



李春泽, 张超, 张皓源, 杨翠翠, 李珊儿, 鄢文聚. 基于XGBoost与地理加权回归的吉林省西部土壤盐渍化反演[J]. 中国农业大学学报, 2024, 29(02): 1-10.
LI Chunze, ZHANG Chao, ZHANG Haoyuan, YANG Cuicui, LI Shaner, YUN Wenju. Inversion of soil salinization in western Jilin Province based on XGBoost and Geographically Weighted Regression[J]. *Journal of China Agricultural University*, 2024, 29(02): 1-10.
DOI: 10.11841/j.issn.1007-4333.2024.02.01

基于XGBoost与地理加权回归的吉林省西部土壤盐渍化反演

李春泽¹ 张超^{1*} 张皓源¹ 杨翠翠¹ 李珊儿¹ 鄢文聚²

(1. 中国农业大学 土地科学与技术学院, 北京 100193;

2. 自然资源部 国土整治中心, 北京 100035)

摘要 为及时、准确地应用多源遥感数据提取干旱和半干旱区域土壤盐渍化反演特征及获取土壤盐渍化程度的空间分布数据,以吉林省西部大安市为研究区,利用Sentinel-1 SAR、Sentinel-2 MSI多源遥感数据和DEM数据,构建土壤盐分含量(Soil Salt Content, SSC)反演特征集,结合BorutaShap算法优选特征,通过耦合地理加权回归(Geographically weighted regression, GWR)与极限梯度提升树(XGBoost)回归构建土壤盐渍化反演模型,并与XGBoost回归、GWR反演结果对比分析。结果表明:SSC反演特征集中,盐分指数、植被指数在BorutaShap算法中取得了较高的重要性排名,是大安市SSC反演的重要特征。GWR模型的 R^2 和RMSE分别为0.48和4.83 g/kg, XGBoost回归模型的 R^2 和RMSE分别为0.54和4.35 g/kg,耦合GWR与XGBoost回归构建的土壤盐渍化反演模型预测精度得到显著提高, R^2 与RMSE分别达到0.63和3.71 g/kg。依据该模型反演结果,大安市各类盐渍土分布存在较强的空间异质性,土壤盐分含量呈现出由东南向西北逐渐递减的趋势,与实地调查基本一致。综上,耦合GWR与XGBoost回归模型充分考虑了反演特征的空间异质性和非线性关系,可有效提高SSC反演精度,可获得更符合实际的SSC空间分布,可用于干旱和半干旱地区土壤盐分含量的反演。

关键词 土壤盐渍化; 遥感反演; 地理加权回归; XGBoost; 特征优选; 吉林省西部

中图分类号 S107

文章编号 1007-4333(2024)02-0001-10

文献标志码 A

Inversion of soil salinization in western Jilin Province based on XGBoost and Geographically Weighted Regression

LI Chunze¹, ZHANG Chao^{1*}, ZHANG Haoyuan¹, YANG Cuicui¹, LI Shaner¹, YUN Wenju²

(1. College of Land Science and Technology, China Agricultural University, Beijing 100193, China;

2. Land Consolidation Center of the Ministry of Natural Resources, Beijing 100035, China)

Abstract In order to timely and accurately applying multi-source remote sensing data to extract soil salinization inversion characteristics in arid and semi-arid areas and obtain spatial distribution data of soil salinization degree. Taking Da'an City in western Jilin Province as the research area, Sentinel-1 SAR, Sentinel-2 MSI multi-source remote sensing data and DEM data were used to construct a soil salt content (SSC) inversion feature set. Combined with the BorutaShap algorithm to optimize features, the coupling Geographically Weighted Regression (GWR) with XGBoost regression was used to construct a soil salinization inversion model. The results were compared with XGBoost regression and GWR inversion results. The results show that: The SSC inversion features of this study are concentrated. Salinity index and vegetation index achieve high importance ranking in the BorutaShap algorithm and are important features of

收稿日期: 2023-06-19

基金项目: 国家重点研发项目(2021YFD1500202)

第一作者: 李春泽(ORCID:0009-0003-5133-9434),硕士研究生, E-mail: 2116016705@qq.com

通讯作者: 张超(ORCID:0000-0002-0597-0861),教授,主要从事农业与国土资源遥感监测研究, E-mail: zhangchaobj@cau.edu.cn

SSC inversion in Da'an City. In the soil salinization inversion model constructed, the R^2 and RMSE of the GWR model are 0.48 and 4.83 g/kg, respectively, and the R^2 and RMSE of the XGBoost regression model are 0.54 and 4.35 g/kg, respectively. The prediction accuracy of the soil salinization inversion model constructed by coupling GWR and XGBoost regression is significantly improved, with R^2 and RMSE reaching 0.63 and 3.71 g/kg, respectively. According to the inversion results of this model, there is strong spatial heterogeneity in the distribution of various types of saline soil in Da'an City. The SSC shows a gradually decreasing trend from southeast to northwest, which is basically consistent with the field survey. In summary, the coupled GWR and XGBoost regression model constructed in this study fully considers the spatial heterogeneity and nonlinear relationship of the inversion characteristics, which effectively improves the SSC inversion accuracy. A more realistic SSC spatial distribution can be obtained. This model can be used for the inversion of soil salinity content in arid and semi-arid areas.

Keywords soil salinization; remote sensing inversion; geographically weighted regression; XGBoost; feature optimization; western part of Jilin Province

土壤盐渍化又称土壤盐碱化,是指由自然因素或人为因素引起的土壤中盐分含量不断累积,并超过一定限度的现象^[1-2]。土壤盐渍化引起土壤结构变差,进一步影响作物生长,使产量下降,严重阻碍了区域农业可持续发展^[3-4]。全球受初级盐渍化影响的土壤面积约为 $9.55 \times 10^8 \text{ hm}^2$,而次生盐渍化影响约为 $7.70 \times 10^7 \text{ hm}^2$,其中 58% 在灌溉区^[5]。我国盐碱地约有 $9.91 \times 10^7 \text{ hm}^2$,主要分布在我国东北、西北、华北内陆地区以及长江以北的沿海地带^[6]。我国盐碱地分布广、面积大,并且土壤盐渍化呈现不断扩张态势^[7]。然而,在我国盐渍化数据贫乏的地区,缺乏准确的土壤盐渍化基础数据。因此,应用多源遥感数据,获取更高精度、高分辨率、更新的土壤盐渍化专题图,对盐碱地的科学开发和利用具有重要意义。

区域盐碱土壤分布特征复杂,包括土壤盐分含量空间异质性以及不同类型盐渍化反演特征之间的非线性关系,给盐渍化土壤的监测带来困难^[8]。多光谱影像、SAR影像等多源遥感数据以及地形数据,结合地面调查,可提高区域土壤盐渍化监测的效率和精度^[9]。光谱波段、植被指数、盐分指数、地形特征、缨帽变换因子、SAR影像及其衍生的指数等均是土壤盐分含量反演的重要特征,特征选择可以通过减少过拟合来提高模型的预测精度和泛化能力,并简化模型以提高可解释性^[10]。杨练兵等^[11]针对随机森林回归算法土壤盐渍化反演,将弹性网络、遗传算法、贝叶斯优化算法应用于特征集、模型参数的同步和分步优化。Ge等^[12]基于 Boruta 特征选择算法和多种机器学习算法反演土壤盐渍化,结果表明光谱信息和环境协变量是土壤盐渍化反演

的重要特征。遥感特征与土壤盐分之间为非线性关系,机器学习方法擅长处理高维数据和解决非线性问题,成为大范围监测土壤盐渍化的主要方法^[13-14]。Zhang等^[15]基于双偏振 SAR 数据,建立了基于一维卷积神经网络的土壤盐分回归模型(DNN-s),实现了土壤盐分含量的反演。Brungard等^[16]对比了 11 种机器学习算法,发现随机森林等更复杂和更精密的建模方法能显著提高土壤属性制图识别的精度。结合辅助变量的地理加权回归可更好地揭示土壤属性空间变异的局部信息,蒙莉娜等^[17]通过对比普通克里金、地理加权回归和随机森林 3 种土壤盐渍化反演模型,发现地理加权回归模型可表现土壤属性指标的局部空间依赖性和异质性。土壤盐渍化程度与环境条件密切相关,模型的预测能力也随着环境条件而变,并且空间距离越近,土壤盐分含量也越相近。选取合适的预测模型,对于准确反演干旱和半干旱地区土壤盐渍化程度尤为重要。GWR 结合机器学习方法已应用于人文地理和统计学领域^[18-20],GWR 算法能充分考虑位置因素对土壤盐分含量的影响,XGBoost 算法可准确模拟多因素非线性关系影响下复杂的土壤盐分含量积累过程^[21]。目前,结合 GWR 与 XGBoost 算法进行土壤盐渍化反演的相关研究鲜有报道,本研究以吉林省西部大安市为研究区,利用多源遥感及衍生数据与土壤表层实测含盐量数据,在 BorutaShap 算法优化反演特征的基础上,耦合 GWR 与 XGBoost 回归进行土壤盐渍化反演,并与 XGBoost、GWR 算法反演结果对比分析,旨在应用多源遥感数据获取及时准确的干旱和半干旱区域土壤盐渍化程度的空间分布数据,以期为干旱和半

干旱区域盐碱地开发利用提供参考。

1 研究区概况与方法

1.1 研究区概况

吉林省大安市地处我国东北地区西部,松嫩平原中南部,位于我国重要的粮食生产基地和世界著名的黄金玉米带。地理坐标为 123°08′45″~124°21′56″ E, 44°57′00″~45°45′51″ N, 总面积约为 4 879 km²。该区域地势平坦,东西略高,中间较低,海拔 120~160 m,土壤盐分易于积聚于中部低洼地区。大安市地处温带大陆性季风气候区,夏季高温多雨,春季干燥多风。降水量季节变化大,造成了土壤雨季脱盐,旱季积盐的交替变化,加快了土壤中盐分的积累过程。第三次全国国土调查数据显示,大安市盐碱地面积共 640.31 km²,占大安市总面积的 13.12%,仅次于耕地和草地,主要分布在中部和南部地区。大安市东北部土壤类型主要为黑钙土和淡黑钙土,西南部土壤类型主要为风沙土、盐碱土和草甸土,具有明显的过渡特征,是开展土壤盐渍化遥感反演研究的典型区域。研究区真彩色遥感影像图、样点分布和主要道路见图 1。

1.2 土壤样本采集及含盐量分析

综合考虑土地利用类型、土壤盐渍化程度和道路通达性,在研究区布设 89 个采样点,具体采样点分布见图 1。2021 年 5 月 16—25 日实地采集土壤样品,采用手持式 GPS 进行采样点定位,基于五点法(距离 10 m)进行土壤采样,采样深度为 0~20 cm。土壤样

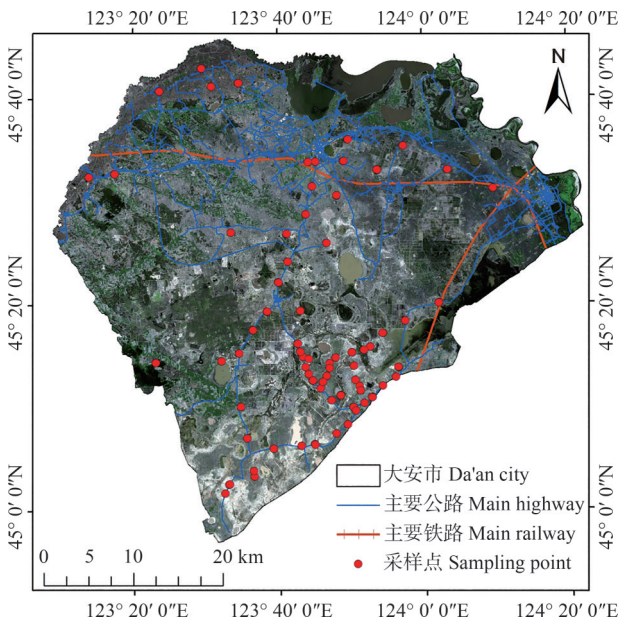


图 1 研究区地理位置

Fig. 1 Location map of study area

品经过自然风干、去除杂质、磨碎、过筛后,测定土壤中 Ca²⁺、K⁺、Mg²⁺、Na⁺、CO₃²⁻、HCO₃⁻、Cl⁻、SO₄²⁻ 8 大离子含量,获得土壤盐分含量(SSC),作为采样点 SSC 的实际测量值。样点 SSC 的统计分析见表 1,SSC 0.37~30.35 g/kg,均值为 8.57 g/kg,整个研究区土壤含盐量较高,标准差为 8.37 g/kg,变异系数接近于 1,说明研究区土壤盐分含量的空间变异性较强。研究区为典型的苏打碱土区,土壤 pH 的变化范围在 6.99~10.22,整个研究区的 pH 偏高,变异系数为 0.08。

表 1 样点土壤属性统计分析表

Table 1 Descriptive statistics of soil properties

指标 Index	最大值 Max	最小值 Min	平均值 Mean	标准差 Standard deviation	变异系数/% Variable coefficient
SSC/(g/kg)	30.35	0.37	8.57	8.37	0.97
pH	10.22	6.99	8.50	0.73	0.08

1.3 遥感数据及反演特征构建

采用的遥感影像数据来源于谷歌地球引擎(Google Earth Engine, GEE)。参考土壤盐渍化反演相关研究,选取与样点采集时间相同的 Sentinel-1 SAR 数据和云量≤10% 的 Sentinel-2 MSI 数据以及 30 m 分辨率的 DEM 数据,并计算植被指数、盐分指数、地形指数、缨帽变换因子和主成分变换因子

等,构成包含 46 个特征的特征集,见表 2。

1.4 研究方法

首先以贝叶斯优化后的 XGBoost 模型作为 BorutaShap 特征选择算法的特征重要性评估器,优化由 Sentinel-1 SAR、Sentinel-2 MSI 多源遥感数据和 DEM 数据构建的反演特征集,作为土壤盐分含量反演模型的输入。然后耦合 GWR 与 XGBoost 回

表2 SSC反演特征集
Table 2 Inversion features of SSC

数据源 Data source	类型 Type	特征 Feature
Sentinel-2 MSI	植被指数 Vegetation index	归一化植被指数(NDVI) ^[12] 、土壤调节植被指数(SAVI) ^[12] 、增强植被指数(EVI) ^[12] 、冠层盐度响应植被指数(CRSI) ^[22] 、三波段指数(TBI) ^[12] 、拓展的归一化植被指数(ENDVI) ^[23] 、拓展的增强植被指数(EEEVI) ^[23] 、差值植被指数(DVI) ^[14] 、大气阻抗植被指数(ARVI) ^[14] 、修改型土壤调节植被指数(MSAVI) ^[14]
	盐分指数 Salinity index	归一化盐分指数(NDSI) ^[12] 、盐分指数(SI) ^[12] 、盐分指数(SI2) ^[12] 、盐分指数(S1) ^[22] 、盐分指数(S2) ^[22]
	缨帽变换因子 Tassel transformation	绿度指数(GVI) ^[25] 、亮度指数(BI) ^[25] 、湿度指数(WI) ^[25]
	主成分分析波段 Principal component analysis	PCA1 ^[25] 、PCA2 ^[25] 、PCA3 ^[25]
	原始波段反射率 Original band reflectivity	Sentinel-2 MSI(蓝波段(B2)、绿波段(B3)、红波段(B4)、近红外波段(B8)、短波红外波段1(B11)、短波红外波段2(B12)) ^[24]
Sentinel-1 SAR	SAR数据及衍生指数 SAR and derivative indices	Sentinel-1 SAR原始后向散射特征(VV、VH) ^[15] 、后向散射系数比(Ratio) ^[15] 、总散射功率(SPAN) ^[15] 、差值指数(NDI) ^[15] 、SAR植被指数(RVI) ^[15] 、平方差指数(SDI) ^[15]
SRTM	地形特征 Terrain features	高程(Elevation) ^[12] 、坡度(Slope) ^[12] 、坡向(Aspect) ^[12] 、地形湿润度指数(TWI) ^[12] 、谷深(VD) ^[22] 、LS因子(LSF) ^[22] 、总流域面积(TCA) ^[22] 、收敛指数(CI) ^[22] 、山体阴影(AH) ^[22] 、相对坡度位置(RSP) ^[22] 、河网基准面(CNBL) ^[22]
	其他反演特征 Other inversion features	土地利用类型(LULC) ^[12] 、土壤类型(Soil_types)

归构建土壤盐渍化反演模型反演SSC,并与XGBoost回归和GWR方法对比分析,具体流程见图2。

1.4.1 BorutaShap特征选择

通过BorutaShap算法^[26]对本研究构建的土壤盐渍化反演特征集进行特征选择,选取反演土壤盐渍化的重要特征,去除冗余特征,减少过度拟合,提高反演精度。BorutaShap是一种包装特征选择方法,该算法主要流程:1)创建特征数据集中所有特征的阴影特征,打乱特征顺序消除变量间的相关性。2)利用贝叶斯优化后的XGBoost模型作为特征重要性评估器,对特征进行排序,以阴影特征的重要性最大值为阈值。3)重要性显著低于阈值的特征被视为“不重要”,并将其从特征集中移除。4)删

除所有阴影特征,并重复该过程,直到为每个特征均分配重要性。

1.4.2 土壤盐渍化反演模型

GWR^[27]是一种在空间非平稳性和位置相关性假设下对空间关系进行建模的局部回归方法。GWR模型如下:

$$\gamma_i = \beta_{i0} + \sum_{k=1}^m \beta_{ik} \omega_{ki} + \epsilon_i; i = \{1, \dots, n\} \quad (1)$$

式中: γ_i 为*i*位置处的因变量; β_{i0} 为*i*位置的估计截距; ω_{ki} 为*i*位置 $k = \{1, \dots, m\}$ 自变量的向量; ϵ_i 为*i*处估计的误差项。在*i*位置的局部参数估计表示为式(2):

$$\hat{\beta}(i) = [X^T W(i) X]^{-1} X^T W(i) \gamma \quad (2)$$

式中: $\hat{\beta}$ 为第*k*个自变量在*i*个位置的空间加权估计

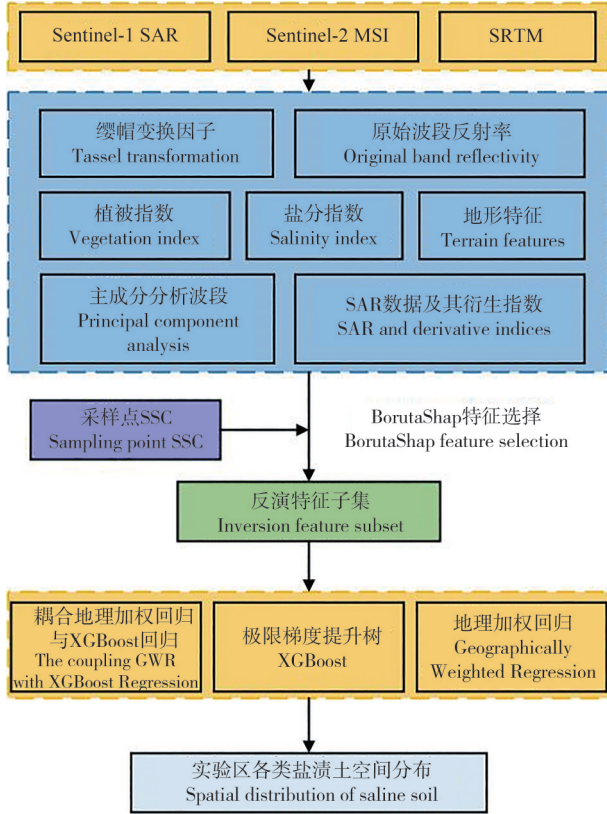


图 2 反演模型流程

Fig. 2 Inversion model process

向量; X 为自变量的 $n \times k$ 矩阵; $W(i)$ 为 $n \times n$ 对角权重矩阵。如式(3), 确保 i 附近的观测值具有最大的权重值, γ 为因变量的 k 个观测值向量。

$$W_{ij} = \exp\left(-\frac{|d_{ij}|}{bw}\right) \quad (3)$$

式中: W_{ij} 为估计 i 而分配给观测 j 的权重; d_{ij} 为 j 和 i 之间的距离; bw 为带宽。

极限梯度提升树(XGBoost)算法是在梯度提升决策树基础上优化得到的集成学习方法^[28], 是梯度提升决策树的高效实现。XGBoost模型更新采用梯度提升(Gradient boosting)策略, 通过更新损失函数的负梯度方向值来更新模型, 其优化方式可以表示为:

$$O(\emptyset) = \sum_i l(\hat{y}_i, y_i) + \sum_n \Omega(f_n) \quad (4)$$

其中

$$\Omega(f_n) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (5)$$

式中: $l(\hat{y}_i, y_i)$ 为损失函数, 即均方根误差; $\Omega(f_n)$ 为正则化项; γ 为模型复杂度; T 为模型中叶子结点个数; λ 为固定系数; ω 为叶子节点量化权重向量。

耦合 GWR 与 XGBoost 回归构建土壤盐渍化反演模型, 具体过程如下: 1) 对于每个样本点 i , 用除 i 之外的数据训练 XGBoost 模型; 2) 使用训练好的模型, 计算样本点 i 数据的预测值; 3) 使用真实值和预测值计算残差, 捕获变量间的非线性关系; 4) 将数据和残差传递给 GWR 模型, 捕获每个数据的空间相关性。5) 将 GWR 模型残差与真实值传递给对所有样本点训练的全局 XGBoost 模型, 计算全局 XGBoost 模型预测结果。土壤盐渍化反演模型见图 3。

1.4.3 模型精度验证

随机选取 30% 的样本作为验证样本, 训练数据集和验证数据集相互独立, 采用决定系数(R^2)和均方根误差(RMSE)评估模型精度。

$$R^2 = 1 - \frac{\sum_{i=1}^n (V_{pi} - V_{oi})^2}{\sum_{i=1}^n (\bar{V}_{oi} - V_{oi})^2} \quad (6)$$

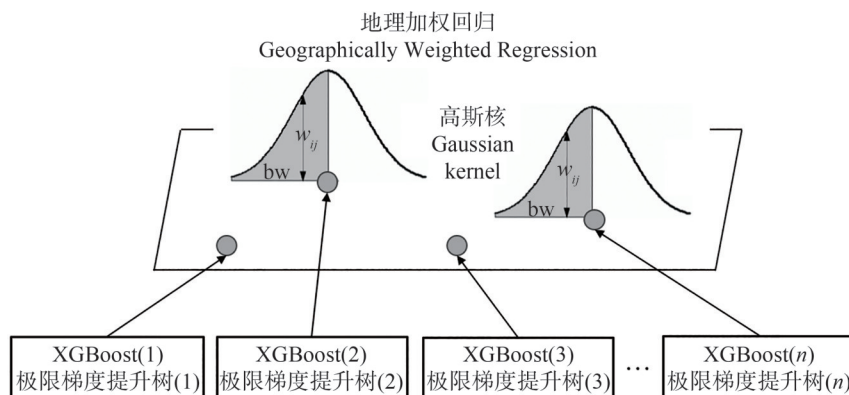


图 3 土壤盐渍化反演模型

Fig. 3 Soil salinization inversion model

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (V_{oi} - V_{pi})^2} \quad (7)$$

式中: V_{oi} 为样本点 i 处的实测值; V_{pi} 为点 i 处的预测值; n 为总样本点个数。 R^2 在 0~1, 越接近 1, 模型的稳定性越高; RMSE 越小, 则模型预测土壤盐分含量的精度越高。

1.4.4 土壤盐渍化程度分级标准

参照文献[29-30]拟定的土壤盐渍化程度分级, 本研究根据 SSC 将大安市土壤盐渍化程度分为盐

渍土 ($SSC > 10 \text{ g/kg}$)、重度盐渍化 ($6 \sim 10 \text{ g/kg}$)、中度盐渍化 ($3 \sim 6 \text{ g/kg}$)、轻度盐渍化 ($1 \sim 3 \text{ g/kg}$)、非盐渍化 ($< 1 \text{ g/kg}$) 5 级, 并依据此标准对大安市土壤盐渍化反演结果绘制 SSC 专题图。

2 结果与分析

2.1 基于 BorutaShap 的反演特征优选

以贝叶斯优化算法优化后的 XGBoost 模型作为 BorutaShap 算法的特征重要性评估器, XGBoost 模型优化参数、优化空间和优化结果见表 3。

表 3 XGBoost 算法参数优化

Table 3 XGBoost algorithm parameter optimization

参数 Parameter	含义 Meaning	优化空间 Optimization space	优化结果 Optimization result
n_estimators	生成树的最大数目	10~1 000	890
max_depth	树的最大深度	3~18	13
gamma	节点分裂所需的最小损失函数下降值	1.00~9.00	1.19
min_child_weight	最小叶子节点样本权重和	0~10	4
colsample_bytree	每棵树随机采样列数的比例	0.50~1.00	0.93

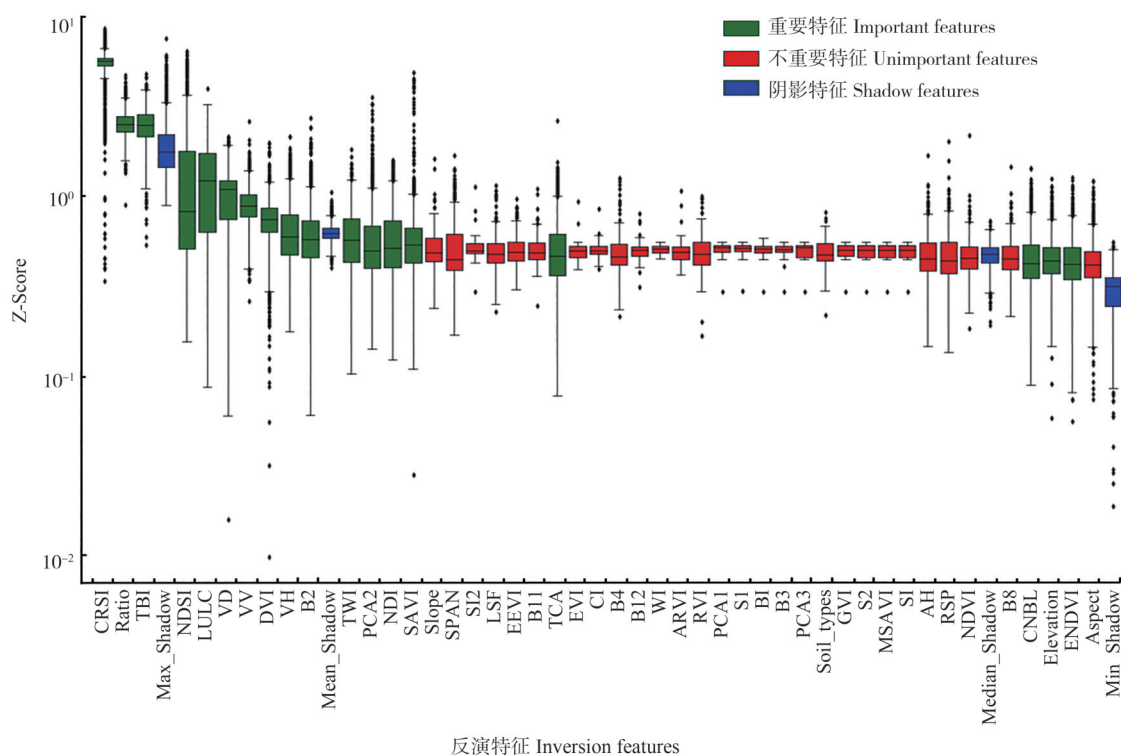
由图 4 可知, 对于大安市土壤盐渍化反演特征集, Sentinel-2 MSI 原始波段反射率中, B2 波段为盐渍化反演重要波段。植被指数中重要的反演特征为 CRSI、TBI、SAVI、ENDVI、DVI, 其中 CRSI 与 TBI 具有较高的重要性排序。盐分指数中, NDSI 是 SSC 反演的重要特征。缨帽变换因子、主成分分析波段等二次特征中, PCA2 对 SSC 反演比较重要。本研究重要的地形特征为 TWI、Elevation、CNBL、TCA 和 VD。但本研究区域地势较为平坦, 地形因素在本研究区土壤盐分反演中未起决定性作用, 因此地形特征的重要性排序较低。本研究构建的 SAR 数据及衍生指数, VV、VH、Ratio、NDI 被选为 SSC 反演的重要特征。本研究中 LULC 取得了较高的重要性排序。

2.2 SSC 反演精度对比分析

由图 5 可知, 耦合 GWR 与 XGBoost 回归方法预测值与测量值更为接近, 精度最高, R^2 和 RMSE 分别为 0.63 和 3.71 g/kg。XGBoost 回归模型的 R^2 和 RMSE 分别为 0.54 和 4.35 g/kg, 而在 GWR 模型中 R^2 和 RMSE 最差, 分别为 0.48 和 4.83 g/kg。

由图 6 可知, 大安市 SSC 呈现出由东南向西北逐渐递减的趋势。大安市东南侧及中部地势较为低洼, 盐分易于累积, 因此南部和中部反演结果主要为盐渍土、重度盐渍化土和中度盐渍化土。大安市西北部主要土地类型为林地与耕地, 该区域反演结果多为轻度盐渍化土壤和无盐渍化土壤。根据耦合 GWR 与 XGBoost 回归方法制作大安市 SSC 专题图, 大安市各类盐渍土分布存在较强的空间异质性。大安市轻度和中度盐渍化土壤面积占比最大, 西北部和北部 SSC 相对较低, SSC 高的区域主要集中在植被覆盖较少的东南部和南部地区。

通过对比 GWR、XGBoost 和耦合 GWR 与 XGBoost 回归 3 种土壤盐渍化反演方法, GWR 方法反演结果平均 SSC 较高, XGBoost 方法反演的 SSC 空间连续性较差。耦合 GWR 和 XGBoost 回归方法预测精度最高, 并且该方法对 SSC 的空间局部变异信息描述更精细, 较好地反映了研究区 SSC 的空间异质性及其变化规律。因此, 结合机器学习算法与地理加权回归可以显著提高土壤盐渍化反演精度。



蓝色变量是阴影特征,绿色变量是重要特征,红色变量是不重要特征。

Blue variables are shaded features, green variables are important features, and red variables are unimportant features.

图 4 BorutaShap 特征选择

Fig. 4 BorutaShap feature selection

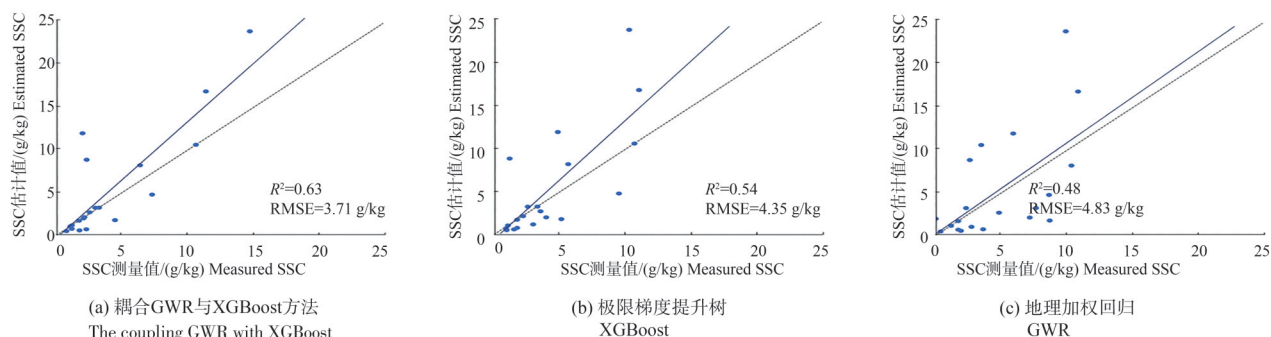


图 5 3 种模型验证集 SSC 测量值与估计值

Fig. 5 SSC measurements and estimates for 3 model validation sets

3 讨论

通过实地调查获得大范围的土壤盐渍化信息通常是费力且困难的,因此大部分土壤盐渍化反演研究中土壤采样数据稀疏。针对这一问题,Wang 等^[13]引入图像处理领域的 Mixup 数据增强方法增加样本多样性。Ge 等^[12]通过 Bootstrap 方法对样本

进行重采样,有效地解决了构建模型时出现的小样本问题。本研究中土壤样本点同样稀疏,下一步应重点考虑采用数据增强的方法,增强土壤盐分含量样本点数据,提高小样本条件下的土壤盐渍化反演精度。

土壤盐分聚集在地势低洼的地区,本研究中地形因素对于吉林省西部大安市的土壤盐分含量累

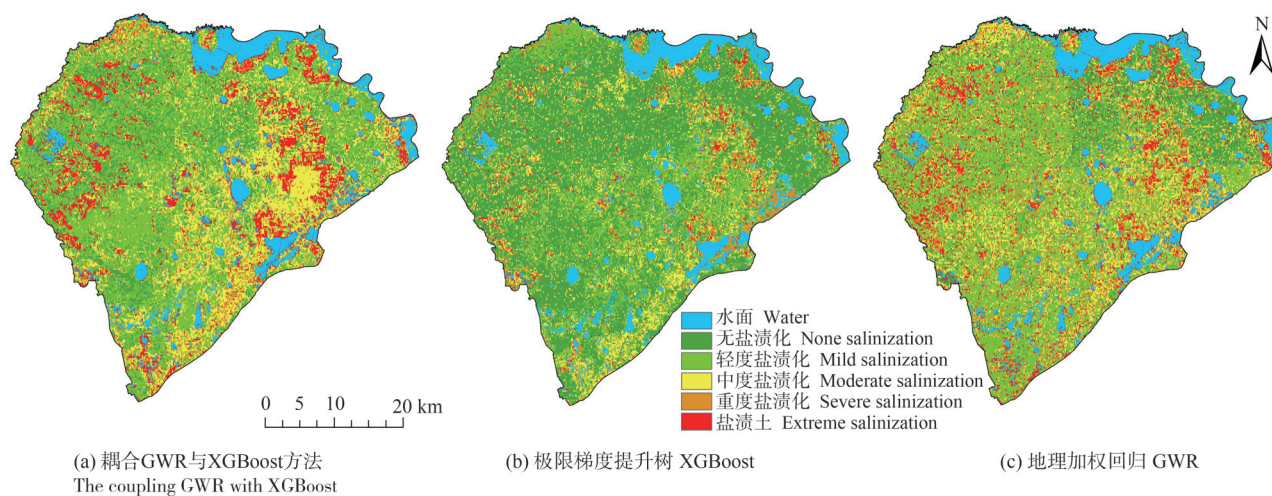


图6 3种模型土壤盐渍化反演

Fig. 6 Inversion of soil salinization in 3 models

积并未起决定性作用,但在 Peng 等^[22]和蒙莉娜等^[17]的研究中,地形因素均取得了较高的重要性排序,这证明不同区域决定土壤盐分含量的主导因素不同。在自然环境中,由于土壤湿度、植被覆盖和数据采集时间等其他因素的干扰,利用遥感技术很难获得盐渍土的纯光谱信息^[10],SSC可以通过作物长势间接表达,因此植被指数是土壤盐渍化反演的重要特征,本研究土壤盐渍化反演取得了较高的重要性排序。土地利用类型是人类活动影响 SSC 的重要表现,土地利用方式与农田灌溉方式是影响土壤盐渍化程度的重要因素^[12],本研究中 LULC 是重要特征之一。目前 GWR 结合机器学习方法主要应用在人文地理和统计学领域,在土壤盐渍化反演方面的应用还十分有限,构建土壤盐渍化反演的空间变系数的机器学习模型需不断地进行研究,以便构建适宜不同区域的土壤盐渍化反演模型。

4 结论

本研究基于实地采样点的土壤盐分含量,结合光学、SAR 遥感数据和地形数据,构建了46个反演特征,耦合 GWR 和 XGBoost 回归方法对吉林省西部大安市表层土壤盐分含量的空间分布进行反演。盐分指数、植被指数在大安市土壤盐分含量的反演中发挥重要作用,反演特征对土壤盐分含量的表征能力存在区域异质性。耦合 GWR 与 XGBoost 回归模型充分考虑了反演特征的空间异质性和非线性关系,可有效提高土壤盐分含量反演精度,且该模

型对土壤盐分含量的空间局部变异信息描述更精细,可用于干旱和半干旱地区土壤盐分含量的反演。

参考文献 References

- [1] Daliakopoulos I N, Tsanis I K, Koutroulis A, Kourgialas N N, Varouchakis A E, Karatzas G P, Ritsema C J. The threat of soil salinity: A European scale review[J]. *Science of the Total Environment*, 2016, 573: 727-739
- [2] 杨劲松. 中国盐渍土研究的发展历程与展望[J]. *土壤学报*, 2008, 45(5): 837-845
Yang J S. Development process and prospect of the research on salt-affected soils in China[J]. *Acta Pedologica Sinica*, 2008, 45(5): 837-845 (in Chinese)
- [3] 朱宏伟,夏军,曹国栋,王绍明,张霞,王进. 盐渍化弃耕地土壤盐分动态及其影响因素[J]. *土壤*, 2013, 45(2): 1339-1345
Zhu H W, Xia J, Cao G D, Wang S M, Zhang X, Wang J. Dynamic change of soil salinity in salinization abandoned farmland and affecting factors[J]. *Soils*, 2013, 45(2): 1339-1345 (in Chinese)
- [4] 于文婧,孙丹峰,曲藏. 基于冗余分析的宁夏平罗耕地土壤盐渍化特征研究[J]. *中国农业大学学报*, 2015, 20(5): 103-110
Yu W J, Sun D F, Qu W. Study of the characteristics of soil salinization based on redundancy analysis in Pingluo, Ningxia[J]. *Journal of China Agricultural University*, 2015, 20(5): 103-110 (in Chinese)
- [5] Metternicht G I, Zinck J A. Remote sensing of soil salinity: potentials and constraints[J]. *Remote Sensing of Environment*, 2003, 85(1): 1-20
- [6] 牛东玲,王启基. 盐碱地治理研究进展[J]. *土壤通报*, 2002, 33(6): 449-455
Niu D L, Wang Q J. Research progress on saline-alkali field control[J]. *Chinese Journal of Soil Science*, 2002, 33(6): 449-455 (in Chinese)
- [7] 王俊枝,薛志忠,张弛,常屹冉. 内蒙古河套平原耕地盐碱化时空演变及其对产额的影响[J]. *地理科学*, 2019, 39(5): 827-835

- Wang J Z, Xue Z H, Zhang C, Chang Y R. Spatial evolution of saline-alkali cultivated land and its impact on productivity in Hetao plain of inner Mongolia[J]. *Geographical Science*, 2019, 39(5):827-835 (in Chinese)
- [8] 王怡婧,贾萍萍,陈睿华,张俊华. 多源遥感对宁夏干旱季土壤含盐量敏感性分析与定量反演[J/OL]. *生态学杂志*, (2023-09-18). DOI:10.13292/j.1000-4890.202309.006
- Wang Y J, Jia P P, Chen R H, Zhang J H. Sensitivity analysis and quantitative inversion of multi-source remote sensing to soil salt content in dry and wet seasons in Ningxia[J/OL]. *Journal of Ecology*, (2023-09-18). DOI:10.13292/j.1000-4890.202309.006 (in Chinese)
- [9] 罗彪,刘潇,郭萍. 基于MODIS数据的河套灌区遥感干旱监测[J]. *中国农业大学学报*, 2020, 25(10):44-54
- Luo B, Liu X, Guo P. Remote sensing drought monitoring of Hetao irrigation area based on MODIS data [J]. *Journal of China Agricultural University*, 2020, 25(10): 44-54 (in Chinese)
- [10] Chen B L, Zheng H W, Luo G, Chen C S, Bao A, Liu T, Chen X. Adaptive estimation of multi-regional soil salinization using extreme gradient boosting with Bayesian TPE optimization [J]. *International Journal of Remote Sensing*, 2022, 43(3): 778-811
- [11] 杨练兵,陈春波,郑宏伟,罗格平,尚白军,Olaf Hellwich. 基于优化随机森林回归模型的土壤盐渍化反演[J]. *地球信息科学学报*, 2021, 23(9):1662-1674
- Yang L B, Chen C B, Zheng H W, Luo G P, Shang B J, Hellwich O. Retrieval of soil salinity content based on random forests regression optimized by Bayesian optimization algorithm and genetic algorithm[J]. *Journal of Geo-Information Science*, 2021, 23(9):1662-1674 (in Chinese)
- [12] Ge X Y, Ding J L, Teng D X, Wang J Z, Huo T C, Jin X Y, Wang J J, He B Z, Han L J. Updated soil salinity with fine spatial resolution and high accuracy: The synergy of Sentinel-2 MSI, environmental covariates and hybrid machine learning approaches[J/OL]. *CATENA*, (2022-02-24). DOI:10.1016/j.catena.2022.106054
- [13] Wang Y, Hu P, Zheng H W, Liu Y, Cao X W, Hellwich O, Liu T, Luo G, Bao A M, Chen X. Integrative modeling of heterogeneous soil salinity using sparse ground samples and remote sensing images [J/OL]. *Geoderma*, (2023-01-04). DOI:10.1016/j.geoderma.2022.116321
- [14] Xu H T, Chen C B, Zheng H W, Luo G P, Yang L A, Wang W S, Wu S X, Ding J L. AGA-SVR-based selection of feature subsets and optimization of parameter in regional soil salinization monitoring [J]. *International Journal of Remote Sensing*, 2020, 41(12): 4470-4495
- [15] Zhang Q Q, Li L, Sun R Z, Zhu D H, Zhang C, Chen Q Q. Retrieval of the soil salinity from Sentinel-1 Dual-Polarized SAR data based on deep neural network regression[J/OL]. *IEEE Geoscience and Remote Sensing Letters*, (2020-12-09). DOI:10.1109/LGRS.2020.3041059
- [16] Brungard C W, Boettinger J L, Duniway M C, Wills S A, Edwards Jr T C. Machine learning for predicting soil classes in three semi-arid landscapes[J]. *Geoderma*, 2015, 239-240: 68-83
- [17] 蒙莉娜,丁建丽,王敬哲,葛翔宇. 基于环境变量的渭干河-库车河绿洲土壤盐分空间分布[J]. *农业工程学报*, 2020, 36(1):175-181
- Meng L N, Ding J L, Wang J Z, Ge X Y. Spatial distribution of soil salinity in the Ugan -Kuqa River delta oasis based on environmental variables [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2020, 36(1): 175-181 (in Chinese)
- [18] Georganos S, Grippa T, Niang Gadiaga A, Linard C, Lennert M, Vanhuysse S, Mboga N, Wolff E, Kalogirou S. Geographical random forests: A spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling [J]. *Geocarto International*, 2021, 36(2): 121-136
- [19] Li L F. Geographically weighted machine learning and downscaling for high-resolution spatiotemporal estimations of wind speed[J/OL]. *Remote Sensing*, (2019-06-10). DOI: 10.3390/rs11111378
- [20] Quinones S, Goyal A, Ahmed Z U. Geographically weighted machine learning model for untangling spatial heterogeneity of type 2 diabetes mellitus (T2D) prevalence in the USA[J/OL]. *Scientific Reports*, (2021-03-26). DOI: 10.1016/j.annepidem.2017.11.002
- [21] Shahneh M R, Oymak S, Magdy A. A-gwr: Fast and accurate geospatial inference via augmented geographically weighted regression [C].// Proceedings of the 29th international conference on advances in geographic information systems. November 2-5, 2021, New York, New York State, USA. Association for Computing Machinery:2021
- [22] Peng J, Biswas A, Jiang Q S, Zhao R Y, Hu J, Hu B F, Shi Z. Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China[J]. *Geoderma*, 2019, 337: 1309-1319
- [23] 陈红艳,赵庚星,陈敬春,王瑞燕,高明秀. 基于改进植被指数的黄河口区盐渍土盐分遥感反演[J]. *农业工程学报*, 2015, 31(5):107-114
- Chen H Y, Zhao G X, Chen J C, Wang R Y, Gao M X. Remote sensing inversion of saline soil salinity based on modified vegetation index in estuary area of Yellow River[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2015, 31(5):107-114 (in Chinese)
- [24] 张素铭,赵庚星,王卓然,肖杨,郎坤. 滨海盐渍区土壤盐分遥感反演及动态监测[J]. *农业资源与环境学报*, 2018, 35(4):349-358
- Zhang S M, Zhao G X, Wang Z R, Xiao Y, Lang K. Remote sensing inversion and dynamic monitoring of soil salinity in coastal saline area[J]. *Journal of Agricultural Resources and Environment*, 2018, 35(4): 349-358 (in Chinese)
- [25] 王飞,杨胜天,丁建丽,魏阳,葛翔宇,梁琛. 环境敏感变量优选及机器学习算法预测绿洲土壤盐分[J]. *农业工程学报*, 2018, 34(22):102-110
- Wang F, Yang S T, Ding J L, Wei Y, Ge X Y, Liang J. Environmental sensitive variable optimization and machine learning algorithm using in soil salt prediction at oasis [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2018, 34(22):102-110 (in Chinese)
- [26] Eoghan K. BorutaShap: A wrapper feature selection method which combines the Boruta feature selection algorithm with Shapley values[EB/OL]. [2023-10-17] <https://zenodo.org/records/4247618>
- [27] Brunson C, Fotheringham S, Charlton M. Geographically weighted regression [J]. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 1998, 47(3): 431-443
- [28] Chen T, Guestrin C. XGBoost: a scalable tree boosting system [C].// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Minin. August 13-17, 2016, San Francisco, California, USA. New York: Association for Computing Machinery, 2016
- [29] 刘全明,成秋明,王学,李相君. 河套灌区土壤盐渍化微波雷达反演[J]. 农

业工程学报, 2016, 32(16): 109-114

Liu Q M, Cheng Q M, Wang X, Li X G. Soil salinity inversion in Hetao Irrigation district using microwave radar[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2016, 32(16): 109-114(in Chinese)

[30] 张子璇, 宋雨桐, 张惠中, 李新举, 牛蓓蓓. 水文气候影响下黄河三角洲土壤盐分时空动态[J]. *应用生态学报*, 2021, 32(4): 1393-1405

Zhang Z X, Song Y T, Zhang H Z, LI X J, Niu B B. Spatiotemporal dynamics of soil salinity in the Yellow River Delta under the impacts of hydrology and climate[J]. *Chinese Journal of Applied Ecology*, 2021, 32(4): 1393-1405(in Chinese)

责任编辑: 吕晓梅



第一作者简介: 李春泽, 中国农业大学土地科学与技术学院 2021 级硕士研究生, 专业: 土地利用与信息技术。参与国家重点研发项目子课题(2021YFD1500202)、新增耕地认定技术规范研究项目、大安市基质调查项目等多项研究课题。2022 年获得中国农业大学校级硕士学业二等奖学金。



通讯作者简介: 张超, 博士, 中国农业大学土地科学与技术学院教授, 研究领域主要为农业与国土资源遥感监测。主要学术兼职为中国土地学会科普工作委员会副主任委员(2017-)、自然资源部首席科学传播专家(2019-)、自然资源部吉林大安盐碱地整治野外科学观测研究站副站长(2021-)等。主持和参加国家重点研发计划项目(2017YFF0206801-2、2021YFD1500202)、国家自然科学基金重大项目(19ZDA096)等研究课题 10 余项。发表学术论文 100 余篇, 获专利 2 项, 软件著作权 6 项。