

基于机器学习算法模型的巫山县洪水灾害研究^{*}

牟风云, 杨猛, 林孝松, 龙秋月, 李梦梅, 何勇
(重庆交通大学建筑与城市规划学院, 重庆 400074)

摘 要: 利用机器学习算法 RF 模型、K-means 模型与 ARMA 模型, 对巫山县范围内 12 369 条径流河段进行分类预测, 研究水文参数在时间序列上的变化规律, 探究降雨-径流演变规律; 并结合 GIS 空间可视化技术, 综合研究区地理环境, 基于 RF 模型预测洪水致灾范围, 分析洪水灾害预测结果的空间特征。结果表明: ① RF 模型能有效预测降雨-径流演变过程中参数的重要性, 当降雨强度为 125 mm、150 ~ 175 mm 时, 预测出水位、流速变化率最大; 降雨强度为 100 ~ 175 mm 时, 预测出流速变化率最为剧烈; ② 利用 ARMA 模型预测出河流比降、流量等水文参数回归性最好, 较低等级河流所预测参数中, 水位、流速变化率最为明显, 流量无明显变化。相较于水位变化率, 流速变化更为强烈, 流速、水位变化率主要集中于等级较高河流; ③ 机器学习算法能有效预测研究区洪水易发程度, 在表征研究区水文参数时, 水位变化主要集中于西北部、中南部, 东北部与中南部水位变化率显著, 预测出部分地区水位可升至 20 m, 处极危险状态。

关键词: 机器学习; RF; K-means; ARMA; 降雨-径流; 洪灾

中图分类号: P954 **文献标志码:** A **文章编号:** 0529-6579 (2020) 01-0105-09

The flood disasters in Wushan county based on machine learning algorithm model

MU Fengyun, YANG Meng, LIN Xiaosong, LONG Qiuyue, LI Mengmei, HE Yong

(School of Architecture and Urban Planning, Chongqing Jiaotong University, Chongqing 400074, China)

Abstract: K-means model and ARMA model was applied to classify 12 369 runoff river reaches in Wushan County, and study the change of hydrological parameters in time series, and probes into the evolution of rainfall runoff process by using the machine learning algorithm RF model. Combined with the GIS spatial visualization technology and the geographical environment of the study are, the flood disaster scope is predicted and the spatial distribution of flood disasters is analyzed based on the RF model. The results show that: 1) The RF model can effectively predict the importance of parameters in the rainfall-runoff process. When the rainfall intensities are 125 mm and 150 - 175 mm, the change rate of water level and velocity is the largest; when the rainfall intensity is 100 - 175 mm, the change rate of velocity is the most intense. 2) The ARMA model is used to predict the river gradient, flow and other hydrological parameters with the best regression. Among the predicted parameters of lower level rivers, the change rate of water level and flow velocity is the most obvious, and the flow quantity has no obvious change. Compared with the change rate of water level, the change rate of flow rate is more intense, and the change rate of

* 收稿日期: 2019-02-27

基金项目: 国家自然科学基金 (41601564); 国家自然科学基金校内培育项目 (2018PY15); 重庆市自然科学基金 (cstc2019jcyj-msxmX0626); 重庆市基础研究与前沿探索项目 (cstc2018jcyjAX0156)

作者简介: 牟风云 (1979 年生), 女; 研究方向: 国土资源遥感, 3S 技术集成与应用; E-mail: mfyds@cqjtu.edu.cn
杨猛 (1993 年生), 男; 研究方向: 3S 技术集成与应用; E-mail: 836819545@qq.com
(以上 2 位作者并列第一作者)

flow rate and water level is mainly concentrated in the river with higher grade. 3) The machine learning algorithm can effectively predict the flood prone degree in the study area. When characterizing the hydrological parameters of the study area, the change of water level is mainly concentrated in the northwest and the south central part, and the change rate of water level in the northeast and the south central part is significant. It is predicted that the water level in some areas can rise to 20 m in extremely dangerous condition.

Key words: machine learning; RF; K-means; ARMA; rainfall runoff; flood disaster

山区地形地貌复杂、气候多变等诸多因素导致洪水灾害频频发生, 如何有效预防洪水灾害是亟待解决的问题。大数据已应用于各大学科领域, 而机器学习具有多学科、多领域、多算法等优势, 能有效地挖掘大数据潜在的信息, 其中, 随机森林模型 (RF)、K-means 聚类模型、自回归滑动模型 (ARMA) 等已广泛应用于大数据平台^[1]。加上山区地貌类型复杂、流域单元数据量大等特性, 利用机器学习能够有效预测出山区洪水演变规律, 实现山区洪水灾害防治与预警。

已有研究中, 在模型运用方面, 胡胜等^[2]、吴裕珍等^[3]采用 SWAT 模型利用 CFSR 气象数据对小流域进行水文模拟, 以及对降雨-径流水质进行研究, 并做出适用性评价; 刘志方等^[4]、顾西辉等^[5], 利用突变检验方法构建小气候对流域产流与汇流研究以及降雨-径流关系, 通过水文综合模拟系统 (HIMS) 考虑流域产流、汇流、蒸发、下渗等水循环过程; 理论研究方面, 现有研究表明, 降雨是导致滑坡主要因素, 降雨^[6-10]导致植被覆盖较少、坡度较陡等区域滑坡泥石流发生。文献 [11-12] 结合多方面因素研究洪水灾害, 进行洪水灾害风险评估。以上学者研究, 多在于数理统计与规律总结, 将研究结果进行空间表达方面有待研究, 且在机器学习方面有待结合。本文针对山区降雨-径流演变所涉及的水文参数大数据, 引入机器学习算法进行数据挖掘与统计分析, 并结合多学科进行地理空间信息可视化分析。

以巫山县为研究对象, 首先, 选取机器学习 RF、K-means 与 ARMA 算法模型, 对研究区降雨-径流演变过程所形成的 12 369 个子河段, 共计 500 条河流。进行水文参数关系拟合, 并利用 GIS 技术, 进行水文参数空间信息表达; 再者, 结合研究区地理环境、土壤类型、土地利用类型等指标, 多方面探讨机器学习算法在山区降雨-径流模拟过程实用性。最后, 综合机器学习优势和研究区地理环境特性, 运用算法进行研究区洪水灾害分析, 最终研究结果可为区域洪水灾害防治、公路洪灾预警

提供科学依据。

1 研究区域与方法

1.1 研究区概况

巫山县位于重庆市东北部, 与湖北接壤, 地处三峡库区腹心, 地理坐标介于 $30^{\circ}46' - 31^{\circ}28'N$, $109^{\circ}33' - 110^{\circ}11'E$ 之间。截至 2017 年底, 户籍人口 63 万, 共 24 个乡镇; 巫山县属亚热带季风气候, 降雨充沛, 四季分明, 年平均温度在 $18.4^{\circ}C$ 左右, 年均降雨量高达 1 041 mm; 山区溪河众多, 常年因降雨致使洪水灾害频发。地形地貌类型复杂, 属典型的喀斯特地貌, 地形起伏大, 最高海拔达 2 698 m, 最低海拔仅 63 m。2015 年 4 月, 巫山县 11 个乡镇遭受不同程度洪水灾害, 农作物受灾面积达 100 hm^2 , 直接经济损失 100 余万元; 2017 年 7 月, 巫山县发生暴雨洪水灾害, 降雨量达 175 mm 以上, 官阳镇、红椿乡等乡镇发生不同程度洪水灾害。

1.2 研究方法

机器学习为人工智能核心算法, 涉及多学科领域, 具有自动组织和模拟功能, 广泛应用于数据统计、资产评估预测等多方面^[12-13]。常见的机器学习算法为监督学习、非监督学习、深度学习 3 类。监督学习算法中, 神经网络、决策树算法最为经典, 常见的模型有随机森林模型、线性回归模型等, K-means 聚类算法模型在非监督学习算法最为常用, 深度学习中马尔可夫链研究较多。

水文参数信息提取提取过程中, 产生大量河段数据, 需要进行有效、精准的分类, 而随机森林模型刚好具备数据精准分类特性, 不仅能够将大量数据河段数据进行分类, 还能进行水文参数重要性预测; 研究区提取的 12 369 条河段, 出现大量数据冗余, 而在机器学习算法模型中, K-mean 模型具有去除冗余数据特性。选取 K-mean 算法对研究区河段数据进行聚类, 去除冗余值; 在研究降雨-径流演变过程中, 由于不同时刻降雨量不同, 水文参数变化率存在着差异, 而 ARMA 模型能够有效地

在时间序列上预测数据变动情况, 故选取 ARMA 模型进行降雨 - 径流时间序列预测。

1) 随机森林模型 Random Forest (RF)^[12-13]。RF 模型为各类决策树算法总和, 而决策树可为大数据统计分类提供有效、精准的分类算法。鉴于此, 结合 RF 模型, 对研究区降雨 - 径流水文参数统计数据, 进行数据挖掘, 获取每一河段预测值, 通过预测值进行河段分类, 最终拟合出最为精确降雨 - 径流关系式, 并以此预测山区洪水致灾范围。

2) K-means 模型^[14-16]。K-means 模型以特定距离对统计数据进行硬性聚类, 在大数据分析中广泛运用。该算法基本思路为: 将降雨 - 径流获取的水文参数中作为 K 个样本, 在 n 个河段中进行分类。K-means 模型的运用, 能够减少数据冗余值, 得到有效的研究参数。

3) ARMA 模型^[15-16]。时间序列 ARMA 模型为滑动自回归模型, 该模型将回归分析与滑动平均模型相结合, 能够有效地预测数据变动情况。结合该模型特性, 将不同河段编号作为时间序列, 以 RF 模型、K-means 模型和水文参数量化指标为回归预测值, 在不同降雨强度下对降雨 - 径流进行时间序列预测分析。

1.3 山区河流水文参数及经验公式建立

受气候、下垫面、人类活动等因素影响, 部分区域不会直接形成径流, 需经下渗、填洼、蒸发等汇流过程形成, 径流地理环境特性, 使其流速、水位变化存在差异。在理想条件下, 综合国内外雨洪计算方法^[13-16], 基于天然河道汇流平均速度与出水口断面平均流速公式, 联立曼宁公式在天然河道构建的平均流速计算公式^[15-18], 考虑岩石、植被、泥沙等因素对山区河流的阻碍作用以及流域汇流累计量的折减效应 (汇流参数); 山区河网密集错综复杂, 基于 DEM 提取的河网虽有等级信息却缺乏高程变化, 然而河流比降却能反映河流的高程变化, 即利用河流比降来表示河流纵断面高程变化。河流比降的表示与坡度相同, 其关系式为河流轴线河底或水面的高程与该河流长度的比值, 考虑到山区地形地貌因素, 将河流划分为若干个子河段求取平均比降; 结合研究区地形地貌、流域地理环境和土地利用等因素, 确定河段阻碍系数并通过研究区 12 369 条河流进行关系拟合, 择取机器学习算法模型, 研究降雨 - 径流演变规律, 并拟合水文参数之间关系。

1.4 数据来源及处理

综合考虑山区地理环境特性, 确定研究区地形

地貌、河流比降、山区河流阻碍系数 (糙率)、汇流参数、土壤特性、植被覆盖等为流域环境指标, 降雨强度 (流速、水位、流量)、径流系数、汇流参数等为水文参数指标。基于 GIS 技术, 以研究区分辨率为 30 m DEM 数据, 进行水文参数信息提取, 并以流域汇流累积流量为河流等级划分依据, 择取栅格单元 2 000 流量将河流划分为 6 个等级, 进而计算出汇流累计量、小流域等水文信息; 选取 2015 年夏季研究区 OIL 遥感影像, 对其进行植被覆盖和土地利用类型信息提取。研究涉及水文参数经验公式, 其理论与推导来源于国内外经验公式, 气象水文数据来源于巫山县观测站点。

2 结果分析

2.1 模型预测结果精度验证

需要对通过机器学习算法模型预测出的结果进行精度验证。巫山县共 52 个观测站点, 由于预测河段样本数量较大, 且每个站点能够统计到 100 以上条河段实际数据。在研究区中进行均匀选取, 并能够覆盖整个河段观测站点, 最终选出 10 个站点进行结果验证。以实测数据为标准进行预测结果对比验证 (图 1)。验证结果显示: 预测数据与实测数据整体上基本吻合, 部分数据存在偏差现象; 流速与水位预测结果准确性比预测流速高, 相较于流速与流量预测结果, 水位预测准确性最高。

2.2 研究区水文参数预测与分析

降雨 - 径流演变包含地形地貌、降雨强度、土壤特征、植被覆盖以及水库湖泊调节等诸多因素, 择取机器学习算法 RF 模型、K-means 模型与 ARMA 模型, 进行研究区降雨 - 径流水文参数拟合, 探究降雨 - 径流演变规律, 测定参数结果如表 1 所示。以研究区 12 369 条河段为研究对象, 通过 RF 模型进行预测统计, 结合 RF 模型与 K-means 算法模型, 进行径流水文参数分类统计, 拟合河段数量与河流等级关系。统计结果中, 不同等级径流流速、水位变化有所差异, 降雨阈值为 25、75、125、200 mm 流速、水位拟合结果最好。

不同降雨强度阈值下, 水文参数随河流等级、径流曲线数 (CN)、河流阻碍系数不同而随之发生变化, 从图 2 (a) 河流等级相对流速变化可以看出, 随着河流等级增加, 流速变化率整体上呈现下降趋势, 且降雨强度越大, 等级较低河流变化幅度越大; 1 ~ 3 级河流变化幅度最大, 降雨量在 25 ~ 75 mm 区间, 其变化幅度最小, 降雨 75 ~ 200 mm, 等级变化相对集中; 预测结果中, 变化最为

剧烈的为 2 ~ 4 级河流。河流等级变化对河流流速变化预测较为客观,需对研究区整体河流进行分类与预测,图 2 (b) 为巫山县整体河流流速变化趋势,在研究区 6 个等级河流中,流速变化主要为 1 ~ 3 等级河流,且降雨阈值为 200 mm 时,流速变化幅度最大;1 ~ 4 等级河流中,部分区域流速突变显著,由此可见,山区河流流速、水位变化率最为明显。

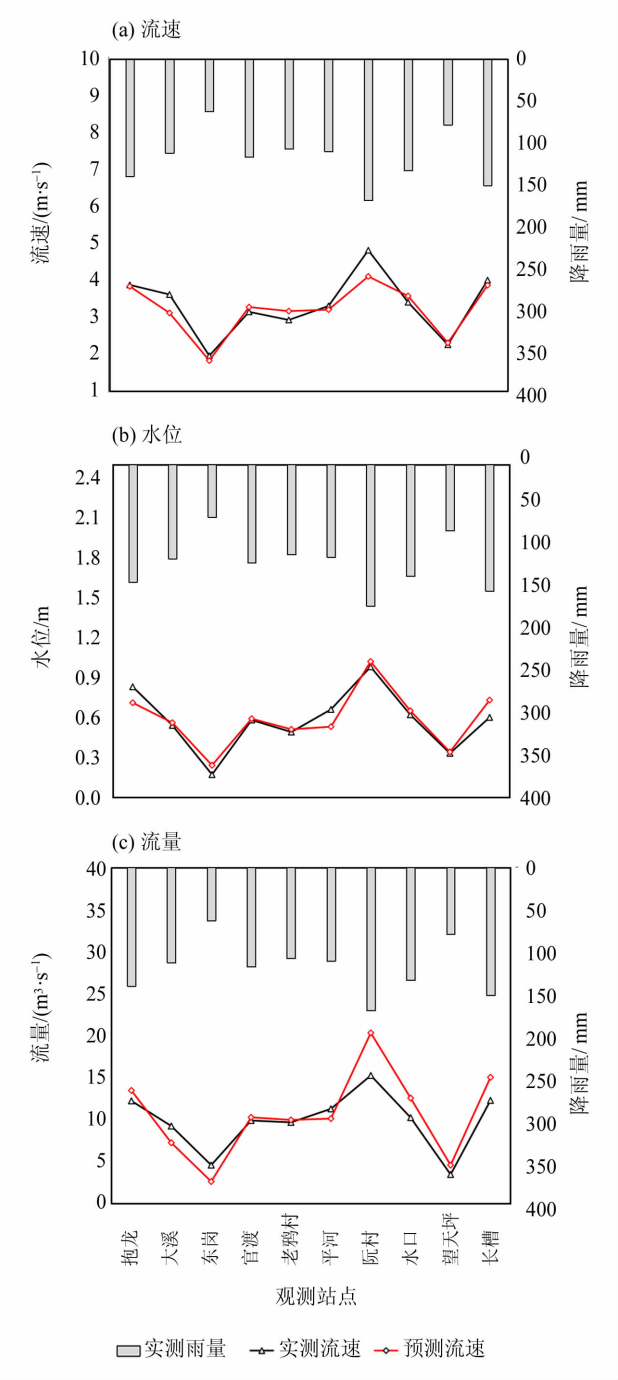


图 1 巫山县模型预测结果对比验证图
Fig. 1 Comparing and authenticating charts of prediction results of Wushan County

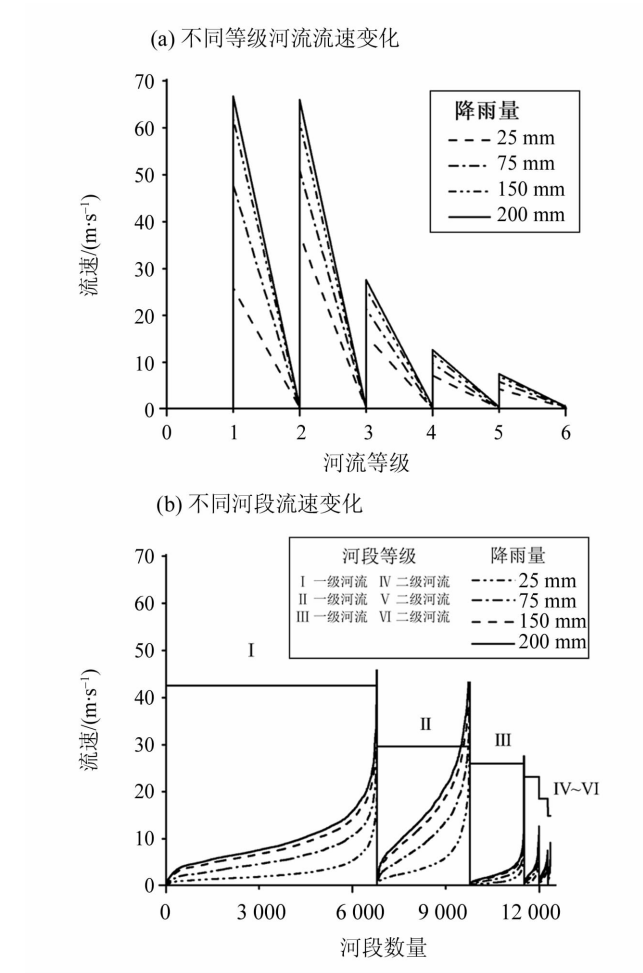


图 2 巫山县不同降雨量下河流与河段流速趋势变化预测
Fig. 2 Prediction of flow velocity trend of river and reach under different rainfall in Wushan County

预测结果中,土壤类型指标重要性最高,其中黄棕壤平均预测值达 8 835,效应值 2 083,影响最为强,而水稻土与黄壤平均预测值为 7 521,效应值为 1 368,影响力次之。而土壤类型在河流、潮土、石灰岩、紫色土、裸岩平均预测值为 6 581,效应值为 -170,影响力较低;土地利用类型中城乡、工矿、居民用地、未利用地、耕地平均预测为 6 751,效应值为 558,林地、水域、草地平均预测值 5 865,效应值为 -327,可见林地、草地保水能力较强,对其水流阻碍能力也强,其预测值低于其他类型用地;当河流比降 > 0.002 时,平均预测值达 8 134,效应值为 612。而当河道比降 ≤ 0.001 时,平均预测值为 5 539,效应值为 -1 438。可见,河流比降预测重要性虽强,但效应值却很低;而分类预测结果中,25、125 和 200 mm 流速预测结果最为精确,统计结果详见表 2。

表 1 巫山县山区河流参数统计表

Table 1 Statistical tables of mountain rivers of Wushan County

河流等级	河道比降	汇流参数	径流系数	阻碍系数	河段周边土地利用类型	土壤类型
1	0.09	68.25	0.23	0.08	林地、草地、耕地	石灰岩、紫色土
2	0.06	69.20	0.24	0.04	林地、耕地	石灰岩、紫色土、黄壤
3	0.11	71.50	0.29	0.28	林地、草地、耕地	石灰岩、黄壤
4	0.13	77.01	0.33	0.28	林地、耕地	石灰岩
5	0.12	87.15	0.59	0.28	林地、水域	石灰岩、紫色土
6	0.09	98.00	0.63	0.28	水域	裸岩、石灰岩

表 2 巫山县不同降雨量的流速、水位、流量巫山县 RF 模型预测结果

Table 2 Prediction results of RF model of velocity, water level and discharge in Wushan County under different rainfall

河流等级	RF 预测值	不同降雨量的流速/(m·s ⁻¹)			不同降雨量的水位/m			不同降雨量的流量/(m ³ ·s ⁻¹)		
		25 mm	100 mm	200 mm	25 mm	100 mm	200 mm	25 mm	100 mm	200 mm
1	5 088	1.09	1.94	3.10	0.94	2.69	4.82	5	74	343
2	5 824	1.40	3.01	4.42	1.25	4.03	6.98	9	169	752
2	6 971	8.15	15.70	21.37	1.67	3.90	5.83	298	1 620	3 852
3	4 761	1.84	4.53	6.17	2.86	9.77	15.19	335	2 892	7 781
3	5 930	4.57	8.86	11.90	3.33	7.68	11.27	641	3 507	8 044
3	6 380	2.00	7.63	10.28	1.00	6.36	9.89	14	1 224	3 970
3	7 258	6.65	18.79	24.81	1.94	7.51	11.20	179	3 719	10 590
3	8 134	4.42	10.63	14.24	1.77	5.79	8.76	193	2 018	5 360
5	5 539	2.46	5.95	7.78	5.21	16.06	23.49	753	6 783	17 220
5	7 373	3.46	8.22	10.73	3.97	12.39	18.14	663	5 774	14 641
6	8 835	2.86	11.14	15.02	0.87	6.45	10.09	20	1 504	2 064

降雨强度达 100 mm 及以上时，研究区流速发生明显变化。而降雨强度在 100 ~ 175 mm 时，流速变化率度最快，且在降雨 175 mm 之后，变化幅度有所减缓；在降雨强度较少情况下，部分流域未能形成径流，出现区域流速为零现象，降雨量小于 25 mm 时，流域流速变化率几乎接近于零值，降雨强度在 25 ~ 50 mm 之间，研究区流速变化率逐渐变快，在 50 ~ 75 mm 之间，流速变化率略为增快；从 12 369 子河流流速变化曲线可以看出，降雨量在 0 ~ 50 mm 之间，径流流速变化集中在 0 ~ 10 m/s，且变化幅度较小；降雨在 50 ~ 100 mm，流速变化集中在 10 ~ 20 m/s 之间，变化幅度逐渐增强；降雨在 100 ~ 200 mm 之间，流速变化集中在 20 ~ 40 m/s 之间，变化幅度明显增强；研究区流速整体上集中在 0 ~ 40 m/s 之间，其变化幅度逐渐增强。

径流水位与流速变化趋势整体相同，降雨强度阈值为 125 mm 时，水位发生明显变化，降雨强度在 150 ~ 175 mm 区间，相较于同等雨量流速相

比，水位变化更为强烈，而降雨在 75 ~ 125 mm 区间时，流速变化幅度大于水位；研究区水位变化集中在 0 ~ 30 m 之间，变化幅度相较于流速变化较小，但水位突变区域比流速突变区域多（集中区域模拟以外曲线即为突变区域）；降雨强度在 0 ~ 75 mm 流速与水位变化基本相同，75 ~ 125 mm 水位变化幅度大于流速，而降雨强度在 125 mm 之后，流速变化幅度则大于水位。

2.3 基于 ARMA 模型的洪水灾害参数指标预测

RF 模型具有分类精度高、指标预测准确等特性。以研究区 12 369 条河段为研究对象，通过 RF 模型进行雨洪流量、流速、水位线性回归关系拟合，进行不同降雨强度情景下流速、水位、流量关系拟合；再者，在不同降雨强度下，进行河流等级回归拟合，最终显著检验 $R^2 > 0.94$ ；最后，通过 K-means 聚类与 RF 统计分析结果，建立流速、水位、断面流量之间转换关系。RF 模型指标重要预测结果中，河流比降、河流等级重要性较高，将其参数以时间散点为时间序列，通过 ARMA 模型参

数预测结果进行对比分析。

从图 3 可以得出，在不同降雨强度情况下，水位变化率最小，而流速变化率最为明显。由此可见，水位变化更易引发洪水灾害，而流速变化，主要表现为构筑物的冲击效应。

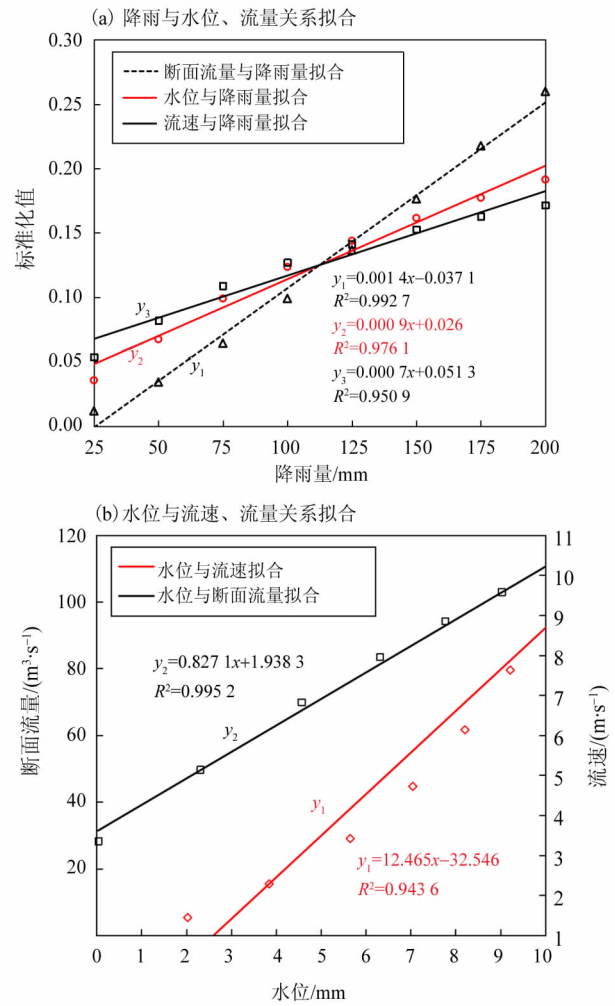


图 3 巫山县山区河流量化关系拟合图

Fig. 3 Fitting diagram of quantitative relationship of mountainous rivers of Wushan County

由于洪水灾害制灾因素在时间序列上存在变动情况，不同时间降雨量不同，导致水文参数变动情况不同。通过 ARMA 模型模拟水文参数在时间序列上的变化。根据水文参数在空间上排序为时间预测目标值，再以不同降雨强度下水文参数量化指标为回归预测值，并对回归残差值进行处理，最终模拟出水文参数指标在时间序列上的回归预测模型（图 4）。

巫山县以山地、丘陵为主，然而河流比降受地形因素影响，河流等级却受汇流累计量影响，因此

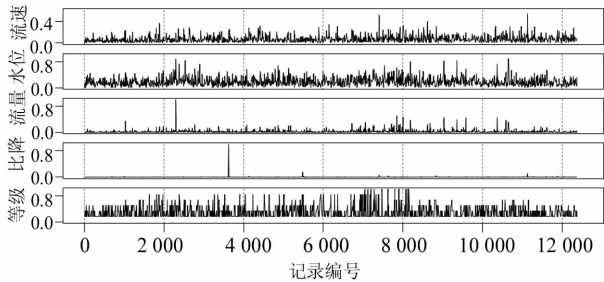


图 4 巫山县 ARMA 模型水文参数拟合图

Fig. 4 Fitting of hydrological parameters based on ARMA model of Wushan County

回归性较差；河流比降回归值介于 0.1 ~ 0.2 之间，回归性较好；河流等级回归性较差区间为在 0 ~ 0.8, 其中 0 ~ 2 000、8 000 ~ 12 000 河段回归拟合较好；通过 ARMA 模型分析，山区径流比降、流量回归拟合较好，流速、水位、河流等级变化幅度大；0 ~ 2 000、8 000 ~ 12 000 段河流回归性较好，且不同河段流速、水位、流量变化与河流等级相关，等级较低河流，其水位、流速变化率最为快，而流量变化率却无显著变化。

ARMA 模型拟合结果显示（图 4），流速整体回归值在 0 ~ 0.4 之间，在 0 ~ 12 369 条河段中，径流流速变化幅度相较于水位变化较小，0 ~ 6 000 条河段变化幅度较小，6 000 ~ 12 000 条河流变化幅度较大；水位回归值在 0 ~ 0.8 之间，整体变化幅度较大，在 0 ~ 2 000、10 000 ~ 12 000 河段变化幅度相对较小，2 000 ~ 8 000 水位变化幅度较大；流量回归变化相对于流速、水位变化幅度最小，0 ~ 6 000 河流流量变化幅度较小，6 400 ~ 8 000 变化幅度最大，8 000 ~ 12 000 变化强度有所减弱；通过 ARMA 模型对研究区河流流速、水位、流量回归模拟可知，流速、水位变化幅度较大，而流量变化幅度较小，可见山区洪水灾害主要为流速与水位作用所致。

2.4 研究区洪水灾害预测结果空间特征分析

巫山县暴雨频发，引发的洪水灾害频繁。根据山区河流洪水突发迅猛、地形地貌复杂等特性，择取统计时间 24h 的实测雨量 25、75、125、200 mm，以巫山县 12 369 河段为研究对象，结合 ArcGIS 空间可视化技术，通过 RF、K-means 与 ARMA 模型模拟结果进行空间可视化（图 5）。综合得出，在不同降雨强度下，巫山县流速水位主要集中在等级较高河流，而在山地区域，流速的空间变化更为剧烈，巫山县中南部与西北部为洪水灾害频发区域。

流速空间变化率图 5 (a, b, c) 可知, 降雨强度临近 25 mm 时, 研究区内径流流速主要集中在中部及长江上游一带变化, 且变化幅度较小, 而在西北部区域, 部分径流流速出现突变现象; 当降雨强度增强至 125 mm 时, 研究区整体流速主要集中在 0 ~ 22 m/s 之内, 部分区域达 46 m/s, 且整体流速为 25 mm 流速一倍, 南部地区流速有增强趋势; 降雨强度达 200 mm 时, 即为大暴雨情景, 流速变化明显增强, 部分区域流速达 58 m/s, 西北部、中南部流速明显增强, 且流速在 20 ~ 58 m/s 之间变化, 洪水灾害发生可能性极大。

相较于流速变化率, 水位变化率主要集中在长江及其西北地区, 图 5 (d, e, f) 可知, 研究区径流整体水位深度变化介于 0 ~ 30 m 之间; 降雨强度越小, 水位变化幅度越小; 等级较低河流在降

雨强度为 25 mm 时, 水位几乎无变化, 主要变化区域集中在长江流域及水库区域; 当降雨强度增强至 125 mm 时, 径流水位发生明显变化, 变化区域主要集中在东北部与中南部, 东北部径流水位呈现上升趋势, 研究区径流最高水位达 20 m; 降雨强度达 200 mm 时, 南部与东北部水位明显上升, 整体水深升至降雨 25 mm 时水位一倍。根据站点统计数据与预测结果对比认证, 以及研究区洪水灾害数据发生情况, 综合得出: 当水位到达 1.2 m 时, 即降雨强度为 125 mm 时, 研究区部分区域出现轻度洪水灾害; 研究区水位升至 10 m 时, 降雨强度介于 125 ~ 175 mm 之间, 出现中度洪水灾害; 当水位超过 10 m 时, 降雨强度在 175 mm 以上时, 研究区出现严重洪水灾害。

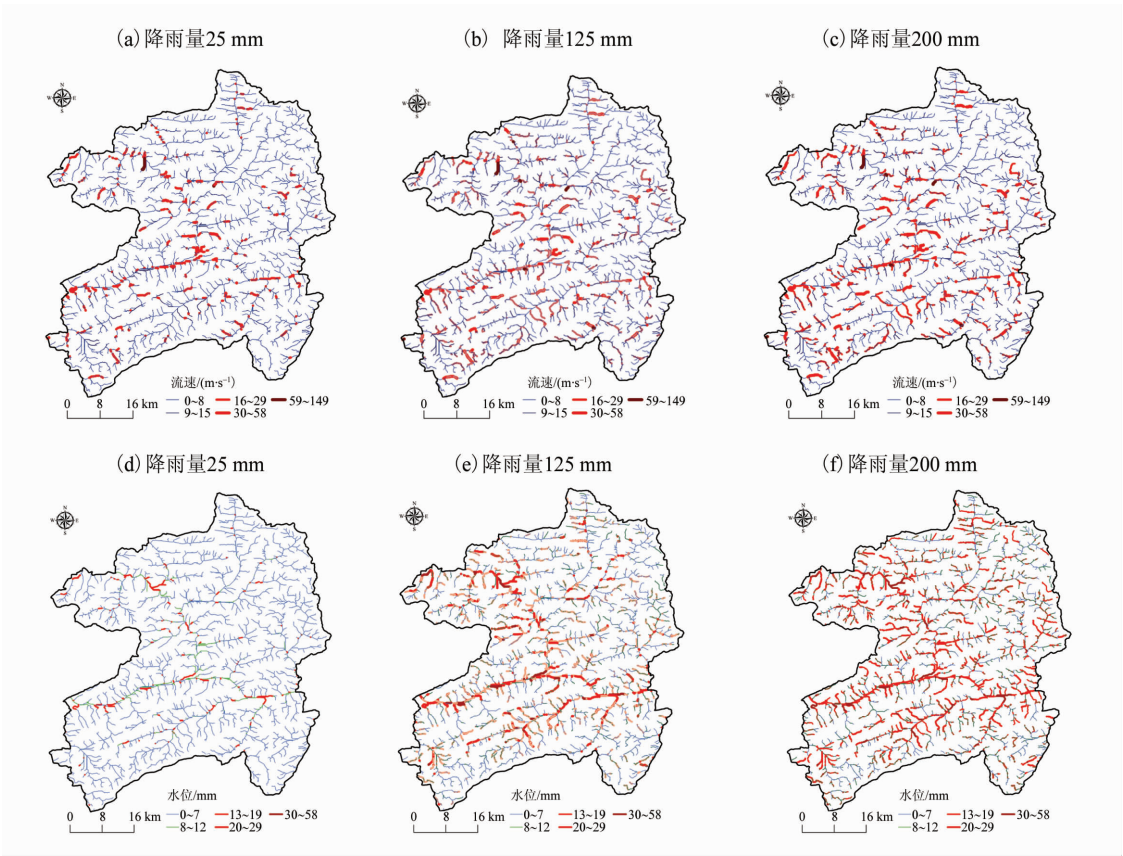


图 5 巫山县流速、水位空间量化图

Fig. 5 Spatial map of velocity and water level of Wushan County

3 结 论

通过 RF、K-mean 与 ARMA 等机器学习算法与模型, 对研究区 12 369 条河段进行统计与分析, 得出了其模型能够精确的拟合出研究区水文参数流

速、水文、流量之间转换关系, 并且有效预测出区域流速、水位变化率。结合 GIS 空间可视化技术, 对降雨 - 径流水文参数空间可视化分析, 综合考虑研究区地理环境要素及孕灾环境, 最终通过 RF 模型对研究区洪水致灾范围进行精细预测。本文根据

机器学习算法特性研究洪水灾害, 由于洪水灾害制灾因素较多, 本文选取指标不够全面, 只在数据分析与算法模型方面研究, 后续研究将综合更多制灾因素进行全面研究, 选取更多适合洪水灾害研究的机器学习算法模型。研究主要结论如下:

1) 研究区指标土壤类型重要性最高, 其影响最为强烈, 其次为土地利用类型、比降; 降雨阈值为 25、125、200 mm 时, 流速分类预测结果最为精确; 通过 RF 模型与 K-means 算法模型对研究区流速与水位预测, 降雨强度在 100 ~ 175 mm 时, 流速变化幅度最为剧烈, 降雨强度阈值达 125 mm 时, 水位变化率最快, 而降雨强度在 150 ~ 175 mm 之间时, 水位变化率最为明显; 在 ARMA 模

型对研究区水文参数指标回归分析中, 河流比降、流量回归性拟合最好, 流速、水位、河流等级变化幅度最大, 且不同河段流速、水位、流量变化率与河流等级相关, 等级较低径流, 其水位、流速变化最为明显, 而较低等级径流, 流量却无显著变化。

2) 机器学习能够有效拟合降雨 - 径流参数关系, 研究区流速、水位变化主要集中于等级较高河流, 而在山区, 流速空间变化更为剧烈; 研究区西北部、中南部流速、水位变化明显, 而当降雨强度增强至 125 mm 时, 即在暴雨、大暴雨及以上情景, 径流水位变化率最为剧烈; 东北部与中南部水位变化率最为明显, 预测水位最高水位达 20 m。

参考文献:

- [1] 张兵. 遥感大数据时代与智能信息提取[J]. 武汉大学学报(信息科学版), 2018, 43(12): 1861 - 1871.
ZHANG B. Remotely sensed Big Data era and Intelligent Information extraction [J]. Geomatics and Information Science of Wuhan University, 2018, 43(12): 1861 - 1871.
- [2] 胡胜, 曹明明, 邱海军, 等. CFSR 气象数据在流域水文模拟中的适用性评价——以澜河流域为例[J]. 地理学报, 2016, 71(9): 1571 - 1586.
HU S, CAO M M, QIU H J, et al. Applicability evaluation of CFSR climate data for hydrologic simulation: A case study in the Bahe River Basin[J]. Acta Geographica Sinica, 2016, 71(9): 1571 - 1587.
- [3] 吴裕珍, 冯志州, 王大刚. 基于贝叶斯模式平均与标准化异常度的东江汛期降水预报[J]. 中山大学学报(自然科学版), 2016, 55(6): 20 - 27.
WU Y Z, FENG Z Z, WANG D G. Precipitation forecasting in flood season over the Dongjiang Basin using Bayesian model averaging and standardized anomaly [J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2016, 55(6): 20 - 27.
- [4] 刘志方, 刘友存, 郝永红, 等. 黑河出山径流过程与气象要素多尺度交叉小波分析[J]. 干旱区地理, 2014, 37(6): 1137 - 1146.
LIU Z F, LIU Y C, HAO Y H, et al. Multi-time scale cross-wavelet transformation between runoff and climate factors in the upstream of Heihe River[J]. Arid Land Geography, 2014, 37(6): 1137 - 1146.
- [5] 顾西辉, 张强, 孔冬冬, 等. 低频气候变化引起的珠江流域年均和洪峰流量变化特征及灵敏度分析[J]. 中山大学学报(自然科学版), 2017, 56(1): 138 - 144.
GU X H, ZHANG Q, KONG D D, et al. Low frequency climate variability induced annual mean and annual maximum streamflow variations and related sensitivity across the Pearl River basin[J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2017, 56(1): 138 - 144.
- [6] 闪丽洁, 张利平, 张艳军, 等. 长江中下游流域旱涝急转事件特征分析及其与 ENSO 的关系[J]. 地理学报, 2018, 73(1): 25 - 40.
SHAN L J, ZHANG L P, ZHANG Y J, et al. Characteristics of dry - wet abrupt alternation events in the middle and lower reaches of the Yangtze River Basin and their relationship with ENSO[J]. Acta Geographica Sinica, 2018, 73(1): 25 - 40.
- [7] 王光朋, 查小春, 黄春长, 等. 基于 HEC - RAS 模型的汉江上游东汉时期古洪水事件研究[J]. 中山大学学报(自然科学版), 2018, 57(3): 44 - 53.
WANG G P, ZHA X C, HUANG C C, et al. Research on palaeoflood event in the Eastern Han Dynasty in upper reaches of Hanjiang River based on HEC-RAS model[J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2018, 57(3): 44 - 53.
- [8] 石振明, 沈丹祎, 彭铭, 等. 考虑多层非饱和土降雨入渗的边坡稳定性分析[J]. 水利学报, 2016, 47(8): 977 - 985.
SHI Z M, SHEN D Z, PENG M, et al. Slope stability analysis by considering rainfall infiltration in multi-layered unsaturated soils[J]. Journal of Hydraulic Engineering,

- 2016, 47 (8): 977 – 985.
- [9] 张社荣, 谭尧升, 王超, 等. 强降雨特性对饱和 – 非饱和边坡失稳破坏的影响[J]. 岩石力学与工程学报, 2014, 33(S2): 4102 – 4112.
- ZHANG S R, TAN Y S, WANG C, et al. Effects of strong rainfall on instability and failure of saturated – unsaturated slopes[J]. Chinese Journal of Rock Mechanics and Engineering, 2014, 33(S2): 4102 – 4112.
- [10] 李骅锦, 许强, 何雨森, 等. “阶跃式”滑坡位移预测及阈值分析的 ARMA – (LASSO – ELM) – Copula 模型[J]. 岩石力学与工程学报, 2017(S2): 4075 – 4084.
- LI H J, XU Q, HE Y S, et al. An ARMA – (LASSO – ELM) – Copula framework for landslide displacement prediction and threshold computing of the displacement of step – like landslides[J]. Chinese Journal of Rock Mechanics and Engineering, 2017 (S2): 4075 – 4084.
- [11] 张国芳, 查小春, 石晓静, 等. 汉江上游汉中市洪水灾害风险评价研究[J]. 中山大学学报(自然科学版), 2016, 55(6): 28 – 34.
- ZHANG G F, ZHA X C, SHI X J, et al. Flood hazard risk assessment of Hanzhong city at the upper reaches of Hanjiang River[J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2016, 55(6): 28 – 34.
- [12] 赖成光, 陈晓宏, 赵仕威, 等. 基于随机森林的洪灾风险评价模型及其应用[J]. 水利学报, 2015, 46(01): 58 – 66.
- LAI C G, CHEN X H, ZHAO S W, et al. A flood risk assessment model based on Random Forest and its application[J]. Journal of Hydraulic Engineering, 2015, 46(1): 58 – 66.
- [13] 刘家福, 李林峰, 任春颖, 等. 基于特征优选的随机森林模型的黄河口滨海湿地信息提取研究[J]. 湿地科学, 2018, 16(2): 97 – 105.
- LIU J F, LI L F, REN C Y, et al. Information extraction of coastal wetlands in Yellow River estuary by optimal feature based Random Forest Model[J]. Wetland Science, 2018, 16(2): 97 – 105.
- [14] 符素华, 王红叶, 王向亮, 等. 北京地区径流曲线数模型中的径流曲线数[J]. 地理研究, 2013, 32(5): 797 – 807.
- FU S H, WANG H Y, WANG X L. The runoff curve number of SCS – CN method in Beijing[J]. Geographical Research, 2013, 32(5): 797 – 807.
- [15] TONG S T Y, SUN Y, RANATUNGAT, et al. Predicting plausible impacts of sets of climate and land use change scenarios on water resources[J]. Applied Geography, 2012, 32(2): 477 – 209.
- [16] LEE S, NI – MESISTER W, TOLL D, et al. Assessing the hydrologic performance of the EPA’s nonpoint source water quality assessment decision support tool using North American Land Data Assimilation System (NL-DAS) products[J]. Journal of Hydrology, 2010, 387(3): 212 – 220.
- [17] 孙东坡, 张晓雷, 陈丹等. 基于错峰型洪水的伊洛河交汇区水动力特性[J]. 水科学进展, 2013, 24(6): 801 – 809.
- SUN D P, ZHANG X L, CHEN D, et al. Hydrodynamic characteristics of staggering peak flows at the confluence section of Yiluo River[J]. Advances in Water Science, 2013, 24(6): 801 – 809.
- [18] 孙周亮, 刘冀, 谈新等. 近 50a 渭河上游汛期降雨径流多尺度时空演变[J]. 长江流域资源与环境, 2018, 26(6): 1324 – 1332.
- SUN Z L, LIU J, TAN X, et al. Multiscale spatio – temporal evolution analysis of rainfall and runoff in flood season in upper Huanhe River in recent 50 years[J]. Resources and Environment in the Yangtze Basin, 2018, 26(6): 1324 – 1332.

(责任编辑 秦社彩)