

doi: 10.6046/zrzyyg.2022047

引用格式: 黄晓宇, 王雪梅, 卡吾恰提·白山. 基于 Landsat8 OLI 影像干旱区绿洲土壤含盐量反演[J]. 自然资源遥感, 2023, 35(1): 189-197. (Huang X Y, Wang X M, Kawuqiati B S. Inversion of soil salinity of an oasis in an arid area based on Landsat8 OLI images[J]. Remote Sensing for Natural Resources, 2023, 35(1): 189-197.)

# 基于 Landsat8 OLI 影像干旱区绿洲土壤含盐量反演

黄晓宇<sup>1</sup>, 王雪梅<sup>1,2</sup>, 卡吾恰提·白山<sup>1</sup>

(1. 新疆师范大学地理科学与旅游学院, 乌鲁木齐 830054; 2. 新疆维吾尔自治区重点实验室“新疆干旱区湖泊环境与资源实验室”, 乌鲁木齐 830054)

**摘要:** 利用遥感技术进行土壤含盐量的快速检测可为土壤盐渍化治理和绿洲农业合理开发提供科学指导。基于渭干河—库车河三角洲绿洲采集的 95 个土壤样品, 采用光谱指数、波段反射率与实测土壤含盐量, 运用多元线性回归、偏最小二乘回归、支持向量机回归和随机森林回归方法构建土壤含盐量估测模型, 并利用最优估测结果对研究区土壤含盐量的空间分布格局进行遥感反演。结果表明: 通过全子集回归法筛选出与土壤含盐量相关显著的 9 个光谱因子, 相关系数均在 0.5 以上 ( $P < 0.01$ )。其中盐分指数中 SI-T 与土壤含盐量的相关系数最大为 0.648; 对比 4 种反演模型的估测精度, 拟合的效果由高到低依次为随机森林回归 > 支持向量机回归 > 偏最小二乘回归 > 多元线性回归。其中随机森林模型拟合精度表现最佳, 训练集和验证集的决定系数分别为 0.870 和 0.766; 相对分析误差分别为 2.792 和 2.105, 值均大于 2, 表明模型反演效果较好, 有稳定的估测能力; 由随机森林模型的反演结果来看, 第 I 等级和第 II 等级占比达到 41.62%, 分布于绿洲内部的耕作区; 第 III, IV 和第 V 等级区共占比 56.41%, 主要分布在绿洲外围与沙漠的交错带和荒漠区。采用随机森林机器学习建模方法对土壤含盐量进行反演, 估测效果明显优于传统的统计模型, 可为干旱区绿洲土壤盐渍化监测提供参考。

**关键词:** 光谱指数; 机器学习模型; 土壤含盐量; 遥感反演; 干旱区绿洲

**中图分类号:** P 935.1; TP 79 **文献标志码:** A **文章编号:** 2097-035X(2023)01-0189-09

## 0 引言

土壤盐渍化是干旱和半干旱地区的主要灾害之一, 是绿洲生态系统退化的一个重要特征, 受气候变化和人类不合理开发利用的影响, 继而危害周围环境和农业生产<sup>[1-3]</sup>。“十四五”规划明确了我国经济社会发展要深入开展土壤污染防治, 指出开展土壤盐渍化防治是农业可持续发展的必由之路。土壤盐渍化会使土壤结构遭到破坏, 抑制农作物的正常生长, 导致农作物产量下降, 进而导致土地荒漠化的形成, 最终将会威胁到人类的生存<sup>[4-6]</sup>。土壤含盐量 (soil salt content, SSC) 是土壤盐渍化形成的来源, 准确了解 SSC 的空间分布和变化特征对治理土壤盐渍化具有重要意义。获取 SSC 的传统方法是通过野外实地采集样本数据并进行室内化学分析, 该方法需要大量样点数据才能合理解释区域土壤盐渍

化状况, 工作量大且耗时耗力。利用遥感技术进行区域尺度 SSC 的快速检测, 对区域土壤盐渍化的动态监测与防治具有十分重要的现实意义<sup>[7]</sup>。

目前, 利用遥感光谱指数构建反演模型获取某一区域土壤盐渍化信息已成为遥感技术在土壤盐渍化研究中的热点。多位学者利用与土壤盐渍化程度相关性较高的多光谱遥感指数构建盐渍化预测模型进而实现对土壤盐渍化的监测<sup>[7-8]</sup>。起初有学者采用传统的统计分析建模方法, 张素铭等<sup>[9]</sup>采用波段和敏感波段组合的方法建立的多元线性回归 (multiple linear regression, MLR) 模型效果最好; 周晓红等<sup>[10]</sup>对艾比湖湿地自然保护区基于多光谱遥感指数分别与土壤盐分建立多种线性回归模型; 张同瑞等<sup>[11]</sup>通过植被指数和实测 SSC 构建线性、二次和三次函数等模型估测黄河三角洲的 SSC; 而奚雪等<sup>[12]</sup>针对黄河三角洲滨海盐渍土问题, 基于多光谱指数采用稳定性和适宜性更好的偏最小二乘法建立土壤

收稿日期: 2022-02-11; 修订日期: 2022-05-21

基金项目: 新疆维吾尔自治区自然科学基金项目“和田地区土地荒漠化时空演变及预警研究”(编号: 2020D01A79)和国家自然科学基金项目“塔里木盆地北缘绿洲-荒漠过渡带植被对土壤盐渍化的响应研究”(编号: 41561051)共同资助。

第一作者: 黄晓宇(1995-), 男, 硕士研究生, 研究方向为资源环境遥感。Email: 18699576547@163.com。

通信作者: 王雪梅(1976-), 女, 教授, 博士, 硕士生导师, 研究方向为干旱区资源环境遥感技术应用研究。Email: wangxm\_1225@sina.com。

反演模型。后来随着建模方法的改进,有学者运用机器学习建模方法,陈红艳等<sup>[13]</sup>基于改进后的多光谱植被指数,采用支持向量机回归(support vector machine regression, SVMR)对黄河入海口处土壤盐分进行遥感反演;Fathizad 等<sup>[14]</sup>针对伊朗中部沙漠的土壤盐分问题,基于 Landsat8 OLI 影像提取的植被盐分指数,利用随机森林回归(random forest regression, RFR)模型反演出沙漠的土壤盐分。从上述研究可以看出,采用多光谱遥感指数构建 SSC 反演模型是一个非常重要的研究方向,从传统统计分析方法到机器学习建模方法的改进对促进遥感技术在环境监控等方面具有非常重要的作用。

基于此,本研究选取典型的干旱区绿洲耕层土壤作为研究对象,采用多光谱遥感指数和光谱反射率并结合传统统计分析模型和机器学习模型进行 SSC 的估测,基于最佳反演模型对研究区土壤盐渍化的空间分布特征进行研究。

## 1 研究区概况及数据源

### 1.1 研究区概况

渭干河—库车河三角洲绿洲(简称渭—库绿洲)位于新疆南部塔里木盆地的中北部(图 1),属于典型的温带大陆性气候。多年平均气温为 11.6℃,多年平均降水量为 52 mm,蒸降比为 54:1。自然条件下,植被盖度较低且多为盐生植被,如芦苇、柽柳、骆驼刺、胡杨、盐节木、盐穗木、盐爪爪、花花柴、苦豆子、白刺和黑刺等。研究区土壤以棕漠土、灌淤土、沼泽土、盐土和风沙土为主。由于该区域自然环境脆弱,加之人为不合理的农业管理,使该地区土壤盐渍化现象发生较为普遍,且存在大面积的潜在次生盐渍化土壤,严重影响了该区域的生态环境,导致绿洲土地退化问题日益严峻,绿洲农业的可持续发展受到了严重威胁<sup>[15]</sup>。

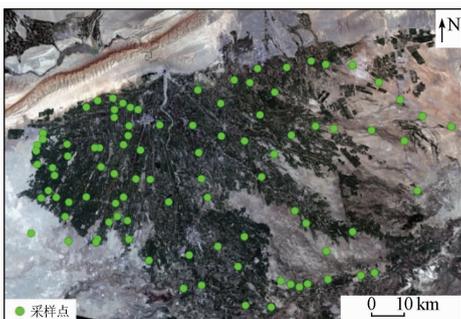


图 1 研究区采样点分布

Fig. 1 Distribution of sampling points in the study area

### 1.2 土样采集与遥感影像预处理

渭—库绿洲夏季蒸发量高能突显地表 SSC 的特征,因此土壤样品采集工作在 2019 年 7 月中下旬进行。利用 GPS 在绿洲不同盐渍化区域进行随机定位,选取样点进行采样。通过剔除,共选取了 95 个样点,采样深度为 0~20 cm。将采集的土样在实验室进行自然风干,剔除杂质,通过磨细过筛,充分混合均匀进行分析。按照水土比 1:5 浸提后,使用电导率仪测定土壤电导率,SSC 和土壤含水量。土壤检测的标准方法采用 NY/T 1121.16—2006 土壤水溶性总盐。结合研究区地理位置,选取云量较少,与采样时间同期的 Landsat8 影像,使用 ENVI 5.3 软件和 R 语言对影像进行辐射定标和大气校正等预处理,提取遥感光谱指数和波段反射率。

## 2 研究方法

### 2.1 遥感光谱指数计算

光谱指数选取与 SSC 有相关性的植被指数、盐分指数和波段反射率<sup>[16]</sup>,如表 1 所示,分别为归一化植被指数(normalized difference vegetation index, NDVI)、扩展增强型植被指数(extended normalized difference vegetation index, ENDVI)、差值植被指数(difference vegetation index, DVI)、冠层响应盐度指数(canopy response salinity index, CRSI)、增强型植被指数(enhanced vegetation index, EVI)、扩展增强型植被指数(extended enhanced vegetation index, EEVI)、归一化盐分指数(normalized difference salinity index, NDSI)、多种盐分指数  $S_3, S_5, SI_4$  和  $SI-T$ 。

表 1 光谱指数计算公式

Tab. 1 Calculation formulas of spectral indexes

光谱指数	计算公式 <sup>①</sup>	参考文献
NDVI	$NDVI = (NIR - Red) / (NIR + Red)$	[10]
ENDVI	$ENDVI = (NIR + SWIR2 - Red) / (NIR + SWIR2 + Red)$	[10]
DVI	$DVI = NIR - Red$	[10]
CRSI	$CRSI = [(NIR \cdot Red - Green \cdot Blue) / (NIR \cdot Red + Green \cdot Blue)]^{0.5}$	[15]
EVI	$EVI = 2.5 [(NIR - Red) / (NIR + 6Red - 7.5Blue + 1)]$	[12]
EEVI	$EEVI = 2.5 [(NIR + SWIR1 - Red) / (NIR + SWIR1 + 6Red - 7.5Blue + 1)]$	[12]
NDSI	$NDSI = (Red - NIR) / (Red + NIR)$	[17]
$S_3$	$S_3 = (Green \cdot Red) / Blue$	[17]
$S_5$	$S_5 = (Blue \cdot Red) / Green$	[17]
$SI_4$	$SI_4 = (Green^2 + Red^2)^{0.5}$	[10]
$SI-T$	$SI-T = (Red/NIR) \times 100$	[18]

①式中: Blue 为蓝光波段反射率; Green 为绿光波段反射率; Red 为红光波段反射率; NIR 为近红外波段反射率; SWIR1 为短波红外 1 波段反射率; SWIR2 为短波红外 2 波段反射率。

## 2.2 SSC 反演模型

由于统计分析建模方法有直观的理解和解释,而机器学习建模方法能够有效地解决非线性问题。本研究以筛选的光谱指数和波段反射率作为反演模型的自变量,SSC 为因变量,采用 MLR、偏最小二乘回归(partial least squares regression, PLSR)、SVMR 和 RFR 模型,通过 R 语言和 Matlab 软件实现模型的构建和验证。MLR 是指 2 个或多个变量的回归分析,MLR 是一元线性回归的扩展,相比较能够充分展示各变量之间的相互关系,更能反映变量的真实情况。PLSR 用于解决多对多线性回归的问题,尤其是当变量个数较多且存在多重相关性时,其分析过程集中了主成分分析和线性回归分析等特点。SVMR 基于统计学习理论是一种相对简单的监督机器学习方法,能够从理论上实现对不同类别进行最优分类,通过建立核函数,寻找超平面,将用于分类问题的损失函数稍加改动,可用于回归问题。RFR 是 1995 年由贝尔实验室提出的,由许多决策树组成,是目前流行的机器学习模型之一。

## 2.3 模型精度评价

为了评价反演模型的效果,采用决定系数  $R^2$ 、均方根误差  $RMSE$  和相对分析误差  $RPD$  进行精度评价。其中  $R^2$  越大,表明模型的拟合程度越好。 $RMSE$  越小,表明模型的预测精度较高。 $RPD$  用以衡量模型的预测能力,当  $RPD < 1.4$  时认为所建模型预测结果较差;当  $1.4 \leq RPD < 2.0$  时认为所建模型预测结果基本满意;当  $RPD \geq 2.0$  时可认为所建模型的预测结果较好<sup>[19]</sup>。各检验指标表达式分别为:

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

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad (2)$$

$$RPD = \frac{SD}{RMSE}, \quad (3)$$

式中:  $n$  为采样点数量;  $\hat{y}_i$  为土壤样点  $i$  的预测值,  $y_i$  为土壤样点  $i$  的实测值;  $\bar{y}$  为土壤样点实测值的平均值;  $SD$  为数据的标准差。

## 3 结果与分析

### 3.1 SSC 的特征分析

将研究区采集的 95 个实测样点进行随机抽样,划分出 70% 的样本作为训练样本,剩余为验证样

本,并进行描述性统计(表 2)。分析得出研究区土壤样本 SSC 在  $0.525 \sim 216.375 \text{ g} \cdot \text{kg}^{-1}$  区间变化。总体样本、训练样本和验证样本的平均值分别为  $26.349 \text{ g} \cdot \text{kg}^{-1}$ ,  $25.576 \text{ g} \cdot \text{kg}^{-1}$  和  $28.108 \text{ g} \cdot \text{kg}^{-1}$ , 说明 SSC 主要集中分布在  $26.349 \text{ g} \cdot \text{kg}^{-1}$  左右。变异系数是反映样本数据的离散程度,3 种样本的变异系数均大于 100%,表明研究区 SSC 的离散程度较大,属于高度空间变异;同时也反映出研究区 SSC 空间分布不均,变化波动较大。

表 2 土壤样本基本统计特征

Tab. 2 Basic statistical characteristics of soil samples

样本类型	样本数/个	SSC/(g · kg <sup>-1</sup> )			变异系数/%	
		最大值	最小值	平均值		
总体样本	95	216.375	0.525	26.349	42.648	161.859
训练样本	66	216.375	0.525	25.576	43.265	169.159
验证样本	29	145.325	0.825	28.108	41.908	149.098

根据研究区土壤 SSC 的实际分布情况,基于乔木等<sup>[20]</sup>研究中的 SSC 分级,制定了本研究 SSC 分级标准(表 3)。实测样点的 SSC 在 5 种等级中均有分布,有 58.95% 的实测样点位于第 I 等级区域中,该区域大部分以农作物为主;第 II 等级区域主要以农作物和草本为主并有 17.89% 的实测样点分布其中;第 III 和 IV 等级区域主要以草本和灌木等植被为主并有 14.74% 的实测样点分布其中,而第 V 等级区域实测样点较少只有 8.42% 的实测样点分布其中,植被类型以灌木和乔木为主。

表 3 SSC 分级标准

Tab. 3 Classification standard of SSC

等级	SSC/(g · kg <sup>-1</sup> )	地表植被类型	生长情况
I	< 10	小麦、玉米和棉花等农作物以及红枣和核桃等经济作物	不受抑制
II	[10,35)	以棉花等农作物以及芦苇等草本等为主	稍有抑制
III	[35,60)	芦苇、白刺等草本以及盐穗木、盐节木等灌木	中等抑制
IV	[60,85]	盐穗木、盐节木等灌木以及柽柳、胡杨等乔木	严重抑制
V	> 85	以盐穗木、柽柳和胡杨等乔木为主	极严重抑制

### 3.2 建模变量筛选

本研究的变量包括 6 个植被指数、5 个盐分指数和 7 个波段反射率,共 18 个建模变量。将建模变量分别与 SSC 作相关性分析如表 4 所示,可以看出,NDVI,ENDVI,DVI,EVI,EEVI,NDSI,S<sub>3</sub>,S<sub>5</sub>,SI<sub>4</sub>,SI-T,红光波段反射率、短波红外 1 波段反射率和短波红外 2 波段反射率均与 SSC 的相关性较强,在  $P < 0.01$  水平上具有显著相关性;其次是 CRSI、海岸波段反射率、蓝光波段反射率、绿光波段反射率、近红

外波段反射率与 SSC 的相关性较弱,但同样在  $P < 0.01$  水平上显著相关。

表 4 建模变量与 SSC 的相关性

Tab. 4 Correlation between modeling variables and SSC

变量	NDVI	ENDVI	DVI	CRSI	EVI	EEVI	NDSI	$S_3$	$S_5$
相关系数	-0.637** <sup>①</sup>	-0.605**	-0.607**	-0.504**	-0.619**	-0.601**	0.637**	0.636**	0.625**
变量	$SI_4$	SI-T	海岸波段	蓝光波段	绿光波段	红光波段	近红外波段	短波红外 1 波段	短波红外 2 波段
相关系数	0.614**	0.648**	0.535**	0.551**	0.577**	0.632**	-0.523**	0.638**	0.657**

① \*\* 表示在 0.01 水平极显著相关。

在多变量回归中,为了得到一个简洁有效的预测模型,通常需要进行变量筛选。统计学中常用的变量筛选方法有逐步回归(向前法、后退法),但由于逐步回归法的局限性,不能保证对建模变量的每一个组合进行评估,而全子集回归法则可以一次性对所有可能的建模变量组合进行检验。研究中使用 R 语言 leaps 包中的 regsubsets 函数,采用全子集回归法的 adjr2 (调整的  $R^2$ ) 参数标准,实现最优建模变量组合的筛选。当考虑 CRSI 和 EEVI 作为建模变量时,adjr2 值为 0.47; 当考虑  $SI_4$ 、蓝光波段反射

率和短波红外 2 波段反射率作为建模变量时,adjr2 值为 0.49; 当考虑 DVI,CRSI,EEVI,NDSI, $S_3$ , $S_5$ ,SI-T,蓝光波段反射率和短波红外 1 波段反射率作为建模变量时,adjr2 值达到最大,为 0.51。选择将 adjr2 参数达到最大时的变量组合作为最佳建模变量。

### 3.3 多光谱遥感光谱指数特征分析

依据筛选出与 SSC 相关性高的建模变量包括 DVI,CRSI,EEVI,NDSI, $S_3$ , $S_5$ ,SI-T,蓝光波段反射率和短波红外 1 波段反射率,并获得各建模指标的空间分布如图 2 所示。

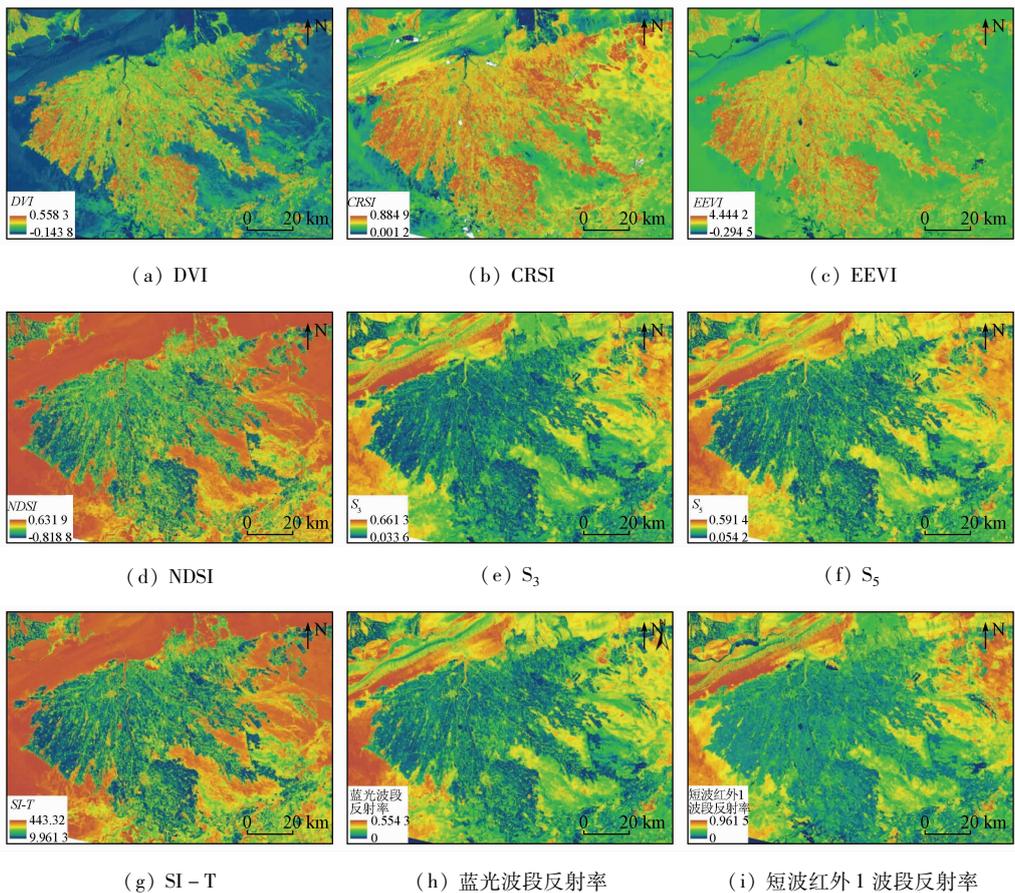


图 2 光谱指数与光谱反射率空间分布

Fig. 2 Spatial distributions of spectral index and spectral reflectance

从图 2 可以看出光谱指数和波段反射率能够反映出该荒漠绿洲地表的光谱特征,由于植被对近红外波段有强烈的反射作用,致使 EEVI 和 CRSI 中绿洲内部呈现出较于绿洲外围荒漠部分更高的数值,

而 DVI 中荒漠和绿洲有明显的差距,是因为植被对红光强烈吸收而近红外强烈反射,因此 DVI 中会发现绿洲和荒漠有清晰的边界;盐分指数  $S_3$  和  $S_5$  中绿洲内部指数值较低,是由于植被对蓝光和红光强烈

吸收作用; NDSI 中绿洲内部反映出的低值与绿洲外部相反,越有植被覆盖的地区值越低; SI - T 中绿洲外围的指数值远远高出内部的值; 蓝光波段反射率中绿洲内部相比绿洲外围荒漠带反射率要低,荒漠带植被覆盖较少,光谱反射特征与裸土相近,绿洲外围的西南部分土壤颜色较浅,波段反射率较高; 短波红外 1 波段反射率相比蓝光波段反射率,绿洲内的植被有较高的反射率,但依旧低于外围的反射率,可能是外围土壤位于荒漠平原颗粒细小,表面平滑,导致具有较高的反射率。结合表 4 和图 2 对 SSC 进行定性分析, DVI, EEVI 和 CRSI 指数中绿洲内部均与 SSC 呈负相关,指数值越高 SSC 越低,而绿洲外围反之。CRSI 指数显示,绿洲内部指数值较高可能造成 SSC 较低,而 DVI 指数中绿洲外围与沙漠的交错带出现极低的负值表明外围部分 SSC 可能会很高。其余 6 个变量均与 SSC 呈正相关,与图中呈现的结果一致,指数值越小的区域 SSC 较低,指数值越大的地区 SSC 较高。

### 3.4 SSC 反演模型的构建与验证

在相关性分析和对研究区 SSC 定性分析的基础上,将 DVI, CRSI, EEVI, NDSI,  $S_3$ ,  $S_5$ , SI - T, 蓝光波段反射率和短波红外 1 波段反射率与 SSC 进行统计分析回归建模和机器学习回归建模分析,由此分别构建 MLR, PLSR, SVMR 和 RFR 这 4 种反演模型,通过  $R^2$ , RMSE 和 RPD 对 SSC 估测结果进行评价,从而寻找出最佳的 SSC 反演模型,精度对比结果如表 5 所示。

表 5 反演模型的精度对比

Tab. 5 Accuracy comparison of inversion models

建模方法	训练集			验证集		
	$R^2$	RMSE/ ( $g \cdot kg^{-1}$ )	RPD	$R^2$	RMSE/ ( $g \cdot kg^{-1}$ )	RPD
MLR	0.480	3.098	1.397	0.599	2.608	1.607
PLSR	0.474	3.114	1.390	0.721	2.512	1.669
SVMR	0.575	2.801	1.545	0.594	2.624	1.598
RFR	0.870	1.550	2.792	0.766	1.991	2.105

MLR 模型使用 R 语言包 stats 中的 lm 函数对训练集进行建模,结果表明,模型整体检验是显著的,因此可以解释因变量的变化。通过 MLR 模型的训练集和验证集的预测结果,可知训练集的  $R^2$  和 RMSE 分别为 0.480 和  $3.098 g \cdot kg^{-1}$ 。研究中利用 plsregress 函数对 PLSR 模型进行训练,通过不断尝试,确定参数  $ncomp = 8$  时,前 3 个主成分分量的贡献率分别为 77.34%, 9.13% 和 5.69%,可以看出前 3 个主成分累计贡献率为 92.16%,模型的精度较好。将训练集和验证集数据使用模型预测,可得训练集的  $R^2$  和 RMSE 分别为 0.474 和  $3.114 g \cdot kg^{-1}$ 。

SVMR 模型的性能取决于核函数及其参数的选择,构建模型使用 R 语言包 e1071,通过训练认为径向基函数 radial 作为核函数效果最佳,  $cost$  和  $gamma$  是非线性 SVM 重要的参数,使用 tune 函数对模型进行优化以寻找最优参数,结果表明。优化后的 SVMR 模型参数为  $cost = 11$ ,  $gamma = 0.09$ ,  $epsilon = 0.1$ ,且使用 40 个支持向量,分别在训练集和验证集上进行预测。并计算训练集和验证集上的平均绝对值误差,结果显示训练集上的绝对误差为 1.302,验证集上的绝对误差为 1.569。通过计算得出训练集的  $R^2$  和 RMSE 分别为 0.575 和  $2.801 g \cdot kg^{-1}$ 。RFR 模型则使用 R 语言包 randomForest 进行训练,并指定决策树数目  $nree = 500$ 。使用训练好的回归模型对验证集进行预测,可以发现,在验证集上的预测 RMSE 只有  $2.178 g \cdot kg^{-1}$ ,这个结果较 MLR, PLSR 和 SVMR 模型的验证集精度要高。但模型的性能除了与  $nree$  有关,还取决于每个决策树包含的变量个数  $mtry$ ,通过使用 tuneRF 函数进一步优化 RFR 模型,来寻找更优的模型参数  $mtry$ ,结果表明,OOBError 取值最小时,  $mtry = 3$ 。再次对验证集进行预测, RMSE 只有  $1.991 g \cdot kg^{-1}$ ,预测误差大大减小,模型的精度得到提升。

通过比较 4 种模型在训练集中的精度,统计分析建模方法与机器学习建模方法在  $R^2$  方面存在显著差异,  $R^2$  最大的是 RFR 模型为 0.870,其次是 SVMR 模型为 0.575,与 SVMR 模型相比  $R^2$  提高了 51.30%,而 MLR 和 PLSR 模型  $R^2$  偏低,分别为 0.480 和 0.474。在 RMSE 方面,MLR 和 PLSR 模型基本一致,分别为 3.098 和 3.114, RFR 模型最低为 1.550,具有最佳的拟合性能,而 SVMR 模型处于二者之间。与其他 3 个模型相比, RFR 模型表现非常好, RMSE 平均降低了 48.4%。在验证集中, RFR 模型具有最大的  $R^2$  和最小的 RMSE; SVMR 模型的估测能力与 MLR 模型相似,但低于 PLSR 模型。此外, MLR, PLSR 和 SVMR 模型验证集的 RPD 值均位于 1.4 ~ 2.0 之间,表明这 3 种模型的预测能力基本可以达到令人满意的结果,可勉强用于 SSC 的估测。然而, RFR 模型的训练集和验证集的 RPD 分别为 2.792 和 2.105,均大于 2,说明 RFR 模型具有良好的估测能力,能够准确地估测出研究区的 SSC。综上分析认为, RFR 模型的拟合能力显著优于 MLR, PLSR 和 SVMR 模型。

基于不同反演模型训练集和验证集的实测值与估测值的散点图如图 3 所示。结合表 3 的分级标准,从图 3 中可以看出,训练集和验证集的样点数据均在各个等级内有分布,突出模型的可行性。从散

点的分布来看, RFR 模型的散点大致在 1:1 线附近, 说明模型的估测值接近于实测值, SVMR 模型少数散点有偏离的倾向, 而从 MLR 模型和 PLSR 模型的散点图来看, 有较大范围的样本点存在偏离 1:1 线的趋势。以总体样本的实测值和估测值计算  $R^2$  和  $RMSE$ , 可以发现 RFR 模型的预测能力最佳, 其

$R^2$  为 0.861, SVMR 模型次之, 其  $R^2$  为 0.644, 同样具有良好的预测能力, 而 MLR 和 PLSR 模型的  $R^2$  分别为 0.521 和 0.525, 模型预测能力一般, 综上所述认为, 4 种反演模型估测效果的优劣排序依次为 RFR > SVMR > PLSR > MLR。

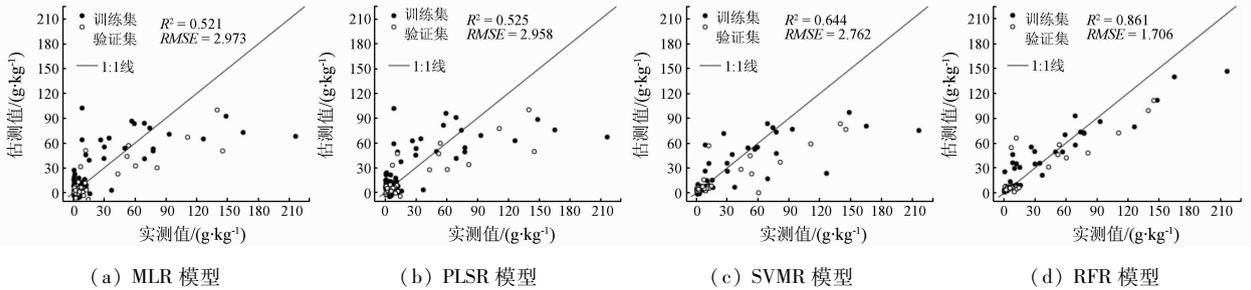


图 3 SSC 不同反演模型的实测值 - 估测值散点图

Fig. 3 Scatter plots of measured and estimated values for different inversion models of SSC

### 3.5 土壤盐渍化空间反演

依据上述建模结果, 由于 RFR 建模方法反演的效果要优于其他建模方法, 因此, 本研究采用 RFR 反演模型使用 R 语言编程结合影像数据对研究区的 SSC 进行空间反演, 并依据表 3 分级标准, 获得研究区 SSC 等级的空间分布情况(图 4)。

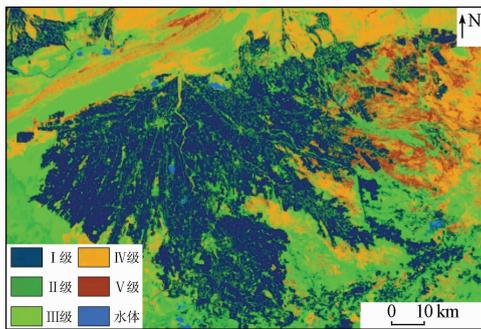


图 4 研究区 SSC 空间分布

Fig. 4 Spatial distribution of SSC in the study area

从图 4 中可以看出, 研究区 SSC 在空间上大致呈现出绿洲内部 SSC 较低, 而绿洲外围 SSC 较高的分布特征。此外, SSC 的等级分布存在明显的空间分异, 这与地表覆盖密切相关, 而建模的光谱因子中, 植被指数和盐分指数能够突显出土地覆盖的空间差异, 所得到的反演结果与地表实际状况吻合。SSC 的空间分布整体表现为自绿洲向外围逐渐增加的趋势, 具体等级空间分布位置以及面积占比为: 第 I 等级分布在绿洲耕地区内部, 所占比例最大, 为 31.87%; 第 II 等级大多呈零星状镶嵌分布在绿洲耕地区, 占研究区总面积较小, 为 9.75%; 第 III 等级较多分布在绿洲外围的西南部和东南部, 其余较少

分布在绿洲内部, 占研究区总面积的 27.48%; 第 IV 等级分布在绿洲边缘和沙漠的交错带之间, 占研究区总面积的 22.83%; 第 V 等级集中分布在绿洲外围东部库车河冲积扇区域, 占比较小, 为 6.10%。绿洲内部 SSC 较低, 而外围 SSC 较高, 这是由于夏季农田植被覆盖程度较高, 而蒸发量相对于外围荒漠区域较低。此外, 绿洲外围 SSC 较高的区域集中分布在东部库车河冲积扇区域, 稍低的区域分布在东南部且靠近河流水系的集中区域, 主要是由于该区地势呈北高南低, SSC 多积聚于地势较低的冲积扇区域, 且地下水和河流汇集在下游, 蒸发量高, 这也是导致绿洲外围 SSC 高的原因之一。

## 4 讨论

本文结果表明干旱区绿洲 SSC 主导因子主要是植被指数和盐分指数, 由于植被与土壤存在密切关系, 通过植被的生长状况可间接地反映出盐分的分布特点, 因此将遥感光谱指数作为训练模型的输入变量。陈红艳等<sup>[13]</sup>、张智韬等<sup>[21]</sup>和苏雯等<sup>[22]</sup>通过提取遥感光谱指数 DVI, CRSI, EEVI 和 SI-T 进行 SSC 反演取得了较好的效果; 厉彦玲等<sup>[23]</sup>的研究进一步发现土壤盐渍化程度低的区域光谱反射率高, 适合进行盐分空间反演。许多学者的研究表明运用机器学习建模比线性关系预测的精度更高, 在本文中机器学习建模同样比传统统计方法建模精度高, 这也反映出植被指数和盐分指数与 SSC 呈现的是非线性关系。同时相比传统统计方法, 在空间分布上也较清晰地表征出其变化特点和趋势。

本研究通过建立 SSC 与基于全子集回归法筛

选的相关性较高的多光谱遥感指数,分别建立 4 种不同的回归模型,进而选择合适的建模变量和拟合度最佳的回归模型方法,构建的 RFR 模型可精确定量反演研究区的 SSC。从杨练兵等<sup>[24]</sup>与冯娟等<sup>[25]</sup>前期的研究来看,研究区内非盐渍土和盐渍化程度较轻的地区位于绿洲的内部,而绿洲外围的盐渍化程度较为严重,本研究反演结果与上述研究结果基本保持一致。由于土壤盐渍化是一个复杂的地表过程,利用光谱指数和光谱反射率进行 SSC 的定量反演将面临诸多困难。综合利用多源遥感数据结合深度学习模型提高 SSC 的估测精度,将会是未来 SSC 定量反演的趋势。

## 5 结论

以渭干河—库车河绿洲为研究区,利用野外实地测量,并结合从多光谱遥感影像中提取的多种遥感光谱指数,选出与 SCC 相关性较高的指数与其实地测量数据分别构建 4 种回归反演模型预测 SSC,探索最佳 SSC 反演模型并实现非线性地理模型在空间尺度上的可视化,可得到如下结论:

1) 基于野外采样点数据结合 Landsat8 影像提取的光谱指数和波段反射率,通过全子集回归法筛选出用于建模的差值植被指数 DVI、冠层响应盐度指数 CRSI、扩展增强型植被指数 EEVI、归一化盐分指数 NDSI、盐分指数  $S_3$ 、盐分指数  $S_5$ 、盐分指数 SI-T、蓝光波段反射率和短波红外 1 波段反射率,这 9 个变量均与 SSC 呈显著的相关关系( $P < 0.01$ )。

2) 通过对比 4 种建模方法的估测结果,RFR 模型精度最高,SVMR 模型次之,PLSR 和 MLR 模型精度最低。最佳模型为 RFR 模型,其训练集  $R^2$ , $RMSE$  和  $RPD$  分别为 0.870,1.550  $g \cdot kg^{-1}$  和 2.792,验证集  $R^2$ , $RMSE$  和  $RPD$  分别为 0.766,1.991  $g \cdot kg^{-1}$  和 2.105,模型估测能力较强,对 SSC 的反演具有较高的准确性。

3) 通过最佳模型对研究区 SSC 进行空间反演分析,该区土壤盐渍化呈现出“内低外高”的空间分异。绿洲内部 SSC 较低,主要以第 I 等级和第 II 等级为主,占研究区总面积的 41.62%; 绿洲外围 SSC 较高,以第 III,IV 和 V 等级为主,共占研究区总面积的 56.41%。

## 参考文献 (References):

[1] Abuelgasim A, Ammad R. Mapping soil salinity in arid and semi-arid regions using Landsat8 OLI satellite data[J]. Remote Sensing Applications: Society and Environment, 2019, 13: 415–425.

[2] Abdenmour M A, Douaoui A, Piccini C, et al. Predictive mapping of soil electrical conductivity as a Proxy of soil salinity in south-east of Algeria[J]. Environmental and Sustainability Indicators, 2020, 8: 100087.

[3] Eswar D, Karuppusamy R, Chellamuthu S. Drivers of soil salinity and their correlation with climate change[J]. Current Opinion in Environmental Sustainability, 2021, 50: 310–318.

[4] Khadim F K, Su H, Xu L, et al. Soil salinity mapping in Everglades National Park using remote sensing techniques and vegetation salt tolerance[J]. Physics and Chemistry of the Earth, 2019, 110: 31–50.

[5] Hammam A A, Mohamed E S. Mapping soil salinity in the East Nile Delta using several methodological approaches of salinity assessment[J]. The Egyptian Journal of Remote Sensing and Space Science, 2020, 23(2): 125–131.

[6] 王遵亲, 祝寿泉, 俞仁培, 等. 中国盐渍土[M]. 北京: 科学出版社, 1993: 7–78.

Wang Z Q, Zhu S Q, Yu R P, et al. Chinese saline soil[M]. Beijing: Science Press, 1993: 7–78.

[7] 周磊, 贺聪聪, 吕爱锋, 等. 柴达木盆地土壤盐渍化程度快速动态监测[J]. 测绘科学, 2021, 46(7): 99–106, 114.

Zhou L, He C C, Lyu A F, et al. Research on quick dynamic monitoring of soil salinization in Qaidam Basin[J]. Science of Surveying and Mapping, 2021, 46(7): 99–106, 114.

[8] 张荣群, 乔月霞, 薛佳妮. 银川平原土壤盐渍化与土地利用强度的空间关联分析[J]. 地球信息科学学报, 2015, 17(5): 598–606.

Zhang R Q, Qiao Y X, Xue J N. Spatial relationship analysis between the soil salinization and land use intensity in Yinchuan Plain[J]. Journal of Geo-Information Science, 2015, 17(5): 598–606.

[9] 张素铭, 赵庚星, 王卓然, 等. 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, et al. Remote sensing inversion and dynamic monitoring of soil salt in coastal saline area[J]. Journal of Agricultural Resources and Environment, 2018, 35(4): 349–358.

[10] 周晓红, 张飞, 张海威, 等. 艾比湖湿地自然保护区土壤盐分多光谱遥感反演模型[J]. 光谱学与光谱分析, 2019, 39(4): 1229–1235.

Zhou X H, Zhang F, Zhang H W, et al. A study of soil salinity inversion based on multispectral remote sensing index in Ebinur Lake wetland nature reserve[J]. Spectroscopy and Spectral Analysis, 2019, 39(4): 1229–1235.

[11] 张同瑞, 赵庚星, 高明秀, 等. 基于近地面多光谱的黄河三角洲典型地区土壤含盐量估算研究[J]. 光谱学与光谱分析, 2016, 36(1): 248–253.

Zhang T R, Zhao G X, Gao M X, et al. Soil salinity estimation based on near-ground multispectral imagery in typical area of the Yellow River Delta[J]. Spectroscopy and Spectral Analysis, 2016, 36(1): 248–253.

[12] 奚雪, 赵庚星, 高鹏, 等. 基于 Sentinel 卫星及无人机多光谱的滨海冬小麦种植区土壤盐分反演研究——以黄三角垦利区为例[J]. 中国农业科学, 2020, 53(24): 5005–5016.

Xi X, Zhao G X, Gao P, et al. Inversion of soil salinity in coastal

- winter wheat growing area based on sentinel satellite and unmanned aerial vehicle multi-spectrum: A case study in Kenli District of the Yellow River Delta [J]. *Scientia Agricultura Sinica*, 2020, 53 (24): 5005 – 5016.
- [13] 陈红艳, 赵庚星, 陈敬春, 等. 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, et al. 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.
- [14] Fathizad H, Ardakani M A H, Sodaiezhadeh H, et al. Investigation of the spatial and temporal variation of soil salinity using random forests in the central desert of Iran [J]. *Geoderma*, 2020, 365: 114233.
- [15] 何宝忠, 丁建丽, 刘博华, 等. 渭库绿洲土壤盐渍化时空变化特征 [J]. *林业科学*, 2019, 55 (9): 185 – 196.
- He B Z, Ding J L, Liu B H, et al. Spatiotemporal variation of soil salinization in Weigan – Kuqa River delta oasis [J]. *Scientia Silvae Sinicae*, 2019, 55 (9): 185 – 196.
- [16] Allbed A, Kumar L. Soil salinity mapping and monitoring in arid and semi-arid regions using remote sensing technology: A review [J]. *Advances in Remote Sensing*, 2013, 2: 373 – 385.
- [17] Wang J Z, Ding J L, Yu D L, et al. Machine learning-based detection of soil salinity in an arid desert region, Northwest China: A comparison between Landsat8 OLI and Sentinel-2 MSI [J]. *The Science of the Total Environment*, 2020, 707 (10): 136092.
- [18] Seifi M, Ahmadi A, Neyshabouri M R, et al. Remote and Vis-NIR spectra sensing potential for soil salinization estimation in the eastern coast of Urmia hyper saline lake, Iran [J]. *Remote Sensing Applications: Society and Environment*, 2020, 20: 100398.
- [19] Liu Y, Zhang F, Wang C, et al. Estimating the soil salinity over partially vegetated surfaces from multispectral remote sensing image using non-negative matrix factorization [J]. *Geoderma*, 2019, 354: 113887.
- [20] 乔木, 周生斌, 卢磊, 等. 新疆渭干河流域土壤盐渍化时空变化及成因分析 [J]. *地理科学进展*, 2012, 31 (7): 904 – 910.
- Qiao M, Zhou S B, Lu L, et al. Causes and spatial-temporal changes of soil salinization in Weigan River basin, Xinjiang [J]. *Progress in Geography*, 2012, 31 (7): 904 – 910.
- [21] 张智韬, 杜瑜燕, 劳聪聪, 等. 基于雷达遥感的不同深度土壤含盐量反演模型 [J]. *农业机械学报*, 2020, 51 (10): 243 – 251.
- Zhang Z T, Du Y Y, Lao C C, et al. Inversion model of soil salt content in different depths based on Radar remote sensing [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51 (10): 243 – 251.
- [22] 苏雯, 丁建丽, 杨爱霞. 基于 GF-1 影像的渭-库绿洲外围土壤含盐量定量反演研究 [J]. *中国农村水利水电*, 2017 (2): 9 – 13.
- Su W, Ding J L, Yang A X. Quantitative inversion of soil salinity in Weigan – Kuqa River oasis based on GF-1 image [J]. *China Rural Water and Hydropower*, 2017 (2): 9 – 13.
- [23] 厉彦玲, 赵庚星, 常春艳, 等. OLI 与 HSI 影像融合的土壤盐分反演模型 [J]. *农业工程学报*, 2017, 33 (21): 173 – 180.
- Li Y L, Zhao G X, Chang C Y, et al. Soil salinity retrieval model based on OLI and HSI image fusion [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2017, 33 (21): 173 – 180.
- [24] 杨练兵, 陈春波, 郑宏伟, 等. 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, et al. 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.
- [25] 冯娟, 丁建丽, 魏雯瑜. 基于 Albedo-MSAVI 特征空间的渭库绿洲土壤盐渍化研究 [J]. *中国农村水利水电*, 2018 (2): 147 – 152.
- Feng J, Ding J L, Wei W Y. A study of soil salinization in Weigan and Kuqa Rivers oasis based on Albedo – MSAVI feature space [J]. *China Rural Water and Hydropower*, 2018 (2): 147 – 152.

## Inversion of soil salinity of an oasis in an arid area based on Landsat8 OLI images

HUANG Xiaoyu<sup>1</sup>, WANG Xuemei<sup>1,2</sup>, KAWUQIATI Baishan<sup>1</sup>

(1. College of Geographic Science and Tourism, Xinjiang Normal University, Urumqi 830054, China; 2. Xinjiang Uygur Autonomous Region Key Laboratory “Xinjiang Arid Lake Environment and Resources Laboratory”, Urumqi 830054, China)

**Abstract:** The rapid detection of soil salinity using remote sensing technology can scientifically guide the soil salinization control and the rational development of oasis agriculture. Based on 95 soil samples from the oasis of the Weigan – Kuqa River delta, this study established four soil salinity estimation models of multiple linear regression, partial least squares regression (PLSR), support vector machine regression (SVR), and random forest regression using the spectral index, band reflectance, and the measured soil salinity. Then, it conducted the remote sensing inversion for the spatial distribution pattern of the soil salinity in the study area using the optimal estimation results. The results are as follows: ① Nine spectral factors that were significantly related to soil salinity were screened using

the all - subsets regression method, with correlation coefficients of all above 0.5 ( $P < 0.01$ ). Among them, the correlation coefficient between salinity index SI - T and the soil salinity was the highest (0.648); ② The comparison of estimation precision show that the fitting effect of the four inversion models was in the order of random forest regression > SVR > PLSR > multiple linear regression. Among these models, the random forest model had the best fitting precision. Its training and validation sets had coefficients of determination ( $R^2$ ) of 0.870 and 0.766, respectively, with relative percent deviation (RPD) of 2.792 and 2.105, respectively, both of which were greater than 2. These results indicate that the random forest model had a good inversion effect and stable estimation capacity; ③ According to the inversion results of the random forest model, grade I and II zones account for 41.62% and are distributed in the cultivated land area inside the oasis; grade III, IV, and V zones account for 56.41% and are primarily distributed in the desert and the desert - oasis ecotones. Therefore, compared with conventional statistical models, the random forest modeling method can yield significantly better estimation effects in the inversion of soil salinity. This study can be used as a reference for the monitoring of soil salinization in oases in arid areas.

**Keywords:** spectral index; machine learning model; soil salinity; remote sensing inversion; oasis in an arid area  
(责任编辑: 陈 理)