

doi:10.6041/j.issn.1000-1298.2021.09.035

基于机器学习的绿洲土壤盐渍化尺度效应研究

陈香月^{1,2} 丁建丽^{1,2} 葛翔宇^{1,2} 王飞^{1,2} 王敬哲³

(1. 新疆大学智慧城市与环境建模自治区普通高校重点实验室, 乌鲁木齐 830046;
2. 新疆大学绿洲生态教育部重点实验室, 乌鲁木齐 830046;
3. 深圳大学自然资源部大湾区地理环境监测重点实验室, 深圳 518060)

摘要: 针对干旱区绿洲土壤盐渍化的生态环境问题,以新疆维吾尔自治区奇台绿洲为研究区,基于 58 个表层土壤盐度数据及与之对应的 Landsat TM 多光谱遥感影像数据,分别选取栅格重采样(空间分辨率为 30 ~ 990 m)和邻域滤波(窗口尺度为 3 × 3、5 × 5、…、31 × 31)两种尺度转换方法获取不同尺度下 Landsat TM 派生数据,并据此计算相应的环境变量(总数为 720);随后利用梯度提升决策树(GBDT)模型在不同尺度下依托环境变量对土壤盐度进行模拟,并分析其定量关系。结果表明:单一尺度下,基于 30 m 空间分辨率的邻域滤波方法对土壤盐度的解析力总体优于栅格重采样模式,其最大解析力分别为 78.55%、75.31%。混合多种尺度下,对土壤盐度的解析效果较单一尺度得到明显提升,解析力最高可达 90.66%,有效实现了信息互补。栅格重采样模式相对于邻域滤波而言,其调整 R^2 波动范围更为宽泛,说明栅格重采样尺度变换方法相较于邻域滤波对土壤盐度-环境变量关系的表征更具灵敏性。

关键词: 土壤盐渍化; 遥感; 尺度效应; 环境变量; 邻域滤波

中图分类号: K909; S1 文献标识码: A 文章编号: 1000-1298(2021)09-0312-09 **OSID:** 

Scale Effect on Soil Salinization Simulation in Arid Oasis Based on Machine Learning Methods

CHEN Xiangyue^{1,2} DING Jianli^{1,2} GE Xiangyu^{1,2} WANG Fei^{1,2} WANG Jingzhe³

(1. Key Laboratory of Smart City and Environment Modelling of Higher Education Institute, Xinjiang University, Urumqi 800046, China
2. Key Laboratory of Oasis Ecology, Xinjiang University, Urumqi 830046, China
3. MNR Key Laboratory for Geo-Environmental Monitoring of Great Bay Area, Shenzhen University, Shenzhen 518060, China)

Abstract: Soil salinity is one of the crucial factors which affects eco-environmental quality in the oasis of arid regions. Consequently, there is a great need to monitor soil salinity for prevention and mitigation of land degradation and further promote regional sustainable development. The variation in soil salinity is affected by environmental factors that occur at different scales with varying intensities. It is critical to adequately consider environmental variables under scale effects for digital soil mapping which has been minimally discussed in previous studies. Totally 58 soil samples were collected from the Qitai Oasis, Xinjiang Uygur Autonomous Region of China. In the laboratory, the soil samples were prepared for analysis of electrical conductivity (EC) when prepared into suspensions 1:5 in soil and distilled water ratio. In addition, the corresponding Landsat - 5 TM data was collected and preprocessed for up-scale transformation by raster resampling (spatial resolution were 30 ~ 990 m) and neighborhood filtering (window size were 3 × 3, 5 × 5, ..., 31 × 31), and the environmental variables (vegetation index (VI), normalized difference infrared index (NDII), principal component analysis (PCA), and tasseled cap transformation (TC)) were further generated. Then, the gradient boosting decision tree (GBDT) model was employed for the estimation of surface soil salinity based on these 720 environmental variables at various spatial scales. The results showed that for individual scale mode, the neighborhood filtering method based on 30 m pretreated data was generally better than those of the raster resampling modes, and

收稿日期: 2020-09-29 修回日期: 2020-11-29
基金项目: 新疆维吾尔自治区自然科学基金项目(2021D01D06)、国家自然科学基金项目(41771470、41661046)和中国博士后科学基金项目(2020M672776)
作者简介: 陈香月(1994—),女,博士生,主要从事干旱区环境遥感研究,E-mail: cxy_1994@126.com
通信作者: 丁建丽(1974—),男,教授,主要从事干旱区资源与环境遥感应用研究,E-mail: watarid@xju.edu.cn

the maximum analytical power reached 78.55% and 75.31%, respectively. In terms of mixed scale, the analytical effect of soil salinity was significantly improved compared with the mode of individual scale, and the analytical power could reach up to 90.66%, which suggested the effective information complementarity. Compared with the neighborhood filtering, the range of adjusts R^2 of the resampling mode was broader, which indicated that the scale transformation of grid resampling was more sensitive to the characterization of the relationship between soil salinity and environmental variables. The research result was helpful for understanding specific scale-dependent relationships and had the potential to reveal the scale control of soil salinity variation in arid regions.

Key words: soil salinization; remote sensing; scale effect; environment variables; neighborhood filtering

0 引言

土壤盐渍化指水溶性盐分向土壤表面迁移和聚集的过程,是自然因素和不合理的人类活动共同作用影响下的产物^[1-3]。在人口快速增长背景下,现有土地资源难以继续承载更多的人口生产生活所需,合理开发利用后备耕地资源,对缓解土地资源稀缺绿洲地区的人地矛盾尤为关键,也成为近年来新的研究热点。如何准确、快速、廉价获取盐渍化信息是科学合理开发盐渍土地资源的重要依据。近年来,遥感技术因其具有回访周期快、覆盖面积广、成本低、获取方便,且非侵入土壤等优势,已成为土壤盐渍化信息数字制图研究中广泛应用的有效工具^[4]。

地表的异质性意味着就单一尺度研究地学问题具有一定局限性,通常基于特定尺度的研究结果较难直接推广至其他尺度,尺度上的细微变化都将引出新的科学问题^[5-6]。尺度问题研究的根本目的是挖掘时空尺度特征信息,从而找到与研究对象的观测尺度相匹配的本征时空尺度。土壤盐渍化往往具有极强的时空变异性,受各类环境要素影响极大,对土壤盐渍化现象发生发展的深刻理解都离不开尺度问题这一环节^[7-8]。由于遥感影像具有多分辨率(多尺度)特性,不同尺度在遥感数据中所呈现的地物信息丰富度存在显著差异,通常某些规律往往在特定的尺度下才被发现,尺度的不匹配将会降低土壤盐度模拟效果,故探寻最优研究尺度对于干旱区绿洲土壤盐渍化研究具有重要意义。

在土壤研究中,基于遥感数据计算的环境变量常被用于预测土壤属性的空间分布。盐分含量及其组成成分是影响盐渍化土壤光谱特征的主要因素,已有研究表明盐分含量增加会引起反射率增加^[9-10]。植被作为土壤盐渍化程度的间接指示器,其光谱特性是各种生态环境因子的综合表达,而为了有效理解这种表达,需要在合适的尺度下开展研究^[11-12]。目前,已有部分学者对土壤属性与遥感数据尺度(栅格单元大小)之间的关系进行了初步分

析^[13-14],并在此基础上对其衍生的生物物理过程开展相关研究^[15-16],但基于植被光谱特征信息间接推理土壤属性研究多停留在单一尺度上,对不同尺度下环境变量对土壤属性的解析尚不够深入。而忽略尺度效应可能会导致难以充分理解“土壤-环境”间的复杂关系,更会限制土壤盐渍化遥感模型的性能^[17]。

鉴于此,本研究针对土壤盐渍化这一典型土壤退化过程开展尺度效应研究,以新疆北部典型绿洲——奇台绿洲为研究区,利用遥感衍生数据与土壤表层实测电导率数据,定量描述土壤盐度与环境变量关系之间的尺度影响,结合机器学习算法通过多环境变量对土壤盐度分尺度进行模拟,并获取不同尺度下各环境变量对土壤盐度的解析能力,以期找出最优响应尺度,充分理解不同尺度下环境变量与土壤盐度之间的复杂关系,对绿洲土壤盐渍化状况进行有效评估,为干旱区绿洲农业可持续发展提供科学参考与数据支持。

1 数据与方法

1.1 研究区概况

奇台绿洲位于新疆维吾尔自治区昌吉回族自治州奇台县境内,地处新疆维吾尔自治区东北部,博格达山北麓,天山山脉东侧。研究区(43.95°~44.20°N, 89.20°~90.00°E,图1)位于奇台县平原地区,属温带大陆性干旱半干旱气候,夏季高温少雨,日温差、年温差较大,多年平均降水量为211.9 mm,多年平均蒸发量为1735.7 mm,蒸降比高达8:1,强烈的蒸散发作用及水资源的不合理利用导致奇台县内广泛存在土壤盐碱化问题。全区约有31%农业土地存在盐碱化现象,严重危害到当地的生态环境与农业生产生活^[18]。

1.2 采样数据

本研究利用的样本采样时间为2009年6月,样本量为58个,样本主要集中在绿洲平原区,主要的土地利用与覆盖类型包括:农田、荒漠、灌丛、稀疏草地、草地等。以梅花五点混合采样方法采取表层土

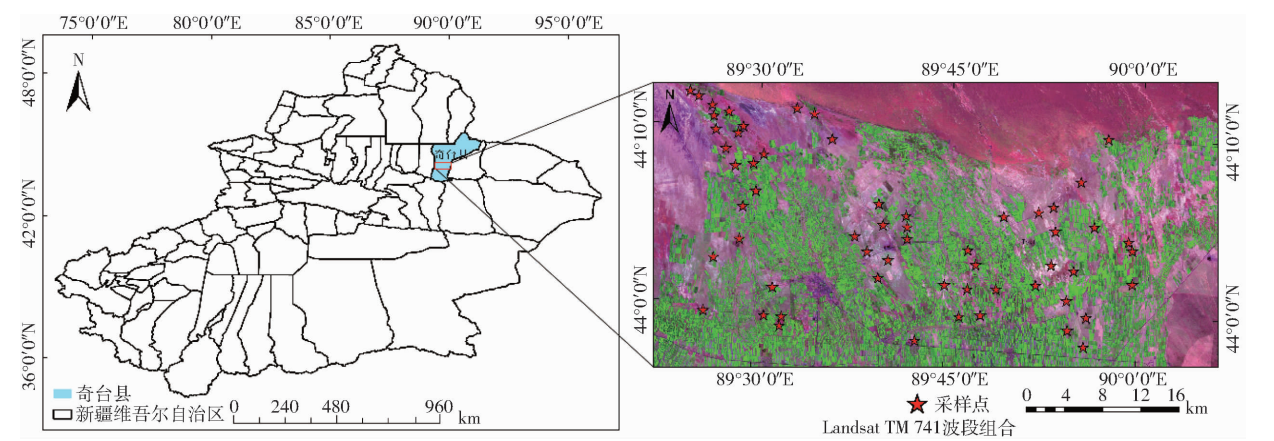


图 1 研究区示意图
Fig. 1 Location of study area

壤(0~20 cm)约 500 g,同时记录样品相应的地理坐标,标记后用自封袋封装带回实验室。采集得到的样本在室内剔除杂质晾干后进行研磨与过筛处理(2 mm),采用土水质量比 1:5 制备土壤悬浊液,利用土壤溶液电导率仪对制备的浸提取液进行电导率(dS/m)测定^[17]。

1.3 遥感数据

Landsat TM 是搭载于 Landsat 4/5 卫星上的光学成像传感器,提供了大量中分辨率和长时间序列的遥感数据。其光谱范围从可见光波长(0.45 μm)到中红外波长(12.5 μm)间共 7 个波段,重访周期为 16 d,空间分辨率为 30 m。从美国 USGS 官网下载 2009 年 6 月过境的 Landsat TM 影像,行列号为 141/29,以完全覆盖所有采样点且云量覆盖低于 10% 为准。通过 ENVI 软件对数据进行辐射定标、大气校正,最后将辐射率转为研究需要的地表反射率以进行后续分析。

1.4 尺度转换

地理空间数据多具有尺度依赖性,尺度效应被认为是地观测中不可忽略的问题,严重制约了遥感、地信等学科的发展,故在研究中考虑尺度问题非常必要^[19-20]。魏阳等^[21]研究表明,就干旱区绿洲土壤而言,在地面分辨率 1 km 左右开展土壤盐度尺度效应研究是有效的。基于此,本研究采用栅格重采样和邻域滤波两种升尺度方法,将最大分辨率控制在 1 km 左右,以此为阈值对影像进行升尺度转换。栅格重采样方法选用最邻近法,基于像元对影像进行重采样计算,邻域滤波方法主要是基于邻域窗口的选择,在特定空间分辨率下按照 $n \times n$ 移动窗口分别计算 3×3 、 5×5 、 $\dots$ 、 31×31 多个邻域滤波,其中窗口尺度为奇数,数值间隔为 2,见图 2。具体而言,利用 ENVI 软件中的重采样模块(Resize data)将空间分辨率为 30 m 的原始遥感数据进行栅格重

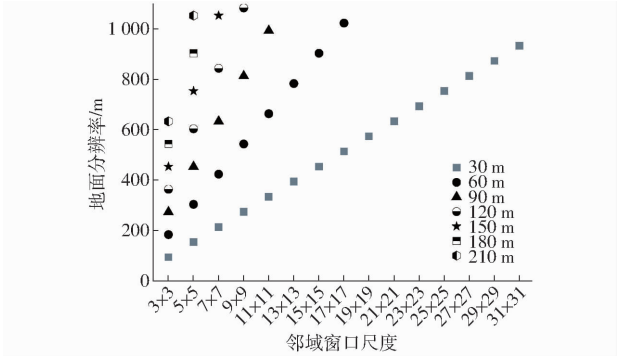


图 2 不同空间分辨率下邻域滤波尺度变换结果
Fig. 2 Transformation of neighborhood filtering scale in different spatial resolutions

采样处理,研究以 30 m 为步长将原始遥感数据依次重采样至 990 m。在邻域滤波计算中,选择了 7 种原始分辨率,分别为 30、60、90、120、150、180、210 m,邻域滤波窗口尺度选择了 3×3 、 5×5 、 $\dots$ 、 31×31 共计 15 个级别,共对应 39 个尺度,利用 ENVI 中的邻域滤波模块(Occurrence measures)实现。

1.5 环境变量

基于尺度变换后的遥感影像,计算应用于土壤盐度评价的植被指数和归一化红外指数^[17]以及图像衍生变换参量(表 1)。植被与土壤关系密不可分,作为地表变化的直接观测因子,其长势能够间接反映土壤盐渍化程度^[22-23]。且已有研究表明,相较于单纯利用土壤信息进行盐度预测,植被光谱信息的引入可以有效改善预测效果^[24]。其中,归一化植被指数(NDVI)、扩展归一化植被指数(ENDVI)以及扩展增强性植被指数(EEVI)已经被证实可以间接用以判断土壤盐度^[25-26]。归一化红外指数(NDII),亦称土壤湿度/水分含量指数,通过融合 Landsat TM 的短波红外波段信息,使得 NDII 对冠层湿度与土壤盐度具有一定指示作用。图像变换衍生因子某种程度上可以对原始波段信息进行高度凝练,对海量数据进行有效降维,

避免维数灾难。其中主成分分析 (Principal components analysis, PCA) 通过生成互不相关的输出波段, 研究表明主成分分析的前 3 个波段 (PCA1、PCA2、PCA3) 占数据整体信息高达 90% 以上, 达到隔离噪声和减少数据集的维数目的; 缨帽变换 (Tasseled cap transformation, TC) 是一种特殊

的主成分分析, 该变换将地表信息有效地转换成 3 个主题因子, 分别为亮度 (TCB)、植被绿度 (TCG) 和土壤湿度 (TCW), 可近似作为土壤属性的相关指示因子。通过上述两种尺度转换方法共产生 720 个环境变量用于耦合多尺度和机器学习算法的土壤盐度估算。

表 1 选取的环境变量

环境变量	特征说明	文献序号
归一化植被指数 (Normalized difference vegetation index, NDVI)	$NDVI = (NIR - R) / (NIR + R)$	[17]
扩展归一化植被指数 (Extended normalized difference vegetation index, ENDVI)	$ENDVI = (NIR + SWIR_2 - R) / (NIR + SWIR_2 + R)$	[21]
扩展增强性植被指数 (Extended enhanced vegetation index, EEVI)	$EEVI = 2.5(NIR - SWIR_1 - R) / (NIR + SWIR_1 + 6R - 7.5B + 1)$	[17]
归一化红外指数 (Normalized difference infrared index, NDII)	$NDII = (NIR - SWIR_1) / (NIR + SWIR_1)$	[22]
主成分分析 (Principal components analysis, PCA)	第 1 主成分, 第 2 主成分, 第 3 主成分	
缨帽变换 (Tasseled cap transformation, TC)	植被绿度, 土壤湿度, 亮度	

注: B 表示蓝光波段, R 表示红光波段, NIR 表示近红波段, $SWIR$ 表示短波红外波段 ($SWIR_1$ 、 $SWIR_2$ 对应波长 $1.55 \sim 1.75 \mu m$ 、 $2.08 \sim 3.35 \mu m$)。

1.6 梯度提升决策树

梯度提升决策树 (Gradient boosting decision tree, GBDT) [27-28] 是一种由多棵决策树组成的集成学习方法, 它采用 Boosting 思想将多个弱分类器组成一个强分类器。该算法的基本原理是通过迭代计算, 不断减小模型的残差, 在梯度方向上建立一个新的模型来代替旧的模型 [29], 并在构建新模型的过程可以迭代地生成基本分类器的组合, 从而将损失函数最小化。作为典型的集成算法, GBDT 减少模型在建模过程中梯度方向上的残差, 相较于传统的机器学习模型具有更好的预测能力 [30], 不仅能够灵活地处理各类离散数据, 同时无需过分调节超参数, 更

为重要的是, GBDT 对输入变量的数量及其自相关不敏感, 还可以使用一些较为稳健的损失函数, 对异常值具有较好的鲁棒性 [31-32]。针对遥感影像这一高维数据源, GBDT 算法可最大程度实现信息的准确提取。因此, 本研究基于 GBDT, 在不同尺度下利用环境变量进行土壤盐度模拟, 并获取不同环境变量对土壤盐度的贡献度。随机抽取建模集, 并固定验证集, 以此优化模型参数, 经多次训练后, GBDT 模型具体参数设置为: 迭代回归树数量为 300, 学习速率为 0.01, 叶子节点的最小个数为 7, 交叉验证折数为 3, 算法依托 R 语言平台实现, 流程如图 3 所示。

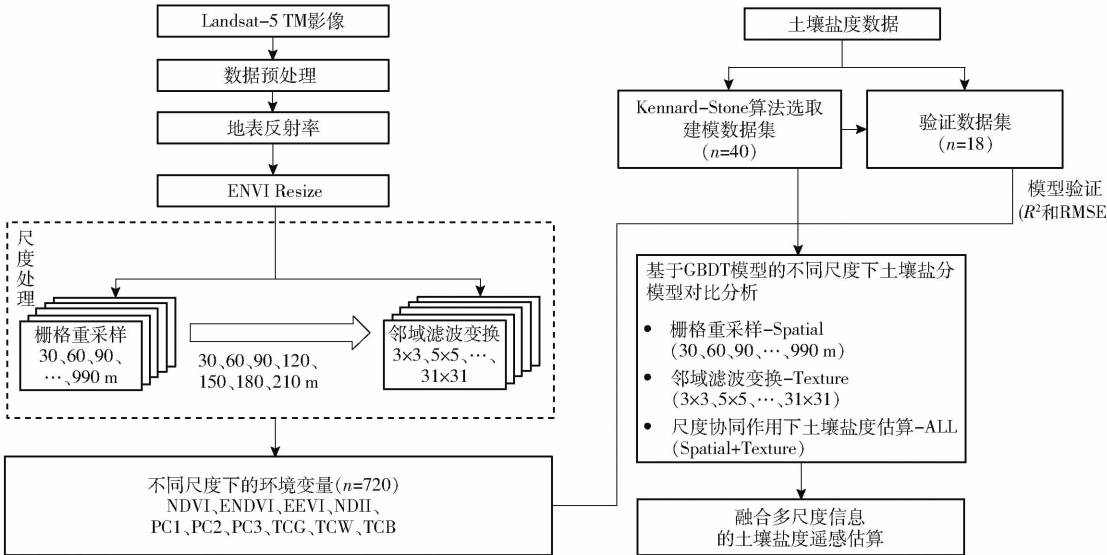


图 3 实验及模型计算过程流程图

Fig. 3 Flow chart of experiment and model calculation process

2 结果与分析

2.1 GBDT 模型构建及验证

本研究区土壤表层盐度最大值为 9.83 dS/m,最小值为 0.16 dS/m,平均值为 3.75 dS/m,标准差为 2.23 dS/m,变异系数达到 59.54%。基于土壤盐度的高变异性,选用 Kennard – Stone (K – S) 算法进行样本划分,选取 40 个样本点作为建模集,18 个样本点为验证集,验证集的数据分布特征与全集一致。为量化基于不同尺度建模的效果和性能,选用决定系数(R^2)、均方根误差(RMSE)对模型进行评估, R^2 越大,表明建模的精度越高,RMSE 表示模型的预测能力,值越小表明模型精度越高。由于模型数量庞大,考虑到篇幅有限,故选取了 $R_c^2 > 0.75$ 进行部分展示(表 2,表中邻域滤波尺度中下划线前数字是基础栅格分辨率,下划线后数字代表具体栅格分辨率下邻域滤波的窗口尺度,T – 30 ~ T – 150 代表不同空间分辨率下所有邻域窗口尺度的环境变量对土壤盐度进行模拟,具体而言,即当初始分辨率为 30 m 时,将 3×3 、 5×5 、 $\cdots$ 、 31×31 共计 15 个邻域窗口尺度输入 GBDT 对土壤盐度进行模拟;当初始分辨率为 60 m 时,将 3×3 、 5×5 、 $\cdots$ 、 17×17 共计 8 个邻域窗口尺度输入 GBDT 对土壤盐度进行模拟;依此类推,当初始分辨率为 150 m 时,将 3×3 、 5×5 、 7×7 共计 3 个邻域窗口尺度输入 GBDT 对土壤盐度进行模拟,Spatial 指集合 30、60、 $\cdots$ 、990 m 所有栅格重采样尺度的环境变量对土壤盐度进行模拟,Texture 指集合邻域滤波 7 个基础分辨率下所有邻域窗口尺度对土壤盐度进行模拟,ALL 指将所有栅格重采样尺度和邻域滤波尺度的 720 个环境变量输入 GBDT 对土壤盐度进行模拟)。从表 2 可见, R_p^2 恒小于 R_c^2 ,且验证集 RMSE 恒小于建模集 RMSE,说明模型没有出现过拟合现象,且 R^2 都不小于 0.73,最高可达 0.92,拟合效果较好,故使用 GBDT 进行土壤盐度预测可行。

2.2 栅格重采样变换后土壤盐度估算

基于 GBDT 算法,对不同栅格重采样尺度下土壤盐度进行预测,并计算相应尺度下不同环境变量对土壤盐度的贡献权重及其总体解析能力。为可视化表达不同尺度下环境变量对土壤盐度的解析效果,本文引入泰勒图进行相关表述(图 4,图中黑色虚线表示调整 R^2 ,蓝色弧线表示 RMSE (dS/m),当调整 R^2 越高、RMSE 越小时模型效果越好)。图 4a 为不同栅格重采样尺度环境变量对土壤盐度的解析能力,尺度范围为 30 ~ 990 m,在不同尺度条件下,不同环境变量对土壤盐度的响应也不尽相同。

表 2 土壤盐度预测效果
Tab.2 Statistical results of estimated soil salinity based on various scale conversion modes

尺度转换模型		建模集		验证集	
		R_c^2	RMSE/ (dS·m ⁻¹)	R_p^2	RMSE/ (dS·m ⁻¹)
栅格重采样	30 m	0.77	1.60	0.74	1.12
	60 m	0.79	1.81	0.75	1.10
	90 m	0.77	1.46	0.75	1.11
	120 m	0.78	1.49	0.75	1.11
	210 m	0.78	1.60	0.74	1.13
	360 m	0.77	1.58	0.73	1.14
	630 m	0.79	1.59	0.74	1.13
	840 m	0.76	1.58	0.74	1.14
邻域滤波	30_3	0.81	1.87	0.78	1.05
	30_5	0.82	1.82	0.77	1.05
	30_7	0.80	1.94	0.78	1.04
	30_9	0.82	1.73	0.77	1.07
	30_11	0.85	2.06	0.79	1.02
	30_13	0.80	2.25	0.76	1.09
	30_27	0.81	2.92	0.78	1.04
	60_3	0.84	1.99	0.76	1.08
	60_5	0.84	1.83	0.78	1.05
	60_7	0.82	1.99	0.76	1.08
	60_9	0.80	2.11	0.75	1.10
	120_5	0.78	2.13	0.76	1.08
多尺度协同作用	T-30	0.87	1.92	0.87	0.81
	T-60	0.85	1.33	0.85	0.85
	T-90	0.83	1.55	0.82	0.94
	T-120	0.81	1.90	0.80	0.99
	T-150	0.80	2.34	0.77	1.06
	Spatial	0.92	1.58	0.91	0.68
	Texture	0.88	1.13	0.88	0.78
	ALL	0.92	1.38	0.91	0.68

由图 4a 可知,上述栅格重采样尺度的调整 R^2 在 0.56 ~ 0.75 之间,RMSE 在 1.10 ~ 1.47 dS/m 之间,其中在栅格重采样尺度 60 m 下,对于解析土壤表层盐度-环境变量效果最好,解析能力达到了 75.31% ($R^2 > 0.75$),为单一栅格重采样尺度下研究土壤表层盐度-环境变量关系的最佳尺度。结合不同尺度下各环境变量对土壤盐度贡献权重(图 5)发现,对应栅格重采样尺度 60 m 下,对土壤表层盐度贡献率从高到低依次为 NDII、EEVI、PC2、ENDVI,其贡献率总和达到了 77.32%,解释了 77.32% 的土壤盐渍化变异,这也佐证了本研究使用植被指数对表层土壤盐度进行表征是可行的。随着尺度的上升,GBDT 算法对于环境变量和土壤盐度的关系描述效果越来越不理想。其中,起主导作用的环境变量从原有与植被相关的 TCG、植被指数等变成了图像衍生变换参量,这可能归因于随着分

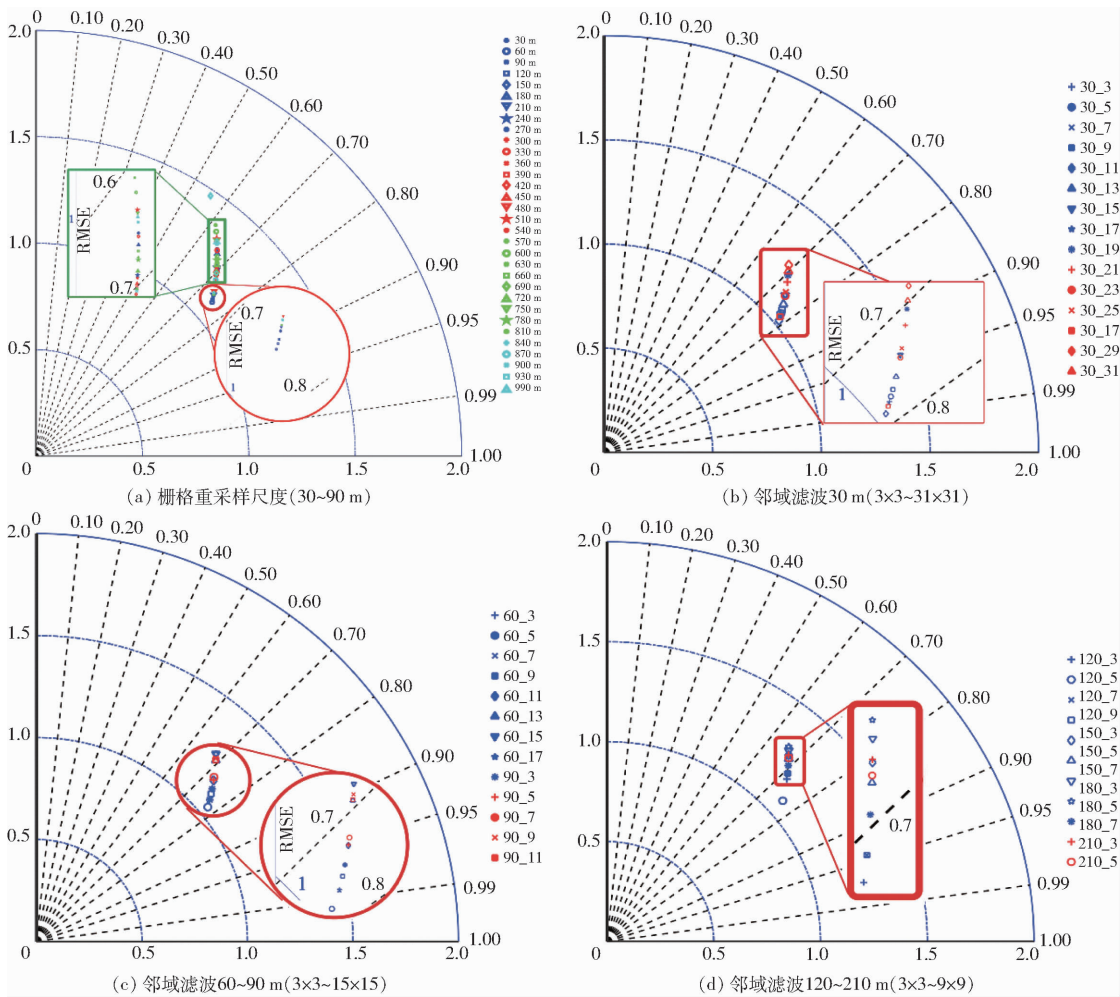


图 4 不同尺度对土壤盐度-环境变量之间关系的影响
Fig. 4 Effects of different scales on soil salinity – environmental variables

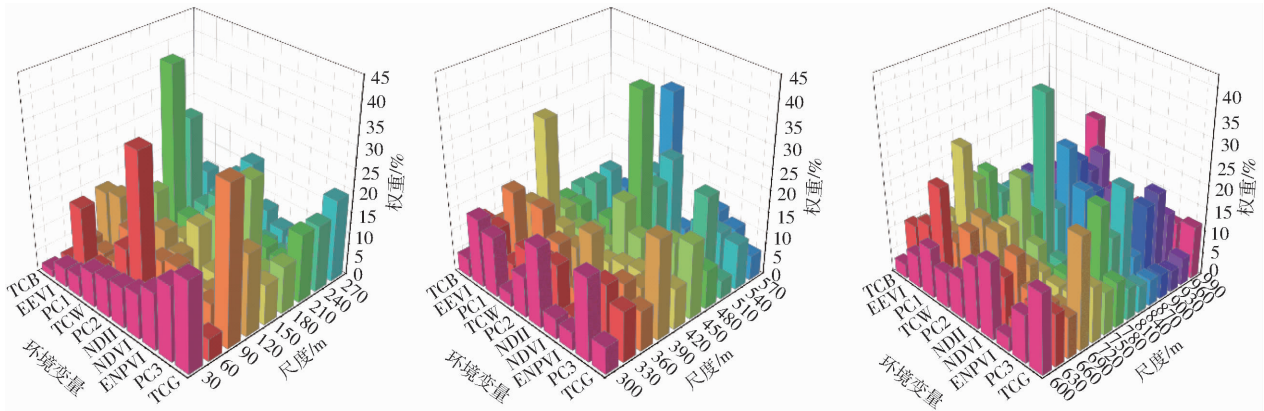


图 5 栅格重采样模式不同尺度下各环境变量对土壤盐度贡献权重
Fig. 5 Contribution weight of each environmental variabale to soil salinity at different raster resampling scales

分辨率的下降,原始数据所包含的信息丰富度损失,植被特征被弱化,而 PC、TC 等参数以其独有的信息优势过渡为主导因素。

2.3 邻域滤波变化后土壤盐度估算

对 30 ~ 210 m 栅格重采样分辨率下对应计算邻域窗口 3 × 3、5 × 5、…、31 × 31 的邻域滤波进行土壤盐度的预测,并获取各环境变量对土壤盐度影响权重,结果见图 4、6。由图 4b 可知,在 30 m 分辨率多

种邻域滤波尺度下,土壤盐度-环境变量之间的调整 R^2 最大值为 0.79, RMSE 为 1.02 dS/m,可解释 78.55% 的土壤盐度变化,对应邻域窗口 11 × 11,地面分辨率为 330 m、响应的环境变量依次为 TCG、ENDVI、PC3、NDVI、NDII,能够解释近 78.57% 的盐度变异性。图 4c、4d 为 60 ~ 210 m 栅格重采样分辨率下多种邻域滤波尺度对土壤盐度与环境变量关系的表征,调整 R^2 为 0.66 ~ 0.78 之间。当基础栅格

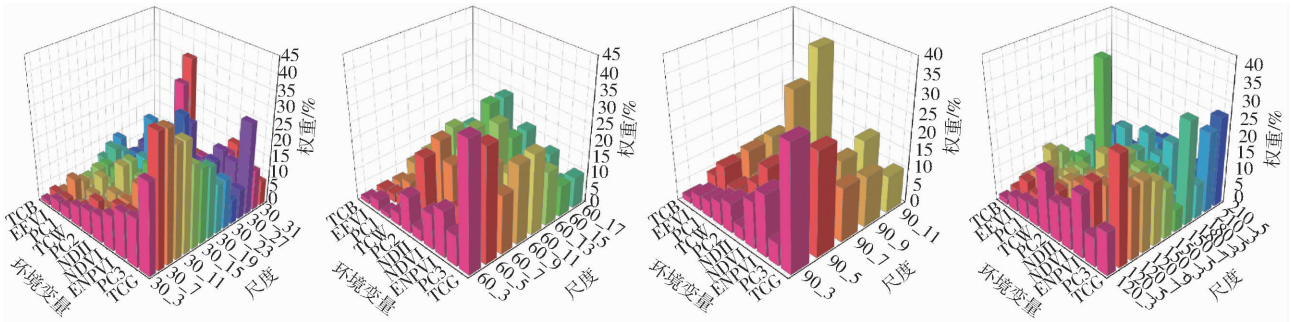


图6 邻域滤波不同尺度下各环境变量对土壤盐度贡献权重

Fig.6 Contribution weight of each environmental variabile to soil salinity at different neighborhood filtering scales

重采样尺度为 60 m 时, R^2 在 0.68 ~ 0.78 之间, RMSE 为 1.05 ~ 1.25 dS/m, 其中当地面分辨率为 300 m、邻域窗口为 5×5 时效果最好; 90 m 时, R^2 在 0.69 ~ 0.74 之间, RMSE 为 1.05 ~ 1.24 dS/m, 其中当地面分辨率为 270 m、邻域窗口为 3×3 时效果最好; 120 ~ 210 m 时, R^2 在 0.66 ~ 0.76 之间, 其中当窗口为 5×5 时效果最好; 对土壤盐度的解析力分别达到 76.01%、68.11%、69.41%、67.97%。可以发现, 当初始分辨率为 30、60、90 m 和 120 ~ 210 m 时, 对应最佳邻域窗口分别为 11×11 、 5×5 、 3×3 、 5×5 , 其中当初始空间分辨率为 30 ~ 90 m 时, 最优土壤盐度表征尺度为 300 m 左右。单一栅格重采样尺度下, 伴随着邻域窗口的变大, 土壤盐度对环境变量的解析程度也在下降; 单一邻域尺度下, 随着栅格重采样尺度的增大其对环境变量的响应也在弱化。即随着尺度上升, 调整 R^2 总体呈下降趋势, RMSE 呈上升趋势, 其中与土壤盐度相关的环境变量集中为与植被相关的 TCG 和植被指数。

2.4 多尺度协同作用下土壤盐度估算

通过 GBDT 算法结合多尺度环境变量对土壤盐度进行模拟, 并获取多尺度下环境变量指征土壤盐度的最优组合。图 7 为多尺度下基于原始分辨率数据和邻域滤波尺度, 考量环境变量对土壤盐度-环境变量的协同影响。在参与分析的 30 ~ 990 m 所有栅格重采样尺度时, 其调整 R^2 可达到 0.91, RMSE 为 0.68 dS/m, 对土壤盐度的解析度达到 90.63%。对邻域滤波, 分别以 30、60、90、120、150、180、210 m 为基础分辨率, 讨论在上述 7 个基础栅格重采样分辨率下随着邻域窗口的变化环境变量与土壤盐度的响应情况, 结果如图 7 所示, 在空间分辨率为 30、60、90、120、150、180、210 m 时, 其对土壤盐度预测的调整 R^2 分别为 0.86、0.85、0.82、0.80、0.77、0.73、0.67, RMSE 分别为 0.81、0.85、0.94、0.99、1.06、1.14、1.27 dS/m, 随着基础空间分辨率的下降, 调整 R^2 降低, 环境变量对土壤盐度的解析力度逐渐降低。当考虑 30 ~ 210 m 所有邻域滤波的情况下, 对土壤

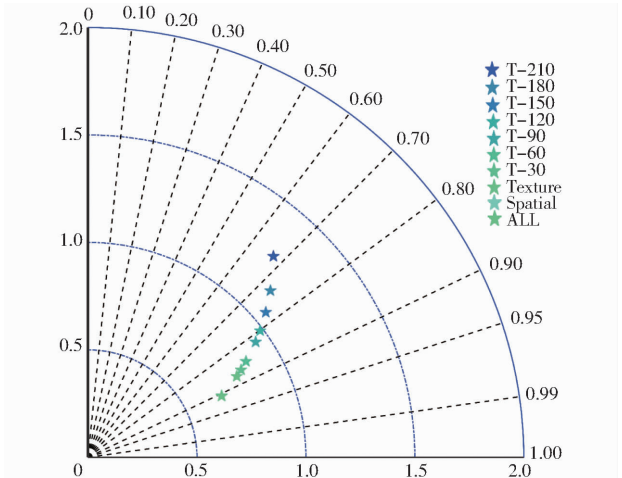


图7 多尺度环境变量协同对土壤盐度估算的影响

Fig.7 Effects of multi-scale environmental variables on soil salinity estimation

盐度的调整 R^2 可达到 0.88, RMSE 为 0.78 dS/m, 由此可见, 相较于邻域滤波变换, 栅格重采样可以更为显著提高土壤盐渍化建模效果。当考虑全尺度情况下, 调整 R^2 达到 0.91, RMSE 低至 0.68 dS/m, 对土壤盐度的解析度达到 90.66%, 与栅格重采样多尺度下性能接近, 故当考虑全尺度时, 增加邻域滤波尺度对环境变量与土壤盐度关系的研究改善有限。综上所述, 多尺度协同极大地提高了环境变量对土壤盐度的解释力, 其中栅格重采样尺度对环境变量与土壤盐度关系的研究起到主导作用。

3 讨论

在植被覆盖度高的区域, 土壤盐渍化会造成植被红光波段反射率增加、近红外波段反射率下降, 故植被指数的变化有助于推测盐渍化的程度和发展趋势^[3]。基于此, 在绿洲地区, 已有许多学者开展了相关研究, 结果表明, 因干旱区地理环境存在显著的差异性, 环境变量在某个区域表现出的高解析力, 其泛化能力尚不可知; 不同盐度等级下的土壤盐度光谱信息存在明显差异, 遥感指数在盐渍化表征上具有复杂的不确定性; 受多种因素影响, 环境变量的数

量和类别上还有待进一步优化^[21]。在本研究中,随着栅格分辨率的上升,环境变量对土壤盐度的解析能力从与植被相关的 TCG、植被指数等变成了图像衍生变换参量,如 PC1、PC2。对邻域滤波而言,TCG、PC3、NDII、TCW 和 ENDVI 可解析 69.83% 的土壤盐度变异,对土壤盐度的预测具有一定的应用优势。不同栅格重采样和邻域滤波窗口尺度下,环境变量与土壤盐度的关系存在较为明显的差异,这说明了环境变量与土壤盐度之间存在强尺度依赖性。MULLER 等^[33]在多种分辨率下对比了植被指数与土壤盐度之间的耦合关系,发现随着分辨率的降低其响应尺度也在降低,这与本研究得到的结论基本一致,如在 30 m 分辨率时,最佳邻域窗口为 11 × 11,随着分辨率变为 60、90 m,最佳邻域窗口分别为 5 × 5、3 × 3。然而,值得注意的是当分辨率在 120 ~ 210 m 时,最佳邻域窗口稳定在 5 × 5,这可能与研究区独特的地理特征及土壤属性的空间变异性有关。

尺度问题是生态学研究中最具挑战的问题之一,也是遥感生态学需要面对的主要问题,反演的准确性和遥感数据的空间精度有很大的关系^[34-35]。本研究以国内外研究者广泛选用的 Landsat 影像为数据源,针对土壤盐渍化现象,分析了土壤盐度-环境变量关系的尺度效应。但该数据的空间分辨率(观测尺度)为 30 m,仅依据单次采样数据结合遥感对土壤盐度进行模拟,仍存在一定的不确定性。随着遥感技术的发展及传感器分辨率的不断提升,更优质的遥感数据的出现(更高空间分辨率、更宽光谱范围等,如欧洲空间局的 Sentinel-2 多光谱遥感数据),为进一步研究尺度对土壤盐度-环境变量之间耦合关系提供了新的选择。同时,本研究使用固

定观测尺度,栅格重采样的步长为 30 m(像元逐个加入),但这可能造成某些尺度被遗漏。而高空间分辨率数据的引入则可以缩短研究间隔,使得更为细致地研究环境变量-土壤盐度的尺度效应成为可能。需要注意的是,由于地表空间异质性的存在会影响土壤属性或者过程模拟的精确程度,故合适的栅格分辨率和邻域滤波尺度的选择最终依赖于研究区的地域特点,即被模拟属性的空间变异性 and 分布规律。

4 结论

(1)单一尺度下,邻域滤波尺度对绿洲土壤盐度的模拟总体优于重采样模式。在单一栅格重采样尺度下,60 m 栅格重采样尺度对土壤表层盐度-环境变量解析效果最好,解析能力达到 75.31%;单一邻域滤波尺度下,初始分辨率为 30 m、邻域窗口为 11 × 11 是土壤表层盐度-环境变量最佳表征尺度,解析能力为 78.55%,随着邻域窗口尺度的外推,其对土壤盐度的响应效果逐渐减弱。

(2)栅格重采样模式相对邻域滤波而言,其调整 R^2 波动范围更为宽泛,分布范围为 0.56 ~ 0.75,而后者变化平缓。一定程度上说明,栅格重采样方法对土壤盐度-环境变量关系的变化较为敏感,邻域滤波对土壤盐度-环境变量关系的敏感程度相对较弱。

(3)当考虑融合多种尺度时,环境变量对土壤盐度的模拟精度相比仅利用单一尺度而言,得到极大程度提升,具体表现在 R^2 从 0.78 提高到 0.91, RMSE 从 1.02 dS/m 降低到 0.68 dS/m,对土壤盐度的解析度最高可达到 90.66%,这也体现了不同环境变量协同作用可最大程度发挥其最大响应效力,同时变量之间的信息进行了有效互补。

参 考 文 献

[1] MA L G, MA F L, LA J D, et al. Characterizing and modeling regional-scale variations in soil salinity in the arid oasis of Tarim Basin, China[J]. *Geoderma*, 2017, 305:1-11.

[2] WANG J, DING J, MA X, 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.

[3] 王飞,丁建丽,魏阳,等. 基于 Landsat 系列数据的盐分指数和植被指数对土壤盐度变异性的响应分析—以新疆天山南北典型绿洲为例[J]. *生态学报*, 2017, 37(15): 5007-5022.

WANG Fei, DING Jianli, WEI Yang, et al. Sensitivity analysis of soil salinity and vegetation indices to detect soil salinity variation by using Landsat series images: applications in different oases in Xinjiang, China[J]. *Acta Ecologica Sinica*, 2017, 37(15):5007-5022. (in Chinese)

[4] MULDER V L, BRUIN S D, SCHAEPMAN M E, et al. The use of remote sensing in soil and terrain mapping—a review[J]. *Geoderma*, 2011, 162(1-2): 1-19.

[5] 李小文,王祎婷. 定量遥感尺度效应刍议[J]. *地理学报*, 2013, 68(9): 1163-1169.

LI Xiaowen, WANG Yiting. Prospects on future developments of quantitative remote sensing[J]. *Acta Geographica Sinica*, 2013, 68(9): 1163-1169. (in Chinese)

[6] MCCARTHY A J P. The Irish national electrification scheme[J]. *Geographical Review*, 1957, 47(4): 539-554.

[7] BEHRENS T, ZHU A X, SCHMIDT K, et al. Multi-scale digital terrain analysis and feature selection for digital soil mapping[J]. *Geoderma*, 2010, 155(3-4): 175-185.

[8] MILLER B A, KOSZINSKI S, WEHRHAN M, et al. Impact of multi-scale predictor selection for modeling soil properties[J]. *Geoderma*, 2015, 239-240: 97-106.

[9] MOUGENOT B, POUGET M, EPEMA G F. Remote sensing of salt affected soils[J]. *Remote Sensing Reviews*, 1993, 7(3-

- 4):241–259.
- [10] RAO B R M, RAVISANKAR T, DWIVEDI R S, et al. Spectral behavior of salt-affected soils[J]. *International Journal of Remote Sensing*, 1995, 16(12): 2125–2136.
- [11] WU H, LI Z L. Scale issues in remote sensing: a review on analysis, processing and modeling[J]. *Sensors*, 2009, 9(3): 1768–1793.
- [12] MCCABE M F, WOOD E F. Scale influences on the remote estimation of evapotranspiration using multiple satellite sensors[J]. *Remote Sensing of Environment*, 2006, 105(4): 271–285.
- [13] VAZE J, TENG J, SPENCER G. Impact of DEM accuracy and resolution on topographic indices[J]. *Environmental Modelling & Software*, 2010, 25(10): 1086–1098.
- [14] WECHSLER S. Uncertainties associated with digital elevation models for hydrologic applications: a review[J]. *Hydrology and Earth System Sciences*, 2007, 3(4): 1481–1500.
- [15] KIM D, ZHENG Y. Scale-dependent predictability of DEM-based landform attributes for soil spatial variability in a coastal dune system[J]. *Geoderma*, 2011, 164(3–4): 181–194.
- [16] PARK S, RUECKER G, AGYARE W. Influence of grid cell size and flow routing algorithm on soil-landform modeling[J]. *Journal of the Korean Physical Society*, 2009, 44(2): 122–145.
- [17] 王飞,杨胜天,魏阳,等. 基于 RF 和 SGT 算法地子区优先建模对绿洲尺度土壤盐度预测精度的影响[J]. *中国农业科学*, 2018, 51(24): 4659–4676.
- WANG Fei, YANG Shengtian, WEI Yang, et al. Influence of sub-region priority modeling constructed by random forest and stochastic gradient treeboost on the accuracy of soil salinity prediction in oasis scale[J]. *Scientia Agricultura Sinica*, 2018, 51(24): 4659–4676. (in Chinese)
- [18] 李宝富,熊黑钢,龙桃,等. 新疆奇台绿洲农田灌溉前后土壤水盐时空变异性研究[J]. *中国沙漠*, 2012, 32(5): 1369–1378.
- LI Baofu, XIONG Heigang, LONG Tao, et al. Irrigation-induced spatio-temporal variability of soil water and salinity in Qitai Oasis, Xinjiang of China[J]. *Journal of Desert Research*, 2012, 32(5): 1369–1378. (in Chinese)
- [19] 郝大磊,肖青,闻建光,等. 定量遥感升尺度转换方法研究进展[J]. *遥感学报*, 2018, 22(3): 40–55.
- HAO Dalei, XIAO Qing, WEN Jianguang, et al. Advances in upscaling methods of quantitative remote sensing[J]. *Journal of Remote Sensing*, 2018, 22(3): 408–423. (in Chinese)
- [20] XU K, TIAN Q, YANG Y, et al. How up-scaling of remote-sensing images affects land-cover classification by comparison with multiscale satellite images[J]. *International Journal of Remote Sensing*, 2019, 40(7): 2784–2810.
- [21] 魏阳,丁建丽,王飞. 基于 Landsat OLI 的绿洲灌区土壤盐度最优预测尺度分析[J]. *中国农业科学*, 2017, 50(15): 2969–2982.
- WEI Yang, DING Jianli, WANG Fei. Optimal scale analysis of soil salinity prediction in oasis irrigated area of arid land based on Landsat OLI[J]. *Scientia Agricultura Sinica*, 2017, 50(15): 2969–2982. (in Chinese)
- [22] HUETE A. A comparison of vegetation indices over a global set of TM images for EOS – MODIS[J]. *Remote Sensing of Environment*, 1997, 59(3): 440–451.
- [23] TEILLET P M, STAENZ K, WILLIAM D J. Effects of spectral, spatial, and radiometric characteristics on remote sensing vegetation indices of forested regions[J]. *Remote Sensing of Environment*, 1997, 61(1): 139–149.
- [24] 王飞,丁建丽,伍漫春. 基于 NDVI–SI 特征空间的土壤盐渍化遥感模型[J]. *农业工程学报*, 2010, 26(8): 168–173.
- WANG Fei, DING Jianli, WU Manchun. Remote sensing monitoring models of soil salinization based on NDVI – SI feature space[J]. *Transactions of the CSAE*, 2010, 26(8): 168–173. (in Chinese)
- [25] DOUAOU A E K, NICOLAS H, WALTER C. Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data[J]. *Geoderma*, 2006, 134(1–2): 217–230.
- [26] ZHANG T T, QI J G, GAO Y, et al. Detecting soil salinity with MODIS time series VI data[J]. *Ecological Indicators*, 2015, 52: 480–489.
- [27] FRIEDMAN J H. Greedy function approximation: a gradient boosting machine[J]. *The Annals of Statistics*, 2001, 29(5): 1189–1232.
- [28] FRIEDMAN J H. Stochastic gradient boosting[J]. *Computational Statistics & Data Analysis*, 2002, 38(4): 367–378.
- [29] JOHNSON R, ZHANG T. Learning nonlinear functions using regularized greedy forest[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(5): 942–954.
- [30] 葛翔宇,丁建丽,王敬哲,等. 基于竞争适应重加权采样算法耦合机器学习的土壤含水量估算[J]. *光学学报*, 2018, 38(10): 393–400.
- GE Xiangyu, DING Jianli, WANG Jingzhe, et al. Estimation of soil moisture content based on competitive adaptive reweighted sampling algorithm coupled with machine learning[J]. *Acta Optica Sinica*, 2018, 38(10): 393–400. (in Chinese)
- [31] WEI Z, MENG Y, ZHANG W, et al. Downscaling SMAP soil moisture estimation with gradient boosting decision tree regression over the Tibetan Plateau[J]. *Remote Sensing of Environment*, 2019, 225: 30–44.
- [32] LU Y, XIAOTONG Z, SHUNLIN L, et al. Estimating surface downward shortwave radiation over China based on the gradient boosting decision tree method[J]. *Remote Sensing*, 2018, 10(2): 185.
- [33] MULLER S J, VAN NIEKERK A. Identification of WorldView–2 spectral and spatial factors in detecting salt accumulation in cultivated fields[J]. *Geoderma*, 2016, 273: 1–11.
- [34] 陈俊英,王新涛,张智韬,等. 基于无人机–卫星遥感升尺度的土壤盐渍化监测方法[J/OL]. *农业机械学报*, 2019, 50(12): 161–169.
- CHEN Junying, WANG Xintao, ZHANG Zhitao, et al. Soil salinization monitoring method based on UAV – satellite remote sensing scale-up[J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2019, 50(12): 161–169. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20191218&flag=1. DOI:10.6041/j.issn.1000-1298.2019.12.018. (in Chinese)
- [35] WANG J Z, DING J L, YU D, et al. Machine learning-based detection of soil salinity in an arid desert region, Northwest China: a comparison between Landsat – 8 OLI and Sentinel – 2 MSI[J]. *Science of the Total Environment*, 2020, 707: 136092.