

基于梯度提升机的干旱趋势提取与 ARIMA 耦合预测模型研究

杨扬¹, 王振霖¹, 刘涛²

1. 湖南省水利水电科学研究院, 湖南 长沙 410007

2. 长沙理工大学水利与环境工程学院, 湖南 长沙 410114

DOI:10.61369/WCEST.2025090006

摘要 : 为实现对大范围旱情的风险精准管控, 对区域尺度气象干旱指数 (SPEI) 进行精准预测至关重要, 本研究构建了一种梯度提升机 (GBM) 与 ARIMA 模型相结合的混合预测框架。该框架通过 GBM 提取 SPEI 序列的内在长期趋势以构建增强型数据集, 利用 ARIMA 模型捕捉序列围绕该趋势的残差波动。以湖南省为案例进行验证模型基于 2012 年及之前的数据进行训练, 并对 2013–2020 年的逐月 SPEI 进行预测。结果表明: 模型综合预测性能优异, 测试集上的均方根误差 (RMSE) 为 0.47, 平均绝对误差 (MAE) 为 0.37, 纳什效率系数 (NSE) 达到 0.75, 趋势一致性指标高达 0.88, 能稳定再现湖南省干旱的宏观时空分布格局。

关键词 : ARIMA 模型; 梯度提升机; SPEI

Research on a Coupled Prediction Model Integrating Gradient Boosting Machine for Drought Trend Extraction with ARIMA

Yang Yang¹, Wang Zhenlin¹, Liu Tao²

1. Hunan Provincial Institute of Water Resources and Hydropower Research, Changsha, Hunan 410007

2. College of Water Resources and Environmental Engineering, Changsha University of Science and Technology, Changsha, Hunan 410114

Abstract : To achieve precise risk management and control over large-scale drought conditions, it is crucial to accurately predict the regional-scale meteorological drought index (SPEI). This study constructs a hybrid prediction framework that combines a Gradient Boosting Machine (GBM) with an ARIMA model. The framework utilizes GBM to extract the inherent long-term trends in the SPEI series to construct an enhanced dataset, and employs the ARIMA model to capture the residual fluctuations of the series around this trend. Using Hunan Province as a case study, the model was trained on data from 2012 and earlier, and then predicted the monthly SPEI from 2013 to 2020. The results demonstrate that the model exhibits excellent comprehensive predictive performance, with a Root Mean Square Error (RMSE) of 0.47, a Mean Absolute Error (MAE) of 0.37, a Nash-Sutcliffe Efficiency (NSE) coefficient reaching 0.75, and a trend consistency index as high as 0.88 on the test set. It can stably reproduce the macroscopic spatiotemporal distribution pattern of drought in Hunan Province.

Keywords : ARIMA model; Gradient Boosting Machine; SPEI

引言

干旱是指长期无雨或少雨, 导致土壤水分亏缺、作物水分平衡破坏, 进而造成农业减产的一种重要农业气象灾害^[1]。随着全球气候变暖与人类经济社会活动的加剧, 水资源短缺问题日益突出, 干旱范围扩大、程度加深, 干旱化趋势已成为全球关注的焦点^[2]。政府间气候变化专门委员会第六次评估报告指出, 受自然强迫和人类活动共同影响, 农业与生态干旱呈加剧趋势, 严重影响区域水资源安全、生态安全和社会经济发展^[3]。中国作为受干旱影响较为严重的国家之一, 近年来极端干旱事件频发, 如 2022 年长江流域特大干旱, 进一步凸显了干旱监测与预测的紧迫性^[4]。

在此背景下, 科学准确地评估和预测干旱成为应对干旱灾害的关键。干旱指数是干旱监测与评估的重要工具, 其中标准化降水蒸散指数由于综合考虑了降水和潜在蒸散发的共同作用, 能够更全面地反映水分供需平衡状况, 尤其适用于全球变暖背景下的干旱识别^[5]。近年来, SPEI 已被广泛应用于中国多个区域的干旱时空特征分析中^[6,7], 显示出其在湿润与干旱半干旱地区均具备良好的适用性^[6]。因此, 本文选取 SPEI 作为干旱监测与预测的核心指标, 旨在为进一步提升干旱预测能力、制定科学抗旱策略提供理论依据。

基金项目: 湖南省水利科技项目 (编号: XSKJ2024064-27)。

一、数据和方法

(一) 数据

SPEI数据来源于CHM_Drought: A New High-Resolution Multi-Drought Indices Dataset for Mainland China

降雨数据来源于彭守璋. (2019). 中国1km分辨率逐月平均气温数据集(1901–2024). 国家青藏高原科学数据中心.

(二) 模型

ARIMA (Autoregressive Integrated Moving Average Model) 模型全称为自回归差分移动平均模型。ARIMA 模型就是试图通过数据的自相关性和差分的方式, 提取出隐藏在数据背后的时间序列模式, 然后用这些模式来预测未来的数据。ARIMA 模型主要由三部分构成, 分别为自回归模型 (AR)、差分过程 (I) 和移动平均模型 (MA)。其中: AR 模型, 即自回归模型, 其优势是对于具有较长历史趋势的数据, AR 模型可以捕获这些趋势, 考虑了过去若干时期的观测值对当前值的影响。I 部分用于使非平稳时间序列达到平稳, 通过一阶或者二阶等差分处理, 消除了时间序列中的趋势和季节性因素。MA 模型, 即移动平均模型, 可以更好地处理那些有临时、突发的变化或者噪声较大的时间序列数据, 但是无法捕捉较长的历史数据, 考虑了过去的预测误差对当前值的影响。

结合这三部分, ARIMA 模型既可以捕捉到数据的趋势变化, 又可以处理那些有临时、突发的变化或者噪声较大的数据。

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \xi_t$$

$$Y_t = \mu + \dot{\varphi}_1 + \theta_1 \dot{\varphi}_{t-1} + \theta_2 \dot{\varphi}_{t-2} + \dots + \theta_q \dot{\varphi}_{t-q}$$

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \theta_1 \dot{\varphi}_{t-1} + \theta_2 \dot{\varphi}_{t-2} + \dots + \theta_q \dot{\varphi}_{t-q} + \dot{\varphi}_t$$

GBM (Gradient Boosting Machine) 算法是 Boosting(提升) 算法的一种。通过串行地生成多个弱学习器, 每个弱学习器的目标是拟合先前累加模型的损失函数的负梯度, 使加上该弱学习器后的累积模型损失往负梯度的方向减少。且它用不同的权重将基学习器进行线性组合, 使表现优秀的学习器得到重用。最常用的基学习器为树模型。

Gradient Boosting 还可以将其理解为函数空间上的梯度下降。我们比较熟悉的梯度下降通常是值在参数空间上的梯度下降。

而在 Gradient Boosting 中, 每轮迭代生成一个弱学习器, 这个弱学习器拟合损失函数关于之前累积模型的梯度, 然后将这个弱学习器加入累积模型中, 逐渐降低累积模型的损失。即参数空间的梯度下降利用梯度信息调整参数, 从而降低损失, 而函数空间的梯度下降利用梯度, 拟合一个新的函数, 从而降低损失。

二、分析方法

基于时空分解与特征增强的混合预测框架。构建湖南省大区域尺度上气象干旱指数 (SPEI) 的精准预测。使用梯度提升机 (Gradient Boosting Machine, GBM) 作为时序信号的解构与预测工具。GBM 模型凭借其强大的非线性拟合能力, 能够有效识别并学习历史数据中隐含的长期演变规律及复杂的季节性波动。

以2012年为时间节点, 将每个栅格2012年及之前的历史 SPEI 序列作为训练样本, 构建 GBM 预测模型。利用训练完成的模型外推2012年后的趋势成分, 根据历史信息预测未来趋势。通过模型获取的非线性长期趋势信号。作为额外变量与对应栅格的原始 SPEI 观测值在时序上进行融合, 生成一个包含原始观测值与趋势辅助变量的增强型数据集。

将每个栅格的增强型数据集输入时序预测模型进行最终学习与预测。鉴于研究涉及大规模栅格运算, 并确保预测结果在空间上的连续性与可比性, 本研究采用了一种“结构约束与动态检验相结合”的 ARIMA 模型参数配置策略。由于将全省区域划分成若干栅格, 为保证模型结构在空间上保持一定的连续性, 设置模型的自回归阶数 (p) 与移动平均阶数 (q) 被统一设定为 1。差分阶数 (d) 通过动态检验确定, 确保了每个栅格输入数据的平稳性。

最终, 该 ARIMA 模型在趋势辅助变量的引导下, 专注于学习并预测原始序列围绕该趋势的残差波动。通过“GBM 提供的非线性趋势基线”与“ARIMA 捕捉的线性残差校正”的有机结合, 本框架实现了对各栅格 2012 年后逐月 SPEI 数据的高鲁棒性预测。

三、结果分析

(一) 模型综合预测精度评估

为全面评估本研究所提出的“分解-优化-集成”混合模型在湖南省 SPEI 预测中的性能, 我们在全域栅格尺度上计算了多项定量评估指标。如图 1 所示, 模型在测试集 (2013–2020 年) 的整体表现优异。

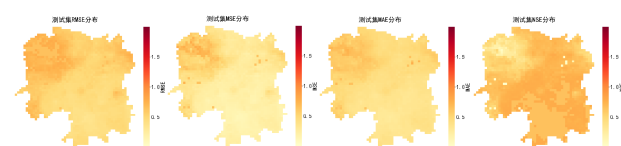


图1 测试集指标分布

(二) 预测结果的时空格局与稳定性分析

模型的预测能力不仅体现在宏观统计指标上, 更需通过其时空预测格局进行验证。空间格局再现能力: 通过对 2013 年与 2017 年 3 月到 10 月的可视化预测结果的示意图进行分析 (图 2–17) 发现, 模型对于湖南省宏观干旱空间分布格局的预测与实际情况高度一致。例如, 对于湘北、湘中地区季节性干旱的空间范围与强度, 模型均能实现准确刻画。这证明模型成功学习了区域干旱的空间分异规律, 具备稳定的空间预测能力。

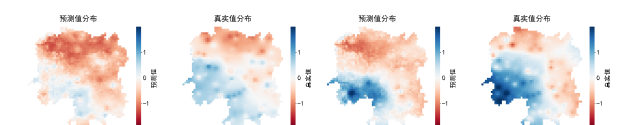


图2 2013年3月

图3 2013年4月

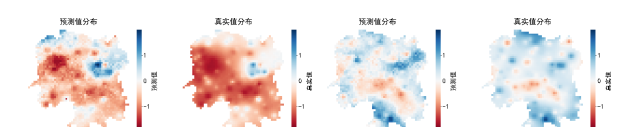


图4 2013年5月

图5 2013年6月

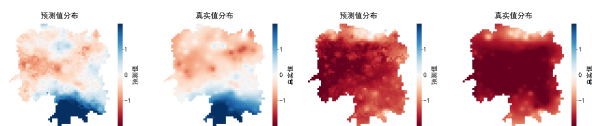


图6 2013年7月

图7 2013年8月

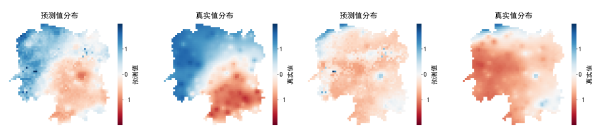


图16 2017年9月

图17 2017年10月

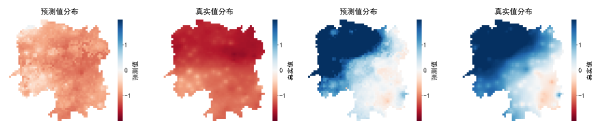


图8 2013年9月

图9 2013年10月

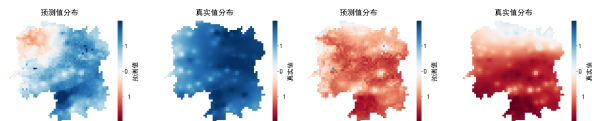


图10 2017年3月

图11 2017年4月

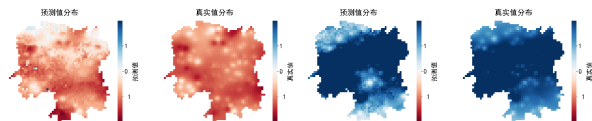


图12 2017年5月

图13 2017年6月

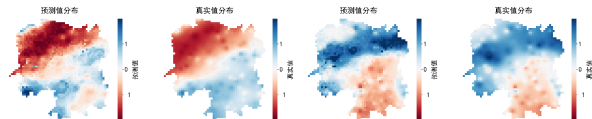


图14 2017年7月

图15 2017年8月

（三）结论：模型有效性与应用价值

本研究构建的混合预测框架，通过梯度提升机（GBM）对非线性趋势的精准提取，与 ARIMA 模型结合，成功实现了对湖南省 2013 - 2020 年 SPEI 的高精度时空预测。统计有效性方面，模型在测试集上表现出优异的综合性能：均方根误差（RMSE）为 0.47，平均绝对误差（MAE）为 0.37，显示出预测值与实测值之间的偏差整体处于较低水平，模型输出稳健可靠；纳什效率系数（NSE）达到约 0.75，说明模型能够解释 SPEI 序列中大部分变异特征，预测效果显著优于均值预测；趋势一致性指标高达 0.88，进一步表明模型不仅在数值上拟合良好，更能准确捕捉各栅格的长期干湿演变规律。

模型成功再现了湖南省干旱的宏观空间分布格局，其误差在空间上的分布与地形、下垫面等地理气候特征相符，反映出模型对干旱驱动物理过程的合理刻画，增强了结果的可解释性与可信度。

参考文献

- [1] 孙嘉敏, 王爱慧. 近 120 年中国干旱年际变化特征及主要的影响因子 [J]. 气候与环境研究, 2024, 29(6): 681-694.
- [2] 高晴, 孙金伟, 赵晓雪, 等. 基于 SPEI 的中国干旱多尺度时空特征分析 [J]. 节水灌溉, 2024(6): 111-120, 127.
- [3] 高启慧, 秦圆圆, 梁媚聪, 等. IPCC 第六次评估报告综合报告解读及对我国的建议 [J]. 环境保护, 2023, 51(Z2): 82-84.
- [4] 李忆平, 张金玉, 岳平, 等. 2022 年夏季长江流域重大干旱特征及其成因研究 [J]. 干旱气象, 2022, 40(5): 733-747.
- [5] VICENTE-SERRANO S M, BEGUERÍA S, LÓPEZ-MORENO J I. A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index[J]. 2010.
- [6] 谢坤, 丁明军, 张华敏, 等. 基于 SPEI 指数的鄱阳湖流域近 60 a 干旱时空变化特征 [J]. 长江流域资源与环境, 2021, 30(8): 2015-2027.
- [7] 王文玉, 李涛, 高泽海, 等. 西北干旱地区气象干旱风险预测模型研究 [J]. 中国水土保持科学 (中英文), 2025, 23(1): 117-130.