

基于多源遥感数据的锡尔河中下游农田土壤水分反演

王浩^{1,2}, 罗格平^{1,2,3}, 王伟胜¹, PACHIKIN Konstantin⁴,
李耀明¹, 郑宏伟^{1,2}, 胡伟杰¹

(1. 中国科学院新疆生态与地理研究所荒漠与绿洲国家重点实验室, 乌鲁木齐 830011; 2. 中国科学院大学, 北京 100049; 3. 中国科学院中亚生态与环境研究中心, 乌鲁木齐 830011; 4. 哈萨克斯坦土壤科学与农业化学研究所, 哈萨克斯坦 阿拉木图 050060)

摘要: 机器学习结合多源遥感数据反演土壤水分含量(SMC)是目前SMC研究的热点, 因较少考虑温度、蒸散等重要SMC影响因子, 反演结果存在一定的不确定性。利用Sentinel-1影像、MODIS产品和SRTM数据, 提取雷达后向散射系数等32个SMC影响因子, 经相关分析选择27个显著的SMC影响因子($P < 0.05$)作为反演因子, 并设计三组因子组合。这三组因子组合分别与随机森林、支持向量回归、BP神经网络三种机器学习方法结合, 发现基于随机森林结合所有因子的方案, 其SMC反演精度最高, 该组合均方根误差RMSE为 $0.039 \text{ m}^3/\text{m}^3$, 将该方案被用于反演2017年生长季锡尔河流域中下游平原区农田SMC。结果表明: 从上部至下部SMC总体呈逐渐增加的态势, 但存在显著时空差异, 春季和秋季SMC较高而夏季较低。SMC差异主要由土壤质地、热量条件和地表植被状况差异引起。春季平原区下部农田SMC要高于上部, SMC的主控因子是土壤质地和地表植被状况; 在夏季, 土壤水分的主控因子是热量条件, 农田灌溉弥补了热量条件差异对土壤水分的影响, 导致空间上平原上部和下部土壤SMC空间差异不显著; 秋季SMC的主控因子植被状况抵消地表温度和土壤质地差异对SMC的影响, 使得秋季SMC空间差异不显著。本文采用的研究方法在一定程度上克服了因考虑SMC影响因子不足而获取更高SMC精度的限制。

关键词: 土壤水分含量; 机器学习; 锡尔河流域中下游; Sentinel-1; MODIS; SRTM

土壤水分含量 (Soil Moisture Content, SMC) 影响地—气能量交换, 是水文模型、气候模型、生态模型和陆面过程等模型的关键变量, 在全球与区域气候变化、能量平衡及水循环的研究中担当重要角色^[1]。SMC直接影响作物生长发育, 是长势监测、产量预报、土壤墒情等农业应用的重要因子。

遥感是大面积精确估计SMC的关键手段。基于遥感的SMC反演基本可分为两大类, 即基于可见光近红外数据的反演和基于微波雷达数据的反演。可见光近红外数据拥有较高空间分辨率且可提供较长的时间序列产品, 其热惯量法适合裸土和低植被覆盖区域, 难以运用于大范围、植被茂密的农田区域^[2]; 而基于指数或特征空间法在特定研究区域反演结果精度较高, 但存在部分参数计算复杂, 泛化性较差等问题^[3]。微波遥感具有全天候模拟能力, 被动微波土壤水分产品如AMSR-E^[4]、SMOS^[5]、AMSR2^[6]、SMAP^[7]支持区

收稿日期: 2019-05-25; 修订日期: 2019-09-09

基金项目: 国家自然科学基金项目 (41877012); 中国科学院特色研究所项目 (TSS-2015-014-FW-1-3)

作者简介: 王浩 (1992-), 男, 贵州贵阳人, 硕士, 主要从事遥感与地理信息系统研究。

E-mail: wangh_xjb@foxmail.com

通讯作者: 罗格平 (1968-), 男, 湖南湘乡人, 博士, 研究员, 主要从事土地利用/覆被变化及其生态与气候效应、遥感和GIS应用研究。E-mail: luogp@ms.xjb.ac.cn

域、全球尺度土壤水分研究,可较好表征大尺度土壤水分的时空特征,但由于空间分辨率太粗,基本不适用于中小尺度的农田SMC监测。基于主动微波雷达(synthetic aperture radar, SAR)的经验与半经验法^[8-9]、物理模型法^[10]在SMC反演中表现出较好适用性,但精度易受到地表粗糙度的影响。诸多学者结合可见光近红外数据与SAR数据的优点,开展基于多源遥感数据的SMC反演取得较好效果^[11-13],但经验关系限制其推广。

利用多源遥感数据,首先寻找SMC与其影响因子的关系;在此基础上,将影响因子与机器学习方法相结合构造SMC反演模型,提取SMC时空信息,具有显著优势,并已成为SMC研究的重要方向之一。在机器学习应用于SMC反演中,BP神经网络(Back Propagation Neural Network, BPNN)^[14-16]具有较好的非线性映射能力和网络柔性,但易陷入局部最小问题,易产生过拟合;支持向量回归(Support Vector Regression, SVR)^[17-19]可较处理好变量高维特征,但是需要寻找合适的核函数。以上两者是基于机器学习的SMC反演中较为常见的方法,显示出良好的精度、运算速度和稳健性。随机森林(Random Forest, RF)泛化能力强,不易过拟合,但对噪音较大的样本比较敏感,RF在SMC反演中应用较少,而在其他生物物理参量反演研究中得到广泛应用^[20]。基于BPNN、SVR的SMC反演因子,通常包括主动微波雷达数据(如RADARSAT-2/Envisat/Sentinel-1)的后向散射系数(如VV、VH)^[16]及其组合特征和基于可见光近红外数据(如MODIS、Landsat)的植被因子如NDVI^[21]以及这两类数据的组合^[22-24],较少考虑温度、蒸散发、地形等因子对SMC的影响,这在一定程度上影响SMC反演精度,增加其反演结果的不确定性。

针对上述机器学习结合遥感的SMC反演研究存在的不足,本研究首先分析干旱区SMC与其主要影响变量的相关性,探讨BPNN、SVR和RF三种机器学习方法在干旱区地表SMC反演适用性;在此基础上,选择一种最适宜的机器学习方法,基于谷歌地球引擎(Google Earth Engine, GEE)平台,反演绿洲灌溉农田生长季SMC时空变化。本文主要研究目标为:(1)基于GEE平台多源遥感数据,提取SMC影响因子,基于相关分析筛选显著的影响因子作为SMC的反演因子;(2)选择中亚锡尔河流域中下游绿洲区,探讨ANN、SVR和RF三种主要机器学习方法反演SMC的适用性,寻找最优机器学习方法和SMC影响变量反演组合方案;(3)反演2017年锡尔河中下游绿洲灌溉农田生长季SMC时空变化,分析其时空动态特征。

1 研究方法与数据来源

1.1 研究区概况

锡尔河流域的中下游位于哈萨克斯坦共和国的南哈萨克斯坦州(简称南哈州)和克孜勒奥尔达州(简称克州)内(61°14'16"~71°12'9"E, 40°29'50"~46°13'59"N),地处图兰低地。锡尔河南距克孜勒库姆沙漠20多km,西临咸海,东接卡拉套山、塔拉斯山和恰特卡尔山(图1)。年均温6~12℃,年降雨量<200mm,年潜在蒸散发约2000mm。该区域气候干燥,降水稀少、潜在蒸散强度大,属于典型的大陆性干旱气候。按海拔高度和分布位置,锡尔河流域中下游地貌类型从东向西分为:低山、山前倾斜平原和中部平原,整体地势除东部山地外较为平缓。锡尔河流域土壤质地在中游段以砂质黏壤土、砂质黏土为主,土壤质地较轻,而下游段则主要为壤质黏土、黏壤土,土壤质地较粘重

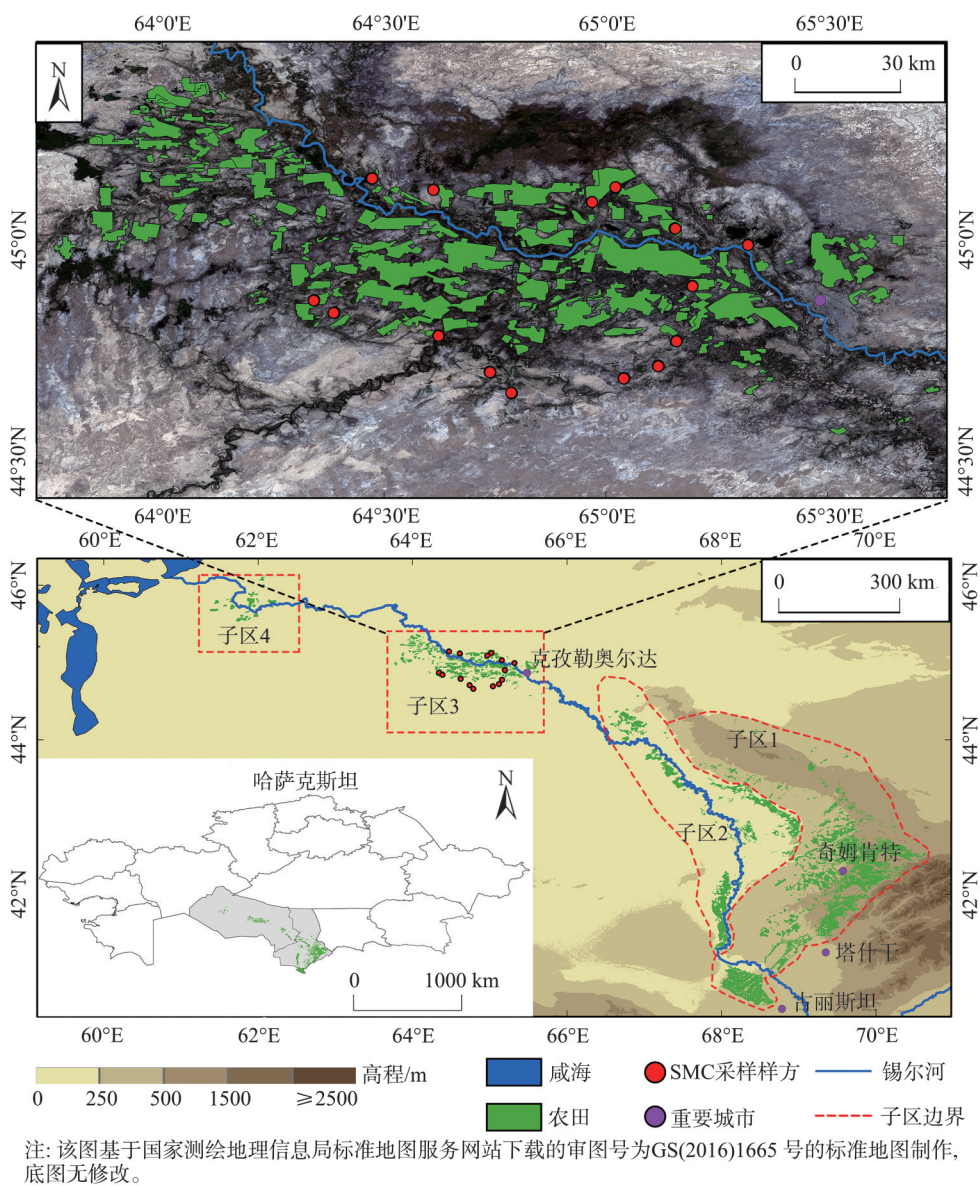


图1 研究区与采样点分布示意

Fig. 1 Location of the sampling points in the middle and lower reaches of Syr Darya River Basin

(图2)。本文研究对象限定为锡尔河流域中下游农田区域SMC。周边荒漠地表覆被以灌木、半灌木及荒漠草地为主。

本文中依据农田的地理位置或地势, 将锡尔河中下游的农田划分为4个研究子区(图1)。在上部的锡尔河流域中游段, 农田分布密集, 不易按照地理位置划分, 但在后续分析中通过观察发现250 m等高线, 可较明显地界定不同时间SMC的高值区和低值区, 所以采用250 m等高线作为分界线将中游段东、西两侧的农田区域分别划分为研究子区1(子区1)和研究子区2(子区2)。对于中下游、下游段, 依据地理位置将分布较为集中的两块农田区域划分为另外2个子区, 中下游克孜勒奥尔达等地区划分为研究子

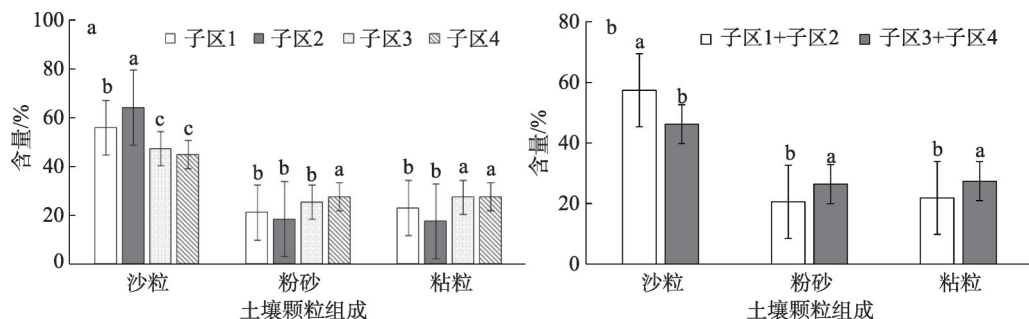


图2 锡尔河流域中下游土壤质地 (0~10 cm) 均值与标准差空间特征

Fig. 2 The spatial characteristics of means and standard deviations of soil texture in the middle and lower reaches of Syr Darya River Basin (0-10 cm)

区3 (子区3), 下游的拜科努尔等地区划分为研究子区4 (子区4)。

1.2 数据来源

1.2.1 Google Earth Engine 平台

谷歌地球引擎 (Google Earth Engine, GEE) 是目前较为成熟的地理数据科学分析云平台^[25]。GEE 拥有大量遥感数据存档, 提供高性能并行计算服务, 支持 JavaScript 和 Python 语言的应用程序编程接口 (API)、算法与分析工具, 基于网络的交互开发环境实现大型数据的处理分析与信息挖掘。本文基于 GEE 平台获取实验所需的遥感数据, 并主要利用该平台所集成的机器学习方法模块完成 SMC 反演。

1.2.2 数据

主要实验数据包括遥感数据 Sentinel-1 SAR 影像、MODIS 产品数据和 SRTM 高程数据以及 SMC 野外实测数据 (表1)。

基于 GEE 平台获取 2017 年生长季 (4-10 月) SMC 时空反演所需要的 MODIS 地表生物物理参量产品数据、Sentinel-1 SAR 影像数据以及 SRTM 高程数据。

GEE 平台中的历史存档遥感产品数据, 均经过了预处理。对于 MODIS 产品数据的预处理, 包括辐射校正、几何校正和大气校正等。DEM 数据为航天飞机雷达地形测绘使命 (Shuttle Radar Topography Mission, SRTM) V3 版本。Sentinel-1 SAR 数据源自欧空局 (European Space Agency, ESA), 本研究采用的 SAR 数据为 C 波段的干涉宽幅 (IW, interferometric wide swath) 地距多视产品 (GRD, ground range detected), 极化模式为垂直极化。GEE 平台中 Sentinel-1 数据预处理步骤包括轨道信息修正、利用噪声向量进行热噪声去除、辐射校准、基于 SRTM 数据的正射校正和将后向散射系数单位转换为分贝, 并利用水云模型去除植被影响^[26]。

本文中遥感数据处理步骤为: 在 GEE 平台的 JavaScript API 界面中将实验所需遥感数据 (表1) 导入。以 2017 年 4 月 1 日和 11 月 1 日分别为起始、终止时间, 按照 16 d 的时间分辨率分别构造 SMC 影响因子数据集 (表2), 每项因子每月产生 2 期, 共产生 14 期因子数据集, 用于反演 SMC 的时间变化。研究区域内 6 月初的 MOD13Q1 植被指数产品存在质量问题无法利用, 本研究使中用同系列的 MYD13Q1 替代。为与实测数据匹配以便后续处理, 使用三次卷积内插法将上述遥感数据重采样至 90 m。基于克州、南哈州的 2017 年 9 月的 Landsat 8 影像, 通过目视解译提取农田分布的矢量数据, 利用该矢量数据对后续的 SMC 反演结果进行裁剪。

表1 锡尔河中下游农田2017年生长季SMC研究所需数据

Table 1 Data of SMC inversion and analysis of farmland in the middle and lower reaches of Syr Darya River Basin during the growing season in 2017

数据	空间分辨率/m 时间分辨率/d		来源	用途	时间
Sentinel-1 SAR	5×20	12	ESA(https://scihub.com/earthdata)	SMC 影响因子 的时空数据	2017年4-10月
MOD09GQ 地表反射率	250	1	NASA(https://lpdaac.usgs.gov/)		2017年4-10月
MOD11A2 v4 地表温度	1000	8			
MOD13Q1v6 植被指数	250	16			
MYD13Q1.v6 植被指数	250	16			
MCD15A3H 叶面积指数	500	4			
MOD16A2 v6 蒸散发	500	8			
MCD43A3 地表反照率	500	1			
SRTM v4 高程数据	90		CRIAR(http://srtm.csi.cgiar.org)		
SMC 实测数据			中国科学院 新疆生地所	模型训练与反演 结果验证	2017年9月18-20日

表2 锡尔河流域中下游SMC影响因子

Table 2 SMC factors in the middle and lower reaches of Syr Darya River Basin

SMC影响因子类型	名称	影响因子
微波物理量	后向散射系数 (backscattering coefficient, <i>BC</i>)	<i>BC</i> 均值/最大值/最小值
植被变量	归一化植被指数 (Normalized differential vegetation index, <i>NDVI</i>)	<i>NDVI</i> 均值/最大值/最小值
	增强型植被指数 (Enhanced Vegetation Index, <i>EVI</i>)	<i>EVI</i> 均值/最大值/最小值
	土壤调整植被指 (Soil Adjusted Vegetation Index, <i>SAVI</i>)	<i>SAVI</i> 均值/最大值/最小值
	叶面积指数 (leaf Area Index, <i>LAI</i>)	<i>LAI</i> 均值/最大值/最小值
	地表温度 (Surface temperature, <i>LST</i>)	<i>LST</i> 均值/最大值/最小值
蒸散发变量	蒸散发 (Evapotranspiration, <i>ET</i>)	<i>ET</i> 均值/最大值/最小值/累积值
下垫面反射特性变量	可见光范围地表反照率 (Albedo in VIS, <i>AIV</i>)	<i>AIV</i> 均值/最大值/最小值
	近红外范围地表反照率 (Albedo in NIR, <i>AIN</i>)	<i>AIN</i> 均值/最大值/最小值
地形变量	高程 (<i>Elevation</i>)	高程/坡度/坡向/地面粗糙度
	坡度 (<i>Slope</i>)	
	坡向 (<i>Aspect</i>)	
	地面粗糙度 (<i>Roughness</i>)	

2017年9月17-20日在锡尔河下游的克孜勒奥尔达市周边进行地表0~10 cm SMC 实测数据采集,用于SMC反演模型的构建与验证。在Google Earth遥感影像上布置采样样方、采样点,沿主要公路每隔5~20 km布置一个采样样方,每个样方按照“梅花形布点法”在边长为150 m的矩形范围内布置3~5个采样点。在野外采用Hydra Probe II型土壤传感器对每个采样点测量5次,取5次测量值的均值作为该采样点土壤水分含量SMC (m³/m³);同时,采用便携式GPS接收机对每个采样点进行定位。3天内布设15个样方共采集61个采样点,比较均匀地分布在100 km×80 km范围内,其中40个采样点位于农田内。

1.3 研究方法

首先,在GEE平台中,按照前文所述的遥感数据处理步骤对Sentinel-1 SAR影像、

MODIS产品、SRTM高程数据进行处理,并获取SMC影响因子。将SMC实测数据与9月第2期的影响因子数据进行相关分析,选择与SMC显著相关的影响因子($P < 0.05$)作为SMC反演因子。

然后,基于9月的SMC反演因子,设计三组试验组合方案,每组实验方案中的反演因子与SMC实测数据作为输入数据,分别带入随机森林(Random Forest, RF)^[27]、支持向量回归(Support Vector Regression, SVR)^[28]、BP神经网络(Back Propagation Neural Network, BPNN)三种机器学习方法,构建不同的SMC机器学习反演模型,并进行精度评价。反演模型构建时,75%的实测SMC数据用于训练、其余25%用于验证。本研究选用3种指标来评价SMC反演组合的效果。判定系数(R^2)、均方根误差(RMSE)和平均相对误差绝对值(MAPE)的计算公式如下:

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

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

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (3)$$

式中: y_i 为实测值; \hat{y}_i 为模型预估值; \bar{y}_i 为是实测值均值; n 为样本数。

在此基础上,选择最佳机器学习方法和试验组合方案,反演2017年生长季锡尔河下游农田SMC时空分布。

采用最小显著性差异法^[29]比较SMC、地表温度LST、植被指数NDVI和土壤质地的时空差异。

RF基于Bagging抽样,从原样本集P中抽样形成 k 个训练样本集,每个训练样本集构建一棵决策树(Classification and Regression Trees, CART)^[30],每棵决策树的节点内部特征mtry从原样本集的 M 个特征中选取,按照节点不纯度最小原则从中选择一个最优特征对节点进行分裂生长。未知样本的回归结果为全部 k 棵决策树回归结果的均值。RF泛化性较强,性能稳定。在GEE平台中使用相应模块与函数完成RF模型构建。

SVM在非线性及高维模式识别问题中具有优势,基本原理是将变量映射到高维空间中,利用核函数将样本的非线性关系变换为线性可分,并选择最优线性分离器以分离不同的类。SVR在SVM的基本原理上加入 ϵ 不敏感函数以拟合实数结果^[31]。在MATLAB语言libsvm工具箱中采用网格搜索法确定SVR参数,后使用GEE平台相应模块与函数,完成SVR的模型构建。

BPNN^[32],具有较好的非线性映射能力与网络柔性。BPNN的结构包括输入层,隐含层和输出层。生成网络时首先依据节点权值、阈值与激活函数正向逐层求值;之后计算样本与期望的相对误差,沿相对误差平方和最快下降方向,反向调整网络的权值和阈值,直到误差符合预设精度。利用MATLAB语言Neural Network工具箱完成BPNN的构建。

2 结果分析

2.1 SMC的显著影响因子

本文所考虑的SMC影响因素,包括微波遥感物理量、植被、温度、蒸散发、地形、

下垫面反射条件6大类因子共32个SMC影响因子(表2)。除地形变量为固定值外,整个生长季中,其余5类影响因子与SMC在自然环境与人类活动干预下均发生动态变化。不仅是变量的均值,而且其极值(最大值、最小值)对SMC同样存在不可忽略的影响,因此处理每个变量所对应的MODIS产品数据或Sentinel-1数据时,统一按照16 d的时间分辨率分别合成均值(mean)、最大值(max)、最小值(min)3项数据,并作为SMC影响因子。需要特别说明的是,蒸散发(ET)是改变SMC的直接驱动力,在整个生长季中,除考虑其均值和极值外,其生长季的累积值(accumulated ET)同样需要考虑。

相关分析结果中(表3),下垫面反射特性变量(Albedo)与SMC的相关性最优,可见光地表反照率(AIV)与近红外地表反照率(AIN)均值、最大值、最小值的 r^2 值域范围为0.67~0.76,极显著($P<0.01$)。植被变量NDVI、EVI、SAVI与SMC相关性较好,三者的均值、最大值、最小值的 r^2 值域为0.35~0.62($P<0.05$);LAI的极小值相关性很好($r^2=0.74$, $P<0.01$)。后向散射系数(BC)的3项数据相关性良好, r^2 值域为0.37~0.44($P<0.05$)。地形变量中高程Elevation与坡向Aspect相关性较好, r^2 分别为-0.68和0.56($P<0.05$)。

表3 锡尔河流域中下游SMC与影响因子相关性分析
Table 3 Correlation analysis of SMC and SMC factors in the middle and lower reaches of Syr Darya River Basin

植被变量			温度或水分变量和微波物理量				地形变量和下垫面反射特性变量			
			r^2		P		r^2		P	
NDVI平均值	0.615	0.006	LST平均值	-0.385	0.034	AIV平均值	0.673	0.002		
NDVI最大值	0.584	0.009	LST最大值	-0.757	0.000	AIV最大值	0.723	0.001		
NDVI最小值	0.584	0.009	LST最小值	-0.492	0.026	AIV最小值	0.701	0.001		
SAVI平均值	0.358	0.042	ET平均值	0.100	0.356	AIN平均值	0.698	0.001		
SAVI最大值	0.354	0.047	ET累积值	-0.328	0.050	AIN最大值	0.726	0.001		
SAVI最小值	0.373	0.039	ET最小值	-0.334	0.048	AIN最小值	0.752	0.000		
LAI平均值	0.183	0.197	ET最大值	0.394	0.031	Elevation 高程	-0.682	0.002		
LAI最大值	0.165	0.270	BS平均值	0.437	0.031	Aspect 坡向	0.556	0.013		
LAI最小值	0.737	0.001	BS最大值	0.379	0.037	Slope 坡度	-0.238	0.188		
EVI平均值	0.380	0.032	BS最小值	0.366	0.040	Rough 地面粗糙度	-0.242	0.184		
EVI最大值	0.526	0.018								
EVI最小值	0.526	0.018								

未达到显著性水平的5个影响因子LAI平均值、LAI最大值、ET平均值、Slope坡度、Rough地面粗糙度被剔除。

2.2 最优SMC反演方案

基于SMC反演因子设计三组试验组合方案(表4):第一组SMC反演因子参考Pasolli等^[17]的研究,仅考虑单一垂直极化VV雷达后向散射系数;第二组结合Paloscia等^[14]的成果,将垂直极化VV后向散射系数与NDVI结合作为SMC反演因子;第三组试验方案是考虑所有的SMC反演因子。基于将上述三组试验方案,分别采用上述RF、SVM、BPNN3种机器学习方法,构建SMC反演模型。基于训练(46个采样点)与验证阶段(15个采样点)精度统计表明随机森林(RF)结合所有反演因子构建的SMC反演模型精度最高(图3),

表4 不同反演组合的SMC反演精度统计

Table 4 Quantitative statistics of the SMC retrieval with different combinations of parameters and models							
机器学习模型	试验组合方案	基于训练样本的精度统计			基于验证样本的精度统计		
		R^2	$RMSE/(m^3/m^3)$	$MAPE/\%$	R^2	$RMSE/(m^3/m^3)$	$MAPE/\%$
RF	VV	0.69	0.041	11.2	0.52	0.065	17.5
	NDVI+VV	0.78	0.038	10.6	0.59	0.054	13.5
	所有因子	0.80	0.035	10.3	0.68	0.039	10.5
SVR	VV	0.71	0.042	11.5	0.53	0.056	13.8
	NDVI+VV	0.75	0.037	11.3	0.55	0.059	13.7
	所有因子	0.78	0.037	10.7	0.64	0.043	10.9
BPNN	VV	0.69	0.042	11.4	0.49	0.086	18.4
	NDVI+VV	0.73	0.039	10.9	0.54	0.059	14.1
	所有因子	0.74	0.038	11.0	0.59	0.052	11.9

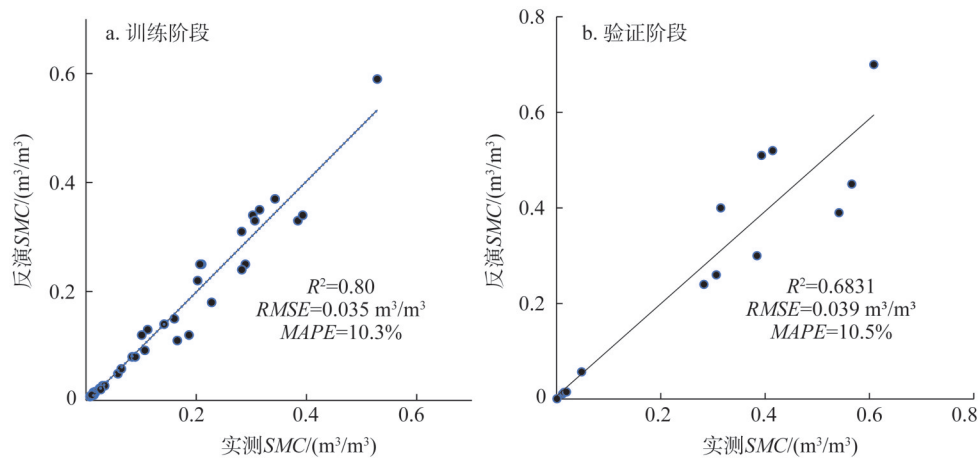


图3 实测SMC和随机森林RF结合多因子的反演SMC散点图

Fig. 3 Scatter plots of observed SMC and retrieved SMC based on The Random forest with the group with all input parameters

因此该方案将被用于反演研究区2017年生长季的农田SMC时空分布。

2.3 锡尔河中下游农田水分的时空特征

锡尔河中下游区域、子区1、子区2、子区3和子区4的2017年生长季SMC分别为0.452±0.093 m³/m³、0.448±0.094 m³/m³、0.450±0.088 m³/m³、0.472±0.088 m³/m³、0.500±0.084 m³/m³。因此，总体上，从上部（子区1和子区2）至中部（子区3）和下部（子区4），锡尔河流域中下游生长季SMC总体呈现增加的趋势，但存在显著的季节差异和一定的空间差异。

空间上，在生长季初期春季（4-5月），SMC的高值区域主要集中在子区2、子区3和子区4，即从恰尔达拉水库周边至下游的克孜勒奥尔达和临近咸海的拜科努尔地区（图4）。而从夏末6月开始至9月下半月，SMC逐渐集中于齐姆肯特与塔什干等区域。子区1、子区2的整个生长季SMC变化趋势与整个区域类似，即“降低—增加—降低”（图5）；子区3、子区4的SMC变化趋势相近，即“降低—保持平稳—增加—减少”。

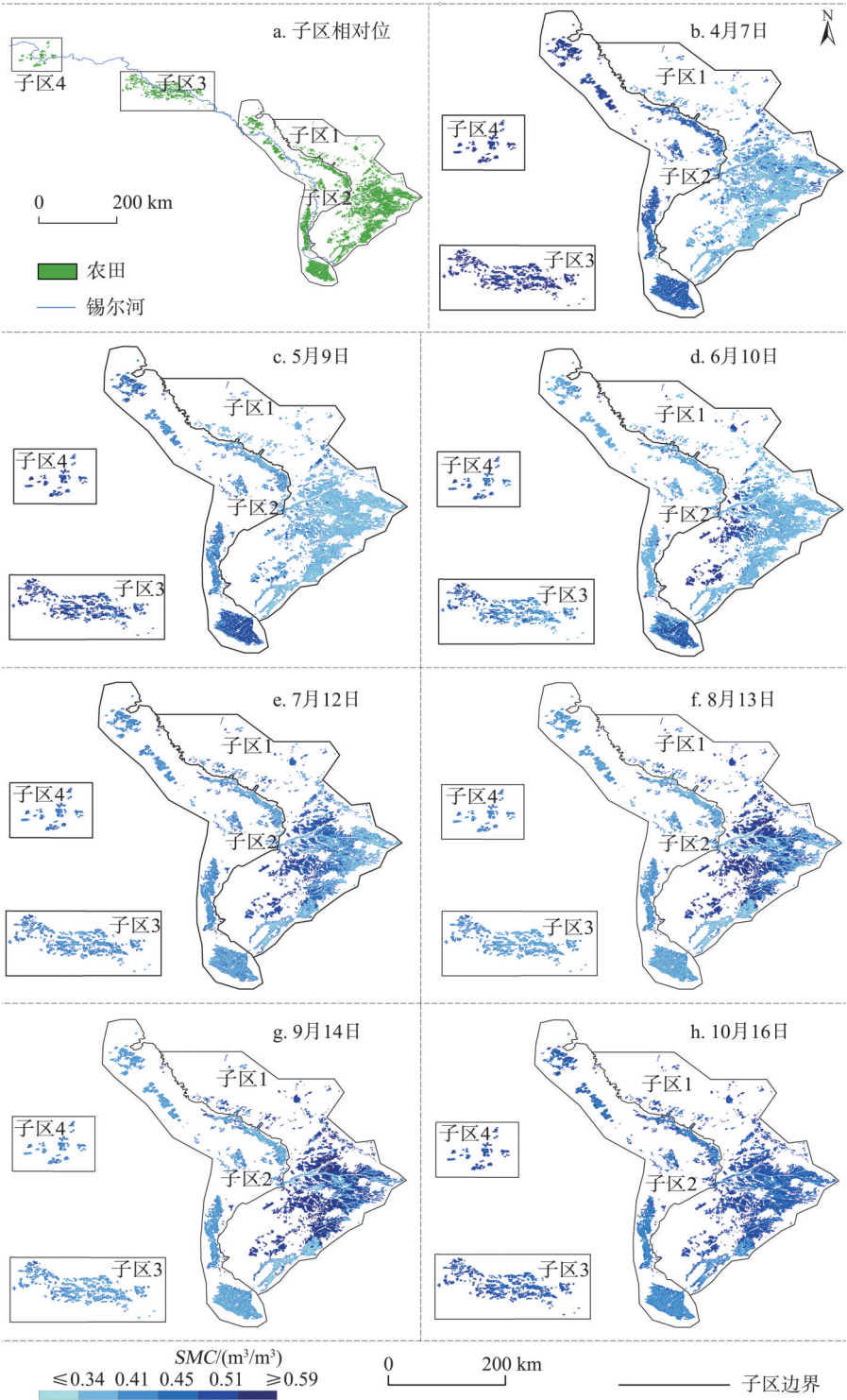


图4 锡尔河流域中下游农田区域2017年生长季SMC时空分布
(通过移动研究子区3和子区4的空间位置, 改善SMC空间显示)

Fig. 4 The retrieved SMC in the farmland in the middle and lower reaches of Syr Darya River Basin during the growing season in 2017 (In order to highlight the SMC spatial distribution, the positions of region 3 and region 4 were moved)

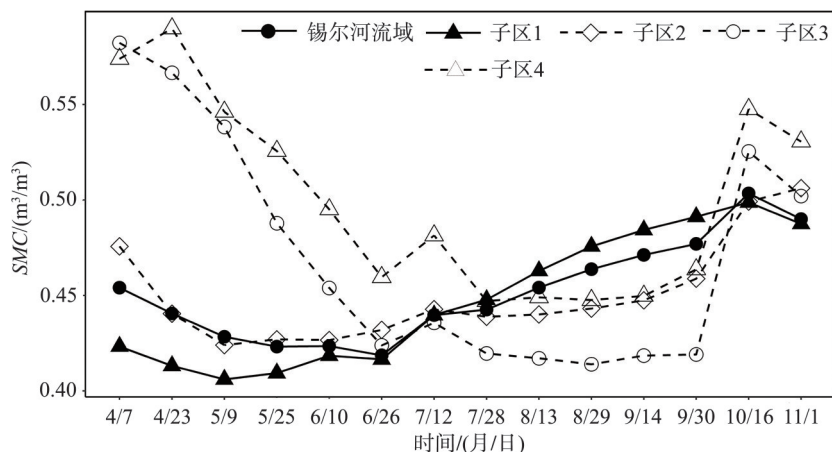


图5 2017年生长季锡尔河流域中下游农田SMC均值变化

Fig. 5 Change of the mean SMC during the growing season in the middle and lower catchments of the farmland in the middle and lower reaches of Syr Darya River Basin in 2017

3 结论与讨论

3.1 结论

(1) 利用 Sentinel-1 影像、MODIS 产品和 SRTM 数据, 提取锡尔河流域中下游平原区雷达后向散射系数、植被指数、地表温度、蒸散发、下垫面特性等影响因子的 32 个 SMC 因子, 经相关分析选择了 27 个显著的 SMC 反演因子 ($P < 0.05$), 参与 SMC 反演模型的构建, 其中 Albedo 与 SMC 相关性最高, 可见光地表反照率 (AIV) 与近红外地表反照率 (AIN) 的均值、最大值、最小值与 SMC 的相关性达到 $P < 0.01$ 信度水平。

(2) 基于显著相关的 27 个 SMC 反演因子设计三组试验方案, 分别与随机森林 RF、支持向量回归 SVR、BP 神经网络三种机器学习模型结合, 发现基于 RF 结合 SMC 所有因子的组合方案, 其 SMC 反演精度最高, 该方案被用于反演 2017 年生长季锡尔河流域平原区农田 SMC 时空动态。

(3) 锡尔河中下游平原区从上部至下部整个生长季 SMC 总体呈逐渐增加的态势, 但存在显著时空差异: 春季和秋季 SMC 较高而夏季较低。这种差异主要由土壤质地、热量条件和地表植被状况季节或空间差异引起的。春季 SMC 的主控因子是土壤质地和地表植被状况, 平原区下部农田 SMC 要高于上部 SMC; 在夏季农田土壤水分的主控因子是热量条件, 导致空间上平原上部和下部土壤 SMC 空间差异不显著; 秋季 SMC 的主控因子植被状况抵消地表温度和土壤质地差异对 SMC 的影响, 其结果是秋季 SMC 空间差异不显著。

(4) 基于多源遥感数据尽可能全面获取 SMC 的影响因子, 在此基础上采用机器学习模型, 可显著提高 SMC 反演的精度, 这在一定程度上克服了因考虑 SMC 影响因子不足获取更高 SMC 精度的限制, 为农田干旱监测提供支持。

3.2 讨论

本文实现了提出的三个研究目标: (1) 基于相关分析从多源遥感数据构造的 SMC 影响因子中筛选反演因子; (2) 寻找最优机器学习方法和 SMC 影响变量反演组合方案; (3) 反演生长季农田 SMC。针对研究目标①, 除考虑 Sentinel-1 的垂直极化后向散射系数

VV和MODIS产品的NDVI以及这两类数据的组合外,还考虑温度、蒸散发、地形等因子对SMC的影响,提高了反演精度(表4),这在一定程度上弥补了机器学习结合遥感的反演SMC的不足。针对研究目标②,我们的结论是RF较常用的其他两种机器学习方法(BPNN和SVR)更适用于干旱区地表SMC反演。对于研究目标③,最后基于构建的RF模型,反演了2017年生长季锡尔河中下游灌溉农田生长季SMC时空分布,并分析其变化特征。

2017年锡尔河中下游灌溉农田生长季SMC的时空差异宏观上主要由土壤质地、热量条件和地表植被状况季节或空间差异引起的。采用最小显著性差异法^[29],比较生长季的不同时段(春、夏、秋季)内锡尔河流域不同子区域之间的绿洲农田SMC差异(图6),并采用该方法分别比较热量条件(地表温度LST)(图7)、植被状况(NDVI)(图8)以及土壤质地(图2)时空差异,及其对的SMC时空格局差异造成的影响。在生长季初期春季(4-5月),整个区域热量条件没有显著差异(图5),但中、下部平原区(子区3和子区4)的农田SMC显著高于上部区域(子区1和子区2)(图5),因此春季影响SMC的主控因子是土壤质地^[33-34]与植被状况。一方面,中、下部平原区的土壤质地比上部区域更加粘重(图2),中、下部地区土壤保墒性更好,有效涵养了降水与大量积雪融水;另一方面,中、下部克孜勒尔达州种植了大量苜蓿等饲料作物,比例高达33%~35%,植被覆盖状况也要优于上部(图8),水分损耗较上部更少,这是中下部SMC较高的另一个重要原因。

在生长季的中期夏季(6-8月),高温炎热,农田灌溉开始实施,作物生长旺盛,地表潜在蒸发强度大,SMC宏观主控因子热量条件抵消土壤质地和地表植被状况差异对SMC的影响^[35];尽管不同子区热量条件存在差异,但在中、下部区域与上部区域之间并

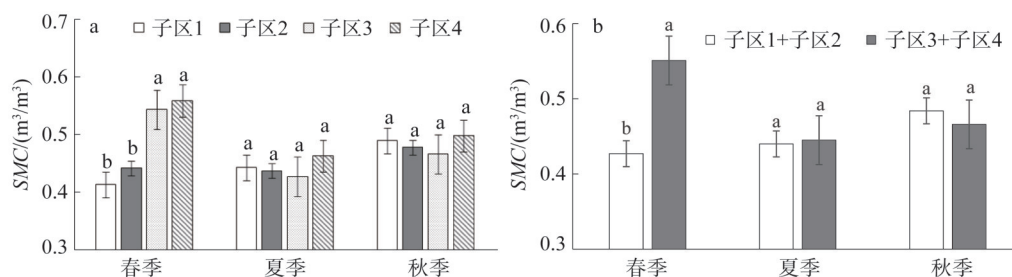


图6 表锡尔河流域中下游SMC均值与标准差时空差异

Fig. 6 The spatiotemporal difference of means and standard deviations of SMC at different time

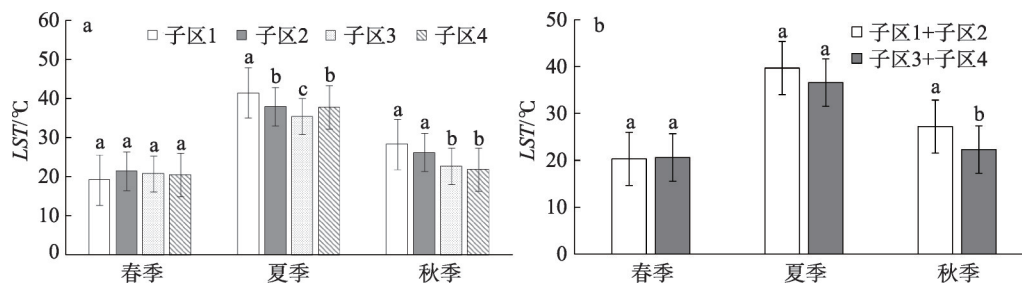


图7 锡尔河流域中下游地表温度均值与标准差时空差异

Fig. 7 The spatiotemporal difference of means and standard deviations of LST at different time

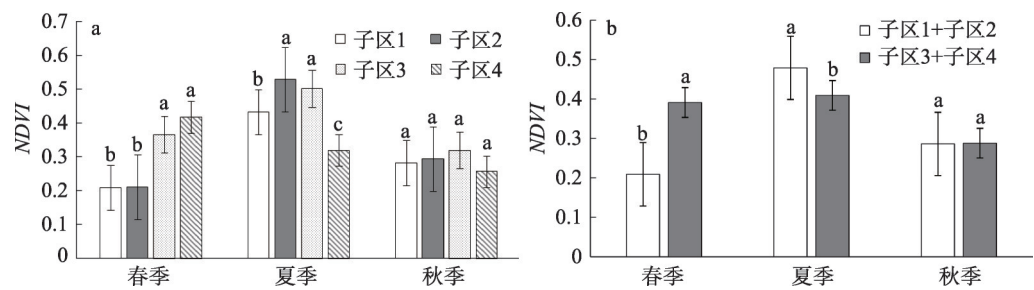


图8 锡尔河流域中下游NDVI均值与标准差时空差异

Fig. 8 The spatiotemporal difference of means and standard deviations of NDVI at different time

不显著，且热量条件差异导致的土壤水分损失由灌溉补水进行补偿，这是导致夏季SMC空间变异不显著的根本原因。

生长季末期秋季（9-10月），农田灌溉基本停止，各区域NDVI下降且无显著差异。地表温度普遍降低，上部的热量条件显著高于中、下部区域，且上部区域土壤质地较轻，砂质粘壤土等较为疏松的土壤保墒较差，更高的温度与较轻的土壤质地本应使上部的SMC更低，但结果却是SMC在上部与中、下部区域无显著差异。所以植被状况NDVI是秋季SMC主控因子，抵消掉热量条件与土壤质地的差异对SMC的影响。

研究结果说明随机森林RF结合多因子具有最高反演精度。RF结合多因子组合的验证精度介于已有的相关研究之间（表5），且优于基于单一因子的研究结果^[17,36]，所以即使不同研究中下垫面存在差异，仍在一定程度上说明本文SMC反演方法具有较高的适用性。同为干旱区农田SMC反演，Zeng^[37]等和Hassan等^[21]的研究，其SMC反演最高精度要优于本文，但他们的采样与反演区仅限于其小面积的试验田。

表5 机器学习结合遥感数据的SMC反演对比

Table 5 The comparison of different machine learning methods for estimating SMC by remotely sensed data					
研究	数据源	建模因子	方法	RMSE	研究区
Pasolli等 ^[36]	RADARSAT-2	HH或VV单个因子	SVR	0.0485 m ³ /m ³	意大利
Pasolli等 ^[17]	RADARSAT-2	HH或VV单个因子	SVR	5.38%~6.85%	意大利
Santi等 ^[24]	AMSR2、Envisat等	VV+NDVI或HH+NDVI双因子	ANN	0.023~0.052 m ³ /m ³	巴西
Paloscia等 ^[14]	Sentinel-1、MODIS等	VV+NDVI或者HH+NDVI双因子	ANN	2.32%~5.47%	意大利等 (部分农田)
Alexakis等 ^[15]	Sentinel-1、Landsat8	NDVI、HV等4个因子	ANN	0.022~0.058 m ³ /m ³	希腊
Zeng等 ^[37]	土样光谱信息	热红外波段相关的3~4个因子	ANN	0.017~0.032 m ³ /m ³	中国内蒙(农田)
Hassan等 ^[21]	AggieAir 无人机高光 谱影像	NDVI等10个因子	ANN	2%	美国犹他州 (农田)
ÖZerdem等 ^[16]	RADARSAT-2	VV等10个因子	ANN	2.84%~9.76%	土耳其(农田)
本文	Sentinel-1、MODIS等	NDVI均值等27个因子	RF	0.039 m ³ /m ³	锡尔河中下游 绿洲农田

注：RADARSAT-2、Envisat、Sentinel-1 为主动雷达 SAR 数据。HH、VV、HV 分别为雷达的水平极化、垂直极化、交叉极化后向散射系数。

本文为中尺度的区域SMC反演，在以后的研究中，可考虑采用空间分辨率更高的数据，如Sentinel-2，以求获取更为精细的SMC时空分布信息。

致谢: 本研究获得国家科技基础条件平台—国家地球系统科学数据共享平台 (www.geo-data.cn)、新疆与中亚科学数据中心 (midasia.data.ac.cn)、中科院中亚生态与环境研究中心和新疆生态与地理研究所超级计算机中心技术支持。

参考文献(References):

- [1] LEGATES D R, MAHMOOD R, LEVIA D F, et al. Soil moisture: A central and unifying theme in physical geography. *Progress in Physical Geography*, 2011, 35(1): 65-86.
- [2] KANG J, JIN R, LI X, et al. High spatio-temporal resolution mapping of soil moisture by integrating wireless sensor network observations and MODIS apparent thermal inertia in the Babao River Basin, China. *Remote Sensing of Environment*, 2017, 191(sC): 232-245.
- [3] ZHU W, JIA S, LYU A. A time domain solution of the Modified Temperature Vegetation Dryness Index (MTVDI) for continuous soil moisture monitoring. *Remote Sensing of Environment*, 2017, 200: 1-17.
- [4] SANTI E, PALOSCIA S, PETTINATO S, et al. Robust assessment of an operational algorithm for the retrieval of soil moisture from AMSR-E data in central Italy. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2016, 9(6): 2478-2492.
- [5] ZENG J, LI Z, CHEN Q, et al. Evaluation of remotely sensed and reanalysis soil moisture products over the Tibetan Plateau using in-situ observations. *Remote Sensing of Environment*, 2015, 163: 91-110.
- [6] KIM S, LIU Y Y, JOHNSON F M, et al. A global comparison of alternate AMSR2 soil moisture products: Why do they differ. *Remote Sensing of Environment*, 2015, 161: 43-62.
- [7] COLLIANDER A, JACKSON T, BINDLISH R, et al. Validation of SMAP surface soil moisture products with core validation sites. *Remote Sensing of Environment*, 2017, 191: 215-231.
- [8] ZENG J, CHEN K-S, BI H, et al. Radar response of off-specular bistatic scattering to soil moisture and surface roughness at L-band. *IEEE Geoscience and Remote Sensing Letters*, 2016, 13(12): 1945-1949.
- [9] XING M, QUAN X, LI X, et al. An extended approach for biomass estimation in a mixed vegetation area using ASAR and TM data. *Photogrammetric Engineering & Remote Sensing*, 2014, 80(5): 429-438.
- [10] KWEON S-K, OH Y. A modified water-cloud model with leaf angle parameters for microwave backscattering from agricultural fields. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 53(5): 2802-2809.
- [11] MERLIN O, AL BITAR A, WALKER J P, et al. An improved algorithm for disaggregating microwave-derived soil moisture based on red, near-infrared and thermal-infrared data. *Remote Sensing of Environment*, 2010, 114(10): 2305-2316.
- [12] PENG Z, JIANLI D, FEI W, et al. Retrieval methods of soil water content in vegetation covering areas based on multi-source remote sensing data. *Journal of Remote Sensing*, 2010, 14(5): 966-981.
- [13] TEMIMI M, LECONTE R, CHAOUCH N, et al. A combination of remote sensing data and topographic attributes for the spatial and temporal monitoring of soil wetness. *Journal of Hydrology*, 2010, 388(1): 28-40.
- [14] PALOSCIA S, PETTINATO S, SANTI E, et al. Soil moisture mapping using Sentinel-1 images: Algorithm and preliminary validation. *Remote Sensing of Environment*, 2013, 134: 234-248.
- [15] ALEXAKIS D D, MEXIS F-D K, VOZINAKI A-E K, et al. Soil moisture content estimation based on Sentinel-1 and auxiliary earth observation products. A hydrological approach. *Sensors*, 2017, 17(6): 1455, Doi: 10.3390/s17061455.
- [16] ÖZERDEM M S, ACAR E, EKINCI R. Soil moisture estimation over vegetated agricultural areas: Tigris Basin, Turkey from Radarsat-2 data by Polarimetric Decomposition Models and a Generalized Regression Neural Network. *Remote Sensing*, 2017, 9(4): 395, Doi: 10.3390/rs9040395.
- [17] PASOLLI L, NOTARNICOLA C, BERTOLDI G, et al. Soil moisture monitoring in mountain areas by using high-resolution SAR images: Results from a feasibility study. *European Journal of Soil Science*, 2014, 65(6): 852-864.
- [18] STAMENKOVIC J, FERRAZZOLI P, GUERRIERO L, et al. Crop backscatter modeling and soil moisture estimation with support vector regression. *Proceedings of the Geoscience and Remote Sensing Symposium (IGARSS)*, 2014 IEEE International. Quebec City, QC, Canada. 2014: 3228-3231.
- [19] ZHANG X, CHEN B, FAN H, et al. The potential use of multi-band SAR data for soil moisture retrieval over bare agri-

- cultural areas: Hebei, China. *Remote Sensing*, 2015, 8(1): 7, Doi: 10.3390/rs8010007.
- [20] KE Y, IM J, PARK S, et al. Downscaling of MODIS one kilometer evapotranspiration using Landsat-8 data and machine learning approaches. *Remote Sensing*, 2016, 8(3): 215, Doi: 10.3390/rs8030215.
- [21] HASSAN-ESFAHANI L, TORRES-RUA A, JENSEN A, et al. Assessment of surface soil moisture using high-resolution multi-spectral imagery and artificial neural networks. *Remote Sensing*, 2015, 7(3): 2627-2646.
- [22] STAMENKOVIĆ J, FERRAZZOLI P, GUERRIERO L, et al. Joining a discrete radiative transfer Model and a kernel retrieval algorithm for soil Moisture estimation From SAR Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015, 8(7): 3463-3475.
- [23] PASOLLI L, NOTARNICOLA C, BRUZZONE L. Multi-objective parameter optimization in support vector regression: General formulation and application to the retrieval of soil moisture from remote sensing data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2012, 5(5): 1495-1508.
- [24] SANTI E, PALOSCIA S, PETTINATO S, et al. Application of artificial neural networks for the soil moisture retrieval from active and passive microwave spaceborne sensors. *International journal of applied earth observation and geoinformation*, 2016, 48: 61-73.
- [25] GORELICK N, HANCHER M, DIXON M, et al. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 2017, 202: 18-27.
- [26] ATTEMA E, ULABY F T. Vegetation modeled as a water cloud. *Radio Science*, 1978, 13(2): 357-364.
- [27] BREIMAN L. Random forests. *Machine Learning*, 2001, 45(1): 5-32.
- [28] CORTES C, VAPNIK V. Support-vector networks. *Machine Learning*, 1995, 20(3): 273-297.
- [29] VOUNDI NKANA J, TONYE J. Assessment of certain soil properties related to different land-use systems in the Kaya watershed of the humid forest zone of Cameroon. *Land degradation & Development*, 2003, 14(1): 57-67.
- [30] BREIMAN L. Random forests. *Machine Learning*, 2001, 45(1): 5-32.
- [31] SCHÖLKOPF B, SMOLA A J, WILLIAMSON R C, et al. New support vector algorithms. *Neural Computation*, 2000, 12(5): 1207-1245.
- [32] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors. *Nature*, 1986, 323(9): 533-536.
- [33] 王修康, 戚兴超, 刘艳丽, 等. 泰山山前平原三种土地利用方式下土壤结构特征及其对土壤持水性的影响. *自然资源学报*, 2018, 33(1): 63-74. [WANG X K, QI X C, LIU Y L, et al. Soil Structure and its effect on soil water holding property under three land use patterns in piedmont plain of Mountain Tai. *Journal of Natural Resources*, 2018, 33(1): 63-74.]
- [34] ZHOU H, CHEN Y, LI W. Soil properties and their spatial pattern in an oasis on the lower reaches of the Tarim River, Northwest China. *Agricultural water management*, 2010, 97(11): 1915-1922.
- [35] 王学春, 李军, 王红妮, 等. 黄土高原冬小麦田土壤水分与小麦产量对降水和气温变化响应的模拟研究. *自然资源学报*, 2017, 32(8): 1398-410. [WANG X C, LI J, WANG X N, et al. Simulation of the response of soil water in winter wheat field and winter wheat yield to rainfall and temperature change on the Loess Plateau. *Journal of Natural Resources*, 2017, 32(8): 1398-4102.]
- [36] PASOLLI L, NOTARNICOLA C, BRUZZONE L, et al. Polarimetric RADARSAT-2 imagery for soil moisture retrieval in alpine areas. *Canadian Journal of Remote Sensing*, 2012, 37(5): 535-547.
- [37] ZENG W, XU C, HUANG J, et al. Predicting near-surface moisture content of saline soils from near-infrared reflectance spectra with a modified Gaussian model. *Soil Science Society of America Journal*, 2016, 80(6): 1496-1506.

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

WANG Hao^{1,2}, LUO Ge-ping^{1,2,3}, WANG Wei-sheng¹, PACHIKIN Konstantin⁴,
LI Yao-ming¹, ZHENG Hong-wei^{1,2}, HU Wei-jie¹

(1. State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, CAS, Urumqi 830011, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China; 3. Central Asian Center for Ecology and Environmental Research, CAS, Urumqi 830011, China; 4. The Kazakh Scientific Research Institute of Soil Science and Agriculture Chemistry, Almaty 050060, Kazakhstan)

Abstract: The use of machine learning method to estimate Soil Moisture Content (*SMC*) from multi-source remotely sensed data is a hot topic in the *SMC* inversion research. However, taking no account of the important variables of *SMC* in the ML method makes the *SMC* results uncertain. The Sentinel-1 and MODIS image products and the STRM data were obtained and used for extracting 32 *SMC* variables, such as backscattering coefficient, vegetation index, surface temperature and evapotranspiration. A total of 27 significant ($P < 0.05$) *SMC* variables were selected as input parameters referring to the correlation analysis result, and the input parameters were assigned to 3 groups. Random forest, Support vector regression and Back Propagation Neural Network were tested with 3 groups parameters. The Random forest with the group with all input parameters showed the best estimation accuracy, with the RMSE being $0.039 \text{ m}^3/\text{m}^3$, and it was used for the inversion of *SMC* in the farmland in the middle and lower reaches of Syr Darya River Basin during the growing season of 2017. The retrieved *SMC* gradually increased in the middle to the lower reaches during the growing season, but there were significant temporal and spatial differences: *SMC* in spring and autumn was higher than that in summer. These differences were mainly caused by seasonal or spatial differences in soil texture, heat conditions (temperature) and vegetation cover. In spring, *SMC* in the lower part of the plain is higher than that in the upper part, and the main *SMC* controlling factors were soil texture and vegetation cover. In summer, the main *SMC* controlling factors were heat condition. Irrigation compensated for the influence of heat condition difference, resulting in no significant spatial difference of *SMC* between upper and lower parts of the plain. The main *SMC* controlling factors in autumn were soil texture and heat conditions, the influence of surface temperature compensated for the influence of soil texture on *SMC*, as a result, there was no significant spatial difference of *SMC* in autumn. With regard to overcoming the limitation of taking no account of the important variables in estimating *SMC*, the research method adopted in this study improves the retrieved *SMC* accuracy to a large extent.

Keywords: soil moisture content; machine learning; middle and lower reaches of Syr Darya River Basin; Sentinel-1; MODIS; SRTM