

DOI: 10. 3880/j. issn. 1004 - 6933. 2022. 02. 005

# 基于机器学习的山洪灾害快速预报方法

周 聂<sup>1</sup>, 侯精明<sup>1</sup>, 陈光照<sup>1</sup>, 马红丽<sup>2</sup>, 洪增林<sup>3</sup>, 李新林<sup>3</sup>

(1. 西安理工大学省部共建西北旱区生态水利国家重点实验室, 陕西 西安 710048;  
2. 鄂尔多斯市水利勘测设计院, 内蒙古 鄂尔多斯 017000; 3. 陕西省地质调查院, 陕西 西安 710054)

**摘要:** 基于高精度水动力模型与机器学习技术, 运用极限随机树(ERT)及 KNN 算法, 构建了高分辨率山洪灾害快速预报模型。利用确定系数、平均绝对误差和均方根误差 3 种指标评估模型的整体可靠性, 同时, 截取流域出口断面流量验证模型的预报性能。结果表明: 所建模型预报结果与水动力模型模拟结果淹没范围基本一致, 流域淹没范围平均相对误差低于 5%, 模型整体稳定可靠; 流域出口断面流量平均相对误差低于 10%, 断面平均水深、流速平均相对误差低于 5%, 模型预报性能良好; 模型可在 10s 内完成最大淹没情况计算并输出淹没范围图, 能为紧急决策提供足够的前置时间, 协助决策者更好地采取应对措施。

**关键词:** 山洪灾害; 快速预报; 机器学习; 极限随机树; KNN 算法; 水动力模型

**中图分类号:** P45 **文献标志码:** A **文章编号:** 1004 - 6933(2022)02 - 0032 - 09

**A rapid forecasting method for mountain flood disaster based on machine learning algorithms**//ZHOU Nie<sup>1</sup>, HOU Jingming<sup>1</sup>, CHEN Guangzhao<sup>1</sup>, MA Hongli<sup>2</sup>, HONG Zenglin<sup>3</sup>, LI Xinlin<sup>3</sup> (1. State Key Laboratory of Eco-Hydraulics in Northwest Arid Region of China, Xi'an University of Technology, Xi'an 710048, China; 2. Ordos Water Conservancy Survey and Design Institute, Ordos 017000, China; 3. Shaanxi Institute of Geological Survey, Xi'an 710054, China)

**Abstract:** Based on high-precision hydrodynamic model and machine learning technology, a high-resolution rapid forecasting model of mountain flood disaster was constructed by using extremely randomized trees and KNN algorithm. The overall reliability of the model was evaluated by three indexes: determination coefficient, mean absolute error and root mean square error. At the same time, the discharge of the outlet section of the basin was intercepted to verify the prediction performance of the model. The results show that the prediction results of the model are basically consistent with the inundation range of the hydrodynamic model simulation results, the average relative error of the basin inundation range is less than 5%, and the model is stable and reliable as a whole. The relative error of discharge at the outlet section of the basin is less than 10%, the average relative error of water depth and velocity at the section is less than 5%, and the prediction performance of the model is good. The model can calculate the maximum inundation situation and output the inundation range map within 10s, which can provide sufficient lead time for emergency decision-making and assist decision-makers to take better countermeasures.

**Key words:** mountain flood disaster; rapid forecasting; machine learning; extremely randomized trees; KNN algorithm; hydrodynamic model

山洪指在山区溪沟中由于强降雨等原因引起水位短时间内暴涨形成的洪水, 具有流速大、冲刷力强、破坏力强等特点, 通常会给交通、农业、居民生命财产安全等造成巨大威胁<sup>[1]</sup>。据《2017 年中国生态环境状况公报》统计, 2017 年洪涝灾害累计受灾人

口 5 515 万人, 因灾直接造成经济损失达 2 143 亿元<sup>[2]</sup>。随着全球气候变化, 极端降雨事件频率增大, 2018 年全国累计 454 条河流出现超过警戒水位的洪水, 其中 24 条河流发生超历史洪水<sup>[3]</sup>。由于山洪灾害具有突发性及破坏性特点, 如何应对山洪灾

基金项目: 国家重点研发计划(2016YFC0402704); 国家自然科学基金(51609199); 陕西省水利科技项目(2017SLKJ-44); 固原海绵城市建设示范区海绵效果数值模拟(SCHM-2018-0104)

作者简介: 周聂(1996—), 男, 硕士研究生, 主要从事水力学及河流动力学研究。E-mail: 2190420070@stu.xaut.edu.cn

通信作者: 侯精明(1982—), 男, 教授, 博士, 主要从事地表水及水力学数值模拟研究。E-mail: jingming.hou@xaut.edu.cn

害降低损失,一直以来都是国内外学者研究的热点与难点。

山洪灾害模拟预报能够有效模拟山洪灾害的受灾范围,为山洪风险管理、应急决策提供依据,是协助防灾减灾工作的重要方式之一。众多学者对此进行了研究,包红军等<sup>[4]</sup>基于新安江模型开发了分布式混合水文模型,模型预报精度较高;胡国华等<sup>[5]</sup>用 HEC-HMS 半分布式水文模型进行山洪模拟,洪峰流量及洪量相对误差可控制在 20% 以内;孟天翔<sup>[6]</sup>基于 MIKE FLOOD 构建水动力模型开展山洪数值模拟,获得各类洪水淹没演进变化情况。然而传统山洪模拟研究都有一个共同的缺陷,即需要求解复杂的方程组,按步时顺序迭代以获取洪水的演进过程,模拟运算耗时长,无法满足紧急决策的需求。近年来,随着人工智能技术的发展,机器学习技术以其普适性与高效性,在医学、化学、数据挖掘等领域都得到了广泛应用并展现出突出优势<sup>[7-9]</sup>,国内外学者也逐渐将人工神经网络、逻辑回归、极限树回归等机器学习算法应用于洪涝灾害的预测中<sup>[10-12]</sup>。阚光远等<sup>[13]</sup>将 ANN 与 KNN 方法相耦合,提高了 ANN 预报能力不佳等问题;张珂等<sup>[14]</sup>将洪水预报智能模型应用于中国半干旱半湿润地区,所建模型在半湿润区典型流域可获得良好的预报结果;张轩等<sup>[15]</sup>基于 BP 神经网络算法建立经验预报,模型预报精度可达乙级以上;Chang 等<sup>[16]</sup>基于 ANN 构建了洪水淹没多步预测模型;Fauzi 等<sup>[17]</sup>同样基于机器学习算法构建了海啸淹没的预报模型。这些研究表明,机器学习技术在洪水预报方面同样具有广阔的应用前景。

机器学习往往需要大量的训练数据,目前洪水灾害快速预报多基于水文站多年的实测数据,依靠完备的大数据进行训练,但对于未发生过洪水或历史淹没资料匮乏的地区无法进行有效预测。为此,本文基于高精度水动力模型与机器学习技术,选取王茂沟流域为研究区域,以数值模型模拟数据作为驱动数据,探索构建实测洪水资料匮乏流域的山洪灾害快速预报模型,旨在为紧急决策提供足够的前置时间,协助决策者更好地采取应对措施。

## 1 研究区概况与降雨资料

### 1.1 研究区概况

本文选取陕西省榆林市绥德县的王茂沟流域作为研究区域,该流域位于东经  $110^{\circ}20'26'' \sim 110^{\circ}22'46''$ ,北纬  $37^{\circ}34'13'' \sim 37^{\circ}36'03''$ ,总面积约为  $5.74 \text{ km}^2$ ,王茂沟主沟长  $3.75 \text{ km}$ ,沟道平均比降为

$2.7\%$ 。流域海拔高度介于  $940 \sim 1200 \text{ m}$ ,属温带半干旱大陆性季风气候,历史最大年降水量为  $735.3 \text{ mm}$ ,多年平均降水量为  $475.1 \text{ mm}$ ,7—9 月降水量可达全年降雨的  $64\%$ ,且多为短历时强降雨,极易引起山洪、泥石流等灾害。2012 年 7 月 15 日及 2017 年 7 月 26 日均出现高强度暴雨,形成山洪,造成巨大经济损失<sup>[18-19]</sup>。

地形精度对数值模拟至关重要,精度不足将无法反映实际地形特征,影响模拟的可靠性;精度过高对模拟精确性提高不大,但会严重延长模型的运行时间。经综合考虑,将研究区域以  $5 \text{ m}$  的水平分辨率划分为  $232037$  个网格,其数字高程图见图 1,土地利用情况见图 2。同时,为更好地表征下垫面信息,将研究区根据不同的土地利用类型,基于最大似然法划分为 8 类。预报模型采用 Green-Ampt 入渗模型,下渗系数根据文献[20]确定,土地的曼宁系数根据文献[21]确定,各土地利用类型具体参数见表 1。

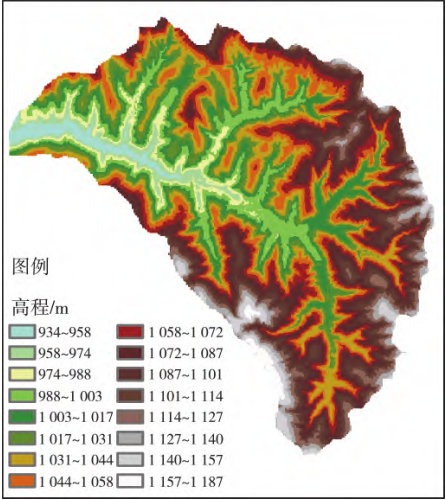


图 1 研究区域数字高程

Fig. 1 DEM of research area

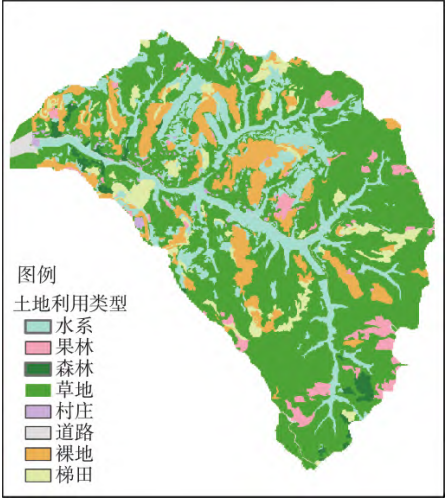


图 2 研究区土地利用情况

Fig. 2 Land use situation of research area

表 1 各类型土地曼宁及下渗参数  
Table 1 Manning and infiltration parameters  
of various types of land

土地利用类型	面积/km <sup>2</sup>	曼宁系数	有效饱和导水率/(cm·s <sup>-1</sup> )	湿润锋平均基质吸力/cm
水系	1.096	0.03	0	0
果林	0.211	0.15	0.008	10.51
森地	0.086	0.20	0.035	19.17
草地	3.336	0.06	0.023	6.39
村庄	0.037	0.01	0.001	2.52
道路	0.055	0.01	0	0
裸地	0.657	0.04	0.005	4.21
梯田	0.323	0.18	0.004	3.59

## 1.2 降雨资料

本研究降雨资料时间分辨率取 1 h。研究区实测降雨资料有限,仅用历史数据无法完成模型训练,因此,在进行模型训练时,同时加入了历史降雨资料与设计降雨资料。张茹等<sup>[22]</sup>研究表明 GPM(global precipitation measurement)在强降水区域表现出较高的命中率和较低的误报率,卫星产品对强降水的探测能力较优,故历史降雨资料从 GPM 网站(<https://gpm.nasa.gov>)获取<sup>[23]</sup>。本研究主要针对短历时强降雨引发的山洪事件,与芝加哥雨型具有一定相似特征,薛宇雷等<sup>[24-25]</sup>通过芝加哥雨型生成器生成短历时设计降雨,应用于周河流域及梅溪流域,均取得良好模拟效果。据此,研究选用芝加哥雨型生成器生成各类情况下的短历时设计降雨。榆林市暴雨公式为

$$i = \frac{8.22(1 + 1.152lgp)}{(t + 9.44)^{0.746}} \quad (1)$$

式中:  $i$  为暴雨强度, mm/h;  $p$  为重现期, a;  $t$  为降雨历时, min。

## 2 快速预报方法

### 2.1 快速预报流程

为实现山洪灾害快速精准预测,基于水动力模型,结合机器学习算法构建了山洪灾害快速预报模型。预报流程图如图 3 所示,首先输入地形数据、降雨数据等驱动水动力模型,获得降雨-致洪数据,并将其划分为训练集与测试集,供机器学习算法拟合使用;同时,从降雨数据中提取特征参数,经相关性分析,获取用于算法拟合的最终参数;通过机器学习算法拟合训练集与特征参数,通过网格搜索算法寻找最优参数,获取初步模型;再通过机器学习算法预报结果与水动力模型模拟结果获得误差矩阵,并将误差作为训练数据生成误差修正模型;预报结果由机器学习算法预报模型与误差修正模型共同生成,同时为降低单一算法自身局限性,将两种算法(KNN 算法和极限随机树(ERT)算法)结果按其

训练集中评价指标得分进行权值分配,生成混合模型预报结果,并以测试集数据检验模型预报可靠性,若误差过大,将参数调整后重新生成模型,直至预报精度达到要求后,得到最终预报模型。

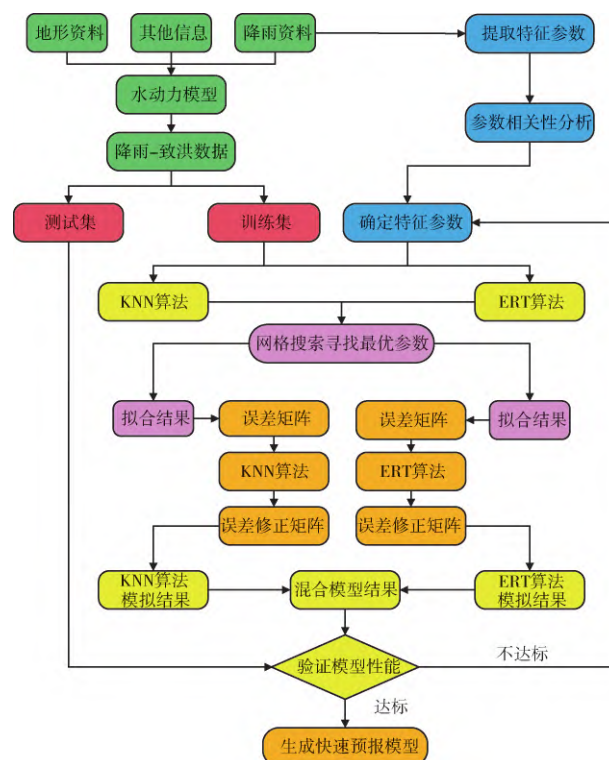


图 3 快速预报流程

Fig. 3 Flow chart of rapid forecasting model

洪水淹没范围、过水断面洪峰流量、最大流速以及最大水深是评价洪水风险的重要指标,因地形数据中可能存在的个别噪点会严重影响过水断面的最大深度,因而本文主要采用洪水淹没面积、断面流量、平均流速及水深来评价模型的可靠性。

### 2.2 水动力模型

模型控制方程为耦合水文过程的二维浅水方程,忽略运动黏性项、科氏力、风应力及紊流黏性项。其对应的二维非线性浅水方程守恒格式的矢量形式表示如下:

$$\frac{\partial \mathbf{q}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} + \frac{\partial \mathbf{G}}{\partial y} = \mathbf{S} \quad (2)$$

$$\text{其中 } \mathbf{q} = \begin{bmatrix} h \\ q_x \\ q_y \end{bmatrix} \quad \mathbf{F} = \begin{bmatrix} uh \\ uq_x + gh^2/2 \\ uq_y \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} vh \\ vq_x \\ vq_y + gh^2/2 \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} r \\ -gh\partial z_b/\partial x \\ -gh\partial z_b/\partial y \end{bmatrix} + \begin{bmatrix} 0 \\ -C_t u \sqrt{u^2 + v^2} \\ -C_t v \sqrt{u^2 + v^2} \end{bmatrix}$$

式中:  $q$  为变量矢量, 包括水深  $h$  以及  $x$  和  $y$  方向的单宽流量  $q_x$  和  $q_y$ ;  $u, v$  分别为  $x$  和  $y$  方向的流速;  $F$  和  $G$  分别为  $x, y$  方向的通量矢量;  $r$  为净雨率;  $S$  为源项矢量;  $z_b$  为河床底面高程;  $C_f = gn^2/h^{1/3}$  为河床糙率系数;  $n$  为曼宁系数。

模型选取对物理问题考虑最全面的动力波法进行地表洪水过程的模拟运算, 通过基于 Godunov 格式的有限体积法进行空间离散求解二维浅水方程<sup>[26-27]</sup>, 并运用 Runge-Kutta 方法构造具有二阶时空精度的 MUSCL (monotonic upwind scheme for conservation) 型格式, 确保物质守恒并有效解决不连续问题<sup>[28]</sup>。针对模拟过程中可能产生急变流与非连续等复杂问题, 模型选用 HLLC 近似黎曼求解器对单元界面上的质量以及动量的通量进行求解。通过静水重构法处理干湿边界处可能出现负水深的问题<sup>[29]</sup>, 并以流速替代单宽流量作为计算变量, 可在水深低于或流速高于一定值时, 将易失稳的二阶格式有效地转换为稳定的一阶计算格式。保证计算精度的同时, 将计算单元中的坡面源项转换为该单元边界上的通量, 确保其在进行复杂地形计算时同样满足全稳条件。摩阻源项使用 Hou 等<sup>[30]</sup> 优化的分裂点隐式法确保运算结果的稳定性。

该模型可对下垫面条件复杂区域进行高精度的雨洪模拟, 并采用 GPU (graphics processing unit) 并行技术加速模拟计算过程, 保障模型运算效率。侯精明等<sup>[31]</sup> 对流域模拟数据与实测数据进行了可靠性对比, 结果表明模型模拟性能良好; 并且随后的文献<sup>[32]</sup> 中另一区域模拟结果与实测数据同样符合, 再次验证了模型的可靠性。

## 2.3 机器学习算法

机器学习是人工智能的核心, 包含有 SVM、逻辑回归、决策树、随机森林、KNN 等算法, 本研究主要选用 ERT 算法及 KNN 算法构建山洪灾害快速预报模型。

### 2.3.1 ERT 算法

ERT 算法是 Geurts 等<sup>[33]</sup> 于 2006 年提出的一种基于决策树的集成学习算法。该方法类似于随机森林算法<sup>[34]</sup>, 通过自上而下的方式生成若干决策树, 运用多棵决策树的综合结果进行预测。训练模型时每棵决策树使用相同的训练样本; 节点分裂时, 省略对 GINI 系数或均方差的计算过程, 采用参数完全随机的方式提升算法的随机性。因节点分裂属性完全随机决定, 单棵决策树的拟合精度通常较低, 但综合多棵决策树结果, 可大幅提升其预报精度。

在 ERT 算法的众多参数中, 模型性能对结点处随机选择的特征个数 Max features 及构成模型的决

策树深度 Max depth 较敏感。对于同一组训练数据, Max features 值越小, 模型随机性越强, 同时对训练样本的输出值的依赖性越弱。预测误差随 Max depth 的增加而减小, 但 Max depth 值过大将造成模型的过拟合, 致使模型在训练集上可达到近乎完美的表现, 但在测试集中会出现较大误差, 模型参数的选取需根据具体数据集确定<sup>[35]</sup>。

### 2.3.2 KNN 算法

KNN 算法是 Cover 和 Hart 基于向量空间模型于 1968 年提出的一种机器学习算法<sup>[36]</sup>, 该算法具有理论成熟、方法简单以及鲁棒性良好的优势, 可有效过滤训练集中的噪声。在 KNN 算法中, 每个样本均被视为  $R^n$  空间中的向量或坐标点, 利用距离公式找出与待分类样本最近的  $K$  个样本, 以此进行样本估计。其主要步骤如下:

步骤 1 对已有样本进行实例化, 转换为  $(x, f(x))$  的形式, 其中  $x$  为样本的特征参数,  $x$  由  $(x_1, x_2, \dots, x_n)$  表示,  $x_n$  为样本的第  $n$  个属性值, 即特征参数的数量等于向量组成的维度, 实例化后所有的样本构成训练集与测试集。

步骤 2 给定一个新的测试样本  $x_i$ , 运用距离公式分别计算  $x_i$  与训练集中各个原本样本之间的距离, 并从中筛选出与  $x_i$  距离最近的  $K$  个样本。其距离公式主要有: 曼哈顿距离公式、欧式距离公式及闵可夫斯基距离公式等, 通过拟合效果综合对比, 最终选择在训练集及测试集上均表现良好的欧式距离公式:

$$L(x_i, x_j) = \left[ \sum_{l=1}^n |x_{il} - x_{jl}|^2 \right]^{1/2} \quad (3)$$

式中:  $x_i, x_j$  为两个样本;  $x_{il}, x_{jl}$  分别为  $x_i$  和  $x_j$  的第  $l$  个特征值;  $L(x_i, x_j)$  为样本  $x_i$  和  $x_j$  之间的欧式距离。

步骤 3 将选择出的  $K$  个样本与未知样本的接近程度, 按权重分配  $K$  个样本的预测结果, 并将其分配给新的测试样本, 作为预报值。

## 2.4 基于网格搜索的预报模型参数优化算法

机器学习算法对算法参数十分敏感, 算法最大深度、最大特征数等参数的选取不合适将直接导致模型预报的失败。为使预报模型能有效拟合所提供数据, 研究采用网格搜索的方法对机器学习算法参数进行优选。网格搜索是一种穷举型算法, 可自动模拟各类参数的组合情况, 进行多次模型训练, 并由交叉验证进行误差对比, 最终找到训练误差最小的模型, 确定出最适合训练数据的最优的参数组合, 进而通过对模型参数的优化提升模型的可靠性。研究中, 交叉验证系数选取为 0.2, 最终确定 ERT 算法的最大深度为 10, 误差公式选取均方根误差 (RMSE), 最小分



裂数为 2; KNN 模型  $K$  邻近个数取为 3, 采用球状树算法进行数据分割, 距离公式选择为欧式距离公式。

### 2.5 误差修正及模型评价指标

为降低由水动力模型及机器学习造成的误差累积, 利用水动力模型模拟结果与机器学习算法初步预报结果生成误差矩阵, 而后以降雨特征参数为输入条件, 误差矩阵为输出结果, 运用机器学习算法建立降雨特征参数与误差矩阵之间的对应关系, 构建误差修正模型。机器学习算法最终预报结果由初步预报结果累加上误差修正模型所得出的对应误差矩阵生成。同时, 为降低由单一机器学习算法本身的缺陷产生的误差, 通过将 KNN 算法模型及 ERT 算法模型的模拟结果依据其在训练数据集中的确定系数( $R^2$ ), 进行加权分配, 获得混合模型预报结果。计算公式为

$$R_C = \frac{1}{S_{ERT} + S_{KNN}}(R_{ERT}S_{ERT} + R_{KNN}S_{KNN}) \quad (4)$$

式中:  $R_C$  为混合模型预报结果;  $R_{ERT}$  和  $R_{KNN}$  分别为 ERT 模型和 KNN 模型各个网格上的预报结果;  $S_{ERT}$  和  $S_{KNN}$  分别为 ERT 模型和 KNN 模型的  $R^2$  值。

选择  $R^2$ 、绝对平均误差(MAE)和 RMSE 作为模型整体可靠性的评价指标。

## 3 结果与分析

在洪水安全应急管理中, 决策者最关心的往往是洪水最大淹没面积及洪峰流量情况, 据此, 通过水动力模型累计模拟 108 场降雨事件, 其中历史降雨事件为 18 场, 设计降雨事件 90 场, 获得其雨洪过程作为样本数据, 运用 ERT 算法及 KNN 算法训练生成快速预报模型, 运用  $R^2$ 、MAE 和 RMSE 评估预报整体性能。此外, 在流域主沟道下游出口处选择断面, 通过计算出口断面的洪峰流量、流速及由流量与流速近似估算出的断面平均水深对模型预报性能进一步进行复核。

### 3.1 水动力模型模拟性能验证

快速预报模型的训练数据均为水动力模型模拟结果, 因此在进行模型构建时, 需先对水动力模型的模拟性能进行验证。模型验证所用降雨资料为王茂沟水文站观测数据, 降雨开始于 2012 年 7 月 15 日 00:25, 降雨历时为 5 h, 流量资料为王茂沟水文站实

测数据, 模拟时长为 10 h, 流域出口处的模拟流量过程与实测流量过程对比见图 4。由图 4 可见, 水动力模型模拟结果与观测流量数据变化趋势基本一致, 流量峰值滞后约 0.5 h, 流量消退过程较观测数据稍快, 因遥测地形数据与实际地形存在差异, 在进行土地利用划分时产生了一定误差, 但模型总体效果良好, 纳什系数可达 0.78, 可有效模拟该流域降雨致洪过程。

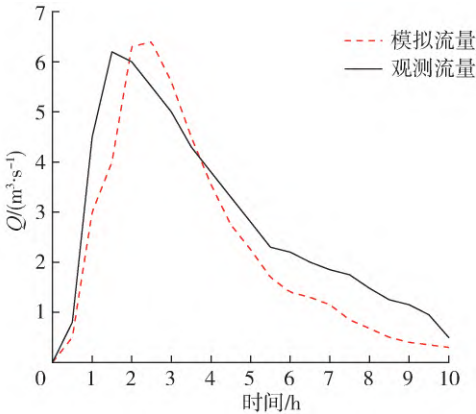


图 4 流域出口流量过程对比

Fig. 4 Comparison of discharge process at outlet

### 3.2 降雨特征参数相关性分析

在进行机器学习训练时, 输入参数的选取至关重要, 合理优选参数即可保证模型精度, 也可增强模型的时效性<sup>[37]</sup>。因此, 研究选用皮尔逊相关性分析方法对模型所选降雨特征参数进行相关性分析。皮尔逊相关系数小于 0.4 时为弱相关、极弱相关或无相关, 0.4~0.6 为中等强度相关, 0.6~1 为强相关或极强相关。计算公式为

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (5)$$

式中:  $\rho_{xy}$  为皮尔逊相关系数;  $\sigma_{xy}$  是参数  $x$  和  $y$  的协方差;  $\sigma_x$ 、 $\sigma_y$  分别为  $x$  和  $y$  的方差。

选用流域总淹没面积、总淹没水量与所选降雨特征参数进行相关性分析, 相关系数计算结果见表 2。降雨重现期、降雨峰值、最大 3 h 降水量、最大 5 h 降水量、累计降水量及峰值前降水量与淹没面积及淹没水量都具有较强的相关性。总淹没面积、总淹没水量与降雨历时的皮尔逊相关系数分别为 0.054 和 0.096, 属于极弱相关或无相关, 因此最终降雨特征参数舍弃降雨历时, 保留其余参数。

表 2 降雨特征参数相关性分析

Table 2 Correlation analysis of characteristic parameters of rainfall

指标	皮尔逊相关系数					
	降雨重现期	降雨峰值	最大 3 h 降水量	最大 5 h 降水量	累计降水量	峰值前降水量
淹没面积	0.709	0.976	0.976	0.975	0.961	0.780
淹没水量	0.696	0.979	0.975	0.975	0.972	0.813

3.3 预报模型整体性能评估

选取 93 场降雨事件作为模型训练数据,15 场作为测试数据,分别运用 ERT 算法、KNN 算法和混合算法构建山洪快速预报模型,3 种模型的统计指标如表 3 所示。由表 3 可知,所有模型  $R^2$  值均可达到 0.90 以上,说明模型能对数据形成较好拟合,可应用于洪水快速预报研究。在对水深进行模拟时,各预报模型均能获得较好的得分,其中 ERT 模型略优, $R^2$  值为 0.958 7,MAE 值为 0.001 4,RMSE 值为 0.010 0;在进行流速预报时,KNN 模型预报效果最佳, $R^2$  值为 0.951 7,模拟效果较好,ERT 模型预报性能略差于 KNN 模型。混合模型在对水深及流速的整体预测上表现最好, $R^2$  值均达到 0.95 以上,MAE 值和 RMSE 值均相对较小。

表 3 快速预报模型评估指标值

Table 3 Evaluation index values of rapid forecasting model

模型	指标	$R^2$	MAE	RMSE
ERT 模型	水深	0.958 7	0.001 4	0.010 0
	流速	0.921 7	0.000 8	0.009 1
KNN 模型	水深	0.907 5	0.001 4	0.014 9
	流速	0.951 7	0.000 5	0.006 3
混合模型	水深	0.954 0	0.001 4	0.010 8
	流速	0.950 1	0.000 6	0.006 8

图 5 为水动力模型、KNN 模型、ERT 模型以及混合模型在出口断面达洪峰流量时刻的水深图,可

以看出,KNN 模型、ERT 模型以及混合模型在出口断面达洪峰流量时刻,对流域范围内水深模拟结果与水动力模型模拟结果基本一致,能准确反映各处水深与各沟道中的行洪情况。

图 6 为水动力模型、KNN 模型、ERT 模型以及混合模型对 15 场测试降雨事件模拟淹没面积和淹没水量的结果。图 7 为 KNN 模型、ERT 模型以及混合模型对淹没面积、淹没水量和平均水深模拟的相对误差,忽略积水深度为 10 cm 以下的区域,在 15 场测试降雨事件中,模型对平均水深的预测最准确,对淹没水量的预测效果稍差。由图 6 可见,KNN 模型、ERT 模型以及混合模型模拟的淹没面积变化情况与水动力模型模拟结果相同,其中 KNN 模型误差较为稳定,平均相对误差为 0.39%,ERT 模型平均相对误差为 1.14%,混合模型平均相对误差为 0.69%,均达到了近乎完美的拟合效果。淹没面积与淹没水量的平均误差也可控制在 5% 以内,表明所构建的学习模型能较好反映山洪灾害整体情况,预报性能良好。

水动力模型需从 0 时刻开始迭代运算,模拟单场降雨 10 h 致洪过程平均用时为 1 688.4 s,KNN 模型对 15 场降雨数据进行最大洪量预测累计用时为 33 s,单场降雨平均用时为 2.2 s。ERT 模型速度略快,单场降雨平均用时为 1.7 s,两类模型均能在极短时间内给出预测结果。混合模型需要综合 ERT 模

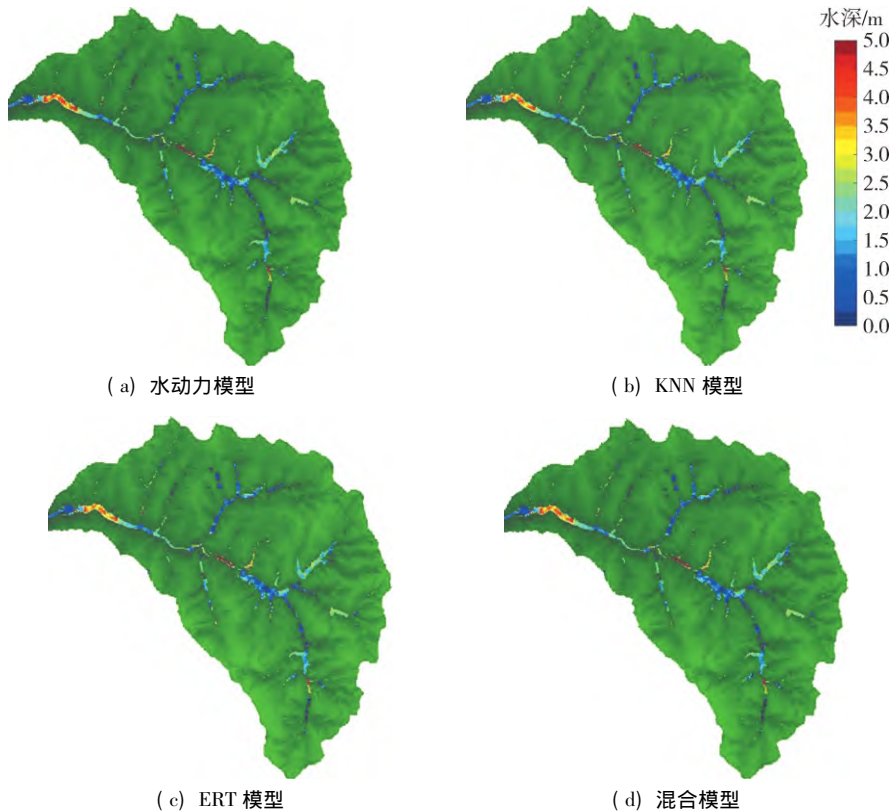


图 5 4 种模型的洪峰时刻预报结果对比

Fig. 5 Comparison of flood peak time prediction results of 4 models

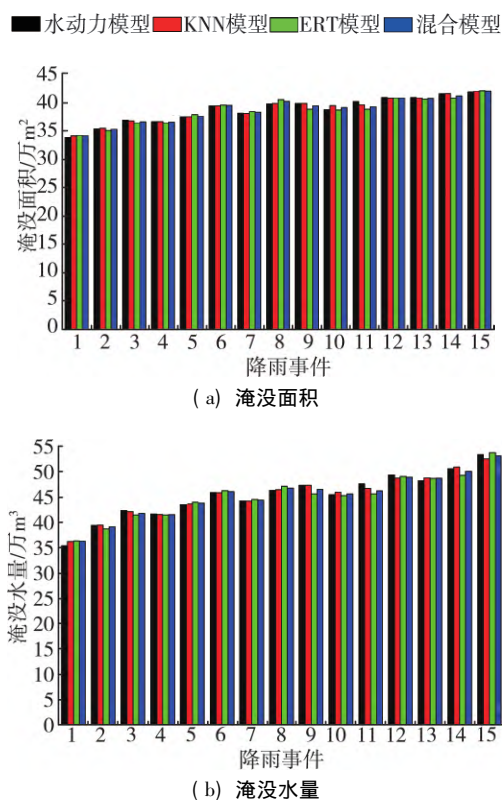


图6 不同方案下4种模型对淹没范围和淹没水量的模拟结果  
Fig. 6 Simulation results of inundation range and inundation water volume of 4 models under different schemes

型及 KNN 模型结果,因此模拟用时稍长,单场降雨平均用时约为 4.1 s,亦可满足快速山洪快速预报需求。

### 3.4 流域出口断面分析

因山区地形起伏较大,在将地形网格化进行洪水模拟计算时,地势高的区域降雨将迅速向低洼区域汇集,这会使地势高的区域出现大量积水量极小

表4 出口断面处平均水深、平均流速和断面流量模拟结果

Table 4 Simulation results of average water depth, average velocity and section flow at outlet section

降雨事件	平均水深/m				平均流速/(m·s <sup>-1</sup> )				断面流量/(m <sup>3</sup> ·s <sup>-1</sup> )			
	水动力模型	KNN模型	ERT模型	混合模型	水动力模型	KNN模型	ERT模型	混合模型	水动力模型	KNN模型	ERT模型	混合模型
1	0.964	0.992	0.965	0.978	0.345	0.359	0.337	0.348	6.652	7.118	6.502	6.801
2	1.346	1.348	1.287	1.317	0.947	0.943	0.883	0.913	12.740	12.707	11.362	12.029
3	1.598	1.628	1.745	1.682	0.722	0.913	0.867	0.890	23.064	22.302	22.698	22.460
4	1.590	1.549	1.563	1.555	0.878	0.856	0.867	0.862	20.926	19.899	20.326	20.101
5	1.712	1.703	1.755	1.728	0.778	0.774	0.792	0.783	26.625	26.365	27.789	27.044
6	1.965	1.962	2.058	2.008	0.839	0.837	0.852	0.845	32.959	32.868	35.084	33.933
7	1.761	1.756	1.692	1.724	0.794	0.792	0.806	0.799	27.963	27.825	27.267	27.556
8	1.942	1.977	1.865	1.921	0.836	0.840	0.834	0.837	32.472	33.213	31.120	32.163
9	2.319	2.317	2.247	2.284	0.880	0.880	0.836	0.858	40.833	40.767	37.567	39.186
10	1.937	1.859	1.779	1.819	0.832	0.819	0.828	0.824	32.215	30.473	29.446	29.964
11	2.267	2.137	1.961	2.050	0.877	0.858	0.836	0.847	39.737	36.687	32.771	34.736
12	2.293	2.284	2.271	2.276	0.881	0.878	0.874	0.876	40.398	40.129	39.702	39.890
13	2.232	2.487	1.998	2.244	0.874	0.895	0.898	0.897	39.001	44.512	35.875	40.226
14	2.486	2.832	2.699	2.767	0.898	0.917	0.906	0.912	44.630	51.962	48.922	50.456
15	3.092	2.878	3.072	2.976	0.541	0.582	0.497	0.533	66.962	58.611	68.680	63.488

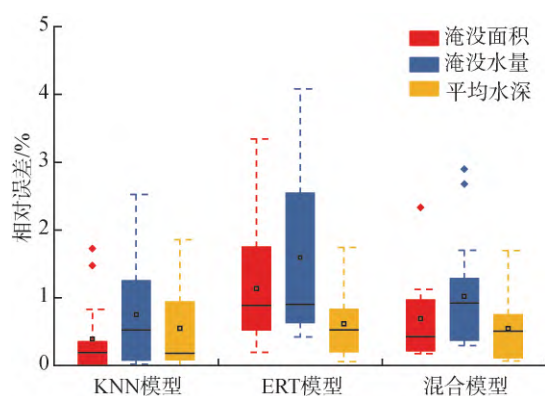


图7 3种模型模拟相对误差

Fig. 7 Simulation relative error of three models

的网格,因此预报模型在该区域将会达到近乎完美的预报效果,在运用传统指标进行模型可靠性评估时将会过高估计模型的实际预报效果。因此,本文以流域出口断面作为研究断面,通过分析出口断面的流量、流速及平均断面水深进一步验证模型预报性能。在15场降雨事件中,尽管模型在整体指标评估中均得到了较高的分数,但在实际的出口断面处,模型性能仍存在一定差异。4种模型在出口断面处平均水深、平均流速和断面流量模拟结果见表4。

图8为模型在出口断面的平均水深、平均流速和断面流量的相对误差。由图8(a)可见,3种模型对断面平均水深预测的平均误差均小于10%,其中KNN模型平均误差为3.51%,ERT模型平均误差为5.98%,KNN模型整体误差较小;两种模型在对个别场次降雨模拟时仍然存在较大误差,KNN模型最大误差为13.90%,ERT模型最大误差为13.51%。混合模型平均误差为3.39%,最大单场降雨误差为11.27%,有效综合了两种算法结果,降低对个别场



次降雨致洪信息的误报。由图 8(b) 可见 3 种模型在进行流速预测时误差均较小。其中, KNN 模型对断面流速预测性能最佳, 平均误差为 1.85%, 比 ERT 模型低 1.03%, 同时, 最大误差也最小, 为 6.00%。混合模型平均误差为 1.92%, 比 KNN 算法模型预报性能略差, 这是由于在对流速进行预测时, KNN 模型拟合效果普遍高于 ERT 模型, 在这种情况下混合模型通过综合两种模型模拟结果对误差的校正效果不明显。由图 8(c) 可见 3 个模型对于断面流速及水深的预测准确性要优于断面流量, 这是由于断面流量是通过构成断面网格上的水深、流速计算求得, 因此存在一定的误差累积现象, 导致模型对流量的预报性能稍差。但所建混合模型对断面流量最大误差可控制在 15% 以内, 平均误差为 4.50%, 低于 10%, 仍具有较高准确性, 可满足紧急决策需求。

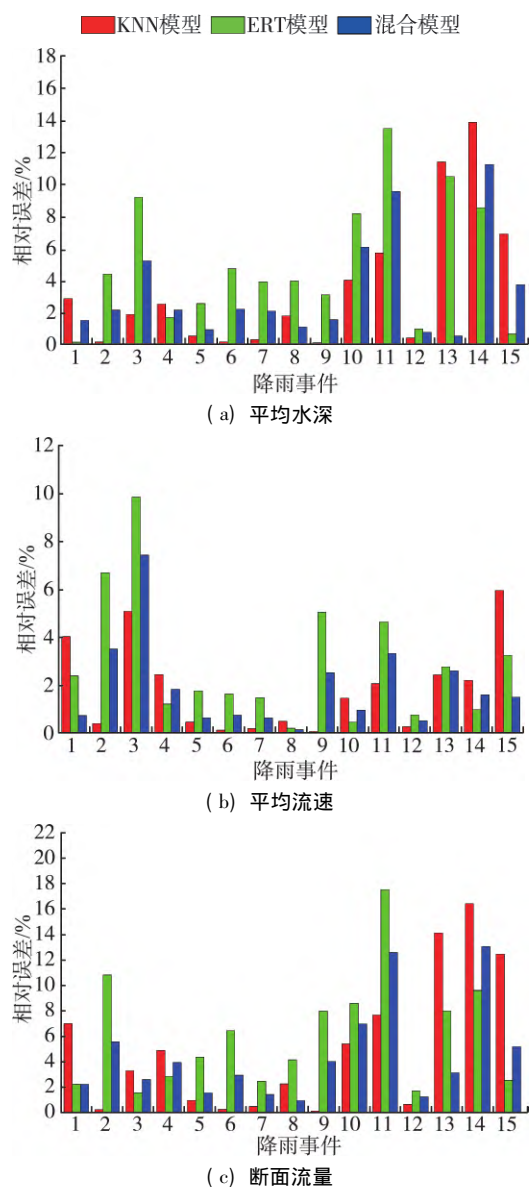


图 8 出口断面平均水深、平均流速和断面流量的相对误差  
Fig.8 Relative error of average water depth, average velocity and section flow at outlet section

## 4 结 论

a. 经评价指标分析, 本文选取的机器学习算法在进行山洪灾害预报时均能有较好性能, 进行水深模拟预测时, 各模型预报均能获得较好的得分, 其中 ERT 模型最佳; 进行流速预报时, KNN 模型预报效果较好, ERT 模型略差于 KNN 模型, 混合模型可有效综合预报结果, 可将整体误差控制在 5% 以内。

b. 根据流域出口断面特征信息, 算法对断面平均流速预测差距不大, KNN 算法表现最佳, 平均相对误差为 1.85%, 混合模型对误差降低不明显; 在进行断面水深及流量预测时, 算法预报效果差距较大, KNN 算法模拟平均水深及断面流量平均误差分别为 3.51% 和 5.10%, ERT 算法表现稍差, 平均误差均分别为 5.98% 和 6.07%, 混合模型对误差校正效果明显, 最终平均误差分别为 3.39% 和 4.50%, 可为应急决策提供可靠依据。

c. 所建模型单场降雨平均模拟时间可控制在 10 s 以内, 可为紧急决策争取大量前置时间, 协助决策者更好地采取应急管理措施。

## 参考文献:

- [1] 张驰, 张家华, 王浩. 基于网格流出修正法的山洪演进数值模拟[J]. 河海大学学报(自然科学版), 2014, 42(2): 107-113. (ZHANG Chi, ZHANG Jiahua, WANG Hao. Numerical simulation of flash flood routing based on grid outflow correction method[J]. Journal of Hohai University(Natural Sciences), 2014, 42(2): 107-113. (in Chinese))
- [2] 2017 年《中国生态环境状况公报》(摘录四)[J]. 环境保护, 2018, 46(14): 69-75. (China Environmental Status Bulletin 2017(Except 4)[J]. Environmental Protection, 2018, 46(14): 69-75. (in Chinese))
- [3] 2018 年《中国生态环境状况公报》(摘录一)[J]. 环境保护, 2019, 47(11): 47-53. (China Environmental Status Bulletin 2018(Except 1)[J]. Environmental Protection, 2019, 47(11): 47-53. (in Chinese))
- [4] 包红军, 李致家, 王莉莉, 等. 基于分布式水文模型的小流域山洪预报方法与应用[J]. 暴雨灾害, 2017, 36(2): 156-163. (BAO Hongjun, LI Zhijia, WANG Lili, et al. Flash flood forecasting method based on distributed hydrological models in a small basin and its application[J]. Torrential Rain and Disasters, 2017, 36(2): 156-163. (in Chinese))
- [5] 胡国华, 陈肖, 于泽兴, 等. 基于 HEC-HMS 的郴江流域山洪预报研究[J]. 自然灾害学报, 2017, 26(3): 147-155. (HU Guohua, CHEN Xiao, YU Zexing, et al. Research on forecast of mountain flood in Chenjiang River



- Basin based on HEC-HMS model [J]. Journal of Natural Disasters 2017 26(3): 147-155. ( in Chinese )
- [6] 孟天翔. 基于 MIKE FLOOD 的清原县海阳河小流域山洪数值模拟 [D]. 大连: 大连理工大学 2017.
- [7] SHEN Y ,ZHANG W ,CHAN B S M ,et al. Detecting risk of suicide attempts among Chinese medical college students using a machine learning algorithm [J]. Journal of Affective Disorders 2020 5: 273.
- [8] 许凡 陈朝阳 李昭 等. 基于机器学习的二甲醚层流燃烧速度预测研究 [J]. 西安交通大学学报, 2020, 54 ( 7 ): 62-67. ( XU Fan ,CHEN Zhaoyang ,LI Zhao ,et al. Predictions on laminar burning velocity of dimethyl ether/air mixtures using machine learning [J]. Journal of Xi'an Jiaotong University 2020 54( 7 ): 62-67. ( in Chinese) )
- [9] 张海山. 基于机器学习的诱导多能干细胞前体细胞识别 [D]. 北京: 中国科学院大学( 中国科学院深圳先进技术研究院) 2020.
- [10] 张幼成 李琼 徐汪洋 等. 基于导向标准断面法的受冲淤影响洪峰水位预报模型 [J]. 水资源保护, 2019, 35 ( 3 ): 53-56. ( ZHANG Youcheng , LI Qiong , XU Wangyang ,et al. Forecasting model of flood peak water level affected by scour and silt based on guide standard section method [J]. Water Resources Protection 2019 35 ( 3 ): 53-56. ( in Chinese) )
- [11] 俞彦 张行南 张鹏 等. 基于 SCS 模型和新安江模型的雨量预警指标综合动态阈值对比 [J]. 水资源保护, 2020 36(3): 28-33. ( YU Yan ZHANG Xingnan ZHANG Peng , et al. Comparison of comprehensive dynamic threshold of rainfall warning indicators based on SCS model and Xinanjiang model [J]. Water Resources Protection 2020 36( 3 ): 28-33. ( in Chinese) )
- [12] 王胜凯. 移动应用在山洪灾害防治中的应用 [J]. 水资源保护, 2017, 33( 增刊 1 ): 48-50. ( WANG Shengkai. Application of mobile application in the prevention and control of mountain torrent disasters [J]. Water Resources Protection 2017 33( Sup1 ): 48-50. ( in Chinses) )
- [13] 阚光远 洪阳 梁珂. 基于耦合机器学习模型的洪水预报研究 [J]. 中国农村水利水电, 2018( 10 ): 165-169. ( KAN Guangyuan ,HONG Yang ,LIANG Ke. Research on the flood forecasting based on coupled machine learning model [J]. China Rural Water and Hydropower, 2018 ( 10 ): 165-169. ( in Chinese) )
- [14] 张珂 牛杰帆 李曦 等. 洪水预报智能模型在中国半干旱半湿润区的应用对比 [J]. 水资源保护, 2021, 37 ( 1 ): 28-35 ( ZHANG Ke ,NIU Jiefan ,LI Xi , et al. Comparison of artificial intelligence flood forecasting models in China's semi-arid and semi-humid regions [J]. Water Resources Protection 2021 37 ( 1 ): 28-35. ( in Chinses) )
- [15] 张轩 张行南 江唯佳 等. 秦淮河流域东站点水位预报研究 [J]. 水资源保护, 2020, 36( 2 ): 41-46 ( ZHANG Xuan ,ZHANG Xingnan ,JIANG Weijia ,et al. Study on water level forecast of Dongshan Station in Qinhuai River Basin [J]. Water Resources Protection 2020 36( 2 ): 41-46. ( in Chinese) )
- [16] CHANG L C ,AMIN M ,YANG S N ,et al. Building ANN-based regional multi-step-ahead flood inundation forecast models [J]. Water 2018 10( 9 ): 1283.
- [17] FAUZI A ,MIZUTANI N. Machine learning algorithms for real-time tsunami inundation forecasting: a case study in Nankai Region [J]. Springer International Publishing , 2020 177( 3 ): 02364.
- [18] 张幸幸 陈祖煜. 小流域淤地坝系的溃决洪水分析 [J]. 岩土工程学报, 2019, 41 ( 10 ): 1845-1853. ( ZHANG Xingxing ,CHEN Zuyu. Breach flood analysis of warping dam system in small watersheds [J]. Chinese Journal of Geotechnical Engineering 2019 41( 10 ): 1845-1853. ( in Chinese) )
- [19] 高海东 李占斌 李鹏 等. 黄土高原暴雨产沙路径及防控: 基于无定河流域 2017-07-26 暴雨认识 [J]. 中国水土保持科学, 2018, 16( 4 ): 66-72. ( GAO Haidong ,LI Zanbin ,LI Peng ,et al. Paths and prevention of sediment during storm-runoff on the Loess Plateau: based on the rainstorm of 2017-07-26 in Wuding River [J]. Science of Soil and Water Conservation 2018 16( 4 ): 66-72. ( in Chinese) )
- [20] ENGMAN E. Roughness coefficients for routing surface runoff [J]. Journal of Irrigation and Drainage Engineering , 1986 112: 39-53.
- [21] ZHONG L I ,CHAO S J ,DONG J Q. Study on ecological environment influence of different age of hippophae in hilly and gully region on the Loess Plateau [J]. Agricultural Research in the Arid Areas ,2008 26( 5 ): 115-120.
- [22] 张茹 雍斌 曾岁康. GPM 卫星降水产品在中国大陆的精度评估 [J]. 人民长江, 2021 52( 5 ): 50-59. ( ZHANG Ru ,YONG Bin ,ZENG Suikang. Evaluation of GPM satellite precipitation products over mainland China [J]. Yangtze River 2021 52( 5 ): 50-59. ( in Chinese) )
- [23] 戴有学 王振华 戴临栋 等. 芝加哥雨型法在短历时暴雨雨型设计中的应用 [J]. 干旱气象, 2017, 35( 6 ): 1061-1069. ( DAI Youxue , WANG Zhenhua , DAI Lindong ,et al. Application of Chicago Hyetograph Method in design of short duration rainstorm pattern [J]. Journal of Arid Meteorology , 2017 , 35 ( 6 ): 1061-1069. ( in Chinese) )
- [24] 薛宇雷. 基于 SWMM 模型对周河流域地表径流和非点源污染的研究 [D]. 西安: 长安大学 2018.
- [25] 蒋卫威 鱼京善 赤穗良辅 等. 基于水文水动力耦合模型的山区小流域洪水预报 [J]. 水文, 2020 40( 5 ): 28-35. ( JIANG Weiwei ,YU Jingshan ,AKOH Ryosuke ,et al.

( 下转第 111 页 )

- modeling[J]. Resources Science 2018 40( 3) : 477-485. ( in Chinese )
- [16] 杜娟娟. 基于不同赋权方法的模糊综合水质评价研究 [J]. 人民黄河 ,2015 ,37( 12) : 69-73. ( DU Juanjuan. Research of fuzzy comprehensive water quality assessment model based on different methods [J]. Yellow River , 2015 37( 12) : 69-73. ( in Chinese ) )
- [17] 傅聪颖 ,赖昭豪 ,郭熙. 基于熵权 TOPSIS 模型的区域资源环境承载力评价及障碍因素诊断 [J]. 生态经济 , 2020 36( 1) : 198-204. ( FU Congying ,LAI Zhaohao ,GUO Xi. Evaluation of regional resource and environment carrying capacity based on the entropy-weight TOPSIS model and the obstacle factors diagnosis [J]. Ecological Economy 2020 36( 1) : 198-204. ( in Chinese ) )
- [18] LI Tianhong ,YANG Songnan ,TAN Mingxin. Simulation and optimization of water supply and demand balance in Shenzhen: a system dynamics approach [J]. Journal of Cleaner Production 2019 207: 882-893.
- [19] 张斌. 基于 SD 模型的深圳市水资源承载力研究 [J]. 水利水电技术 2010 41( 12) : 89-93. ( ZHANG Bin. SD model based study on water resources carrying capacity of Shenzhen [J]. Water Resources and Hydropower Engineering 2010 41( 12) : 89-93. ( in Chinese ) )
- [20] 范嘉伟 ,黄锦林 ,袁明道 ,等. 灰色关联-熵模型在珠三角水资源承载力评估中的应用 [J]. 中国农村水利水电 2019( 7) : 35-39. ( FAN Jiawei ,HUANG Jinlin ,YUAN Mingdao ,et al. The application of grey correlation-entropy model in water resources carrying capacity evaluation to Pearl River Delta [J]. China Rural Water and Hydropower 2019( 7) : 35-39. ( in Chinese ) )
- ( 收稿日期: 2021-01-08 编辑: 施业)

#### ( 上接第 40 页)

- Flood forecasting of small watershed in mountainous area based on hydrological-hydrodynamic coupling model [J]. Journal of China Hydrology ,2020 ,40 ( 5) : 28-35. ( in Chinese ) )
- [26] HUBBARD M E J J O C P. Multidimensional slope limiters for MUSCL-Type finite volume schemes on unstructured grids [J]. Journal of Computational Physics , 2012 155( 1) : 54-74.
- [27] LAI T ,YI T H ,LI H N ,et al. An explicit fourth-order Runge-Kutta method for dynamic force identification [J]. International Journal of Structural Stability and Dynamics , 2017 17( 10) : 1750120.
- [28] HOU J ,LIANG Q ,ZHANG H ,et al. An efficient unstructured MUSCL scheme for solving the 2D shallow water equations [J]. Environmental Modelling Software , 2015 66( C) : 131-152.
- [29] SIVAKUMAR P ,HYAMS D G ,TAYLOR L K ,et al. A primitive-variable Riemann method for solution of the shallow water equations with wetting and drying [J]. Journal of Computational Physics , 2009 , 228 ( 19) : 7452-7472.
- [30] HOU J ,LIANG Q ,SIMONS F ,et al. A stable 2D unstructured shallow flow model for simulations of wetting and drying over rough terrains [J]. Computers & Fluid , 2013 82( 17) : 132-147.
- [31] 侯精明 ,李桂伊 ,李国栋 等. 高效高精度水动力模型在洪水演进中的应用研究 [J]. 水力发电学报 2018 37 ( 2) : 96-107. ( HOU Jingming ,LI Guiyi ,LI Guodong ,et al. Application of efficient high-resolution hydrodynamic model to simulations of flood propagation [J]. Journal of Hydroelectric Engineering ,2018 ,37 ( 2) : 96-107. ( in Chinese ) )
- [32] 侯精明 ,王润 ,李国栋 等. 基于动力波法的高效高分辨率城市雨洪过程数值模型 [J]. 水力发电学报 2018 37 ( 3) : 40-49. ( HOU Jingming ,WANG Run ,LI Guodong ,et al. High-performance numerical model for high-resolution urban rainfall-runoff process based on dynamic wave method [J]. Journal of Hydroelectric Engineering ,2018 , 37( 3) : 40-49. ( in Chinese ) )
- [33] GEURTS P ,ERNST D ,WEHENKEL L. Extremely randomized trees [J]. Machine Learning ,2006 ,63 ( 1) : 3-42.
- [34] JOHN V ,ZHENG L ,GUO C ,et al. Real-time lane estimation using deep features and extra trees regression [C]//Pacific Rim Symposium on Image and Video Technology. Cham: Springer 2015.
- [35] YU Z ,LUO Y ,DENG X ,et al. Satellite-based estimates of daily NO<sub>2</sub> exposure in China using hybrid random forest and spatiotemporal kriging model [J]. Environmental Science & Technology 2018 52( 7) : 4180.
- [36] KELLER J M ,GRAY M R ,GIVENS J A. A fuzzy K-nearest neighbor algorithm [J]. IEEE Transactions on Systems Man and Cybernetics 1985 15( 4) : 12-20.
- [37] HU X ,SHI L ,LIN L ,et al. Improving surface roughness lengths estimation using machine learning algorithms [J]. Agricultural and Forest Meteorology 2020 287: 107956.
- ( 收稿日期: 2020-12-15 编辑: 王芳)