

基于全子集一分位数回归的土壤含盐量反演研究

张智韬^{1,2} 韩佳^{1,2} 王新涛^{1,2} 陈皓锐³ 魏广飞^{1,2} 姚志华^{1,2}

(1. 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;

2. 西北农林科技大学水利与建筑工程学院, 陕西杨凌 712100;

3. 中国水利水电科学研究院水利研究所, 北京 100038)

摘要: 为提高植被覆盖条件下卫星遥感对土壤含盐量的估测精度,以河套灌区解放闸灌域为研究区,以高分一号卫星影像为数据源,同步采集不同深度土壤含盐量,通过全子集筛选法(Best subset selection)分析不同波段和光谱指数对于不同深度土壤含盐量的敏感性,并采用人工神经网络(Artificial neural network, ANN)、支持向量机(Support vector machine, SVM)和分位数回归(Quantile regression, QR)3种方法,构建全子集筛选前后0~20 cm、20~40 cm、0~40 cm、40~60 cm、0~60 cm等不同深度下的土壤含盐量反演模型。结果表明,B4、BI、SI1、SI3是0~20 cm、0~40 cm处土壤含盐量的敏感变量组合,B4、BI、NDVI为20~40 cm、40~60 cm、0~60 cm处土壤含盐量的敏感变量组合;在各深度下,分位数回归模型的精度最高,模型的确定系数 R_{cl}^2 、 R_{vl}^2 均在0.4以上,均方根误差 R_{MSEcl} 、 R_{MSEvl} 均小于0.4%,SVM次之,ANN最差;在20~40 cm深度下QR反演模型效果优于其他深度,为本文土壤含盐量估算的最优模型,其建模和验证的确定系数 R_{cl}^2 、 R_{vl}^2 分别为0.611和0.671,建模和验证均方根误差 R_{MSEcl} 、 R_{MSEvl} 分别为0.177%和0.160%。本研究可为卫星遥感大范围监测植被覆盖条件下土壤盐渍化程度提供参考。

关键词: 土壤含盐量; 遥感; 全子集; 分位数回归; 反演模型

中图分类号: S127; S156.4 文献标识码: A 文章编号: 1000-1298(2019)10-0142-11

Soil Salinity Inversion Based on Best Subsets – Quantile Regression Model

ZHANG Zhitao^{1,2} HAN Jia^{1,2} WANG Xintao^{1,2} CHEN Haorui³ WEI Guangfei^{1,2} YAO Zhihua^{1,2}

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

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

3. Department of Irrigation and Drainage, China Institute of Water Resources and Hydropower Research, Beijing 100038, China)

Abstract: The soil salinity is essential for the morphological development, growth process and final yield of crops in the irrigation area. With present methods, satellite remote sensing though was non-invasive, dynamic, rapid and macroscopic, estimated soil salinity of soil covered by vegetation have less significant effect, yet. In order to improve the estimation effect, soil salinity at different depths at Hetao Irrigation Area was collected. GF-1 image simultaneous was downloaded as the data source. Best subset selection was used to analyze the sensitivity of different bands and spectral indices to soil salinity at different depths. R_{MSE} , R^2 , A_{IC} and S_{IC} were used to determine the optimal combination mode of the sensitive independent variables number at different depths. Based on these, artificial neural network (ANN), support vector machine (SVM) and quantile regression (QR) were used to construct soil salinity inversion model at such depths as: 0~20 cm, 20~40 cm, 0~40 cm, 40~60 cm and 0~60 cm before and after best subset selection. The determination coefficient for calibration set before best subset selection (R_{c0}^2), determination coefficient for calibration set after best subset selection (R_{cl}^2), determination coefficient for verification set before best subset selection (R_{v0}^2), determination coefficient for verification set after best subset selection (R_{vl}^2), root mean square error for calibration set before best subset selection (R_{MSEc0}), root mean square error for calibration set after best subset selection (R_{MSEcl}), root mean square error for verification set before best subset selection (R_{MSEv0}) and root mean square error

收稿日期: 2019-07-06 修回日期: 2019-08-02

基金项目: 国家重点研发计划项目(2017YFC0403302)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

for verification set after best subset selection ($R_{MSE_{el}}$) were used to evaluate the effects of the models. The results showed that B4, BI, SI1 and SI3 were sensitive variable combinations of soil salinity at depths of 0 ~ 20 cm and 0 ~ 40 cm. B4, BI and NDVI were sensitive variable combinations of soil salinity at depths of 20 ~ 40 cm, 40 ~ 60 cm and 0 ~ 60 cm. QR inversion model showed its good performance because of its strong robustness. With R_{cl}^2 and R_{vl}^2 were both above 0.4, and $R_{MSE_{el}}$ and $R_{MSE_{el}}$ were both under 0.4%; followed by SVM, and ANN was the worst. Compared with other depths, the QR inversion model performed best at depths of 20 ~ 40 cm, with R_{cl}^2 of 0.611, R_{vl}^2 of 0.671, $R_{MSE_{el}}$ of 0.177%, and $R_{MSE_{el}}$ of 0.160%. The combination of best subset selection and QR method in the modeling analysis of soil salinity provided a new approach to optimize the satellite multispectral model and quickly measure the soil salinity. The research result provided a reference for the wide-scale soil salinity monitoring of soil covered by vegetation.

Key words: soil salinity; remote sensing; best subset selection; quantile regression; inversion models

0 引言

土壤盐渍化程度是影响灌区土地资源经济效益和作物产量的关键因素,及时、准确地监测盐渍化动态是进行科学防治的重要前提。与传统的土壤盐渍化监测方法相比,卫星遥感具有无创、动态、综合、快速、宏观等优点,已成为现阶段大面积监测土壤盐渍化的重要研究方向。

国内外学者在利用卫星遥感监测土壤盐渍化程度方面取得了一定的进展。目前,常用方法主要有遥感图像分类及解译法^[1]、光谱指数法^[2-3]、特征空间分析法^[4]。其中,光谱指数法是一种便捷高效的方法。陈红艳等^[2]在传统的植被指数基础上加入短波红外波段,明显提高了遥感反演土壤盐渍化的精度。EL HARTI 等^[3]在盐分指数(SI)的基础上加入蓝波段,构建了新的盐分指数 OLI-SR,提高了 Morocco 灌区盐渍化反演的精度。但是,光谱指数易受到下垫面条件和大气状况等多种因素的影响,具有明显的地域性和时效性。因此,筛选敏感光谱指数对于土壤含盐量监测具有重要意义。目前,常用的变量筛选方法有灰度关联法^[5]、岭回归^[6]、LASSO 回归^[7]、变量投影重要性分析^[8]等,这些筛选方法均为局部最优筛选。全子集筛选法列举全部可能组合方式,建立全局最优模型,以包含最少自由变量的模型解释因变量,进而消除共线性的影响。李长春等^[9]在估测叶面积指数时发现,采用全子集筛选法建立的回归模型优于 PLSR、SVM 和 RF 模型。冯克伟^[10]经过全子集筛选,对野生二粒小麦和硬粒小麦的耐盐指数分别拟合,预测决定系数达到 0.95 以上,进一步说明全子集筛选法对于筛选光谱指数具有一定的可行性。

在以往的土壤盐渍化监测中,大多采用经典线性回归模型^[11-12](如最小二乘法和偏最小二乘法)和机器学习模型^[13-14]。经典线性回归对随机误差的分布特征要求严格,对于一些实际问题,难以得到

无偏、有效的参数估计值^[15]。人工神经网络模型和支持向量机模型是常用的机器学习算法。人工神经网络模型基于梯度学习方法,在层间传播的非参数非线性模型,因其具有较强的自学习和自适应能力、非线性映射能力和容错能力^[16],而被广泛用于遥感反演中。支持向量机模型主要用于对高维度、小样本的分类和回归,将有限样本信息在模型的复杂性和学习能力之间寻求最佳折中,在保证泛化能力的前提下达到最优学习效果^[17]。与机器学习方法相比,分位数回归模型所需参数少,训练时间短,不易产生过拟合现象。分位数回归可以在任意一个分位点准确地描述自变量对于因变量的变化范围以及条件分布形状的影响,适合处理波动性大、异质性强的数据^[18],广泛应用于经济、医学等领域,但在监测盐渍化方面的研究还很少^[19]。

本文利用河套灌区解放闸灌域 GF-1 卫星遥感影像数据,计算多种光谱指数,将通过全变量及全子集筛选法得到的最优光谱指数组合方式作为自变量,分别构建不同深度下的人工神经网络、支持向量机和分位数回归 3 种模型,进而对比分析得到土壤含盐量反演的最优模型,以期提高植被覆盖条件下的土壤含盐量反演精度。

1 材料与方法

1.1 研究区概况

研究区设在内蒙古自治区河套灌区解放闸灌域(40°30'~41°17'N,106°33'~107°31'E),是河套灌区的第二大灌域,位于河套灌区西部,西接一干灌域,东邻永济灌域,灌溉面积 1 420 km²,灌溉水主要来源于过境的黄河水。主要农作物包括玉米、小麦、葵花。该灌域地处干旱半干旱气候区,年均降水量约为 140 mm,年均蒸发量约为 2 000 mm^[20]。研究区内平均坡度为 0.2%,地势平坦使得排水系统的排水能力降低,地下水埋深较浅。加之降雨少、蒸发大、高灌低排,使得研究区内土壤盐渍化问题较为严重。

1.2 遥感数据获取及预处理

以国产高分一号卫星影像(GF-1 WFV 相机)为数据源,通过中国资源卫星应用中心进行下载(www.cresda.com)。卫星影像的成像时间为2018年7月27日,与实测数据日期基本同步,高分一号卫星数据的重访周期为4 d,空间分辨率为16 m,包括4个波段,分别为蓝(0.45~0.52 μm)、绿(0.52~0.59 μm)、红(0.63~0.69 μm)和近红外波段(0.77~0.89 μm)。对下载的影像经过几何精校正、辐射校正、大气校正、镶嵌、裁剪预处理过程。并对影像进行低通滤波处理^[21],保存图像的低频部分,使图像平滑,在一定程度上消除了高频随机噪声,提高数据的信噪比。

1.3 土壤样本采集及化学分析

采样点的选定考虑到灌域内盐分分布、植被覆盖种类(包括葵花35个、玉米22个、小麦8个、果蔬8个和荒地7个),并顾及点位分布的均匀性,本次试验在解放闸灌域共设置了80个不同土壤盐渍化程度的采样点,采样点分布情况见图1。采样时间为2018年7月12—16日,根据该时期灌区内主要作物的根系活动层所在深度,每个采样点按照0~20 cm、20~40 cm、40~60 cm分层采样,并通过手持GPS记录采样点位置信息及周围环境信息。采用五点法采样,采样单元为16 m×16 m,每个采样点取5个土样,编号后带回实验室。

将野外收集的土样干燥研磨处理后,配置土水比为1:5的土壤溶液,经搅拌、静置、沉淀、过滤后,采用电导率仪(DDS-307A型,上海佑科仪器公司)

测定土壤溶液电导率 $EC_{1:5}$,对每个采样点的5个土样电导率取平均值作为该样本样点处的电导率,并通过经验公式计算土壤含盐量 S ($S = (0.2882EC_{1:5} + 0.0183) \times 100\%$),剔除图像中小麦收割后的8个样本,剩余72个土壤样本用于本次植被覆盖条件下的灌区土壤含盐量反演。将测得的土壤含盐量从大到小排序,每隔2个样本取出1个作为验证集样本。可保证建模样本和验证样本范围一致且分布均匀。土壤含盐量数据的统计特征如表1所示。土壤盐渍化等级划分参照文献[22]。表1中0~40 cm土壤含盐量为0~20 cm和20~40 cm深度实测含盐量的平均值;0~60 cm土壤含盐量为0~20 cm、20~40 cm、40~60 cm深度实测含盐量的平均值。

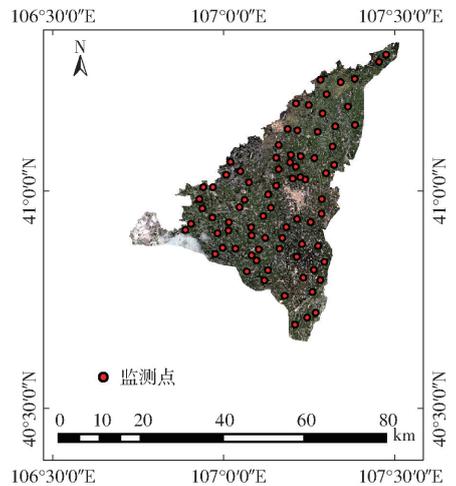


图1 研究区地理示意图及采样点分布

Fig.1 Schematic of study area and distribution of sampling points

表1 采样点的土壤含盐量统计

Tab.1 Summary statistics of soil salinity of sampling points

深度/cm	类型	样本数				含盐量					
		非盐土 (0~0.2%)	轻盐渍化土 (0.2%~0.5%)	重盐渍化土 (0.5%~1.0%)	盐土 (大于1.0%)	最小值/ %	最大值/ %	均值/ %	标准差/ %	变异 系数	偏度
0~20	总体	33	26	7	6	0.060	2.531	0.378	0.468	1.238	3.071
	建模集	22	17	5	4	0.060	2.531	0.369	0.446	1.209	3.169
	验证集	11	9	2	2	0.070	2.416	0.395	0.520	1.316	3.052
20~40	总体	33	27	8	4	0.068	1.289	0.316	0.274	0.867	1.942
	建模集	22	18	5	3	0.068	1.289	0.318	0.281	0.884	2.049
	验证集	11	9	3	1	0.076	1.062	0.314	0.265	0.844	1.801
0~40	总体	35	25	7	5	0.064	1.828	0.347	0.360	1.037	2.539
	建模集	23	17	5	3	0.064	1.828	0.346	0.357	1.032	2.541
	验证集	12	8	2	2	0.080	1.734	0.349	0.375	1.074	2.697
40~60	总体	35	21	14	2	0.053	1.229	0.311	0.253	0.814	1.799
	建模集	23	14	10	1	0.053	0.929	0.299	0.220	0.736	1.293
	验证集	12	7	4	1	0.080	1.197	0.312	0.261	0.837	2.010
0~60	总体	34	24	10	4	0.073	1.452	0.335	0.313	0.934	2.162
	建模集	23	16	6	3	0.073	1.452	0.336	0.319	0.949	2.186
	验证集	11	8	4	1	0.080	1.381	0.332	0.306	0.922	2.214

1.4 光谱指数计算及全子集筛选

高分一号遥感影像自身的光谱分辨率不高,对于植被覆盖条件下的土壤盐渍化反演模型,仅用多个波段的反射率对比分析提取土壤盐渍化信息有明显的局限性,本研究利用高分一号卫星数据的 4 个波段反射率和 12 个光谱指数(包括 5 个土壤指数、3 个盐分指数和 4 个植被指数,见表 2)建立遥感图像与土壤含盐量的定量关系。

表 2 光谱指数

Tab. 2 Spectral index summary

光谱指数	计算公式	文献序号
BI	$\sqrt{R_R^2 + R_{NIR}^2}$	
SI	$\sqrt{R_B R_R}$	
SI1	$\sqrt{R_R R_G}$	
SI2	$\sqrt{R_G^2 + R_R^2 + R_{NIR}^2}$	[23]
SI3	$\sqrt{R_G^2 + R_R^2}$	
S1	R_B/R_R	
S2	$(R_B - R_R)/(R_B + R_R)$	
S6	$R_R R_{NIR}/R_G$	
RVI	R_R/R_{NIR}	[24]
NDVI	$(R_{NIR} - R_R)/(R_{NIR} + R_R)$	
MSAVI	$0.5 \{2R_{NIR} - 1 - [(2R_{NIR} + 1)^2 - 8(R_{NIR} - R_R)]^{0.5}\}$	[25]
ARVI	$(R_{NIR} - 2R_R + R_B)/(R_{NIR} + 2R_R - R_B)$	[26]

注: R_B 、 R_G 、 R_R 和 R_{NIR} 分别为高分一号卫星数据对应 4 个波段的光谱反射率。

全子集筛选是基于不同自变量的所有可能的组合方式,对缩减后的变量组合通过最小二乘法进行拟合,在所有可能的模型中选择一个最优模型。该方法简单直观,适合自变量较少的条件下使用。其主要计算步骤如下:记 K 为子变量数目 ($K = 1, 2, \dots, P$);拟合 $1 \sim P$ 个预测变量的模型;在 $1 \sim P$ 个模型中,根据“调整后 R^2 最大”准则来选择 P 个最优模型;根据验证集调整后决定系数 (Coefficient of determination, R^2)、均方根误差 (Root mean squared error, R_{MSE}) 从 P 个模型中选择一个最佳自变量组合。全子集筛选的模型建立和预测通过 Matlab R2017b 完成。

1.5 模型建立

1.5.1 分位数回归模型

分位数回归研究自变量与因变量的条件分位数之间的关系,通过最小化离差绝对值的加权和,依据因变量的条件分位数对自变量进行回归,进而得到所有分位数下的回归模型^[27]。相较于传统的回归方法,其优点主要表现在:①无需对模型中变量进行随机扰动或进行正态变化,在非正态的干扰下,分位数模型精度更高。②模型中异常点对于模型整体精度影响较小,模型稳定性强。③可以给出任意分位

点的参数估计,解释自变量对不同水平因变量的影响^[28]。分位数回归模型的建立与预测在 SPSS 23.0 软件中完成。

对于因变量 Y 的一个随机样本 $\{y_1, y_2, \dots, y_n\}$,通常 τ 分位数的样本分位数线性回归要求满足

$$\min_{\beta \in \mathbf{R}} \left(\sum_{i=1}^n y_i - x_i' \beta(\tau) \right)$$

其中, y_i 为因变量, β 为自变量的回归系数, $\beta(\tau)$ 为特定分位数 τ 的最终回归系统, x_i' 为自变量。

对于任意的 $0 < \tau < 1$, 求解参数估计值的公式为

$$\beta' = \arg \min \sum_{i=1}^n (y_i - x_i' \beta(\tau)) \quad (1)$$

将检验函数定义式代入 $\min \left(\sum_{i=1}^n y_i - x_i' \beta(\tau) \right)$, 改写后为

$$\min_{\beta \in \mathbf{R}} \left(\sum_{i: y_i \geq x_i' \beta(\tau)} \tau |y_i - x_i' \beta(\tau)| + \sum_{i: y_i < x_i' \beta(\tau)} (1 - \tau) |y_i - x_i' \beta(\tau)| \right)$$

求得参数 $x_i' \beta(\tau)$ 即为唯一的第 τ 回归分位数。

1.5.2 人工神经网络模型

人工神经网络模型从信息处理的角度对人脑的神经元进行抽象,建立某种简单模型,按不同的连接方式组合不同的网络。该模型拓扑结构由输入层、隐含层、输出层构成。经大量试算,本文输入层为不同深度下全变量及经过全子集筛选后的敏感光谱指数组合,输出层为土壤含盐量,隐含层中的网络层数为 1。其中隐含层采用双曲正切激活函数,输出层采用恒等激活函数,可以在一定程度上防止梯度消失且结果可靠;训练拟合目标误差拟定为 0.1%;培训类型采用适用于小数据集且精度高的批处理,并选用对应的标度共轭梯度估计突出权重进行网络训练;为了消除量纲、数量级等对数据分析造成的影响,将光谱指数标准化处理。本文采用 SPSS 23.0 软件中的多层感知器神经网络进行模型的建立与预测。

1.5.3 支持向量机模型

支持向量机模型是在结构风险最小化原理的基础上,通过一个非线性变换将数据变换到一个高维特征空间,并在该特征空间建立线性模型来拟合回归函数,很大程度上克服了“离散值多”和“过学习”等问题^[29]。径向基核 (RBF) 是应用最广泛的核函数。惩罚参数 C 和 RBF 核参数 δ 是影响 SVM 的主要参数,其中参数 C 直接影响模型的稳定性,避免模型在学习和推广过程中产生欠学习和过学习问题,决定了适应误差的最小化和平滑程度;参数 δ 反

表 3 全子集筛选最佳组合方式结果统计
Tab. 3 Best combination result after total subset selection

深度/cm	自变量数/个	最佳组合	R^2	$R_{MSE}/\%$	A_{IC}	S_{IC}
0~20	2	B4*、BI	0.37	0.41	0.81	0.93
	3	B4*、BI、MSAVI**	0.41	0.40	0.82	0.98
	4	B4*、BI、S11**、S13**	0.53	0.37	0.79	0.99
	5	B4*、BI、S11**、S13**、NDVI**	0.46	0.39	0.81	1.05
	6	B4*、BI、S11**、S13**、NDVI**、B2**	0.52	0.38	0.82	1.09
	7	B4*、BI、S11**、S13**、NDVI**、B2**、RVI**	0.53	0.37	0.86	1.17
	20~40	2	B4*、BI	0.59	0.17	-0.40
3		B4*、BI、NDVI**	0.67	0.15	-0.53	-0.37
4		B4*、BI、NDVI**、ARVI**	0.63	0.16	-0.51	-0.32
5		B4*、BI、NDVI**、S2**、ARVI**	0.62	0.17	-0.49	-0.25
6		B4*、BI、NDVI**、S2**、S11**、S1**	0.63	0.16	-0.46	-0.18
7		B4*、BI、NDVI**、S2**、S11**、S1**、S6	0.59	0.17	-0.44	-0.12
0~40		2	B4*、BI	0.55	0.26	0.28
	3	B4*、BI、NDVI**	0.51	0.27	0.19	0.35
	4	B4*、BI、S11**、S13**	0.63	0.24	0.23	0.43
	5	B4*、BI、S11**、S13**、NDVI**	0.57	0.26	0.22	0.45
	6	B4*、BI、S11**、S13**、NDVI**、B2**	0.40	0.29	0.21	0.48
	7	B4*、BI、S11**、S13**、NDVI**、B2**、S6	0.24	0.33	0.20	0.51
	40~60	2	B4、BI	0.51	0.19	-0.71
3		B4、BI、NDVI**	0.55	0.18	-0.78	-0.62
4		B4、BI、NDVI**、S1**	0.55	0.18	-0.78	-0.59
5		B4、BI、NDVI**、S1**、ARVI**	0.54	0.18	-0.74	-0.51
6		B4、BI、NDVI**、S1**、ARVI**、B1**	0.50	0.19	-0.72	-0.45
7		B4、BI、NDVI**、S1**、B2**、ARVI**、S12	0.21	0.25	-0.74	-0.43
0~60		2	B4*、BI	0.48	0.23	-0.10
	3	B4*、BI、NDVI**	0.49	0.23	-0.25	-0.09
	4	B4*、BI、S11**、S13**	0.47	0.22	-0.22	-0.02
	5	B4*、BI、S11**、S13**、NDVI**	0.48	0.22	-0.24	-0.01
	6	B4*、BI、S11**、S1**、NDVI**、S2**	0.48	0.23	-0.21	0.07
	7	B4*、BI、S11**、S1**、NDVI**、S2**、S6	0.39	0.27	-0.17	0.14

注: **表示 0.01 显著性水平, * 表示 0.05 显著性水平。

看出,在相关性分析中 B4 和 BI 与不同深度土壤含盐量没有呈现极显著的相关关系,经过全子集筛选后 B4 和 BI 的组合方式在不同深度下与土壤含盐量均呈现良好的相关性,并成为各深度下自变量数目为 2 的最优自变量组合。在植被遥感中应用最为广泛的归一化植被指数 NDVI 与不同深度的土壤含盐量均呈现极显著的相关性,在 20~40 cm、0~40 cm、40~60 cm、0~60 cm 土壤深度下与 B4、BI 组合后成为该深度下自变量数目为 3 的最优自变量组合。在 0~20 cm 处,其最优组合方式为 B4、BI、MSAVI,这是由于在 0~20 cm 处多是土壤和主要作物的侧根,土壤背景和植被的相互作用通过改进型土壤调整植被指数 MSAVI 减小土壤亮度的影响^[33]。当自变量数目增加到 4,主要增加了由 B2 和 B3 计算得出的 S11、S13 两种光谱指数,这是由于土壤一般都有很高的溶解性盐分,在 520~770 nm 平均反射率最

高^[34],且属于确定不同盐渍化过程中积盐状态特征的 6 个光谱区间之一^[35]。随着模型复杂程度的增加, A_{IC} 和 S_{IC} 逐渐增大,但 R^2 减小、 R_{MSE} 增大,模型灵敏度降低。综合分析全子集筛选的各个评价指标,确定在 0~20 cm、0~40 cm 处选择 B4、BI、S11、S13 共 4 个自变量,在 20~40 cm、40~60 cm、0~60 cm 处选择 B4、BI、NDVI 共 3 个自变量作为各深度下最优自变量组合方式。

2.3 神经网络模型建立与分析

分别以不同深度下筛选前后的反射率及光谱指数为自变量、土壤含盐量为因变量,运用神经网络模型进行筛选前后不同深度下土壤含盐量估算,其建模及验证结果如表 4 所示。

从表 4 可以看出,未进行全子集筛选条件下不同深度土壤含盐量的人工神经网络模型中,0~20 cm 深度的模型效果相对较差, R_{0}^2 、 R_{10}^2 分别为

表4 基于不同深度土壤含盐量的人工神经网络模型

Tab.4 ANN model of soil salinity at different depths

深度/cm	建模集(未筛选)		验证集(未筛选)		建模集(筛选后)		验证集(筛选后)	
	R_{c0}^2	$R_{MSEc0}/\%$	R_{v0}^2	$R_{MSEv0}/\%$	R_{c1}^2	$R_{MSEc1}/\%$	R_{v1}^2	$R_{MSEv1}/\%$
0~20	0.484	0.314	0.243	0.463	0.395	0.304	0.433	0.387
20~40	0.558	0.183	0.583	0.165	0.583	0.201	0.623	0.178
0~40	0.516	0.247	0.453	0.277	0.483	0.254	0.436	0.277
40~60	0.481	0.156	0.413	0.202	0.485	0.155	0.411	0.194
0~60	0.432	0.240	0.483	0.217	0.481	0.234	0.492	0.216

0.484和0.243, R_{MSEc0} 和 R_{MSEv0} 在0.4%左右,模型误差最大。0~40 cm、0~60 cm深度的 R_{c0}^2 和 R_{v0}^2 都在0.45左右且相差较小, R_{MSEc0} 和 R_{MSEv0} 在0.25%左右,误差较小。40~60 cm深度下的 R_{MSEc0} 最小,但是 R_{c0}^2 和 R_{v0}^2 、 R_{MSEc0} 和 R_{MSEv0} 的差距较大,分别达到了0.068%和0.046%,表明该模型稳定性较差。20~40 cm深度下模型的 R_{c0}^2 和 R_{v0}^2 均为最高且其模型的 R_{c0}^2 和 R_{v0}^2 最为接近,分别为0.558和0.583, R_{MSEc0} 和 R_{MSEv0} 最小,分别为0.183%和0.165%,总体来看,该深度下建立的模型误差最小。

表5 基于不同深度土壤含盐量的支持向量机模型

Tab.5 SVM model of soil salinity at different depths

深度/cm	建模集(未筛选)		验证集(未筛选)		建模集(筛选后)		验证集(筛选后)	
	R_{c0}^2	$R_{MSEc0}/\%$	R_{v0}^2	$R_{MSEv0}/\%$	R_{c1}^2	$R_{MSEc1}/\%$	R_{v1}^2	$R_{MSEv1}/\%$
0~20	0.452	0.327	0.487	0.432	0.377	0.353	0.376	0.453
20~40	0.674	0.157	0.601	0.173	0.659	0.160	0.608	0.171
0~40	0.477	0.267	0.503	0.295	0.466	0.268	0.486	0.293
40~60	0.456	0.205	0.450	0.163	0.523	0.174	0.563	0.155
0~60	0.584	0.208	0.491	0.209	0.499	0.240	0.458	0.224

从表5可以看出,未进行全子集筛选条件下不同深度土壤含盐量的支持向量机模型中, R_{c0}^2 、 R_{v0}^2 均不小于0.45,拟合效果较好。0~20 cm深度处的模型效果相对较差, R_{MSEc0} 和 R_{MSEv0} 在0.4%左右,模型误差最大。随着深度的增加(0~20 cm、0~40 cm、0~60 cm),模型的 R_{MSE} 逐渐减小至0.2%左右。在20~40 cm处 R_{c0}^2 和 R_{v0}^2 最高,分别为0.674和0.601; R_{MSEc0} 和 R_{MSEv0} 较小,分别为0.157%和0.173%。总体来看,该模型误差最小。

经过全子集筛选后不同深度土壤含盐量的支持向量机模型中,除40~60 cm深度外,其他深度筛选后的模型精度均低于筛选之前的模型精度。除0~20 cm深度外,其余相差幅度不大,其中 R_v^2 相差0.007、-0.017、-0.033, R_{MSEv} 相差-0.002%、-0.002%、0.015%。因此,经过全子集筛选,自变量数目减少后,模型精度基本保持不变。

2.5 分位数回归模型建立与分析

对于不同深度选定的敏感植被指数建立分位

经过全子集筛选后不同深度土壤含盐量的人工神经网络模型中,与未筛选结果相似,20~40 cm深度下模型最优,但精度略有不同。 R_{c1}^2 和 R_{v1}^2 分别达到了0.583和0.623,拟合精度明显提升,虽然 R_{MSEc1} 和 R_{MSEv1} 较未筛选之前有所增大,但相差较小。

2.4 支持向量机模型建立与分析

以不同深度下筛选前后的反射率及光谱指数为自变量,土壤含盐量为因变量,运用支持向量机模型进行筛选前后不同深度下土壤含盐量估算。其建模和验证结果如表5所示。

数回归模型,由于采样点土壤含盐量变异性较大,选取 $\tau = 0.5$ 分位点可以较好地解决最小二乘法中某些“离群值”影响回归显著性的问题。同时,由于0.5分位点处于因变量的中间位置,在对所有的数据进行拟合时较为适宜^[36]。故本文以不同深度下筛选前后的反射率及光谱指数为自变量,以土壤含盐量为因变量,运用分位数回归模型中的0.5分位点进行筛选前后不同深度下土壤含盐量估算,其建模及验证结果如表6所示。

从表6可以看出,未进行全子集筛选条件下不同深度土壤含盐量的模型中,在5个深度下, R_{c0}^2 均大于0.15, R_{MSEc0} 在0.2%左右,与筛选后 R_{MSEc1} 相差不多,但是 R_{v0}^2 明显小于 R_{c0}^2 , R_{MSEv0} 明显大于 R_{MSEc0} ,模型均出现过拟合现象。

经过全子集筛选后不同深度土壤含盐量的分位数回归模型均未产生过拟合现象。在5个深度的模型中,0~20 cm深度的模型效果相对较差, R_{c1}^2 和 R_{v1}^2 分别为0.415和0.442, R_{MSEc1} 和 R_{MSEv1} 均小于0.4%,

表 6 基于不同深度土壤含盐量的分位数回归模型

Tab. 6 QR model of soil salinity at different depths

深度/cm	建模集(未筛选)		验证集(未筛选)		建模集(筛选后)		验证集(筛选后)	
	R_{c0}^2	$R_{MSEc0}/\%$	R_{v0}^2	$R_{MSEv0}/\%$	R_{c1}^2	$R_{MSEc1}/\%$	R_{v1}^2	$R_{MSEv1}/\%$
0~20	0.255	0.280	0.090	0.507	0.415	0.346	0.442	0.397
20~40	0.366	0.127	0.090	0.370	0.611	0.177	0.671	0.160
0~40	0.252	0.233	0.020	0.523	0.453	0.259	0.505	0.251
40~60	0.198	0.139	0.105	0.351	0.499	0.153	0.551	0.181
0~60	0.356	0.119	0.023	0.694	0.499	0.155	0.457	0.210

模型误差最大。0~40 cm 深度的 R_{c1}^2 和 R_{v1}^2 相差较大,但 R_{MSEc1} 和 R_{MSEv1} 在 0.255% 左右,模型误差较小且相对稳定。0~60 cm、40~60 cm 深度下的 R_{MSEc1} 较小,但是 R_{c1}^2 和 R_{v1}^2 以及 R_{MSEc1} 和 R_{MSEv1} 的差距较大,分别达到了 0.042、-0.052 和 -0.055%、-0.028%,表明该模型稳定性较差。20~40 cm 深度下模型的 R_{c1}^2 和 R_{v1}^2 均为最高,分别为 0.611 和 0.671; R_{MSEc1} 和 R_{MSEv1} 最小,分别为 0.177% 和 0.160%,总体来看,该模型拟合效果最好且误差最小。

2.6 模型综合评价

以 R_{c0}^2 、 R_{v0}^2 、 R_{c1}^2 、 R_{v1}^2 、 R_{MSEc0} 、 R_{MSEv0} 、 R_{MSEc1} 和 R_{MSEv1} 为指标,利用建模集 48 个样本数据和验证集 24 个样本数据,对比评价全子集筛选前后人工神经网络模型、支持向量机模型、分位数回归模型的预测能力,筛选最优估测模型。通过对表 4~6 的分析可见,采用分位数回归建立土壤盐渍化估算模型中 R_{v0}^2 接近于 0,且 R_{MSEv0} 偏差很大,该模型产生严重的过拟合现象,支持向量机和人工神经网络模型在全子集筛选前后均未产生过拟合现象,但在相同深度下全子集筛选前后的人工神经网络模型和支持向量机模型的精度略有不同。在人工神经网络模型中,除 0~40 cm 深度模型以外,其余深度下均为全子集筛选后的模型估算含盐量精度高。在支持向量机模型中,除 40~60 cm 深度的模型以外,其余深度均为全子集筛选之前模型拟合效果好,但其 R_c^2 、 R_v^2 、 R_{MSEc} 和 R_{MSEv} 全子集筛选前后并未产生显著的变化。这表明全子集筛选结果真实可靠,选取的自变量对各个深度土壤含盐量敏感性高。

对比 5 个深度的土壤含盐量估算模型,发现在 20~40 cm 深度下,3 种模型的 R_c^2 和 R_v^2 均达到最高,其中 R_c^2 和 R_v^2 均在 0.58 以上, R_{MSEc} 和 R_{MSEv} 均在 0.21% 以下,3 种模型在该深度下均取得了良好的建模和预测能力。从模型的稳定性 (R_c^2/R_v^2) 来看,人工神经网络为 93.6%、分位数回归模型与支持向量机相差不大,分别为 109.8% 和 112.1%。在模型的拟合度方面,分位数回归模型略高于支持向量机模型,人工神经网络模型最差。支持向量机模型的

R_{c0}^2 分别比分位数回归 R_{c1}^2 和人工神经网络模型 R_{c1}^2 高 0.063 和 0.091;分位数回归模型的 R_{v1}^2 分别比支持向量机 R_{v0}^2 和人工神经网络模型 R_{v1}^2 高 0.07 和 0.048。从模型的预测效果方面考虑,支持向量机模型略高于分位数回归模型,人工神经网络模型最差。人工神经网络模型的 R_{MSEc1} 、 R_{MSEv1} 分别比分位数回归 R_{MSEc1} 、 R_{MSEv1} 和支持向量机模型 R_{MSEc0} 、 R_{MSEv0} 高 0.044、0.005 个百分点和 0.024、0.018 个百分点。

综上所述,人工神经网络模型的反演效果最差,分位数回归模型和支持向量机模型在本次土壤含盐量估算中就模型效果而言,二者基本相同。但是分位数回归模型仅采用 3 个自变量,支持向量机模型采用 16 个自变量。因此,本文选择拟合效果好、验证精度高、模型稳定性强、简洁高效的 20~40 cm 深度的全子集-分位数回归模型作为本次最佳土壤盐渍化估算模型。

基于全子集-分位数回归方法建立模型,进行解放闸灌域土壤含盐量估测,结果如图 3 所示。图中白色部分为城镇所在区域不参与本次含盐量反演。在该时期,灌域内盐渍化程度较轻,有利于作物的生长发育。灌域内轻度盐渍化土所占比例最大为 54%,非盐土所占比重次之,为 26%。这是由于该时期处于作物生长关键期,灌水次数多且灌水量大,土壤盐分淋洗到下层土壤,土壤处于夏季脱盐状态。重盐渍化土所占比重较少,为 13%,对于灌水次数

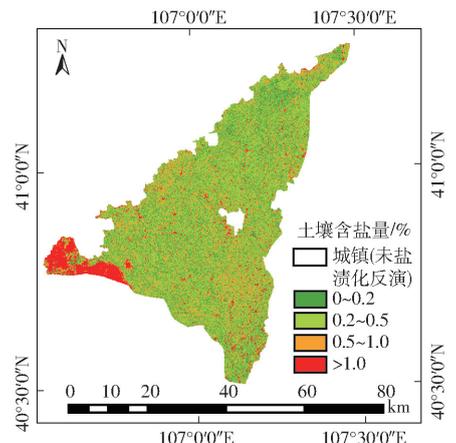


图 3 土壤含盐量空间分布图

Fig. 3 Soil salinity distribution map

较少的地方,土壤仍产生积盐现象。盐土占总面积的比例最少,为7%,主要集中在灌域西部(多为盐荒地),其余盐土零散分布于灌域内,主要包括部分盐荒地和弃耕地。

3 讨论

土壤盐渍化问题在中国北方干旱灌区较为突出,已成为制约灌区农业可持续发展的重要因素。黄河流域上游的河套灌区是受盐渍化影响的典型区,土壤盐渍化面积约占总面积的69%^[37]。利用卫星遥感对土壤含盐量进行大面积监测,符合未来精准农业发展的要求。在本次植被覆盖条件下,作物长势可间接识别土壤根域盐渍化水平,研究结果显示20~40 cm深度下的土壤含盐量与高分一号卫星数据相关性最高,土壤盐分含量对作物主根系的影响明显。史海滨等^[38]对比各生育期内不同含盐处理的各层主根重的变化情况发现,向日葵的苗期后期(7月27日)主根深度在30 cm之内,土壤盐分含量对作物主根系层的生长影响明显。与本文研究成果基本一致。由此可见,应用卫星遥感大范围监测植被覆盖条件下不同土层的盐渍化水平具有一定的可行性。

采用相关性分析只能获得单一自变量与含盐量的相关关系,本研究采用全子集法列举全部组合方式,得到不同深度下对于土壤含盐量最优自变量数目的敏感变量组合,筛选后的结果与未筛选的结果在拟合优度和模型精度方面略有提升。表明全子集筛选过程简单、高效、灵敏度强,适用于本次盐分含量的估测。OKCU等^[39]采用全子集法生成一个新的适用于河流和水槽数据的泥沙运移公式,精准预测泥沙运移的非线性关系。但是由于自变量数目的增加,全子集筛选的运行效率减弱,HOFMANN等^[40]改进和扩展全子集回归的计算方法,采用基于回归树的RadiusBBA和HBBA方法解决大规模模型选择问题的穷举和启发式策略。全子集法和相关性分析法的筛选结果可作为自变量寻优的一种参考,而筛选结果与因变量是否有内在联系,需要做更严谨的物理机制和数学推理工作。

本文对于不同深度选定的敏感指数建立的ANN、SVM、QR 3种模型进行土壤含盐量反演,发现QR模型的反演精度最高。这是由于QR采用非对称权重的方法估计参数,适用于离群值多、变量间关系较弱和需要了解因变量分布的情况,且无需对模型进行任何分布的假定,对于强变异性(变异系数

约100%)的实测土壤盐分具有抗耐性。AMAKOR^[19]进行钙质盐碱土电导率模拟时,实测电导率变异系数在100%左右,采用QR模型决定系数 R^2 达到了0.88。MUELLER等^[41]、GERBER等^[42]和CARSLAW等^[43]采用QR模型,均很好地解决了实测数据重尾分布、变异性强的问题。 $\tau=0.5$ 分位数处于土壤含盐量的中间位置,采用该分位数进行各个深度的土壤盐渍化反演可以保证轻、重度盐渍化土权重较大,非盐土和盐土权重较小,从而减小离群值对于模型精度的影响。前人也得到类似的结果,如王蕾等^[36]发现在0.5分位数下进行冬小麦估测结果最为可靠。

研究植被覆盖条件下不同深度的解放闸灌域土壤盐渍化模型,对改善灌区土壤盐渍化情况具有现实意义。然而最优土壤含盐量反演模型会因作物的种类、生育期、天气状况、灌溉情况、甚至所使用的遥感平台而异。本文所得的最佳含盐量反演模型也仅限于本次测量结果,研究区内其他生育阶段和其他地区还可尝试更多图像特征(如颜色特征^[44]、地形特征^[45]),寻找更优的含盐量反演指标,从而建立精度更高且实用性更强的土壤含盐量反演模型。

4 结论

(1)全子集筛选具有全局性,经过全子集筛选后敏感变量组合方式不尽相同。B4、BI、SI1、SI3是0~20 cm、0~40 cm处土壤含盐量的敏感变量组合,B4、BI、NDVI为20~40 cm、40~60 cm、0~60 cm处土壤含盐量的敏感变量组合。筛选后的敏感变量组合方式仅用3~4个自变量,与筛选之前的效果相差不多,因而全子集筛选具有灵敏度强、简单、高效的优点。

(2)在不同深度下的土壤含盐量反演模型中,分位数回归、人工神经网络、支持向量机模型均取得良好的建模验证精度。分位数回归模型由于其抗异性较强,其精度与支持向量机模型相差不多,优于人工神经网络模型。土壤深度对含盐量反演模型精度具有一定的影响。在使用相同的建模方法时,20~40 cm深度下反演模型效果优于其他深度。

(3)在20~40 cm深度下建立的模型效果最为理想,且分位数回归抗异性、稳健性强,在保证模型估测精度的前提下,比支持向量机模型和人工神经网络模型操作简单、结果可靠。因此,全子集-分位数回归模型是本次估算植被覆盖条件下土壤含盐量的最优模型。

参 考 文 献

- [1] 吴加敏,姚建华,张永庭,等. 银川平原土壤盐渍化与中低产田遥感应用研究[J]. 遥感学报, 2007, 11(3): 414-419.
WU Jiamin, YAO Jianhua, ZHANG Yongting, et al. Remote sensing application study on the soil salinity and the medium & low yield croplands in Yinchuan plain(YCP)[J]. Journal of Remote Sensing, 2007, 11(3): 414-419. (in Chinese)
- [2] 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 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): 107-114. (in Chinese)
- [3] HARTI A E, 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.
- [4] 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 in NDVI-SI feature space [J]. Transactions of the CSAE, 2010, 26(8): 168-173. (in Chinese)
- [5] 王海峰,张智韬,ARNON Karnieli,等. 基于灰度关联-岭回归的荒漠土壤有机质含量高光谱估算[J]. 农业工程学报, 2018, 34(14): 124-131.
WANG Haifeng, ZHANG Zhitao, ARNON Karnieli, et al. Hyperspectral estimation of desert soil organic matter content based on gray correlation-ridge regression model[J]. Transactions of the CSAE, 2018, 34(14): 124-131. (in Chinese)
- [6] 张智韬,王海峰,ARNON Karnieli,等. 基于岭回归的土壤含水率高光谱反演研究[J/OL]. 农业机械学报, 2018, 49(5): 240-248.
ZHANG Zhitao, WANG Haifeng, ARNON Karnieli, et al. Inversion of soil moisture content from hyperspectra based on ridge regression[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(5): 240-248. http://www.jcsam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20180528&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2018.05.028 (in Chinese)
- [7] UPADHYAYA S, RAMSANKARAN R. Modified-INSAT multi-spectral rainfall algorithm (M-IMSRA) at climate region scale: development and validation[J]. Remote Sensing of Environment, 2016: 186-187.
- [8] 张海威,张飞,张贤龙,等. 光谱指数的植被叶片含水量反演[J]. 光谱学与光谱分析, 2018, 38(5): 1540-1546.
ZHANG Haiwei, ZHANG Fei, ZHANG Xianlong, et al. Inversion of vegetation leaf water content based on spectral index[J]. Spectroscopy and Spectral Analysis, 2018, 38(5): 1540-1546. (in Chinese)
- [9] 李长春,牛庆林,杨贵军,等. 基于无人机数码影像的大豆育种材料叶面积指数估测[J/OL]. 农业机械学报, 2017, 48(8):147-158.
LI Changchun, NIU Qinglin, YANG Guijun, et al. Estimation of leaf area index of soybean breeding materials based on UAV digital images[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(8): 147-158. http://www.jcsam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20170816&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2017.08.016 (in Chinese)
- [10] 冯克伟. 野生二粒小麦优异耐盐种质的鉴定及其相关基因的发掘[D]. 杨凌:西北农林科技大学, 2018.
FENG Kewei. Identification and analysis of salt tolerance germplasms and related genes in wild emmer wheat[D]. Yangling: Northwest A&F University, 2018. (in Chinese)
- [11] FAN X, WENG Y, TAO J. Towards decadal soil salinity mapping using Landsat time series data[J]. International Journal of Applied Earth Observation and Geoinformation, 2016, 52: 32-41.
- [12] YAO R, YANG J. Quantitative evaluation of soil salinity and its spatial distribution using electromagnetic induction method [J]. Agricultural Water Management, 2010, 97(12): 1961-1970.
- [13] 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)
- [14] VERMEULEN D, NIEKERK A V. Machine learning performance for predicting soil salinity using different combinations of geomorphometric covariates[J]. Geoderma, 2017, 299: 1-12.
- [15] CHEN X, LI X. Using quantile regression to analyze the stressor-response relationships between nutrient levels and algal biomass in three shallow lakes of the Lake Taihu Basin, China[J]. Chinese Science Bulletin, 2014, 59(28): 3621-3629.
- [16] LIANG L, DI L, ZHANG L, et al. Estimation of crop LAI using hyperspectral vegetation indices and a hybrid inversion method[J]. Remote Sensing of Environment, 2015, 165: 123-134.
- [17] 胡丽,赵南京,刘文清,等. 基于多元校正的水体 Pb 元素 LIBS 定量分析[J]. 光学学报, 2015, 35(6): 322-328.
HU Li, ZHAO Nanjing, LIU Wenqing, et al. Quantitative analysis of Pb in water based on multivariate calibration with LIBS [J]. Acta Optica Sinica, 2015, 35(6): 322-328. (in Chinese)
- [18] 陈小华,李小平,钱晓雍,等. 基于分位数回归的洱海藻类对氮、磷及水温的响应特征[J]. 环境科学, 2017, 38(1): 113-120.
CHEN Xiaohua, LI Xiaoping, QIAN Xiaoyong, et al. Response characteristics of algal chlorophyll-a to Nitrogen, Phosphorus and water temperature in lake Erhai based on quantile regression[J]. Environmental Science, 2017, 38(1): 113-120. (in Chinese)
- [19] AMAKOR X N. Using electromagnetic induction sensing to understand the dynamics and interacting factors controlling soil salinity[D]. Logan: Utah State University, 2013.

- [20] 李亮. 内蒙古河套灌区耕荒地间土壤水盐运移规律研究[D]. 呼和浩特:内蒙古农业大学, 2008.
LI Liang. Study on the transport of soil water and salt between irrigated land and salineland in Hetao Irrigation District[D]. Huhhot: Inner Mongolia Agricultural University, 2008. (in Chinese)
- [21] ALLBED A, KUMAR L, SINHA P. Mapping and modelling spatial variation in soil salinity in the al hassa oasis based on remote sensing indicators and regression techniques[J]. *Remote Sensing*, 2014, 6(2): 1137–1157.
- [22] 王伦平, 陈亚新, 曾国芳. 内蒙古河套灌区灌溉排水与盐碱化防治[M]. 北京: 水利电力出版社, 1993.
- [23] 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.
- [24] SIDIKE A, ZHAO S, WEN Y. Estimating soil salinity in Pingluo County of China using QuickBird data and soil reflectance spectra[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2014, 26: 156–175.
- [25] CHO K H, BEON M, JEONG J. Dynamics of soil salinity and vegetation in a reclaimed area in Saemangeum, Republic of Korea[J]. *Geoderma*, 2018, 321: 42–51.
- [26] 冯娟, 丁建丽, 杨爱霞, 等. 干旱区土壤盐渍化信息遥感建模[J]. 干旱地区农业研究, 2018, 36(1): 266–273.
FENG Juan, DING Jianli, YANG Aixia, et al. Remote sensing modeling of soil salinization information in arid areas[J]. *Agricultural Research in the Arid Areas*, 2018, 36(1): 266–273. (in Chinese)
- [27] MERCIER G, MOSER G, SERPICO S. Conditional copulas for change detection in heterogeneous remote sensing images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2008, 46(5): 1428–1441.
- [28] KRISHNASWAMY J, BAWA K S, GANESHIAH K N, et al. Quantifying and mapping biodiversity and ecosystem services: utility of a multi-season NDVI based Mahalanobis distance surrogate[J]. *Remote Sensing of Environment*, 2009, 113(5): 904–906.
- [29] SCHUG F, OKUJENI A, HAUER J, et al. Mapping patterns of urban development in Ouagadougou, Burkina Faso, using machine learning regression modeling with bi-seasonal Landsat time series[J]. *Remote Sensing of Environment*, 2018, 210: 217–228.
- [30] FEILHAUER H, ASNER G P, MARTIN R E. Multi-method ensemble selection of spectral bands related to leaf biochemistry [J]. *Remote Sensing of Environment*, 2015, 164: 57–65.
- [31] ATKINSON P M, JEGANATHAN C, DASH J, et al. Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology[J]. *Remote Sensing of Environment*, 2012, 123: 400–417.
- [32] HANGSHING L, DABRAL P P. Multivariate frequency analysis of meteorological drought using copula[J]. *Water Resources Management*, 2018, 32(5): 1741–1758.
- [33] 徐雯靓, 王少军. PROSAIL 模型模拟下的植被指数土壤调节能力比较与适用环境分析[J]. 遥感学报, 2014, 18(4): 826–842.
XU Wenjing, WANG Shaojun. Soil-adjusted power comparison and application conditions of vegetation indices based on PROSAIL model[J]. *Journal of Remote Sensing*, 2014, 18(4): 826–842. (in Chinese)
- [34] 赵英时. 遥感应用分析原理与方法[M]. 北京: 科学出版社, 2003.
- [35] CSILLAG F, PÁSZTOR L, BIEHL L L. Spectral band selection for the characterization of salinity status of soils[J]. *Remote Sensing of Environment*, 1993, 43(3): 231–242.
- [36] 王蕾, 王鹏新, 李俐, 等. 基于 VTCI 和分位数回归模型的冬小麦单产估测方法[J/OL]. 农业机械学报, 2017, 48(7): 167–173, 166.
WANG Lei, WANG Pengxin, LI Li, et al. Winter wheat yield estimation method based on quantile regression model and remotely sensed vegetation temperature condition index [J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2017, 48(7): 167–173, 166. http://www.jcsam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20170721&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2017.07.021. (in Chinese)
- [37] 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)
- [38] 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, 2009.
- [39] OKCU D, PEKTAS A O, UYUMAZ A. Creating a non-linear total sediment load formula using polynomial best subset regression model[J]. *Journal of Hydrology*, 2016, 539: 662–673.
- [40] HOFMANN M, GATU C, KONTOGHIORGHES E J. Efficient algorithms for computing the best subset regression models for large-scale problems[J]. *Computational Statistics & Data Analysis*, 2007, 52(1): 16–29.
- [41] MUELLER N D, RHINES A, BUTLER E E, et al. Global relationships between cropland intensification and summer temperature extremes over the last 50 years[J]. *Journal of Climate*, 2017, 30(18): 7505–7528.
- [42] GERBER F, JONG R D, SCHAEPMAN M E, et al. Predicting missing values in spatio-temporal remote sensing data[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(5): 2841–2853.
- [43] CARSLAW D C, WILLIAMS M L, TATE J E, et al. The importance of high vehicle power for passenger car emissions[J]. *Atmospheric Environment*, 2013, 68: 8–16.
- [44] 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 HSI image fusion[J]. *Transactions of the CSAE*, 2017, 33(21): 173–180. (in Chinese)
- [45] PENG J, BISWAS 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.