

基于环境变量和机器学习的土壤水分反演模型研究

王思楠¹ 李瑞平^{1,2} 吴英杰³ 赵水霞³ 王秀青⁴

- (1. 内蒙古农业大学水利与土木建筑工程学院, 呼和浩特 010018;
- 2. 内蒙古自治区农牧业大数据研究与应用重点实验室, 呼和浩特 010018;
- 3. 中国水利水电科学研究院牧区水利科学研究所, 呼和浩特 010020;
- 4. 内蒙古自治区测绘地理信息中心, 呼和浩特 010050)

摘要: 为利用多源数据构建毛乌素沙地腹部土壤含水率建模指示因子,通过微波后向散射系数、地表温度、缨帽变换要素、波段反射率、干旱指数和地形要素等 17 个变量为建模因子,分别以偏最小二乘 (Partial least squares regression, PLSR)、极限学习机 (Extreme learning machine, ELM) 和随机森林 (Random forest, RF) 3 种方法构建土壤含水率反演模型,对模型进行验证和对比,并对研究区的土壤水分分布进行制图。结果表明:温度植被干旱指数是土壤水分空间变异性的最重要的预测因子 (决定系数为 0.64),其次是地表温度 (0.6)、 σ_{vv} (0.38)、植被指数 (0.38)、波段 7 反射率 (0.35)、 σ_{vh} (0.32)、波段 6 反射率 (0.3) 和反照率 (0.26)。相比于未筛选变量所构建的模型,利用最优子集筛选 (Best subset selection, BSS) 变量所构建的模型精度均有所提升。其中 PLSR 在处理共线性方面表现最优,ELM 回归模型最稳定。RF 模型具有最高的精确度,4 月,决定系数为 0.74,均方根误差为 8.85%,平均绝对误差为 7.86%;8 月,决定系数为 0.75,均方根误差为 8.86%,平均绝对误差为 7.41%。不同方法反演的土壤水分分布趋势没有显著差异,高土壤含水率出现在研究区的北部和东南部,中北部平坦地区的土壤含水率较低。利用光谱指数、环境因子和地形数据构建的多因子、多指数综合模型能较高精度地反演毛乌素沙地腹部表层土壤水分,对研究该地区土地荒漠化和生态环境治理具有参考价值。

关键词: 土壤水分; 遥感反演; 随机森林; 极限学习机; 偏最小二乘回归; 环境变量

中图分类号: TP79 **文献标识码:** A **文章编号:** 1000-1298 (2022) 05-0332-10

OSID: 

Soil Moisture Inversion Based on Environmental Variables and Machine Learning

WANG Sinan¹ LI Ruiping^{1,2} WU Yingjie³ ZHAO Shuixia³ WANG Xiuqing⁴

- (1. College of Water Conservancy and Civil Engineering, Inner Mongolia Agricultural University, Huhhot 010018, China
- 2. Inner Mongolia Autonomous Region Key Laboratory of Big Data Research and Application of Agriculture and Animal Husbandry, Huhhot 010018, China
- 3. Institute of Water Resources for Pastoral Area, China Institute of Water Resources and Hydropower Research, Huhhot 010020, China
- 4. Inner Mongolia Autonomous Region Surveying and Mapping Geographic Information Center, Huhhot 010050, China)

Abstract: In order to construct the modeling indicators of soil moisture content in the Mu Us sandy land using multi-source data, totally 17 variables, including microwave backscattering coefficient, surface temperature, silk hat transform factor, band reflectance, drought index and topographic factor were used as modeling factors. PLSR, extreme learning machine (ELM) and random forest (RF) were used to construct soil water content inversion models, verify and compare the models, and map soil water distribution in the study area. The results showed that the determination coefficient of temperature vegetation drought index was 0.64, followed by land surface temperature (0.6), σ_{vv} (0.38), vegetation index (0.38), band 7 reflectance (0.35), σ_{vh} (0.32), band 6 reflectance (0.3) and Albedo (0.26). Compared with the model constructed with unscreened variables, the accuracy of the model constructed with best subset selection (BSS) variables was improved. PLSR had the best performance in

收稿日期: 2021-05-24 修回日期: 2021-10-21

基金项目: 内蒙古自治区科技计划项目 (201802123) 和国家自然科学基金项目 (52069021, 51839006)

作者简介: 王思楠 (1993—), 男, 博士生, 主要从事水利遥感信息技术研究, E-mail: nmgnxdx2016@163.com

通信作者: 李瑞平 (1973—), 男, 教授, 博士生导师, 主要从事节水灌溉与农业水利遥感信息技术研究, E-mail: nmglrp@163.com

collinearity, and ELM regression model was the most stable. RF model had the highest accuracy, with a determination coefficient of 0.74, root mean square error of 8.85% and mean absolute error of 7.86% in April. In August, the determination coefficient was 0.75, the root mean square error was 8.86%, and the mean absolute error was 7.41%. There was no significant difference in soil water distribution trend between different methods. The highest soil water content occurred in the north and southeast of the study area, and the lower soil water content occurred in the flat area in the central and northern part of the study area. Using spectral index, environmental factor and topographic data, the multi-factor and multi-index comprehensive model can accurately retrieve the surface soil moisture in the Mu Us sandy land, which had reference value for the study of land desertification and ecological environment control in this area.

Key words: soil moisture; remote sensing inversion; random forest; extreme learning machine; partial least squares regression; environmental variables

0 引言

土壤含水率是地表能量平衡的重要决定参数,在全球水循环中起着重要作用^[1]。同时在农业应用中,土壤水分含量不仅是作物生长发育的基本条件,也是作物产量估算、干旱监测的关键参数^[2]。因此,准确、及时地反演土壤水分具有重要意义。

土壤水分的分布受多个相互作用的因素影响,如土壤特性、植被覆盖和气候条件^[3]。因此,用传统的单点测量方法,如烘干法、数字探头等,相对难以有效地获得大规模的土壤水分信息^[4]。与地面单点测量方法相比,遥感技术因其覆盖面广、时效性强、成本低,逐渐得到应用^[5-6]。光学遥感作为最早和最成熟的地球观测技术,一直发挥着重要的作用。学者们通过地物反射辐射特征变化来模拟地表覆盖类型、地表温度、土壤热惯量及地表蒸散发与土壤含水量的经验关系,从而实现土壤水分的反演^[7-9]。然而,上述大多数方法都是经验性的,对于回归分析需要做出的一些统计假设可能会导致使用受限,如离群数据、非线性、异方差和多共线性,不能对土壤水分进行强有力的预测。机器学习方法能够发现数据之间有意义的关联、模式和规则,可以克服上述问题,如人工神经网络^[10]、卷积神经网络^[11]、支持向量机^[12]、极限学习机(Extreme learning machine, ELM)^[13]、随机森林(Random forest, RF)^[14],已经成功地应用于干旱半干旱地区土壤水分反演。研究证明基于机器学习方法能够构建综合多特征变量的土壤水分反演模型,可以提高土壤水分反演的时空分辨率以及反演精度^[15-17]。研究表明,偏最小二乘回归法、极限学习机和随机森林是定量反演的可行方法^[18-19]。总体而言,上述模型方法各具优缺点,对不同研究区域有其不同的要素要求和适用范围,有的土壤水分反演方法虽精度较高,但所需测定要素容量大,有的模型虽解决了大范围应用的难题,却在参数获取和模型建立方面要求很高。

本文以内蒙古自治区乌审旗作为研究区,对不同环境变量与表层土壤含水率进行相关分析,利用最优子集筛选环境变量作为模型输入变量,利用偏最小二乘(Partial least squares regression, PLSR)、极限学习机和随机森林等方法构建不同的土壤含水率反演模型,通过比较,探索毛乌素沙地腹部土壤含水率的高效反演模型及方法,在此基础上,反演研究区不同月份土壤水分并分析其时空变化。

1 数据与方法

1.1 研究区概况

乌审旗位于鄂尔多斯市西南部(图 1),地处毛乌素沙地腹部(37°38'~39°23'N, 108°17'~109°40'E),地势由西北向东南倾斜,平均海拔 1 305 m。年平均降雨量 350~400 mm,年平均气温 6.8℃。主要土壤类型有栗钙土、草甸土、盐碱土、沼泽潜育土以及各类风沙土。主要土地利用类型为草地、沙地、林地、耕地、水体和建筑用地。

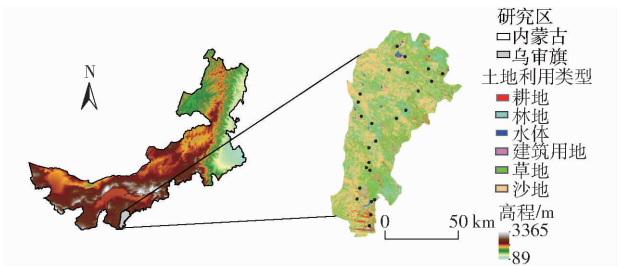


图 1 研究区地理位置与采样区分布图
Fig. 1 Geographical location and sampling layout of study area

1.2 数据

1.2.1 遥感数据

选取 2016 年 4 月 21 日、8 月 27 日的 Landsat8 OLI 影像,数据来源于美国地质调查局网站(<https://earthexplorer.usgs.gov>),通过 ENVI 5.3 软件完成影像预处理,具体包括:辐射定标、大气校正、影像镶嵌、裁剪和波段运算。从而进一步得到地表温度、地

表反照率、缨帽变换要素、反射率、植被指数、水体指数、建筑指数和干旱指数等环境变量。其中植被指数和地表温度是表述地表特征的两个重要参数,以及二者构成的温度植被干旱指数均与地表土壤水分存在着紧密关系,是土壤水分监测的常用参数。反照率是影响地表辐射平衡的一个重要参数,其可以进一步影响土壤水分的变化。不同的光谱反射率对土壤水分的敏感性也不一样。缨帽变换要素的湿度对不同植被覆盖的土壤水分也有一定的表征。水体指数与建筑指数可以直接影响地表温度来间接影响土壤水分。哨兵 1A 数据是 Level-1 地距影像 (Ground range detected, GRD), 成像方式为干涉宽幅 (Interferometric wide swath, IW) 模式、极化方式 (Vertical vertical (VV) 和 Vertical horizontal (VH)), 经过辐射标定等得到所需要的后向散射系数。考虑地形条件对土壤水分的影响,本研究还通过地理空间数据云网站 (<http://www.gscloud.cn/>) 获取 ASTER 数字高程模型。

1.2.2 野外实测数据

根据卫星过境时间在 2016 年 4 月 20—22 日、8 月 26—28 日进行了两次地面采样实验。采样区域共 24 个,每个采样区域之间的最小间隔为 1 km。采样时,首先记录每个采样区域 5 个点的 GPS 位置信息,各点间隔 30 m,然后用土钻从土壤表面 0~10 cm 处分别采集土壤样品。最后,在现场对这 5 个点的土壤样品进行混合和称量。使用烘干称量法进行土壤含水率的测量,计算公式为

$$W = \frac{W_2 - W_3}{W_3 - W_1} \times 100\%$$

式中 W ——土壤质量含水率, %
 W_1 ——空土盒质量, g
 W_2 ——土盒和湿土质量, g
 W_3 ——土盒和干土质量, g

1.3 最优子集筛选

全子集回归是对所有预测变量的可能组合模型都进行拟合,然后根据贝叶斯信息准则 (Bayesian information criterion, BIC) 筛选出现有变量条件下的最佳模型,又叫最优子集筛选,计算公式为

$$BIC = m \ln c - 2 \ln L$$

式中 m ——模型参数个数
 c ——样本数量 L ——最大似然函数

1.4 建模方法及模型评价指标

偏最小二乘回归利用对系统中的数据信息进行分解和筛选的方式,提取对因变量的解释性最强的综合变量,辨识系统中的信息与噪声,从而更好地克服变量多重相关性在系统建模中的不良作用^[20]。

极限学习机是在单隐层前馈神经网络上发展起来的一种机器学习方法,可以随机初始化输入权重和偏置并得到相应的输出权重,弥补了传统神经网络中运行时间长等缺点^[21]。随机森林是多重决策树的组合,输出类别由个别树输出的类别众数而决定的分类回归模型。不需要关于响应协变量关系的分布假设,该过程通过平均决策树进行统计上的可靠估算,降低过拟合风险^[22]。

采用决定系数 R^2 、均方根误差 (RMSE) 和平均绝对误差 (MAE) 验证反演模型的精度。RMSE、MAE 越小, R^2 越大, 表明模型反演精度越高。

2 结果与分析

2.1 描述性统计

为了对预测模型进行训练和验证,分别将 2016 年 4 月 21 日、8 月 27 日土壤含水率数据 (120 个) 随机分为两组:建模集 (70%, 84 个样点) 对模型进行训练,验证集 (30%, 36 个样点) 对模型进行检验。全集、建模集和验证集描述性统计结果如图 2 所示。4 月,全集土壤含水率均值为 4.84%, 标准差为 5.18%。建模集 (0.45%~20.13%) 和验证集 (0.61%~15.89%) 均值分别为 4.73% 和 5.90%; 8 月,全集土壤含水率均值为 7.51%, 标准差为 6.98%。建模集 (0.82%~24.67%) 和验证集

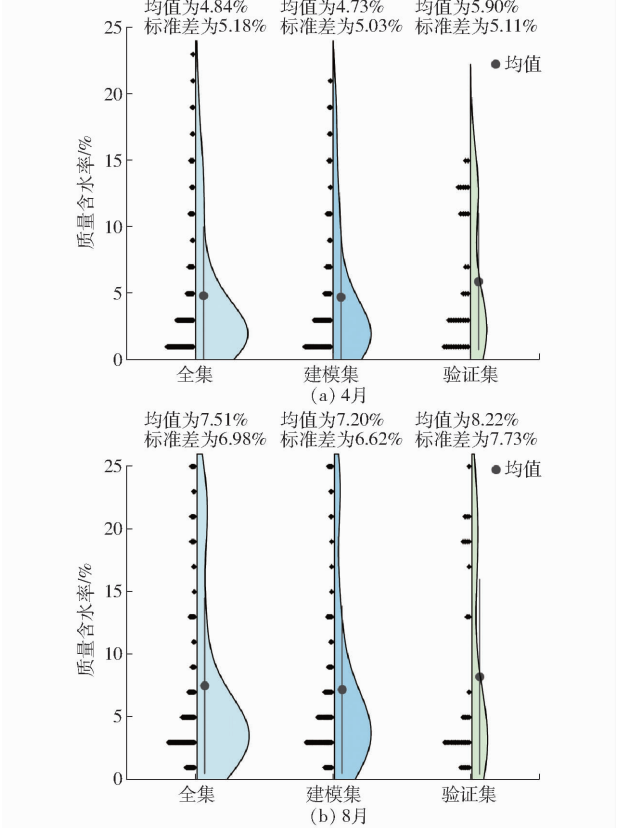


图2 土壤含水率描述性统计

Fig.2 Descriptive statistics of soil moisture content

(1.19% ~ 24.00%) 均值分别为 7.20% 和 8.22%。表明建模集和验证集与全集土壤含水率保持类似的统计分布,在确保代表性样本的同时尽可能缩减了建模集和验证集中存在偏差的估计。

2.2 环境变量与表层土壤含水率的相关性分析

基于机器学习环境变量的数据来源与性质,包括 17 个变量:后向散射系数 (σ_{VV} 、 σ_{VH})、地表温度 (Land surface temperature, LST)、地表反照率

(Albedo)、亮度 (Bright)、绿度 (Green)、湿度 (Wet)、波段反射率 (红光、近红外、短波红外)、植被指数 (Normalized difference vegetation index, NDVI)、水体指数 (Normalized differential water index, NDWI)、建筑指数 (Normalized differential building index, NDBI)、高程 (DEM)、坡度 (Slope) 和温度植被干旱指数 (Temperature vegetation drought index, TVDI) 等要素 (图 3)。

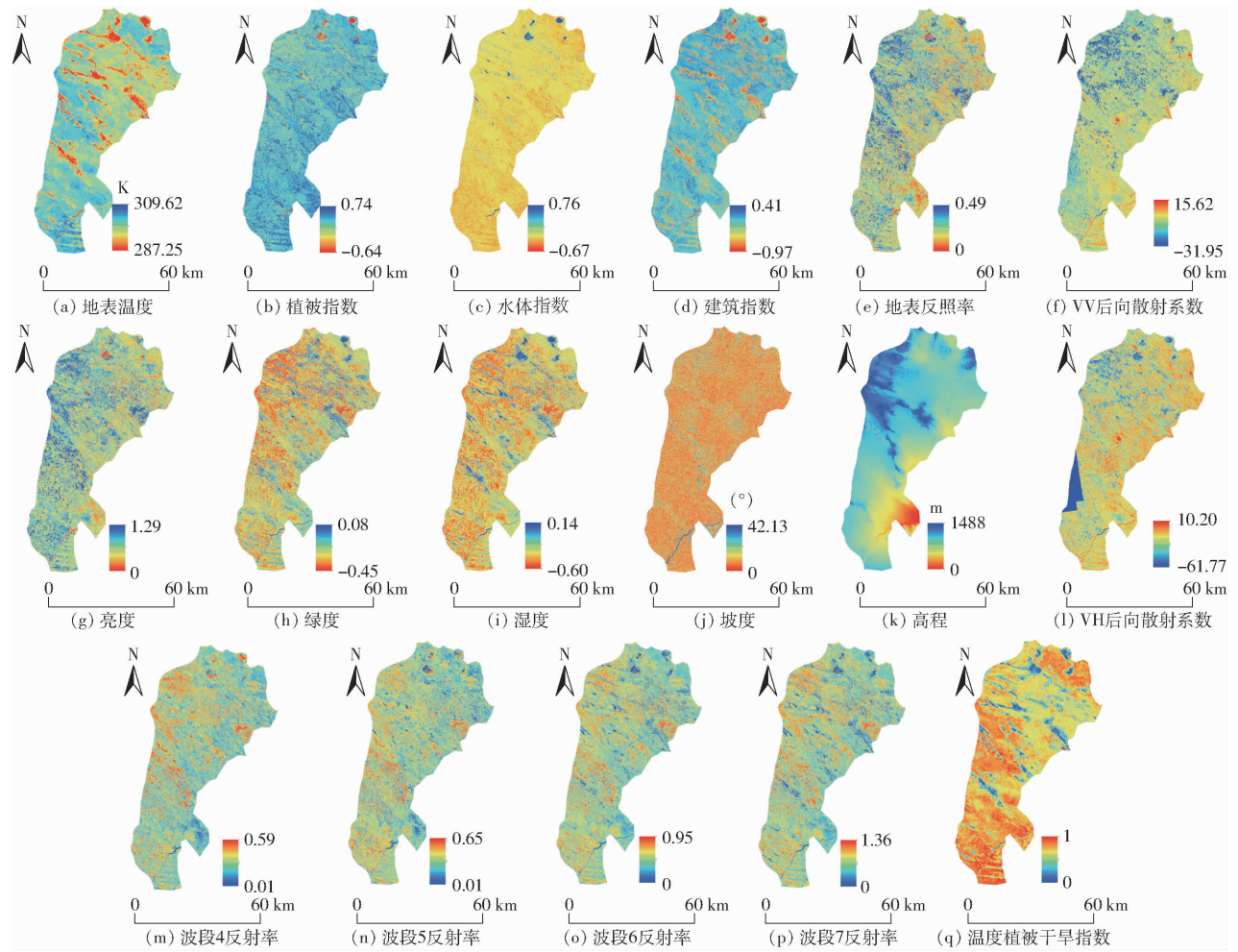


图 3 环境变量空间分布

Fig. 3 Surface biophysical characteristics and topographic parameters

通过对不同环境变量与表层土壤含水率进行相关分析,结果发现不同的环境变量与表层土壤含水率之间的相关程度不同 (图 4 中 *、** 分别表示 $P < 0.05$ 、 $P < 0.01$)。首先,温度植被干旱指数与土壤含水率相关性最高,决定系数为 0.64,主要原因是温度植被干旱指数综合植被和地表温度信息,从裸地到植被全覆盖可以准确体现土壤湿度状况^[23]。其次为地表温度与土壤含水率相关性,决定系数为 0.6,地表温度在土壤湿度和地、气相互循环过程中有重要作用,同时能间接反映土壤水分状况,表征旱情分布,因而对地表温度数据的研究是不可或缺的内容^[24],本研究地表温度与土壤含水率为高相关

性,对土壤含水率高精准反演有重要价值。最后, σ_{VV} 、植被指数、波段 7 反射率、 σ_{VH} 、波段 6 反射率、建筑指数、水体指数、反照率、亮度、绿度和湿度与土壤含水率之间的决定系数分别为 0.38、0.38、0.35、0.32、0.30、0.18、0.13、0.26、0.14、0.15、0.06,其中 NDVI 与土壤含水率的相关性从 4 月到 9 月增加最明显,主要原因是 NDVI 在植被覆盖度高的月份更加敏感,反映的信息更为丰富;NDBI 与土壤含水率的相关性从 4 月到 9 月减小最明显,主要原因是 NDBI 在植被覆盖度低的月份更加敏感。 σ_{VV} 、 σ_{VH} 在植被覆盖度低的月份对土壤含水率更加敏感。地形因素中 DEM 和坡度的正向相关性较好,这主要是地

形会影响地表径流与该地所受的太阳光照、植被生长环境等^[25],从而对土壤水分产生影响,此外乌审旗的地形呈现西北高东南低的特征,这对DEM和坡度间相关性影响较大。反射率与土壤含水率也有很好的相关性,主要是由于土壤水分的变化会引起土壤颗粒物理性质和反射率也发生变化,通常随着土壤水分的减少,其反射率也会相应增大^[25];而植被的生长状态能够间接反映出土壤水分,通常来说,植被密度越大,生长态势越好的地方的土壤含水率越高。其中,波段反射率中B6、B7波段反射率与表层土壤含水率的相关性高于B4、B5波段,主要原因是短波红外光谱域包括水分吸收带,影响植被和土壤中水分含量的反射率敏感性^[26]。

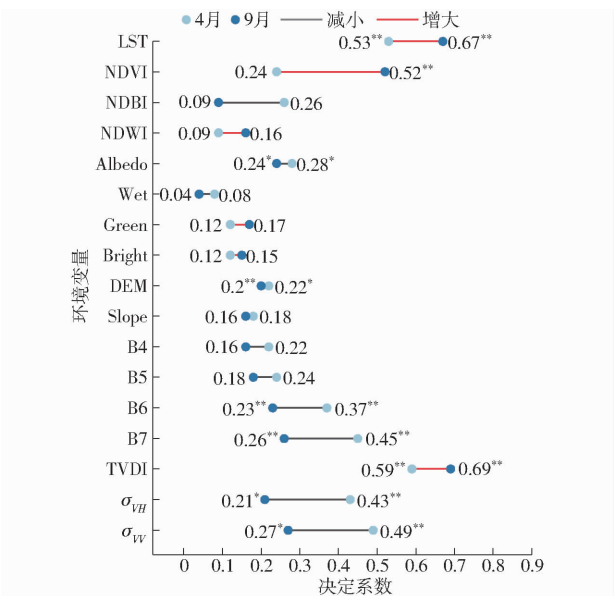


图4 不同月份环境变量与土壤含水率的相关性分析
Fig.4 Correlation analysis of environmental variables and soil moisture in different months

总体而言,2个月间表层土壤含水率与后向散射系数均值 R^2 (0.35)、波段反射率均值 R^2 (0.26)、地表温度、地表反照率、缨帽变换要素、植被指数、水体指数、建筑指数均值的 R^2 (0.24)以及干旱指数均值的 R^2 (0.6)高于表层土壤含水率与地形参数的均值 R^2 (0.19)。这一结果表明,表层土壤含水率对研

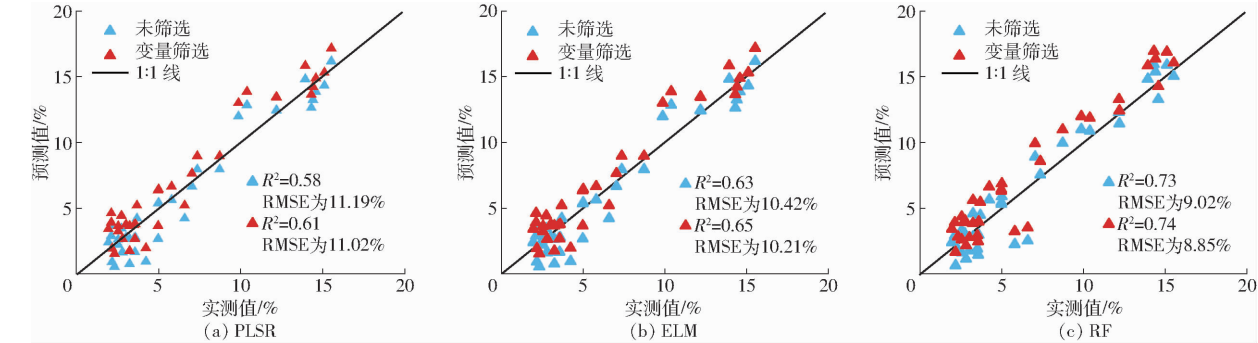


图5 4月土壤含水率实测值与预测值比较

Fig.5 Comparison of measured and predicted soil moisture content in April

究区后向散射系数、干旱指数、波段反射率和地表生物物理特征的依赖性较高。

2.3 最优子集筛选特征变量

基于BIC最小的原则,利用最优子集筛选法(Best subset selection, BSS)选取不同月份最优的环境变量组合,如表1所示。

表1 全子集筛选结果统计
Tab.1 Statistics of full subset selection

时期	最优变量组合	R^2	贝叶斯信息准则
4月	σ_{VV} 、TVDI、LST、DEM、B6、B7	0.73	-56
8月	TVDI、LST、DEM、NDVI	0.68	-43

2.4 土壤水分反演模型的精度评价与分析

为了验证3类模型的可靠性和适用性,使用2期验证集中的36个样点实测数据分别进行精度检验。图5、6分别为2016年4月21日和2016年8月27日PLSR、ELM、RF3种模型预测值与实测数据的散点图。由图5、6可知,预测值与实测值都具有很高的相关性。由表2可知,4月,建模集中3种模型方法的精度差异较大,相比于未筛选变量所构建的模型,利用BSS筛选变量所构建的模型精度均有所提升,且在RF模型的 R^2 最高,为0.88,而RMSE和MAE均为最小值,分别为8.25%、5.17%;PLSR模型的 R^2 最低,为0.66, RMSE和MAE分别为10.07%、9.43%,均为建模集中最大值;ELM模型性能介于二者间。由表3可知,8月,建模集中3种模型方法的精度差异较大,相比于未筛选变量所构建的模型,利用BSS筛选变量所构建的模型精度均有所提升,且在RF模型的 R^2 最高,为0.89,而RMSE和MAE均为最小值,分别为7.36%、4.61%;PLSR模型的 R^2 最低,为0.69, RMSE和MAE分别为9.74%、9.37%,均为建模集中最大值;ELM模型性能介于二者间。综上所述,建模集中,RF模型性能明显优于PLSR和ELM方法,反演效果由优到劣依次为RF、ELM、PLSR。在验证集中,3种模型的各项指标相较于建模集均无明显下降,表明模型比较稳

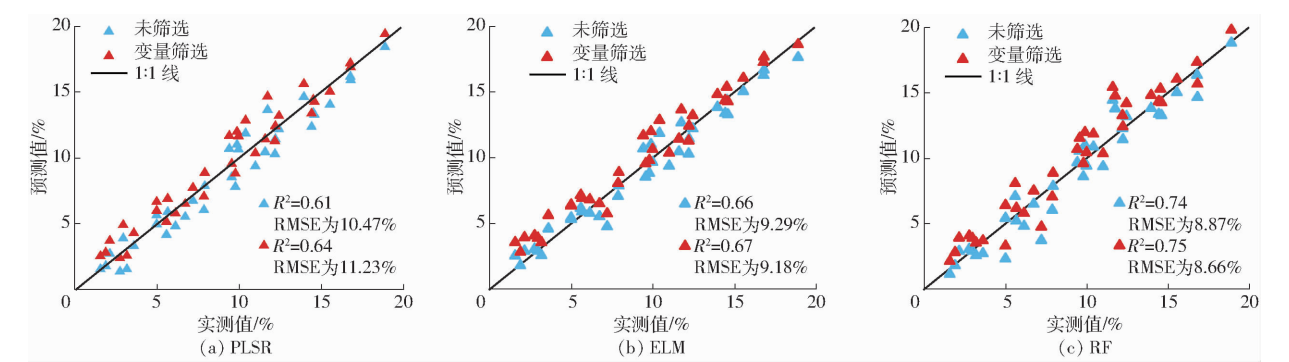


图 6 8 月土壤含水率实测值与预测值比较

Fig. 6 Comparisons of measured and predicted soil moisture content in August

表 2 4 月土壤含水率反演模型与精度

Tab. 2 Soil moisture content retrieval models and accuracies in April

模型	变量选择方法	建模集			验证集		
		R^2	RMSE/ %	MAE/ %	R^2	RMSE/ %	MAE/ %
PLSR	—	0.64	10.18	9.56	0.58	11.19	11.24
	BSS	0.66	10.07	9.43	0.61	11.02	11.15
ELM	—	0.73	9.54	7.57	0.63	10.42	9.57
	BSS	0.74	9.47	7.54	0.65	10.21	9.42
RF	—	0.86	8.34	5.28	0.73	9.02	7.97
	BSS	0.88	8.25	5.17	0.74	8.85	7.86

注：“—”和“BSS”分别表示未使用变量选择方法筛选变量和使用最优子集筛选变量，下同。

定,但三者相比较而言,仍然是 RF 模型的各项指标明显优于 ELM 和 PLSR,而 ELM 和 PLSR 相比,ELM 模型的 R^2 明显高于 PLSR 模型,而 RMSE 和 MAE 低于 PLSR 模型。综合考察各模型建模集与验证集的评价指标,3 种模型的预测性能和稳定性从高到低排序依次为 RF、ELM、PLSR,基于 PLSR 的模型精度最低,这是由于土壤的组分非常复杂,功能团多样,偏最小二乘回归方法仅仅将环境变量与土壤含水率进行线性回归,同时没有将部分与土壤含水率非线性相关的关系考虑进来,导致所构建的模型具有一

表 3 8 月土壤含水率反演模型与精度

Tab. 3 Soil moisture content retrieval models and accuracies in August

模型	变量选择方法	建模集			验证集		
		R^2	RMSE/ %	MAE/ %	R^2	RMSE/ %	MAE/ %
PLSR	—	0.67	9.88	9.57	0.61	10.47	9.63
	BSS	0.69	9.74	9.37	0.64	10.23	9.47
ELM	—	0.75	8.79	7.14	0.66	9.29	8.57
	BSS	0.76	8.61	7.06	0.67	9.18	8.46
RF	—	0.87	7.58	4.66	0.74	8.87	7.46
	BSS	0.89	7.36	4.61	0.75	8.66	7.41

定的缺陷。而 ELM、RF 模型充分考虑了环境变量对表层土壤含水率的影响,并且具有对非线性问题的强解析能力和模型的稳健性。研究区表层土壤含水率模型预测精度结果进一步表明了 RF 模型自身的优越性和利用环境变量预测研究区表层土壤水分空间分布的可行性。

2.5 土壤水分的空间分布特征

由图 7、8 可知,不同月份 ELM 和 RF 模型预测的土壤水分分布图与 PLSR 模型预测的土壤水分分布图比较相似。高土壤含水率出现在研究区的北部和东南部,主要由于这些地区有密集的植被覆盖,土

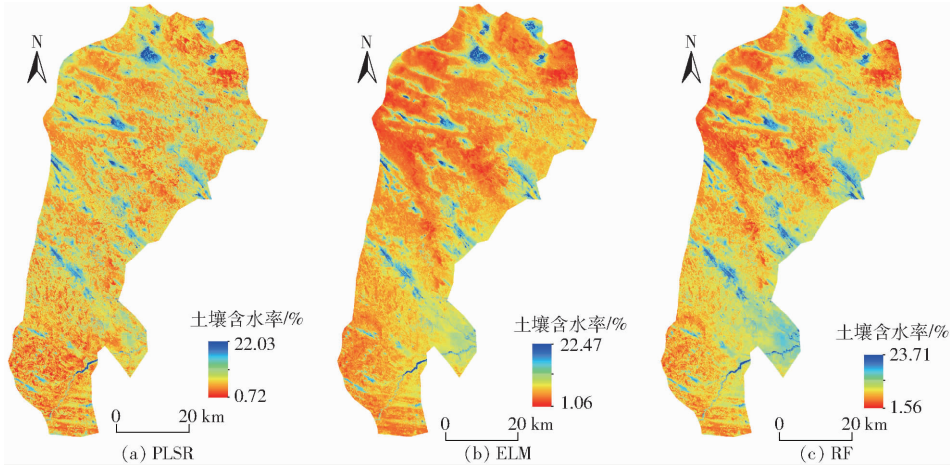


图 7 4 月表层土壤水分空间分布特征

Fig. 7 Spatial distribution characteristics of surface soil moisture in April

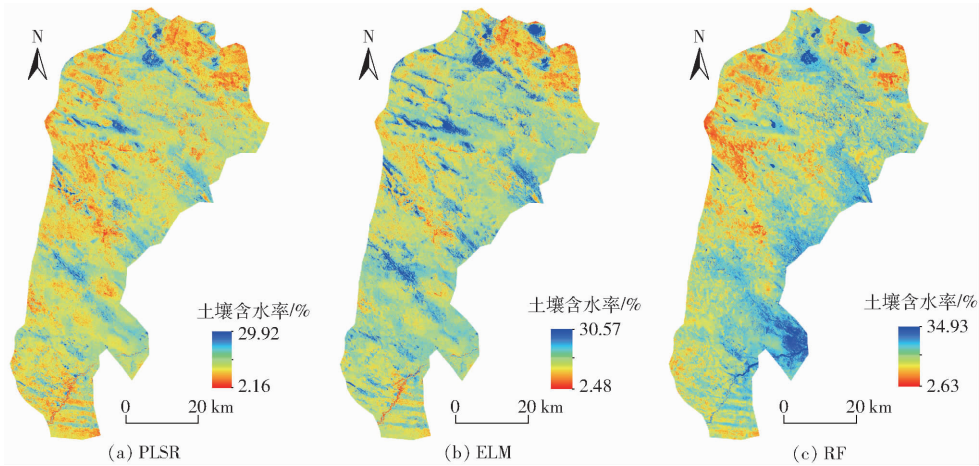


图 8 8 月表层土壤水分空间分布特征

Fig. 8 Spatial distribution characteristics of surface soil moisture in August

壤持水性好。中北部平坦地区的土壤含水率较低，主要由于这些地区植被覆盖度低，地表蒸散强烈。此外，ELM 反演的土壤水分分布（4 月，1.06% ~ 22.47%；8 月，2.48% ~ 30.57%）与 RF 模型（4 月，1.56% ~ 23.71%；8 月，2.63% ~ 34.93%）相似。而 PLSR 模型（4 月，0.72% ~ 22.03%；8 月，2.16% ~ 29.92%）的土壤水分分布图更强烈地表现出所有部分的低土壤含水率。一般情况下，PLSR 方法反演土壤水分的空间变异性低于 ELM 和 RF 方法，其主要原因是 ELM 和 RF 方法更能识别和反演土壤水分中的局部微小变化，能够更详细地表达地表信息和空间异质性。

3 讨论

土壤水分是了解地表过程、陆-气相互作用、干旱预测、作物生长模式等的基础，是一个动态变量，受多种因素的影响，如植被覆盖、地表粗糙度、土壤类型、地形等，在不同的时空尺度上，即使在较小的区域内也会发生显著的变化。本文在建模方法上选择了 PLSR、ELM 和 RF 构建研究区土壤含水率反演模型，经过对比分析发现，在相同条件下 ELM、RF 算法效果优于 PLSR 算法，因为土壤含水率与环境变量之间并非简单的线性关系，PLSR 模型在处理土壤含水率与环境之间复杂关系时具有一定的局限性^[27]，而 ELM、RF 算法在非线性问题中具有较强的解析力和较高的模型鲁棒性，这与葛翔宇等^[18]和蔡亮红等^[13]研究结果一致。但在 ELM、RF 土壤含水率反演模型中，本文建模集和验证集 R^2 均不小于 0.65，其中 RF 模型的精度最高，这与王浩等^[28]利用 RF 方法考虑温度、蒸散发、地形等因子对土壤水分影响的结果相一致。土壤水分不仅与这 3 个变量有关，还与坡度、波段

反射率、地表反照率和植被指数等其他变量有关^[29]。本研究与传统的经验模型相比，基于机器学习的检索算法避免了复杂的公式，提高了土壤水分的检索效率^[30]。但各环境变量对研究区土壤水分的影响不同，在红色、近红外、短波红外 1 和短波红外 2 波长范围内，土壤含水率和土壤光谱反射率之间存在不显著的线性关系，主要由于水在非饱和砂中的水力特性，土壤水分低时光谱反射率的降低非线性^[26]，本研究发现短波红外 1 和短波红外 2 波长与土壤水分之间线性相关。主要原因是短波红外波段裸土和植被土壤的反射率与表面粗糙度的相关性最好。但是光谱反射率会受到土壤固有因素的影响，例如有机质含量、粒径分布、矿物成分、表面粗糙度和土壤元素的颜色^[31]。温度植被干旱指数对土壤水分的影响最大，它可以将植被指数和地表温度结合起来，从水分胁迫开始到植被指数出现变化的时刻之间的时间延迟是最小的，从而避免了红波段和近红外波段光谱响应的延迟^[32]，这一结果与王思楠等^[33]利用不同干旱指数研究土壤水分的结果较为相似。地表温度对土壤水分的影响也很大，因为它直接影响能量平衡分量和地表蒸散发。随着地表温度的增大，土壤水分含量减小。BABAEIAN 等^[34]和 CARLSON 等^[35]也表明，影响土壤水分空间变异性的最重要参数是地表温度。因此，在基于遥感数据常用的 TVDI^[36]、条件温度植被指数（Vegetation temperature condition index, VTCI）^[37]、植被供水指数（Vegetation supply water index, VSWI）^[38]等反演土壤水分时，都使用了地表温度参数。因此，有研究表明，土壤水分也受到这些环境变量因素的直接和间接影响。在环境变量中，研究区 NDVI 和绿度的增加表明植被覆盖度增加，蒸腾作用增加，因此土壤水分增加。增加 NDBI 和亮度会增加地表反照率，

从而降低土壤水分。地形参数也是一个控制地表温度的重要影响因素,高程的增加导致气温下降,地表蒸散发减少,从而土壤水分增加。在该区域,表面生物物理性质对土壤水分的影响大于地形的影响。这些结果与其他一些研究结果相似^[30]。

土壤水分受不同环境因素的显著影响,不同模型在不同月份的反演结果相似。在毛乌素沙地腹部,4月降水量变化小,而且农作物处于生长初期,蒸发量小,土壤水分变化量不大;8月降水高度集中,沙壤土入渗率高,降水很快能渗入土壤,故很少能形成径流,土壤水分出现明显增加。土地利用在确定土壤水分变化的空间变异性方面非常重要,因为它影响植被覆盖、入渗和径流速率、蒸散过程、土壤表面特征。土地利用甚至可以消除地形相关参数对土壤含水率的影响。一些研究提供了通过反射图像和辅助地理空间数据估计土壤湿度的解决方案。乌审旗南部地区的高土壤水分可能是由于土壤表面的蒸发潜力相对较低,降水多,植被对降水的截留作用比较大。北部地区土壤含水率低可能是降水稀少,植被覆盖率低,基本上为沙地,水分下渗非常快。当沙丘被梭梭稳定后,粘土和淤泥的数量迅速增加,因为悬浮颗粒堆积和细颗粒是由沙子的风化机制产生的。粘土和淤泥含量的增加也与土壤含水率有轻微的正相关关系^[39]。沙地的表层土壤含水率低是因为由于高太阳辐射,土壤表面蒸发潜力高,植物的生长主要受土壤水分的控制,而土壤水分是光合作用

所必需的。此外,土壤湿度调节植物蒸腾和蒸发的速率,进而影响近地表温度、湿度和大气水蒸气^[40]。本研究由于其他土壤特性(例如土壤孔隙度、堆积密度和土壤有机质含量)很难获取,仅使用了17个环境变量,可能会导致反演结果的不确定性。除此之外,光学遥感受天气的影响较大,本研究选取的遥感影像获取时间范围内研究区处于晴空状态下,因此未考虑降雨条件下的不同模型的土壤水分估算表现。

4 结论

(1)通过对不同环境变量与表层土壤含水率进行相关分析,发现不同的环境变量与表层土壤含水率之间的相关程度不同。环境变量温度植被干旱指数和地表温度均与土壤含水率相关性较高。

(2)对比 PLSR、ELM 和 RF 3 种模型评价指标发现,4月,RF 模型的 R^2 为 0.74, RMSE 为 8.85%; 8月,RF 模型的 R^2 为 0.75, RMSE 为 8.66%,表明 RF 为土壤水分反演最优模型。

(3)不同月份 ELM 和 RF 模型预测的土壤水分分布图与 PLSR 模型预测的土壤水分分布图比较相似。高土壤含水率出现在研究区的北部和东南部,中北部平坦地区的土壤含水率较低。此外,ELM 与 RF 模型反演的土壤水分分布,突出了研究区地理位置的高低土壤含水率。而 PLSR 模型更强烈地表现出所有的低土壤含水率。

参 考 文 献

- [1] 王春梅,王鹏新,朱向明,等. 区域蒸散和表层土壤含水量遥感模拟及影响因子[J]. 农业工程学报, 2008, 24(10): 127-133.
WANG Chunmei, WANG Pengxin, ZHU Xiangming, et al. Estimations of evapotranspiration and surface soil moisture based on remote sensing data and influence factors[J]. Transactions of the CSAE, 2008, 24(10): 127-133. (in Chinese)
- [2] 黄友昕,刘修国,沈永林,等. 农业干旱遥感监测指标及其适应性评价方法研究进展[J]. 农业工程学报, 2015, 31(16): 186-195.
HUANG Youxin, LIU Xiuguo, SHEN Yonglin, et al. Advances in remote sensing derived agricultural drought monitoring indices and adaptability evaluation methods[J]. Transactions of the CSAE, 2015, 31(16): 186-195. (in Chinese)
- [3] 王雅婷,孔金玲,杨亮彦,等. 基于 SVR 的旱区稀疏植被覆盖下土壤水分遥感反演[J]. 地球信息科学学报, 2019, 21(8): 1275-1283.
WANG Yating, KONG Jinling, YANG Liangyan, et al. Remote sensing inversion of soil moisture in vegetation-sparse arid areas based on SVR[J]. Journal of Geo-information Science, 2019, 21(8): 1275-1283. (in Chinese)
- [4] 孔金玲,李菁菁,甄珮珮,等. 微波与光学遥感协同反演旱区地表土壤水分研究[J]. 地球信息科学学报, 2016, 18(6): 857-863.
KONG Jinling, LI Jingjing, ZHEN Peipei, et al. Inversion of soil moisture in arid area based on microwave and optical remote sensing data[J]. Journal of Geo-information Science, 2016, 18(6): 857-863. (in Chinese)
- [5] 何连,秦其明,任华忠,等. 利用多时相 Sentinel-1 SAR 数据反演农田地表土壤水分[J]. 农业工程学报, 2016, 32(3): 142-148.
HE Lian, QIN Qiming, REN Huazhong, et al. Soil moisture retrieval using multi-temporal Sentinel-1 SAR data in agricultural areas[J]. Transactions of the CSAE, 2016, 32(3): 142-148. (in Chinese)
- [6] 周鹏,丁建丽,王飞,等. 植被覆盖地表土壤水分遥感反演[J]. 遥感学报, 2010, 14(5): 959-973.
ZHOU Peng, DING Jianli, WANG Fei, et al. Retrieval methods of soil water content in vegetation covering areas based on multi-source remote sensing data[J]. National Remote Sensing Bulletin, 2010, 14(5): 959-973. (in Chinese)

- [7] 虞文丹, 张友静, 郑淑倩. 基于作物缺水指数的土壤含水量估算方法[J]. 国土资源遥感, 2015, 27(3): 77–83.
YU Wendan, ZHANG Youjing, ZHENG Shuqian. Estimation of soil moisture based on crop water stress index[J]. Remote Sensing for Land & Resources, 2015, 27(3): 77–83. (in Chinese)
- [8] 赵杰鹏, 张显峰, 廖春华, 等. 基于 TVDI 的大范围干旱区土壤水分遥感反演模型研究[J]. 遥感技术与应用, 2011, 26(6): 742–750.
ZHAO Jiepeng, ZHANG Xianfeng, LIAO Chunhua, et al. TVDI based soil moisture retrieval from remotely sensed data over large arid areas[J]. Remote Sensing Technology and Application, 2011, 26(6): 742–750. (in Chinese)
- [9] 杨永民, 邱建秀, 苏红波, 等. 基于热红外的四种土壤含水量估算方法对比[J]. 红外与毫米波学报, 2018, 37(4): 459–467.
YANG Yongmin, QIU Jianxiu, SU Hongbo, et al. Estimation of surface soil moisture based on thermal remote sensing: intercomparison of four methods[J]. Journal of Infrared and Millimeter Waves, 2018, 37(4): 459–467. (in Chinese)
- [10] 蔡庆空, 李二俊, 陶亮亮, 等. 基于改进粒子群神经网络的麦田土壤水分反演[J]. 土壤通报, 2018, 49(6): 1333–1340.
CAI Qingkong, LI Erjun, TAO Liangliang, et al. soil moisture retrieval in wheat fields based on modified particle swarm optimization and neural network[J]. Chinese Journal of Soil Science, 2018, 49(6): 1333–1340. (in Chinese)
- [11] 谭建灿, 毛克彪, 左志远, 等. 基于卷积神经网络和 AMSR2 微波遥感的土壤水分反演研究[J]. 高技术通讯, 2018, 28(5): 399–408.
TAN Jiancan, MAO Kebiao, ZUO Zhiyuan, et al. Soil moisture retrieval based on convolutional neural network and AMSR2 microwave remote sensing[J]. Chinese High Technology Letters, 2018, 28(5): 399–408. (in Chinese)
- [12] 薛晓萍, 王新, 张丽娟, 等. 基于支持向量机方法建立土壤湿度预测模型的探讨[J]. 土壤通报, 2007, 38(3): 427–433.
XUE Xiaoping, WANG Xin, ZHANG Lijuan, et al. Prediction model of soil moisture based on support vector machines[J]. Chinese Journal of Soil Science, 2007, 38(3): 427–433. (in Chinese)
- [13] 蔡亮红, 丁建丽. 基于变量优选和 ELM 算法的土壤含水量预测研究[J]. 光谱学与光谱分析, 2018, 38(7): 2209–2214.
CAI Lianghong, DING Jianli. Prediction for soil water content based on variable preferred and extreme learning machine algorithm[J]. Spectroscopy and Spectral Analysis, 2018, 38(7): 2209–2214. (in Chinese)
- [14] 李平湘, 刘致曲, 杨杰, 等. 利用随机森林回归进行极化 SAR 土壤水分反演[J]. 武汉大学学报(信息科学版), 2019, 44(3): 405–412.
LI Pingxiang, LIU Zhiyu, YANG Jie, et al. Soil moisture retrieval of winter wheat fields based on random forest regression using quad-polarimetric SAR images[J]. Geomatics and Information Science of Wuhan University, 2019, 44(3): 405–412. (in Chinese)
- [15] 贾艳昌, 谢谟文, 姜红涛. 全球 36 km 格网土壤水分逐日估算[J]. 地球信息科学学报, 2017, 19(6): 854–860.
JIA Yanchang, XIE Mowen, JIANG Hongtao. Daily estimate of global 36 km grid soil moisture[J]. Journal of Geo-information Science, 2017, 19(6): 854–860. (in Chinese)
- [16] 杨婷, 陈秀万, 万玮, 等. 基于光学与被动微波遥感的青藏高原地区土壤水分反演[J]. 地球物理学报, 2017, 60(7): 2556–2567.
YANG Ting, CHEN Xiuwan, WAN Wei, et al. Soil moisture retrieval in the Tibetan plateau using optical and passive microwave remote sensing data[J]. Chinese Journal of Geophysics, 2017, 60(7): 2556–2567. (in Chinese)
- [17] 程渊, 李玉霞, 李凡, 等. 基于极端随机树的闪电河流域土壤水分反演[J]. 遥感学报, 2021, 25(4): 941–951.
CHENG Yuan, LI Yuxia, LI Fan, et al. Soil moisture retrieval using extremely randomized trees over the Shandian river basin [J]. National Remote Sensing Bulletin, 2021, 25(4): 941–951. (in Chinese)
- [18] 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)
- [19] 谭丞轩, 张智韬, 许崇豪, 等. 无人机多光谱遥感反演各生育期玉米根域土壤含水率[J]. 农业工程学报, 2020, 36(10): 63–74.
TAN Chengxuan, ZHANG Zhitao, XU Chonghao, et al. Soil water content inversion model in field maize root zone based on UAV multispectral remote sensing[J]. Transactions of the CSAE, 2020, 36(10): 63–74. (in Chinese)
- [20] 张东兴, 刘江, 杨丽, 等. 基于 VIS–NIR 的播种沟内土壤水分测量传感器研究[J]. 农业机械学报, 2021, 52(2): 218–226.
ZHANG Dongxing, LIU Jiang, YANG Li, et al. Soil moisture measurement sensor research in seeding ditch based on VIS–NIR[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(2): 218–226. (in Chinese)
- [21] 周鹏, 杨玮, 李民赞, 等. 基于灰度关联-极限学习机的土壤全氮预测[J]. 农业机械学报, 2017, 48(增刊): 271–276.
ZHOU Peng, YANG Wei, LI Minzan, et al. Soil total nitrogen content prediction based on gray correlation-extreme learning machine[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(Supp.): 271–276. (in Chinese)
- [22] 包青岭, 丁建丽, 王敬哲. 利用随机森林方法优选光谱特征预测土壤水分含量[J]. 激光与光电子学进展, 2018, 55(11): 470–476.
BAO Qingling, DING Jianli, WANG Jingzhe. Prediction of soil moisture content by selecting spectral characteristics using random forest method[J]. Laser & Optoelectronics Progress, 2018, 55(11): 470–476. (in Chinese)
- [23] ZARE M, DRASTIG K, ZUDE-SASSE M. Tree water status in apple orchards measured by means of land surface temperature

- and vegetation index (LST – NDVI) trapezoidal space derived from Landsat 8 satellite images[J]. Sustainability, 2020, 12(1):1 – 19.
- [24] LIU Y, YUE H. The temperature vegetation dryness index (TVDI) based on bi-parabolic NDVI – TS space and gradient-based structural similarity (GSSIM) for long-term drought assessment across Shaanxi Province, China (2000—2016) [J]. Remote Sensing, 2018, 10(6):959.
- [25] CHEN S, SHE D, ZHANG L, et al. Spatial downscaling methods of soil moisture based on multisource remote sensing data and its application[J]. Water, 2019, 11(7):1401.
- [26] NGO THI D, HA N T T, TRAN DANG Q, et al. Effective band ratio of Landsat 8 images based on vnir-swir reflectance spectra of topsoils for soil moisture mapping in a tropical region[J]. Remote Sensing, 2019, 11(6):716.
- [27] 陈文倩, 丁建丽, 谭娇, 等. 干旱区绿洲植被高光谱与浅层土壤含水率拟合研究[J]. 农业机械学报, 2017, 48(12): 229 – 236.
- CHEN Wenqian, DING Jianli, TAN Jiao, et al. Fitting of hyperspectral reflectance of vegetation and shallow soil water content in oasis of arid area[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(12): 229 – 236. (in Chinese)
- [28] 王浩, 罗格平, 王伟胜, 等. 基于多源遥感数据的锡尔河中下游农田土壤水分反演[J]. 自然资源学报, 2019, 34(12): 2717 – 2731.
- WANG Hao, LUO Geping, WANG Weisheng, et al. Inversion of soil moisture content in the farmland in middle and lower reaches of Syr Darya River Basin based on multi-source remotely sensed data[J]. Journal of Natural Resources, 2019, 34(12): 2717 – 2731. (in Chinese)
- [29] FATHOLOLOUMI S, VAEZI A R, ALAVIPANAH S K, et al. Effect of multi-temporal satellite images on soil moisture prediction using a digital soil mapping approach[J]. Geoderma, 2021, 385(6):114901.
- [30] FATHOLOLOUMI S, VAEZI A R, ALAVIPANAH S K, et al. Comparison of spectral and spatial-based approaches for mapping the local variation of soil moisture in a semi-arid mountainous area[J]. Sci. Total Environ., 2020, 724:138319.
- [31] ADAB H, MORBIDELLI R, SALTALIPPI C, et al. Machine learning to estimate surface soil moisture from remote sensing data[J]. Water, 2020, 12(11):3223.
- [32] DU L, SONG N, LIU K, et al. Comparison of two simulation methods of the temperature vegetation dryness index (TVDI) for drought monitoring in semi-arid regions of China[J]. Remote Sensing, 2017, 9(2):177.
- [33] 王思楠, 李瑞平, 李夏子. 基于综合干旱指数的毛乌素沙地腹部土壤水分反演及分布[J]. 农业工程学报, 2019, 35(13): 113 – 121.
- WANG Sinan, LI Ruiping, LI Xiazi. Inversion and distribution of soil moisture in belly of Maowusu sandy land based on comprehensive drought index[J]. Transactions of the CSAE, 2019, 35(13): 113 – 121. (in Chinese)
- [34] BABAEIAN E, SADEGHI M, FRANZ T E, et al. Mapping soil moisture with the optical trapezoid model (OPTRAM) based on long-term MODIS observations[J]. Remote Sensing of Environment, 2018, 211(4):425 – 440.
- [35] CARLSON T N, GILLIES R R, PERRY E M. A method to make use of thermal infrared temperature and NDVI measurements to infer surface soil water content and fractional vegetation cover[J]. Remote Sensing Reviews, 1994, 9(1 – 2): 161 – 173.
- [36] 蔡庆空, 陶亮亮, 蒋瑞波, 等. 基于理论干湿边与改进 TVDI 的麦田土壤水分估算研究[J]. 农业机械学报, 2020, 51(7): 202 – 209.
- CAI Qingkong, TAO Liangliang, JIANG Ruibo, et al. Soil moisture estimation of wheat field based on theoretical dry-wet edge and improved TVDI[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(7): 202 – 209. (in Chinese)
- [37] 王鹏新, 吴高峰, 白雪娇, 等. 基于 Landsat 数据的条件植被温度指数升尺度转换方法[J]. 农业机械学报, 2015, 46(7): 264 – 271.
- WANG Pengxin, WU Gaofeng, BAI Xuejiao, et al. Up-scaling transformation methods for vegetation temperature condition index retrieved from Landsat data[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(7): 264 – 271. (in Chinese)
- [38] 曹广真, 侯鹏, 范锦龙, 等. TM 与 MODIS 植被供水指数反演及对比分析[J]. 遥感技术与应用, 2010, 25(1): 63 – 68.
- CAO Guangzhen, HOU Peng, FAN Jinlong, et al. VSWI retrieved and compared between TM and MODIS[J]. Remote Sensing Technology and Application, 2010, 25(1): 63 – 68. (in Chinese)
- [39] NASERI S, ADIBI M A, JAVADI S, et al. Investigation of the effect of biological stabilization practice on some soil parameters (North East of Iran)[J]. Journal of Rangeland Science, 2012, 2(4):643 – 653.
- [40] FAN B, ZHANG A, YANG Y, et al. Long-term effects of xerophytic shrub haloxylon ammo dendron plantations on soil properties and vegetation dynamics in Northwest China[J]. PLoS One, 2016, 11(12): e0168000.