

doi:10.6041/j.issn.1000-1298.2019.12.017

基于无人机多光谱遥感的土壤含盐量反演模型研究

张智韬^{1,2} 魏广飞^{1,2} 姚志华^{1,2} 谭丞轩^{1,2} 王新涛^{1,2} 韩佳^{1,2}

(1. 西北农林科技大学水利与建筑工程学院, 陕西杨凌 712100;

2. 西北农林科技大学旱区农业水土工程教育部重点实验室, 陕西杨凌 712100)

摘要: 为探究无人机多光谱遥感技术快速监测植被覆盖下的土壤含盐量问题, 以内蒙古河套灌区沙壕渠灌域内4块不同盐分梯度的耕地为研究区域, 利用无人机搭载多光谱传感器获取2018年8月遥感影像数据, 并对0~40 cm的土壤进行盐分测定。分别引入敏感波段组、光谱指数组、全变量组作为模型输入变量, 采用支持向量机(Support vector machine, SVM)、BP神经网络(Back propagation neural network, BPNN)、随机森林(Random forest, RF)、多元线性回归(Multiple linear regression, MLR)4种回归方法, 建立基于3组输入变量下的土壤盐分反演模型, 并进行精度评价, 比较不同输入变量、不同回归方法对模型精度的影响, 评价并优选出最佳盐分反演模型。结果表明, 通过分析3个变量组的 R^2 和RMSE, 光谱指数组在4种回归方法中均取得了最佳的反演效果, 敏感波段组和全变量组在不同的回归方法中反演效果不同。4种回归方法中, 3种机器学习算法反演精度明显高于MLR模型, 且MLR模型中的敏感波段组和全变量组均出现了“过拟合”现象, RF算法在3种机器学习算法中表现最优, SVM算法和BPNN算法在基于不同变量组的模型中表现也不相同。基于光谱指数组的RF的盐分反演模型在12个模型中取得了最佳的反演效果, R^2 和 R^2_v 分别达到了0.72和0.67, RMSE_v仅为0.112%。

关键词: 土壤含盐量; 无人机; 多光谱遥感; 变量组; 机器学习; 多元线性回归

中图分类号: S127 **文献标识码:** A **文章编号:** 1000-1298(2019)12-0151-10

Soil Salt Inversion Model Based on UAV Multispectral Remote Sensing

ZHANG Zhitao^{1,2} WEI Guangfei^{1,2} YAO Zhihua^{1,2} TAN Chengxuan^{1,2} WANG Xintao^{1,2} HAN Jia^{1,2}

(1. College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China

2. The Key Laboratory of Agricultural Soil and Water Engineering in Arid Areas, Ministry of Education, Northwest A&F University, Yangling, Shaanxi 712100, China)

Abstract: Fast acquisition of soil salt content under vegetation cover is the objective need of saline soil management and utilization. Four kinds of croplands with different salinization values in Shahaoqu District of Hetao Irrigation Area were set as the study areas. The UAV equipped with a multi-spectral camera obtained the remote sensing image data of August, meanwhile, the soil salinity with depth of 0~40 cm was tested. The sensitive band group, spectral index group and full variable group were introduced as model input variables. Four regression methods, including support vector machine (SVM), BP neural network (BPNN), random forest (RF) and multiple linear regression (MLR), were used to establish soil salinity inversion models which were based on three groups of input variables, respectively. Firstly, the model precision was evaluated, and then the effects of different input variables and different regression methods on the model precision were compared, finally the best salt inversion model was evaluated and optimized. The results indicated that comparing the R^2 and RMSE of three variable groups, the spectral index group achieved the best inversion effect between the four regression model methods, and the sensitive band group and the full variable group had advantages and disadvantages in different regression algorithms. Between the four regression methods, the inversion accuracy of three machine learning regression algorithms was significantly higher than that of the MLR model. Moreover, both the sensitive band group and the full variable group in the MLR model showed the phenomenon of “overfitting”. And RF algorithm performed best between the three machine learning algorithms. Besides,

收稿日期: 2019-04-23 修回日期: 2019-06-14

基金项目: 国家重点研发计划项目(2017YFC0403302)、杨凌示范区科技计划项目(2018GY-03)和国家自然科学基金项目(41502225)

作者简介: 张智韬(1976—), 男, 副教授, 博士, 主要从事遥感技术在节水灌溉及水资源中的应用研究, E-mail: zhitaozhang@126.com

SVM algorithm and BPNN algorithm performed better and worse in the model with different variable groups. The RF salt inversion model based on the spectral index group achieved the best inversion effect among the 12 models, the R_c^2 and R_v^2 reached 0.72 and 0.67, respectively, and the RMSE_v error was only 0.112%. The research result can provide a theoretical reference for soil salinity monitoring in arid and semi-arid areas.

Key words: soil salt; unmanned aerial vehicle; multispectral remote sensing; variable group; machine learning; multiple linear regression

0 引言

土壤盐渍化已成为干旱-半干旱地区最严重的土壤问题之一,严重制约了区域经济健康发展^[1]。快速、准确地获取盐渍土时空分布信息成为合理防治盐渍化的必要前提。传统的土壤含盐量测定方法费时、费力、成本高,且无法全面获取数据。近年来,随着光谱分析技术的发展,遥感技术在农业中的应用越来越广泛,尤其在土壤盐渍化监测方面^[2-7]。已有研究结果显示,在干旱和半干旱地区,植被的生长状况可以间接反映土壤盐渍化程度^[8-10]。ALLBED等^[11]研究表明,土壤含盐量超过植被生长阈值时,致使叶片构造发生变化,在可见光和近红外波段,植被冠层光谱反射率对土壤含盐量响应敏感。ZHANG等^[12]利用MODIS卫星影像数据构建增强性植被指数和EVI,反演了黄河三角洲植被区土壤盐分。陈红艳等^[13]基于Landsat8 OLI多光谱影像引入第七波段对植被指数进行改进,构建土壤含盐量的支持向量机模型,获得了良好的土壤盐分空间分布反演效果。以上研究大多以植被指数和盐分指数作为指示因子建立与实测土壤含盐量的关系^[14-17]。王飞等^[18]通过引入波段组、植被指数变量组、优选变量组等6个变量组,进行多变量分析,通过算法筛选植被指数、土壤盐度指数等盐度敏感变量,得出最优变量组合,并利用多个机器学习算法预测绿洲土壤盐分,研究发现,在涉及的变量个数较多、且与各土壤含盐量相关性较低时,机器学习算法的挖掘能力得到了充分的体现。BP神经网络、支持向量机和随机森林是3种常用的机器学习算法,BP神经网络算法具有非线性拟合能力和自学习能力,支持向量机算法可以避免从归纳到演绎的传统过程,随机森林算法具有可处理非线性数据、实现简单、训练速度快、抗拟合能力强的特点。王明宽等^[16]研究表明,遥感影像的反射率与土壤含盐量并不是单纯的线性关系,BP神经网络模型能很好地模拟土壤含盐量与光谱数据的关系。厉彦玲等^[19]利用遥感数据建立统计分析模型(多元线性回归、偏最小二乘回归),用机器学习模型(BP神经网络、支持向量机模型和随机森林)反演盐分,结果表明,

3种机器学习模型反演效果均取得了较高的精度,且反演效果远优于统计分析模型。

国内外学者对区域盐分定量监测已取得较多的成果,但这些成果多偏向于卫星遥感和高光谱遥感,IVUSHKIN等^[20]通过在无人机上装载热成像、高光谱、激光雷达3种传感器发现,无人机遥感在农田盐分监测方面有较大潜力。但基于无人机多光谱遥感监测植被覆盖下的土壤含盐量研究较少,同时缺乏无人机多光谱遥感对植被覆盖下的土壤含盐量反演模型。

本文以河套灌区的沙壕渠灌域为研究区域,基于无人机搭载多光谱相机获取的遥感数据和同步测定的土壤盐分数据,引入敏感波段组、光谱指数组 and 全变量组作为模型输入变量,利用多元线性回归、BP神经网络、支持向量机和随机森林算法,构建土壤盐分的定量反演模型,评定不同模型输入变量和不同模型方法的土壤含盐量反演模型的精度,优选土壤含盐量最佳反演模型,以期对干旱及半干旱地区无人机遥感土壤盐分定量反演提供参考。

1 材料与方法

1.1 研究区概况

沙壕渠灌域是内蒙古河套灌区西北部解放闸灌域内部的一个独立单元,位于40°52′~41°00′N,107°05′~107°10′E,灌域形状近似为一个狭长的倒三角形(图1),南窄北宽,南北约为15 km,东西平均宽度约4 km,地面较为平整,地势走向为南高北低,海拔为1 034~1 037 m,总控制面积为52.4 km²,当地属干旱-半干旱气候,冬季严寒少雪,夏季高温少

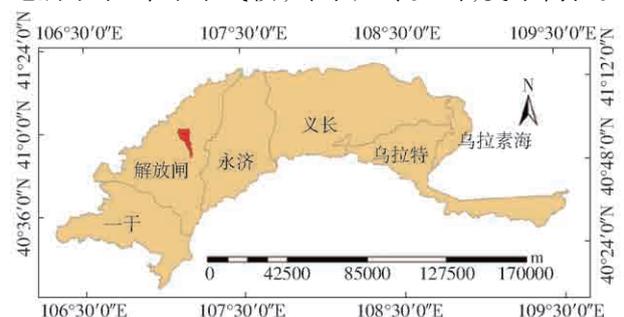


图1 研究区位置示意图

Fig.1 Sketch of study area

雨, 年均降雨量 140 mm, 年均蒸发量 2 000 mm, 年平均气温 7℃。灌区土壤类型以粉壤土、砂壤土和壤土为主。研究区无霜期为 120 ~ 150 d, 主要农作物包括小麦、玉米、葵花、白菜等。沙壕渠灌域南部盐渍化程度相对较轻, 作物种植以小麦、玉米为主, 而北部盐渍化程度较重, 以种植耐盐的向日葵为主。

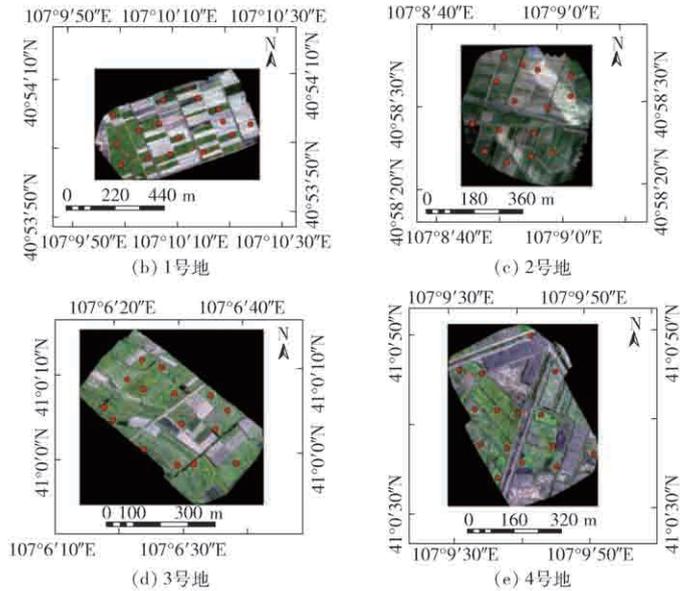
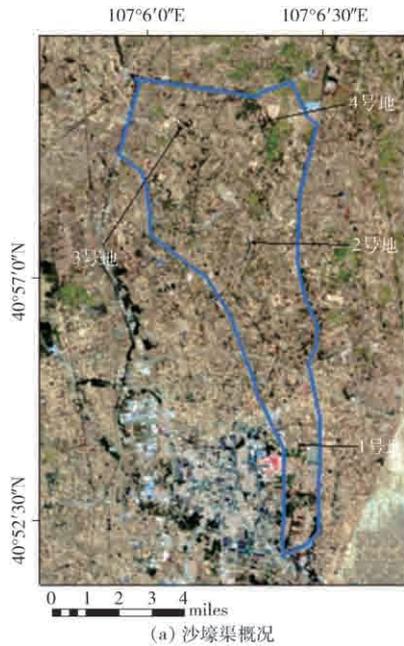


图 2 采样点分布示意图

Fig. 2 Sampling point distribution maps

1.3 多光谱遥感影像的获取与处理

1.3.1 多光谱遥感影像获取

所用遥感平台为大疆创新公司生产的经纬 M600 型六旋翼无人机, 其搭载的传感器为美国 Tetracam 公司生产的 Micro-MCA 多光谱相机 (简称 MCA), 具有质量轻、体积小及远程触发的特点, 非常适合在中小型无人机上进行搭载及拍摄, MCA 相机包括 490 nm (蓝光)、550 nm (绿光)、680 nm (红光)、720 nm (红边)、800 nm (近红外)、900 nm (近红外) 6 个波长的光谱采集通道。2018 年 8 月 12—16 日于河套灌区沙壕渠灌域进行试验, 天气情况分别为: 晴转多云、晴、晴、晴、晴转多云, 影像获取时间选为 11:00—15:00 之间 (均在晴朗时拍摄), 无人机飞行模式按照提前规划的航线飞行, 拍照时相机镜头与地面呈 90°, 拍照模式为等时间间隔, 主航线上和主航线间图像重叠率均设置为 80% 以上, 飞行高度经多次试飞后选定为 120 m (此时影像地面分辨率为 6.5 cm, 影像刈幅宽度为 66.5 m), 采集光谱前用标准白板进行标定, 4 块区域分别得到 341、320、390、344 幅遥感图像。所用无人机与多光谱相机如图 3 所示, 无人机飞行路线 (以 4 号地为例) 及现场作业如图 4 所示。

1.2 试验设计

2018 年 5 月, 对沙壕渠灌域的耕地进行实地调研, 根据土壤的盐渍化程度, 在灌域确定 4 块不同盐渍化梯度的试验区域, 并依次进行编号, 每块研究区域为 16 hm², 每块区域均匀布设 15 个作物 (玉米、西葫芦和葵花) 覆盖下的地面数据采集点, 共计 60 个采样点, 如图 2 所示。



(a) 大疆经纬 M600 型无人机 (b) Micro-MCA 多光谱相机

图 3 M600 型无人机及 Micro-MCA 多光谱相机

Fig. 3 M600 unmanned aerial vehicle and micro-MCA multispectral camera



(a) 飞行路线 (b) 现场作业

图 4 无人机飞行路线及现场作业图

Fig. 4 UAV flight path and application scene of UAV over cropland

1.3.2 多光谱遥感影像处理

利用多光谱相机配套的 pixel Wrench2 软件对

获取的多光谱正射影像进行提取、配准与合成,然后将获得的 TIF 格式的遥感图像和每幅图像的 GPS 数据导入 pix4Dmapper 软件中完成 4 块研究区域的图像拼接,把拼接完成的图像输入 ENVI Classic 中作进一步处理,为排除单一点反射率产生的随机误差,由地面数据采集点的 GPS 裁剪出以采集点为中心的 200 像素 × 200 像素图像,采用监督分类的最大似然法剔除土壤背景后,将此区域 6 个波长平均灰度除以标准白板的灰度作为该采集点对应 6 个波长的反射率。

1.4 土壤盐分数据获取

每块研究区域均匀布设 15 个地面数据采集点,为保证遥感影像与地面数据的时间一致性,在获取多光谱遥感影像的同时,用对角线 5 点取样法在作物根系附近采集深度为 0 ~ 20 cm、20 ~ 40 cm 的土样,并用手持式 GPS 仪记录每个采样点的位置,取出的土样放置在标记的铝盒中,并带回实验室干燥,采用土水比 1:5 法配置土壤溶液,经过滤后静置 8 h,采用电导率仪(雷磁 DDS-307A 型,上海佑科仪器分公司)测定土壤溶液电导率($EC_{1:5}$, dS/cm),并通过经验公式计算土壤含盐量(SSC,%)^[21]:

($SSC = 0.288 2EC_{1:5} + 0.018 3$)。该时期作物的根系主要活动层深度为 0 ~ 40 cm^[22],故将此深度的含盐量作为该试验的土壤含盐量数据(此时,0 ~ 40 cm 的含盐量为 0 ~ 20 cm、20 ~ 40 cm 含盐量的平均值)。

1.5 光谱指数构建

采用遥感影像提取的各类光谱指数构建特征空间进行土壤含盐量反演和检测是目前土壤盐分遥感监测的主要方法之一。本研究选择 10 个光谱指数,分别为归一化植被指数(Normalized difference vegetation index, NDVI)、差值植被指数(Difference vegetation index, DVI)、增强性植被指数(Enhanced vegetation index, EVI)、增强性植被指数 2(Enhanced vegetation index 2, EVI2)、三角形植被指数(Triangular vegetation index, TVI)、土壤调节植被指数(Soil-adjusted vegetation index, SRVI)、归一化差异绿度植被指数(Normalized difference greenness vegetation index, NDGI)、归一化盐分指数(Normalized difference soil index, NDSI)、盐分指数(Salinity index, SI-T)、简单比值指数(Simple ratio index, SR),其计算公式如表 1 所示。

表 1 光谱指数及相关计算公式

Tab. 1 Spectral indexes and related formula

光谱指数	计算公式	文献序号
归一化植被指数(NDVI)	$NDVI = (NIR - R) / (NIR + R)$	[18]
差值植被指数(DVI)	$DVI = NIR - R$	[23]
增强性植被指数(EVI)	$EVI = 2.5(NIR - R) / (NIR + 6R - 7.5B + 1)$	[18]
增强性植被指数 2(EVI2)	$EVI2 = 2.5(NIR2 - R) / (NIR2 + 6R - 7.5B + 1)$	[18]
三角形植被指数(TVI)	$TVI = 0.5[120(NIR - G) - 200(R - G)]$	[23]
土壤调节植被指数(SRVI)	$SRVI = (1 + L)(NIR - R) / (NIR + R + L)$	[23]
归一化差异绿度植被指数(NDGI)	$NDGI = (G - R) / (G + R)$	[24]
归一化盐分指数(NDSI)	$NDSI = (R - NIR) / (R + NIR)$	[11]
盐分指数(SI-T)	$SI - T = 100(R - NIR)$	[11]
简单比值指数(SR)	$SR = NIR / R$	[18]

注: G 、 R 、 NIR 、 $NIR2$ 分别为 550、680、800、900 nm 波长处的光谱反射率。 L 为盖度背景调节因子,取 0.5。

1.6 盐分反演模型构建

将 60 个研究样本按照含盐量由小到大排序分组,根据建模集与验证集为 2:1 的比例进行等间隔取样。选取 40 个样本用于建模,其余 20 个样本用于验证。

本研究将模型输入变量分为 3 组:敏感波段变量组、光谱指数变量组和全变量组。采用多元线性回归模型、支持向量机、反神经网络和随机森林等 4 种回归方法分别建立基于 3 种模型输入变量的土壤盐分反演模型,共计 12 个盐分反演模型。本研究的 MLR 建模与分析使用 IBM SPSS 23 软件完成,

3 种机器学习算法模型均在 R3.4.0 软件中完成。这 4 种回归方法如下:

(1) 多元线性回归

MLR 是用于两个或两个以上自变量时的一种多元回归分析,由多个自变量的最优组合共同来预测因变量,比只用一个自变量进行预测或估计更有效、更符合实际^[16],是遥感盐分反演最常用的回归方法之一^[25-26]。

(2) 支持向量机

SVM 是根据统计学理论,以结构风险最小化为原则为理论基础从线性可分扩展到线性不可分的一

种新的机器学习方法,在图像识别与分类上应用较广,近年来在回归问题上也得到了一些应用^[27-28]。本文设定核函数类型为“poly”,采用训练集交叉验证和网络搜索法(Grid search)进行参数寻优,根据方差最小原则确定惩罚系数为 $C = 10\ 000$, $\gamma = 0.01$ 。

(3) 反神经网络

BPNN 是一种按误差逆传播算法训练的多层前馈神经网络,也已应用到了盐分反演问题中^[29-31]。本文采用 3 层网络拓扑结构构建 BPNN 模型,按照训练结果误差相对较小的原则,权值的衰减参数和隐层节点数通过多次反复试验确定,最终确定衰减参数为 0.001、隐层节点数为 4,其他参数均使用默认值。

(4) 随机森林

RF 算法是将 Bagging 算法与决策树算法进行结合所得到的集成学习算法,近年来多位学者将其应用到遥感技术中^[31-32],此算法主要参数为决策树的个数,采用训练集交叉验证和网络搜索法(Grid search)进行决策树个数寻优,经过多次训练确定决策树个数为 4。

1.7 反演模型检验与评价

利用 20 个验证样本对各类方法构建的 12 个盐分反演模型进行检验分析,采用验证集将模型预测值和实测值进行拟合,对模型建模和验证的精度采用决定系数(Coefficient of determination, R^2)和均方根误差(Root mean square error, RMSE)两个指标来评价。 R^2 越接近 1, RMSE 越小,说明模型精度效果越好。

2 结果与分析

2.1 土壤盐分统计特征

将 60 个土样测得的土壤含盐量分为 3 个等级^[21],即:非盐渍化($< 0.2\%$)、轻度盐渍化($0.2\% \sim 0.5\%$)和重度盐渍化($0.5\% \sim 1.0\%$),土壤含盐量的总样本、建模集、验证集的分析结果如表 2 所示。河套灌区总样品统计结果显示,非盐渍土、轻度盐渍土、重度盐渍土占比分别为 16.7%、63.3%、20%,且变异系数为 0.49,表明表层土壤含盐量呈现中度变异性。

表 2 土壤盐分的描述性统计分析

Tab. 2 Descriptive statistical analysis on soil salinity

数据集	样本数	非盐渍化	轻度盐渍化	重度盐渍化	最小值/%	最大值/%	均值/%	标准差/%	变异系数
总样本	60	10	38	12	0.11	0.89	0.37	0.18	0.49
建模集	40	7	25	8	0.11	0.81	0.36	0.17	0.47
验证集	20	3	13	4	0.12	0.89	0.38	0.19	0.50

2.2 土壤含盐量与多光谱六波段光谱反射率、光谱指数的相关性分析

在干旱和半干旱地区,生长在盐渍化环境中的植物,其根系接触到土壤溶液中各种可溶性盐离子,当阴阳离子超过一定含量时,就会对植物生长产生各种危害,即土壤盐渍化会在一定程度上影响地表作物生长长势,进而影响冠层的光谱特征^[33]。分别

对多光谱六波段反射率、不同光谱指数分别与土壤含盐量进行相关性分析,根据皮尔逊(Person)相关系数界值表, $n = 60$ 时,当 $|r| > 0.250$ 时,表示在 0.05 水平上显著, $|r| > 0.325$ 时,表示在 0.01 水平上显著, $|r| > 0.408$ 时,表示在 0.001 水平上显著,二者与土壤盐分的 Pearson 相关系数如表 3、4 所示。

表 3 多光谱六波段反射率与 SSC 的相关系数($n = 60$)

Tab. 3 Correlation coefficient of reflectivity in six bands of multispectral and SSC($n = 60$)

波段	B1(490 nm)	B2(550 nm)	B3(680 nm)	B4(720 nm)	B5(800 nm)	B6(900 nm)
相关系数 $ r $	0.321 *	0.334 **	0.515 ***	0.094	0.565 ***	0.615 ***

注: * 表示显著性检验 $p < 0.05$, ** 表示 $p < 0.01$, *** 表示 $p < 0.001$,下同。

从表 3 可以看出,除了红边波长(720 nm)处相关性仅为 0.094 外,其他 5 个波长与 SSC 均表现出一定的相关性。在红光波长(680 nm)、近红外波长(800 nm)和近红外波长(900 nm)处显著性检验 $p < 0.001$,且相关系数均达到了 0.5 以上。六波长与 SSC 相关性由大到小依次为 B6、B5、B3、B2、B1、B4。

从表 4 可看出,选取的 10 种光谱指数与 SSC 均呈现出良好的相关性,相关性由大到小依次为 SI-T、

TVI、EVI2、SRVI、NDVI、EVI、NDGI、NDSI、SR、DVI。其中,TVI、EVI2 和 SI-T 相关系数均超过了 0.6, NDVI 和 SRVI 相关系数都接近 0.6,说明 SI-T、TVI、EVI2、SRVI、NDVI 在一定程度上可表征土壤含盐量。

2.3 土壤盐分反演模型

将模型输入变量分为 3 组。通过 2.2 节中多光谱六波长与 SSC 的相关性分析,选择与土壤盐分极

表4 光谱指数与SSC的相关系数($n=60$)Tab.4 Correlation coefficient of spectral indexes and SSC($n=60$)

光谱指数	相关系数 $ r $
NDVI	0.576***
DVI	0.452***
EVI	0.555***
EVI2	0.601***
TVI	0.624***
SRVI	0.596***
NDGI	0.541***
NDSI	0.513***
SI-T	0.627***
SR	0.482***

显著且 Person 相关系数大于 0.5 的 B3、B5 和 B6 波段,将这 3 个敏感波段作为第 1 组模型输入变量,记

表5 基于不同变量组的多元线性回归模型

Tab.5 MLR model based on different groups of variables

变量组	表达式	建模集($n=40$)		验证集($n=20$)	
		R_c^2	RMSE _c /%	R_v^2	RMSE _v /%
敏感波段	$Y = -0.37X_1 + 0.542X_2 - 0.173X_3 + 0.292$	0.47	0.138	0.26	0.168
光谱指数	$Y = 2.337X_1 - 0.008X_2 + 1.673X_3 + 0.032X_4 - 6.676X_5 + 0.294$	0.56	0.129	0.47	0.141
全变量	$Y = -0.101X_1 + 0.458X_2 - 0.227X_3 + 2.561X_4 - 0.01X_5 + 1.302X_6 + 0.028X_7 - 6.604X_8 + 0.263$	0.64	0.119	0.26	0.184

注:敏感波段变量组中: X_1 、 X_2 、 X_3 分别表示 B3、B5 和 B6 波长对应的反射率;光谱指数组中: X_1 、 X_2 、 X_3 、 X_4 、 X_5 分别表示 SI-T、TVI、EVI2、SRVI、NDVI;全变量组中: X_1 、 X_2 、 X_3 、 X_4 、 X_5 、 X_6 、 X_7 、 X_8 分别表示 B3、B5、B6、SI-T、TVI、EVI2、SRVI、NDVI。

量组的 R_c^2 达到了 0.64,光谱指数组次之,敏感波段组的 R_c^2 最差,仅为 0.47,呈现出自变量个数越多,建模精度效果越优的趋势。同时可以看出,全变量组的 RMSE_c 为 0.119%,光谱指数组的 RMSE 次之,敏感波段的误差最大,为 0.138%。综合发现,全变量组作为模型输入变量时取得了最佳的建模效果。

对上述模型进行精度检验时,敏感波段组和全变量组的 R_c^2 都为 0.26,且 RMSE_v 分别高达 0.168%、0.184%,此两组的 R_c^2 与 R_v^2 相差较大,即出现了较严重的“过拟合”现象,这可能是由于作物受到盐分胁迫后,冠层的光谱反射特性与土壤盐分存在着复杂的非线性函数关系。而在光谱指数变量组, R_v^2 和 RMSE_v 分别达到了 0.47 和 0.141%,均为 3 个模型输入变量组中的最优值。

综合对比 MLR 模型的 3 个变量输入组建模和验证参数结果,发现基于光谱指数组建立的 MLR 模型较为稳定,是 3 个变量组中最佳的多元线性回归盐分反演模型。

2.3.2 基于 3 个模型输入变量组的机器学习盐分反演模型

将敏感波段、光谱指数、全变量组作为模型输

为敏感波段组;通过 2.2 节中 10 种光谱指数与 SSC 的相关性分析,选择 SI-T、TVI、EVI2、SRVI 和 NDVI 共 5 个相关性较高的光谱指数作为第 2 组模型输入变量,记为光谱指数组;为了尽量减少因重要变量缺失而引起的模型偏差,将上述 3 个敏感波长和 5 个敏感光谱指数作为第 3 组模型输入变量,共计 8 个变量,记为全变量组。

2.3.1 基于 3 个变量组的多元线性回归土壤盐分反演模型

将敏感波段组、光谱指数和全变量组分别作为模型输入变量,建立基于多元线性回归的盐分反演模型,并对模型进行精度验证,基于 3 种变量组的多元线性回归模型结果见表 5。

通过表 5 可以看出,3 个模型输入变量组构建的 MLR 模型决定系数 R^2 不小于 0.47。其中,全变

入变量,分别建立 SVM、BPNN、RF 等 3 种机器学习回归模型并进行模型精度验证,模型结果如表 6 所示。

表6 基于不同变量组的机器学习模型

Tab.6 Machine learning models based on different variable groups

变量组	模型 算法	建模集($n=40$)		验证集($n=20$)	
		R_c^2	RMSE _c /%	R_v^2	RMSE _v /%
敏感波段	SVM	0.48	0.129	0.44	0.200
	BPNN	0.55	0.116	0.53	0.133
	RF	0.63	0.107	0.61	0.118
光谱指数	SVM	0.59	0.113	0.57	0.132
	BPNN	0.65	0.102	0.57	0.148
	RF	0.72	0.094	0.67	0.112
全变量	SVM	0.51	0.123	0.53	0.141
	BPNN	0.48	0.126	0.42	0.146
	RF	0.72	0.093	0.56	0.135

由表 6 可以发现,同一机器学习算法下,光谱指数组的 R_c^2 和 R_v^2 均是 3 个输入变量组中的最高值,都不小于 0.57,且其所建模型的 R_c^2 和 R_v^2 都比较接近,说明模型并未发生“过拟合”现象, RMSE_v 平均为 0.131%,亦为 3 个变量组中的最低值。在 SVM 算法中,敏感波段组的 R_c^2 和 RMSE_v 分别为 0.44 和

0.200%, 全变量组 R_v^2 和 $RMSE_v$ 分别为 0.53 和 0.141%, 此时敏感波段组模型精度表现差于全变量组; 而在 BPNN 和 RF 算法中, 敏感波段组 R_c^2 和 R_v^2 分别为 0.53 和 0.61, 高于全变量组的 0.42 和 0.56, 且敏感波段组的 $RMSE_c$ 和 $RMSE_v$ 分别为 0.133%、0.118%, 低于全变量组的 $RMSE_c$ 和 $RMSE_v$ 。在 3 个变量组中, 光谱指数组模型精度最佳; 而对比全变量组, 敏感波段组在 BPNN 算法和 RF 算法中表现出一定优势, 但在 SVM 算法中, 敏感波段组精度表现次于全变量组。

从表 6 可以看出, 同一变量组作为模型输入变量下, RF 算法的 R_c^2 和 R_v^2 都为 3 个算法中最高值, 且均没有出现“过拟合”现象, 在 3 个输入变量组模型中, R_v^2 分别达到了 0.61、0.67 和 0.56, $RMSE_v$ 分别仅为 0.118%、0.112% 和 0.135%, 均为 3 个算法中的最低值。在光谱指数组构建的模型中, SVM 算法和 BPNN 算法 R_v^2 均为 0.57, 二者的 $RMSE_v$ 仅相差 0.016%; 在敏感波段组中, BPNN 算法的 R_c^2 为 0.55, 高于 SVM 算法中的 0.48, 且 BPNN 算法 $RMSE_v$ 为 0.133%, 低于 SVM 算法中的 0.200%; 在全变量组中, SVM 算法 R_v^2 为 0.53, 高于 BPNN 算法中 R_v^2 的 0.42, 同时 SVM 算法中 $RMSE_v$ 为 0.141%, 低于 BPNN 算法 $RMSE_v$ 的 0.146%。由此得出, 在 3 种机器学习算法中, 基于 RF 算法的盐分模型反演精度效果最佳。

2.4 盐分反演模型的综合评价分析

采用敏感波段组、光谱指数组和全变量组作为模型输入变量组, 分别使用 MLR、SVM、BPNN、RF 共 4 种回归方法, 共构建 12 个土壤盐分反演模型。将 12 个反演模型的土壤盐分预测值与验证集土壤盐分实测值进行比较, 结果如图 5 所示。

由图 5 可知, 整体而言, 除去基于敏感波段组、全变量组构建的 2 个多元线性回归模型外, 其余 10 个模型 R_v^2 都达到了 0.4 以上, 其 R_c^2 与 R_v^2 都较为接近, 且 $RMSE_v$ 也处于合理的范围内, 表现出了良好的稳定性。比较 3 个模型输入变量组对模型精度效果的影响发现, 光谱指数组在 3 个变量组中呈现出明显的优势, 即使在多元线性回归模型中, 其 R_v^2 也达到了 0.47, 在 3 种机器学习回归模型中亦表现最佳; 敏感波段组和全变量组在多元线性回归模型中均出现了“过拟合”现象, 二者在 3 种机器学习模型中精度互有高低, 并无明显差距。比较 4 种回归方法对模型精度效果的影响发现, 3 种机器学习算法模型相比多元线性回归模型精度明显提高, 且表现出了较好的稳定性; 在 3 种机器学习算法中, RF 算法表现最优, 在基于 3 个模型输入变量组中 R_v^2 分别

达到了 0.61、0.67 和 0.56, $RMSE_v$ 分别为 0.118%、0.112% 和 0.135%, 均为 3 个机器学习算法中的最优值; RF 算法在 4 种回归方法中表现最优, MLR 表现最差, SVM 算法和 BPNN 算法表现较为接近。对比分析 12 个模型精度发现, 基于光谱指数组的 RF 盐分反演模型取得了最佳的精度效果, R_c^2 与 R_v^2 分别达到了 0.72 和 0.67, $RMSE_v$ 仅为 0.112%。

3 讨论

基于 3 个模型输入变量组, 使用 4 种回归方法建立了 12 个盐分反演模型, 发现选择合适的模型输入变量和模型回归方法, 无人机遥感可定量反演土壤含盐量。前人对土壤盐分的定量反演大多集中于对裸土的直接反演^[34-36], 而本研究主要针对于作物覆盖下的土壤盐分反演研究, 研究发现植被在近红外波长 900、800 nm 处对土壤含盐量变化响应较为敏感, 在“红边”720 nm 处反映最不明显, 而 ALLBED 等^[11]研究也表明, 当土壤含盐量超过植被生长阈值时, 植被在近红外波段会出现反射率降低的现象, 因此近红外波段可作为监测作物覆盖下的土壤含盐量的重点研究波段。此外, 本研究发现光谱指数组反演精度优于敏感波段组, 这与王飞等^[18]比较不同变量组对预测绿洲土壤盐分的精度影响时, 研究结果相一致; 而全变量组表现不如光谱指数组, 可能是由于过多自变量的引入产生了过度拟合^[37]。同时, 基于 3 种机器学习算法构建的 9 个盐分反演模型均表现稳定, 而 3 个基于多元线性回归的线性盐分反演模型表现极不稳定, 除光谱指数变量组外, 其余两组 R_c^2 与 R_v^2 相差太大, 即存在“过拟合”现象, 这可能是由于土壤盐渍化发生机理的复杂性, 植被冠层的光谱特征与土壤含盐量存在复杂的非线性关系, 此时, 基于 3 种机器学习构建的盐分反演模型相比多元线性回归模型表现出强大的非线性拟合能力和出色的数据挖掘能力, 加大了光谱反射信息的利用程度。厉彦玲等^[19]反演黄河三角洲盐分时, 发现 3 类机器学习模型 (BP 神经网络、支持向量机和随机森林) 反演效果明显优于统计分析模型, 与本文研究结果一致。

本研究取得了良好的盐分反演效果, 但是由于不同作物对盐分的敏感响应程度不同, 从而造成获取的光谱特征也不同, 所以下一步可以考虑对不同的作物进行分开盐分反演。此外, 本研究的盐分反演模型局限于小灌域尺度, 对大区域甚至是全国尺度的统一盐分定量盐分反演模型亟待研究。

4 结论

(1) 对比了基于 3 个模型输入变量组构建的土

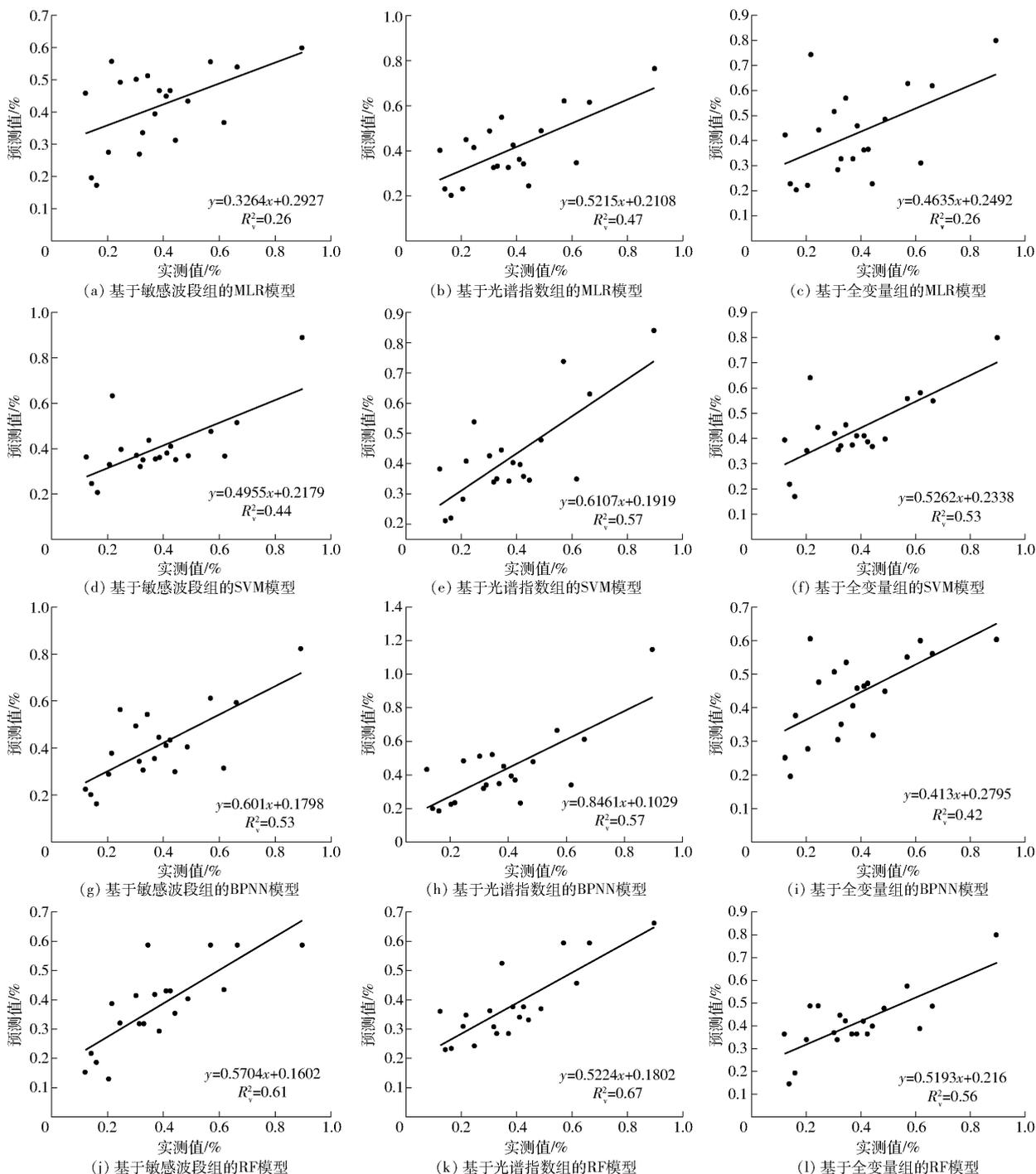


图5 实测值与预测值的散点图

Fig. 5 Scatter plots between measured and predicted values

壤含盐量反演模型,光谱指数组取得了最优的反演精度。

(2)对比分析了土壤盐渍化发生机理的复杂性,在多元线性回归模型中,反演精度最佳的一组 R^2 为0.47,相较于多元线性回归模型,3种机器学习模型的鲁棒性更好,模型精度得到较大提高;在基于机器学习算法的9个模型中, R^2 平均值达到了

0.54,呈现出强大的非线性拟合能力,且在3种机器学习模型中,随机森林(RF)算法模型表现最佳。

(3)对比筛选12个盐分反演模型,基于光谱指数组的随机森林算法反演模型 R^2 为0.67, RMSE_v为0.112%,为12个模型中的最优值。故基于光谱指数组的随机森林模型为最佳土壤含盐量反演模型。

参 考 文 献

- 2016, 53(6): 1399 – 1409.
- CAO Lei, DING Jianli, UMUT Halik, et al. Extraction and modeling of regional soil salinization based on data from GF – 1 satellite [J]. *Acta Pedologica Sinica*, 2016, 53(6): 1399 – 1409. (in Chinese)
- [2] DEHNI A, LOUNIS M. Remote sensing techniques for salt affected soil mapping: application to the Oran region of algeria [J]. *Procedia Engineering*, 2012, 33(3): 188 – 198.
- [3] GE X, WANG J, DING J, et al. Combining UAV – based hyperspectral imagery and machine learning algorithms for soil moisture content monitoring [J]. *PeerJ*, 2019, 7: e6926.
- [4] 王敬哲, 丁建丽, 马轩凯, 等. 基于光谱指数的绿洲农田土壤含水率无人机高光谱检测[J/OL]. *农业机械学报*, 2018, 49(11): 171 – 179.
- WANG Jingzhe, DING Jianli, MA Xuankai, et al. Detection of soil moisture content based on UAV – derived hyperspectral imagery and spectral index in oasis cropland [J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2018, 49(11): 171 – 179. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20181119&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2018.11.019. (in Chinese)
- [5] XIAO Z Z, LI Y, FENG H. Hyperspectral models and forecasting of physico-chemical properties for salinized soils in northwest China [J]. *Spectroscopy and Spectral Analysis*, 2016, 36(5): 1615 – 1622.
- [6] IVUSHKIN K, BARTHOLOMEUS H, BREGT A K, et al. Soil salinity assessment through satellite thermography for different irrigated and rainfed crops [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2018, 68: 230 – 237.
- [7] WANG H, CHEN Y, ZHANG Z, et al. Quantitatively estimating main soil water-soluble salt ions content based on visible-near infrared wavelength selected using GC, SR and VIP [J]. *PeerJ*, 2019, 7: e6310.
- [8] 迟春明, 王志春. 苏打盐渍土草原植被群落分布与土壤理化性质相互关系 [J]. *生态学杂志*, 2013, 32(9): 2245 – 2249.
- CHI Chunming, WANG Zhichun. Interrelationships between grassland plant community distribution and soil physical and chemical properties in sodasaline soil regions of Songnen Plain, Northeast China [J]. *Chinese Journal of Ecology*, 2013, 32(9): 2245 – 2249. (in Chinese)
- [9] DEHAAN R L, TAYLOR G R. Field-derived spectra of salinized soils and vegetation as indicators of irrigation-induced soil salinization [J]. *Remote Sensing of Environment*, 2002, 80(3): 406 – 417.
- [10] 林文鹏, 李厚增, 黄敬峰, 等. 上海城市植被光谱反射特征分析 [J]. *光谱学与光谱分析*, 2010, 30(11): 3111 – 3114.
- LIN Wenpeng, LI Houzeng, HUANG Jingfeng, et al. Analysis on urban vegetations reflectance characteristics in Shanghai [J]. *Spectroscopy and Spectral Analysis*, 2010, 30(11): 3111 – 3114. (in Chinese)
- [11] 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 – 231: 1 – 8.
- [12] 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.
- [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 – 112.
- CHEN Hongyan, ZHAO Gengxing, CHEN Jingchun, et al. Remote sensing inversion of saline soil salinity based on modified vegetation index in estuary area of Yellow River [J]. *Transactions of the CSAE*, 2015, 31(5): 107 – 112. (in Chinese)
- [14] DEHAAN R, TAYLOR G R. Image-derived spectral endmembers as indicators of salinisation [J]. *International Journal of Remote Sensing*, 2010, 24(4): 775 – 794.
- [15] 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)
- [16] 范晓梅, 刘高焕, 唐志鹏, 等. 黄河三角洲土壤盐渍化影响因素分析 [J]. *水土保持学报*, 2010, 24(1): 139 – 144.
- FAN Xiaomei, LIU Gaochuan, TANG Zhipeng, et al. Analysis on main contributors influencing soil salinization of Yellow River delta [J]. *Journal of Soil and Water Conservation*, 2010, 24(1): 139 – 144. (in Chinese)
- [15] 翁永玲, 宫鹏. 土壤盐渍化遥感应用研究进展 [J]. *地理科学*, 2006, 26(3): 369 – 375.
- WENG Yongling, GONG Peng. A review on remote sensing technique for salt-affected soils [J]. *Scientific Geographica Sinica*, 2006, 26(3): 369 – 375. (in Chinese)
- [16] 王明宽, 莫宏伟, 陈红艳. 基于多光谱影像反演土壤盐分的建模方法研究 [J]. *土壤通报*, 2016, 47(5): 1036 – 1041.
- WANG Mingkuan, MO Hongwei, CHEN Hongyan. Study on model method of inversion of soil salt based on multispectral image [J]. *Chinese Journal of Soil Science*, 2016, 47(5): 1036 – 1041. (in Chinese)
- [17] 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.
- [18] 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)

- [19] 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 Yanling, ZHAO Gengxing, CHANG Chunyan, et al. Soil salinity retrieval model based on OLI and HIS image fusion [J]. Transactions of the CSAE, 2017, 33(21):173-180. (in Chinese)
- [20] IVUSHKIN K, BARTHOLOMENUS H, BREGTA K, et al. UAV based soil salinity assessment of cropland [J]. Geoderma, 2018, 338:502-512.
- [21] 黄权中, 徐旭, 吕玲娇, 等. 基于遥感反演河套灌区土壤盐分分布及对作物生长的影响 [J]. 农业工程学报, 2018, 34(1):102-109.
HUANG Quanzhong, XU Xu, LÜ Lingjiao, et al. Soil salinity distribution based on remote sensing and its effect on crop growth in Hetao Irrigation District [J]. Transactions of the CSAE, 2018, 34(1):102-109. (in Chinese)
- [22] 史海滨, 杨树青, 李瑞平, 等. 作物水盐联合胁迫效应与水分高效利用研究[M]. 北京: 中国水利水电出版社, 1999:89-90.
- [23] 韩婷婷, 刁晓环, 王成, 等. 基于 TM 数据的西双版纳地区森林叶面积指数反演[J]. 遥感信息, 2014, 29(2):26-30.
- [24] 陈兵, 李少昆, 王克如, 等. 基于 TM 影像光谱指数的棉花病害严重度估测[J]. 红外与毫米波学报, 2011, 30(5):451-457.
CHEN Bing, LI Shaokun, WANG Keru, et al. Estimating severity level of cotton disease based on spectral indices of TM image [J]. J. Infrared Millim. Waves, 2011, 30(5):451-457. (in Chinese)
- [25] MARSHALL F E, SMITH D T, NICKERSON D M. Empirical tools for simulating salinity in the estuaries in Everglades National Park, Florida [J]. Estuarine Coastal and Shelf Science, 2011, 95(4):377-387.
- [26] LESCH S M, STRAUSS D J, RHOADES J D. Spatial prediction of soil salinity using electromagnetic induction techniques: 1. statistical prediction models: a comparison of multiple linear regression and cokriging [J]. Water Resources Research, 1995, 31(2):373-386.
- [27] DIAO S, LIU C, ZHANG T, et al. Extraction of remote sensing information for lake salinity level based on SVM: a case from Badain Jaran desert in Inner Mongolia [J]. Remote Sensing for Land & Resources, 2016, 28(4):114-118.
- [28] 梁栋, 管青松, 黄文江, 等. 基于支持向量机回归的冬小麦叶面积指数遥感反演[J]. 农业工程学报, 2013, 29(7):117-123.
LIANG Dong, GUAN Qingsong, HUANG Wenjiang, et al. Remote sensing inversion of leaf area index based on support vector machine regression in winter wheat [J]. Transactions of the CSAE, 2013, 29(7):117-123. (in Chinese)
- [29] YAO Rongjiang. BP neural network model for spatial distribution of regional soil water and salinity [J]. Acta Pedologica Sinica, 2009, 46(5):788-794.
- [30] ZHENG Z, ZHANG F, CHAI X, et al. Spatial estimation of soil moisture and salinity with neural Kriging [J]. Computer and Computing Technologies in Agriculture II, 2008, 2:1227-1237.
- [31] 宋荣杰, 宁纪锋, 常庆瑞, 等. 基于小波纹理和随机森林的猕猴桃果园遥感提取 [J/OL]. 农业机械学报, 2018, 49(4):222-231.
SONG Rongjie, NING Jifeng, CHANG Qingrui, et al. Kiwifruit orchard mapping based on wavelet textures and random forest [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(4):222-231. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20180425&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2018.04.025. (in Chinese)
- [32] WANG J, DING J, ABULIMITI A, et al. Quantitative estimation of soil salinity by means of different modeling methods and visible-near infrared (vis-NIR) spectroscopy, Ebinur Lake Wetland, Northwest China [J]. Peerj, 2018, 6:e4703.
- [33] 贾科利, 张俊华. 不同碱化程度土壤对其上覆植被冠层光谱特征的影响 [J]. 光谱学与光谱分析, 2014, 34(3):782-786.
JIA Keli, ZHANG Junhua. Impact of different alkaline soil on canopy spectral characteristics of overlying vegetation [J]. Spectroscopy and Spectral Analysis, 2014, 34(3):782-786. (in Chinese)
- [34] 马红章, 刘素美, 彭爱华, 等. L 波段主被动微波协同反演裸土土壤水分 [J]. 农业工程学报, 2016, 32(19):133-138.
MA Hongzhang, LIU Sumei, PENG Aihua, et al. Active and passive cooperative algorithm at L-band for bare soil moisture inversion [J]. Transactions of the CSAE, 2016, 32(19):133-138. (in Chinese)
- [35] 孙亚楠, 李仙岳, 史海滨, 等. 河套灌区土壤水溶性盐基离子高光谱综合反演模型 [J/OL]. 农业机械学报, 2019, 50(5):344-355.
SUN Ya'nan, LI Xianyue, SHI Haibin, et al. Hyperspectral integrated inverse model for water-soluble salt ions content in Hetao Irrigation District [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(5):344-355. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190539&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2019.05.039. (in Chinese)
- [36] KHAN N M, RASTOSKUEV V V, SATO Y, et al. Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators [J]. Agricultural Water Management, 2005, 77(1):96-109.
- [37] MORELLOS A, PANTAZI X E, MOSHOUS D, et al. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using vis-NIR spectroscopy [J]. Biosystems Engineering, 2016, 152:104-116.