

DOI:10.3876/j.issn.1000-1980.2024.03.002

# 基于机器学习模型的多层土壤湿度反演

刘 娣<sup>1,2,3</sup>, 孙佳倩<sup>1,2,3</sup>, 余钟波<sup>1,2,3,4</sup>

(1. 河海大学水灾害防御全国重点实验室, 江苏 南京 210098; 2. 河海大学水文水资源学院, 江苏 南京 210098;  
3. 河海大学全球变化与水循环国际合作联合实验室, 江苏 南京 210098;  
4. 河海大学长江保护与绿色发展研究院, 江苏 南京 210098)

**摘要:** 为了获取深层土壤湿度缺测值, 采用支持向量机、BP 神经网络和随机森林 3 种机器学习算法, 在表层至深层土壤中利用主成分分析法选择与土壤湿度相关性显著的气象因子作为输入数据, 建立多层土壤湿度反演模型反演了不同深度的土壤湿度。结果表明: 随机森林模型模拟结果更加稳定, 反演效果更佳; 受气象因子驱动的影响, 3 种机器学习模型对地表 0~10 cm 深度内土壤湿度的反演效果更佳, 对深层土壤湿度的反演效果随着深度增加而变差; 增加表层土壤湿度及不同深度土壤湿度作为驱动因子可以有效提高机器学习模型对深层土壤湿度的反演能力。

**关键词:** 土壤湿度; 机器学习; 支持向量机; BP 神经网络; 随机森林; 主成分分析法

中图分类号: S152.7<sup>+1</sup>

文献标志码: A

文章编号: 1000-1980(2024)03-0007-08

## Multi-layer soil moisture inversion based on machine learning models

LIU Di<sup>1,2,3</sup>, SUN Jiaqian<sup>1,2,3</sup>, YU Zhongbo<sup>1,2,3,4</sup>

(1. The National Key Laboratory of Water Disaster Prevention, Hohai University, Nanjing 210098, China;

2. College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China;

3. Joint International Research Laboratory of Global Change and Water Cycle, Hohai University, Nanjing 210098, China;

4. Yangtze Institute for Conservation and Development, Hohai University, Nanjing 210098, China)

**Abstract:** In order to obtain soil moisture data at deep soil layers and to complete the missing soil moisture measurements at different depths, machine learning technique is used to perform the soil moisture inversion at multi-layers from surface to root zone. Three machine learning algorithms, support vector machine (SVM), back propagation (BP) neural network and random forest (RF), are applied to construct the inversion models for the train and inversion of soil moisture data at different soil layers, and the meteorological factors, which have high correlation coefficient with the soil moisture, are selected as the input factors based on the principal component analysis (PCA) method in each layer of soil. The major findings are: the simulation results of random forest are more stable and the inversion effect is the best; due to the impact of meteorological variables, the three machine learning models perform best for the inversion of surface soil moisture at the surface soil layer within 0 to 10 cm depth; however, the inversion effect for the soil moisture at the deep zone is gradually reduced with the depth of soil layer; the addition of surface soil moisture and soil temperature at different soil layers as driven factors could improve the inversion capacity of machine learning model for the soil moisture at the deep zone.

**Key words:** soil moisture; machine learning; support vector machine; back propagation neural network; random forest; principal component analysis

土壤湿度是研究气象变化中的重要一环, 其变化过程与气象因子的变化有着密切的关系。气象因子之间的变化并不一致, 是非线性多样化的, 而气象因子的变化对土壤湿度的变化有着决定性的作用<sup>[1]</sup>。因此, 在对土壤湿度建模分析时, 选择合适的气象因子作为输入数据是极其重要的。运用主成分分析法 (principal component analysis, PCA) 选择对土壤湿度影响显著的气象因子进行建模与计算, 能够大幅度提高计算的效率。解雪等<sup>[2]</sup>利用主成分分析法和粒子群算法优化的支持向量机 (support vector machine) 建立了冻融土壤蒸发量的预报模型, 对冻融土壤蒸发量进行预报, 结果证明主成分分析法能够有效消除各影响因素间的冗余

**基金项目:** 国家自然科学基金项目 (U2240217); 水灾害防御全国重点实验室基本科研业务费自主研究项目 (520004412, 521013122)

**作者简介:** 刘娣 (1986—), 女, 副研究员, 博士, 主要从事水文循环演变机理研究。E-mail: liudi@hhu.edu.cn

**通信作者:** 余钟波 (1964—), 男, 教授, 博士, 主要从事水文水资源研究。E-mail: zyu@hhu.edu.cn

**引用本文:** 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.

LIU Di, SUN Jiaqian, YU Zhongbo. Multi-layer soil moisture inversion based on machine learning models[J]. Journal of Hohai University (Natural Sciences), 2024, 52(3): 7-14.

与相关性,同时提高模型预报的准确度。

土壤湿度也会对土壤表面的反射率以及土壤热容量产生影响,进而通过潜热和感热传导的方式对气候变化产生影响<sup>[3]</sup>。陈钢等<sup>[4]</sup>通过计算土壤含水量并对降雨径流事件进行分析,证明了表层土壤水分动态变化对降雨极为敏感,随着土壤深度的增加,土壤水分随降水量增加的增幅逐渐减少,响应时间也相应延长;何涯舟等<sup>[5]</sup>对土壤蓄水量和径流评估结果表明,降雨会优先补给土壤缺水,对土壤含水量产生直接的影响;吴志勇等<sup>[6]</sup>使用 VIC-EPIC 水文作物耦合模型,利用气温、降水和风速等驱动数据模拟土壤含水量,结果显示 0~40 cm 土层模拟土壤含水量与实测值的相关性最好,属于强相关,表明气象因子能作为反演土壤湿度的驱动要素。

随着机器学习技术的发展,机器学习模型被广泛应用于水文领域的科学研究。机器学习可以利用有限甚至缺失的数据建立相关模型并进行训练,进而进行反演与预测。李彬楠等<sup>[7]</sup>利用支持向量机模型对土壤水分特征曲线进行预测,证实支持向量机模型适用于小样本和非线性问题的预测。王彤彤等<sup>[8]</sup>利用 BP 神经网络和支持向量机算法建立了施加生物炭的土壤水分预测模型,为半干旱地区生物炭还田土壤水分的模拟预测以及管理提供了理论支撑;韩慧敏等<sup>[9]</sup>利用梯度提升机、BP 神经网络(back propagation neural networks)、随机森林(random forest)及 Stacking 集成学习算法对土壤湿度进行降尺度获取高时空分辨率的土壤湿度估算值,并通过对比分析得出 Stacking 集成学习算法估算的土壤湿度精度更高;屈忠义等<sup>[10]</sup>通过建立 BP 神经网络预测模型对多年地下水埋深变化进行模拟,预测了未来地下水水位下降的趋势;沈润平等<sup>[11]</sup>利用随机森林算法综合多因子构建遥感干旱监测模型,并证实该模型适用于监测研究区域的旱情。目前,国内外直接利用气象资料对土壤湿度数据进行反演的研究还较少,本文旨在利用气温、相对湿度等气象资料,通过机器学习算法建立气象因子与土壤湿度之间的反演模型,从而实现对不同深度土壤湿度数据的模拟。

## 1 研究区域与数据

本文以位于美国南卡罗来纳州的 Blackville 和 McClellanville 两个气象站为研究对象<sup>[12]</sup>。Blackville 站位于北纬 33°21'18"、西经 81°19'40";McClellanville 站位于北纬 33°5'20"、西经 79°28'2"。两个气象站内均有基于实时传感器的土壤湿度监测站和 NOAA 美国气候参考网络站。

选择两个气象站 2010 年 1 月 1 日至 2013 年 12 月 31 日的观测数据进行建模研究。所用数据包括:①每日总太阳能( $E_{\text{总}}$ );②每日平均降水量( $P$ );③每日最高、最低以及平均红外太阳能( $E_{\text{max}}$ 、 $E_{\text{min}}$ 、 $E_{\text{avg}}$ );④每日最大、最小以及平均相对湿度( $H_{\text{Rmax}}$ 、 $H_{\text{Rmin}}$ 、 $H_{\text{Ravg}}$ );⑤每日最高、最低、加权平均以及平均气温( $T_{\text{max}}$ 、 $T_{\text{min}}$ 、 $T_{\text{mean}}$ 、 $T_{\text{avg}}$ );⑥5、10、20、50、100 cm 土层深度处每日平均土壤温度( $T_{\text{S5}}$ 、 $T_{\text{S10}}$ 、 $T_{\text{S20}}$ 、 $T_{\text{S50}}$ 、 $T_{\text{S100}}$ );⑦5、10、20、50、100 cm 土层深度处每日平均土壤湿度( $M_{\text{S5}}$ 、 $M_{\text{S10}}$ 、 $M_{\text{S20}}$ 、 $M_{\text{S50}}$ 、 $M_{\text{S100}}$ )。其中,2012 年为 366 天,其余各年为 365 天,共 1461 天的有效数据,选择其中 2010 年 1 月 1 日至 2012 年 12 月 31 日(1096 天)的数据作为训练集,用以建立模型并进行训练,2013 年 1 月 1 日至 2013 年 12 月 31 日(365 天)的数据作为测试集,用以检验模型的反演精度。

## 2 研究方法

### 2.1 主成分分析法

主成分分析也称主分量分析或者矩阵数据分析,是 1901 年由 Karl 提出的多元统计分析方法。主成分分析法本质上是利用相关性提取重要数据来降低问题复杂性和难度<sup>[13]</sup>,其基本思想是利用较少的变量对原始数据中大部分的变异进行解释,并且通过线性变化从一组相关的数据中产生互不相关的变量。这些新变量就是筛选出来的主成分因子,它们能够基本解释原始数据的变化。其基本步骤为:①将原始数据标准化;②建立相关系数矩阵,计算特征向量以及对应的特征值;③确定主成分因子的个数,建立主成分表达式。

### 2.2 机器学习

#### 2.2.1 支持向量机

支持向量机是 Cortes 等<sup>[14]</sup>在 1995 年基于结构风险最小化原理提出的机器学习算法。支持向量机算法分为回归和分类两种,本文选择支持向量机回归算法,回归算法的基本思想是寻找一条最优拟合曲线,使得曲线在误差范围内能够尽可能多地包含数据点。支持向量机回归算法求解原理是通过添加偏差构建误差间

隔带,只要预测值落入此间隔带内,则认为预测结果是正确的。核函数的选择会对支持向量机的性能产生直接影响,而径向基函数通过定义一个比线性核函数或多项式核函数更大的函数空间,稳定性更好<sup>[15]</sup>。因此,本文选用径向基函数作为核函数构建支持向量机算法。

2.2.2 BP 神经网络

BP 神经网络算法在泛化及非线性映射方面的能力出众,由输入层、隐藏层和输出层三部分组成,是一种通过误差逆向传播算法来进行训练的多层前馈神经网络<sup>[16]</sup>。

构建 BP 神经网络模型的基本步骤为:①正向传播。输入样本数据,隐藏层对样本数据进行非线性运算处理,通过输出层将结果输出并与已知结果进行对比,若输出结果与已知结果的误差在设置阈值范围内,则结束训练;若误差在设置阈值范围之外,则进行误差反向传播。②反向传播。将输入数据与输出数据之间的误差进行反向传播,经由隐藏层对权值和阈值进行多次优选,不断对神经网络进行更新。当输出结果的误差小于设定误差或者模型的训练次数达到设定的最大训练次数时,模型停止训练,输出预测结果。

2.2.3 随机森林

随机森林是一种基于决策树的集成算法,具有效率较高、计算成本低的优点<sup>[17]</sup>。随机森林每棵树使用的样本和特征都不相同,训练出的结果也不相同,能够大大降低在运行时出现过拟合的概率<sup>[18]</sup>。

构建随机森林模型的基本步骤为:①从样本中随机有放回的抽取  $n$  个样本生成  $n$  个训练集,没有被选中的样本自动作为测试集;②每一个训练集的子集都是一棵决策树,每一棵决策树对样本特征都进行随机有放回的抽取并进行独立的训练;③取所有决策树预测结果的平均值作为预测值。

2.2.4 机器学习模型构建

3 种机器学习模型的建模流程如图 1 所示。主要步骤包括:①获取数据;②对数据进行归一化处理,将归一化后的数据划分为训练集和测试集;③使用支持向量机算法对模型进行初始化,在选择出最优参数后对训练集数据进行训练,利用训练后的支持向量机模型对测试集数据进行预测;④使用 BP 神经网络算法建立神经网络模型,在设置模型中的网络参数后对训练集数据进行训练,利用训练后的 BP 神经网络模型对测试集数据进行仿真计算;⑤使用随机森林算法构建反演模型时,在确定回归树的数量后,对训练集数据进行回归训练,利用训练后的随机森林模型对测试集数据进行预测;⑥将 3 种机器学习模型的计算结果分别进行反归一化处理,再对数据进行误差计算。

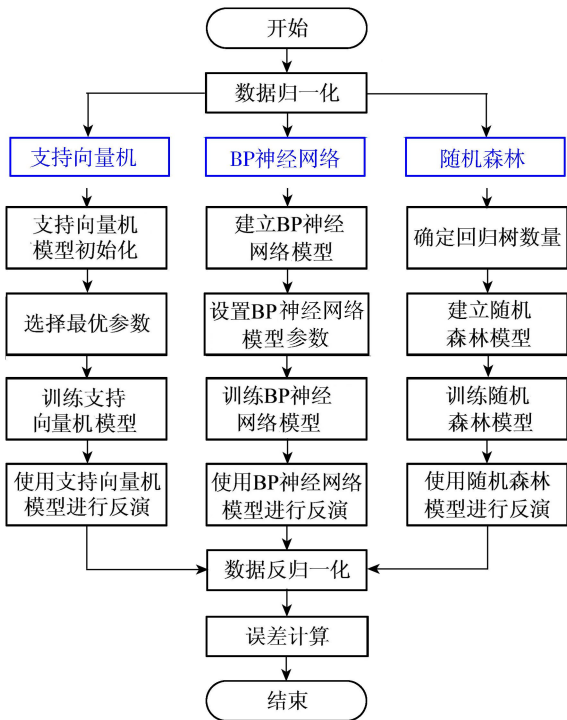


图 1 3 种机器学习模型建模流程

Fig.1 Flow chart of model construction based on three different machine learning models



2.3 评估指标

采用均方根误差(root mean square error, RMSE)、平均绝对误差(mean absolute error, MAE)和相关系数( $R$ )来评估3种机器学习模型对不同深度土壤湿度的反演效果。

3 结果与分析

3.1 主成分分析结果

卫星遥感技术的发展为高精度表层土壤湿度数据的获取提供了条件<sup>[19]</sup>,但深层土壤湿度信息的获取仍是目前的难题。受植被覆盖率以及云层遮盖的影响,卫星遥感土壤湿度产品在大范围的裸土区域内,一般只能反演0~5 cm土层的土壤湿度,与土壤湿度有关的研究也多集中在地表土壤垂直剖面的顶部。与表层土壤垂直剖面相比,深层土壤湿度对全球水循环同样重要,高质量的深层土壤湿度数据对研究气候变化和植被生长具有重要价值。已有研究表明深层土壤湿度跟气象因子(如降雨)之间存在滞后效应<sup>[20]</sup>,因此本文在反演深层土壤湿度时考虑将表层土壤湿度 $M_{ss}$ 作为驱动因子之一,探究表层土壤湿度对各机器学习模型的改进能力。不同深度土壤湿度的主成分因子如表1所示,可见表层土壤湿度 $M_{ss}$ 是10 cm及更深层土壤湿度的主成分因子之一。此外,不同深度土壤湿度也与土壤温度密切相关,这可能是研究站点不同深度土壤温度在时间序列上变化较弱的缘故。在研究时段内,两个气象站不同深度土层的土壤温度时间序列较为一致,偏差较小。研究区域的水汽含量较为丰富,这会在很大程度上影响土壤水分的含量,并且植被覆盖率较低,太阳辐射对土壤湿度有着直接的影响,因此不同深度土壤湿度的主成分因子均有相对湿度以及每日总太阳能。

表1 不同深度土壤湿度的主成分因子  
Table 1 Principal component factors of soil moisture at different depths

| 土层深度/cm | 主成分因子                                                                                                                |
|---------|----------------------------------------------------------------------------------------------------------------------|
| 5       | $T_{max}、T_{min}、P、E_{总}、E_{max}、E_{min}、E_{avg}、H_{Rmax}、H_{Rmin}、H_{Ravg}、T_{S5}、T_{S10}、T_{S20}、T_{S50}、T_{S100}$ |
| 10      | $T_{max}、T_{min}、P、E_{总}、E_{max}、H_{Rmax}、H_{Rmin}、H_{Ravg}、M_{ss}、T_{S5}、T_{S10}、T_{S20}、T_{S50}、T_{S100}$          |
| 20      | $T_{max}、E_{总}、E_{max}、H_{Rmax}、H_{Rmin}、H_{Ravg}、M_{ss}、T_{S5}、T_{S10}、T_{S20}、T_{S50}、T_{S100}$                    |
| 50      | $T_{max}、E_{总}、E_{max}、H_{Rmax}、H_{Rmin}、H_{Ravg}、M_{ss}、T_{S5}、T_{S10}、T_{S20}、T_{S50}、T_{S100}$                    |
| 100     | $T_{max}、E_{总}、E_{max}、H_{Rmax}、H_{Rmin}、H_{Ravg}、M_{ss}、T_{S5}、T_{S10}、T_{S20}、T_{S50}、T_{S100}$                    |

3.2 BP神经网络模型参数调优

本文构建的BP神经网络模型输入层有11~15个数据,输出层有5个数据,训练函数选择“trainlm”,隐藏层的传输函数选择“tansig”,输出层的传输函数选择“purelin”,最大训练次数设置为1 000,学习速率为0.01,训练目标为0.001<sup>[21]</sup>。

因为BP神经网络设置的参数较多,为了降低过拟合发生的概率,同时提高模型的泛化能力,需要结合训练集和测试集的平均相关系数( $\bar{R}$ )以及平均均方根误差(aRMAE)对BP神经网络模型的拟合效果进行综合判断。隐藏层层数会对神经网络的模拟结果产生较大影响,当隐藏层层数过多时,模型训练的时间会大幅度增加,训练效率降低,网络的泛化能力下降;当隐藏层层数过少时,不能正确反映前驱值与后续值的相关关系,致使建模不够准确。隐藏层层数可以由以下公式计算得到<sup>[22]</sup>:

$$h = \sqrt{u + v} + c$$

(1)

式中: $h$ 为隐藏层层数; $u$ 为输入层层数; $v$ 为输出层层数; $c$ 为[1,10]之间的任意常数。

计算结果如表2所示,两个气象站隐藏层层数均设置为10层时,BP神经网络模型的反演效果最佳。为了控制变量一致,对两个气象站不同深度土壤湿度进行反演时,将BP神经网络模型隐藏层层数均设置为10层。

3.3 模型驱动要素设计及结果分析

为了研究表层土壤湿度在10、20、50、100 cm土层中的反演效果,本文在构建机器学习模型时,分别设计了3组试验对比不同机器学习模型在有实测表层土壤湿度(试验1)、有反演表层土壤湿度(试验2)和无表层土壤湿度(试验3)驱动下对深层土壤湿度反演能力的影响。此外,由主成分分析可得土壤温度与不同深度土壤湿度密切相关,是重要的影响要素。为了分析土壤温度对不同深度土壤湿度反演精度的影响,在试验1的基础

表2 BP神经网络模型参数计算结果  
Table 2 Calculation results of BP neural network model parameter

| 土层深度/cm | 隐藏层最佳层数      |                  |
|---------|--------------|------------------|
|         | Blackville 站 | McClellanville 站 |
| 5       | 10           | 10               |
| 10      | 10           | 10               |
| 20      | 7            | 9                |
| 50      | 11           | 10               |
| 100     | 9            | 6                |

上设计了试验 4~6。试验 4 分析无土壤温度驱动后各机器学习模型对不同深度土壤湿度的反演能力,试验 5 分析去除试验 1 中 10 cm 及以下土层土壤温度情况下机器学习模型的反演能力,试验 6 分析同时去除表层土壤湿度和各土层的土壤温度驱动对机器学习模型反演能力的影响。试验 1~6 的驱动要素和输出要素如表 3 所示。在不同试验条件下,3 种机器学习模型对两个气象站测试集中不同深度土壤湿度的反演值与观测值的 RMSE、MAE 及  $R$  分别如表 4、表 5 和表 6 所示。

表 3 试验设计

Table 3 Experimental design

| 试验编号 | 驱动要素                              | 输出数据                                                      |
|------|-----------------------------------|-----------------------------------------------------------|
| 1    | 表 1 中所有主成分因子                      | $M_{s5}$ 、 $M_{s10}$ 、 $M_{s20}$ 、 $M_{s50}$ 、 $M_{s100}$ |
| 2    | 用反演 $M_{s5}$ 替代试验 1 中的实测 $M_{s5}$ | $M_{s10}$ 、 $M_{s20}$ 、 $M_{s50}$ 、 $M_{s100}$            |
| 3    | 去除试验 1 中的实测 $M_{s5}$              | $M_{s10}$ 、 $M_{s20}$ 、 $M_{s50}$ 、 $M_{s100}$            |
| 4    | 去除试验 1 中各土层的土壤温度                  | $M_{s5}$ 、 $M_{s10}$ 、 $M_{s20}$ 、 $M_{s50}$ 、 $M_{s100}$ |
| 5    | 去除试验 1 中 10 cm 及以下土层的土壤温度         | $M_{s5}$ 、 $M_{s10}$ 、 $M_{s20}$ 、 $M_{s50}$ 、 $M_{s100}$ |
| 6    | 去除试验 1 中实测 $M_{s5}$ 及各土层的土壤温度     | $M_{s10}$ 、 $M_{s20}$ 、 $M_{s50}$ 、 $M_{s100}$            |

表 4 支持向量机模型对测试集中不同深度土壤湿度的反演评估结果

Table 4 Results of inverse assessment of soil moisture at different depths in the test set by support vector machine models

| 站点 | 土层深度/<br>cm | RMSE/( $\text{m}^3/\text{m}^3$ ) |       |       |       |       |       | MAE/( $\text{m}^3/\text{m}^3$ ) |       |       |       |       |       | $R$   |       |       |       |       |       |
|----|-------------|----------------------------------|-------|-------|-------|-------|-------|---------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|    |             | 试验 1                             | 试验 2  | 试验 3  | 试验 4  | 试验 5  | 试验 6  | 试验 1                            | 试验 2  | 试验 3  | 试验 4  | 试验 5  | 试验 6  | 试验 1  | 试验 2  | 试验 3  | 试验 4  | 试验 5  | 试验 6  |
| B  | 5           | 0.005                            | 0.005 | 0.005 | 0.011 | 0.011 | 0.011 | 0.004                           | 0.004 | 0.004 | 0.009 | 0.010 | 0.009 | 0.992 | 0.992 | 0.992 | 0.982 | 0.993 | 0.982 |
|    | 10          | 0.003                            | 0.033 | 0.006 | 0.029 | 0.017 | 0.005 | 0.003                           | 0.014 | 0.005 | 0.023 | 0.014 | 0.004 | 0.999 | 0.776 | 0.992 | 0.883 | 0.923 | 0.994 |
|    | 20          | 0.030                            | 0.044 | 0.028 | 0.026 | 0.026 | 0.028 | 0.022                           | 0.034 | 0.022 | 0.021 | 0.021 | 0.023 | 0.852 | 0.660 | 0.815 | 0.762 | 0.811 | 0.770 |
|    | 50          | 0.034                            | 0.060 | 0.043 | 0.038 | 0.046 | 0.049 | 0.023                           | 0.048 | 0.035 | 0.031 | 0.038 | 0.042 | 0.685 | 0.578 | 0.643 | 0.664 | 0.630 | 0.649 |
|    | 100         | 0.045                            | 0.056 | 0.048 | 0.047 | 0.047 | 0.044 | 0.039                           | 0.046 | 0.040 | 0.036 | 0.038 | 0.036 | 0.397 | 0.300 | 0.406 | 0.384 | 0.258 | 0.431 |
| M  | 5           | 0.010                            | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.006                           | 0.006 | 0.006 | 0.007 | 0.006 | 0.007 | 0.991 | 0.991 | 0.991 | 0.990 | 0.990 | 0.990 |
|    | 10          | 0.006                            | 0.013 | 0.008 | 0.011 | 0.009 | 0.011 | 0.005                           | 0.011 | 0.005 | 0.008 | 0.007 | 0.010 | 0.997 | 0.990 | 0.992 | 0.986 | 0.991 | 0.993 |
|    | 20          | 0.017                            | 0.229 | 0.097 | 0.045 | 0.024 | 0.017 | 0.011                           | 0.035 | 0.036 | 0.012 | 0.015 | 0.011 | 0.978 | 0.276 | 0.517 | 0.831 | 0.946 | 0.978 |
|    | 50          | 0.041                            | 0.226 | 0.105 | 0.075 | 0.044 | 0.045 | 0.032                           | 0.036 | 0.058 | 0.033 | 0.033 | 0.034 | 0.845 | 0.251 | 0.398 | 0.663 | 0.831 | 0.856 |
|    | 100         | 0.133                            | 0.230 | 0.172 | 0.130 | 0.131 | 0.162 | 0.105                           | 0.055 | 0.132 | 0.107 | 0.108 | 0.133 | 0.615 | 0.181 | 0.143 | 0.468 | 0.463 | 0.423 |

注:表中 B 为 Blackville 站;M 为 McClellanville 站。下同。

表 5 BP 神经网络模型对测试集中不同深度土壤湿度的反演评估结果

Table 5 Results of inverse assessment of soil moisture at different depths in the test set by BP neural network models

| 站点 | 土层深度/<br>cm | RMSE/( $\text{m}^3/\text{m}^3$ ) |       |       |       |       |       | MAE/( $\text{m}^3/\text{m}^3$ ) |       |       |       |       |       | $R$   |       |       |       |       |       |
|----|-------------|----------------------------------|-------|-------|-------|-------|-------|---------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|    |             | 试验 1                             | 试验 2  | 试验 3  | 试验 4  | 试验 5  | 试验 6  | 试验 1                            | 试验 2  | 试验 3  | 试验 4  | 试验 5  | 试验 6  | 试验 1  | 试验 2  | 试验 3  | 试验 4  | 试验 5  | 试验 6  |
| B  | 5           | 0.007                            | 0.007 | 0.007 | 0.011 | 0.019 | 0.011 | 0.006                           | 0.006 | 0.006 | 0.009 | 0.012 | 0.009 | 0.983 | 0.983 | 0.983 | 0.980 | 0.922 | 0.980 |
|    | 10          | 0.004                            | 0.026 | 0.008 | 0.015 | 0.006 | 0.010 | 0.004                           | 0.018 | 0.006 | 0.012 | 0.003 | 0.008 | 0.997 | 0.808 | 0.985 | 0.935 | 0.991 | 0.989 |
|    | 20          | 0.027                            | 0.029 | 0.030 | 0.028 | 0.029 | 0.031 | 0.022                           | 0.022 | 0.025 | 0.023 | 0.024 | 0.026 | 0.743 | 0.613 | 0.722 | 0.728 | 0.716 | 0.720 |
|    | 50          | 0.055                            | 0.040 | 0.046 | 0.046 | 0.048 | 0.043 | 0.047                           | 0.031 | 0.038 | 0.038 | 0.041 | 0.036 | 0.721 | 0.334 | 0.680 | 0.602 | 0.644 | 0.603 |
|    | 100         | 0.060                            | 0.045 | 0.046 | 0.046 | 0.061 | 0.048 | 0.046                           | 0.037 | 0.040 | 0.037 | 0.059 | 0.042 | 0.307 | 0.163 | 0.266 | 0.372 | 0.386 | 0.461 |
| M  | 5           | 0.011                            | 0.011 | 0.011 | 0.013 | 0.012 | 0.013 | 0.007                           | 0.007 | 0.007 | 0.009 | 0.008 | 0.009 | 0.988 | 0.998 | 0.988 | 0.983 | 0.986 | 0.983 |
|    | 10          | 0.005                            | 0.016 | 0.009 | 0.012 | 0.010 | 0.020 | 0.006                           | 0.012 | 0.005 | 0.011 | 0.008 | 0.018 | 0.992 | 0.981 | 0.990 | 0.984 | 0.989 | 0.982 |
|    | 20          | 0.025                            | 0.031 | 0.092 | 0.043 | 0.024 | 0.102 | 0.017                           | 0.024 | 0.037 | 0.018 | 0.016 | 0.036 | 0.942 | 0.868 | 0.560 | 0.836 | 0.941 | 0.467 |
|    | 50          | 0.044                            | 0.043 | 0.101 | 0.045 | 0.045 | 0.102 | 0.033                           | 0.034 | 0.042 | 0.035 | 0.034 | 0.043 | 0.853 | 0.843 | 0.530 | 0.823 | 0.823 | 0.461 |
|    | 100         | 0.125                            | 0.150 | 0.159 | 0.128 | 0.127 | 0.173 | 0.096                           | 0.120 | 0.132 | 0.106 | 0.105 | 0.142 | 0.515 | 0.482 | 0.512 | 0.463 | 0.466 | 0.313 |

由表 4~6 可见,试验 1 反演效果最佳, $R$  在各组试验中最高,而 RMSE 和 MAE 最低。试验 2 中除 McClellanville 站 BP 神经网络模型外,反演能力迅速下降,反演效果在 6 组试验中最差,尤其是对  $M_{s50}$  和  $M_{s100}$  的反演。试验 3 对  $M_{s20}$ 、 $M_{s50}$  和  $M_{s100}$  的反演效果较试验 1 变差,但除 BP 神经网络模型外,其余模型对深层土壤湿度的反演效果均优于试验 2。综合对比试验 1~3,增加实测表层土壤湿度作为驱动因子可以有效提高模型对 20 cm 及更深层土壤湿度的反演能力。

试验 4 的反演效果较试验 1 也明显变差,尤其是在 20 cm 及更深土层,表明土壤温度对模型在表层至深层土壤湿度反演中起着重要作用。试验 5 的反演效果较试验 1 也略有变差,表明 10 cm 及更深层土壤温度

表6 随机森林模型对测试集中不同深度土壤湿度的反演评估结果

| Table 6 Results of inverse assessment of soil moisture at different depths in the test set by random forest models |             |                                  |       |       |       |       |       |                                 |       |       |       |       |       |       |       |       |       |       |       |
|--------------------------------------------------------------------------------------------------------------------|-------------|----------------------------------|-------|-------|-------|-------|-------|---------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 站点                                                                                                                 | 土层深度/<br>cm | RMSE/( $\text{m}^3/\text{m}^3$ ) |       |       |       |       |       | MAE/( $\text{m}^3/\text{m}^3$ ) |       |       |       |       |       | R     |       |       |       |       |       |
|                                                                                                                    |             | 试验1                              | 试验2   | 试验3   | 试验4   | 试验5   | 试验6   | 试验1                             | 试验2   | 试验3   | 试验4   | 试验5   | 试验6   | 试验1   | 试验2   | 试验3   | 试验4   | 试验5   | 试验6   |
| B                                                                                                                  | 5           | 0.005                            | 0.005 | 0.005 | 0.011 | 0.011 | 0.011 | 0.004                           | 0.004 | 0.004 | 0.009 | 0.009 | 0.009 | 0.996 | 0.996 | 0.996 | 0.978 | 0.981 | 0.978 |
|                                                                                                                    | 10          | 0.004                            | 0.050 | 0.006 | 0.005 | 0.005 | 0.005 | 0.003                           | 0.012 | 0.004 | 0.002 | 0.005 | 0.004 | 0.999 | 0.603 | .991  | 0.994 | 0.991 | 0.992 |
|                                                                                                                    | 20          | 0.023                            | 0.059 | 0.029 | 0.030 | 0.029 | 0.030 | 0.024                           | 0.032 | 0.023 | 0.024 | 0.024 | 0.024 | 0.813 | 0.456 | 0.815 | 0.818 | 0.812 | 0.813 |
|                                                                                                                    | 50          | 0.031                            | 0.072 | 0.048 | 0.063 | 0.050 | 0.048 | 0.027                           | 0.046 | 0.040 | 0.049 | 0.041 | 0.039 | 0.734 | 0.340 | 0.651 | 0.513 | 0.634 | 0.685 |
|                                                                                                                    | 100         | 0.043                            | 0.067 | 0.050 | 0.064 | 0.138 | 0.049 | 0.036                           | 0.045 | 0.040 | 0.047 | 0.047 | 0.043 | 0.548 | 0.147 | 0.250 | 0.238 | 0.342 | 0.369 |
|                                                                                                                    |             | 5                                | 0.005 | 0.005 | 0.006 | 0.011 | 0.010 | 0.011                           | 0.006 | 0.006 | 0.006 | 0.008 | 0.006 | 0.008 | 0.995 | 0.995 | 0.995 | 0.987 | 0.998 |
| M                                                                                                                  | 10          | 0.002                            | 0.011 | 0.012 | 0.009 | 0.009 | 0.012 | 0.004                           | 0.009 | 0.010 | 0.007 | 0.007 | 0.010 | 0.998 | 0.990 | 0.996 | 0.994 | 0.994 | 0.991 |
|                                                                                                                    | 20          | 0.016                            | 0.047 | 0.017 | 0.036 | 0.042 | 0.017 | 0.010                           | 0.014 | 0.011 | 0.012 | 0.014 | 0.010 | 0.979 | 0.820 | 0.977 | 0.892 | 0.843 | 0.977 |
|                                                                                                                    | 50          | 0.028                            | 0.243 | 0.045 | 0.050 | 0.059 | 0.046 | 0.021                           | 0.039 | 0.034 | 0.037 | 0.039 | 0.035 | 0.854 | 0.235 | 0.840 | 0.782 | 0.703 | 0.836 |
|                                                                                                                    | 100         | 0.119                            | 0.237 | 0.149 | 0.141 | 0.144 | 0.154 | 0.094                           | 0.061 | 0.120 | 0.114 | 0.115 | 0.127 | 0.633 | 0.170 | 0.494 | 0.367 | 0.301 | 0.460 |
|                                                                                                                    |             |                                  |       |       |       |       |       |                                 |       |       |       |       |       |       |       |       |       |       |       |

在不同深度土壤湿度反演中的作用不可忽略。试验6的反演效果与试验3和试验4较为一致,较试验1变差,尤其是在20 cm及更深土层。试验4~6的反演效果进一步表明土壤温度和表层土壤湿度对深层土壤湿度的反演效果影响显著。在各组试验中,3种机器学习模型对两个气象站表层土壤湿度 $M_{ss}$ 和 $M_{s10}$ 的反演效果优于深层土壤湿度 $M_{s20}$ 、 $M_{s50}$ 和 $M_{s100}$ 。

综合以上分析,试验1中利用所有主成分因子作为驱动的机器学习模型对深层土壤湿度的反演效果显著优于其余5组试验,增加表层土壤湿度数据及各土层的土壤温度作为驱动因子可以有效提升机器学习模型对深层土壤湿度的反演能力。因此下面结合试验1深入分析3种机器学习模型对不同深度土壤湿度的反演性能。

3.4 反演精度分析

由表4~6可知,在试验1中,3种机器学习模型对 $M_{ss}$ 和 $M_{s10}$ 的反演效果非常好,相关系数 $R$ 均高于0.98, RMSE均低于 $0.011\text{ m}^3/\text{m}^3$ ,而MAE均低于 $0.007\text{ m}^3/\text{m}^3$ 。其中,随机森林模型对 $M_{ss}$ 的反演效果最优,表现在 $M_{ss}$ 的RMSE与MAE较支持向量机模型和BP神经网络模型略低,而 $R$ 略高。在Blackville站, $M_{s10}$ 的 $R$ 均高于0.997(支持向量机模型和随机森林模型高达0.999),而RMSE和MAE均低于 $0.00\text{ m}^3/\text{m}^3$ (支持向量机模型和随机森林模型MAE低至 $0.003\text{ m}^3/\text{m}^3$ );在McClellanville站,随机森林反演模型对 $M_{s10}$ 的反演效果最优, RMSE与MAE在3种机器学习模型中最低而 $R$ 最高。

对20 cm及更深层土壤湿度的反演,随机森林模型的反演效果显著优于支持向量机模型和BP神经网络模型,主要表现在 $M_{s20}$ 、 $M_{s50}$ 和 $M_{s100}$ 的RMSE与MAE最低而 $R$ 最高。综上所述,在3种机器学习模型中,随机森林模型对不同深度土壤湿度的反演能力优于支持向量机模型和BP神经网络模型,稳定性与适配性更高。

3.5 不同深度土壤湿度反演效果分析

由表4~6可知,在相同试验条件下,3种机器学习模型对不同深度土壤湿度的反演效果存在差异。以试验1为例,不同深度土壤湿度的反演效果在5、10 cm土层较优,而随着土层深度的增加(20、50、100 cm),反演效果迅速变差。例如,对于支持向量机模型,Blackville站 $M_{ss}$ 和 $M_{s10}$ 的RMSE分别为0.005、 $0.003\text{ m}^3/\text{m}^3$ , MAE分别为0.004、 $0.003\text{ m}^3/\text{m}^3$ ,而 $R$ 分别为0.992、0.999;随着土层深度的增加, RMSE和MAE迅速增大而 $R$ 迅速减小,  $M_{s50}$ 和 $M_{s100}$ 的RMSE分别为0.034、0.45  $\text{m}^3/\text{m}^3$ , MAE分别为0.023、 $0.039\text{ m}^3/\text{m}^3$ ,  $R$ 分别为0.685、0.397。

土壤湿度数据与气象因子的关系密切,通过多种气象因子对土壤湿度数据进行反演,能够获取较为准确的土壤湿度值。分析表4~6可知,3种机器学习模型对土层5、10 cm处的土壤湿度反演效果最优。在Blackville站, $M_{ss}$ 和 $M_{s10}$ 的RMSE介于 $0.003\sim0.007\text{ m}^3/\text{m}^3$ 之间, MAE介于 $0.003\sim0.006\text{ m}^3/\text{m}^3$ 之间,  $R$ 介于 $0.983\sim0.999$ 之间;在McClellanville站, $M_{ss}$ 和 $M_{s10}$ 的RMSE介于 $0.002\sim0.011\text{ m}^3/\text{m}^3$ 之间, MAE介于 $0.004\sim0.007\text{ m}^3/\text{m}^3$ 之间,  $R$ 介于 $0.988\sim0.998$ 之间。可见在0~10 cm的土层深度内,气温、相对湿度等气象因子对土壤湿度的影响显著。随着土层深度的增加,模型的反演值与观测值的误差越来越大。在Blackville站,  $M_{s100}$ 的RMSE和MAE最高达 $0.060\text{ m}^3/\text{m}^3$ 和 $0.046\text{ m}^3/\text{m}^3$ ,而 $R$ 最低值降至0.307;在

McClellanville 站,  $M_{s100}$  的 RMSE 和 MAE 最高达  $0.133\text{ m}^3/\text{m}^3$  和  $0.105\text{ m}^3/\text{m}^3$ ,  $R$  最低值降至 0.515。其主要原因是随着土层向下延伸,气象因子对土壤湿度的影响能力逐渐减弱,降低了机器学习模型对深层土壤湿度的反演能力,使得深层土壤湿度的反演值与观测值之间的偏差增大。3 种机器学习模型对地表 5、10 cm 土层的土壤湿度反演效果非常好,反演值与观测值误差非常小;而对 20 cm 及更深层的土壤湿度(50、100 cm),3 种机器学习模型的反演效果变差,反演值与观测值误差显著增大,仅能展现土壤湿度的平均状况。

4 结 论

- a. 在 3 种机器学习模型中,相较于支持向量机模型和 BP 神经网络模型,随机森林模型对不同深度土壤湿度的反演效果更佳,更适用于不同深度土壤湿度的反演。
- b. 3 种机器学习模型对地表 0 ~ 10 cm 深度内土壤湿度的反演效果更佳,但对深层土壤湿度的反演效果随着土层深度的增加呈变差趋势。这主要是气象因子对机器学习模型的影响造成的;随着土层深度的增加,气象因子对深层土壤湿度的影响逐渐减弱,使得以气象因子驱动为主的机器学习模型对深层土壤湿度的反演能力减弱。
- c. 增加表层土壤湿度及不同深度土壤温度作为驱动因子可以有效提高机器学习模型对深层土壤湿度(20、50、100 cm)的反演能力。

参考文献:

[ 1 ] 郭连云. 三江源区高寒草原土壤湿度变化特征及与气候因子的关系[J]. 草业科学,2018,35(1):46-53. ( GUO Lianyun. Alpine grassland soil moisture variation characteristics and its relationship with climate factors in Three River Source [J]. Pratacultural Science,2018,35(1):46-53. ( in Chinese ))

[ 2 ] 解雪,陈军锋,郑秀清,等. 基于主成分分析和粒子群算法优化支持向量机的冻融土壤蒸发预报模型[J]. 节水灌溉,2020(1):61-65. ( XIE Xue, CHEN Junfeng, ZHENG Xiuqing, et al. Freeze-thaw soil evaporation forecast model based on PCA-PSO-SVM [J]. Water Saving Irrigation,2020(1):61-65. ( in Chinese ))

[ 3 ] CHAHINE M T. The hydrological cycle and its influence on climate[J]. Nature,1992,359:373-380.

[ 4 ] 陈钢,王船海,翟月,等. 太湖流域平原地区田间尺度产流机理试验研究[J]. 河海大学学报( 自然科学版 ),2021,49(1):57-63. ( CHEN Gang, WANG Chuanhai, ZHAI Yue, et al. Field-scale experiment study of runoff generation mechanism in plain area of Taihu Basin, China [J]. Journal of Hohai University ( Ntural Sciences ),2021,49(1):57-63. ( in Chinese ))

[ 5 ] 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. ( HE Yazhou, ZHANG Ke, CHAO Lijun, et al. Watershed runoff simulation based on multi-source remotely sensed soil moisture and data assimilation [J]. Water Resources Protection,2023,39(2):145-151. ( in Chinese ))

[ 6 ] 吴志勇,张静杰,程丹丹,等. 考虑地表-土壤-作物水分联合亏缺的综合旱情指数研究[J]. 水资源保护,2023,39(3):55-64. ( WU Zhiyong, ZHANG Jingjie, CHENG Dandan, et al. Study of comprehensive drought index based on water deficits of surface, soil, and crop [J]. Water Resources Protection,2023,39(3):55-64. ( in Chinese ))

[ 7 ] 李彬楠,樊贵盛. 基于支持向量机方法的土壤水分特征曲线预测模型[J]. 节水灌溉,2019(1):108-111. ( LI Binnan, FAN Guisheng. Prediction model of soil moisture characteristic curve based on support vector machine [J]. Water Saving Irrigation,2019(1):108-111. ( in Chinese ))

[ 8 ] 王彤彤,翟军海,何欢,等. BP 神经网络和 SVM 模型对施加生物炭土壤水分预测的适用性[J]. 水土保持研究,2017,24(3):86-91. ( WANG Tongtong, ZHAI Junhai, HE Huan, et al. Applicability of BP neural network model and SVM model to predicting soil moisture under incorporation of biochar into soils [J]. Research of Soil and Water Conservation,2017,24(3):86-91. ( in Chinese ))

[ 9 ] 韩慧敏,沈润平,黄安奇,等. 基于集成学习方法的 CLDAS 土壤湿度降尺度研究[J]. 南京信息工程大学学报( 自然科学版 ),2021,13(6):693-706. ( HAN Huimin, SHEN Runping, HUANG Anqi, et al. Downscaling of CLDAS soil moisture based on ensemble learning method [J]. Journal of Nanjing University of Information Science & Technology ( Natural Science Edition ),2021,13(6):693-706. ( in Chinese ))

[ 10 ] 屈忠义,陈亚新,史海滨,等. 地下水文预测中 BP 网络的模型结构及算法探讨[J]. 水利学报,2004,35(2):88-93. ( QU Zhongyi, CHEN Yaxin, SHI Haibin, et al. Structure and algorithm of BP network for underground hydrology forecasting [J]. Journal of Hydraulic Engineering,2004,35(2):88-93. ( in Chinese ))

[ 11 ] 沈润平,郭佳,张婧娴,等. 基于随机森林的遥感干旱监测模型的构建[J]. 地球信息科学学报,2017,19(1):125-133.



(SHEN Runping, GUO Jia, ZHANG Jingxian, et al. Construction of a drought monitoring model using the random forest based remote sensing[J]. Journal of Geo-information Science, 2017, 19(1): 125-133. (in Chinese))

[12] LIU Di, MISHRA A K, YU Zhongbo. Evaluating uncertainties in multi-layer soil moisture estimation with support vector machines and ensemble Kalman filtering[J]. Journal of Hydrology, 2016, 538: 243-255.

[13] JAHIRUL M I, RASUL M G, BROWN R J, et al. Investigation of correlation between chemical composition and properties of biodiesel using principal component analysis (PCA) and artificial neural network (ANN) [J]. Renewable Energy, 2021, 168: 632-646.

[14] CORTES C, VAPNIK V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273-297.

[15] 张哲铭, 李晓瑜, 姬建. 基于LS-SVM的TBM掘进参数预测模型[J]. 河海大学学报(自然科学版), 2021, 49(4): 373-379. (ZHANG Zheming, LI Xiaoyu, JI Jian. TBM excavation parameter prediction model based on LS-SVM method[J]. Journal of Hohai University (Ntural Sciences), 2021, 49(4): 373-379. (in Chinese))

[16] 杨丽萍, 侯成磊, 苏志强, 等. 基于机器学习和全极化雷达数据的干旱区土壤湿度反演[J]. 农业工程学报, 2021, 37(13): 74-82. (YANG Liping, HOU Chenglei, SU Zhiqiang, et al. Soil moisture inversion in arid areas by using machine learning and fully polarimetric SAR imagery[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(13): 74-82. (in Chinese))

[17] 张珂, 牛杰帆, 李曦, 等. 洪水预报智能模型在中国半干旱半湿润区的应用对比[J]. 水资源保护, 2021, 37(1): 28-35. (ZHANG Ke, NIU Jiefan, LI Xi, et al. Comparison of artificial intelligence flood forecasting models in China's semi-arid and semi-humid regions[J]. Water Resources Protection, 2021, 37(1): 28-35. (in Chinese))

[18] GISLASON P, BENEDIKTSSON J, SVEINSSON J. Random forests for land cover classification[J]. Pattern Recognition Letters, 2006, 27(4): 294-300.

[19] 何涯舟, 张珂, 晁丽君, 等. 基于多源卫星遥感产品的土壤湿度融合与降尺度研究[J]. 河海大学学报(自然科学版), 2022, 50(6): 40-46. (HE Yazhou, ZHANG Ke, CHAO Lijun, et al. Study on soil moisture merging and downscaling based on multi-source satellite remote sensing products[J]. Journal of Hohai University (Ntural Sciences), 2022, 50(6): 40-46. (in Chinese))

[20] 刘永伟, 王文, 刘元波, 等. 水文模型模拟预报的多源数据同化方法及应用研究进展[J]. 河海大学学报(自然科学版), 2021, 49(6): 483-491. (LIU Yongwei, WANG Wen, LIU Yuanbo, et al. Advances in multi-source data assimilation approach and application in simulation and forecast of hydrological model[J]. Journal of Hohai University (Ntural Sciences), 2021, 49(6): 483-491. (in Chinese))

[21] 王青青, 张珂, 叶金印, 等. 安徽省土壤湿度时空变化规律分析及遥感反演[J]. 河海大学学报(自然科学版), 2019, 47(2): 114-118. (WANG Qingqing, ZHANG Ke, YE Jinyin, et al. Spatiotemporal analysis and remote sensing retrieval of soil moisture across Anhui Province, China [J]. Journal of Hohai University (Ntural Sciences), 2019, 47(2): 114-118. (in Chinese))

[22] 沈花玉, 王兆霞, 高成耀, 等. BP神经网络隐含层单元数的确定[J]. 天津理工大学学报, 2008, 24(5): 13-15. (SHEN Huayu, WANG Zhaoxia, GAO Chengyao, et al. Determining the number of BP neural network hidden layer units[J]. Journal of Tianjin University of Technology, 2008, 24(5): 13-15. (in Chinese))

(收稿日期: 2023-06-13 编辑: 熊水斌)