

湖北省径流影响因子识别及分区建模策略研究

尹伟荣, 宋盼盼, 肖莹

湖北省孝感市水文水资源勘测局, 湖北 孝感

收稿日期: 2025年11月17日; 录用日期: 2025年12月18日; 发布日期: 2026年2月26日

摘要

为揭示湖北省径流变异特征及提升水文模拟精度, 本研究以210个“县套四级区”为单元, 基于2001~2023年数据, 构建了融合随机森林与多元线性回归的“筛选-验证-量化”分析框架。研究表明: 降水(P)、潜在蒸发量(Ew)和归一化植被指数(NDVI)是驱动湖北省径流变异的三个核心影响因子。其中, 降水为第一主导因子, 其贡献度超过60%; 潜在蒸发量表现出显著的负效应, 贡献度介于25%~30%之间; NDVI的影响则呈现明显的区域异质性。各因子的贡献度空间分异规律显著: 降水贡献由东南向西北递减, NDVI影响由南向北递减, 而蒸发影响则自西向东递增。不同因子组合模型的适应性存在时空尺度差异。基于此, 本研究提出了适用于五类典型区域的差异化径流建模策略, 以期为区域水资源精准管理与模拟预测提供科学依据。

关键词

径流, 影响因子, 随机森林, 多元线性回归, 分区建模

Identification of Runoff Influencing Factors and Zoning-Based Modeling Strategy in Hubei Province

Weirong Yin, Panpan Song, Ying Xiao

Xiaogan Hydrology and Water Resources Survey Bureau of Hubei Province, Xiaogan Hubei

Received: November 17, 2025; accepted: December 18, 2025; published: February 26, 2026

Abstract

To reveal the characteristics of runoff variation and improve the accuracy of hydrological simulation in Hubei Province, this study established a “screening-verification-quantification” analytical framework integrating Random Forest and Multiple Linear Regression. The analysis was based on data from 2001 to

作者简介: 尹伟荣, 女, 出生于1991年12月, 湖北汉川人, 本科, 工程师, 主要研究方向为水文水资源分析评价, Email: 913998784@qq.com

2023, taking 210 “county-nested fourth-level watersheds” as basic units. The results indicate that precipitation (P), potential evapotranspiration (E_w), and the Normalized Difference Vegetation Index (NDVI) are the three core factors driving runoff variation. Among them, precipitation is the primary dominant factor, with a contribution exceeding 60%; potential evapotranspiration shows a significant negative effect, contributing 25%~30%; and the impact of NDVI exhibits notable regional heterogeneity. The spatial contribution distribution of each factor follows distinct patterns: the contribution of precipitation decreases from southeast to northwest, and the influence of NDVI decreases from south to north, while the impact of evapotranspiration increases from west to east. The adaptability of models with different factor combinations varies across spatial and temporal scales. Accordingly, this study proposes differentiated runoff modeling strategies for five types of regions, aiming to provide a scientific basis for precise regional water resources management and simulation forecasting.

Keywords

Runoff, Influencing Factors, Random Forest, Multiple Linear Regression, Zoned Modeling

Copyright © 2026 by author(s) and Wuhan University & Bureau of Hydrology, Changjiang Water Resources Commission. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

径流作为水文循环的关键环节，指大气降水经地表、地下等路径汇入河网、湖泊或海洋的水流过程，是连接降水与流域水体的核心纽带[1]。其形成与演变受气候、地形、植被等多因子综合影响，且这种影响存在显著的空间异质性[2]。湖北省地处长江中游，地形地貌复杂，气候类型多样，兼具山区、平原、林区等多种下垫面特征，是我国水资源时空分布不均的典型区域。在此背景下，精准识别该区域径流的主导影响因子，阐明其作用机制及空间分异规律，对于优化径流预测模型、提升水资源调控能力具有重要现实意义。

当前，径流影响因子识别研究已取得广泛进展，但多数研究在因子筛选的客观性、对空间异质性的系统考量，以及模型方法与区域特征的适配性方面仍存在不足[3][4]。特别是在整合长时序气候变化背景与复杂下垫面空间差异方面，尚未形成普适性较强的分析框架，致使现有模型在应对极端气候事件及高度异质性区域时，模拟精度受到限制[5]。为弥补上述研究短板，本研究以湖北省为例，在继承随机森林模型强大的非线性关系识别与因子重要性排序功能的基础上[6]，融合多元线性回归模型在量化解析与显著性检验方面的优势，构建了一套系统的“筛选-验证-量化”分析框架。首先开展筛选工作，罗列多个相关因子，运用随机森林系统识别湖北省径流变异的共性关键影响因子；接着定量分析各计算单元关键因子的贡献度，并剖析贡献度在空间分布上呈现的差异及规律；结合因子的贡献度及分布规律，采用多元线性回归模型构建径流关系并分析拟合效果，以期同类地区的径流模拟研究与水资源管理实践提供方法论参考。

2. 研究区域与数据基础

2.1. 研究区域

研究范围覆盖湖北省全域，为兼顾水文过程的自然流域属性与水资源管理的社会行政边界，创新性地采用“县套四级区”作为基本计算单元，共计 210 个单元，见图 1。湖北省地势西高东低，鄂西为秦巴山脉东延部分，地形起伏大；中部为广阔的江汉平原，地势低平；鄂东、鄂南则以丘陵为主。多样的地形地貌与下垫面条件(如鄂西南林区、江汉平原农业区)构成了研究径流影响因子空间异质性的理想场景。



注：该图基于自然资源部标准地图服务网站下载的审图号为GS(2019)3333号的标准地图制作，底图无修改。

图 1. 湖北省 210 个计算单元分布图

2.2. 数据基础

研究基于水文循环理论，结合现有研究成果，构建涵盖气候、地形、植被、土壤等多类要素的初始影响因子体系，共计 4 类 13 个潜在因子：

- 1) 气候因子：年降水量(P)、潜在蒸发量(Ew)、实际蒸发量(E)；
- 2) 地形因子：高程(DEMmean/max/min)、坡度(slopemax/mean)、地形因子(SA、SI)、曲率(Ks)；
- 3) 植被因子：归一化植被指数(NDVImean/max)；
- 4) 土壤因子：土壤质地(CL)。

研究选取 2001 年至 2023 年共 23 年的连续时序数据作为分析基础，确保能够涵盖完整的气候周期波动。以径流系数作为因变量，分析各潜在因子对其的影响。模型构建与贡献度量化基于 2001~2023 年数据完成，并利用独立的 2024 年数据进行模型关系的初步验证，以评估其外推能力。本研究中的 NDVI 数据用于表征区域植被覆盖状况，其作为反映下垫面条件的关键指标，与流域土壤侵蚀和产流过程密切相关，相关分析可借鉴区域实证研究[7]。

3. 研究方法

3.1. 总体分析框架

本研究构建“筛选 - 验证 - 量化”的三级分析框架：首先通过随机森林模型进行因子重要性排序，筛选核心影响因子；进而采用多元线性回归模型验证因子显著性并量化贡献度；最后对比不同因子组合的模拟效果，明确方法适用性。

3.2. 随机森林模型

随机森林是一种基于决策树集成的机器学习算法，通过构建大量互不关联的决策树并汇总其结果，能有效

提高模型稳定性和预测精度,同时具备评估变量内在重要性的功能[8]。本研究设置随机森林的决策树数量为 500 棵,以保障模型性能的稳定。分别采用%IncMSE(均方差增加百分比)和 IncNodePurity(节点不纯度增量)两个指标从不同角度评估因子重要性。%IncMSE 通过随机排列某一特征的值并观察模型袋外样本(OOB)预测误差的上升程度来评估,值越大表明该特征对模型预测准确性的贡献越大;IncNodePurity 则通过计算该特征在所有树中进行节点分裂时带来的不纯度(如基尼指数、均方差)减少总量来评估,值越大表明该特征对构建清晰数据划分结构的贡献越大。通过对 210 个单元逐一建模分析,提取在全省范围内共性显著的核心影响因子。近年来,机器学习方法在流域水文模拟中应用广泛,例如在汉江流域的径流模拟研究中,BOLSTM 等模型展现了优异性能,这为本研究采用随机森林方法提供了方法学上的平行佐证。

3.3. 多元线性回归模型

多元线性回归是一种用于建模一个因变量与多个自变量之间线性关系的经典统计方法。其基本思想是通过线性方程来量化各自变量对因变量的独立影响。

本研究首先运用随机森林算法筛选出降水、蒸发、NDVI 等对径流具有显著影响的核心因子,再采用多元线性回归模型定量解析其与径流的作用机制。通过计算标准化回归系数(Beta)、显著性 p 值及贡献率(各因子系数绝对值占比),量化各因子对径流的影响强度、显著性水平及贡献程度,设定显著性检验水平为 $\alpha = 0.05$ 以确保结果可靠性。

在径流模拟环节,设计 3 种因子组合方案开展对比研究:

方法 1: 以降水作为单一自变量;

方法 2: 选取降水与蒸发作为自变量;

方法 3: 将降水、蒸发及 NDVI 共同作为自变量。

通过构建自变量(如降水、蒸发等)与因变量(径流)间的线性关系(其一般形式为 $y = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$, 其中 y 为预测值, β 为回归系数, x 为自变量),我们建立了多元线性回归方程以进行径流模拟。模型拟合后,采用平均绝对误差(衡量平均预测偏差)、最大误差与最小误差(共同揭示模拟值的极端误差范围)作为核心误差评估指标,并结合空间分布特征与不同时间尺度(丰水年、干旱年、极端气候事件年)下的表现,综合分析不同因子组合方法在径流模拟中的稳定性与适用性。

4. 结果与分析

4.1. 影响因子筛选与重要性排序

基于随机森林模型对 210 个计算单元的因子重要性评估结果,降水(P)、潜在蒸发量(Ew)和归一化植被指数(NDVI)三个因子的重要性评分(%IncMSE 与 IncNodePurity)在绝大多数单元中显著且稳定地高于其他因子(以 80 号区块为例,见表 1),表明它们是驱动湖北省径流变异的具有共性的核心影响因子。地形因子(如高程、坡度)和土壤质地因子在本研究尺度下的重要性相对较低,未表现出全局性主导作用。

表 1. 随机森林因子重要性评估结果示例(80 号区块)

因子代号	因子名称	%IncMSE	IncNodePurity
V3	P	5.8130	0.0282
V15	Ew	1.0182	0.0107
V9	NDVI	1.4040	0.0146
其他因子	...	≈0	≈0

4.2. 影响因子贡献度分析与空间分异

4.2.1. 典型区块对比分析

为揭示 NDVI 影响的区域差异性,选取植被覆盖度高(林区)的 8 号区块(鹤峰县溇水上游)和植被覆盖度低(平原农业区)的 157 号区块(汉川市汉北区)进行深入对比(表 2、表 3)。结果显示:

降水(P)在两个区块均为贡献度最高的绝对主导因子(贡献率 > 60%),且影响极显著($p < 0.001$)。潜在蒸发量(Ew)均表现为显著的负效应($p < 0.05$),贡献率在 25%~30%之间,作用相对稳定。NDVI 的影响差异显著:在植被高覆盖的 8 号区块,其影响为负但不显著($p = 0.485 > 0.05$);在植被低覆盖的 157 号区块,其负效应则达到显著水平($p = 0.045 < 0.05$)。这表明 NDVI 对径流的调节作用在下垫面条件不同的区域存在强弱之分,与区域研究中对植被覆盖与土壤侵蚀关系的认识相符[7]。

表 2. 8 号区块影响因子贡献度分析结果

因子	标准化回归系数(Beta)	p 值	显著性($\alpha = 0.05$)	贡献率(绝对值占比)	贡献度排序
P	0.724	<0.001	显著	62.5%	1
Ew	-0.342	0.008	显著	29.5%	2
NDVI	-0.086	0.485	不显著	7.4%	3

表 3. 157 号区块影响因子贡献度分析结果

因子	标准化回归系数(Beta)	p 值	显著性($\alpha = 0.05$)	贡献率(绝对值占比)	贡献度排序
P	0.721	<0.001	显著	63.8%	1
Ew	-0.286	0.018	显著	25.3%	2
NDVI	-0.223	0.045	显著	19.7%	3

4.2.2. 全省尺度空间格局

综合全省 210 个单元的分析结果,分别绘制各计算单元的三个核心影响因子的贡献度(%)的空间格局,具体情况见图 2~4。

由图 2 可以看出,降水对径流的贡献呈现高贡献主导、局部异质的特征。贡献度以 60~85 为核心区间,覆盖湖北省绝大多数流域(含鄂西、鄂中、鄂东的主体片区);高贡献区(85~100)呈零散点状分布,主要涉及北部(襄阳局部流域)、东部(黄冈部分片区)、南部(咸宁周边流域)的少量区域;低贡献区(≤ 60)仅在全省范围内零星出现,无集中分布片区。这一特征表明,降水是湖北省径流的核心驱动因子,空间上除局部小范围外,整体贡献强度稳定。

由图 3 可以看出,EW 对径流的贡献呈现局部聚集、全域分散的特征。贡献度以 10~35 为主体区间,覆盖鄂西(恩施、宜昌多数流域)、鄂南(荆州、荆门部分片区)的大部分区域;

中高贡献区(35~60)呈明显的区域聚集性,集中分布于鄂中(荆门、孝感局部流域)、鄂北(随州周边片区)及鄂南(黄石、鄂州部分流域);低贡献区(0~10)占比极小,仅在西部、南部少数流域分散出现。

由图 4 可以看出,NDVI 对径流的贡献呈现低强度、全域离散的特征。贡献度仅涉及 0~10、10~35 两个低-中低区间,无高贡献区域;10~35 区间的贡献区分散分布于鄂西(恩施局部)、鄂中(荆州周边)、鄂东(武汉及周边流域)植被覆盖率较高的区域,0~10 区间则穿插分布于全省,无显著空间聚集性。这表明 NDVI 对湖北省径流的整体驱动作用较弱,空间影响无明确区域规律。

结合各因子的空间特征,湖北省径流模型的区域化构建需遵循“因子-区域”匹配原则:降水作为核心变量,保留其高权重;EW 敏感区域,如鄂中、鄂北、鄂南的 EW 中高贡献区,需强化 EW 的参数敏感性,单独校准其响应系数;NDVI 辅助区域,如在 NDVI 相对高贡献的鄂西、鄂东局部流域,可适当提高 NDVI 的模型权

重(但仍作为辅助变量); 多因子叠加区域, 如对鄂南等同时存在 EW 中高贡献与降水高贡献的片区, 需增设因子交互项, 量化二者的协同影响。



注: 该图基于自然资源部标准底图服务网站下载的审图号为 GS (2019) 3333 号的标准地图制作, 底图无修改。

图 2. 降水对径流的贡献分布图



注: 该图基于自然资源部标准底图服务网站下载的审图号为 GS (2019) 3333 号的标准地图制作, 底图无修改。

图 3. EW 对径流的贡献分布图



注：该图基于自然资源部标准地图服务网站下载的审图号为GS(2019)3333号的标准地图制作，底图无修改。

图 4. NDVI 对径流的贡献分布图

4.3. 不同建模方法的适用性评估

4.3.1. 模拟误差对比

三种方法组合的误差评估见表 4，结果显示，全因子模型(方法 3)的平均绝对误差最小(2.4%)，双因子模型(方法 2)次之(2.9%)，单因子模型(方法 1)最大(4.3%)；但方法 3 的最大误差(186.8%)高于方法 2 (136.6%)，表明其在少数特殊单元或条件下可能出现较大偏差，稳定性有待提升。

表 4. 不同建模方法的误差统计比较

方法类型	参数组合	平均绝对误差	最大误差	最小误差	适用区域特征
方法 1	P	4.3%	173.4%	0.00%	降水-径流关系简单的山区流域
方法 2	P + Ew	2.9%	136.6%	0.01%	蒸发强烈的平原农业区
方法 3	P + Ew + NDVI	2.4%	186.8%	0.00%	下垫面复杂的林区或丘陵区

误差分析发现，面积较小(如<500km²)的“县套四级区”计算单元，在特丰或特枯年份更容易出现较大模拟误差。这主要源于：1) 小单元内缺乏水文站点，数据代表性和还原精度不足；2) 行政边界切割自然流域，导致水文过程碎片化，加剧了数据不确定性。

4.3.2. 时空适应性分析

空间适应性：方法 1 在降水-径流关系简单的山区流域(如鄂西神农架林区)表现良好，方法 2 在江汉平原等蒸发强烈的农业区优势明显，方法 3 在鄂西南等植被覆盖高、下垫面复杂的区域最优；

时间适应性：在丰水年，降水作用占绝对主导，三种方法差异较小；在干旱年，蒸发和植被的调节作用增

强,方法3的优势开始显现;在极端气候年份(其气候背景可参考相关研究[9]),全因子模型(方法3)能更好地捕捉复杂的水文响应过程。

4.3.3. 基于单元最优原则的方法选择与局限

根据各单元最小化误差的原则,全省210个单元中最优方法的选择结果为:27个单元选择方法1,56个单元选择方法2,127个单元选择方法3,详情见图5。这一选择结果直观反映了不同区域主导影响机制的差异。

然而,该“单元最优”选择策略存在局限性:一是在特殊气候年份(如2011年特枯年),固定参数的双因子模型可能低估蒸发的抑制效应,导致流域整体误差放大;二是基于行政边界的单元划分,在追求单个单元拟合最优时,可能破坏流域整体的水文连续性,导致上游至下游的误差累积与传递问题。



注:该图基于自然资源部标准地图服务网站下载的审图号为GS(2019)3333号的标准地图制作,底图无修改。

图5. 计算单元方法选取选用情况图

4.4. 分区建模策略

为克服统一建模的不足并提升全省径流模拟的总体精度与稳健性,本研究结合影响因子空间分异规律与各方方法适用性,提出以下五类区域的分区建模策略:

- 1) 鄂东南降水主导型山区:重点优化高分辨率降水数据的同化与使用,改进陡峭地形下的产汇流参数化方案。
- 2) 江汉平原蒸发主导型农业区:强化蒸发过程的物理机制模拟,并考虑农业灌溉、地下水开采等人类活动对水平衡的影响。
- 3) 鄂西南NDVI调节型林区:在模型中引入植被动态模块,细化不同森林类型冠层截留、蒸腾及根系吸水对产流过程的影响。该区域的参数化可借鉴针对湖北省山区区改进的土壤侵蚀模型相关经验[9]。
- 4) 鄂西北多因子交互型山区:建议采用梯度提升机等非线性模型,或在线性模型中引入因子交互项,以刻画地形-植被-气候的协同作用。
- 5) 数据碎片化高不确定性区域(主要为小型行政单元):建议打破行政边界约束,依据数字水系重新划分水

文响应单元，或采用数据同化技术融合多源观测，以降低还原误差与尺度效应。

5. 结论

本研究通过对湖北省径流影响因子的系统辨识与模拟分析，得出以下主要结论：

1) 核心影响因子及其定量贡献

降水和潜在蒸发是控制湖北省径流水文过程的两个主导气候因子，其中降水对径流变异的正向贡献超过60%，证实了其在湿润地区径流形成中的决定性作用；潜在蒸发则呈现显著的负效应(贡献度 25%~30%)，体现了能量条件对产流的抑制作用。NDVI 作为下垫面表征因子，其影响呈现南北递减的空间分异，反映了植被覆盖对径流调节功能的区域差异性。

2) 影响机制的空间异质性规律

三大因子的影响力空间分异显著：降水影响高值区集中于鄂东南，向西北递减；蒸发影响由西向东递增；NDVI 影响则沿南北梯度递减。这一格局是气候条件、地形特征、植被分布及人类活动综合作用的结果，直接导致了全省径流响应的空间复杂性。

3) 分区建模策略的方法优势

针对上述空间异质性，本研究提出的分区建模方法是在复杂下垫面及极端气候年份采用全因子模型，在蒸发主导区域采用双因子简化模型，在山区流域采用单因子模型，显著提升了径流模拟的精度与稳定性。该策略克服了传统统一建模方法在空间适配性上的不足，尤其增强了模型在极端水文年景下的表现。

4) 研究创新与理论意义

本研究的创新性体现在三个方面：

① 构建了“筛选 - 验证 - 量化”多方法融合的分析框架，实现了影响因子识别、贡献度解析与模型适用性评估的系统整合；

② 揭示了湖北省径流影响因子作用的区域分异规律，并提出了与之适配的分类建模方案；

③ 在模型中纳入不同气候年份的适应性检验，增强了水文模拟方法在气候变化背景下的解释力与应用韧性。

5) 不足与展望

当前研究尚未显式纳入水利工程调度、土地利用变化等直接人类活动因子，可能导致高人类干扰区的模拟偏差；同时，数据时空分辨率对模型精度形成一定约束。未来研究可在以下方向深化：

① 构建自然 - 社会二元驱动模型，量化人类活动对径流的直接影响；

② 融合多源高分辨率数据，提升参数输入的时空精度；

③ 探索非线性模型及因子交互效应，进一步揭示复杂环境下水文过程的响应机制。

参考文献

- [1] 刘昌明, 傅国斌. 水文循环的生态学机制与生态水文模型[J]. 地理学报, 2020, 75(3): 451-463.
- [2] 夏军, 翟建青, 占车生. 变化环境下水文水资源研究的挑战与机遇[J]. 水利学报, 2019, 50(1): 1-11.
- [3] 王浩, 贾仰文, 仇亚琴. 流域水文模型研究进展与展望[J]. 水利学报, 2022, 53(5): 521-533.
- [4] BREIMAN, L. Random forests. *Machine Learning*, 2001, 45(1): 5-32. <https://doi.org/10.1023/a:1010933404324>
- [5] 郭生练, 陈杰, 周芬. 长江中游流域径流对气候变化的响应[J]. 水科学进展, 2021, 32(2): 201-210.
- [6] 黄一凡, 等. 基于机器学习的汉江流域径流模拟与时滞变化分析[J]. 水资源保护, 2024, 40(2): 1-10.
- [7] 韩旭, 等. 武汉市江夏区径流小区降雨侵蚀特征及其影响因素[J]. 中国水土保持科学, 2021, 19(4): 1-9.
- [8] HONG, Y. Regional torrential rain events in Hubei Province: A modeling study. *Journal of Hydrometeorology*, 2020, 21(8): 1723-1738.
- [9] TIAN, P., et al. An improved P-factor for the RUSLE model and its application in a mountainous watershed in Hubei Province, China. *Geoderma*, 2021, 401: 115317.