (in Chinese)

[3] ZHANG X Y, ZHONG T Y, LIU M, et al. Chromium occurrences in arable soil and its influence on food production in China [J]. Environment Earth Sciences, 2016, 75:257.

[4] 江振蓝, 杨玉盛, 沙晋明. 福州市土壤铬含量高光谱预测的 GWR 模型研究[J]. 生态学报, 2017, 37(23):8117-8127.

JIANG Zhenlan, YANG Yusheng, SHA Jinming. Study on GWR model applied for hyperspectral prediction of soil chromium in Fuzhou City[J]. Acta Ecologica Sinica, 2017, 37(23):8117-8127. (in Chinese)

[5] ARAÚJO S R, DEMATTÊ J A M, VICENTE S. Soil contaminated with chromium by tannery sludge and identified by vis-NIR-mid spectroscopy techniques[J]. International Journal of Remote Sensing, 2014, 35(10):3579-3593.

[6] 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]. Transaction of the Chinese Society for Agricultural Machinery, 2022, 53(8):231-239. (in Chinese)

[7] 于雷, 洪永胜, 耿雷, 等. 基于偏最小二乘回归的土壤有机质含量高光谱估算[J]. 农业工程学报, 2015, 31(14): 103-109.

YU Lei, HONG Yongsheng, GENG Lei, et al. Hyperspectral estimation of soil organic matter content based on partial least squares regression[J]. Transactions of the CSAE, 2015, 31(14):103-109. (in Chinese)

[8] 江晓宇, 李福生, 王清亚, 等. 基于惩罚最小二乘算法的土壤重金属检测光谱基线校正[J]. 农业机械学报, 2021, 52(8):205-212.

JIANG Xiaoyu, LI Fusheng, WANG Qingya, et al. Spectrum baseline correction for soil heavy metal detection based on penalized least squares algorithm[J]. Transaction of the Chinese Society for Agricultural Machinery, 2021, 52(8):205-212. (in Chinese)

[9] 钟亮, 钱家炜, 储学远, 等. 利用高光谱遥感技术监测小麦土壤重金属污染[J]. 农业工程学报, 2023, 39(5):265-270.

ZHONG Liang, QIAN Jiawei, CHU Xueyuan, et al. Monitoring heavy metal contamination of wheat soil using hyperspectral remote sensing technology[J]. Transactions of the CSAE, 2023, 39(5):265-270. (in Chinese)

[10] 林楠, 刘翰霖, 孟祥发, 等. 基于高光谱的黑土区土壤重金属含量估测[J]. 农业机械学报, 2021, 52(3):218-225.

LIN Nan, LIU Hanlin, MENG Xiangfa, et al. Hyperspectral estimation of heavy metal contents in black soil region[J]. Transaction of the Chinese Society for Agricultural Machinery, 2021, 52(3):218-225. (in Chinese)

[11] 苗瑞雪, 贺银海, 刘孝阳, 等. 利用机器学习模型预测土壤重金属空间分布的研究进展[J]. 环境科学研究, 2025, 38(5):1087-1097.

MIAO Ruixue, HE Yin Hai, LIU Xiaoyang, et al. Predicting the spatial distribution of heavy metals in soil using machine learning models: a review[J]. Research of Environmental Sciences, 2025, 38(5):1087-1097. (in Chinese)

[12] 毛继华, 赵恒谦, 金倩, 等. 河北铅锌尾矿库区土壤重金属含量高光谱反演方法对比[J]. 农业工程学报, 2023, 39(22):144-156.

MAO Jihua, ZHAO Hengqian, JIN Qian, et al. Comparative study on the hyperspectral inversion methods for soil heavy metal

contents in Hebei lead-zinc tailings reservoir areas[J]. Transactions of the CSAE, 2023, 39(22):144 – 156. (in Chinese)

[13] 田安红, 李智缘, 付承彪, 等. 云南墨江矿区周边山地农田土壤重金属的高光谱反演[J]. 农业工程学报, 2025, 41(5):191 – 200.

TIAN Anhong, LI Zhiyuan, FU Chengbiao, et al. Hyperspectral inversion of soil heavy metals in mountain farmland around Mojiang Mining Areas in Yunnan, China[J]. Transactions of the CSAE, 2025, 41(5):191 – 200. (in Chinese)

[14] WANG Y L, ZHAO J H, GUO Z W, et al. Soil moisture inversion based on data augmentation method using multisource remote sensing data[J]. Remote Sensing, 2023, 15(7):1899.

[15] SHI T Z, CHEN Y Y, LIU Y L, et al. Visible and near-infrared reflectance spectroscopy-an alternative for monitoring soil contamination by heavy metals[J]. Journal of Hazardous Materials, 2014, 265:166 – 176.

[16] JABER S M, AL-QINNA M I. Global and local modeling of soil organic carbon using Thematic Mapper data in a semi-arid environment[J]. Arabian Journal of Geosciences, 2015, 8(5): 3159 – 3169.

[17] 江振蓝, 杨玉盛, 沙晋明. GWR 模型在土壤重金属高光谱预测中的应用[J]. 地理学报, 2017, 72(3):533 – 544.

JIANG Zhenlan, YANG Yusheng, SHA Jinming. Application of GWR model in hyperspectral prediction of soil heavy metals [J]. Acta Geographica Sinica, 2017, 72(3):533 – 544. (in Chinese)

[18] 雷宇斌, 刘宁, 郭云开, 等. 高光谱组合变换下土壤 Cd 含量 GWR 模型反演研究[J]. 测绘工程, 2018, 27(11):71 – 76.

LEI Yubin, LIU Ning, GUO Yunkai, et al. Inversion of soil Cd content with GWR model under hyper-spectral combined transformation[J]. Engineering of Surveying and Mapping, 2018, 27(11):71 – 76. (in Chinese)

[19] LIN X, SU Y C, SHANG J L, et al. Geographically weighted regression effects on soil zinc content hyperspectral modeling by applying the fractional-order differential[J]. Remote Sensing, 2019, 11:636.

[20] 阿依努尔·麦提努日, 麦麦提吐尔逊·艾则孜, 李新国. 基于地理加权回归模型的葡萄园土壤砷含量高光谱反演[J]. 环境监测管理与技术, 2021, 33(4):45 – 48.

MATNURI Ayur, EZIZ Mamtursum, LI Xinguo. Hyperspectral inversion of arsenic content in vineyard soil based on geographically weighted regression model[J]. The Administration and Technique of Environmental Monitoring, 2021, 33(4): 45 – 48. (in Chinese)

[21] 刘琼峰, 李明德, 段建南, 等. 农田土壤铅、镉含量影响因素地理加权回归模型分析[J]. 农业工程学报, 2013, 29(3): 225 – 234.

LIU Qiongfeng, LI Mingde, DUAN Jiannan, et al. Analysis on influence factors of soil Pb and Cd in agricultural soil of Changsha suburb based on geographically weighted regression model[J]. Transactions of the CSAE, 2013, 29(3):225 – 234. (in Chinese)

[22] LU B B, CHARLTON M, HARRIS P, et al. Geographically weighted regression with a non-euclidean distance metric: a case study using hedonic house price data[J]. International Journal of Geographical Information Science, 2014, 28(4):660 – 681.

[23] YANG Q Y, JIANG Z C, MA Z L, et al. Spatial prediction of soil water content in Karst area using prime terrain variables as auxiliary coKriging variable[J]. Environmental Earth Sciences, 2014, 72(11):4303 – 4310.

[24] NORDHAUSEN K. An introduction to statistical learning-with applications in R by GARETH James, DANIELA Witten, TREVOR Hastie & ROBERT Tibshirani[J]. International Statistical Review, 2014, 82(1):156 – 157.

[25] 杨奇勇, 李文军, 肖琼, 等. 基于高光谱协同的土壤重金属 Hg 含量空间预测[J]. 农业机械学报, 2025, 56(11):651 – 657,676.

YANG Qiyong, LI Wenjun, XIAO Qiong, et al. Spatial prediction of soil heavy meta Hg content based on hyperspectral data as covariates[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(11):651 – 657,676. (in Chinese)