

达拉特旗黄河南岸盐碱化土壤不同含盐量估算模型对比

刘霞 胡宇 张圣微 白燕英 张欢

(内蒙古农业大学水利与土木建筑工程学院, 呼和浩特 010018)

摘要: 内蒙古自治区鄂尔多斯市达拉特旗黄河南岸由于气候干旱, 降水量少, 年蒸发量远大于年降水量, 靠近黄河地下水位高, 导致土壤盐渍化问题突出。以达拉特旗黄河南岸盐碱地为研究对象, 基于 Sentinel-1、Sentinel-2、Landsat-8 和 SRTM DEM 多源数据, 采取相关性分析和连续变量投影结合索套回归 (Lasso)、随机森林回归 (Random forest, RF)、轻量梯度提升机模型 (Light gradient boosting machine, LightGBM)、极端梯度提升模型 (Extreme gradient boosting, XGBoost)、一维卷积神经网络 (One dimensional convolutional neural networks, 1DCNNs)、深度神经网络 (Deep neural network, DNN) 6 种模型进行春季裸土期与植被覆盖期土壤含盐量估算。结果表明: XGBoost 模型精度最高, 春季裸土期、植被覆盖期测试集决定系数 (R^2) 为 0.76、0.58; 均方根误差 (RMSE) 为 5.76、7.22 g/kg; 平均绝对误差 (MAE) 为 3.38、4.33 g/kg。多源遥感数据结合变量筛选方法利用 XGBoost 模型揭示研究区不同季节土壤盐分空间分布最有效, 含盐量反演结果与野外实际调查分析结果基本吻合。变量重要性分析表明春季裸土期、植被覆盖期重要反演因子分别为: 盐分指数 (48.3%)、地形因子 (33.8%); 植被指数 (22%)、地形因子 (47.9%)。本研究为达拉特旗黄河南岸盐碱地遥感反演提供了有效方法, 为春季裸土期与植被覆盖期盐碱化土壤监测及预防提供了理论依据。

关键词: 干旱盐碱地; 土壤含盐量估算; 机器学习模型; 深度学习模型

中图分类号: S127 文献标识码: A 文章编号: 1000-1298 (2024) 10-0360-11

OSID:



Comparison of Different Salinity Estimation Models for Salinized Soils on South Bank of Yellow River in Dalat Banner

LIU Xia HU Yu ZHANG Shengwei BAI Yanying ZHANG Huan

(College of Water Conservation and Civil Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China)

Abstract: The south bank of the Yellow River in Dalate Banner, Ordos City, Inner Mongolia Autonomous Region, is characterized by arid climate, low precipitation, annual evaporation much larger than annual precipitation, and the proximity to the Yellow River leads to a high water table, which leads to prominent soil salinization. Taking the saline soil along the south bank of the Yellow River in Dalate Banner as the research object, based on the multi-source data of Sentinel-1, Sentinel-2, Landsat-8 and SRTM DEM, correlation analysis and continuous variable projection combined with Lasso regression (Lasso), random forest regression (RF), light gradient boosting machine model (LightGBM), extreme gradient boosting (XGBoost), one dimensional convolutional neural networks (1DCNNs), and deep neural network (DNN) were used to estimate soil salinity during spring bare soil period and vegetation cover period. The results showed that the XGBoost model had the highest accuracy, and the coefficients of determination (R^2) of the test sets were 0.76 and 0.58 for the spring bare soil period and vegetation cover period, the root mean square errors (RMSE) were 5.76 g/kg and 7.22 g/kg, and the mean absolute errors (MAE) were 3.38 g/kg and 4.33 g/kg. The combination of multi-source remote sensing data and the variable screening method by using the XGBoost model revealed that the soil salinity spatial distribution in different seasons in the study area was the most effective, and the results of salinity

收稿日期: 2024-05-24 修回日期: 2024-07-11

基金项目: 内蒙古自治区科技计划项目 (2021GG0369)、国家自然科学基金项目 (52369009)、鄂尔多斯市科技局项目 (2021EEDSCXSQZD01) 和内蒙古自然科学基金项目 (2023MS05024)

作者简介: 刘霞 (1971—), 女, 副教授, 主要从事农业水土资源利用与水土环境调控研究, E-mail: liukasumi@126.com

inversion basically coincided with the results of the actual field investigation and analysis. The variable importance analysis showed that the important inversion factors in the spring bare soil period and vegetation cover period were salinity index (48.3%) and topography factor (33.8%), vegetation index (22%) and topography factor (47.9%), respectively. The research result can provide an effective method for remote sensing inversion of saline and alkaline land on the south bank of the Yellow River in Dalat Banner, and provide a theoretical basis for monitoring and preventing salinized soil in the spring bare soil period and vegetation cover period.

Key words: arid saline soil; soil salinity estimation; machine learning models; deep learning models

0 引言

土壤盐碱化不仅制约农业可持续发展,还对生态环境造成潜在威胁。达拉特旗北部黄河冲积平原是当地主要的农业种植区^[1],由于降水量少、年蒸发量大的气候条件,地下水位高且矿化度高以及耕作粗放等因素导致土壤盐碱化加重^[2]。达拉特旗盐渍化土壤总面积为 718.34 hm², 占总土地面积的 9.23%, 其中沿河平原盐渍化面积 686.38 hm², 占平原面积的 37.16%^[3]。如何快速获取大尺度、高精度的土壤盐分信息对于盐碱化土壤监测和预防具有重要意义。

土壤盐渍化成因复杂,主要取决于植被、土壤类型、地形要素、气候因素和人类活动等因素^[4]。近些年土壤盐渍化数字制图得到显著发展,主要为线性回归模型与机器学习模型。线性模型主要有多元线性回归^[5]、偏最小二乘回归^[5]、索套回归^[6]等,索套回归在特征选择、处理高维数据和防止过拟合方面具有显著优势。机器模型主要有 BP 神经网络^[5,7]、支持向量机^[5,7]、随机森林^[7]、轻量梯度提升机模型^[8]、极端梯度提升模型^[9]等。机器学习模型在研究区土壤盐分分布不均、环境变量复杂时,可以较好估算含盐量,弥补线性模型对于非线性回归的不足^[10];其中集成机器学习在相关土壤监测研究中表现良好,泛化能力优于单个学习器的机器模型^[8]。研究者从最初的只使用多光谱盐分指数^[10-13]、植被指数^[14]、雷达数据及其衍生因子^[15-16]单源数据发展为采取多光谱数据、雷达数据、地形数据等多源遥感数据^[17-19]结合机器学习的方法进行土壤含盐量估算。这种将遥感数据及相关协变量与机器学习算法的结合,显著提高了土壤含盐量估算模型精度^[20],但利用机器学习方法构建反演模型时,随着反演因子的增加会产生大量冗余信息,特征选取可以减少模型过拟合现象,提高模型精度和泛化能力^[18,20],因而反演因子筛选至关重要。研究表明连续变量投影可以解决变量之间的共线性问题,筛选出最优组合^[21-22]。目前大部分学者^[23-24]针对土壤含盐量估算采用机器学习模型进

行对比分析,并未考虑利用深度学习模型进行土壤含盐量反演,进而对比分析线性回归模型、机器学习及深度学习模型估算效果。

针对以上研究存在的不足,采取多源遥感数据结合变量筛选方法,评估线性、机器学习、深度学习模型含盐量反演精度,以达拉特旗黄河南岸盐碱地为研究对象,基于 Sentinel-1、Sentinel-2、Landsat-8 和 SRTM DEM 多源数据及衍生数据,结合 2022、2023 年春季裸土期与植被覆盖期土壤含盐量实际采样数据,采用相关性筛选变量,再利用连续变量投影算法二次筛选,构建线性(索套回归(Lasso))、集成机器学习(随机森林回归(RF)、轻量梯度提升机模型(LightGBM)、极端梯度提升模型(XGBoost))及深度学习(一维卷积神经网络(1DCNNs^[25])、深度神经网络(DNN^[26]))土壤含盐量估算模型,探究研究区盐碱地土壤盐分反演与分级精度,以期实现土壤盐分快速监测,为盐碱地综合治理及规划提供科学依据。

1 材料与方 法

1.1 研究区域概况

研究区位于内蒙古自治区达拉特旗(东经 108°59'~109°59',北纬 40°16'~40°33'),地处黄河冲积平原南部,鄂尔多斯高原北端,地势南高北低^[1],黄河中游“几”字湾南岸(图 1)。研究区占地面积约为 1 437.12 km²,土壤类型主要为风沙土,主要水系由黄河和哈什拉川、西柳沟、黑赖沟、卜尔色太沟、毛不拉孔兑 5 条季节性一级支流河川构成,发源于鄂尔多斯台地,流经库布其沙漠,通过冲积平原汇入黄河^[27]。地下水类型主要为松散岩类孔隙水,埋深为 1~20 m。该地区属于中温带半干旱地区,多年平均气温 6.1~7.1℃,年平均降雨量为 240~360 mm(多集中于 7—8 月),蒸发量 1 450~3 250 mm^[3],平均日照时数 3 000 h,平均海拔 1 031 m。

1.2 土壤数据采集与处理

在研究区随机采样,于 2022 年 5 月 13—18 日、2022 年 8 月 24—29 日、2023 年 4 月 15—19 日、2023 年 9 月 11—14 日分别取样 85、117、219、61 个,

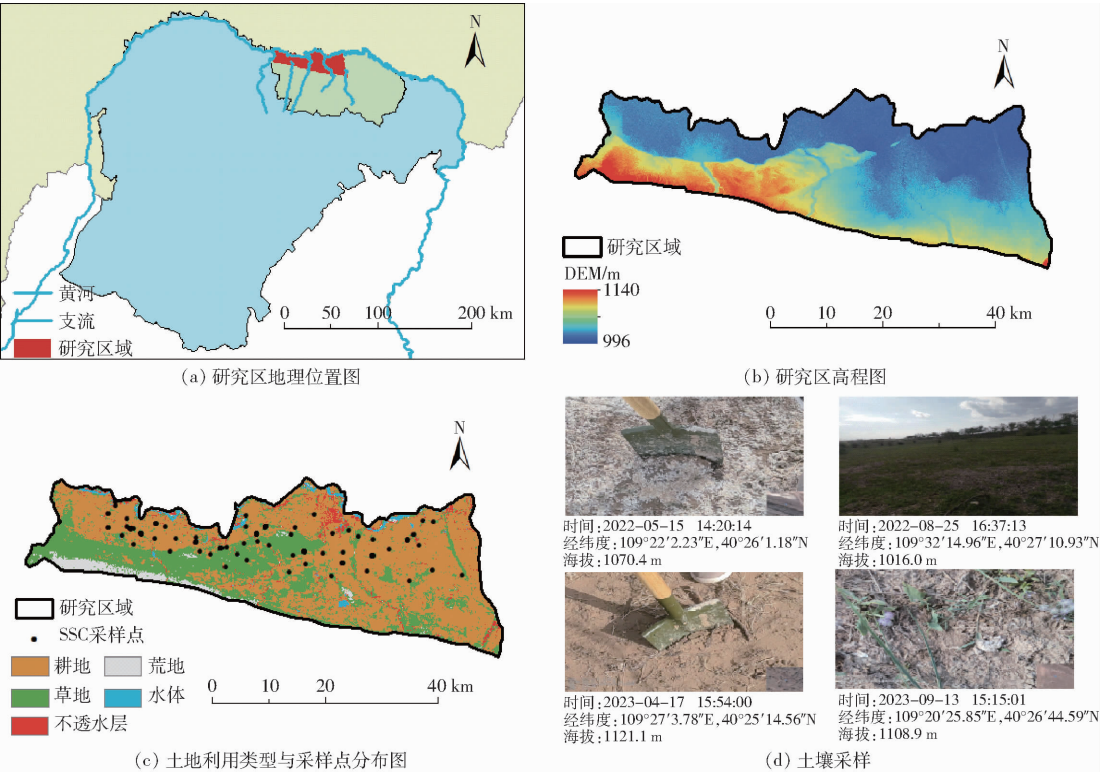


图 1 研究区概况和采样点分布

Fig.1 Overview of study area and distribution of sampling sites

取样深度为 0 ~ 10 cm,使用 GPS 记录采样点经纬度。采集的土壤样本,在自然条件下风干经研磨过 1 mm 筛,按水土质量比 5 : 1 调配土壤浸提液,测定土壤含盐量 (SSC) [28]。由于随机采样个别月份点位较少以及植被覆盖影响卫星反射率,将土壤样本划分为春季裸土期(2022 年 5 月 13—18 日、2023 年 4 月 15—19 日)与植被覆盖期(2022 年 8 月 24—29 日、2023 年 9 月 11—14 日)两组。

1.3 研究框架

研究框架如图 2 所示。首先计算各类多源遥感数据反演因子,结合皮尔逊相关分析 (PCC) 筛选与 SSC 存在显著相关因子;其次使用连续变量投影算法 (SPA) 二次筛选;然后利用评价指标对不同模型

的数据集进行评估,确定最优含盐量估算模型;最后利用最优模型完成土壤盐分的空间制图。具体反演流程为:首先将训练好的模型保存;其次利用 GDAL 读取对应变量的 TIFF 文件,调用春季裸土期及植被覆盖期各自的模型完成土壤含盐量逐像元反演;最后剔除研究区水体、不透水层,借助 ArcGIS 完成可视化。

1.4 遥感数据获取与预处理

遥感数据采用光学、雷达和地形数据,均采用 GEE 进行预处理。Sentinel-1 数据选取 IW 模式下的 Level-1 级 GRD 产品,GEE 平台已经对 Sentinel-1 进行热噪声去除、辐射校正、基于 SRTM 30 m 分辨率数据的地形校正等预处理 [29],再利用 Python 版

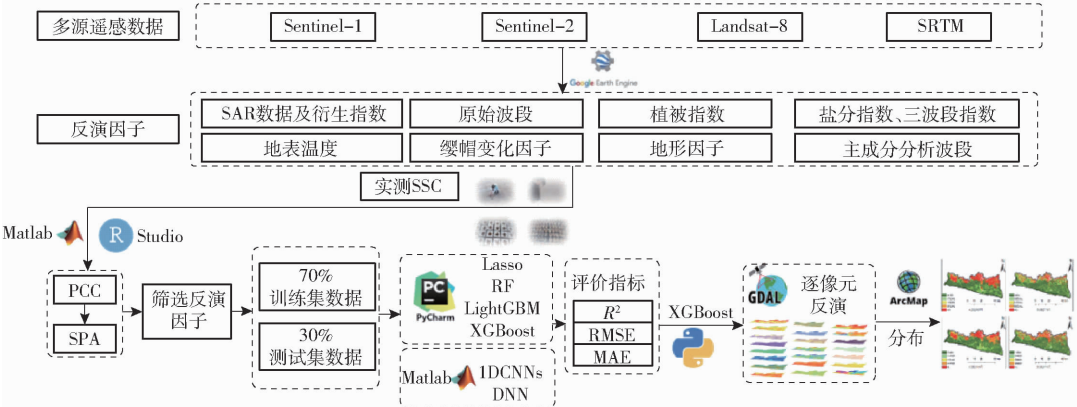


图 2 研究框架图

Fig.2 Research framework diagram

GEE 采用 5×5 Refined - Lee 滤波减少图像噪声^[30],将分辨率重采样到 10 m。Sentinel - 2 选取经大气处理和几何校正数据集,筛选与取样时间一致的影像,经过 QA60 去云、中值合成(减少影像数据量便于后续快速分析)、裁剪、重采样、下载本地。Landsat - 8 采用 SR 级影像,利用单窗算法反演地表温度。地形数据采用 USGS/SRTMGL1_003 数据集,

空间分辨率 1"(约 30 m),为后续分析将影像重采样至 10 m,使用 SAGA 计算多种地形因子。

1.5 反演因子

采取 SAR 数据及衍生指数、原始波段、植被指数、盐分指数、缨帽变化因子、主成分分析波段、三波段指数、地表温度以及地形因子 65 种反演因子,如表 1 所示。

表 1 反演因子
Tab.1 Inversion factors

数据来源	类型	变量	总计
Sentinel - 1	SAR 数据及衍生指数	原始后向散射特征(VV、VH)、后向散射系数比(Ratio) ^[31] 、总散射功率(SPAN) ^[31] 、差值指数(NDI) ^[31] 、SAR 植被指数(RVI) ^[31] 、平方差指数(SDI) ^[31]	7
	原始波段	Aerosols、Blue、Green、Red、Red Edge 1、Red Edge 2、Red Edge 3、Nir、Red Edge 4、Water vapor、SWIR1、SWIR2	12
Sentinel - 2	植被指数	归一化植被指数(NDVI) ^[32] 、插值植被指数(DVI) ^[32] 、联合光谱响应指数(COSRI) ^[32] 、修改型土壤调节植被指数(MSAVI) ^[32] 、归一化插值植被指数(NDWI) ^[32] 、叶绿素指数-绿光(ciGREEN) ^[32] 、绿光归一化植被指数(GNDVI) ^[33] 、优化土壤调节植被指数(OSAVI) ^[33] 、土壤调节植被指数(SAVI) ^[33] 、可见光和短波红外干旱指数(VSDI) ^[34] 、增强植被指数(EVI) ^[35] 、修正比值植被指数(RDVI) ^[35] 、比值植被指数(RVI) ^[35] 、三角植被指数(TVI) ^[35] 、修正简单比率(MSR) ^[36] 、归一化差异指数(NDI) ^[37] 、拓展的归一化植被指数(ENDVI) ^[19] 、全球植被湿度指数(GVMI) ^[38] 、大气阻抗植被指数(ARVI) ^[38] 、短波红外水分胁迫指数(SIWIS) ^[39]	20
	盐分指数	盐分指数(SI4) ^[24] 、盐分指数(SI) ^[35] 、盐分指数(S2) ^[35] 、盐分指数(IBM_temp) ^[35] 、盐分指数(CAEX) ^[38] 、土壤盐分和碱度指数(SSSI - 1) ^[38] 、盐分指数(SAIO) ^[38] 、归一化盐分指数(NDSI) ^[38]	8
	缨帽变化因子	亮度指数(Brightness) ^[40] 、绿度指数(Greenness) ^[40] 、湿度指数(Wetness) ^[40]	3
	主成分分析波段	第一主成分(Pca1) ^[40] 、第二主成分(Pca2) ^[40] 、第三主成分(Pca3) ^[40]	3
	三波段指数	三波段指数(TBI3) ^[37] 、三波段指数(TBI4) ^[37] 、三波段指数(TBI7) ^[37]	3
Landsat - 8	地表温度	LST	1
NASA SRTM	地形因子	坡向(AS)、河网基准面(CNBL)、渠道网络距离(CND)、高程(DEM)、地形湿润度指数(RSP)、坡度(S)、地形湿润度指数(TWI)、山谷深度(VD)	8

1.6 筛选算法构建

采用相关性分析结合连续变量投影算法筛选反演因子。相关性分析采取皮尔逊相关分析,其指标为相关系数(R)。 R 取值范围为 $[-1,1]$, R 绝对值越大,说明这两个变量的相关性越强。连续变量投影算法是一种前向特征变量选择方法,采用向量的投影分析,它通过在光谱矩阵中任选一列向量并计算其在剩余变量中的最大投影变量,并将该向量作为下一轮的投影向量依次计算并循环直至达到所设定的波长数量 N 后结束,最后得到 N 组波段组合,分别建立多元线性回归模型,并选出各模型中均方根误差(RMSE)最小波段组合作为最优变量组合^[41]。

首先采用相关性分析对原始变量进行筛选,然后再利用连续变量投影算法确定最终建模的反演因子。裸土期、植被覆盖期分别剔除 R 介于 $[-0.30, 0.30]$ 、 $[-0.20, 0.20]$ 的变量(植被生长期由于覆盖的影响导致反演因子与土壤含盐量相关性较低,因而 R 的选取较低),利用 corrplot 包进行相关性分

析;连续变量投影算法采用 Matlab R2018a 软件实现。

1.7 预测模型与验证

1.7.1 预测模型

选取 Lasso、RF、LightGBM、XGBoost、1DCNNs、DNN 共 6 种模型进行土壤含盐量估算。其中 RF、LightGBM、XGBoost、Lasso 使用 Python 3.9.18 实现;1DCNNs、DNN 利用 Matlab R2018a 实现。

1.7.2 验证

将春季裸土期、植被覆盖期采集的土壤分别随机选取 30% 样本为测试集,其余样本为训练集。数据集划分如表 2、3 所示,训练集、测试集盐分分布接近,保证了样本具有代表性的同时,还在一定程度上避免了模型构建和验证中的偏差估计^[42]。

使用决定系数(R^2)、均方根误差(RMSE)和平均绝对误差(MAE)对反演模型进行精度评价。决定系数越接近 1,均方根误差和平均绝对误差越小,模型精度越高。

表 2 裸土期土壤含盐量数据集划分

Tab.2 Division of soil salinity dataset during bare soil period						
数据集	样本数	最小值/ (g·kg ⁻¹)	最大值/ (g·kg ⁻¹)	平均值/ (g·kg ⁻¹)	变异 系数	方差/ (g ² ·kg ⁻²)
全集	304	0.38	51.42	8.78	1.28	126.63
训练集	213	0.38	51.42	8.79	1.25	121.07
测试集	91	0.39	47.83	8.73	1.35	139.64

表 3 植被覆盖期土壤含盐量数据集划分

Tab.3 Division of soil salinity dataset during vegetation cover period						
数据集	样本数	最小值/ (g·kg ⁻¹)	最大值/ (g·kg ⁻¹)	平均值/ (g·kg ⁻¹)	变异 系数	方差/ (g ² ·kg ⁻²)
全集	178	0.63	44.72	6.62	1.61	113.91
训练集	124	0.63	44.72	6.52	1.60	108.97
测试集	54	0.63	41.34	6.85	1.63	125.18

2 结果与分析

2.1 实测土壤含盐量统计

研究区土壤盐分以氯化物为主,根据第三次全国土壤普查盐渍度分类分级标准^[43]将盐碱地划分为 5 个盐渍化等级,分级标准见表 4。参考此标准划分,研究区实测土壤盐渍化程度如图 3 所示,春季裸土期及植被覆盖期土壤含盐量为 0.38 ~ 51.42 g/kg、0.63 ~ 44.72 g/kg;平均值为 8.78、6.62 g/kg;土壤盐分时空变化较大,均属于强变异。土壤样本中非盐土占比较大,非盐土样本中以非盐碱土为主;盐碱化土壤样本中盐土居多。春季裸土期及植被覆盖期非盐碱土与盐土样本分别占比 45.7%、29.9% 和 57.3%、18.5%。

表 4 土壤盐渍化程度分级标准

Tab.4 Criteria for grading degree of soil salinization					
含盐量/(g·kg ⁻¹)	0 ~ 2	2 ~ 4	4 ~ 6	6 ~ 10	> 10
土壤盐渍化类型	非盐碱土	轻度盐碱土	中度盐碱土	重度盐碱土	盐土

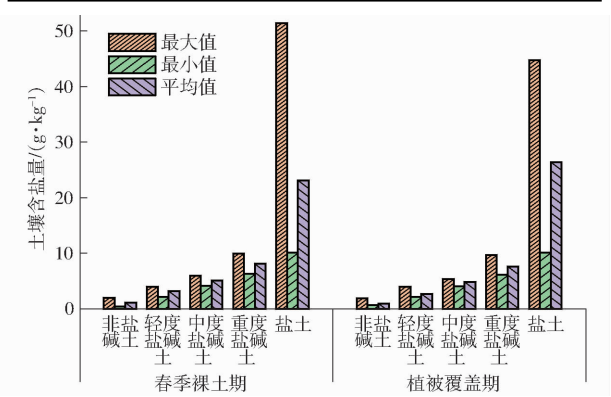


图 3 土壤含盐量统计

Fig.3 Soil salinity statistics

2.2 土壤含盐量与反演因子相关性

由图 4a 可知,土壤含盐量并未与 Sentinel-1 因子存在明显的相关性,本研究剔除此因子。

由图 4b 可知,春季裸土期土壤含盐量与短波红外波段 SWIR1、SWIR2 呈负相关,其余波段呈正相关。其中与 Aerosols 波段相关系数最大,为 0.27;其次为 Blue 波段,为 0.26;与 SWIR1、SWIR2 波段相关系数分别为 -0.13、-0.22。植被覆盖期的相关性低于裸土期,植被覆盖期土壤含盐量与可见光中的 Blue 波段相关性最大(0.24),其次为 Red 波段(0.22)。

由图 4c 可知,春季裸土期筛选出 SI4(0.38)、S1(0.42)、S2(0.42)、IBI_temp(-0.43)、CAEX(-0.48)、TBI3(0.36)、TBI4(0.33)、TBI7(0.40);植被覆盖期筛选出 SSSI-1(0.21)、SAIO(0.24)、CAEX(-0.31)、NDSI(0.23)。

由图 4d 可知,春季裸土期筛选出 GVM1(0.36)、VSDI(0.42)、SIWIS(-0.38)、NDI(-0.36);植被覆盖期筛选出 RVI(-0.20)、ENDVI(-0.25)、DVI(-0.21)、EVI(-0.20)、SAVI(-0.22)、OSAVI(-0.23)、NDVI(-0.23)、GNDVI(-0.22)、TVI(-0.24)、ARVI(-0.21)、ciGREEN(-0.22)、MSAVI(-0.22)、COSRI(0.21)、NDWI(0.23)、MSR(-0.21)、RDVI(-0.20)。

由图 4e 可知,地形因子筛选出 CNBL、CND、DEM、RSP、VD。春季裸土期及植被覆盖期相关系数分别为 -0.36、-0.30;-0.35、-0.26;-0.42、-0.34;-0.38、-0.29;0.47、0.35。

由图 4f 可知,裸土期土壤含盐量与 Wetness 相关系数为 0.38,植被覆盖期与 Wetness(0.36)、Pca3(0.33)、LST(-0.27)相关性较大。

综上,春季裸土期及植被覆盖期分别筛选出 18 种及 31 种反演因子。

2.3 敏感反演因子筛选

春季裸土期、植被覆盖期相关性筛选出 18、31 种变量,记为集合 A、B。对集合 A、B 使用连续变量投影算法(图 5),裸土期选出 10 种变量,分别为 RSP、CAEX、CND、SI4、Wetness、IBI_temp、TBI7、VD、S1、CNBL。植被覆盖期筛选出 GNDVI、RSP、TVI、CND、Wetness、VD、Pca3、CNBL、EDVI、CAEX、ENDVI、Blue、LST 共 13 种变量。

2.4 土壤含盐量估算模型对比与筛选

由表 5 可知,Lasso 模型土壤含盐量估算效果较差;RF、LightGBM、XGBoost、1DCNNs、DNN 模型相较

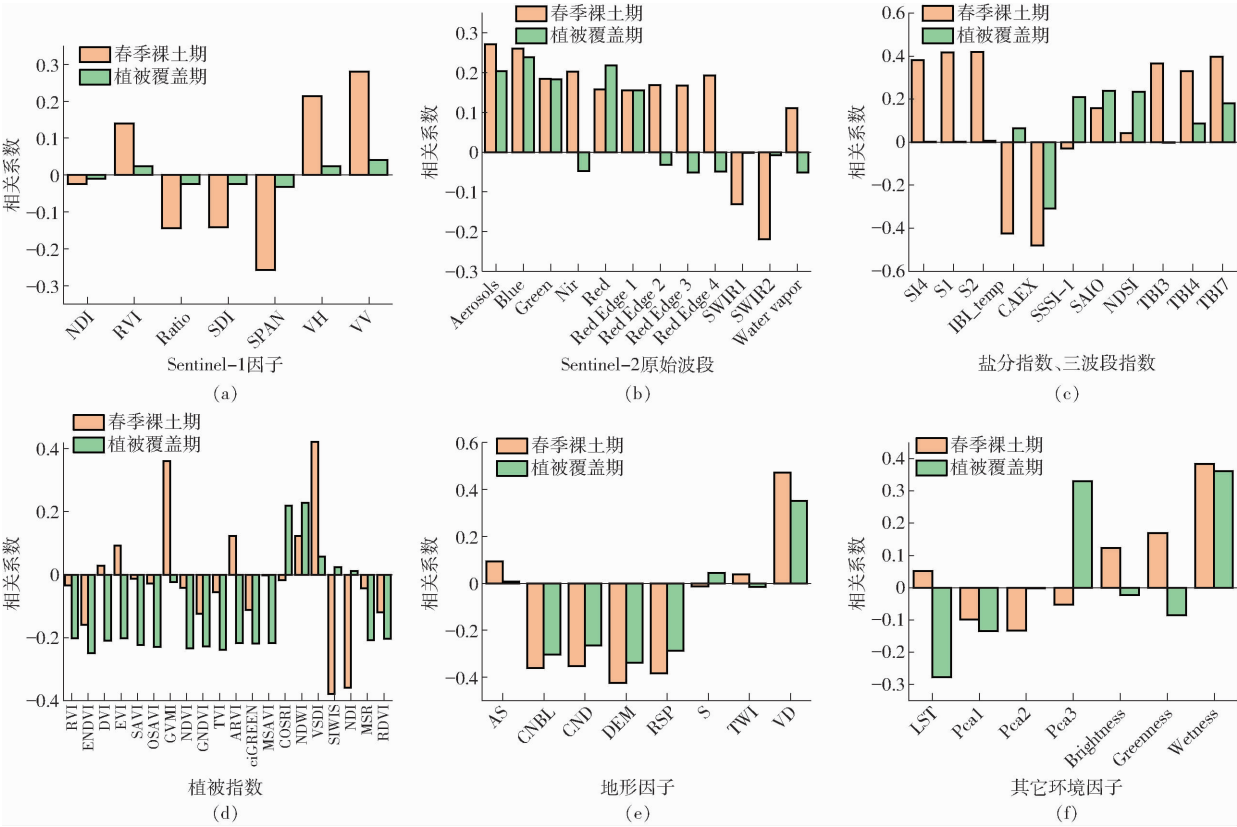


图 4 土壤含盐量与反演因子相关性分析

Fig. 4 Correlation analysis between soil salinity and inversion factors

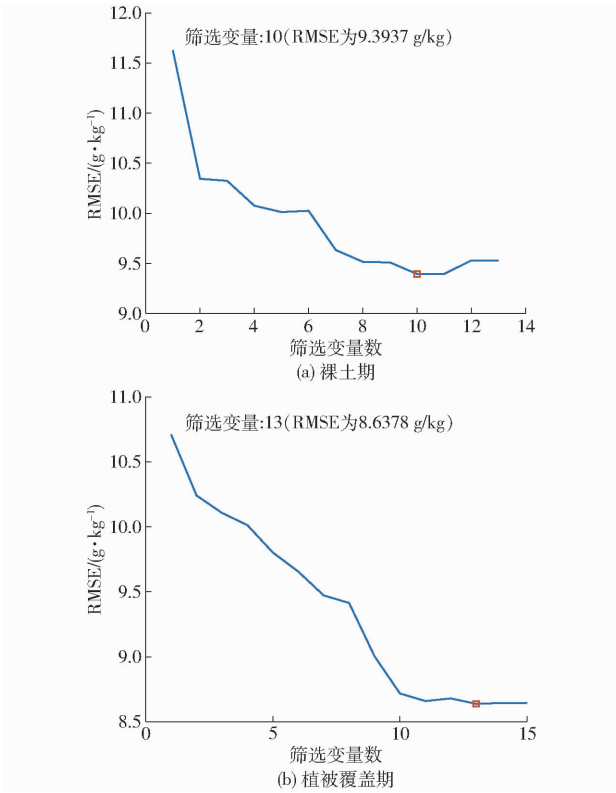


图 5 土壤含盐量反演因子 SPA 筛选结果分析
Fig. 5 Analysis of soil salinity inversion factor SPA screening results

表 5 春季裸土期土壤含盐量反演模型评价指标

Tab.5 Evaluation indexes of soil salinity inversion model during spring bare soil period

模型算法	训练集			测试集		
	R^2	RMSE/ ($\text{g}\cdot\text{kg}^{-1}$)	MAE/ ($\text{g}\cdot\text{kg}^{-1}$)	R^2	RMSE/ ($\text{g}\cdot\text{kg}^{-1}$)	MAE/ ($\text{g}\cdot\text{kg}^{-1}$)
Lasso	0.36	8.78	5.98	0.45	8.72	5.86
RF	0.73	5.71	3.31	0.65	6.98	3.96
LigtGBM	0.87	3.96	2.52	0.72	6.19	3.74
XGBoost	0.94	2.66	1.51	0.76	5.76	3.38
1DCNNs	0.89	3.60	2.22	0.60	7.53	5.06
DNN	0.86	4.13	2.70	0.75	5.86	3.88

测试集 R^2 、RMSE、MAE 分别为 0.45、8.72 g/kg、5.86 g/kg;XGBoost 模型最优,其测试集 R^2 、RMSE、MAE 分别为 0.76、5.76 g/kg、3.38 g/kg(图 6)。

由表 6 可知,植被覆盖期 Lasso 模型最差,测试集 R^2 、RMSE、MAE 分别为 0.32、9.20 g/kg、6.71 g/kg;XGBoost 模型最优,其测试集 R^2 、RMSE、MAE 分别为 0.58、7.22 g/kg、4.33 g/kg(图 7)。本研究对比分析 Lasso、RF、LightGBM、XGBoost、1DCNNs、DNN 共 6 种模型对于土壤含盐量的估算效果,结果表明对于春季裸土期、植被覆盖期 XGBoost 模型具有较好的估算能力,最终选取 XGBoost 模型作为研究区土壤含盐量估算模型。

而言可以较好地估算土壤含盐量,其中 XGBoost 模型效果最优。对于春季裸土期Lasso模型最差,其

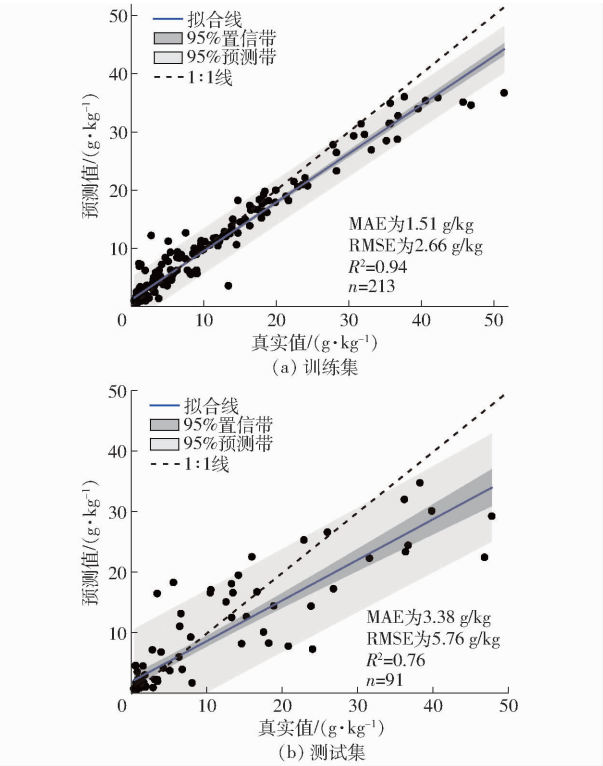


图6 XGBoost模型春季裸土期土壤含盐量反演精度验证

Fig.6 Validation of accuracy of soil salinity inversion during spring bare soil period for XGBoost modeling

表6 植被覆盖期土壤含盐量反演模型评价指标

Tab.6 Evaluation indexes of soil salinity inversion model during vegetation cover period

模型算法	训练集			测试集		
	R^2	RMSE/ ($\text{g}\cdot\text{kg}^{-1}$)	MAE/ ($\text{g}\cdot\text{kg}^{-1}$)	R^2	RMSE/ ($\text{g}\cdot\text{kg}^{-1}$)	MAE/ ($\text{g}\cdot\text{kg}^{-1}$)
Lasso	0.25	9.05	6.27	0.32	9.20	6.71
RF	0.77	4.99	3.03	0.48	8.05	4.97
LightGBM	0.76	5.05	3.33	0.55	7.53	5.05
XGBoost	0.88	3.54	2.15	0.58	7.22	4.33
1DCNNs	0.72	5.58	3.73	0.42	8.62	5.60
DNN	0.60	6.62	4.06	0.56	7.46	4.92

3 讨论

3.1 不同含盐量估算模型对比

本研究结果表明 XGBoost 模型优于 DNN、LightGBM、RF、1DCNNs、Lasso 模型,春季裸土期模型精度优于植被覆盖期。春季裸土期土壤裸露,卫星影像反射率直接反映了土壤盐分特征;植被覆盖期由于作物覆盖影响,土壤盐分只能通过反射率间接反映;并且由于作物覆盖导致植被覆盖期反演因子相关性低于裸土期;因而植被覆盖期预测精度较低。

虽然线性模型可以解释土壤含盐量与反演因子之间的部分关系,但是对数据分布和线性关系的要

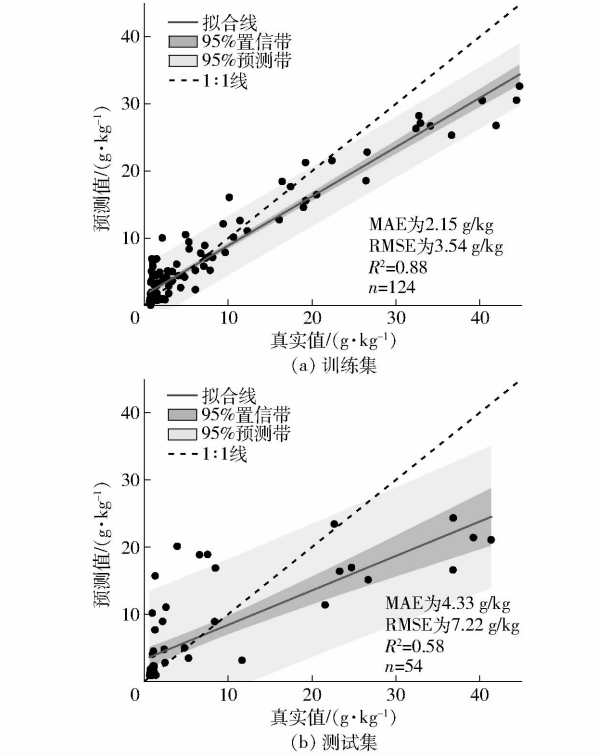


图7 XGBoost模型植被覆盖期土壤含盐量反演精度验证

Fig.7 Validation of accuracy of soil salinity inversion during vegetation cover period for XGBoost model

求限制了其使用^[39]。由于研究区土壤盐分分布高度离散,不符合正态分布,土壤盐分积累受到多种环境变量的影响,环境变量与含盐量的关系往往为非线性的^[23],因此用线性回归模型对研究区土壤含盐量进行估算并不适宜。本文选取机器学习模型中的3种集成学习模型,RF模型属于Bagging算法,XGBoost、LightGBM模型属于Boosting算法,其中XGBoost模型性能较优。ZHOU等^[44]基于全球土壤光谱库,比较了PLSR、Cubist、RF、XGBoost模型估算土壤含盐量精度,得出XGBoost模型的预测效果最优。ZAREI等^[45]研究表明,使用XGBoost、RF、GBM利用Sentinel-2图像进行土壤含盐量估算,XGBoost模型优于RF、GBM模型,而且指出XGBoost模型使用二阶泰勒展开来近似损失函数,提供了更多关于梯度方向以及如何达到损失函数最小值的信息;损失函数引入L1、L2正则化提高了泛化能力,训练速度非常快,具有可以跨批次并行处理等优点。深度学习引入了1DCNNs、DNN模型,结果表明DNN优于1DCNNs模型,但DNN模型精度低于XGBoost(机器学习模型)。机器学习模型已被证明是估算土壤含盐量的有效手段,并且在小样本数据集上表现良好^[46]。深度学习模型参数较多、并且需要大量样本进行训练^[25],由于研究区样本点较少,为了防止出现过拟合现象,训练次数不宜设置过多。因此1DCNNs模型没有在小样本数据中发挥优势;DNN

模型由于参数较少、训练速度较快,模型精度仅次于 XGBoost。

本研究只是进行线性模型、机器学习模型与简单深度学习模型对比,并未考虑改进之后的深度学习模型。MIAO 等^[47]对比分析 LSTM、LSTM - CNN 与 LWR 模型利用近红外光谱遥感反演有机质含量,结果表明由于 LSTM - CNN 可以有效提取空间和时间特征,因而其验证集精度最高 ($R^2 = 0.96$)。上述研究表明改进的深度学习模型用于土壤分析的可行性,在后续研究中应深入对比分析深度学习模型与机器学习模型对土壤性质的遥感反演精度。

3.2 变量重要性分析

土壤盐渍化受到多方面因素的共同影响,本研究采用多源遥感数据结合 PCC、SPA 筛选方法建立了研究区春季裸土期及植被覆盖期卫星尺度土壤含盐量估算模型。由于境内只有 Sentinel - 1A 影像,重访周期为 12 d,下载的 Sentinel - 1 影像时间为 2022 年 5 月 9 日及 8 月 1 日、2023 年 4 月 22 日及 9 月 13 日,由于部分影像时间与采样时间不一致、植被覆盖影响导致相关性较低,因而本研究并未考虑此因子。如图 8 所示,研究区裸土期土壤盐分的敏感因子主要为盐分指数 (占 48.3%),其次为地形因子、TBI7、Wetness,分别为 33.8%、14.5%、3.4%。植被覆盖期由于覆盖的影响含盐量反演精度较低,其主要反演因子为地形因子 (47.9%) 和植被指数 (22%),见图 9。盐分指数为盐分反演常用的一种指数^[13],许多学者依托盐分指数建立含盐量反演模型;地形直接影响了地下水的径流运动,进而间接影响了盐分的迁移及

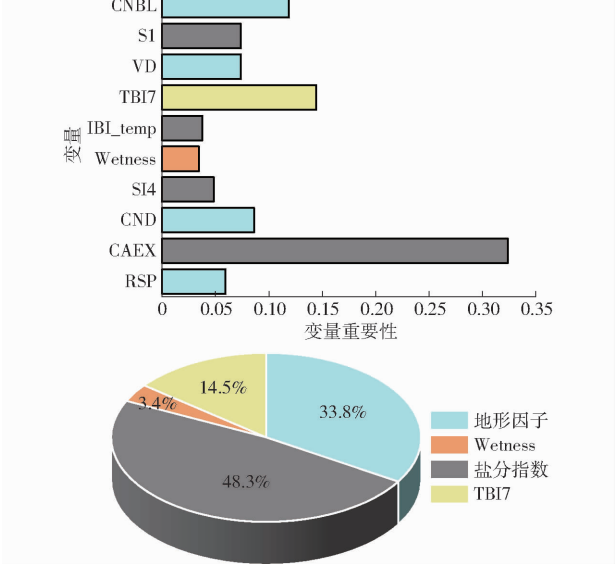


图 8 春季裸土期因子重要性评估

Fig. 8 Assessment of importance of factors during spring bare soil period

分布^[48];Wetness 反映土壤湿度,土壤水分是植被生长的重要因素,水分胁迫影响植物生长,当缺水严重致使植被生长稀疏及死亡,进而导致蒸发强烈,盐分开始聚集^[49];已有学者证明三波指数是反演土壤含盐量的重要因子^[37];植被指数反映植被生长情况和覆盖度,盐分抑制了植被生长,可以通过植被指数间接反映盐分的空间分布。

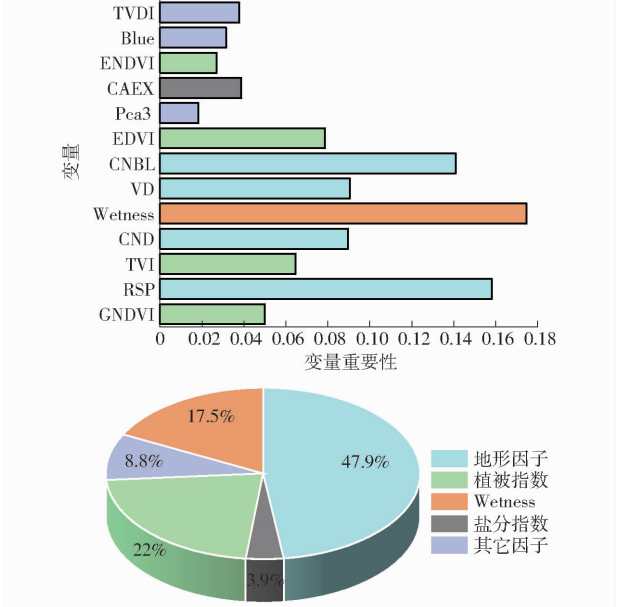


图 9 植被覆盖期因子重要性评估

Fig. 9 Assessment of importance of vegetation cover period factors

本研究得出春季裸土期重要反演因子为盐分指数和地形因子,植被覆盖期为植被指数及地形因子。WANG 等^[25]在新疆南部利用遥感数据、地形因子、盐分指数和植被指数开展土壤盐分研究,指出地形因子对土壤盐分的影响最大。蒙莉娜等^[17]指出 CRSI、VD、CNBL、DEM 和 Greenness 是 RF 模型含盐量估算中最重要变量;光学遥感影像数据对含盐量的解释能力最高,其次是地形指数。LI 等^[24]指出 Sentinel - 2 因子中盐分指数是主要反演因子,地形因子和雷达数据对土壤盐分影响不大;这是因为采样区域地势平缓,地形因子对土壤盐分分布影响较小。本文并未考虑降水量、地下水埋深、与河道距离等对土壤盐分时空变化影响,今后研究应当考虑多种全面的环境因子,从土壤盐分形成机理探讨盐分的空间分布。

3.3 土壤盐分空间分布

各类盐碱土面积统计如图 10 所示。选取 XGBoost 模型反演土壤含盐量,进而分析春季裸土期及植被覆盖期土壤盐分时空分布,如图 11 所示。结果表明,研究区土壤以非盐碱土为主,分布于南部,面积占比 50% 以上;轻度盐碱土主要零散分布

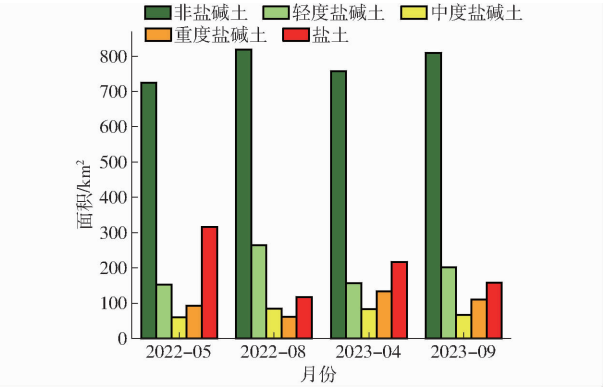


图 10 2022—2023 年研究区各类盐碱土面积统计
Fig. 10 Area statistics of various types of saline soils in study area in 2022—2023

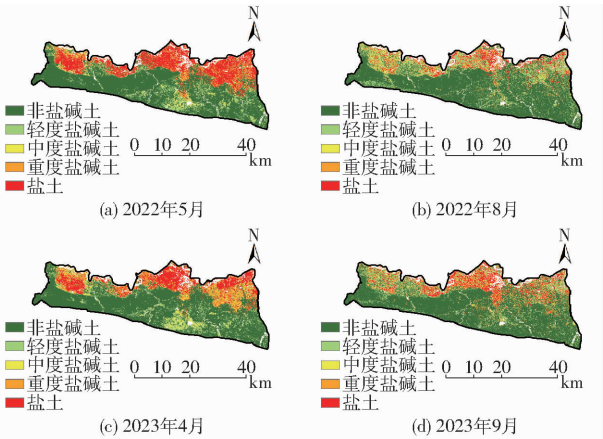


图 11 研究区土壤盐分空间分布图

Fig. 11 Spatial distribution of soil salinity in study area

于东、西部,面积占比 10% 以上;中、重度盐碱土占比较小,零散分布于研究区;盐土主要分布于北部,面积占比约 15%,与野外实际调查分析结果基本吻合。从空间分布来看,土壤盐碱化程度从南向北呈现逐渐增加趋势,靠近黄河一侧最为严重。由于黄河侧渗和凌汛影响导致地下水位高以及地势较低导致北部土壤盐碱化问题突出。春季裸土期土壤盐渍化面积大于植被覆盖期,由于裸土期气候干燥、蒸发强烈,盐分随着包气带、毛管水向上运动,最终聚集在表层,出现积盐现象;研究区雨量集中于 7、8 月,由于地下水位年中最低,有助于土壤脱盐,因此植被覆盖期盐渍化减轻,这一现象符合“盐随水来,水随气散,气散盐存”。

4 结论

- (1) XGBoost 模型为研究区最佳土壤含盐量估算模型,具有较高的精度(春季裸土期、植被覆盖期测试集 R^2 分别为 0.76、0.58;RMSE 为 5.76、7.22 g/kg)。
- (2) 春季裸土期重要反演因子为盐分指数和地形因子、植被覆盖期为植被指数及地形因子;春季裸土期土壤含盐量估算精度高于植被覆盖期。
- (3) 根据 XGBoost 模型反演估算得到 2022—2023 年春季裸土期及植被覆盖期盐分空间分布,盐碱土主要分布于研究区北部,南部土壤盐碱化相对较轻。

参 考 文 献

[1] 杨泽龙,苗百岭,郑晓艳. 内蒙古达拉特旗黄河灌区春季返盐气象条件分析[J]. 中国农学通报,2019,35(6):70-76.
YANG Zelong, MIAO Bailing, ZHENG Xiaoyan. Meteorological condition of slat accumulation in spring of Yellow River irrigational area in Dalate Banner, Inner Mongolia[J]. Chinese Agricultural Science Bulletin, 2019, 35(6): 70-76. (in Chinese)

[2] 李平,牛燕冰,杜占春,等. 达拉特旗盐碱地改良调查报告[J]. 现代农业,2019(4):48-51.

[3] 魏韬书,牛洁,景宇鹏,等. 内蒙古达拉特旗黄河沿岸土壤盐渍化特征研究[J]. 内蒙古农业大学学报(自然科学版),2016,37(2):55-61.
WEI Taoshu, NIU Jie, JING Yupeng, et al. The principal component analysis of salinized in Dalate of Inner Mingolia[J]. Journal of Inner Mongolia Agricultural University(Natural Science Edition),2016,37(2):55-61. (in Chinese)

[4] BUTCHER K, WICK A F, DESUTTER T, et al. Soil salinity: a threat to global food security[J]. Agronomy Journal, 2016, 108(6): 2189-2200.

[5] 刘旭辉,白云岗,柴仲平,等. 基于多光谱遥感的典型绿洲棉田春季土壤盐分反演及验证[J]. 干旱区地理,2022,45(4):1165-1175.
LIU Xuhui, BAI Yungang, CHAI Zhongping, et al. Inversion and validation of salinity based on multispectral remote sensing in typical oasis cotton field in spring[J]. Arid Land Geography, 2022, 45(4): 1165-1175. (in Chinese)

[6] TIBSHIRANI R. Regression shrinkage and selection via the lasso[J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 1996, 58(1): 267-288.

[7] MA Y, TASHPOLAT N. Current status and development trend of soil salinity monitoring research in China[J]. Sustainability, 2023, 15(7): 5874.

[8] WANG L, HU P, ZHENG H, et al. Integrative modeling of heterogeneous soil salinity using sparse ground samples and remote sensing images[J]. Geoderma, 2023, 430: 116321.

[9] 韩文霆,崔家伟,崔欣,等. 基于特征优选与机器学习的农田土壤含盐量估算研究[J]. 农业机械学报,2023,54(3):328-337.
HAN Wenting, CUI Jiawei, CUI Xin, et al. Estimation of farmland soil salinity content based on feature optimization and

- machine learning algorithms[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(3): 328–337. (in Chinese)
- [10] ALLBED A, KUMAR L, ALDAKHEEL Y Y. Assessing soil salinity using soil salinity and vegetation indices derived from IKONOS high-spatial resolution imageries: applications in a date palm dominated region[J]. Geoderma, 2014, 230: 1–8.
- [11] EL H A, LHISSOU R, CHOKMANI K, et al. Spatiotemporal monitoring of soil salinization in irrigated Tadla Plain (Morocco) using satellite spectral indices[J]. International Journal of Applied Earth Observation and Geoinformation, 2016, 50: 64–73.
- [12] WU W, AL-SHAFIE W M, MHAIMEED A S, et al. Soil salinity mapping by multiscale remote sensing in Mesopotamia, Iraq [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 7(11): 4442–4452.
- [13] 李志刚, 许强, 赵宽耀, 等. 基于高分一号卫星影像的延安市顾电流域耕地盐渍化定量反演[J]. 科学技术与工程, 2021, 21(13): 5228–5235.
- LI Zhigang, XU Qiang, ZHAO Kuanyao, et al. Quantitative inversion of salinization of cultivated land in Gutun watershed of Yan'an City based on Gaofen-1 satellite image[J]. Science Technology and Engineering, 2021, 21(13): 5228–5235. (in Chinese)
- [14] RAMOS T B, CASTANHEIRA N, OLIVEIRA A R, et al. Soil salinity assessment using vegetation indices derived from Sentinel-2 multispectral data. Application to Lezíria Grande, Portugal[J]. Agricultural Water Management, 2020, 241: 106387.
- [15] 弓永利. 基于微波遥感的裸露地表土壤盐分含量的反演[J]. 江苏农业科学, 2015, 43(11): 442–444.
- [16] TAGHADOSI M M, HASANLOU M, EFTEKHARI K. Soil salinity mapping using dual-polarized SAR Sentinel-1 imagery [J]. International Journal of Remote Sensing, 2019, 40(1): 237–252.
- [17] 蒙莉娜, 丁建丽, 王敬哲, 等. 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 Li'na, DING Jianli, WANG Jingzhe, et al. Spatial distribution of soil salinity in Ugan-Kuqa River delta oasis based on environmental variables[J]. Transactions of the CSAE, 2019, 36(1): 175–181. (in Chinese)
- [18] 李春泽, 张超, 张皓源, 等. 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(2): 1–10.
- LI Chunze, ZHANG Chao, ZHANG Haoyuan, et al. Inversion of soil salinization in western Jilin Province based on XGBoost and geographically weighted regression[J]. Journal of China Agricultural University, 2024, 29(2): 1–10. (in Chinese)
- [19] 杨丽萍, 任杰, 王宇, 等. 基于多源遥感数据的居延泽地区土壤盐分估算模型[J]. 农业机械学报, 2022, 53(11): 226–235.
- YANG Liping, REN Jie, WANG Yu, et al. Soil salinity estimation model in Juyanze based on multi-source remote sensing data[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(11): 226–235. (in Chinese)
- [20] WANG J, ZHEN J, HU W, et al. Remote sensing of soil degradation: progress and perspective[J]. International Soil and Water Conservation Research, 2023, 11(3): 429–454.
- [21] 姜传礼, 赵健赞, 丁圆圆, 等. SPA 算法与机器学习的黄河源土壤水分反演[J]. 光谱学与光谱分析, 2023, 43(6): 1961–1967.
- JIANG Chuanli, ZHAO Jianyun, DING Yuanyuan, et al. Study on soil water retrieval technology of Yellow River source based on SPA algorithm and machine learning[J]. Spectroscopy and Spectral Analysis, 2023, 43(6): 1961–1967. (in Chinese)
- [22] 许童羽, 陆佳欣, 白驹驰, 等. 基于无人机偏振遥感的水稻冠层氮素含量反演模型[J]. 农业机械学报, 2023, 54(10): 171–178.
- XU Tongyu, YANG Jiaxin, BAI Juchi, et al. Inversion model of nitrogen content of rice canopy based on UAV remote sensing [J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(10): 171–178. (in Chinese)
- [23] PENG J, BIASWAS A, JIANG Q, et al. Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China[J]. Geoderma, 2019, 337: 1309–1319.
- [24] LI J, ZHANG T, SHAO Y, et al. Comparing machine learning algorithms for soil salinity mapping using topographic factors and Sentinel-1/2 data: a case study in the Yellow River delta of China[J]. Remote Sensing, 2023, 15(9): 2332.
- [25] WANG N, XUE J, PENG J, et al. Integrating remote sensing and landscape characteristics to estimate soil salinity using machine learning methods: a case study from Southern Xinjiang, China[J]. Remote Sensing, 2020, 12(24): 4118.
- [26] LEE S J, CHOI C, KIM J, et al. Estimation of high-resolution soil moisture in Canadian croplands using deep neural network with Sentinel-1 and Sentinel-2 images[J]. Remote Sensing, 2023, 15(16): 4063.
- [27] 姬宝霖, 吕忠义, 申向东, 等. 内蒙古达拉特旗十大孔兑综合治理方案研究[J]. 人民黄河, 2004(1): 31–32, 36.
- [28] 王宇璇, 屈忠义, 白燕英, 等. 基于 Sentinel-2 影像的黄河南岸典型改良示范区土壤含盐量反演模型[J]. 农业机械学报, 2024, 55(4): 290–299, 439.
- WANG Yuxuan, QU Zhongyi, BAI Yanying, et al. Soil salt inversion of typical improvement demonstration area of south bank of Yellow River based on Sentinel-2 images[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(4): 290–299, 439. (in Chinese)
- [29] POORTINGA A, TENNESON K, SHAPIRO A, et al. Mapping plantations in Myanmar by fusing Landsat-8, Sentinel-2 and Sentinel-1 data along with systematic error quantification[J]. Remote Sensing, 2019, 11(7): 831.
- [30] LEE J S, WEN J H, AINSWORTH T L, et al. Improved Sigma filter for speckle filtering of SAR imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2008, 47(1): 202–213.

[31] ZHANG Q, LI L, SUN R, et al. Retrieval of the soil salinity from Sentinel-1 dual-polarized SAR data based on deep neural network regression[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 19: 1-5.

[32] 王银凤. 基于无人机多光谱遥感的土壤含水率反演模型研究[D]. 兰州:兰州理工大学,2023.
WANG Yinfeng. Research on soil moisture content inversion model based on UAV multi-spectral remote sensing[D]. Lanzhou: Lanzhou University of Science and Technology,2023. (in Chinese)

[33] DENG J, PAN S, ZHOU M, et al. Optimum sampling window size and vegetation index selection for low-altitude multispectral estimation of root soil moisture content for Xuxiang kiwifruit[J]. Agricultural Water Management, 2023, 282: 108297.

[34] WANG S, WANG W, WU Y, et al. Surface soil moisture inversion and distribution based on spatio-temporal fusion of MODIS and Landsat[J]. Sustainability, 2022, 14(16): 9905.

[35] GUO Z, LI Y, WANG X, et al. Remote sensing of soil organic carbon at regional scale based on deep learning: a case study of agro-pastoral ecotone in northern China[J]. Remote Sensing, 2023, 15(15): 3846.

[36] ZHOU J, XU Y, GU X, et al. High-precision mapping of soil organic matter based on UAV imagery using machine learning algorithms[J]. Drones, 2023, 7(5): 290.

[37] WANG J, DING J, YU D, et al. Capability of Sentinel-2 MSI data for monitoring and mapping of soil salinity in dry and wet seasons in the Ebinur Lake region, Xinjiang, China[J]. Geoderma, 2019, 353: 172-187.

[38] WANG N, PENG J, XUE J, et al. A framework for determining the total salt content of soil profiles using time-series Sentinel-2 images and a random forest-temporal convolution network[J]. Geoderma, 2022, 409: 115656.

[39] GE X, DING J, TENG D, et al. Updated soil salinity with fine spatial resolution and high accuracy: the synergy of Sentinel-2 MSI, environmental covariates and hybrid machine learning approaches[J]. Catena, 2022, 212: 106054.

[40] 王飞,杨胜天,丁建丽,等. 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 Fei, YANG Shengtian, DING Jianli, et al. Environmental sensitive variable optimization and machine learning algorithm using in soil salt prediction at oasis[J]. Transactions of the CSAE,2018,34(22):102-110. (in Chinese)

[41] 章海亮,罗微,刘雪梅,等. 应用遗传算法结合连续投影算法近红外光谱检测土壤有机质研究[J]. 光谱学与光谱分析, 2017,37(2):584-587.
ZHANG Hailiang, LUO Wei, LIU Xuemei, et al. Measurement of soil organic matter with near infrared spectroscopy combined with genetic algorithm and successive projection algorithm[J]. Spectroscopy and Spectral Analysis, 2017, 37(2): 584-587. (in Chinese)

[42] 杨练兵,郑宏伟,罗格平,等. 基于遗传算法优化 BP 神经网络的土壤盐渍化反演[J]. 地理与地理信息科学,2021, 37(2):12-21,37.
YANG Lianbing, ZHENG Hongwei, LUO Geping, et al. Retrieval of soil salinity content based on BP network optimized by genetic algorithm[J]. Geography and Geo-Information Science, 2021, 37(2):12-21,37. (in Chinese)

[43] 国务院第三次全国土壤普查领导小组办公室. 第三次全国土壤普查土壤类型名称校准技术规范(修订版)[EB/OL] (2023-02-27)[2024-07-05]. https://www.moa.gov.cn/ztlz/dscqgtrpc/zywj/202307/t20230720_6432535.htm

[44] ZHOU Y, CHEN S, HU B, et al. Global soil salinity prediction by open soil Vis-NIR spectral library[J]. Remote Sensing, 2022, 14(21): 5627.

[45] ZAREI A, HASANLOU M, MAHDIANPARI M. A comparison of machine learning models for soil salinity estimation using multi-spectral earth observation data[J]. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2021, 3: 257-263.

[46] GORJI T, SERTEL E, TANIK A. Monitoring soil salinity via remote sensing technology under data scarce conditions: a case study from Turkey[J]. Ecological Indicators, 2017, 74: 384-391.

[47] MIAO T, JI W, LI B, et al. Advanced soil organic matter prediction with a regional soil NIR spectral library using long short-term memory-convolutional neural networks: a case study[J]. Remote Sensing, 2024, 16(7): 1256.

[48] MCBRATNEY A B, SANTOS M L M, MINASNY B. On digital soil mapping[J]. Geoderma, 2003, 117(1-2): 3-52.

[49] XU H, CHEN C, ZHENG H, et al. 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.