

DOI: 10.20103/j.stxb.202209182672

朱婉怡, 张振克, 郭新亚, 冯首铭, 蒋大亮, 江飞. 马拉河流域植被生态需水特征及估算. 生态学报, 2023, 43(18): 7523-7535.

Zhu W Y, Zhang Z K, Guo X Y, Feng S M, Jiang D L, Jiang F. Characteristics and estimation of vegetation ecological water demand in the Mara River Basin. Acta Ecologica Sinica, 2023, 43(18): 7523-7535.

马拉河流域植被生态需水特征及估算

朱婉怡^{1,2}, 张振克^{1,2,*}, 郭新亚^{1,2}, 冯首铭^{1,2}, 蒋大亮^{1,2}, 江飞¹

¹ 南京大学地理与海洋科学学院, 南京 210023

² 南京大学非洲研究所, 南京 210023

摘要:生态需水是生态用水控制和区域生态环境恢复建设的基本依据。马拉河流域拥有世界著名的生态系统, 植被生态需水占流域总需水量的很大一部分。基于 1980—2020 年 ERA5 气象数据、叶面积指数 (LAI) 与世界土壤数据库数据, 采用 Penman-Monteith 法计算了马拉河流域四个季节 (短旱季、长雨季、长旱季、短雨季) 植被生态需水量的时空变化特征。在此基础上, 使用支持向量机 (SVM)、随机森林 (RF) 和卷积神经网络 (CNN) 3 种机器学习方法与 7 个环境因子 (气温、降水、10 m 风速、LAI、太阳辐射、相对湿度、地形) 建立了回归模型, 分别估算了 2011—2020 年逐年不同季节的植被生态需水量, 并与 Penman-Monteith 法计算结果进行时间序列拟合度和空间相似性的比较。结果表明: 马拉河流域植被生态需水量在过去 40 年所有季节都呈现为波动变化, 植被生态需水量长雨季 > 长旱季 > 短雨季 > 短旱季, 长雨季的植被生态需水量约为短旱季的 1.5 倍。不同季节均呈现出上下游高、中游低的植被生态需水量空间分布格局。LAI 为最大的正影响因子, 风速为最大的负影响因子。就不同方法估算的植被生态需水量准确性而言, RF 表现最为优异, 主要体现在最值估算误差最小, 时间变化序列的拟合度最高, 空间分布最为相似, 相对误差最小, 而 SVM 的预测结果相对最差。RF 是相对最适用于马拉河流域植被生态需水量估算的算法。使用 3 种不同的机器学习方法估算马拉河流域不同季节植被生态需水量, 并对结果进行比较, 可为生态需水的估算提供技术参考。

关键词: 马拉河流域; 植被生态需水量; 支持向量机 (SVM); 随机森林 (RF); 卷积神经网络 (CNN); 估算

Characteristics and estimation of vegetation ecological water demand in the Mara River Basin

ZHU Wanyi^{1,2}, ZHANG Zhenke^{1,2,*}, GUO Xinya^{1,2}, FENG Shouming^{1,2}, JIANG Daliang^{1,2}, JIANG Fei¹

¹ School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China

² Institute of African Studies, Nanjing University, Nanjing 210023, China

Abstract: Ecological water demand (EWD) is the basis for regionally ecological water control and environment restoration. The Mara River Basin has a world famous ecosystem, and the vegetation EWD accounts for a large part of the total water demand in the Basin. Based on the ERA5 meteorological data, leaf index data (LAI) data from 1980 to 2020, and HSWD world soil data, we calculated the vegetation EWD temporal and spatial variation characteristics in the four seasons (short dry season, long rainy season, long dry season, and short rainy season) of Mara River Basin by using the Penman-Monteith method. Then, three different machine learning methods of support vector machine (SVM), random forest (RF), convolutional neural network (CNN) and seven environmental factors (temperature, precipitation, 10 m wind speed, LAI, solar radiation, relative humidity and terrain) were used to establish regression models. Thus, the vegetation EWD in different seasons from 2011 to 2020 was estimated by the machine learning regression models, and the estimation results were compared with the results calculated by the Penman-Monteith method in the fitting degree of time change series and the

基金项目: 国家重点研发计划项目 (2018YFE0105900)

收稿日期: 2022-09-18; 网络出版日期: 2023-05-08

* 通讯作者 Corresponding author. E-mail: zhangzk@nju.edu.cn

similarity in the space. The results showed that the EWD of vegetation in the Mara River Basin showed fluctuating trends in all seasons in the past 40 years. The vegetation EWD in different seasons from the most to the least was long rainy season, long dry season, short rainy season, and short dry season, the EWD in the long rainy season was about 1.5 times of that in the short dry season. The EWD in the upstream and downstream was larger than that in the midstream in all the seasons. LAI was the largest positive influence factor and wind speed was the largest negative influence factor. In terms of the accuracy of vegetation EWD estimation, RF performed the best, which mainly reflected in the minimum estimation error of the maximum, mean and minimum values, and the highest fitting degree of time change series. In the space, the best performance reflected in the most similar spatial distribution, and the smallest relative error. However, the estimation of SVM was relatively the worst. RF was the most suitable method for estimating vegetation EWD in the Mara River Basin. In this study, the three different machine learning methods were used to estimate the vegetation EWD in different seasons in the Mara River Basin, and the results can provide technical reference for the EWD estimation.

Key Words: Mara River Basin; vegetation ecological water demand; support vector machine (SVM); random forest (RF); convolution neural network (CNN); estimation

生态环境是人类生存和发展的基本要求,水对维持生态环境具有关键作用^[1]。水资源的日益短缺加剧了水资源的供需矛盾,目前大多数研究只关注工业和生活的需水要求,忽视了生态用水需求^[2]。植被是生态系统的重要组成部分,是生态系统中物质循环与能量流动的中枢,是连接生物圈、水圈、大气圈的重要一环^[3]。植被生长必然消耗一定的水资源,其需水量由于植被类型、生长状况差异等而呈现出时空变化差异^[4]。研究植被生态需水量的时空变化并建立模型预测其变化对区域水资源规划和管理、灌溉调度、植被恢复等方面具有重要意义。

植被生态需水量普遍定义为保障植被正常生长、发育或者维持植被生态系统健康并发挥正常生态系统服务功能而需要消耗的最低水量^[5],是植被生态系统是否会发生退化或已退化植被生态系统能否重建的关键^[6]。在植被生态需水计算方面,不同学者根据研究区植被状况、水资源现状、气象监测数据质量选择了不同的研究方法。在气象监测数据相对完整,植被状况较为单一的情况下,学者多采用面积定额法^[7];在植被生长主要依靠地下水的干旱区,潜水蒸发模型是更为常用的计算方法^[4];此外还有使用水量平衡法计算植被恢复期生态需水量大小^[8],但是该方法未考虑植被生理特性,导致其在应用上具有一定的局限性;在气象站点监测数据较少的情况下,一般多采用 RS、GIS 技术结合 Penman-Monteith 公式^[9]计算植被生态需水量,其中 Penman-Monteith 公式基于植被的潜在蒸散量并结合土壤水分和植被面积以及植被特性等影响因素对生态需水量进行计算,适用性较广。然而植被生态需水受气象和环境的影响而发生变化,根据“地理环境越相似,地理特征越相近”的规律可知,可通过探索建立植被生态需水量与气象、环境因子关系的模型来估算植被生态需水量^[10],遗憾的是目前学术界在该领域鲜有研究。机器学习算法是目前应用广泛、预测性能较为稳定的算法之一。基于机器学习算法的模型是从观测数据(样本)中寻找统计规则,并利用这些规则建立模型对目标数据进行估算,常用的机器学习算法包括支持向量机(SVM)、随机森林(RF)、卷积神经网络(CNN)。机器学习模型以相同空间结构的时间序列作为模型的输入,同时建模过程中,采用对关键参数进行优化以减少计算量^[11],能更为全面地表达数据的细节和突变,从而提高预测精度^[12]。然而对于哪种模型更适合生态需水量的估算还没有一致的意见。因此,本文拟分别利用 SVM、RF 和 CNN 这三种机器学习算法开展植被生态需水量建模对比,探索不同机器学习算法对植被生态需水量的估算能力。

不同地区植被生态需水已存在大量的研究,但是有关非洲植被生态需水的研究起步较晚,研究也较少,且集中在南非地区^[13]、非洲热带稀疏草原气候区^[5]等有限研究区域,鲜有关于马拉河植被生态需水的研究。马拉河流域具有丰富的生物多样性和农业生态系统多样性,是世界著名的马赛马拉国家保护区和塞伦盖蒂国家公园的一部分,植被生态需水量占据了生态需水乃至水资源需求的绝大部分^[14]。然而由于气象监测数据和

植被统计数据不足,该流域的植被生态需水量时空动态变化特征尚未得到研究。遥感监测数据和全球再分析天气数据集提供了以低成本研究植被生态需水动态变化的渠道,为更好地理解影响马拉河流域水资源分配和水量平衡关键因素提供了机会。

在此背景下,本文使用 1980—2020 年 ERA5-Land 月平均再分析气象数据、叶面积指数(LAI)以及世界土壤数据库(HSWD)土壤数据,结合 Penman-Monteith 法计算了马拉河流域短旱季、长雨季、长旱季、短雨季的植被生态需水量。在此基础上,利用支持向量机(SVM)、随机森林(RF)、卷积神经网络(CNN)估算了 2011—2020 年马拉河流域不同季节植被的生态需水量。本研究应用多方法对马拉河流域植被生态需水量展开对比研究,旨在为估算植被生态需水的变化提供一种新的智能估算手段,从而更好地为该区域植被建设及生态环境保护提供综合决策,以促进该地区水资源与生态环境的可持续发展。

1 研究区概况与数据源

1.1 研究区域概况

马拉河发源于肯尼亚马乌森林,流经马赛马拉国家保护区、盖伦盖蒂国际公园,最终在坦桑尼亚穆索马农村流入维多利亚湖(图 1),流域总面积 13750 km²(33°88′—35°90′ E, 0°28′—1°97′ S)。流域上游为国家级森林保护区和农垦区;中游主要为草原和牧场,包含两个国家级国家野生动物保护区——肯尼亚的马赛马拉国家保护区和坦桑尼亚的塞伦盖蒂国家保护区,下游为马拉湿地和其他稀疏草原区^[15]。

马拉河流域内的降雨量变化很大,最高和最低年平均降雨量分别为 1750mm/a(马乌森林地区)和 600 mm/a(流域东南部)。流域降水呈双峰型,存在两个雨季,长雨季(3—5 月)和短雨季(10—12 月);两个旱季,长旱季(6—9 月)和短旱季(1—2 月)。年平均气温约为 25.5℃,温度由南向北递减,地势北高南低^[16]。土壤以火山灰土、黑土和粘磐土为主(图 1)。从流域上游到下游植被分布区变化明显,从高山森林、分散林地到耕地,再到大草原、灌木丛和灌草混交区,主要植被类型为森林、农作物、草地、灌木等,是众多野生动物生存的家园,具有重要的生态价值。

1.2 数据源

数据可分为两部分,第一部分包括 1980—2020 年逐月气象因子数据(地面净辐射、平均气温、10 m 平均风速、平均空气湿度)、逐月叶面积指数(LAI)和土壤属性数据,该部分数据用于计算植被生态需水量。气象因子和叶面积指数来源于 ERA5-Land 月平均再分析数据,土壤属性数据来源于世界土壤数据库(HSWD)。第二部分的地形因子数据来源于 STRM 全球地形数据库,空间分辨率为 90 m×90 m。为使两部分数据空间分辨率保持一致,将所获取数据使用双线性内插法重采样为 1 km×1 km。

2 研究方法

2.1 植被生态需水量

植被生态需水主要受气候因子、土壤水分含量和植被类型差异三个因素的影响,因此,植被生态系统所需消耗的水量可以表示为^[17]:

$$ET_q = ET \times S \times 10000 \quad (1)$$

$$ET = ET_0 \times K_c \times K_s \quad (2)$$

式中, ET_q 为生态需水量(m³); S 为植被面积(m²); ET 为单位时间、单位面积内生态需水定额(mm); ET_0 为潜在蒸散(mm),按 Penman-Monteith 公式计算得到; K_c 为植被系数,与植被种类和生长状况有关,其与 LAI 具有高度相关性,拟合模型为^[18]:

$$K_c = 0.4280 LAI^{0.6988} \quad (3)$$

式中, K_s 为土壤水分系数,与土壤质地和土壤含水量有关。土壤含水量低于土壤临界含水量(S^*),土壤水分含量就成为林草地实际蒸散量的主要限制因素。根据土壤水分有效性的划分,暂时凋萎含水量(S_l)是保

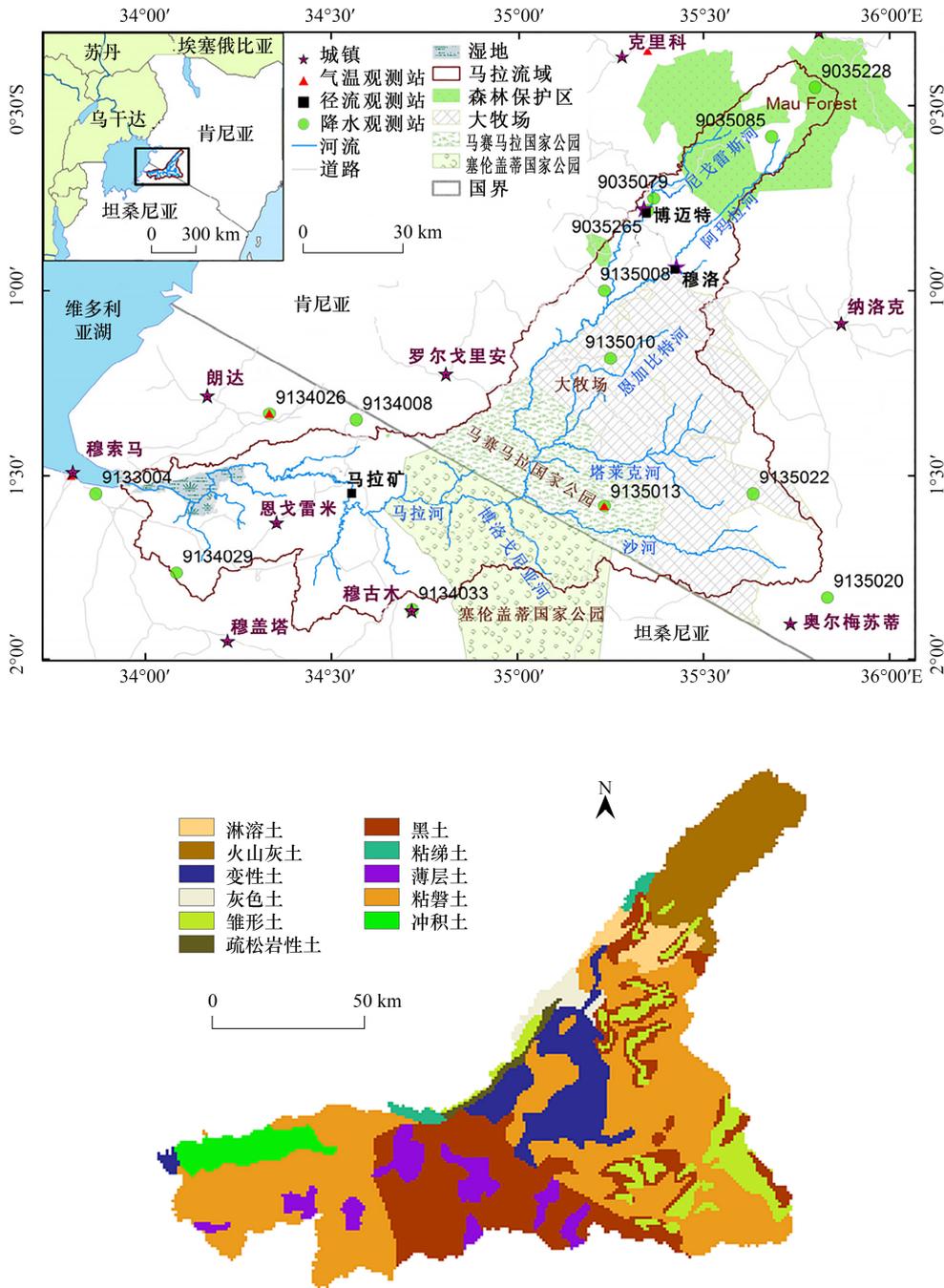


图1 马拉河流域与土壤类型分布图

Fig.1 The location of Mara River Basin and soil type distribution

证林草地基本生存的下限,是最小生态需水量。根据闵庆文等人^[19]的研究,本文使用凋萎系数的120%作为暂时凋萎系数,采用 Saxton 计算研究区不同质地类型各种状态下的土壤水分,然后根据 Jensen 公式^[20]计算最小土壤水分系数(K_s):

$$K_s = \ln \left[\frac{(S - S_w)}{(S^* - S_w)} \times 100 + 1 \right] / \ln 101 \quad (4)$$

式中, S 为土壤实际含水量, S_w 为土壤凋萎含水量, S^* 为土壤临界含水量。将 $S = S_l$ 代入上式即可得植被最小生态需水的土壤水分系数(表1)。

表 1 不同土壤质地类型的土壤水分系数

Table 1 Soil moisture coefficients of different soil texture types

| 土壤类型 Soil types | 土壤凋萎含水量 $S_w / (\text{m}^3/\text{m}^3)$ Wilting soil water content | 土壤临界含水量 $S^* / (\text{m}^3/\text{m}^3)$ Critical soil water content | K_s |
|-----------------------------|---|--|--------|
| 淋溶土 Alfisols | 0.3717 | 0.4550 | 0.6825 |
| 火山灰土 Volcanic ash soil | 0.3317 | 0.439356 | 0.6063 |
| 变性土 Vertisol | 0.3813 | 0.463422 | 0.6905 |
| 灰色土 Ashen-grey soil | 0.3416 | 0.460146 | 0.5926 |
| 锥形土 Prototype soil | 0.3560 | 0.435576 | 0.6827 |
| 疏松岩性土 Loose lithologic soil | 0.3458 | 0.4134 | 0.7107 |
| 黑土 Black soil | 0.3838 | 0.4481 | 0.5602 |
| 粘缋土 Nitosol | 0.3826 | 0.4641 | 0.6930 |
| 薄层土 Ranker | 0.3393 | 0.4701 | 0.5715 |
| 粘磐土 Planosol | 0.3534 | 0.4290 | 0.6920 |
| 冲积土 Alluvial soil | 0.3570 | 0.4363 | 0.6841 |

K_s : 最小土壤水分系数

2.2 植被生态需水量估算

为建立环境因子与马拉河流域不同季节植被生态需水量的模型, 本文将 1980—2010 年不同季节植被生态需水与其相关的环境因子栅格数据, 包括 1980—2010 年不同季节植被生态需水量、气温、降水、10 m 风速、LAI、太阳辐射、相对湿度以及地形, 使用皮尔逊相关性, 利用 SVM、RF、神经网络重要性得分来评价不同环境因子对植被生态量的影响程度, 以此删选因子来建立回归模型。

2.2.1 支持向量机(SVM)

支持向量机(SVM)是寻找使类别间距最大化的最佳超平面, 并利用超平面上支持向量来构建模型^[21]。在解决线性、非线性预测中都具有较好的适应性。对于一组线性可分的向量 $x_i (i = 1, 2, \dots, n)$, x_i 表示环境因子, 通过确定超平面最大间距来对植被生态需水量进行分类, 其最大间距为 $\frac{1}{2} \|s\|^2$ 。对于线性不可分离数据, 利用松弛变量 δ_i 来控制分类误差, 相应约束条件为 $y_i (s \times x_i + b) \geq 1 - \delta_i$ 。此外, 引入 $\Delta(0, 1)$ 考虑错误的分类。本文使用广泛应用且性能相对最佳的径向基函数(RBF)核函数作为 SVM 的核函数^[22], SVM 算法借助 Python 中的 libsvm 库实现。本文对 1980—2020 年不同季节植被生态需水量与环境数据在二维空间上进行了线性回归, 模型估算结果通过纳什效率系数来评价, 纳什效率系数在区间(0, 1)范围内, 表明模型整体结果可信。

2.2.2 随机森林(RF)

随机森林算法(RF)是将 Bagging、随机子空间与决策树方法相结合^[23—24]的方法, 算法融入自助聚集(bootstrap aggregating)方法生成子集, 据此构建多棵决策树。在对决策树的每个节点进行分裂时, 引入随机子空间, 从全部 K 个特征中均匀随机地抽取一个特征子集, 从该子集中选择一个最优分裂特征^[25]。最后采用多棵决策树并行的方式, 求取多棵决策树结果的均值作为最终计算结果。本文的 RF 算法通过调用 R 语言平台 Random Forest 包实现。本文按 3:7 的比例划分训练集和测试集, 随机使用 10 年不同季节的植被生态需水量与环境数据作为训练集, 训练 RF 估算模型, 剩下 31 年数据用于评估模型运算能力。RF 中决策树的数量是直接关系到模型运行准确率的参数。经过大量的实验, 当决策树为 300 时, 模型误差已保持平稳, 模型运行结果使用 R^2 评价, R^2 越高, 模型效果越好。

2.2.3 卷积神经网络(CNN)

CNN 可用于处理具有已知网格拓扑结构的数据^[26], 其核心结构为卷积层、池化层。卷积层中的卷积核可用于提取数据的初级特征, 池化操作可以降低卷积层输出特征图的维度, 层间参数也会相应的减少, 从而达到

降低模型复杂度,提高计算效率的目的。

本文将原始数据映射为二维 CNN 数据,特征映射结构激活函数采用 Sigmoid 函数,采用最大池采样法减少数据量,降低时间复杂度,最后使用 Softmax 回归进行植被生态需水量的估算。本文的 CNN 算法借助 Python 中 Keras 和 TensorFlow 库实现。本文按 1:1:2 的比例划分训练集、验证集和测试集,随机使用 8 年不同季节的环境因子和植被用作训练集,输入 CNN 中进行模型训练,随机 8 年数据作为验证集,用于模型估算能力的初步评估,剩余 15 年数据用于评估模型的最终泛化能力。模型结果使用 RMSE 来评价, RMSEF 越小,模型拟合度越高。

在所有模型建立到最佳参数后,将 2011—2020 年不同季节所有环境因子栅格数据导入模型分别输出逐年不同季节的植被生态需水量,同时将模型估算出的植被生态需水量与 Penman-Monteith 公式计算出的植被生态需水量进行比较,以分析不同模型估算结果的准确性。

3 研究结果

3.1 植被生态需水量时空变化

马拉河流域植被生态需水量季节差异较大。如图 2 所示,1980—2020 年短旱季生态需水量波动变化幅度最大;长雨季植被生态需水量相对最为平缓。以 2000 年为分界点,长旱季和短雨季的生态需水量均呈现为先快速波动后逐渐平稳的变化趋势。雨季的植被生态需水量一般高于旱季。1987 年各季节植被生态需水量极少;1990 年长旱季、短旱季生态需水量大于短雨季、长雨季。2000 年以后,长旱季植被生态需水量逐渐多于长雨季。各个季节植被生态需水量最小值和最大值的变化趋势与平均值的变化趋势一致。

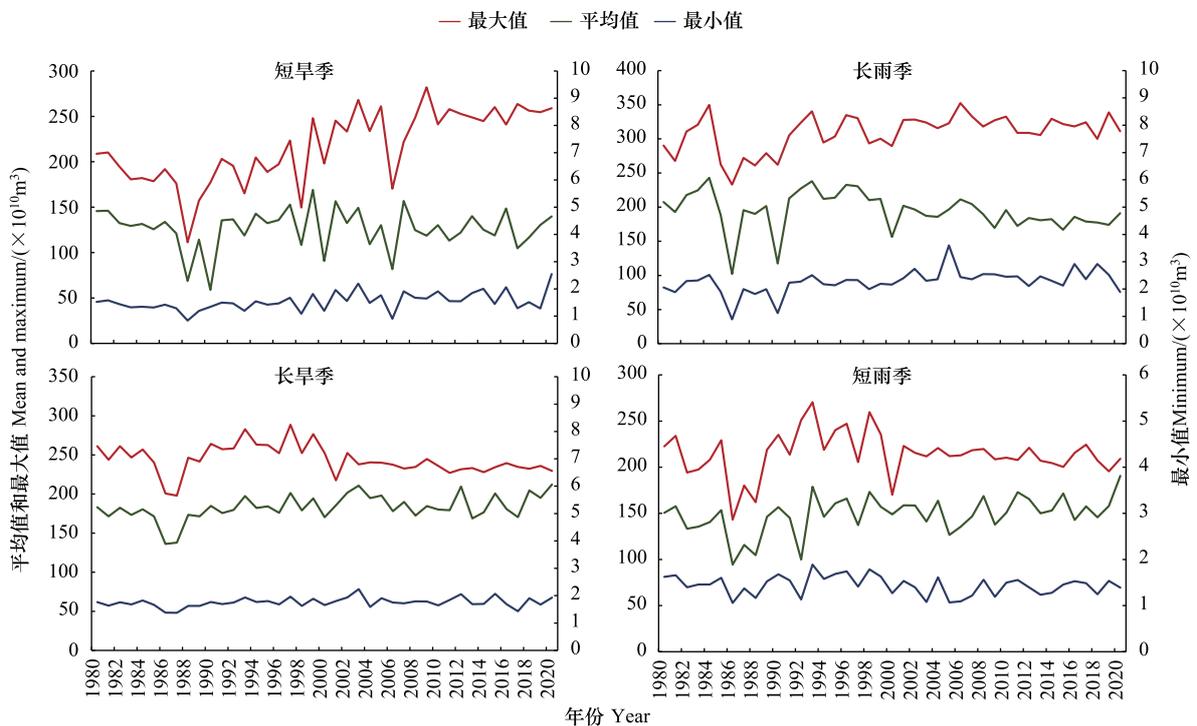


图 2 马拉河流域 1980—2020 年植被生态需水量季节变化特征

Fig.2 Seasonal variation characteristics of vegetation ecological water demand in the Mara River Basin from 1980 to 2020

为便于比较不同季节植被生态需水量的差异,采用等间距法对马拉河流域不同季节的生态需水量进行分类,间距约为 $5.0 \times 10^{11} \text{ m}^3$ (图 3)。马拉河流域所有季节植被生态需水量均表现为上游较高,中游东部低的空间分布格局。上游马乌森林区植被生态需水量最大,中游大牧场东部生态需水量最小。从植被生态需水量的

季节性变化来看,长雨季需水量最大,短雨季需水量最少。短旱季中游和下游广大区域植被生态需水量为 $1.0\text{—}1.5\times 10^{12}\text{ m}^3$,长旱季中游为 $1.0\text{—}1.5\times 10^{12}\text{ m}^3$,下游广大区域为 $1.5\text{—}2.0\times 10^{12}\text{ m}^3$,旱季植被生长状态较为一致,植被生态需水量也较为一致。

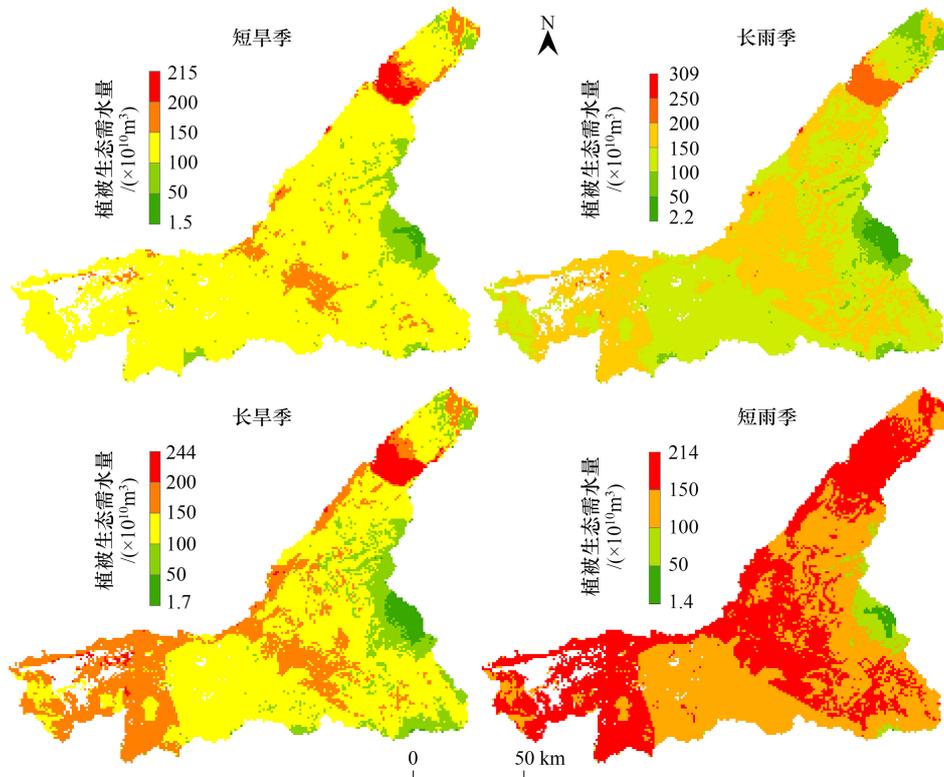


图3 马拉河流域 1980—2020 年不同季节植被平均生态需水量空间分布

Fig.3 Spatial distribution of average vegetation ecological water demand in the Mara River Basin in different seasons from 1980 to 2020

3.2 植被生态需水量估算

3.2.1 环境因子重要性得分

需水量是植被耗水分析的重要指标,与环境因子关系息息相关。表2显示了7个环境因子(降水、气温、10 m 风速、LAI、太阳辐射、相对湿度、地形)的SVM、RF和神经网络重要性得分以及Person相关系数(r),以分析环境因子对植被生态需水量的影响程度,判断所选取环境因子进行估算的可行性。

根据Person相关分析结果,所有环境因子的相关系数均通过了显著性检验,表明所选取的因子与植被生态需水都具有较好的相关性。SVM模型重要性得分结果显示环境因子对植被生态需水量的影响相差不大,约为13%。RF重要性得分显示除了LAI重要性得分大于50以外,其余因子多约为20。神经网络重要性得分显示除LAI大于0.4外,其余因子多为0.1左右。由此可知,本文选取的环境因子较为合理,可作为植被生态需水的回归影响因子进行估算分析。

3.2.2 植被生态需水量估算结果

本文使用1980—2010年逐月植被生态需水量与环境因子数据,结合SVM、RF、CNN估算了马拉河流域2011—2020年逐年各个季节的植被生态需水量。短旱季、长雨季、长旱季、短雨季的植被生态需求量SVM的纳什效率系数分别为0.0309、0.0316、0.0322、0.0310;RF的 R^2 分别为0.905、0.942、0.902、0.949;CNN的RMSE分别为0.152、0.142、0.116、0.115。所有模型的估算结果均较好。

图4显示了马拉河流域2011—2020年植被生态需水量SVM、RF和CNN三种算法估算的不同季节流域最大值、平均值、最小值与Penman-Monteith计算值的时间序列对比情况。从起伏性来看,SVM的估算较为平

缓,RF 估算结果波动幅度最大。从估算差值来看,SVM 最大值、最小值、平均值估算值的平均相对误差分别为 0.52%、2.64%、1.64%;RF 分别为 0.21%、1.13%、0.64%;CNN 分别为 0.26%、0.17%、0.73%;SVM 的估算差值最大,RF 的差值最小。从与 2011—2020 年线性变化趋势来看,RF 的估算结果与 1980—2020 年的线性趋势最为拟合,SVM 拟合度最差。整体而言,RF 的估算结果最好,SVM 最差。

表 2 环境因子的重要性得分

Table 2 Importance score of environmental factors

| 季节 Seasons | 环境因子 Environmental factors | SVM | RF | 神经网络 Neural network | 相关系数 <i>r</i> |
|---------------------------|-------------------------------|------|-------|------------------------|------------------|
| 短旱季 Short dry season | 降水 | 0.13 | 18.37 | 0.049 | 0.705(+) |
| | 气温 | 0.15 | 22.73 | 0.196 | 0.113(+) |
| | 风速 | 0.20 | 23.71 | 0.040 | 0.500(-) |
| | LAI | 0.11 | 55.49 | 0.440 | 0.788(+) |
| | 太阳辐射 | 0.14 | 10.99 | 0.086 | 0.206(+) |
| | 相对湿度 | 0.14 | 20.90 | 0.133 | 0.121(+) |
| | 地形 | 0.13 | 21.37 | 0.055 | 0.146(-) |
| 长雨季 Long rainy season | 降水 | 0.13 | 17.97 | 0.088 | 0.701(+) |
| | 气温 | 0.13 | 20.71 | 0.161 | 0.268(+) |
| | 风速 | 0.13 | 24.91 | 0.126 | 0.175(-) |
| | LAI | 0.24 | 50.63 | 0.294 | 0.778(+) |
| | 太阳辐射 | 0.13 | 11.19 | 0.103 | 0.154(-) |
| | 相对湿度 | 0.14 | 17.92 | 0.098 | 0.209(+) |
| | 地形 | 0.09 | 23.89 | 0.130 | 0.302(-) |
| 长旱季 Long dry season | 降水 | 0.10 | 20.08 | 0.067 | 0.506(+) |
| | 气温 | 0.12 | 23.53 | 0.125 | 0.508(+) |
| | 风速 | 0.10 | 21.34 | 0.044 | 0.359(-) |
| | LAI | 0.31 | 51.38 | 0.466 | 0.700(+) |
| | 太阳辐射 | 0.10 | 15.04 | 0.031 | 0.290(+) |
| | 相对湿度 | 0.17 | 16.80 | 0.075 | 0.428(+) |
| | 地形 | 0.10 | 23.23 | 0.193 | 0.532(-) |
| 短雨季 Short rainy season | 降水 | 0.11 | 16.22 | 0.037 | 0.700(+) |
| | 气温 | 0.14 | 15.59 | 0.122 | 0.269(+) |
| | 风速 | 0.19 | 28.64 | 0.127 | 0.806(-) |
| | LAI | 0.20 | 58.66 | 0.405 | 0.811(+) |
| | 太阳辐射 | 0.10 | 10.40 | 0.127 | 0.808(+) |
| | 相对湿度 | 0.12 | 17.72 | 0.092 | 0.111(+) |
| | 地形 | 0.14 | 26.26 | 0.091 | 0.293(-) |

SVM:支持向量机 Support Vector Machines;RF:随机森林 Random forest

图 5 显示了马拉河流域 2011—2020 年植被平均生态需水量的 Penman-Monteith 计算值与 RF、SVM、CNN 三种算法估算的 2011—2020 年植被平均生态需水量空间分布情况。三种算法反演的空间分布情况与公式计算值相似度很高,但也存在一些不同。

RF 的空间估算结果与计算结果很接近,仅最小值的分布区预测结果相差较大,较公式计算值少了 3.09%,不同植被生态需水量分区的面积与计算值相差的最少。短旱季仅上游马乌森林处植被生态需水量为 $(1.5-2.0) \times 10^{12} \text{ m}^3$ 范围稍增加,下游 $(0.5-1.0) \times 10^{12} \text{ m}^3$ 面积减少。长雨季流域内部分植被生态需水量为 $(2.0-2.5) \times 10^{12} \text{ m}^3$ 的区域变为 $(1.5-2.0) \times 10^{12} \text{ m}^3$ 。长旱季需水量为 $(1.0-1.5) \times 10^{12} \text{ m}^3$ 的面积减少较多,中游需水量为 $(0-5.0) \times 10^{11} \text{ m}^3$ 的区域面积扩大较多。短雨季植被生态需水量为 $(1.0-1.5) \times 10^{12} \text{ m}^3$ 的在上游和下游增长较多;中游东部部分计算值为 $(0-5.0) \times 10^{11} \text{ m}^3$ 的区域部分变为 $(0.5-1.0) \times 10^{12} \text{ m}^3$ 。

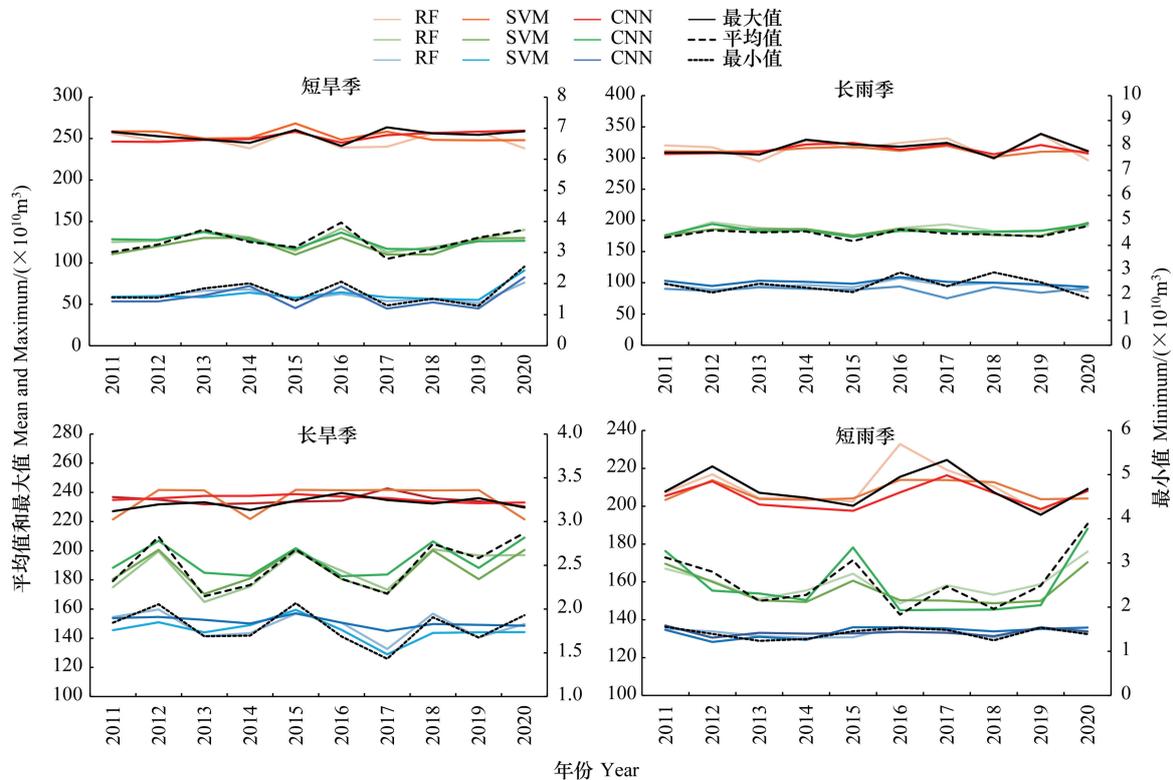


图4 马拉河流域 2011—2020 年不同季节植被生态需水量最小值、平均值、最大值估算结果

Fig.4 Estimation results of minimum, average and maximum vegetation ecological water demand in the Mara River Basin in different seasons from 2011 to 2020

SVM 不同季节植被生态需水量的估算值与计算值总相对误差为 6.68%, 雨季的预测结果较旱季更好。短旱季中游东部大面积植被生态需水量为 $(0.5-1.0) \times 10^{12} \text{ m}^3$ 变为 $(1.0-1.5) \times 10^{12} \text{ m}^3$ 。长雨季植被生态需水量为 $(2.0-2.5) \times 10^{12} \text{ m}^3$ 的区域扩大, 流域东部应为 $(0.5-1.0) \times 10^{12} \text{ m}^3$ 的区域大部分变为 $(1.0-1.5) \times 10^{12} \text{ m}^3$ 。长旱季中游需水量为 $(1.0-1.5) \times 10^{12} \text{ m}^3$ 区域分布区减少较多, 上游和下游生态需水量为 $(1.0-1.5) \times 10^{12} \text{ m}^3$ 分布区增大。短雨季流域中游东部生态需水量为 $(0-5.0) \times 10^{11} \text{ m}^3$ 的区域扩大; $(1.0-1.5) \times 10^{12} \text{ m}^3$ 的区域在中游和下游略有减少, 而 $(1.5-2.0) \times 10^{12} \text{ m}^3$ 的区域在上游减少。

CNN 的不同季节植被生态需水量估算值与计算值的总相对误差为 4.33%, 短旱季中游东部植被生态需水量为 $(0-5.0) \times 10^{11} \text{ m}^3$ 的范围扩大较多, 中游为 $(0.5-1.0) \times 10^{12} \text{ m}^3$ 的区域略有扩大。长雨季中游为 $(2.0-2.5) \times 10^{12} \text{ m}^3$ 的分布区较真实情况更为分散, 上游为 $(1.5-2.0) \times 10^{12} \text{ m}^3$ 分布区面积扩大且更为集中。长旱季需水量为 $(0-5.0) \times 10^{11} \text{ m}^3$ 的分布区更大, $(1.0-1.5) \times 10^{12} \text{ m}^3$ 的区域面积减小, 尤其是在马乌森林上游区域。短雨季需水量为 $(1.0-1.5) \times 10^{12} \text{ m}^3$ 区域面积减少最多, 流域东部需水量为 $(0-5.0) \times 10^{11} \text{ m}^3$ 面积增加较多。

4 讨论

马拉河流域上世纪为促进经济发展, 大规模发展农业和旅游业, 植被退化严重。作为生态需水研究的重要内容, 植被生态需水可为区域植被生态恢复和重建提供重要科学依据^[27], 因此开展该流域植被生态需水量研究显得尤为重要。

4.1 马拉河流域植被生态需水量变化特征

马拉河流域降水呈现双峰型, 降水集中在长雨季, 长旱季降水较少。植被常受降水时空分布不均的限制,

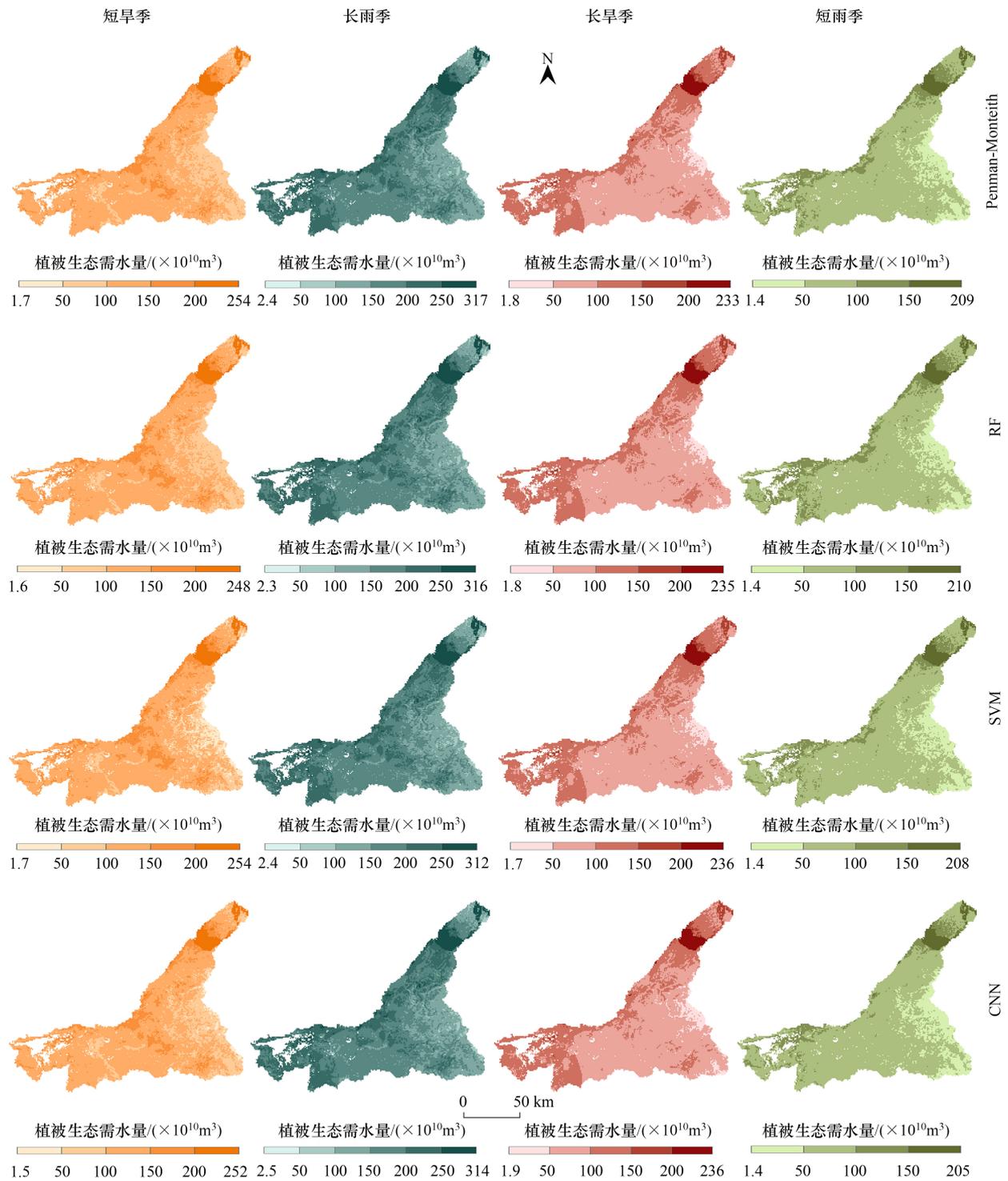


图5 马拉河流域 2011—2020 年不同季节植被生态需水量空间分布估算结果

Fig.5 Estimation results of spatial distribution of vegetation ecological water demand in the Mara River Basin in different seasons from 2011 to 2020

生长状态季节差异很大,形成了独特的需水习性^[14]。植被生态需水量的季节大小关系为短旱季<短雨季<长旱季<长雨季。长雨季降水多,植被长势最好,植被系数最大,植被生态需水最大。短旱季热量和土壤湿度较大,光照充足,土壤蒸发和叶面蒸腾作用加快,蒸散量较高^[28],且持续时间最短,因而植被生态需水量最小。

马拉河流域上游为马乌森林和农作物,中游主要为草地和灌木,下游为湿地森林。不同植被类型的潜在蒸散相差大导致不同植被类型的生态需水量相差较大。湿地森林的潜在蒸散比马乌森林高 24%—29%,农业的潜在蒸散比灌木和草地高 52%^[28],因而森林的植被生态需水量大于农作物,也大于草地和灌木。中游的土壤水分系数小于上游和下游。上下游植被潜在蒸散系数、土壤水分系数大,中游小的空间分布使得马拉河流域植被生态需水量呈现为上下游高,中游较低的空间分布格局。

4.2 环境因子对植被生态需水量影响

生态需水量与环境因子有很好的相关性,最主要的正影响因子为 LAI,负影响因子为风速。LAI 与潜在蒸散存在强正相关性,即 LAI 的增长意味着植被生长态势变好,土壤水分充足,植被蒸散量变大^[29],植被生态需水量增加。降水的增加可以促进植被生长,使植被生态需水量增加。气温上升可以促进潜在蒸散的增加,增加植被生态需水量。马拉河流域地形为上游高下游低,高低不利于水分集聚^[30],因而地形对植被生态需水量的空间分布呈现负相关关系。雨季太阳辐射的增加会提高植被的光合作用,促进有机质分解提高土壤养分,进而促进植被生长;旱季太阳辐射的增加会加快蒸散,使土壤干化,抑制大部分区域植被生长,因而太阳辐射在雨季为正相关影响,旱季为负相关影响。相对湿度与植被潜在蒸散之间呈现为正相关关系^[31],旱季的正相关性大于雨季,说明旱季空气湿度的增加能更明显地促进植被生长。风速的增加和风强度的加大可能会引起气孔关闭,蒸腾速率变慢,从而又引起液流速率加快,疾风将气孔外相对湿度较高的空气吹走,使蒸汽扩散层变薄或消失,阻碍了植被的蒸腾作用^[32],使得植被生态需水量减少。马拉河流域风速较大,植被蒸腾速率变慢,潜在蒸散量减少,呈现为明显的负相关关系。

4.3 植被生态需水量估算

本文使用了 3 种机器学习方法对马拉河流域不同季节植被生态需水量进行了估算,机器学习方法是目前应用最为广泛的学习算法,能充分高效挖掘不完备信息进行优质建模训练^[33]。而传统植被生态需水量计算方法其基本步骤均为先划分研究区的植被类型,并确定其范围和面积,然后根据不同植被类型的生态需水定额进行计算。以应用最为广泛的面积定额法为例,该方法需要进行耗水实验以获得植被需水定额。但是在缺乏实验条件的区域该方法可行性不大。此外,面积定额法也应用在对植被未来生态需水量的预测方面,即通过规划文件中规划的未来植被面积计算植被生态需水量,该预测方法虽简单便捷,但忽视了其他环境变化可能对植被生态需水量变化的影响。温家鸣等^[34]使用 BPNN 预测了桂林植被生态需水量,证实了神经网络可用于预测植被生态需水量。

本文 SVM 模型选用径向基核函数,估算结果较好,但仍是三种方法中误差最大的算法。基于栅格数据计算得出的马拉河流域植被生态需水量结构较为复杂,使用在线性预测表现结果优良的径向基核函数会出现高值和低值预测值偏差较大,更偏向于平均值^[35]。图 4 的最大值、最小值、平均值平稳的变化趋势也验证了这一缺点。二维 CNN 的估算结果较 SVM 有较大幅度的提升,误差少了 2.35%。神经网络的训练结果受池化层窗口的大小、平移不变性影响较大^[36]。池化层的存在导致 CNN 算法会使一些有价值的信息丢失,同时也会忽略掉整体与部分之间的关联。已有研究表明,三维 CNN 可以进一步提高 CNN 模型的预测精度^[37],但同时会使模型运行的时间大幅度增加。RF 的估算结果与 Penman-Monteith 公式计算结果最为接近。RF 既能处理离散数据又能处理连续型数据,对数据集适应性强且具有较强的稳定性,适合用于连续型变量的建模。此外,郭巧玲等^[38]、Saha 等^[39]等学者使用 SVM 和更为传统的预测模型,如线性等,结果均显示 RF 的性能优于 SVM 等更传统的数据驱动模型,这也证实了 RF 在预测马拉河流域植被生态需水量方面的可行性。

5 结论

本文使用 Penman-Monteith 公式计算了马拉河流域不同季节 1980—2020 年植被生态需水量,并在此基础上分析了环境因子(气温、降水、10 m 风速、LAI、太阳辐射、相对湿度、地形)对生态需水的影响,最后使用了 RF、SVM、CNN 算法估算了 2011—2020 年逐年不同季节的植被生态需水量。结果表明:

(1) 1980—2020 年马拉河流域所有季节植被生态需水量处于波动变化中。长雨季植被生态需水量最大,短旱季生态需水量最小,平均值相差约 1.5 倍。所有季节植被生态需水量均呈现出上游和下游较高、中游较低的空间分布格局。

(2) 气温、降水、10 m 风速、LAI、太阳辐射、相对湿度、地形这 7 个环境因子与植被生态需水量都具有较好的相关性。其中 LAI 对植被生态需水量正影响作用最大,而风速的负影响最大。

(3) 根据 RF、SVM、CNN 算法估算结果,从时间序列趋势和最小值、平均值、最大值来看,RF 算法的估算结果最好,SVM 估算结果最差。从空间分布情况来看,RF 和 CNN 的估算结果与公式计算值更为吻合。因而本研究认为 RF 是相对最适用于马拉河流域植被生态需水量估算的算法。

本文利用机器学习算法,挖掘了植被生态需水量与环境之间更深层次的因果关系,验证了机器算法可用于植被生态需水量的估算,为后续植被生态需水量的预测提供强有力的支撑,也为国内缺乏植被观测数据的地区提供一种新的估算思路。

参考文献(References):

- [1] Han Y Q. Analysis on ecological environment construction and policy in Yi-jin-huo-luo-qi. *Environment protection of Neimenggu*, 2005, 17(4): 50-52.
- [2] Lu C Y, Feng J S, Chen Y D, Liu W, Lin Z C, Yan S C. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(4): 925-938.
- [3] 王改玲, 王青杵, 石生新. 山西省永定河流域林草植被生态需水研究. *自然资源学报*, 2013, 28(10): 1743-1753.
- [4] Ye Z X, Chen Y N, Li W H. Ecological water demand of natural vegetation in the lower Tarim River. *Journal of Geographical Sciences*, 2010, 20(2): 261-272.
- [5] Wilson T B, Witkowski E.. Water requirements for germination and early seedling establishment in four African savanna woody plant species. *Journal of Arid Environments*, 1998, 38(4): 541-550.
- [6] Thevs N, Peng H Y, Rozi A, Zerbe S, Abdusalih N. Water allocation and water consumption of irrigated agriculture and natural vegetation in the Aksu-Tarim River Basin, Xinjiang, China. *Journal of Arid Environments*, 2015, 112: 87-97.
- [7] Xu Z Q, Zhang N X, Wang R, Yang X, Sun S J, Yan T F. The ecological water demand of different vegetation types in the Bashang area, northwest Hebei Province. *Journal of Resources and Ecology*, 2022, 13(1): 113-119.
- [8] García-Leoz V, Villegas J C, Suescún D, Flórez C P, Merino-Martín L, Betancur T, León J D. Land cover effects on water balance partitioning in the Colombian *Andes*: improved water availability in early stages of natural vegetation recovery. *Regional Environmental Change*, 2018, 18(4): 1117-1129.
- [9] Chi D, Wang H, Li X B, Liu H H, Li X H. Estimation of the ecological water requirement for natural vegetation in the Ergune River Basin in Northeastern China from 2001 to 2014. *Ecological Indicators*, 2018, 92: 141-150.
- [10] Zhang L L, Wei Y F, Wang Z Z. Prediction on ecological water demand based on support vector machine. 2008 International Conference on Computer Science and Software Engineering. December 12-14, 2008, Wuhan, China. IEEE, 2008: 1032-1035.
- [11] Huang F M, Cao Z S, Guo J F, Jiang S H, Li S, Guo Z Z. Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. *CATENA*, 2020, 191: 104580.
- [12] Zhao L H. Prediction model of ecological environmental water demand based on big data analysis. *Environmental Technology & Innovation*, 2021, 21: 101196.
- [13] Adams J B, Cowie M, Van Niekerk L. Assessment of completed ecological water requirement studies for South African estuaries and responses to changes in freshwater inflow. Pretoria: Water Research Commission, 2016:1-57.
- [14] Alemayehu T, van Griensven A, Woldegiorgis B T, Bauwens W. An improved SWAT vegetation growth module and its evaluation for four tropical ecosystems. *Hydrology and Earth System Sciences*, 2017, 21(9): 4449-4467.
- [15] Mati B M, Mutie S, Gadain H, Home P, Mtalo F. Impacts of land-use/cover changes on the hydrology of the transboundary *Mara River*, Kenya/Tanzania. *Lakes & Reservoirs: Science, Policy and Management for Sustainable Use*, 2008, 13(2): 169-177.
- [16] WREM International Inc. *Mara River Basin Monograph: Final Report*; WREM International Inc.: Atlanta, GA, USA, 2008, 446.
- [17] Larcher W. *Physiological plant ecology*. 3rd edition. Berlin, 1995: 170-236.
- [18] FAO. *ET0 calculator: land and water digital media series 36*. Rome: FAO, 2009.

- [19] 闵庆文, 何永涛, 李文华, 李贵才. 基于农业气象学原理的林地生态需水量估算——以泾河流域为例. *生态学报*, 2004, 24(10): 2130-2135.
- [20] Saxton K E, Rawls W J, Romberger J S, Papendick R I. Estimating generalized soil-water characteristics from texture. *Soil Science Society of America Journal*, 1986, 50(4): 1301-1306.
- [21] Huang Y, Zhao L. Review on landslide susceptibility mapping using support vector machines. *CATENA*, 2018, 165: 520-529.
- [22] He Q F, Shahabi H, Shirzadi A, Li S J, Chen W, Wang N Q, Chai H C, Bian H Y, Ma J Q, Chen Y T, Wang X J, Chapi K, Ahmad B B. Landslide spatial modelling using novel bivariate statistical based Naïve Bayes, RBF Classifier, and RBF Network machine learning algorithms. *The Science of the Total Environment*, 2019, 663: 1-15.
- [23] Breiman L. Random forests. *Machine Learning*, 2001, 45(1): 5-32.
- [24] Ho T K. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(8): 832-844.
- [25] 王奕森, 夏树涛. 集成学习之随机森林算法综述. *信息通信技术*, 2018, 12(1): 49-55.
- [26] LeCun Y, Boser B, Denker J S, Henderson D, Howard R E, Hubbard W, Jackel L D. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1989, 1(4): 541-551.
- [27] Zhang H, Huang C, Hu X D, Mei H B, Hu R F. Evaluating water resource carrying capacity using the deep learning method: a case study of Yunnan, Southwest China. *Environmental Science and Pollution Research*, 2022, 29(32): 48812-48826.
- [28] Alemayehu T, van Griensven A, Senay G B, Bauwens W. Evapotranspiration mapping in a heterogeneous landscape using remote sensing and global weather datasets: application to the *Mara basin*, east Africa. *Remote Sensing*, 2017, 9(4): 390.
- [29] Mutiga J K, Su Z, Woldai T. Using satellite remote sensing to assess evapotranspiration: case study of the upper Ewaso Ng'iro North Basin, Kenya. *International Journal of Applied Earth Observation and Geoinformation*, 2010, 12: S100-S108.
- [30] Ogutu J O, Owen-Smith N. ENSO, rainfall and temperature influences on extreme population declines among African savanna ungulates. *Ecology Letters*, 2003, 6(5): 412-419.
- [31] Fan X, Xu Z, Liu X, Tang C, Wang L, Han X. Effects of light intensity on the growth and leaf development of young tomato plants grown under a combination of red and blue light. *Scientia Horticulturae*, 2013, 153: 50-55.
- [32] 杨文新, 何新林, 杨广, 李俊峰, 杜玉娇. 荒漠区柽柳液流特征及其与环境因子的关系研究. *节水灌溉*, 2012(12): 26-29, 36.
- [33] Pham B T, Tien Bui D, Prakash I, Dholakia M B. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *CATENA*, 2017, 149: 52-63.
- [34] 温家鸣, 郭纯青, 李新建, 李文斌. 基于 BP 神经网络的桂林生态城市建设需水量预测. *水资源保护*, 2012, 28(3): 47-50.
- [35] Xu C, Shen L L, Wang G L. Soft computing in assessment of earthquake-triggered landslide susceptibility. *Environmental Earth Sciences*, 2016, 75(9): 767.
- [36] 杨颖. 基于卷积神经网络的语音识别研究[D]. 广州: 华南理工大学, 2018.
- [37] 王毅, 方志策, 牛瑞卿, 彭令. 基于深度学习的滑坡灾害易发性分析. *地球信息科学学报*, 2021, 23(12): 2244-2260.
- [38] 郭巧玲, 杨云松, 李建林, 陈志辉. 额济纳绿洲生态需水及其预测研究. *干旱区资源与环境*, 2011, 25(5): 135-139.
- [39] Saha S, Saha M, Mukherjee K, Arabameri A, Ngo P T T, Paul G C. Predicting the deforestation probability using the binary logistic regression, random forest, ensemble rotational forest, REPTree: a case study at the Gumani River Basin, India. *Science of the Total Environment*, 2020, 730: 139197.