

基于不同插值方法的 遥感土壤湿度数据重建精度比较

胡若轩¹, 郑丽虹², 李涛¹, 刘睿璇¹

(1. 南京市滁河河道管理处, 江苏南京 210044; 2. 江苏省水利科学研究院, 江苏南京 210017)

摘要: 为重构缺测的遥感土壤湿度信息, 基于气象、地形、植被、土壤等数据, 采用统计空间插值(普通克里金法)、线性回归模型(多元线性回归法)、机器学习模型(人工神经网络、随机森林法)不同类型方法, 模拟构建了多套数据模型。结合其他土壤湿度产品, 从统计特征、空间分布规律等方面评估不同方法对重建精度的影响。结果显示: 在数据质量优、缺测比例低的情形下, 普通克里金法优势显著; 而数据缺测比例高、缺测空间分布不均时, 多元线性回归法和机器学习模型精度更高。

关键词: 土壤湿度; 插值法; 神经网络; 线性回归

中图分类号: [TV123] **文献标识码:** B **文章编号:** 1007-7839(2025)11-0034-0005

Comparison of reconstruction accuracy for remote sensing soil moisture data based on different interpolation methods

HU Ruoxuan¹, ZHENG Lihong², LI Tao¹, LIURuixuan¹

(1. Nanjing Chuhe River Channel Management Office, Nanjing 210044, China;

2. Jiangsu Hydraulic Research Institute, Nanjing 210017, China)

Abstract: To reconstruct the missing remote sensing soil moisture information, multiple sets of data models were simulated and constructed using different types of methods—including statistical spatial interpolation (ordinary Kriging), linear regression models (multiple linear regression), and machine learning models (artificial neural networks, random forest)—based on data such as meteorology, topography, vegetation, and soil. Combined with other soil moisture products, the impact of different methods on reconstruction accuracy was evaluated in terms of statistical characteristics and spatial distribution patterns. The results show that ordinary Kriging has significant advantages when data quality is high and the missing data ratio is low; however, when the missing data ratio is high and the spatial distribution of missing data is uneven, multiple linear regression and machine learning models achieve higher accuracy.

Key words: soil moisture; interpolation method; neural network; linear regression

收稿日期: 2025-08-04

作者简介: 胡若轩(1997—), 男, 硕士, 研究方向为干旱监测与评估。15094302022。E-mail: 779730312@qq.com

通信作者: 郑丽虹(1996—), 女, 助理工程师, 研究方向为干旱监测与评估。E-mail: rainbow0428@163.com

1 概述

土壤湿度亦称土壤含水量、土壤水分,土壤是地球表层系统的核心要素,参与流域水文循环过程和地表能量循环过程^[1-4]。土壤湿度是陆地过程模式中不可或缺的关键参量,被广泛应用于水资源管理^[5-7]、水旱灾害监测预警^[8-9]、农业生产^[10]、天气预报^[11]和气候变化^[12]等研究中。

微波遥感是监测大尺度土壤湿度的主要方法,其观测覆盖范围大,持续时间长,能够不分昼夜实时、动态地监测地表变化,捕捉土壤湿度空间异质性信息。近年来,卫星遥感土壤湿度反演技术快速发展,遥感土壤湿度数据产品不断丰富,相继研发出多种高时空分辨率的全球卫星遥感土壤湿度产品。然而,受云层大气^[13]、地表植物覆盖、自身卫星轨道和传感器限制^[14]等因素影响,遥感观测可能存在条带缺失^[15]、数据缺测^[16]、时空序列不完整^[17]等问题,导致遥感数据包含大量缺失信息,限制了土壤湿度在水文、水环境等研究领域的应用。

重新构建遥感土壤湿度数据中缺失的信息是近年来的研究热点,旨在消除数据缺失、数据质量低带来的不利影响,插补缺失监测值,为相关研究提供时空序列完整的基础数据。主要通过空间插值模型,结合气象数据、地形地貌、植被信息、土壤质地等土壤湿度的影响因素,提高其模拟精度^[18]。常用的土壤湿度插值方法主要有传统的线性回归模型、统计空间插值法^[19]和机器学习算法^[20],包括多元线性回归法、反距离权重法、普通克里金插值法、回归克里金法、径向基神经网络、人工神经网络、随机森林法等。

2 数据分析

2.1 遥感土壤湿度数据

长时间序列的主被动结合微波遥感土壤湿度数据产品(ESA CCI)的时间范围是从1978年至今,空间分辨率为25 km,时间分辨率为1 d。该数据集中包含主动、被动、主被动结合3种不同的产品类型。有关研究表明,主动土壤湿度数据适用于植被密集区,被动遥感数据适用于干旱、半干旱区域,主被动结合数据产品更适合大尺度区域的应用中。本研究选取主被动结合数据,同时仅使用表层土壤湿度。

2.2 其他土壤湿度信息

(1) ERA

ERA-Interim 再分析数据集(ECMWF Re-

Analysis)是基于站点观测数据和四维变分数据同化方法构建的第四代全球大气产品。该数据集时间跨度为1978—2019年,空间分辨率为78 km×78 km,时间分辨率为6 h,包含0~7 cm、7~28 cm、28~100 cm、100~289 cm共4层不同深度的土壤湿度,本文仅使用表层(0~7 cm)土壤湿度。目前,ERA-Interim 土壤湿度数据集的可靠性已得到普遍认可。

ERA与ESA CCI数据集的相关系数分布见图1,两者相关系数均值为0.719。其中,0.7~0.8区间占比最高(52.6%),大于0.6区间占90.3%,而小于0.4区间仅占1.2%,表明两者存在显著相关性。

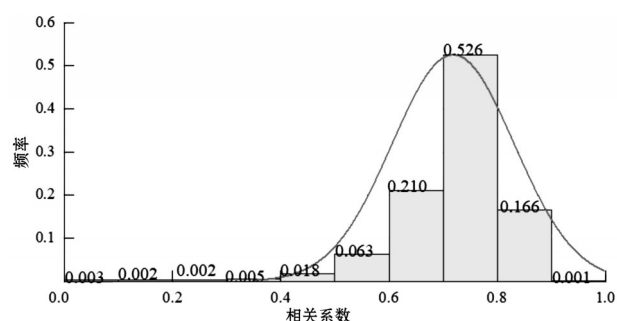


图1 ERA与ESA CCI数据集的相关系数分布

(2) ISMN

ISMN (International Soil Moisture Network) 国际土壤湿度网是规模较大的站点土壤湿度共享数据库,时间分辨率为1 h。该数据集按土壤深度划分为11个层级,本研究选取0~5 cm表层土壤湿度数据用于分析。经统计验证,ISMN与ESA CCI的相关系数为0.499,均方根误差为0.088,表明两者存在一定程度的相关性。

2.3 相关数据信息

(1) 气象数据

本文采用的降水和气温数据为气象站实测日气象数据,时间长度为1982—2019年,利用反距离加权法将其插值为 $0.25^{\circ} \times 0.25^{\circ}$ 。

(2) 植被信息

研究采用的植被指数(Normalized Difference Vegetation Index, NDVI)由两部分构成。1981—2015年数据源于第三代(NDVI3g)全球覆盖产品数据集,其数据更新至2015年,时间分辨率为15 d,空间分辨率为 0.083° 。2016—2019年植被信息数据和叶面积指数源于相关资源环境科学数据中心。其中,1981—1999年数据由反演而来,时间分辨率为15 d,2000年至今的数据是基于MODIS地表反射率

产品以 GLOBCARBON LAI 算法反演得到,时间分辨率为 8 d,空间分辨率均为 8 km。

(3) 土壤质地

土壤质地数据库分为 2 层,顶层土壤深度为 0~30 cm,底层深度为 30~100 cm,根据土壤分离物(砂土、粉砂土、黏土)的相对比例将土壤类型分为 13 类。

(4) 地形信息

采用了 5 种地理数据:经度、纬度、数字高程模型(DEM)、坡度和坡向。DEM 数据来源于 Shuttle Radar 地形任务数字高程模型(SRTM),空间分辨率为 0.0833°,基于 DEM 数据在地理空间信息系统(ArcGIS)平台上导出研究区域各网格的坡度和坡向。

3 研究方法

3.1 普通克里金法

克里金法是一种考虑样本点空间位置关系,对邻近已知样本加权平均来估计其他未知点,并使估计值与真实值的数学期望相同且方差最小的空间局部插值方法,通过建立变异函数来表征变量的主要结构特征,对未知点进行最佳线性无偏估计。常用的克里金法主要有:简单克里金法(simple Kriging)、普通克里金法(ordinary Kriging, OK)、泛克里金法(universal Kriging)和协同克里金法(co-Kriging)等,其中普通克里金法应用最为广泛。

3.2 多元线性回归法

多元线性回归法(multiple linear regression, MLR)是一种通过多个自变量的线性回归模型来解释因变量变化的统计方法,该方法简单便捷、应用较广。在多元线性回归模型中,共引入 26 个自变量,包含日降水量、3 d 累积降水量、5 d 累积降水量、日气温、3 d 平均气温、5 d 平均气温、NDVI、LAI、纬度、经度、高程、坡度、坡向以及 FAO90 土壤分类系统中 13 类土壤类型。

3.3 人工神经网络

人工神经网络(artificial neural network, ANN)是模拟人脑神经网络信息处理机制构建的人工智能算法,由大量神经元通过权重相互连接形成复杂神经网络。神经元作为基本功能单位,类似于脑神经细胞,可实现信息的计算、短暂存储与传递,连接权重是神经元间的“纽带”,代表着输入特征的重要性。ANN 是多层感知器模型,由输入层、一个或多个隐含层及输出层组成。理论上,只要神经元数量足够,带有一个隐含层的三层神经网络足以逼近任

意一个复杂的非线性函数。

3.4 随机森林法

随机森林(random forest, RF)以决策树为基学习器,是一种功能强大的集合学习算法,结合了 bagging 和 random subspace method 算法。为提升预测精度并抑制过拟合,RF 通过不同的数据集子样本及初始变量独立构建若干棵决策树。

4 结果分析与讨论

4.1 不同插值方法的误差分析

由不同插值方法与 ESA CCI、ERA、ISMN 原始土壤湿度数据的散点关系可知,基于 ESA CCI 数据的插值结果具有高度的一致性,其相关系数(CC)介于 0.8068(多元线性回归法)至 0.9694(随机森林法)区间,均方根误差(RSME)介于 0.0165(随机森林法)至 0.0495(多元线性回归法)区间。随机森林法精度最高,多元线性回归法精度较低且插值结果在极值处均略微偏大。

对于 ERA 数据集,随机森林法仍然拟合效果最好,而普通克里金法在土壤湿度数据处于极小值时拟合较差,多元线性回归法则在极大值时拟合较差。就 ISMN 数据而言,人工神经网络表现较差,模拟结果最离散,其模拟结果中出现少量极低值,相关系数小于 0.50,效果欠佳。相比之下,多元线性回归法的表现超出预期,其相关系数达到 0.6059,均方根误差为 0.0460,拟合效果较好。

为分析不同插值算法空间插值效果,选取数据缺失程度不同的历史时期:1990 年、2002 年、2019 年 5 月上旬(数据缺失率 65.41%、39.93%、18.68%),对比不同算法插值结果与原 ESA CCI 土壤湿度数据的空间分布。当数据缺失率高且空间距离大时,基于空间自相关法的普通克里金法无法有效插值。多元线性回归法插值结果较保守,低值处偏大,高值处偏小,在高原地区和沿海地区表现尤为明显,且随数据质量提升效果有所改善。两种机器学习算法插值结果表现最佳:数据缺失率高时,人工神经网络细节刻画更优(如长江流域与原始数据相似度高);而数据缺失率低时,随机森林算法能精准捕捉到细微变化,表明其对数据质量依赖性更强,当数据充足时模拟结果更优。

由不同插值方法与 ESA CCI 和 ERA 数据的 CC、RMSE 值(图 2)可以看出,不同插值方法的精度差异明显。随机森林法表现最佳,CC 均值大于 0.95,RSME 均值为 0.15;普通克里金法的 CC 值高达

0.95~0.99, RSME 值为 0.01~0.02, 表明两者相关性较好, 模拟精度较高; 线性回归法表现相对欠佳, 但其 CC 均值大于 0.8, RSNE 均值为 0.045, 结果尚可满足预期。同时, 各插值结果与 ERA 的 CC、RMSE 值接近 ERA 与 ESA CCI 的统计值, 表明其精度均较高。

4.2 不同数据缺失情况的模拟表现

进一步分析不同程度数据缺失情况的模拟表现, 如图 3 所示, 选取高原地区(Q1, 数据质量差, 缺失率 81.3%)与平原地区(Q2, 数据质量好, 缺失率 44.7%)为例进行对比。

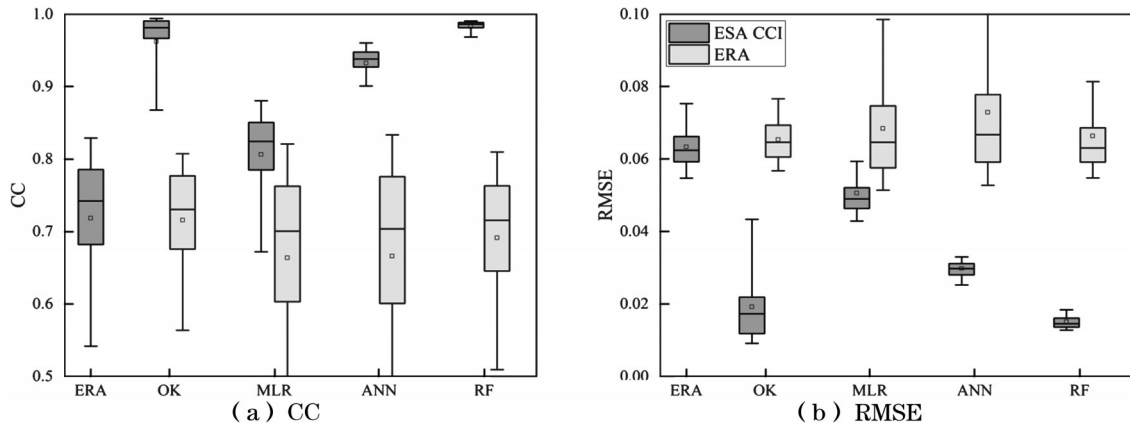


图 2 不同插值方法与 ESA CCI 和 ERA 数据的箱线图

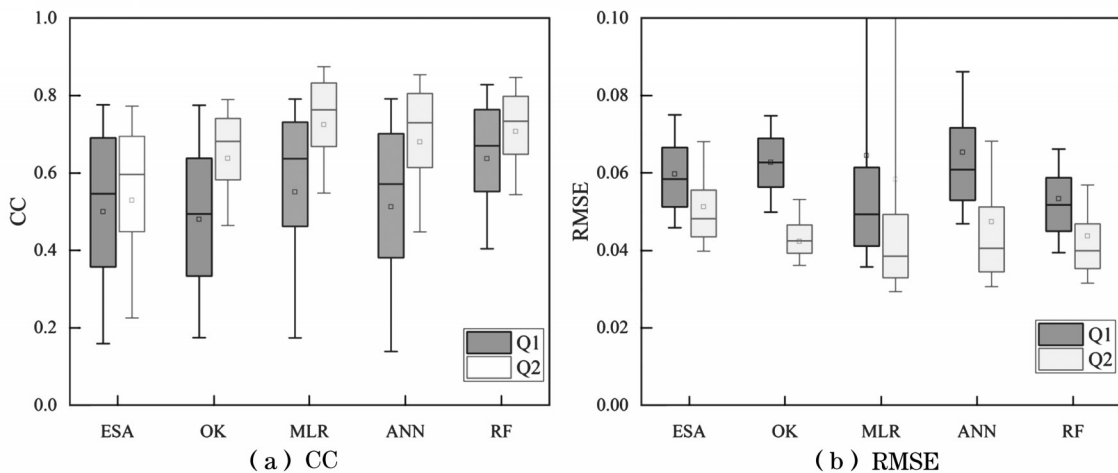


图 3 不同插值方法在 Q1、Q2 处数据的箱线图

图 3(a) 显示, 数据质量提升后, 各种空间插值方法模拟精度均显著提高, 相关系数约提升了 0.2~0.3, 表明训练数据充足可显著改善模拟效果较差的数据精度。整体而言, 随机森林法模拟效果较好, $CC \geq 0.4$, $RSME \leq 0.07$ 。多元线性回归法模拟结果较好, 但图 3(b) 显示其存在 $RSME > 0.10$ 情形, 表明部分结果偏差较大。

5 结 语

综上所述, 机器学习插值结果精度最高, 相关系数为 0.90~0.95, 均方根误差为 0.03~0.04。其中, 随机森林算法精度略优于人工神经网络, 普通克里

金法仅在空间距离小且数据质量高时优势显著, 数据缺失率高时难以有效填补空缺, 而多元线性回归法计算便捷、效率高、精度较好。数据质量对拟合结果的统计特征有显著影响, 数据缺失率由 81.3% 降低至 44.7% 时, 相关系数提升 0.1~0.2, 均方根误差降低 0.01。而对于部分高原等数据稀缺地区, 多元线性回归法和随机森林算法表现突出, 5—9 月非冻融期与 ISMN 站点实测数据拟合度相对较高。

参考文献:

[1] ILLSTON B G, BASARA J B, FIEBRICH C A, et al. Mesoscale monitoring of soil moisture across a statewide

- network [J]. *Journal of Atmospheric and Oceanic Technology*, 2008, 25(2): 167–182.
- [2] KERR Y H. Soil moisture from space: Where are we [J]. *Hydrogeology Journal*, 2007, 15(1): 117–120.
- [3] WANG C, FU B, ZHANG L, et al. Soil moisture–plant interactions: an ecohydrological review [J]. *Journal of Soils and Sediments*, 2019, 19(1): 1–9.
- [4] 何涯舟, 张珂, 晁丽君, 等. 基于多源遥感土壤湿度与模型数据同化的流域径流模拟 [J]. *水资源保护*, 2023, 39(2): 145–151, 189.
- [5] DOBRIYAL P, QURESHI A, BADOLA R, et al. A review of the methods available for estimating soil moisture and its implications for water resource management [J]. *Journal of Hydrology*, 2012(458): 110–117.
- [6] MAO K, SHI J, TANG H, et al. A neural network technique for separating land surface emissivity and temperature from ASTER imagery [J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2007, 46(1): 200–208.
- [7] ROBINSON D A, CAMPBELL C S, HOPMANS J W, et al. Soil moisture measurement for ecological and hydrological watershed–scale observatories: A review [J]. *Vadose Zone Journal*, 2008, 7(1): 358–389.
- [8] AGHAKOUCHAK A, FARAHMAND A, MELYON F S, et al. Remote sensing of drought: Progress, challenges and opportunities [J]. *Reviews of Geophysics*, 2015, 53(2): 452–480.
- [9] BINDLISH R, CROW W T, JACKSON T J. Role of Passive Microwave Remote Sensing in Improving Flood Forecasts [J]. *IEEE Geoscience & Remote Sensing Letters*, 2009, 6(1): 0–116.
- [10] CHEN C F, SON N T, CHANG L Y, et al. Monitoring of soil moisture variability in relation to rice cropping systems in the Vietnamese Mekong Delta using MODIS data [J]. *Applied Geography*, 2011, 31(2): 463–475.
- [11] DAI A, TRENBERTH K E, QIAN T. A global dataset of Palmer Drought Severity Index for 1870–2002: Relationship with soil moisture and effects of surface warming [J]. *Journal of Hydrometeorology*, 2004, 5(6): 1117–1130.
- [12] ANDERSON M C, NORMAN J M, MECIKALSKI J R, et al. A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation [J]. *Journal of Geophysical Research: Atmospheres*, 2007, 112(10): 73–84.
- [13] SHEN H, LI X, CHENG Q, et al. Missing information reconstruction of remote sensing data: A technical review [J]. *IEEE Geoscience and Remote Sensing Magazine*, 2015, 3(3): 61–85.
- [14] LIN C H, LAI K H, CHEN Z B, et al. Patch–based information reconstruction of cloud–contaminated multi–temporal images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2013, 52(1): 163–174.
- [15] MELGANI F. Contextual reconstruction of cloud–contaminated multitemporal multispectral images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2006, 44(2): 442–455.
- [16] XIAO Z, JIANG L, ZHU Z, et al. Spatially and temporally complete satellite soil moisture data based on a data assimilation method [J]. *Remote Sensing*, 2016, 8(1): 49.
- [17] WANG G, GARCIA D, LIU Y, et al. A three–dimensional gap filling method for large geophysical datasets: Application to global satellite soil moisture observations [J]. *Environmental Modelling & Software*, 2012(30): 139–142.
- [18] 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]. *Science of the Total Environment*, 2020(724): 138319.
- [19] ZHANG L Q, LIU Y, REN L L, et al. Reconstruction of ESA CCI satellite–derived soil moisture using an artificial neural network technology [J]. *Science of the Total Environment*, 2021(782): 146602.
- [20] 刘娣, 孙佳倩, 余钟波. 基于机器学习模型的多层土壤湿度反演 [J]. *河海大学学报(自然科学版)*, 2024, 52(3): 7–14.