引文格式：吴雄辉,赵若言,王健. 基于“RF-EnKF”的土壤水分反演及降尺度数据融合[J]. 地理空间信息, 2025, 23(8):1-8. doi:10. 3969/j.issn. 1672-4623.2025.08.001

基于“RF-EnKF”的土壤水分反演及降尺度数据融合

吴雄辉[1]，赵若言[1]，王　健[1*]
（1. 成都理工大学 地理与规划学院，四川 成都 610059）

摘　要：当前，基于机器学习的土壤水分反演与土壤水分产品数据的降尺度已成为获取土壤水分连续空间分布的2种常规手段，然而，如何有效综合2种手段的优势，获得更加准确、可靠的土壤水分分布亟需开展进一步的研究。鉴于此，以内蒙古自治区闪电河流域为研究区，以NDVI（normalized differential vegetation index）、LST（land surface temperature）和后向散射系数等为辅助数据，耦合随机森林（random forest，RF）和集合卡尔曼滤波（ensemble kalman filter，EnKF）算法，将反演和降尺度得到的土壤水分数据进行融合得到研究区1 km分辨率土壤水分数据，结合实地观测数据对模型效果进行评估和验证。结果表明：①基于EnKF算法的数据融合方法能够综合利用反演和降尺度的优势，进一步提升土壤水分空间分布预测的准确性。反演结果 R 为0.52，RMSE为0.018 $m^3 \cdot m^{-3}$，降尺度结果 R 为0.37，RMSE为0.022 $m^3 \cdot m^{-3}$，融合两者后的 R 提升至0.71，RMSE降低为0.010 $m^3 \cdot m^{-3}$；②流域内大部分区域土壤水分低于0.20 $m^3 \cdot m^{-3}$，南部边缘高海拔区域的土壤水分含量较高；结合随机森林特征重要性分析结果，NDVI、降水量以及DEM等对土壤水分空间分布的控制作用相对最为显著。该研究为准确获取土壤水分空间分布提供了一种新的思路，也为闪电河流域水资源管理等应用提供一定参考依据。

关键词：反演；降尺度；数据融合；随机森林；集合卡尔曼滤波
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2025）08-0001-08

Soil Moisture Inversion and Downscaled Data Fusion Based on RF-EnKF

WU Xionghui[1], ZHAO Ruoyan[1], WANG Jian[1]
(1. College of Geography and Planning, Chengdu University of Technology, Chengdu 610059, China)

Abstract: Soil moisture is not only an important variable in land surface interaction, but also plays an important role in water cycle and water resources management. At present, the soil moisture inversion based on machine learning and the downscaling of soil moisture product data have become two popular methods to obtain the continuous spatial distribution of soil moisture. However, how to effectively integrate the advantages of two methods to obtain more accurate and reliable soil moisture distribution needs further research. In view of this, taking the Inner Mongolia Lightning River basin as the study area, taking NDVI, LST and backscattering coefficient as supplementary data, combining the RF and EnKF, we fused the soil moisture data obtained from the inversion and downscaling to obtain a 1 km resolution soil moisture in the study area, and evaluated and verified the effectiveness of model by combining the field observation data. The results show that ①the data fusion method based on EnKF algorithm can simultaneously take advantage of the strengths of inversion and downscaling, and further improve the accuracy of spatial distribution prediction of soil moisture. The correlation coefficient R for the inversion result is 0.52, and RMSE is 0.018 m^3/m^3, while for the downscaling result R is 0.37 and RMSE is 0.022 m^3/m^3. After fusing the two results, R increases to 0.71 and RMSE decreases to 0.010 m^3/m^3. ②The soil moisture in most areas of the basin is less than 0.20 m^3/m^3, and the soil moisture in the high-altitude area of the southern edge is higher. Combined with the importance analysis results of RF features, NDVI, precipitation and DEM have significant control effects on the spatial distribution of soil moisture. This study provides a new idea for accurately obtaining the spatial distribution of soil moisture, and provides a reference for the application of water resources management in Shandian River basin.

Key words: inversion, downscaling, data fusion, RF, EnKF

近年来，随着计算机科学技术的发展，机器学习算法被许多学者运用于土壤水分产品数据的降尺度研究中[1-5]。机器学习可以有效捕捉复杂地表环境条件下土壤水分和环境因子之间的非线性关系，为不同数据源的信息整合和海量数据处理提供了便利[6]。而在众多机器学习算法中，随机森林（RF）是一种既可用于分类又可用于回归的机器学习方法。RF模型中涉及的自适应、随机和去相关特性[7-8]可以自动对预测变量中的非线性关系进行建模，被许多学者研究表明其更适合运用于土壤水分降尺度研究中[9-10]。

项目来源：国家自然科学基金资助项目（42002295）。
第一作者简介：吴雄辉（1999—），硕士研究生，研究方向为土壤水分遥感反演，E-mail：2022020036@stu.cdut.edu.cn。
通信作者：王健（1990—），博士，副教授，研究方向为数学地质与遥感地质，E-mail：jwang@cdut.edu.cn。

然而，不论是土壤水分反演或是土壤水分产品降尺度，都易受到观测误差等诸多不确定性因素的影响而显示出不同的性能[11]。数据融合可以有效综合不同方法所得结果的优势，从而获得更高精度的土壤水分空间分布[12]。数据融合技术是将多源数据进行有效集成并剔除无用信息的过程，融合后的数据通常具有更高的完整性和准确性。而EnKF算法作为经典的数据融合方法之一，已经被许多学者运用于土壤水分数据的融合中[13-16]。这些研究表明，经过EnKF算法融合得到的土壤水分数据相较于原土壤水分数据不仅精度有明显的改善，且融合后的土壤水分数据与真实参考土壤水分数据的相关性相对较高[17]。然而，有关将土壤水分反演数据和土壤水分产品降尺度数据进行融合的研究报道相对较少。

鉴于此，本文为了更好地估算闪电河流域土壤水分分布情况，运用RF-EnKF耦合模型将反演与降尺度得到的土壤水分数据进行融合，最终得到闪电河流域1 km空间分辨率的土壤水分空间分布情况。

1 研究区域和数据集

1.1 研究区概况

闪电河流域(图1)是滦河流域的一部分，经纬度范围为115°51′47″~117°17′40″E、41°46′48″~42°39′05″N，全长约254 km，总流域面积4 105 km²[18]。

闪电河流域地形地貌复杂，地形主要为平原、丘陵和洼地，海拔1 072~2 229 m之间，地势南高北低。地表覆盖类型主要为林地、农田和草地几种。闪电河流域内四季分明，昼夜温差大。流域内最高气温为27.9℃，最低气温为-14.5℃。流域内植被分布较好，湿地较多。整体来说，流域内多风少雨，蒸发量大，水资源匮乏。

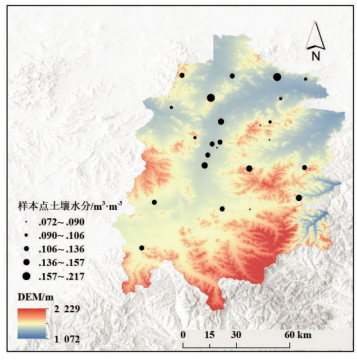


图1 研究区位置及样本点分布

审图号：GS(2023)2767

1.2 数据描述

本文涉及的数据主要包括SMAP土壤水分产品数据、MODIS产品数据、Sentinel-1A数据、土壤特性数据、地形数据和站点数据。其中，SMAP土壤水分数据用作土壤水分降尺度；MODIS产品数据用于计算相关的环境变量参数；Sentinel-1A数据用于提取后向散射系数VV和VH、土壤特性数据获取土壤砂粒含量Sand和土壤黏粒含量Clay、地形数据中的DEM。上述数据均用作土壤水分反演和土壤水分降尺度的辅助变量，站点数据用于土壤水分反演中模型的输入以及最终土壤水分的验证分析。主要数据集介绍如下。

1.2.1 SMAP数据

SMAP卫星是美国的地球观测卫星之一，于2015年1月31日发射[19]。SMAP共提供了4种土壤水分产品数据，分别为被动、主动、主动-被动和增强型被动土壤水分产品数据。有研究表明，SMAP土壤水分产品数据已经满足了卫星发射任务的精度要求[20]。因此，SMAP产品数据在全球土壤水分检测中发挥着重要作用[21-23]。

SMAP L4产品数据提供时间分辨率3 h，空间分辨率9 km的表面（0~5 cm）土壤水分数据，它利用了从星载仪器获取到的测量结果和陆面过程模型LSM（Land Surface Model）的输出结果[24]。

本文选取2020年3月17日、6月21日、9月25日和12月18日4 d SMAP L4 SM用作土壤水分降尺度的数据源。SMAP数据预处理过程在HEG tools中进行。上述数据可从美国国家航空航天局的地球科学数据系统（ESDS）的地球科学数据收集处下载（Earthdata | Earthdata （nasa.gov））。

1.2.2 MODIS数据

MODIS数据包含36个光谱波段，扫描幅宽为2 330 km，具有波段多、光谱分辨率高、重访周期短等优点[25]。本文所需要的数据包括MOD11A2的LST数据、MOD12Q1的VI数据以及MOD09GA的每日反射率产品数据。其中，MOD11A2产品数据提供了时间分辨率为8 d，空间分辨率为1 km的LST数据。NDVI数据由MOD13Q1产品数据提供，时间分辨率为16 d，空间分辨率为250 m。剩余相关指数计算所需的表面光谱反射率 d 来自MOD09GA数据产品，空间分辨率为500 m。利用反射率来得到的相关指数（表1），表中 R_{bi} 表示MODIS产品第 $i(i=1,2,\ldots,7)$ 个波段的地表反射率。本文借助ENVI软件中的MCTK插件对MODIS产品数据进行预处理。MODIS产品数据可从美国国家航空航天局的地球科学数据系统（ESDS）的地球科学数据收集处下载（Earthdata | Earthdata （nasa.gov））。

表1　相关指数的计算公式

相关指数	计算公式	变量说明
NDTI	$\frac{R_{b6}-R_{b7}}{R_{b6}+R_{b7}}$	归一化差值温度指数[26]（Normalized difference tillage index）
NSDSI	$\frac{R_{b6}-R_{b7}}{R_{b6}}$	归一化短波红外差值裸土水分指数[22]（Normalized shortwave-infrared difference bare soil moisture indices）

1.2.3　Sentinel-1A 数据

Sentinel 系列卫星为欧空局“哥白尼计划”的地球观测卫星系列之一，其在693 km的极地同步轨道上运行[7]。该系列包括2颗Sentinel-1卫星，搭载有 C 波段（5.405 GHz）合成孔径雷达，可全天时、全天候提供图像，用以监测地表形变（滑坡、地震等）、全球海洋环境变化以及陆地森林、地表水等动态变化[27]。本文选取的Sentinel-1A SAR数据使用的是IW模式的GRD数据，其成像幅宽为250 km，空间分辨率为5 m×20 m[28]。本文选取的为VV极化方式和VH极化方式，并通过欧空局开发的SNAP软件对Sentinel-1A影像进行基本的预处理工作，主要为：热噪声去除、轨道校正、消除GRD边框噪声、辐射定标、相干斑滤波和地形校正等，最后将得到后向散射系数转化为分贝格式。Sentinel-1A数据下载网址为欧空局（https://sci-hub.copernicus.eu/）。

1.2.4　土壤特性数据

本文使用的土壤性质变量是砂土和黏土的含量。这些数据来自国际土壤参考信息中心（ISRIC）发布的世界土壤信息数据（Soil Grids, SoilGrids250m 2.0）。该数据由机器学习方法制作的全球尺度的250 m空间分辨率数据[29]。

1.2.5　地形数据

地形可以影响土壤水分在不同尺度上的空间分布[30]。DEM以及相关衍生变量已经在土壤水分降尺度与反演研究中获得广泛应用。本文所使用的90 m空间分辨率的DEM数据来自地理空间数据云（https://www.gscloud.cn/）。

1.2.6　降水量数据

本文所使用的降水量数据来自国家地球系统科学数据中心（https://www.geodata.cn/）。该数据集为中国逐月降水量数据，空间分辨率为0.008 333 3°（约为1 km），其中降水量单位为0.1 mm[31]。

1.2.7　站点数据

本文所使用的站点实测土壤水分数据来自闪电河流域水循环和能量平衡遥感综合试验：2020年土壤温湿度飞行同步测量数据集[32]。该数据集包含深度为3、5、10、20 cm和50 cm的土壤水分和土壤温度数据[33]。测量数据稳定后定期针对每个站点的每一层土壤采集土壤样品，分析重量/体积含水量、容重和土壤质地等，以对原始测量数据进行校准。

2　方　法

本文以DEM、LST、NDVI、土壤性质等环境因子为预测变量，分别以实测土壤水分和SMAP土壤水分为响应变量，基于随机森林建立土壤水分反演和降尺度模型，得到研究区域内1 km分辨率土壤水分分布，进而借助EnKF方法将前述2种模型的结果进行融合，作为该区域土壤水分含量的最终结果，并采用交叉验证方法对上述反演、降尺度以及最后融合结果进行评估和对比。下面将对所用方法进行介绍。

2.1　随机森林

随机森林（RF）是由Breiman[34]提出的一种基于树的机器学习方法，可以应用于分类和回归任务[35]。RF模型基于Bootstrapping和随机属性划分2种随机机制，实现对目标变量的准确预测，并且可以有效防止模型过拟合。由于目标变量——土壤水分是一个连续变化的随机变量，所以本文采用随机森林回归拟合各个环境变量与土壤水分之间的定量关系，组成RF模型的决策树可表示为：

$$\left\{h\left(x,\Theta_i\right),i=1,2,3\ldots m\right\} \tag{1}$$

式中，Θ_i 为生成的独立同分布的决策树参数；x 为自变量；m 为生成的决策树的数量；$h\left(x,\Theta_i\right)$ 为该决策树的预测值。RF通过集成各个决策树的预测结果对目标变量进行合理预测。

有关随机森林算法原理以及执行步骤等，可参考文献[35]。本文使用的RF模型是使用python中的scikit-learn包建立，并使用Grid Search法对超参数进行优化选择。

2.2　集合卡尔曼滤波

集合卡尔曼滤波（EnKF）是Evensen[36]根据Epstein的随机动态预报理论，提出的通过Monte Carlo法（总体积分法）来计算状态变量的预测误差协方差的一种数据同化算法[11]。EnKF是将模型状态变量的预测看作一个近似随机动态的预测，用状态总体去代表随机动态预测中的概率密度函数，通过向前积分计算不同时刻的概率密度函数所对应的协方差[37]。该方法的优点在于其便于计算，因此，被广泛应用于大气与水文的数据融合研究中。

其非线性过程如下：

1）初始化背景场：

$$X_0=\left(X_{0,1}^a,X_{0,2}^a,\cdots,X_{0,n}^a\right) \tag{2}$$

2）计算 $k+1$ 时刻的预测值和预测误差协方差：

$$X_{t+1,i}^f=M\left(X_{t,i}^a\right)+W_i \tag{3}$$

$$P_{t+1}^f=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_{t+1,i}^f-\overline{X_{t+1}^f}\right)\left(X_{t+1,i}^f-\overline{X_{t+1}^f}\right)^T \tag{4}$$

3）计算 $t+1$ 时刻的增益矩阵：

$$K_{t+1}=p_{t+1}^f H_{t+1}^T\left(H_{t+1}P_{t+1}^f H_{t+1}^T+R_{t+1}\right)^{-1} \tag{5}$$

4）计算 $k+1$ 时刻分析值和分析误差协方差：

$$X_{t+1,i}^a=X_{t+1,i}^f+K_{t+1}\left(Z_{t+1}-H\left(X_{t+1,\ i}^f\right)+U_i\right) \tag{6}$$

$$P_{t+1}^a=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_{t+1,i}^a-\overline{X_{t+1}^a}\right)\left(X_{t+1,i}^a-\overline{X_{t+1}^a}\right)^T \tag{7}$$

5）完成预测和更新步骤后，进入下一时刻。上述公式中各符号及其背后涵义，可参考文献[11]。本文运用 EnKF 来实现数据融合的思路为：将通过卫星遥感反演得到的 1 km 土壤水分数据作为模型状态变量，将降尺度得到的 1 km 的 SMAP 土壤水分数据作为观测数据，借助 EnKF 进行数据融合。

2.3 交叉验证

交叉验证是一种评价方法，它不需要任何前提假设可对模型的泛化误差进行估计，同时操作简便，因此在模型评价实践中获得广泛应用。

留一法交叉验证[38]（leave-one-out cross-validation，LOO-CV）是交叉验证中较为常见的方法。它的基本思路为：每次从一个样本数为 N 的数据集中取出一个作为验证集，剩下的 $N-1$ 个样本数据则作为训练集，重复上述步骤 N 次，依次取遍样本数据集中的数据作为验证集，最后，将上述 N 次结果的平均作为泛化误差估计。泛化误差估计如下：

$$R_{\mathrm{LOO}}(f,D_N)=\frac{1}{N}\sum_{i=1}^{N}L\left(f\left(D_N^{(-i)}\right),y_i\right) \tag{8}$$

式中，D_N 为 N 个样本的数据集；$D_N^{(-i)}$ 为取出 i 后剩下的样本数据；y_i 为第 i 次取出的样本的观测值；$L(*)$ 为损失函数。

本文由于进行验证的实测站点数据有限，因此采用留一法交叉验证对土壤水分反演数据进行评估验证。

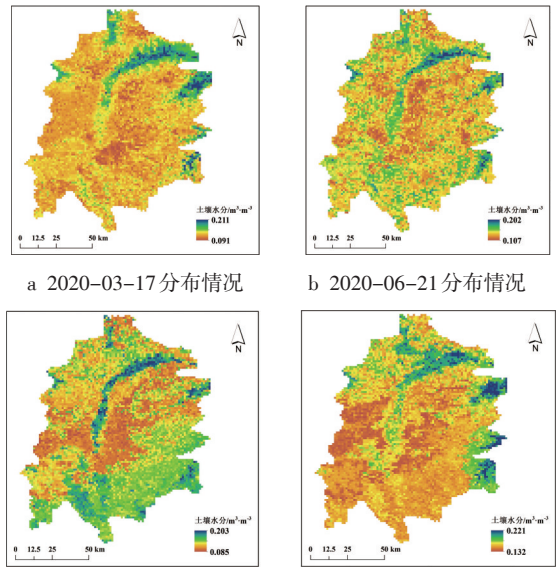
3 结果与分析

3.1 基于随机森林的土壤水分反演

3.1.1 土壤水分反演结果

本文以 MODIS 产品数据、Sentinel-1A 后向散射系数、土壤特性数据和地形数据作为模型的输入参数，基于随机森林模型计算出各输入参数的特征重要性。其中，NDVI 为归一化植被指数，NDTI 位归一化差值温度指数，LST 为地表温度，NSDSI 为短波红外差分裸露土壤水分指数，VV 和 VH 为 Sentinel-1A 的极化参数，Sand 和 Clay 为土壤特性数据，DEM 为地形数据，pre 为降水量数据，输入参数共计 10 种。

随机森林的超参数设置如下：n 为总的特征数，则构建决策树的最大特征数目“max_features”设置为“sqrt”，即 $\sqrt{n}$；树的最大深度“max_depth”为 30；树的数量“n_estimators”为 400；叶子节点的最小样本数目“min_samples_leaf”为 3；节点样本最小数量“min_samples_split”为 5。反演得到土壤水分空间分布如图 2 所示。图 2 显示，闪电河流域内土壤水分普遍较低，这可能是由于闪电河流域内土壤质地主要为砂土，而砂土的蓄水能力较差所导致的。可能因为黏土的蓄水能力较好，有一定的持水性，故而黏土含量较高部分区域的土壤水分较高。



a 2020-03-17 分布情况　b 2020-06-21 分布情况

c 2020-09-25 分布情况　d 2020-12-18 分布情况

图2 土壤水分反演空间分布

审图号：GS(2023)2767

采用留一法交叉验证，通过对比反演得到的土壤水分与站点实测土壤水分，定量评估反演结果的准确性（图 3）。图 3 显示，反演得到的土壤水分与站点实测土壤水分的相关系数 R 为 0.52，均方根误差 RMSE 为 0.018 $m^3 \cdot m^{-3}$。

3.1.2 特征重要性

本文基于影响土壤水分的 10 个环境变量建立了随机森林模型。图 4 显示了各个环境变量在随机森林模型中的重要性得分，得分越高，特征越重要。如图 4

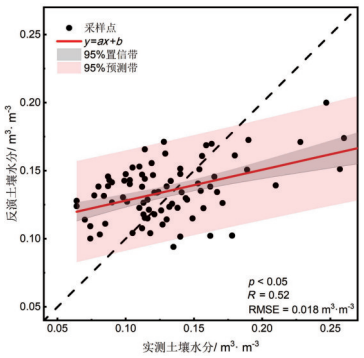


图3　反演土壤水分与站点实测对比

所示，土壤性质与土壤水分的相关性最高，其次为NDVI、LST、NDTI、降水量和后向散射系数。

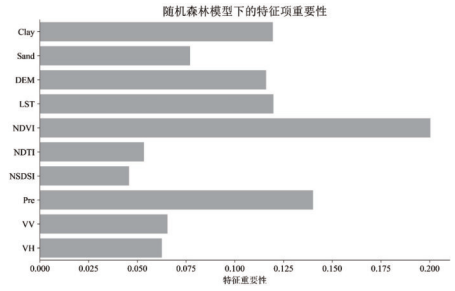


图4　随机森林模型特征重要性

归一化植被指数NDVI与地表温度LST构建的特征空间可以很好地解释陆地表面水热交换过程中的差异，因此两者通常被选取为土壤水分研究中常用的地表变量。其中NDVI能够反映地表植被覆盖情况，而植被可以通过影响太阳入射辐射、土壤蒸散发进而影响了土壤水分，而地表温度LST也是通过土壤蒸散发进而影响土壤水分，LST越高，土壤蒸散发效应越显著，地表土壤水分越低，这与之前的研究一致。对于降水量而言，它对土壤水分的影响是一个复杂的过程，一般情况下，降水量的增加会直接导致土壤水分的增加。但是，土壤水分并不是简单地随着降水量的增加而增加，当降水量逐渐增加，土壤水分接近饱和，土壤的水分吸收能力大大降低，导致土壤水分渗透率降低，降水量很大一部分随着地表径流流失而不是被土壤吸收。此外，在干旱地区，空气过于干燥，导致降水在到达地面之前蒸发，并且强烈的地表蒸散发作用也会导致土壤可吸收的降水量的减少，进而导致土壤水分减少[39-41]。就土壤性质来说，土壤的粒径和孔隙率对土壤水分的垂直渗透性有显著影响，进而影响水分的空间分布状况。土壤的粒径和孔隙率越大，地表水向地下水位更深的地方转移的速度越快。因此，砂土蓄水能力差，而黏土蓄水能力较好，具有一定的持水性[22, 29]。DEM的影响也不容忽视，由于地形起伏和拦截程度的差异会造成不同的产汇流路径，影响地表径流量，进而影响土壤水分；其次是不同的坡向所对应的水热条件与植被生长环境不同，阳坡相对于阴坡受到较多的太阳辐射，温度状况较好，但水分状况较差，进而导致土壤水分的差异[40-43]。雷达后向散射系数VV和VH对土壤水分也较为显著，其原因为雷达后向散射系数对地表土壤水分的变化比较敏感[44-45]。此外，NSDSI和NDTI 2个指数是由2个SWIR（1 550~1 750 nm和2 100~2 300 nm）波段计算所得，而在SWIR波段观察到光谱吸收与土壤水分显著相关，因此NSDSI和NDTI对土壤水分也具有较高的重要性[46-48]。

3.2　土壤水分数据产品降尺度

SMAP土壤水分数据产品具有高时间分辨率，但是其较低的空间分辨率限制了其在小范围区域的适用性。本文采用随机森林算法，以DEM、LST、NDVI、Sand、VV和VH等辅助数据作为环境变量，将9 km空间分辨率的SMAP土壤水分产品数据降尺度至1 km（图5）。

由图5可知，通过随机森林模型降尺度不仅能够提高SMAP土壤水分产品数据的空间分辨率（9 km到1 km），丰富细节信息，而且降尺度后的土壤水分在整体空间分布模式与降尺度前SMAP土壤水分保持高度一致。

使用土壤水分站点实测数据对降尺度前后的SMAP土壤水分数据进行评估，结果如图6所示。降尺度前SMAP土壤水分与站点实测土壤水分相关系数R为0.04，均方根误差RMSE为0.024 $m^3 \cdot m^{-3}$；而降尺度后的SMAP土壤水分与站点实测土壤水分相关系数R的为0.37，均方根误差RMSE为0.022 $m^3 \cdot m^{-3}$。以上数据表明，降尺度后的R值有所提升，说明降尺度后的土壤水分与站点实测土壤水分相关性有所增强。

降尺度前SMAP土壤水分与站点实测土壤水分呈现出较低的相关性，这可能是由于用于验证的站点数据较少，才会造成验证结果出现异常。此外，卫星数据与站点数据之间的空间尺度存在着较大的差异，可能会使两者结果不能相匹配，这也会导致最后的验证结果出现异常情况[49]。

3.3　基于EnKF的土壤水分数据融合

将通过随机森林模型反演得到的1 km空间分辨率土壤水分数据，以及随机森林模型降尺度得到的1 km空间分辨率的土壤水分数据，采用EnKF算法进行数据融合，所得结果如图7所示。

从图7可以看出，通过EnKF算法融合得到的土壤水分与降尺度得到的土壤水分在空间分布上保持一致。

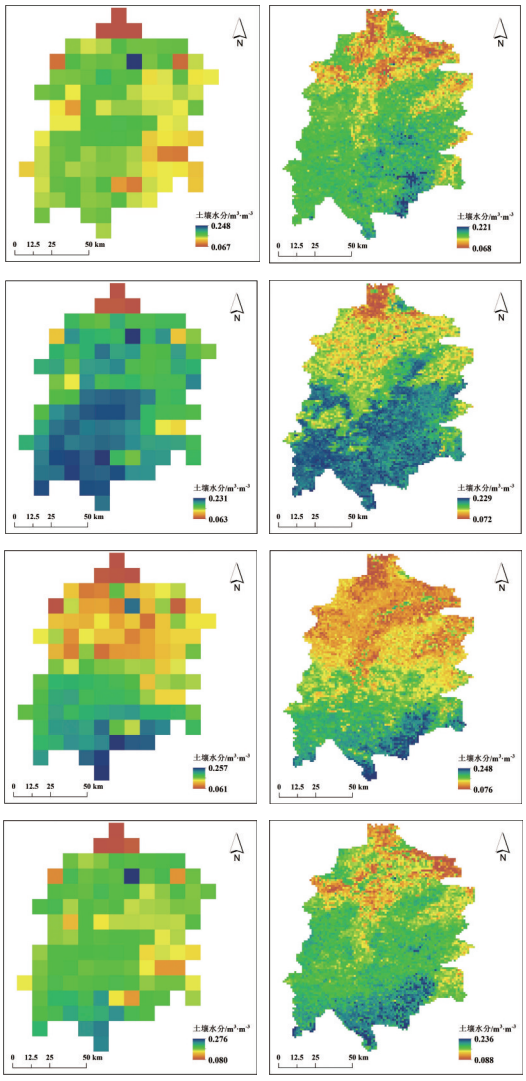
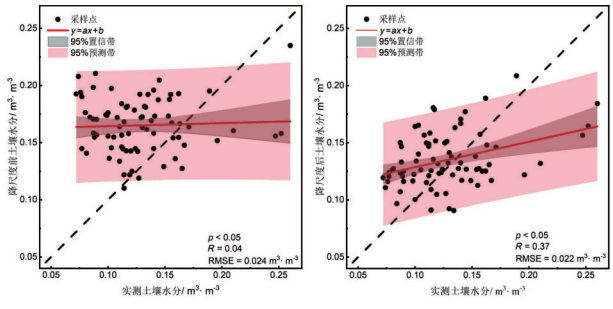


图5　降尺度前后SMAP土壤水分空间分布

审图号：GS(2023)2767

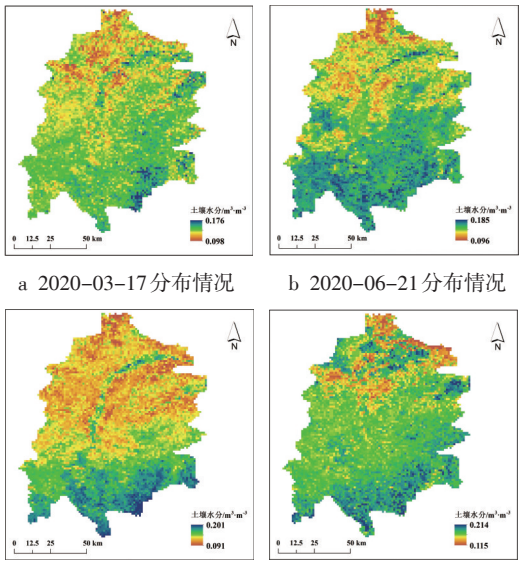
注：左侧为降尺度前，右侧为降尺度后，从上至下依次为2020-03-17、2020-06-21、2020-09-25、2020-12-18。



a 降尺度前　　b 降尺度后

图6　SMAP土壤水分降尺度前后与站点实测对比

从图3和图6（b）可知，反演得到的土壤水分与站点实测土壤水分的相关系数 R 为0.52，RMSE为0.018 $m^3 \cdot m^{-3}$，而降尺度得到的土壤水分与站点实测土壤水分相关系数 R 为0.37，RMSE为0.022 $m^3 \cdot m^{-3}$，相比较而言，将降尺度和反演得到的土壤水分利用En-KF算法进行数据融合，得到融合后的土壤水分与实测土壤水分对比结果如图8所示，融合后的土壤水分与站点实测土壤水分相关系数 R 为0.71，RMSE为0.010 $m^3 \cdot m^{-3}$。因此，融合后的土壤水分数据在精度方面得到了一定的改善，准确性也得到了可观的提升。



a 2020-03-17分布情况　　b 2020-06-21分布情况

c 2020-09-25分布情况　　d 2020-12-18分布情况

图7　融合后的土壤水分空间分布

审图号：GS(2023)2767

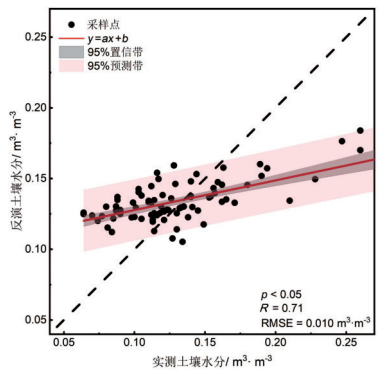


图8　融合后土壤水分与站点实测对比

3.4　闪电河流域土壤水分空间分布特征

从最终土壤水分融合的结果来看（图7），闪电河流域内土壤水分大部分区域低于0.20 $m^3 \cdot m^{-3}$。究其原因，主要是闪电河流域内气候地形复杂，草地为该流域内主要的植被覆盖类型，但其草质一般，产草量普遍较低，对土壤水分的保持作用较低。其次，闪电河流域位于干旱半干旱过渡带，日照时间长，蒸发旺盛，导致流域内土壤水分含量较低。此外，流域内不同时间段的土壤水分含量也有所差异，这可能与流域内降水相关。闪电河流域内降水偏少，主要集中在夏季，季节分配极其不均。因此，流域内3月份（0.098～0.176 $m^3 \cdot m^{-3}$）土壤水分含量最低，9月份（0.091～0.201 $m^3 \cdot m^{-3}$）土壤水分含量较高。值得注意的是，12月

份（0.115~0.214 $m^3 \cdot m^{-3}$）土壤水分含量最高，这可能与流域内冬季气温较低，日照时间减少，蒸发量减少有关。

闪电河流域内土壤水分空间分布特征整体上呈现为区域北部较南部干旱[25]，空间变化特征与流域内地势分布一致。从东西分布来看，流域内西部较东部干旱。这种空间分布特征可能与土壤质地和地表温度LST有关。根据土壤砂粒含量分布特征，闪电河流域内砂粒分布广泛，整体上呈现西北高东南低的特征，而砂粒的粒径和孔隙率较大，能够加快地表水向地下水位更深的地方迅速转移，进而导致流域内土壤水分呈现出东南高西北低的空间分布特征。地表温度LST分布特征为，流域内地表温度总体呈现西北高东南低的特征，而地表温度LST通过影响水分的蒸散发从而影响土壤水分含量，地表温度越高则土壤水分越低，从而导致流域内土壤水分呈现出东南高西北低的空间分布特征。

综上，闪电河流域内土壤水分的空间分布主要与土壤质地、地表温度和地表覆盖状况等因素有关，并且融合后的土壤水分含量与闪电河流域内相关地理要素的空间分布模式一致，间接表明本文所提的融合方法能够较好地反映闪电河流域的土壤水分空间分布情况。

4　结　论

本文提供了一种基于随机森林与集成卡尔曼滤波的反演与数据融合方法，来估计闪电河流域的土壤水分空间分布情况，并用实测数据对土壤水分推断结果进行了对比和评价。得到的结论如下：

1）运用EnKF算法进行融合得到的土壤水分相较于融合前的土壤水分在准确性上得到了提升，并充分利用反演和降尺度的优势，表明利用EnKF算法对反演和降尺度数据进行融合得到更高的精度的土壤水分数据的可行性；

2）闪电河流域内大部分区域土壤水分低于0.20 $m^3 \cdot m^{-3}$，南部边缘部分高海拔区域的土壤水分相对较高，结合随机森林特征重要性分析来看，NDVI对土壤水分影响最为显著，其次为降水量、LST、DEM、土壤性质和后向散射系数。

3）本文在对常用的土壤水分获取方法（反演和降尺度）的基础上，提出了将两者数据融合的方法，为获取更高准确性的土壤水分提供了一种新思路，同时为闪电河流域水资源管理等应用提供一定的参考依据。

参考文献

[1] Abbaszadeh P, Moradkhani H, Zhan X. Downscaling SMAP Radiometer Soil Moisture over the Conus Using an Ensemble Learning Method[J]. Water Resources Research, 2019,55(1):324-344

[2] Kolassa J, Reichle R H, Liu Q, et al. Estimating Surface Soil Moisture from Smap Observations Using a Neural Network Technique[J]. Remote Sensing of Environment,2018, 204:43-59

[3] Pelletier C, Valero S, Inglada J, et al. Assessing the Robustness of Random Forests to Map Land Cover with High Resolution Satellite Image Time Series over Large Areas[J]. Remote Sensing of Environment,2016,187:156-168

[4] Rodriguez-fernandez N J, Aires F. Soil Moisture Retrieval Using Neural Networks:Application to SMOS[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53 (11):5 991-6 007

[5] Wei Z S, Meng Y Z, Zhang W, et al. Downscaling SMAP Soil Moisture Estimation with Gradient Boosting Decision Tree Regression over the Tibetan Plateau[J]. Remote Sensing of Environment,2019,225:30-44

[6] 曹煜,方秀琴,杨露露,等. 基于随机森林的西辽河流域CCI土壤湿度降尺度研究[J]. 地球信息科学学报,2023,25(8): 1 669-1 681

[7] Bai J, Cui Q, Zhang W, et al. An Approach for Downscaling SMAP Soil Moisture by Combining Sentinel-1 SAR and MODIS Data[J]. Remote Sensing,2019,11(23):2 736-2 738

[8] Ezzahar J, Chehbouni A, Ouaadi N, et al. Sentinel-1 Backscatter and Interferometric Coherence for Soil Moisture Retrieval in Winter Wheat Fields Within a Semiarid Southmediterranean Climate:Machine Learning Versus Semiempirical Models[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024, 17: 2 256-2 271

[9] Im J, Park S, Rhee J, et al. Downscaling of AMSR-e Soil Moisture with MODIS Products Using Machine Learning Approaches[J]. Environmental Earth Sciences, 2016, 75 (15):1-19

[10] Long D, Bai L, Yan L, et al. Generation of Spatially Complete and Daily Continuous Surface Soil Moisture of High Spatial Resolution[J]. Remote Sensing of Environment, 2019,233:111 364

[11] 王雨诗,闵馨童,王成,等. 基于EnKF的湘江流域多源遥感土壤水分数据分析[J]. 中国水土保持科学(中英文),2022, 20(2):40-48

[12] Zhu Y, Zhang L, Li F, et al. Comparison of Data Fusion

Methods in Fusing Satellite Products and Model Simulations for Estimating Soil Moisture on Semi-arid Grasslands[J]. Remote Sensing,2023,15(5):3 789

[13] 曹炯玮,魏加华. 集合卡尔曼滤波法在土壤湿度遥感数据融合的研究[J]. 青海师范大学学报(自然科学版),2019,35(1):72-76

[14] Draper C S,Reichle R H,De L G J M,et al. Assimilation of Passive and Active Microwave Soil Moisture Retrievals[J]. Geophysical Research Letters,2012,39:L04401

[15] Lievens H,Martens B,Verhoest N E C,et al. Assimilation of Global Radar Backscatter and Radiometer Brightness Temperature Observations to Improve Soil Moisture and Land Evaporation Estimates[J]. Remote Sensing of Environment,2017,189:194-210

[16] Loizu J,Massari C,Álvarez-mozos J,et al. On the Assimilation Set-up of ASCAT Soil Moisture Data for Improving Streamflow Catchment Simulation[J]. Advances in Water Resources,2018,111:86-104

[17] 王雨诗. 基于SVM-EnKF和多源遥感土壤水分的湘江流域骤旱预测研究[D]. 南京:东南大学,2021

[18] 姚苏红,朱仲元,张圣微,等. 基于SWAT模型的内蒙古闪电河流域径流模拟研究[J]. 干旱区资源与环境,2013,27(1):175-180

[19] Mao T,Shang G W,Li Q,et al. A Spatial Downscaling Method for Remote Sensing Soil Moisture Based on Random Forest Considering Soil Moisture Memory and Mass Conservation[J]. Remote Sensing,2022,14(16):3 858-3 868

[20] Chen Q,Zeng J,Cui C,et al. Soil Moisture Retrieval from SMAP:A Validation and Error Analysis Study Using Ground-based Observations over the Little Washita Watershed[J]. IEEE Transactions on Geoscience and Remote Sensing,2018,56(3):1 394-1 408

[21] Karami A,Moradi H R,Mousivsnd A,et al. Using Ensemble Learning to Take Advantage of High-resolution Radar Backscatter in Conjunction with Surface Features to Disaggregate SMAP Soil Moisture Product[J]. International Journal of Remote Sensing,2022,43(3):894-914

[22] Hu F,Wei Z,Zhang W,et al. A Spatial Downscaling Method for SMAP Soil Moisture Through Visible and Shortwave-infrared Remote Sensing Data[J]. Journal of Hydrology,2020,590:125 360

[23] Colliander A,Jackson T J,Bindlish R,et al. Validation of SMAP Surface Soil Moisture Products with Core Validation Sites[J]. Remote Sensing of Environment,2017,191:215-231

[24] Abowarda A S,Bai L,Zhang C,et al. Generating Surface Soil Moisture at 30 m Spatial Resolution Using Both Data Fusion and Machine Learning Toward Better Water Resources Management at the Field Scale[J]. Remote Sensing of Environment,2021,255:112 301

[25] 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

[26] Vandeventer A P,Ward A D,Gowda P H,et al. Using Thematic Mapper Data to Identify Contrasting Soil Plains and Tillage Practices[J]. Photogrammetric Engineering and Remote Sensing,1997,63(1):87-93

[27] 杨金明,刘志辉. Sentinel-1卫星数据产品应用探讨[J]. 地理空间信息,2016,14(12):18-20

[28] 焦月正,张安兵,王贺封,等. 基于Sentinel-1A数据的土壤水分反演[J]. 测绘与空间地理信息,2022,45(2):40-44

[29] Bond-lamberty B,Hengl T,Mendes D J J,et al. SoilGrids 250 m:Global Gridded Soil Information Based on Machine Learning[J]. Plos One,2017,12(2):e0169 748

[30] Crow W T,Berg A A,Cosh M H,et al. Upscaling Sparse Ground-based Soil Moisture Observations for the Validation of Coarse-resolution Satellite Soil Moisture Products[J]. Reviews of Geophysics,2012,50(2):1-20

[31] Peng S,Ding Y,Liu W,et al. 1 km Monthly Temperature and Precipitation Dataset for China from 1901 to 2017[J]. Earth System Science Data,2019,11(4):1 931-1 946

[32] 赵天杰,姬大彬,蒋玲梅,等. (2023) 闪电河流域土壤温湿度无线传感器网络观测数据集(2020) [DS]. 国家青藏高原科学数据中心,2023

[33] Zheng J,Zhao T,Lu H,et al. Assessment of 24 Soil Moisture Datasets Using a New in Situ Network in the Shandian River Basin of China[J]. Remote Sensing of Environment,2022,271:112 891

[34] Breiman L. Random Forests[J]. Machine Learning,2001,45(1):5-32

[35] Liu Z,Min M,Li J,et al. Local Severe Storm Tracking and Warning in Pre-convection Stage from the New Generation Geostationary Weather Satellite Measurements[J]. Remote Sensing,2019,11(4):383-389

[36] Evensen G. Sequential Data Assimilation with a Nonlinear Quasi-geostrophic Model Using Monte-carlo Methods to Forecast Error Statistics[J]. Journal of Geophysical Research Atmospheres,1994,99(C5):10 143-10 162

[37] 师春香,谢正辉,钱辉,等. 基于卫星遥感资料的中国区域土壤湿度EnKF数据同化[J]. 中国科学:地球科学,2011,41(3):375-385

[38] Geisser S. A Predictive Approach to the Random Effect Model[J]. Biometrika,1974,61(1):101-107

(下转第14页)