

基于机器学习的东非植被变化因子重要性分析

张秀梅¹, 马波^{1,2}, 张怡捷¹

(1.西北农林科技大学 水土保持研究所, 陕西 杨凌 712100; 2.中国科学院 水利部 水土保持研究所, 陕西 杨凌 712100)

摘要: [目的] 基于机器学习算法对东非植被变化进行因子重要性分析, 测度不同算法在各情况下的精度差异及适用性, 为保护、恢复和促进可持续森林管理、水土流失综合防治提供科学依据。[方法] 以东非9个国家2001—2020年的归一化植被指数(normalized difference vegetation index, NDVI)变化为研究对象, 选取影响东非植被变化的2个气候因子及5个人类活动因子作为自变量, 利用随机森林(random forest, RF)、BP神经网络(BP neural networks, BP)、支持向量机(support vector machines, SVM)、遗传算法(genetic algorithm, GA)、径向基神经网络(radial basis function, RBF)、卷积神经网络(convolutional neural networks, CNN)6种机器学习算法建立NDVI预测模型, 以决定系数(R^2)、平均绝对误差(MAE, mean absolute error)、平均相对误差(MRE, mean relative error)3个指标评价评估6种机器学习算法预测NDVI变化的潜力, 并基于所得的最优模型即对选取的7个因子进行重要性分析。[结果] 精度验证结果表明, 研究区内在全因子的情况下, CNN算法的回归精度最差; 经逐轮删除一个综合表现不佳的算法后, RF算法建立的模型在东非NDVI变化分析中回归精度较高; 基于随机森林算法的不同因子变量对NDVI变化的重要性表明, 年降水量、 N_2O 排放量、 CH_4 排放量、牲畜数量4个变量对NDVI变化回归的结果影响较大。[结论] 随机森林算法的回归能力在东非NDVI模拟中具有相对优势, 降水量是影响植被变化最重要的气候因子, 同时, 温室气体的排放对于东非植被的变化也具有一定的影响。东非各国应提高植被变化对气候环境、社会经济和政治制度相互依存关系的认识和理解, 并制定适当的政策以促进可持续森林管理、防治荒漠化。

关键词: 归一化植被指数(NDVI); 机器学习; 精度评估; 因子重要性; 东非

文献标识码: B

文章编号: 1000-288X(2023)06-0227-10

中图分类号: F301.24

文献参数: 张秀梅, 马波, 张怡捷. 基于机器学习的东非植被变化因子重要性分析[J]. 水土保持通报, 2023, 43(6): 227-236. DOI: 10.13961/j.cnki.stbctb.2023.06.028; Zhang Xiumei, Ma Bo, Zhang Yijie. Importance analysis of vegetation change factors in East Africa based on machine learning [J]. Bulletin of Soil and Water Conservation, 2023, 43(6): 227-236.

Importance Analysis of Vegetation Change Factors in East Africa Based on Machine Learning

Zhang Xiumei¹, Ma Bo^{1,2}, Zhang Yijie¹

(1. Institute of Soil and Water Conservation, Northwest A&F University,

Yangling, Shaanxi 712100, China; 2. Institute of Soil and Water Conservation, Chinese

Academy of Sciences and Ministry of Water Resources, Yangling, Shaanxi 712100, China)

Abstract: [Objective] A factor importance analysis of vegetation changes in East Africa based on different machine learning algorithms was conducted to measure the accuracy and applicability of the different algorithms in order to provide a scientific basis for protecting, restoring, and promoting sustainable forest management and comprehensive prevention and control of soil erosion. [Methods] Changes in normalized difference vegetation index (NDVI) for nine countries in East Africa from 2001 to 2020 were determined. The independent treatment variables were two climatic factors and five human activity factors affecting vegetation changes in East Africa. Six machine learning algorithms were used to establish NDVI prediction models: random forest (RF), BP neural networks (BP), support vector machines (SVM), genetic algorithm (GA),

收稿日期: 2022-11-14

修回日期: 2023-03-31

资助项目: 国家自然科学基金项目“尼罗河上游丘陵区水土保持与高产高效农业研究”(41561144011), “黄土高原北部冻融作用下物理结皮特性变化对风蚀的影响机理”(42277342), “黄土高原冻融对水蚀过程的作用机理研究”(41771311)

第一作者: 张秀梅(1998—), 女(汉族), 山西省晋城市人, 硕士研究生, 研究方向为机器学习与植被变化。Email: xzmzhang@nwfufu.edu.cn.

通信作者: 马波(1982—), 男(回族), 陕西省西乡县, 博士, 副研究员, 博士生导师, 主要从事土壤侵蚀机理与区域土壤侵蚀评价研究。Email: soilcrop@163.com.

radial basis function (RBF), and convolutional neural networks (CNN). Coefficient of determination (R^2), mean absolute error (MAE), and mean relative error (MRE) were used as error indicators to evaluate the potential of the six machine learning algorithms for predicting NDVI changes. Based on the optimal model (RF), the importance of the selected seven factors was determined. [Results] The accuracy verification results showed that the regression accuracy of the CNN algorithm was the worst for the full factor case in the study area. After deleting an algorithm with poor comprehensive performance in each round of testing, the model established by the RF algorithm had the highest regression accuracy for NDVI change analysis in East Africa. The importance of different factor variables to NDVI change based on RF showed that annual precipitation, N_2O emission, CH_4 emission, and livestock number had the greatest influence on the results of the NDVI change regression. [Conclusion] RF had a comparative advantage for NDVI simulation in East Africa. Precipitation was the most important climatic factor affecting vegetation changes. At the same time, greenhouse gas emissions also had an important impact on vegetation changes in East Africa. East African countries should raise awareness and understanding of the interdependence of vegetation changes on climate, environment, socio-economic conditions, and political systems, and develop appropriate policies to promote sustainable forest management and to combat desertification.

Keywords: normalized difference vegetation index (NDVI); machine learning; accuracy evaluation; factor importance; East Africa

作为全球环境变化生态指示器,植被覆盖被认为是生态系统非常重要,并且对环境变化极为敏感的能够影响土壤侵蚀的一部分重要元素^[1-2],其变化在一定程度上可反映区域自然和人类活动的变化^[3]。尤其在土壤侵蚀的发生、发育过程中,地表植被的覆盖和保护是削弱土壤侵蚀发展的关键性环节^[4],其演变方向及程度为土壤侵蚀评价的导向性要素^[5]。近年来,卫星传感器已成为研究区域到全球尺度植被动态和趋势的重要工具^[6]。为表现植被状况的时空综合特征而产生的归一化植被指数(normalized difference vegetation index, NDVI)是分析实时绿色植被生长状况,反映植被动态变化的重要指标^[7]。

目前,国内外已经发展了许多方法来估计植被覆盖变化的驱动因素,包括逐步线性回归、相关分析、残差法等,这些方法主要是分析气象因子等与 NDVI 或植被覆盖度之间的相关性,而不是定量评价各因子的重要性并监测其在长时间序列中的重要性变化规律^[8]。与此同时,为了解决非线性问题和大数据处理问题的机器学习方法应运而生。机器学习是一个跨学科的研究领域,主要研究如何利用计算机来模拟或实现人类的学习行为,从而获得新的知识或技能,为数据的分析和未来趋势的预测提供了一种数据驱动的、经验的途径^[9]。该方法已经应用于植被恢复领域,并取得了良好的效果^[10]。黄生志等^[11]引入基于支持向量机的模型量化气候变化和人类活动对植被覆盖动态的影响;吴书普等^[12]利用随机森林方法提取出非洲萨赫勒地区荒漠、草地生物群落过渡带的范围,利用元胞自动机(CA)—马尔科夫模型预测荒漠/

草地过渡带的空间位置;Adam 等^[9]提出了一种基于卫星的植被状况指标预测方法,采用高斯过程模型(GP)和线性自回归模型(AR)对 NDVI 异常等进行了预测,划定了肯尼亚国家干旱管理局所使用的干旱预警红线。

由于全球气候变暖、人口压力增加、快速城市化以及全球对初级农产品和矿物资源的需求增加等因素^[13-15],非洲的环境和植被的退化越来越严重^[13]。东非横跨赤道,植被类型广泛,一直都被认为是非洲植被严重退化的重点关注地区。该区域植被覆盖动态对气候变化及人类活动的敏感性各不相同,确定这种敏感性的影响因素对于评估整个区域对气候变化和人类活动的敏感性至关重要。现有研究针对植被动态对气候变化、人类活动的区域响应评估在数据一致性、时间序列分析技术和统计建模等方面尚未有一致定论。尤其在东非这种植被覆盖具有强烈异质性、人类活动急剧增加的大尺度空间范围内,目前尚缺乏人类活动增强对植被覆盖变化的定量研究。如何推导出有意义的环境指标对于整个区域植被变化的驱动力研究至关重要^[14]。因此,本文基于东非 9 个国家 2001—2020 年的 MODIS NDVI 数据,考虑年降水量、年均地表温度 2 个气候变化因子以及 CH_4 排放量、 N_2O 排放量、林业产品消耗量等 5 个人类活动因子,使用 6 种机器学习算法建立东非 NDVI 估算模型并对其进行精度验证,同时对影响 NDVI 变化的各因子进行重要性分析。本研究结果可以为东非各国 NDVI 估算精度的提高和方法的选取提供科学建议。

1 研究区概况

选取的东非(28.8°—51.4°E, -18.0°—12.0°S)面积约 $3.71 \times 10^6 \text{ km}^2$, 包含埃塞俄比亚、布隆迪、厄立特里亚、肯尼亚、吉布提、卢旺达、索马里、坦桑尼亚和乌干达9个国家(图1)。尽管位于热带,但主要被归类为干旱和半干旱地区^[15], 主要表现为不同的季节间雨量差异大^[16], 不同年份间降水模式差异大, 整个研究区域的年降雨量在0~2 340 mm之间变化, 年均降水量为51.73 mm。热带间辐合区(ITCZ)的季节性迁移主导使其形成了强烈的降水周期, 即3—5月的“长雨季”和10—12月的“短雨季”^[13], 介于两个雨季之间的即为旱季。温度年变和缓、年较差小, 多年平均气温为24.85℃。地形以高原为主, 海拔自埃塞俄比

亚达纳基尔洼地(Danakil Depression)海平面以下197 m到肯尼亚山(Mount Kenya)海平面以上5 880 m不等, 平均海拔高度为1 200~1 500 m。东非地貌类型多样, 北部为东非湖群高原, 呈圆形, 东、西为两支裂谷带, 裂谷带中有湖群, 并有被充填熔岩分割成的盆地, 中间高原面平坦而辽阔。土壤包含黏土、斑脱土、硅藻土、漠质土、棕钙土等。研究区域内植被类型丰富, 植被覆盖的类型变化率高, 从森林到沙漠生态系统不等, 但是以草地为主, 夹杂着木本植物^[17]。全域NDVI均值最高值为0.922 2, 均值约为0.557 0, NDVI高值主要分布在主要为海拔较高的埃塞俄比亚西南部、肯尼亚西北部、坦桑尼亚北部以及乌干达的部分区域, 低值则集中在厄立特里亚东北部及肯尼亚中部地区, 这些地区多为裸地、稀疏草地, 植被较为稀疏。

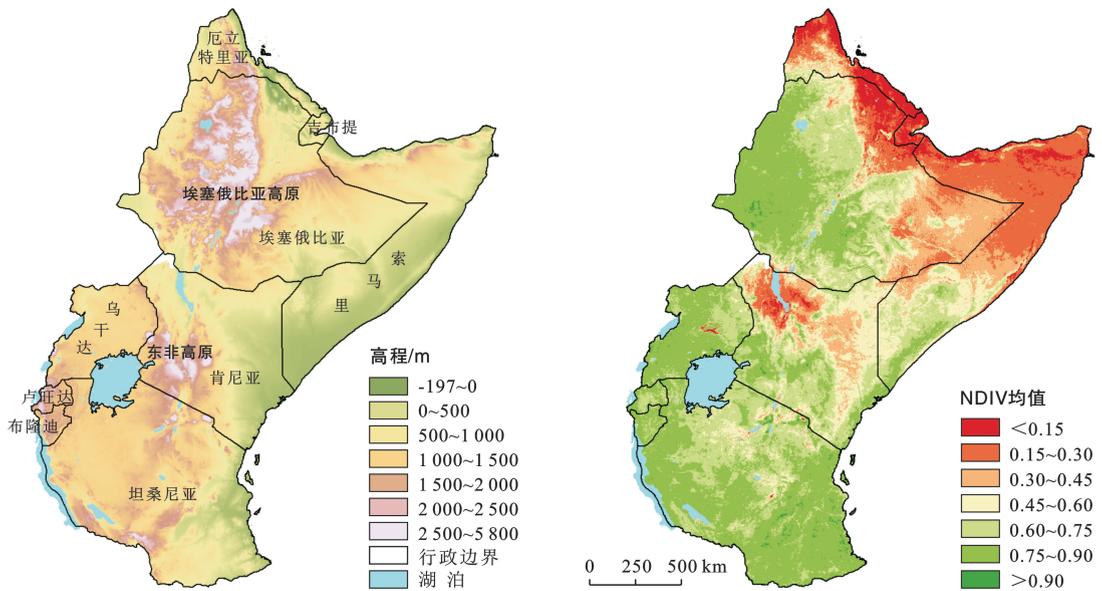


图1 研究区域高程和2001—2020年NDVI均值分布

Fig.1 Elevation of study area and mean NDVI distribution in East Africa from 2001 to 2020

2 数据来源与方法

2.1 数据来源及预处理

2.1.1 NDVI数据 为反映植被生长周期的真实特征, 避免掩盖植被周期内的变化, 本研究选用2001—2020年美国国家航空航天局地球观测系统数据和信息系统(<https://earthexplorer.usgs.gov/>)发布的按16 d合成的MOD13A2产品, 轨道编号为h20v08-h23v10, 采用国际上通用的最大合成法(MVC)对其进行处理得到年值, 通过ArcGIS 10.8裁剪得到以国家级行政区为单位的逐年NDVI。

2.1.2 气象数据 本研究所使用的降水数据为气候灾害组红外降水站数据集(climate hazards group

infrared precipitation with station, CHRIPS), 是目前东非地区最高分辨率的长时间序列降水产品^[18], 该数据集的原始分辨率为 $0.05^\circ \times 0.05^\circ$, 时间跨度为1981年至今。温度数据选用来自中分辨率成像光谱辐射计(MODIS)的分辨率为1 km的地表温度(land surface temperature, LST)产品MOD11A2, 其值为8 d中晴好天气下的地表温度/发射率的平均值, 数据下载及预处理同上述NDVI数据下载地址及处理方法。即本研究选用的降水量为各国年内降水量的总和, LST, NDVI为各数据原有精度基础上各栅格的最大值; 并选用全域的平均值为该国该年的变量值。

2.1.3 人类活动相关数据 本研究选取的人类活动因子为9个国家2001—2020年内的 CH_4 排放量、

N_2O 排放量、林业产品消耗量 (forestry, FRS)、生物质燃烧排放量 (biomass burned emissions, BBE)、牲畜数量 (production crops livestock, PCL) 5 种, 均可通过联合国粮食及农业组织 (Food and Agriculture Organization of the United Nations, FAO) 提供的包含全球各国粮食和农业的数据库页面 (<https://www.fao.org/faostat/zh/#home>) 下载得到, 数据以年内总值为计量单位, 格式为 XML。

2.2 研究方法

2.2.1 机器学习算法

① 随机森林 (random forest, RF) 算法。由 Leo Breiman^[19] 于 2001 年提出, 其原理是将 Bagging 集成学习思想与随机子空间方法相结合。随机森林以决策树为基础学习器构建 Bagging 集成, 并在决策树的训练历程中添加了随机属性的选取, 以此加大分类模型之间的相异性, 进而增强该模型的泛化能力以及预测能力^[20]。② BP 神经网络 (BP neural networks, BP)。是人工神经网络的一种类型, 是一种误差反向传递学习算法的多层前馈网络^[21]。BP 神经网络模型为 3 层前馈神经网络, 包括输入层、隐含层和输出层。BP 算法的学习过程由正向传播和反向传播组成。在正向传播过程中, 输入信号从输入层经隐含层逐层处理后传向输出层。如果输出层不能得到期望的输出, 则转入反向传播, 将误差信号沿原来的连接通道返回, 通过修改各层节点的权值, 使得误差信号最小, 是目前应用最广泛的神经网络模型之一^[22]。

③ 支持向量机 (support vector machines, SVM) 算法。该算法可以通过核函数将训练数据转换为高维特征空间^[23], 确定一个最佳超平面, 使平面与最近的正反面之间的距离达到最大^[24]。本研究采用径向基函数作为支持向量机的核函数, 并利用 “tune.svm” and “svm” 两个函数调整参数 cost 和 gamma^[25], 确定两个参数的最佳组合。④ 遗传算法 (genetic algorithm, GA)。该算法是一种借鉴生物界自然选择和自然遗传机制的随机搜索算法^[26]。它通过模拟自然进化过程来搜索最优解, 并按照某种方式进行编码, 作用于称为染色体的字符串, 模拟由这些串组成的群体进化过程, 最后通过有组织、随机的信息交换来重新组合那些适应性较好的串, 生成新的群体^[27]。⑤ 径向基神经网络 (radial basis function, RBF)。只具有三层前向网络: 第 1 层为输入层; 第 2 层为隐藏层, 对输入信息进行空间映射变换; 第 3 层为输出层, 结构相对简单, 涉及的神经元连接数目较少^[28]。与其他类型的神经网络相比, RBF 神经网络具有更强的逼近能力, 拓扑结构简单, 逼近精度高且速度更快, 可逼

近任意非线性函数, 应用广泛^[29]。⑥ 卷积神经网络 (convolutional neural networks, CNN)。是一种可以自动从数据库中学习并更新权重参数的模型^[30], 其是深度学习的代表模型之一。卷积神经网络由输入层、隐藏层和输出层组成, 而权重参数“隐藏”在隐藏层中, 隐藏层的数量越多则表明深度学习的深度越深。卷积神经网络具有结构简单、应用性强和识别率高的特点, 在比较模型性能的方面具有很好的对比意义^[31]。

2.2.2 评价指标 本文以研究区内 9 个国家 2001—2020 年逐年年均 NDVI, 共 180 个样本作为因变量, 选择影响植被生长的 7 个要素为自变量。以每 10 次为 1 轮模拟, 逐轮删除 1 个综合表现不佳的算法, 同时基于随机森林算法所得的因子重要性每轮剔除一个相对重要性较低的因子, 直至仅有两项对比即可得到最优算法并对影响 NDVI 变化的各因子重要性进行排序。整体样本采用简单随机法随机抽取 126 个、即 70% 的样本数据作为训练集 (training sets, TRS), 54 个、即 30% 的样本作为测试集 (testing sets, TES), 采用 SVM, BP, RF, RBF, CNN, GA 6 种方法建立东非 NDVI 回归模型, 用决定系数 (R^2)、平均绝对误差 (MAE)、平均相对误差 (MBE) 3 个误差评价指标来评价机器学习算法的准确性, 计算公式为:

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

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |\bar{y}_i - y_i| \quad (2)$$

$$\text{MBE} = \frac{\left| \frac{\sum_{i=1}^m \hat{y}_i - y_i}{y_i} \right|}{m} \quad (3)$$

式中: i 代表年份; y_i 代表东非实际 NDVI; \hat{y}_i 代表预测 NDVI 即根据回归方程计算值; \bar{y}_i 为平均 NDVI。

3 结果与分析

3.1 模型评价

3.1.1 全因子全算法的模型评价 不同机器学习算法在东非 NDVI 模拟预测的适应性不同, 在全因子的模拟情况下各算法的训练测试结果如图 2 所示。可以看出, SVM 算法建立的回归模型分布较为集中, 拟合优度较高, 训练集、测试集的 R^2 均值分别达到 0.994, 0.990, 训练集 MAE 均值最低、仅为 0.011; CNN 算法的精度最低, 训练集、测试集的数据较为分

散, R^2 均值仅有 0.828 和 0.858, 其训练集 MAE, MBE 也均为 7 种算法的最大值, 分别为 0.054, 0.001。GA, BP 算法表现次之, 训练集的 R^2 相较于 SVM 分别减少了 0.17% 和 0.27%, MAE 则相对增加了 8.44%, 17.00%; RF 算法在本模拟情景下, 其测试集拟合优度表现出较大的波动性, R^2 , MAE 均值与

中位数相差达 0.40%, 4.78%; RBF 算法训练集、测试集的 MAE, MBE 则分别为同种情况下的绝对值最大值, 表现相对一般。综合来看, 6 种机器学习算法分别在不同的评价指标下具有各自的相对优势, 若想求得适用于东非 NDVI 预测的最佳回归模型还需进一步模拟并重新评估。

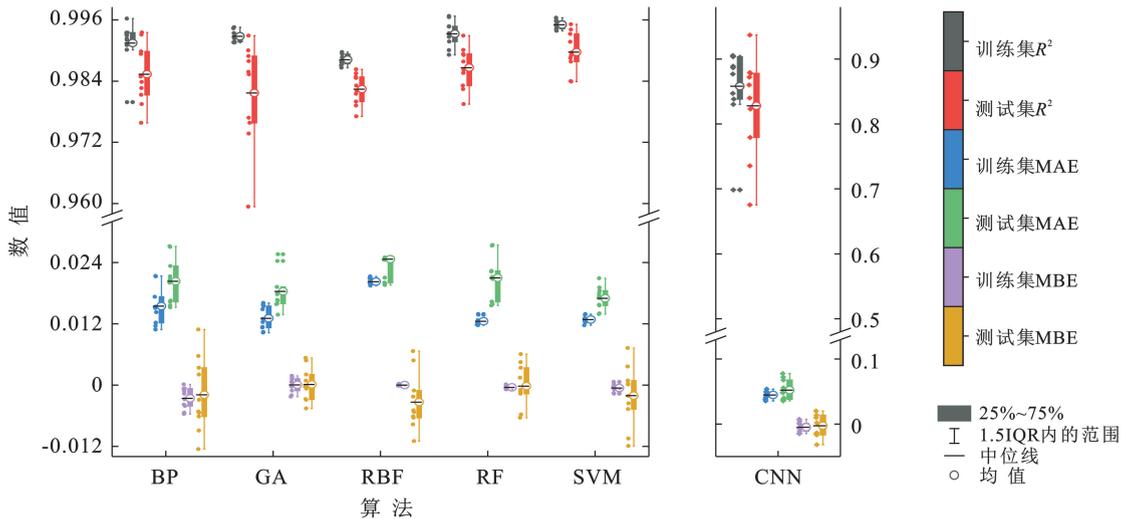


图 2 不同机器学习算法首轮训练测试结果

Fig.2 First round of training test results of different machine learning algorithms

3.1.2 最优模型的探求 为寻找适用于东非 NDVI 计算回归的最佳模型, 同时评估基于不同机器学习算法建立的 NDVI 估算模型是否存在差异, 本文基于各算法拟合所得的评价指标值(图 2), 逐轮删除 2 个及 2 个以上指标均劣于其他算法同期表现的算法^[32], 直至仅有两项对比即可得到最优算法。图 3 为分析 5 轮模拟中各方法训练集、测试集 R^2 一致性的 Bland-Altman 图。表 1 展示了 6 种算法分别在 5 轮模拟汇总所得的 MAE, MBE 结果, 最优指标用粗体表示。结果显示, 第 1, 2 轮拟合过程中, CNN 算法、RBF 算法先后因 R^2 几何均值(0.840, 0.976) 低于同期其他算法、训练集 MAE 几何均值(0.048, 0.023) 及测试集 MAE 几何均值(0.053, 0.027) 高于同期其他算法均被淘汰; 在第 3 轮中, BP 算法训练值与预测值的最大误差为 0.032, 差分值可达 0.024, 相较其在第 2 轮的表现误差增加 29.61%, 其测试集 MAE、训练集 MBE 也逼近同期最大值; 第 4 轮拟合时, SVM 算法误差有 40%(2/10) 点在 95% 一致性界限外, 几何均值仅有 0.967, 其余四项指标均达各项极值, 表现极不稳定, 同时, GA 算法的 R^2 几何均值也仅有 0.976, 低于同期 RF 算法 R^2 几何均值 0.70%。

尽管前 4 轮中 RF 算法、GA 算法的训练集、测试集 R^2 误差均在 95% 置信区间, 但在最后一轮拟合中

二者均表现出强烈的不稳定性, 所得误差值均有 2/10 点在界限外, RF 算法误差值达全轮最高、为 0.1228, 超出置信区间上限值 45.40%; GA 算法所得训练集与测试集 R^2 的几何均值最小为 0.886, 低于 RF 算法所得最小值 4.35%; 其剩余指标没有表现出明显的优势, 但整体效果略差, 说明就预测模拟结果的定量精度而言, 更少的指标作为驱动因子在一定程度上降低了模拟结果的准确性。综合 5 轮结果来看, RF 算法各项指标在训练集、测试集中的协同性较强、同步率较高, 为模拟东非 NDVI 变化的最优算法。

3.2 东非植被变化的因子重要性分析

机器学习算法可以在建立模型时以定量比较分析东非气候因子和人类活动对其 NDVI 变化的影响, 本文基于随机森林模型对选取的 7 个因子进行重要性分析(图 4)。在各种影响东非植被生长的要素中, 降水对植被的贡献率达到 31.36%, 表明降水对于东非植被的生长具有至关重要的作用; 同为气象要素的地表温度则相对重要性程度较低, 其次最大重要性为 0.395, 仅占同期模拟过程中的 8.78%; 牲畜数量对于东非 NDVI 的变化影响也较大, 重要性比值约为 13.96%; CH_4 排放、 N_2O 排放在本次模拟中重要程度相当, 其值均稳定在 0.49 上下; 林业产品消耗的重要程度较次。

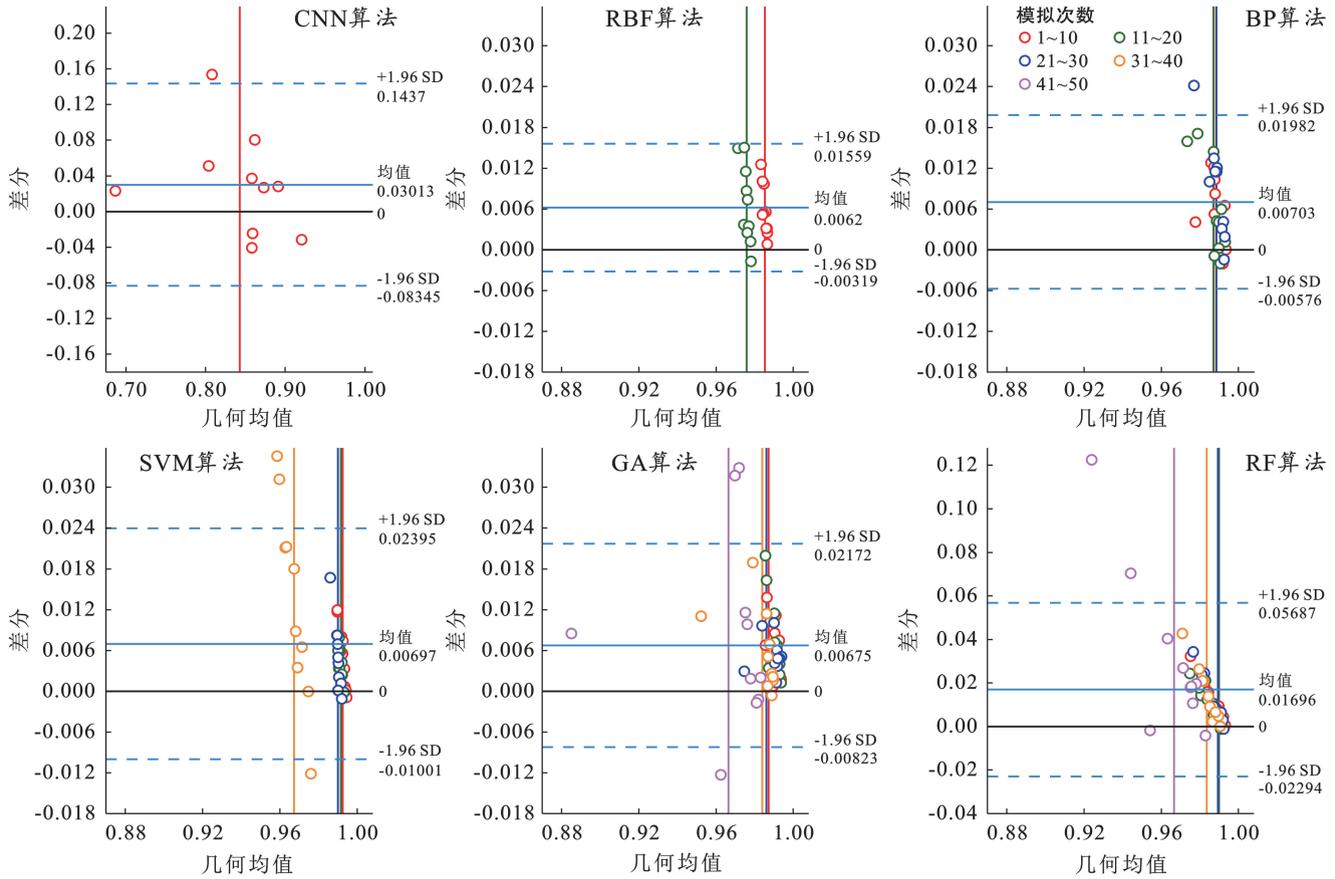


图 3 5 轮模拟中各方法训练集、测试集 R^2 Bland-Altman

Fig.3 R^2 Bland-Altman plots of training set and test set of each method in 5 round simulations

表 1 5 轮模拟中各算法所得训练集、测试集的 MAE, MBE 几何均值

Table 1 MAE and MBE geometric mean of training set and test set obtained by each algorithm in 5 round simulations

指标	算法	模拟轮数				
		1	2	3	4	5
TRS MAE	CNN	0.048	—	—	—	—
	RBF	0.017	0.023	—	—	—
	BP	0.013	0.014	0.013	—	—
	SVM	0.012	0.013	0.013	0.023	—
	GA	0.012	0.012	0.013	0.017	0.027
	RF	0.012	0.014	0.012	0.013	0.018
TES MAE	CNN	0.053	—	—	—	—
	RBF	0.021	0.027	—	—	—
	BP	0.017	0.017	0.018	—	—
	SVM	0.015	0.016	0.017	0.028	—
	GA	0.017	0.016	0.016	0.020	0.027
	RF	0.018	0.019	0.018	0.021	0.028
TRS MBE	CNN	0.006	—	—	—	—
	RBF	0.000 0	0.000 0	—	—	—
	BP	0.001	0.000 4	0.001	—	—
	SVM	0.001	0.001	0.001	0.005	—
	GA	0.001	0.001	0.001	0.002	0.001
	RF	0.000 3	0.000 4	0.000	0.000	0.000
TES MBE	CNN	0.012	—	—	—	—
	RBF	0.003	0.002	—	—	—
	BP	0.002	0.002	0.005	—	—
	SVM	0.002	0.002	0.004	0.007	—
	GA	0.003	0.005	0.001	0.002	0.003
	RF	0.001 6	0.001 9	0.005	0.004	0.005

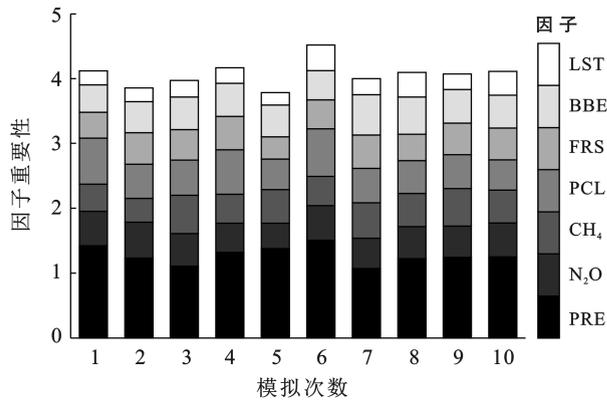


图 4 基于随机森林算法的首轮模拟各因子重要性

Fig.4 Importance diagram of each factor in first round of simulation based on random forest algorithm

为深入研究因子组合对东非 NDVI 变化的解释程度,在结合全因子重要性定量评价结果的基础上,逐轮删除一个重要值几何均值最低的因子,重新组合各因子对东非 NDVI 的变化进行重要性分析,分析结果如图 5 所示。通过对比 5 轮模拟结果可知,对东非 NDVI 变化影响最大的因子为年降水量,其平均重要

性在 31%~42%之间,且其重要性大小随着因子数量的减少而降低。仅次于降水量的为 N₂O 排放、CH₄ 排放,且 N₂O 排放的重要性始终高于 CH₄;与降水量不同,二者重要性随模拟次数的递增其值呈递增状态,增加速率分别为 0.063/10 次和 0.054/10 次;值得注意的是,第二轮中生质燃烧排放量的重要性仅有 0.475,为同期最低值而被剔除。在第三轮模拟中其分量分别有 0.121,0.135 被转移至 N₂O,CH₄ 排放量,尽管非洲不是因商业及工业能源需求而产生的温室气体的主要排放国,但由于森林砍伐和植被覆盖变化,其碳氮排量可达全球排放量的约 20%~30%^[33]。林业产品消耗量的重要性则始终低于东非牲畜数量的重要性,三轮相差比例分别为 2.81%,2.45%,4.41%,东非牲畜数量的贡献率也由最初的 0.597,占比 14.01%上涨至 0.604,占比 18.39%,在第四轮模拟中基本与 CH₄ 排放量的重要性相持平;第三轮中则更是相差仅有 4.41%,这是由于非洲人口对非洲牧场资源的依赖将持续增加,更多牧场内的牲畜资源将会转变为农田以满足粮食需要^[34]。

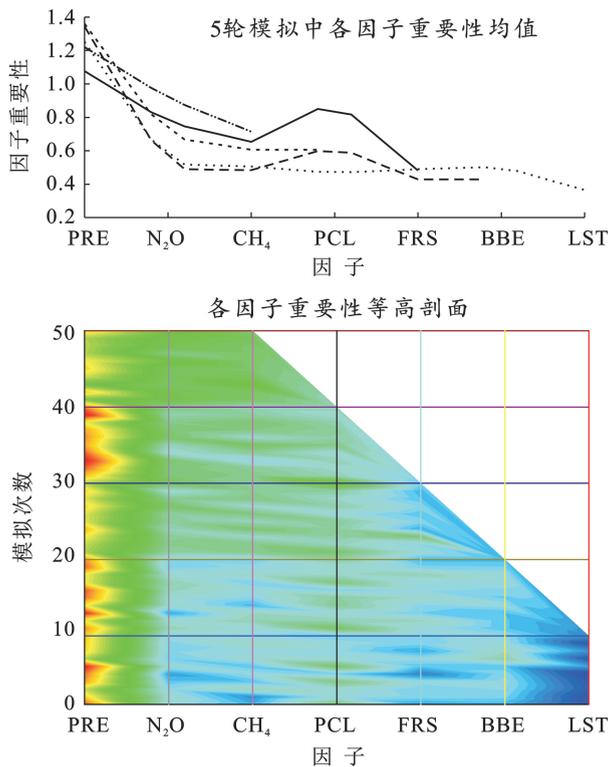


图 5 5 轮模拟中各因子重要性等高线剖面

Fig.5 Contour profile of importances of each factor in 5 rounds of simulation

4 讨论与结论

4.1 讨论

全面分析气候变化和人类活动对植被覆盖动态研究的主要困难是,驱动因素重要性大小的准确性在

很大程度上由计算方法的选择所决定^[35]。机器学习方法通过不断优化迭代表达数据中的非线性关系,拓展了气候变化和人类活动对植被变化影响研究的深度和广度,可有效弥补目前常用研究方法的局限。本文实践了多种机器学习算法在东非 NDVI 变化中的

适用性分析,从模拟预测结果角度来看,6种机器学习算法对 NDVI 的预测能力有所差异,在影响 NDVI 变化的因子为全因子的情况下,除 CNN 外其余 5 种算法均表现出较强的稳定性、准确性;各算法的稳定性随因子数量的减少而降低、模拟精度也逐步小幅度向劣,尤以 SVM 算法为首开始出现误差点在 95% 一致性以外。

按照本文试验设计,第 5 轮模拟中仅有降水量、CH₄ 排放量、N₂O 排放量 3 个因子,此时遗传算法、随机森林算法分别呈现出模拟精度低、误差值增大的劣势,遗传算法则在第 3 轮、第 4 轮呈现出该轮次模拟精度低于同期值约 0.02,0.03 的现象,故在综合考虑的前提下,本文认为随机森林算法在东非 NDVI 模拟与预测的能力更强、适应性更高,这与之前学者得出的结论一致^[36-39]。

随机森林算法可以准确描述生物物理参数与复杂环境因素之间的复杂非线性关系^[40],但在其应用

过程中,样本、训练集等的的数据量对计算结果的影响一直为机器学习的研究热点之一。本文选取东非 9 个国家 2001—2020 年逐年年均 NDVI,经由 7 个影响因子在 5 轮模拟中对比筛选出对 NDVI 变化影响最大的因子。由表 2 可知,在全因子的情况下,各因子重要性平均偏差值较大、表现出较强的离势程度;随着模拟次数的增加,按照本文逐个删除重要性最低因子的试验设计,除首轮模拟中平均偏差值本就最小的 N₂O 排放量重要性平均偏差值呈增大趋势,由 0.036 上升至 0.048 之外,其余各因子的重要性大小多呈平均偏差值减小,离势程度减弱的趋势,即在引起因变量变化过程中,越少的自变量其重要性大小越稳定。这与前人多认为训练集样本特征足够多时精度往往较好的研究所得并不一致^[41-42],这是由于在不同模拟轮数中,数据样本总量在逐步减少,而当数据量偏少时,采样方式为随机采样可以减弱数据量不足对结果精度的负面影响^[43]。

表 2 5 轮模拟中各因子重要性的平均偏差值

Table 2 Average deviation of importance of each factor in 5 rounds of simulation

模拟轮数	因子重要性平均偏差值						
	PRE	N ₂ O	CH ₄	PCL	FRS	BBE	LST
1	0.105	0.036	0.058	0.084	0.046	0.039	0.064
2	0.106	0.047	0.043	0.064	0.044	0.030	—
3	0.056	0.044	0.047	0.068	0.053	—	—
4	0.081	0.054	0.056	0.034	—	—	—
5	0.090	0.048	0.033	—	—	—	—

东非地域的特殊性也是本文试验设计的一大难点。东非盛行大面积的降雨异常和长期干旱^[44],Mike 等^[45]预测整个非洲大陆在每 10 a 将会增温 0.2~0.5 °C 的温度条件下,东非可能预计到 2050 年长雨季降水增加约 5%~20%,短雨季降水减少约 5%~10%,降水量的减少直接导致区域干旱频率和强度的增加,将对植被健康直接造成损害,甚至导致土壤侵蚀程度增加、土壤质量恶化^[46]。此外,东非社会经济和政治变化以及人口压力的骤增^[47],使得该地区通过城市扩张引起的植被退化面积持续显著增加^[48]、林业产品消耗量及出口量大幅增长、牲畜数量增加带来的人畜争粮日趋严重,废弃农作物等生物质燃烧排放量的增加、碳排放骤增引起的虫灾更是对农田的直接破坏^[49],这是本文选取 CH₄ 排放量、N₂O 排放量、林业产品消耗等因子作为人类活动因子的重要原因。另外,本研究选取该系列等因子作为人类活动因子的主要目的为保证数据同源、尽可能降低误差。另有影响东非植被变化的部分气候因子(如干旱指数、蒸散发数据等)因数据时间跨度不一致、人类活动

指标(如 GDP、永久性作物耕地面积等)受数据来源限制没有被纳入训练分析。同时,由于本研究分析的基本单元为国家级行政区,地形因子及土地利用方式由栅格尺度的二维数据向单维度的数据转换时误差程度远远高于气象数据的转换,这也是本研究的不足之处,在后续的研究中,拟决定进一步将栅格数据导入机器学习模型,最终以各像元为基本单元分析机器学习算法的适应性,同时寻找更多能代表农业活动因子的变量参与模型预测,以尽可能达到最接近真实情况的气候及人类活动变化对东非植被覆盖演变的影响程度定量估算。

4.2 结论

(1) 不同机器学习算法在不同因子组合的情况下其模拟精度存在差异,使用 6 种机器学习算法建立的模型中 CNN 算法建立的模型精度最低($R_{G.M.}^2 = 0.840$, TRS MAE_{G.M.} = 0.048}),RF 算法稳定性最高、准确性较强($R_{G.M.}^2 = 0.988$),为模拟东非 NDVI 变化的最优算法,且在逐轮删除一个 NDVI 驱动因子的情况下,指标数量的增加在一定程度上可以保证模拟

结果的准确性。

(2) 选取的 7 个影响东非 NDVI 变化的自变量中降水对 NDVI 的重要性最高,且其重要性大小随着因子数量的减少而降低,其次为 N_2O 排放量、 CH_4 排放量、牲畜数量,重要性占比依次为 20.55%,19.72% 和 18.39%,年均地表温度对研究区 NDVI 的变化影响最小,仅有 6.74%。

[参 考 文 献]

- [1] Kalisa W, Igbawua T, Henchiri M, et al. Assessment of climate impact on vegetation dynamics over East Africa from 1982 to 2015 [J]. *Scientific reports*, 2019,9(1): 16865.
- [2] 杜佳梦,包刚,佟斯琴,等.1982—2015 年蒙古国植被覆盖变化及其与气候变化和人类活动的关系[J].*草业学报*,2021,30(2):1-13.
- [3] Ma Bo, Wang Shanshan, Mupenzi C, et al. Quantitative contributions of climate change and human activities to vegetation changes in the Upper White Nile River [J]. *Remote Sensing*, 2021,13(18): 3648.
- [4] 郭敏杰.基于 NDVI 的黄土高原地区植被覆盖度对气候变化响应及定量分析[D].北京:中国科学院研究生院(教育部水土保持与生态环境研究中心),2014.
- [5] 孙艳萍.黄土高原水蚀风蚀交错带植被覆盖动态变化及其与气候因子的关系[D].陕西 杨凌:西北农林科技大学,2012.
- [6] Gao Mengdi, Piao Shilong, Chen Anping, et al. Divergent changes in the elevational gradient of vegetation activities over the last 30 years [J]. *Nature Communications*, 2019,10(1):2970.
- [7] Fangang Yu, Price K P, Ellis J, et al. Response of seasonal vegetation development to climatic variations in Eastern Central Asia [J]. *Remote Sensing of Environment*, 2003,87(1):42-54.
- [8] Meng Yanrong, Yang Mingxia, Liu Shan, et al. Quantitative assessment of the importance of bio-physical drivers of land cover change based on a random forest method [J]. *Ecological Informatics*, 2021,61:101204.
- [9] Barrett A B, Duivenvoorden S, Salakpi E E, et al. Forecasting vegetation condition for drought early warning systems in pastoral communities in Kenya [J]. *Remote Sensing of Environment*, 2020,248:111886.
- [10] Shi Yu, Jin Ning, Ma Xuanlong, et al. Attribution of climate and human activities to vegetation change in China using machine learning techniques [J]. *Agricultural and Forest Meteorology*, 2020,294:108146.
- [11] Huang Shengzhi, Zheng Xudong, Ma Lan, et al. Quantitative contribution of climate change and human activities to vegetation cover variations based on GA-SVM model [J]. *Journal of Hydrology*, 2020,584:124687.
- [12] Wu Shupu, Gao Xin, Lei Jiaqiang, et al. Spatial and temporal changes in the normalized difference vegetation index and their driving factors in the desert/grassland biome transition zone of the Sahel Region of Africa [J]. *Remote Sensing*, 2020,12(24):4119.
- [13] Hawinkel P, Thiery W, Lhermitte S, et al. Vegetation response to precipitation variability in East Africa controlled by biogeographical factors [J]. *Journal of Geophysical Research: Biogeosciences*, 2016,121(9):2422-2444.
- [14] Hoscilo A, Balzter H, Bartholomé E, et al. A conceptual model for assessing rainfall and vegetation trends in sub-Saharan Africa from satellite data [J]. *International Journal of Climatology*, 2015,35(12):3582-3592.
- [15] Ayugi B, Tan G, Niu R, et al. Evaluation of meteorological drought and flood scenarios over Kenya, East Africa [J]. *Atmosphere*, 2020,11(3):307.
- [16] Payet R, Obura D. The negative impacts of human activities in the Eastern African region: An international waters perspective [J]. *AMBIO: A Journal of the Human Environment*, 2004,33(1):24-33.
- [17] Pricope N G, Husak G, Lopez C D, et al. The climate-population nexus in the East African Horn: Emerging degradation trends in rangeland and pastoral livelihood zones [J]. *Global Environmental Change*, 2013,23(6): 1525-1541.
- [18] Agutu N O, Awange J L, Zerihun A, et al. Assessing multi-satellite remote sensing, reanalysis, and land surface models' products in characterizing agricultural drought in East Africa [J]. *Remote Sensing of Environment*, 2017,194:287-302.
- [19] Breiman L. Random forests [J]. *Machine learning*, 2001,45:5-32.
- [20] Zhong Shifa, Zhang Kai, Bagheri M, et al. Machine learning: new ideas and tools in environmental science and engineering [J]. *Environmental Science & Technology*, 2021,55(19):12741-12754.
- [21] Zipser D, Andersen R A. A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons [J]. *Nature*, 1988,331(6158):679-684.
- [22] Xue Zhenghai, Feng Wenkai, Li Botao, et al. Landslide susceptibility mapping based on the coupling of two correlation methods and the BP neural network model: a case study of the Baihetan Reservoir area, China [J]. *Frontiers in Environmental Science*, 2022, 10:1039985.
- [23] Giorgos M, Jungho I, Caesar O. Support vector machines in remote sensing: a review [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2011,66(3): 247-259.

- [24] Corinna C, Vladimir V. Support-vector networks [J]. *Machine Learning*, 1995,20: 273-297.
- [25] Rodriguez G V, Sanchez C M, Chica O M, et al. Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines [J]. *Ore Geology Reviews*, 2015,71:804-818.
- [26] Dao S D, Abhary K, Marian R. Maximising performance of genetic algorithm solver in Matlab [J]. *Engineering Letters*, 2016,24(1).
- [27] Wu Shengshen, Zhao Gaofeng, Wu Bisheng. Real-time prediction of the mechanical behavior of suction caisson during installation process using GA-BP neural network [J]. *Engineering Applications of Artificial Intelligence*, 2022,116:105475.
- [28] Whirter J G M C, Broomhead D S, Shepherd T J. A systolic array for nonlinear adaptive filtering and pattern recognition [J]. *Journal of VLSI signal processing systems for signal, image and video technology*, 1991, 3(1-2):69-75.
- [29] Li Tang, He Kaiyue, Chen Yang, et al. Integral BLF-based adaptive neural constrained regulation for switched systems with unknown bounds on control gain [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.
- [30] Petsche T, Dickinson B W. Trellis codes, receptive fields, and fault tolerant, self-repairing neural networks [J]. *IEEE Transactions on Neural Networks*, 1990,1(2):154-166.
- [31] Kim H. Convolution neural network based degenerated mode decomposition via near-field images from linear and circular polarizers [J]. *Optics and Lasers in Engineering*, 2023,160:107310.
- [32] Athanasopoulos G, Hyndman R J. Modelling and forecasting Australian domestic tourism [J]. *Tourism Management*, 2007,29(1):19-31.
- [33] Yirga A, Addisu Legesse S, Mekuriaw A. Carbon stock and mitigation potentials of Zeghie natural forest for climate change disaster reduction, Blue Nile Basin, Ethiopia [J]. *Earth Systems and Environment*, 2020,4(1):27-41.
- [34] Van Ittersum M K, Van Bussel L G J, Wolf J, et al. Can sub-Saharan Africa feed itself? [J]. *Proceedings of the National Academy of Sciences*, 2016, 113(52): 14964-14969.
- [35] 章明,张培松,刘洪斌,等.基于 SPOT VEGETATION 数据的海南岛年际植被变化研究[J].*西南大学学报(自然科学版)*,2009,31(3):148-153.
- [36] 卜灵心,来全,刘心怡.不同机器学习算法在草原草地生物量估算上的适应性研究[J].*草地学报*,2022,30(11): 3156-3164.
- [37] Hamimeche M, Niculescu S, Billey A, et al. Identification and mapping of Algerian island vegetation using high-resolution images (Pléiades and SPOT 6/7) and random forest modeling [J]. *Environmental Monitoring and Assessment*, 2021,193(9):617.
- [38] Gage E A, Cooper D J. Urban forest structure and land cover composition effects on land surface temperature in a semi-arid suburban area [J]. *Urban Forestry & Urban Greening*, 2017,28:28-35.
- [39] Ali I, Cawkwell F, Dwyer E, et al. Satellite remote sensing of grasslands: from observation to management [J]. *Journal of Plant Ecology*, 2016,9(6):649-671.
- [40] Barrett B, Nitze I, Green S, et al. Assessment of multi-temporal, multi-sensor radar and ancillary spatial data for grasslands monitoring in Ireland using machine learning approaches [J]. *Remote sensing of environment*, 2014,152:109-124.
- [41] 陈柯兵,李圣伟,何奇锴,等.关键水文信息对三峡水库调度决策的重要性分析[J].*水电与新能源*,2022,36(11):45-49.
- [42] 张子慧,吴世新,赵子飞,等.基于机器学习算法的草地上生物量估测:以祁连山草地为例[J].*生态学报*, 2022,42(22):8953-8963.
- [43] 何坤龙,赵伟,刘晓辉,等.云雾覆盖下地表温度重建机器学习模型的训练集敏感性分析[J].*遥感学报*,2021, 25(8):1722-1734.
- [44] Nicholson S E. An analysis of recent rainfall conditions in Eastern Africa [J]. *International Journal of Climatology*, 2016,36(1):526-532.
- [45] Hulme M, Doherty R, Ngara T, et al. African climate change: 1900—2100 [J]. *Climate research*, 2001, 17(2):145-168.
- [46] Pfeiffer M, Kumar D, Martens C, et al. Climate change will cause non-analog vegetation states in Africa and commit vegetation to long-term change [J]. *Biogeosciences*, 2020.
- [47] Serneels S, Linderman M, Lambin E F. A multilevel analysis of the impact of land use on interannual land-cover change in East Africa [J]. *Ecosystems*, 2007,10(3):402-418.
- [48] Elagouz M H, Abou-Shleel S M, Belal A A, et al. Detection of land use/cover change in Egyptian Nile Delta using remote sensing [J]. *The Egyptian Journal of Remote Sensing and Space Science*, 2019,1(23):57-62.
- [49] Hosea M M, Stefan J, Sopan D P, et al. Relative contribution of land use change and climate variability on discharge of upper Mara River, Kenya [J]. *Journal of Hydrology: Regional Studies*, 2016,5: 244-260.