

基于聚类及决策树的长江中上游流域洪水类型动态识别研究

丁 伟¹,赵美晨²,王浅宁¹,李 敏¹,张金男¹,周惠成¹,彭 勇¹

(1. 大连理工大学建设工程学院; 2. 淮河水利委员会水文局(信息中心))

摘要:为应对长江中上游流域洪水类型复杂且遭遇规律多变给流域防洪调度带来的挑战,构建了耦合 K-Means++ 聚类算法与决策树算法的洪水类型动态识别模型,对长江中上游流域的三峡入库洪水类型进行了识别。结果表明:依据次洪总量占比和次降雨总量占比 2 个指标分别将长江中上游流域洪水划分为岷沱嘉陵江遭遇型、岷沱金沙江及向寸区间遭遇型、岷沱嘉陵江及向寸与寸三区间遭遇型、乌江及向寸与寸三区间遭遇型 4 种类型,分类一致性高达 88.4%,各类型洪水特征显著;依据降雨信息动态识别洪水类型的整体精度为 83.3%,不同类型洪水的精确度、召回率、F1 分数 3 个评价指标的加权平均值均高于 0.70,模型对多数洪水类型的识别性能良好。

关键词:K-Means++ 聚类算法;决策树算法;洪水类型识别;长江中上游流域

Research on dynamic identification of flood types in middle and upper reaches of the Yangtze River Basin based on clustering and decision trees//Ding Wei¹, Zhao Meichen², Wang Qianning¹, Li Min¹, Zhang Jinnan¹, Zhou Huicheng¹, Peng Yong¹(1. School of Infrastructure Engineering, Dalian University of Technology; 2. Hydrological Bureau (Information Center) of Huaihe River Commission)

Abstract: To address the challenges posed by the complex and varied flood types in the upper and middle reaches of the Yangtze River Basin, a dynamic identification model for flood types was constructed by coupling K-Means++ clustering algorithm and decision tree algorithm. The model was used to identify the types of Three Gorges inflow floods in the upper and middle reaches of the Yangtze River Basin. The results indicate that based on two indicators of the proportion of total secondary floods and the proportion of total secondary net rainfall floods, floods in the upper and middle reaches of the Yangtze River can be classified into four types, including the Min River, the Tuo River, and the Jialing River encounter type, the Min River, the Tuo River, the Jialing River, and the Xiangcun interval encounter type, the Min River, the Tuo River, the Jialing River and Xiongchun, Chunsan intervals encounter type, and the Wujiang River and Xiongchun, Chunsan intervals encounter type. The classification consistency is as high as 88.4%, and the characteristics of various types of floods are significant. The overall accuracy of dynamically identifying flood types based on rainfall information is 83.3%. The weighted average values of the accuracy, recall rate, and F1 score for different types of floods are all above 0.70, indicating that the model has good recognition performance for most flood types.

Key words: K-Means++ clustering algorithm; decision tree algorithm; flood type identification; upper and middle reaches of the Yangtze River Basin

长江支流多,流域面积大,洪水频发且遭遇类型复杂等问题使中下游,尤其是荆江河段,成为洪灾最频繁的地区。目前长江流域水工程联合调度体系的水工程总数已扩大至 135 座,水工程联合调度提高了长江流域防洪兴利水平,对于削峰拦洪、调节上下游洪水作用至关重要。但联合调度范围的扩大带来一个调度难题,不同雨洪区水库的任务需求不一致,导致水库群联合调度难度增加,影响水库防控措施

的部署。根据一定原则进行洪水分类,在实际调度中提前判断未来可能面临的洪水类型,可以帮助决策者更有针对性地进行水库调度决策。依据洪水影响范围不同,《长江流域综合规划(2012—2030 年)》将长江成灾大洪水归为区域性洪水和全流域性洪水两种类型^[1]。若提前识别区域性洪水的时空分布,主雨区的水工程联合调度抵御洪水,非主雨区的水工程在有余力的情况下进行洪水资源利

基金项目:国家自然科学基金项目(U2240204)
作者简介:丁伟(1987—)女,副教授,博士,主要从事流域水资源管理研究。E-mail:weiding@dlut.edu.cn

用,既可保证主雨区防洪安全,又可提高非主雨区的洪水资源利用,可充分发挥水工程联合调度优势,因此亟须开展长江流域洪水类型识别研究。

洪水分类是洪水类型识别的前置条件,其主要依据历史洪水数据,按照一定的原则和标准,将具有相同规律的洪水过程归为同一类型。关于洪水分类的研究起步较早,如 Merz 等^[2]使用不同来源的信息组合,手动进行了洪水分类。洪水分类研究主要有两类:一类是对洪水类型的描述,从不同角度出发描述多种洪水形式,如依据洪水事件成因相似性描述为暴雨洪水、冰雪融水洪水、河流洪水等;另一类是研究方法的探索,主要有成因分析法、智能优化算法、聚类分析法等^[3]。早期研究以成因综合分析为主,通过相似度计算将洪水归为不同类型,如马寅午等^[4]用相关系数表示相似度,演绎归纳各类相似洪水;张艳平等^[5]通过计算相似系数,实现了大伙房水库洪水分类。随着计算能力的提高,混合蛙跳算法、人工神经网络算法^[6]等机器学习算法常被用来提高分类精度。随着水文数据的丰富和可用性的提高,聚类算法在洪水分类领域中广泛使用,如 Dinesh^[7]通过模糊聚类对多个流域的历史洪水过程进行分类,证明了方法的良好适用性;Turkington 等^[8]利用 *K*-Means 聚类法对阿尔卑斯山两个不同的区域进行洪水分类,并评估了未来气候情景下的洪水类型变化。随着研究不断深入,洪水分类研究趋于综合方法运用,如 Zou 等^[9]提出利用基于 IGWO 算法优化的组合权重模糊迭代模型进行洪水分类;甘甜等^[10]结合密度聚类与随机森林算法,显著提高了徒骇河流域洪水分类效果。现有方法多聚焦中小流域,对特大流域多指标特点响应不足,缺乏对特大流域洪水多来源、多遭遇类型的系统分类。相较于传统方法,*K*-Means++ 聚类算法可通过概率性选择初始聚类中心,使聚类中心间距更大,有效避免局部最优解,且对多指标的分类适应性更强,能提升特大流域洪水类型划分的稳定性和准确性。

准确的洪水类型识别是洪水分析的基础和关键步骤,洪水类型识别主要指通过相关降雨和洪水信息,结合已获取的长期历史数据,检索与当前洪水最相似的历史洪水过程或类型。目前,长江流域洪水类型识别研究主要集中于历史大洪水的时空变化特征与遭遇规律分析,多数学者在识别过程中以洪水分类结果作为参考,筛选最相似洪水后进行预报和模型优化。例如: Huang 等^[11]基于 1998 年全流域大洪水和 2020 年区域性大洪水资料,分析了洪水过程的形态、洪峰、洪水总量等洪水时空演变规律;丁启等^[12]对长江上游进行了水文相似区识别;赵圆元

等^[13]对嘉陵江流域进行了相似洪水分析,以相似特征识别结果为洪水下一阶段特征或水文模型参数提供了更精确地预测方向。但此类方式多为事后分析,洪水类型识别研究多集中于使用洪水发生后的个别特征信息进行识别^[14],对于完整识别洪水事件类型、在洪水发展过程中实时识别洪水类型及如何提前预测洪水类型研究相对较少,难以满足实时调度需求。

鉴于此,本文采用 *K*-Means++ 聚类算法对三峡水库入库洪水过程进行分类,进而利用决策树算法建立基于降雨信息的洪水类型识别模式,开展长江中上游流域洪水类型动态识别,以为各流域防洪调度和水能资源的高效利用提供决策依据。

1 研究区概况与数据来源

1.1 研究区概况

长江中上游流域位于我国西南地区,涵盖从源头至江西、湖南交界的区域,全长逾 4 500 km,集水面积约 100 万 km²,占长江流域总面积的 60%^[15],本文以该区域为研究区。研究区地理特征复杂多样,包括 6 个子流域(岷江、嘉陵江、雅砻江、沱江、渠江、涪江)和 8 个水文区间(向寸区间(向家坝至寸滩)、寸三区间(寸滩至三峡)、金沙江上游、金沙江中游、金沙江下游、乌江上游、乌江中游、乌江下游),洪水组成及其遭遇规律错综多变,尤其是每年 6—8 月雨季及台风季节洪水频发。研究区流域出口控制站为宜昌站,位于此的三峡水库是世界著名水利工程。研究区概况如图 1 所示。

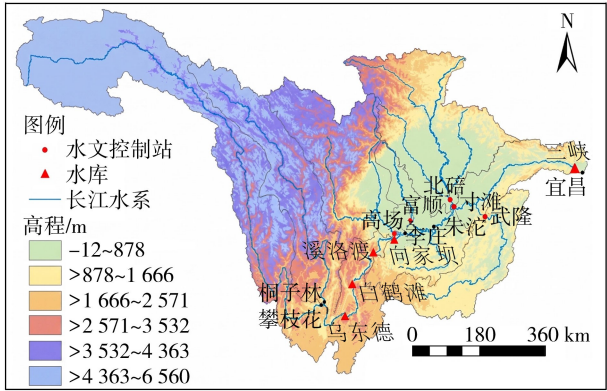


图 1 研究区概况

Fig. 1 Overview of study area

1.2 数据来源

本文涉及的历史降雨与洪水资料均源自水利部长江水利委员会,所用的预报降雨数据源自 TIGGE 网站 (<https://www.ecmwf.int/>) 获取的 ECMWF 集合预报数据。研究中历史数据使用 2014—2021 年及 1954 年、1981 年、1982 年、1998 年汛期(6—9 月)

共 43 场降雨和洪水信息,预报降雨数据使用 2014—2021 年预见期为 7 d 的面平均降水量,构建场次洪水分类及识别模式,实现洪水类型实时动态识别。

本文所选的 43 场洪水均为三峡入库洪水及各子流域对应的流量过程。43 场洪水采用“流量超阈值法+持续时间法”,以宜昌站日流量超过 4 万 m^3/s (超过 5 万 m^3/s 的为编号洪水)为阈值且持续时间大于或等于 3 d 作为场次洪水识别标准,并通过退水曲线分割来消除上一场洪水退水影响。洪水发生前 7~15 天内流域面平均日降水量连续 3 天小于 2 mm 视为降雨开始,直至连续 3 天面平均日降水量小于 2 mm 视为降雨结束,则该降雨过程为场次洪水对应的降雨过程。

2 研究方法

2.1 洪水分类指标选取

研究区幅员辽阔,子流域及水文区间众多,暴雨洪水时空变化导致流域内洪水遭遇组合复杂、洪水类型丰富。不同场次洪水中,干支流与区间洪水遭遇时,主来水区不同会影响三峡的入库洪水。同一出口断面的洪水地区组成格局相似,依据干支流洪水组成实施分类,既能规避因分类指标选取不同引发的结果差异,又利于区分主来水区。同时,洪水规模与暴雨降水量之间关联紧密,通过对暴雨降水量的深入剖析认知洪水类型,有助于显著延长洪水预见期,实现对洪水类型的提前识别。根据洪水组成决定类型、降雨组成决定识别的逻辑,选取洪水分类、降雨分类、洪水特征描述指标,进行典型洪水分类,进而开展洪水类型动态识别。

a. 洪水分类指标。选取次洪总量占比作为洪水分类指标,其反映场次洪水各子流域洪水总量在三峡入库洪水总量中的占比,能够精准刻画不同子流域洪水对三峡入库洪水的贡献程度,计算公式为:

$$P_{w,ij} = W_{ij}/W_j \quad (i = 1, 2, \dots, 6; j = 1, 2, \dots, 43) \quad (1)$$

式中: $P_{w,ij}$ 为第 i 个子流域在第 j 场洪水的次洪总量占比; W_{ij} 为第 i 个子流域在第 j 场洪水中汇入控制断面的洪量; W_j 为第 j 场洪水三峡入库总洪量。

b. 降雨分类指标。选取次降雨总量占比作为降雨分类指标,其反映降雨空间分布特征以及致洪暴雨中某区域相对于全流域贡献水量的大小,能够剔除非致洪水分损耗,更能反映降雨与洪水的成因关联,结合分区面积加权后,可体现降雨空间分布对洪水类型的影响,计算公式为:

$$P_{R,kj} = R_{kj}A_k / \sum_{k=1}^8 R_{kj}A_k \quad (k = 1, 2, \dots, 8) \quad (2)$$

式中: $P_{R,kj}$ 为第 k 个水文分区在第 j 场洪水的次降雨总量占比; R_{kj} 为第 k 个水文分区在第 j 场洪水的总净雨深; A_k 为第 k 个水文分区的面积。

c. 洪水特征描述指标。选取洪峰时间偏度作为形态指标,涨洪速率、落洪速率作为速率指标,高脉冲历时占比作为强度指标分析不同类型洪水的特征,计算公式为:

$$T_S = t_R/T_R \quad (3)$$

$$v_R = (Q_m - Q_0)/(T_m - T_0) \quad (4)$$

$$v_D = (Q_m - Q_{\text{end}})/(T_{\text{end}} - T_m) \quad (5)$$

$$P_H = T_H/T_R \quad (6)$$

式中: T_S 为洪峰时间偏度,取值越小,峰现时间越靠前; t_R 为涨洪历时; T_R 为洪水总历时; v_R 为涨洪速率,取值越大表明洪水起涨速度越快,易在短时间内形成灾害性洪水; Q_m 、 Q_0 分别为场次洪水洪峰流量、起涨流量; T_0 为洪水起涨时刻; T_m 为洪峰出现时刻; v_D 为落洪速率; Q_{end} 为场次洪水终止流量; T_{end} 为洪水终止时刻; P_H 为高脉冲历时占比,值越大洪峰形态越宽; T_H 为洪水流量大于 75% 洪峰流量值的持续时间。

2.2 洪水类型动态识别方法

2.2.1 基于 K-Means++ 聚类算法的洪水分类

K-Means++ 聚类算法通过一种特定的概率性方法选择初始聚类中心,使选择的中心点之间距离尽可能远,显著改善了经典 K-Means 聚类算法的性能,从而达到更好的聚类效果^[16]。应用其进行分类时,采用肘法估计最佳聚类数的基本思想是:聚类分析中,随着分类数量增加,数据划分更精细,各簇内紧密程度提高,误差平方和逐渐减小。误差平方和的计算公式为:

$$S_{SE} = \sum_{i=1}^n \sum_{x \in C_i} |x - C_i|^2 \quad (7)$$

式中: S_{SE} 为误差平方和; C_i 为第 i 个簇; x 为属于 C_i 的数据点; n 为簇的总数量。

基于 K-Means++ 聚类算法进行场次洪水分类的具体步骤为:

步骤 1 初始化。对输入的 43 场洪水数据集中随机选择 1 个数据作为第一个聚类中心。

步骤 2 确定聚类中心。对数据集中的每一个数据点(每一场洪水) x ,计算其与已有聚类中心中最近中心的距离 $D(x)$,并以概率的方式选择新的聚类中心。概率计算公式为:

$$P(c_i = x) = D(x)^2 / \sum_x D(x')^2 \quad (8)$$

式中: c_i 为第 i 个新的聚类中心; x' 为新的聚类中心数据点; $\sum_{x'} D(x')^2$ 为所有数据点到其最近聚类中心距离的平方和。

步骤 3 重复选择。重复步骤 2,直到选择出所有的 K 个聚类中心。

步骤 4 聚类分配。为每个数据点分配到最近聚类中心,形成 K 簇。

步骤 5 更新聚类中心。计算每个簇的平均点,作为新的聚类中心。

步骤 6 迭代优化。重复步骤 4、5,直到聚类中心不变化或达到预设的迭代次数,算法结束。

2.2.2 基于决策树算法的洪水类型动态识别

决策树算法是常用的分类和回归方法,其将数据分成多个子集,识别数据间的潜在结构和模式关系,并以树形结构展示决策过程^[17-18]。结构每个内部节点和分支分别表示一个特征属性和该特征的取值,每个叶子节点是最终决策结果,即洪水类型。

决策树算法通过选择最优特征进行数据分割,常用信息熵、信息增益率、基尼指数等指标评估不同属性对分类贡献大小。本文采用基尼指数评价贡献度,其取值为 0~1,代表在样本集中随机抽取两个样本属于不同类型的概率,数值越小,样本集中数据集中程度越高,计算公式为:

$$G = 1 - \sum_{i=1}^N p_i^2 \tag{9}$$

式中: G 为基尼指数; p_i 为数据集属于第 i 个类别的概率; N 是数据集类别数。

基于决策树算法构建洪水类型识别模型的具体步骤为:

步骤 1 数据集与特征选择。使用 43 场洪水分类结果及其对应的降雨分类指标、洪水分类指标构建数据集。

步骤 2 数据划分。为确保模型训练的公平性与代表性,按 7 : 3 比例划分训练集和测试集,对所有场次洪水数据按类型进行分层抽样,每个类型随机选取 3 个样本(共 12 场)为测试集,剩余样本为训练集。

步骤 3 模型建立与训练。使用决策树分类器为基本模型,为避免规则冗杂导致识别过程繁复及过拟合情况,限制决策最大深度不超过 6 层,在训练集上进行拟合,利用训练数据特征和目标类型建立动态识别模型。

步骤 4 特征排序与模型优化。模型训练过程中测试不同特征组合对模型性能的影响,对分类特征进行排列组合及分析,区分其重要度顺序。每次使用不同特征排列顺序重新训练决策树,并计算相

应准确率,通过对比不同特征排列下模型表现,选定表现最优特征排列为最终模型输入。

步骤 5 识别结果评估与验证。利用精确度、召回率、F1 分数 3 个指标对不同类型洪水的识别结果进行评价,其中,精确度指模型判定为某一类型的洪水中,实际属于该类型的比例;召回率指实际属于某一类型洪水中,模型能够正确识别的比例; $F1$ 分数为精确度和召回率的调和平均值。准确率指模型正确识别样本数量占总样本数量的比例,反映了模型总体分类精度,计算准确率以评价模型整体性能,取准确率达 70% 以上为规则可用。评价指标的计算公式为:

$$P_r = T_p / (T_p + F_p) \tag{10}$$

$$R_c = T_p / (T_p + F_N) \tag{11}$$

$$F_1 = 2P_r R_c / (P_r + R_c) \tag{12}$$

$$A_c = N_c / N_T \tag{13}$$

式中: P_r 为精确度; T_p 为真实类别为正类且识别为正的样本数; F_p 真实类别为负类但识别为正的样本数; R_c 为召回率; F_N 为真实类型为正,但识别为负的样本数; F_1 为 F1 分数; A_c 为准确率; N_c 为正确识别样本数; N_T 为总样本数。

在洪水分类阶段,本文侧重于挖掘历史场次洪水的空间来源特性。选取 7 个核心子流域、水文区间的次洪总量占比作为聚类输入特征,包括金沙江、岷江、沱江、嘉陵江、乌江、向寸区间、寸三区间。次洪总量占比有效排除了绝对洪量大小的干扰,能够更纯粹地反映洪水的空间组合结构。通过将历史洪水样本的 7 维特征向量输入 K -Means++ 聚类模型进行无监督划分,对三峡入库洪水过程进行分类。在洪水分类的基础上,利用基于决策树的洪水类型动态识别模型,以上述 4 类洪水为目标变量,以研究区 14 个子流域、水文区间的次降雨总量占比作为输入特征进行监督训练,动态识别各场次洪水类型。

3 结果与分析

3.1 基于洪水指标的洪水分类

基于 K -Means++ 聚类算法的洪水分类的结果表明,聚类数为 2~9 时,误差平方和分别为 0.83、0.65、0.56、0.49、0.42、0.38、0.35、0.30,其变化率分别为 0.43、0.21、0.13、0.12、0.13、0.11、0.08、0.13。可见,误差平方和变化率开始处于较高状态,随着聚类数增加,逐渐减小,数据点聚类紧密度提高。当聚类数为 8 时,误差平方和变化率降到 0.08,随后回升至 0.13。误差平方和变化率从聚类数 2 到 3 时变化较大,从 3 到 4 后趋于平稳,说明进一步增加聚类数对改善聚类效果的影响有限。因此

选择最佳聚类数为 4,既能保证不同类型洪水场次足够,又能避免过度细分。因此,将三峡入库洪水过程分为:岷沱嘉陵江遭遇型(类型 I)、岷沱金沙江及向寸区间遭遇型(类型 II)、岷沱嘉陵江及向寸与寸三区间遭遇型(类型 III)、乌江及向寸与寸三区间遭遇型(类型 IV) 4 种类型。

图 2 为不同洪水类型的次洪总量占比及其与

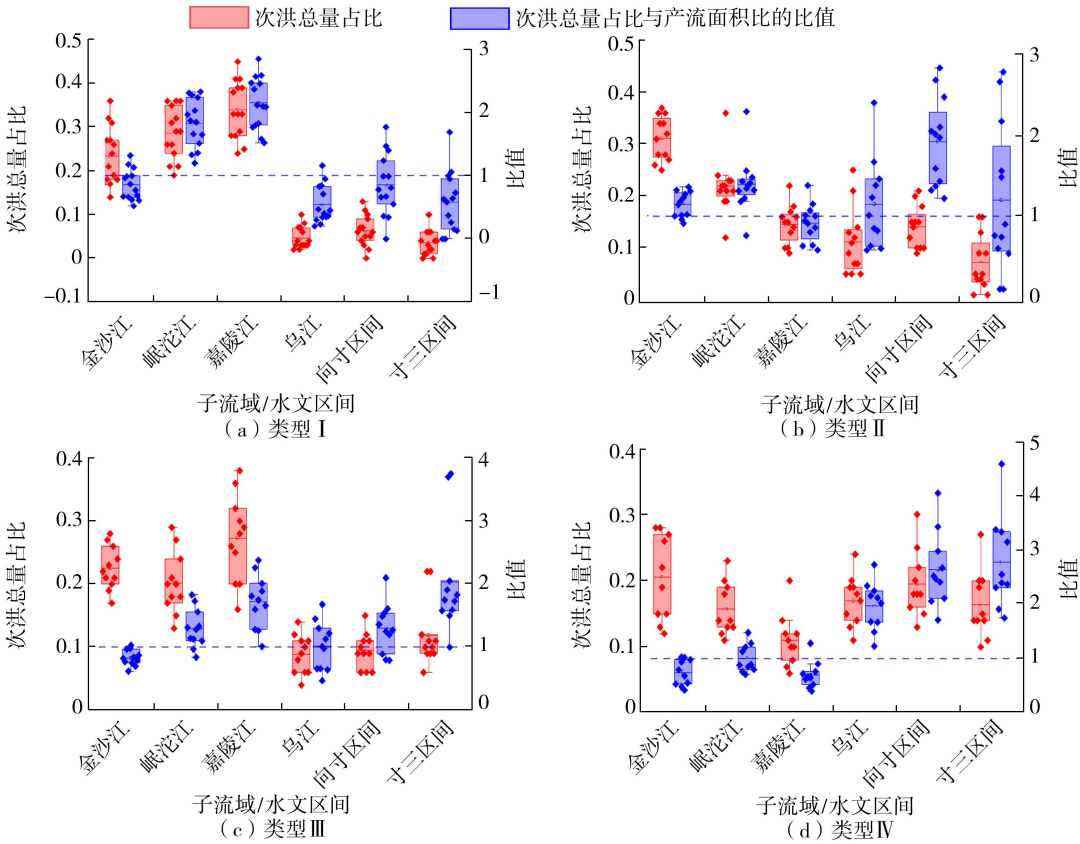


图 2 不同洪水类型的次洪总量占比及其与产流面积比的比值

Fig.2 Proportion of secondary floods for different flood types and their ratio to runoff-generating area ratio

图 3 为核心子流域、水文区间不同洪水类型的次洪总量。由图 3 可见,不同洪水类型的次洪总量分布与其占比分布特征一致,表现为:发生遭遇时主来水区洪量显著偏高,除乌江、向寸、寸三区间发生

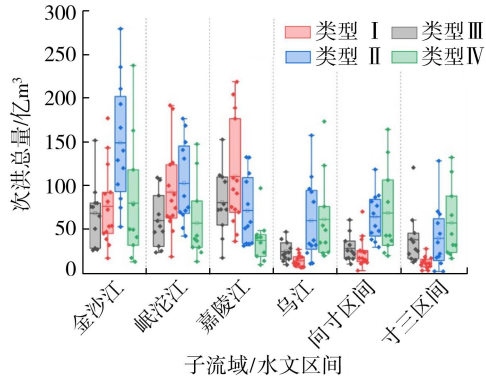


图 3 核心子流域、水文区间不同洪水类型的次洪总量
Fig.3 Total amount of secondary floods of different flood types in core sub-basins and hydrological intervals

产流面积比的比值,其中岷江与沱江合并分析,记为岷沱江。不同洪水分类中,遭遇来水区域的次洪总量占比突出,其与产流面积比的比值远超 1.0,显著偏高。其中,类型 I 与类型 III 具有相关性,当区间来水增加,岷沱江和嘉陵江的次洪总量与产流面积比的比值削弱至与区间洪水同一数值范围时,洪水从类型 I 转为类型 III。

与岷沱嘉陵江遭遇洪水时洪量变化范围较小外,其他子流域不同洪水类型发生时,洪量整体变化幅度均较大。

洪水历时为 10 d 是水利、防洪、灾害评估的常用临界标准,从洪水历时是否超过 10 d 的角度对不同洪水类型的特征进行分析,特征指标值均值见表 1,特征描述见表 2。由表 1、表 2 可见,类型 I 洪水以岷沱江和嘉陵江为主来水区,洪水历时短,表现为洪峰流量高、涨落洪速率快、洪峰时间偏度靠前的特点,形成了陡涨陡落、高峰尖锐的典型洪水过程,易对中下游地区造成突出风险;类型 II 洪水洪峰流量高且出现时间早,但洪水历时长,整体涨落洪速度缓慢,具体表现为缓涨缓落的平稳大洪水过程,对中下游地区风险适中;类型 III 洪水与类型 I 洪水具有较高相似性,但其洪峰流量相对降低,洪水历时延长,洪量更大,整体表现为涨落较快且对称性高的洪

表 1 不同洪水类型的特征指标均值

Table 1 Mean values of characteristic indicators of different flood types

洪水类型	洪水历时是否超过 10 d	洪水次数	洪水历时/d	$Q_m/$ (m^3/s)	T_s	$v_R/$ ($m^3/(s \cdot h)$)	$v_D/$ ($m^3/(s \cdot h)$)	P_H	三峡入库总量/亿 m^3
I	否	7	5.4	54 657	0.44	407.16	250.47	0.42	209.83
	是	7	13.7	52 086	0.38	233.71	180.69	0.37	400.67
II	否	5	7.2	51 740	0.39	262.05	205.08	0.52	283.07
	是	7	19.4	50 471	0.38	169.26	96.37	0.39	551.58
III	否	6	6.7	44 500	0.46	306.63	243.79	0.48	193.36
	是	5	16.5	49 682	0.51	224.10	215.52	0.38	432.63
IV	否	7	6.5	49 214	0.40	389.91	253.83	0.49	209.78
	是	3	15.7	51 250	0.47	172.92	102.90	0.52	504.35

表 2 不同洪水类型的特征描述

Table 2 Characteristic description of different types of floods

洪水类型	主来水区	洪水历时	涨落洪速率	洪水形态	洪水量级
I	岷沱江、嘉陵江	短	陡涨陡落	洪峰靠前且尖锐	洪峰高、洪量大
II	岷沱江、金沙江、向寸区间	长	缓涨缓落	洪峰靠前且集中	洪峰较高、洪量较大
III	岷沱江、嘉陵江、向寸区间、寸三区间	偏长	涨落较快	洪峰居中、洪水对称性高	洪峰适中、洪量较大
IV	乌江、向寸区间、寸三区间	适中	涨落适中	峰形平缓、洪峰靠前	洪峰适中、洪量适中

水特征,风险程度仅次于类型 I 洪水;类型 IV 洪水的高脉冲历时占比较大,其他指标处于适中水平,除 3 场洪水历时较长的洪水外,洪水总量普遍偏小,该类型洪水主要表现为峰形平缓的中洪水过程,风险相对较低。

3.2 基于降雨指标的洪水分类

图 4 为核心子流域、水文区间不同洪水类型的次降雨总量占比。由图 4 可见,不同洪水类型的降雨过程呈现主雨带区域性集中特征,不同子流域、水文区间次降雨总量占比有所差异。金沙江流域次降雨总量占比在 4 种洪水类型中始终不小于 0.20,按该子流域是否与其他流域遭遇,次降雨总量占比呈分级特征,遭遇时占比超 0.35 且集中;乌江流域、向寸区间、寸三区间次降雨总量占比相对较低且变化范围小,多数情况低于 0.20;岷沱江和嘉陵江的次

降雨总量占比相似性高,类型依赖性强,在不同类型间波动范围均较大,为 0.07~0.45。本文结果与张为等^[16]对三峡水库洪水过程研究结论一致,说明按次降雨总量占比进行洪水类型划分具有一定依据。

进一步地,将寸三区间分为寸万区间(寸滩到万县)和万宜区间(万县到宜昌站,即三峡坝址),分析不同洪水类型累积次降雨深在各子流域、水文区间的分布,结果如图 5 所示。发生类型 I 洪水时,主雨带分布区域性特征明显,主来水区岷沱江与嘉陵江降雨集中,流域内各分区的累计次降雨深普遍较高,其他分区降雨特征则有所不同,如:寸万、万宜区间的次降雨深虽然保持有较高水平,但降水分布存在较大不均匀性,降雨深差异较大;除此之外的分区次降雨深则普遍较低,仅为 30~100 mm,降水强度低。发生类型 II 洪水时,不同分区的次降雨深特征差异较小,但整体降水波动性较大且在空间上分散,意味着降水量变化具有较强的不规则性,识别过程中易产生混淆。类型 III 洪水的降雨特征与类型 I 高度相似,两者存在的明显区别在于类型 III 洪水在向寸、寸万、万宜区间的降水分散程度有所降低、降雨分布更均衡,使得区间水量有所提升,一定程度上削弱了类型 III 洪水主雨带的集中性特征。类型 IV 洪水主雨带分布的区域特征突出,主要表现为:在起主导作用的乌江流域和向家坝至三峡流域范围内,次降雨深显著偏高,且分布一致性明显,其他分区的降雨深偏低,在洪水形成过程中贡献相对有限。整体而言,不同洪水类型下,各子流域、水文区间的次降雨深分布展现出与流域整体一致的规律,主雨区在不同洪水类型下呈现出集中或分散性的显著高降雨水平。

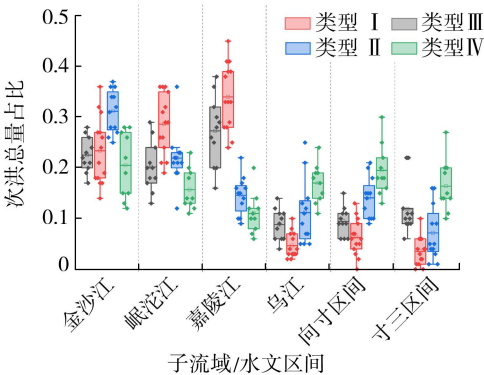


图 4 核心子流域、水文区间不同洪水类型的次降雨总量占比

Fig. 4 Proportion of total secondary net rainfall of different flood types in core sub-basins and hydrological intervals

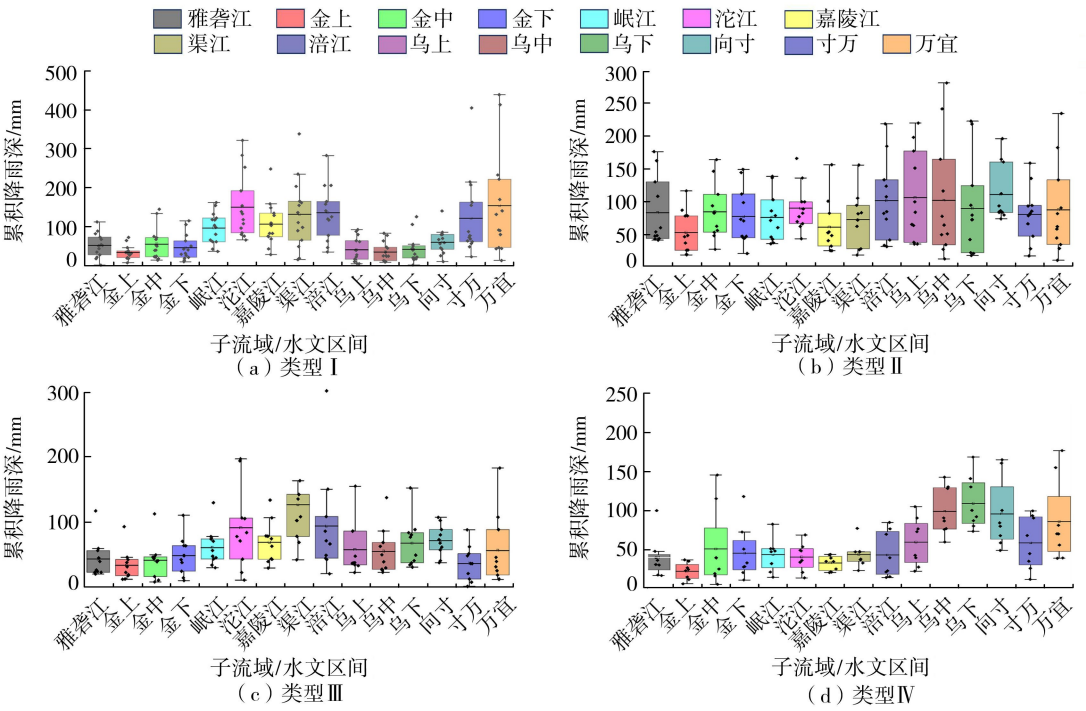


图 5 不同洪水类型的累积次降雨深在各子流域、水文区间的分布

Fig. 5 Distribution of cumulative rainfall depth of different flood types in various sub-basins and hydrological intervals

3.3 分类结果一致性

图 6 为基于洪水分类指标和降雨分类指标的场次洪水分类结果对比。基于次洪总量占比和次降雨总量占比两种指标的分类结果一致性达到 88.4%。尽管基于两种指标的分类结果存在一定偏差,有 5 场洪水分类结果不一致,但表现出高度一致性,且不同类型洪水的暴雨洪水特征均较为显著,表明降雨分类指标在洪水分类中具有较强的可用性。依据不同洪水类型的特征及风险程度,将 5 场分类结果存在差异的洪水从防洪安全角度确认其最终类型归属于对防洪危险更大的类型,供识别使用。

3.4 洪水类型动态识别

3.4.1 识别规则

通过特征排序与模型优化,各子流域、水文区间次降雨总量占比的最佳重要度顺序为:岷江、嘉陵江、寸三区间、向寸区间、雅砻江、沱江、金沙江下游、乌江下游、渠江、涪江、金沙江中游、金沙江上游、乌江中游、乌江上游。挖掘各分区次降雨总量占比特

征与 4 种洪水类型之间的潜在关系,形成识别规则,通过多层次条件判断,实现根据不同降雨数据动态识别洪水类型。

对洪水类型识别结果进行评价,得到不同洪水类型的精确度、召回率和 F1 分数。结果表明,4 种洪水类型中,类型 I 洪水具有最优分类表现,其精确度、召回率和 F1 分数均为 1.00,表明模型在该类型洪水上能够实现完全正确识别,具有完美的分类识别能力。类型 II 洪水的精确度、召回率和 F1 分数都为 0.67,识别结果尚未完全失败,但识别准确度不高,说明该类型洪水的区分存在一定困难,模型在该类别的识别存在较大挑战,一定情况下有误分类问题;该类型洪水主来水区为金沙江及向寸区间,金沙江流域地形复杂(海拔 2 500~4 000 m)、植被覆盖率高,下渗损耗大,导致降雨转化为径流的效率存在空间异质性;金沙江降雨多为持续性小雨,与岷沱江暴雨区的降雨特征差异不显著,加之天气系统(西南涡与西风带系统)叠加导致降雨时空分布分散,增

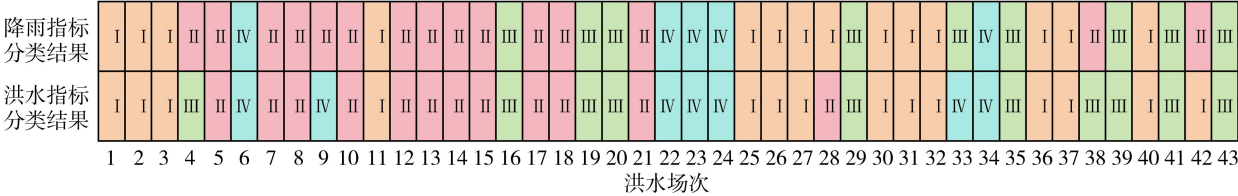


图 6 基于洪水分类指标和降雨分类指标的场次洪水分类结果对比

Fig. 6 Comparison of flood classification results based on flood classification index and rainfall classification index for each event

加识别难度。类型Ⅲ洪水的召回率为 1.00,表明模型成功识别了所有实际属于该类型的洪水样本,但其识别精确度为 0.75 表示有 25%的样本被错误地分为该类别,而 F1 分数为 0.86 说明模型对该类型洪水具有较为平衡的分类性能。类型Ⅳ洪水的精确度为 1.00,表明所有预测为该类型的洪水样本均为正确预测,然而召回率为 0.67 表明仅有 67%的实际该类型的洪水被成功识别,导致 F1 分数为 0.80,具有相对较好的综合性能。

整体而言,虽然本文构建的模型在个别类型洪水中表现有所欠缺,但模型整体的识别准确率结果可达 83.33%,表现良好,尤其是类型Ⅰ洪水上达到完美分类。以各类型洪水的样本数为权重,计算精确度、召回率和 F1 分数 3 个指标的平均值,结果均为 0.83,表明模型保持有较均衡的水平 and 较高的适用性,能有效支持后续分析。

3.4.2 滚动识别过程

以 2014—2021 年长江中上游流域预报降雨信息充足的 28 场洪水为例,按时段进行滚动识别,滚动识别过程可以概括为以下几个关键阶段:①利用预见期为 7 d 的降雨预报信息,在降雨事件发生前 7 天获得降雨第 1 天的降雨信息,在降雨事件发生前 6 天获得降雨第 1、2 天的降雨信息,以此类推,直至降雨事件发生时,获取当天落地雨信息及未来

7 天的降雨信息;综合利用第 1 天的实际落地雨信息和未来 7 天的预报降雨信息进行洪水类型的初步识别。②随着实际降雨过程发展,通过更多累积落地雨信息和未来 7 天的预报降雨信息,实时更新、调整洪水识别模型,进行更精确地洪水类型实时滚动识别,提高准确性。③当主降雨过程已经完成,降水量逐渐趋于稳定,洪水过程基本形成;此时,后续的降雨通常已不具备 7 d 的预见期,因此模型将利用已发生的累积落地雨信息与短预见期的预报降雨信息进行识别,对洪水类型进行最后确认。

图 7~10 为不同类型洪水在降雨过程中按时段滚动识别的结果。由图 7 可见,类型Ⅰ洪水的识别过程中,主来水区域在识别规则中的重要性较高,主要依赖于第一层和第二层识别规则中的寸三区间、岷江、嘉陵江流域的特征进行识别。少数情况下,识别结果位于第四识别层,该类型洪水的识别阈值相对明显,因而识别准确率较高。通过实时滚动各时段的识别结果进行分析,可以观察到识别结果呈现一定的时间分布规律。在使用预报降雨数据进行初期洪水识别时,对于部分洪水事件,存在少量误将洪水识别为区间遭遇型的情况。随着降水过程的逐步发展及降水量的积累,识别结果逐渐趋于稳定,并最终能够实现约 70%的准确率。

由图 8 可见,类型Ⅱ洪水的主来水区涵盖了金

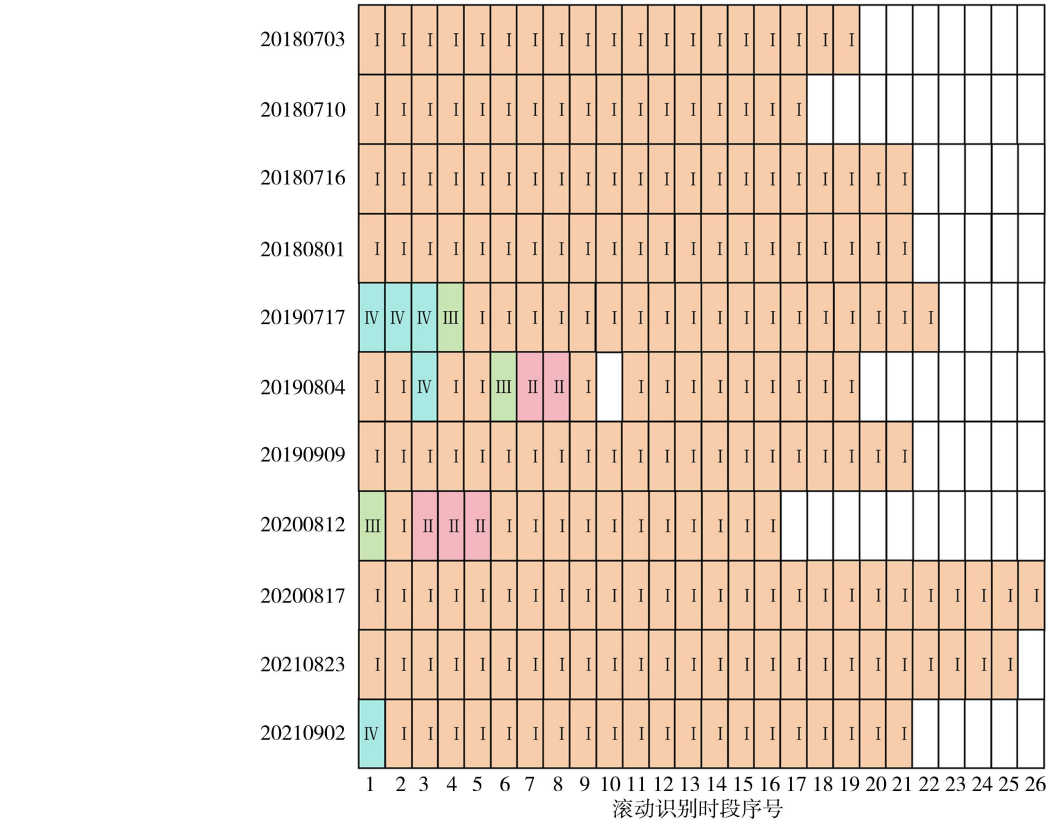


图 7 类型Ⅰ洪水各时段识别结果
Fig. 7 Identification results of type I flood at different time periods

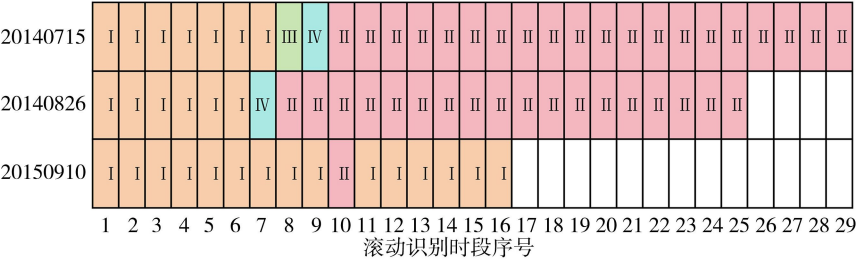


图 8 类型 II 洪水各时段识别结果

Fig. 8 Identification results of type II flood at different time periods

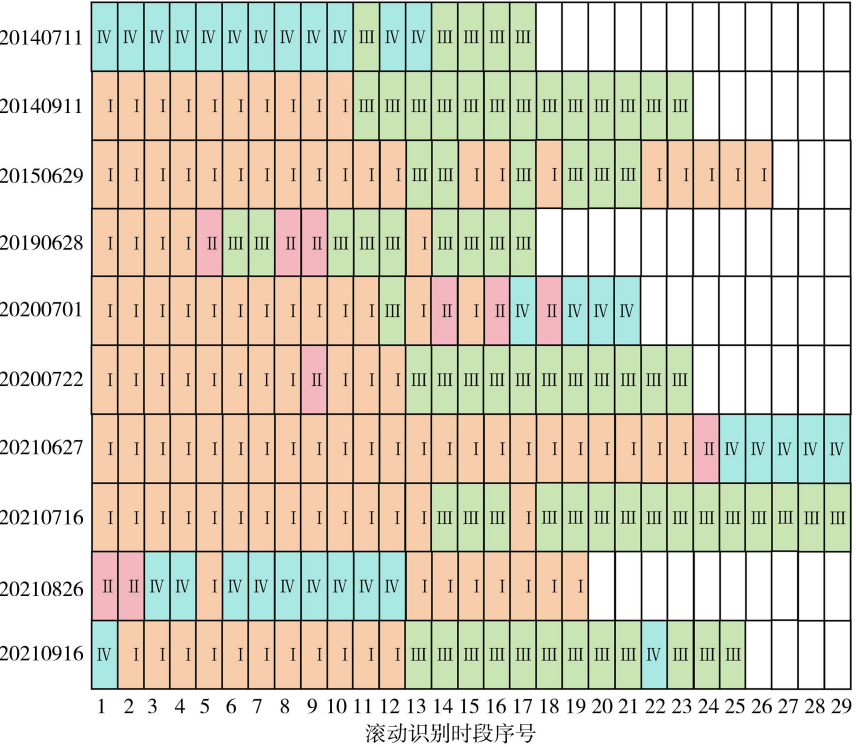


图 9 类型 III 洪水各时段识别结果

Fig. 9 Identification results of type III flood at different time periods

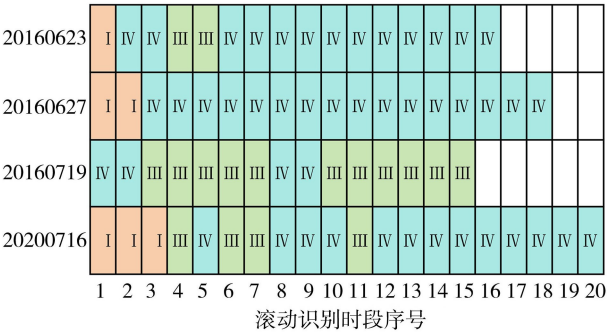


图 10 类型 IV 洪水各时段识别结果

Fig. 10 Identification results of type IV flood at different time periods

沙江、岷江、沱江等重要流域。其中,金沙江在整个类型 II 洪水的形成机制中主要承担起底水作用,与其他流域相比,降水特征相对较弱。因此在降水过程初期,当降水量相对较少时,金沙江流域特征不够鲜明,其降水过程相对平缓,而嘉陵江和岷沱江暴雨

区的降水显著集中特征,使得识别系统仅凭借有限的信息难以有效区分类型 II 洪水与类型 I 洪水,易导致将类型 II 洪水误判为类型 I 洪水的情况发生。

由图 9 可见,类型 III 洪水在滚动时段内的识别精度较低,且呈现较强的分区特征,通常在各流域落地降雨量充分的时段,才能实现较为完整的识别。同时识别结果伴随有滞后现象,在洪水识别的初期,该类型洪水常被误判为类型 I 洪水。一方面是类型 III 洪水与类型 I 洪水在特征上存在较为显著的相似性,该类型洪水主要由类型 I 洪水与区间发生遭遇时形成;另一方面,区间主降雨时段相对较晚,使得该类型洪水与类型 I 洪水的显著差异,通常出现在场次洪水后期。源于两种类型洪水特征的相似性,系统前期难以准确区分两者的不同,识别误差大。

由图 10 可见,类型 IV 洪水通常位于洪水识别规则的末层,且其特征较为集中。该类型的洪水在识别过程中可使用场次较少,洪水情况相对单一,因此

在不同场次的洪水识别结果和各滚动时段的识别结果中,通常表现出适中的识别程度,当各流域的预报信息较为完整时,该类型洪水识别结果较为准确。

不同类型洪水的滚动识别与场次洪水整体识别结果的稳定性有所差异。类型Ⅰ洪水在滚动时段内的识别准确率普遍较高,且识别结果准确,表明这种类型的洪水更容易被准确预测;类型Ⅱ洪水在滚动时段内的识别准确率具有一定波动,其中,第3场洪水准确率极低,最终导致识别结果错误;类型Ⅲ洪水在滚动时段内识别准确率普遍较低,可能由该类型洪水的降雨特征不明显、易发生混淆所导致;类型Ⅳ洪水在滚动时段内识别准确率波动较大,在第3场出现了识别错误。并且,不同洪水类型之间的识别结果也存在差异,表现为:类型Ⅰ洪水在预报降雨初期,提前3~5天可以达到准确识别;类型Ⅱ洪水在落地雨产生1~3天达到准确识别;类型Ⅲ洪水易与类型Ⅰ混淆,在落地雨产生4~6天可以达到较准确的类型识别结果;类型Ⅳ洪水可以提前2~4天较准确识别,但识别结果在类型Ⅱ洪水间存在波动。

对28场洪水动态识别结果进行评价,识别准确率为78.57%。类型Ⅰ洪水的精确度为0.79,召回率为1.00,表明模型在该类型洪水的识别中能在准确性和覆盖面上取得较好的平衡。尽管精确度略低于召回率,但召回率达到1.00、F1分数达到0.88表明模型能够识别所有类型Ⅰ洪水事件,表现较为优异。类型Ⅱ洪水的识别结果在精确度方面表现突出,达到1.00,表明模型对该类型的洪水具有非常高的准确性,但召回率仅为0.67,意味着模型未能全面识别出所有类型Ⅱ洪水事件,因此F1分数仅为0.80,整体表现相对较弱。类型Ⅲ洪水识别的精确度为0.86、召回率为0.60、F1分数为0.71,表明模型对该类型洪水的识别具有一定的优势,对已分类洪水的识别较准确。类型Ⅳ洪水识别精确度为0.60、召回率为0.75、F1分数为0.67,召回率较高,表明模型能识别出更多的类型Ⅳ洪水事件,但精确度较低说明一些被识别为类型Ⅳ的洪水并不符合实际情况。精确度、召回率、F1分数的加权平均值为0.81、0.79、0.78,普遍较高,表明模型对多数洪水能较好地识别,综合表现良好。

4 结 论

本文引入次洪总量占比和次降雨总量占比两种指标,进行了长江中上游流域典型洪水分类,并利用决策树算法构建基于降雨信息的洪水类型识别模式,主要结论如下:

a. 利用K-Means++聚类算法,根据次洪总量占

比,长江中上游流域洪水划分为岷沱嘉陵江遭遇型、岷沱金沙江及向寸区间遭遇型、岷沱嘉陵江及向寸与寸三区间遭遇型、乌江及向寸与寸三区间遭遇型4种类型洪水,不同类型洪水特征区分显著。

b. 根据次降雨总量占比,将场次降雨分为与洪水类型相对应的4种类型,基于洪水指标与降雨指标的分类结果一致性达到88.4%,说明利用降雨指标进行洪水分类可行,能够有效支持洪水类型识别研究。

c. 构建了洪水类型动态识别模式,整体精度为83.3%,对多数洪水能较好识别其实际类型,综合表现良好。对于不同类型洪水,类型Ⅰ洪水识别准确率普遍较高,且识别结果准确;类型Ⅱ洪水识别准确率具有一定波动;类型Ⅲ洪水识别准确率普遍较低,主要由该类洪水降雨特征不明显、易混淆导致;类型Ⅳ洪水识别准确率波动较大。

参考文献:

[1] 周新春,许银山,杨雁飞,等. 长江流域性洪水辨识与特征对比分析[J]. 人民长江,2020,51(12):111-115. (Zhou Xinchun, Xu Yinshan, Yang Yanfei, et al. Identification and characteristic comparison analysis of Basin-wide flood in Yangtze River Basin [J]. Yangtze River,2020,51(12):111-115. (in Chinese))

[2] Merz R, Blöschl G. Regional flood risk: what are the driving processes? [C]//Proceedings of Symposium on Water Resources Systems-Hydrological Risk, Management and Development. Sapporo: IAHS, 2003:49-58.

[3] 李常青. 黄河中游典型流域洪水分类实时预报研究[D]. 郑州:郑州大学,2020.

[4] 马寅午,周晓阳,尚金成,等. 防洪系统洪水分类预测优化调度方法[J]. 水利学报,1997(4):1-8. (Ma Yanwu, Zhou Xiaoyang, Shang Jincheng, et al. A new method for flood types prediction and optimal control [J]. Shuili Xuebao,1997(4):1-8. (in Chinese))

[5] 张艳平,周惠成. 基于暴雨洪水相似性分析的洪水分类研究[J]. 水电能源科学,2012,30(9):50-54. (Zhang Yanping, Zhou Huicheng. Study on flood classification based on storm flood similarity analysis [J]. Water Resources and Power,2012,30(9):50-54. (in Chinese))

[6] Sahoo G B, Ray C, De Carlo E H. Use of neural network to predict flash flood and attendant water qualities of a mountainous stream on Oahu, Hawaii [J]. Journal of Hydrology,2006,327(3/4):525-538.

[7] Dinesh S. Fuzzy classification of simulated droughts and floods of water bodies[J]. Journal of Applied Sciences, 2007,7(18):2610-2616.

[8] Turkington T, Breinl K, Ettema J, et al. A new flood type classification method for use in climate change impact

studies[J]. Weather and Climate Extremes, 2016, 14: 1-16.

[9] Zou Qiang, Liao Li, Ding Yi, et al. Flood classification based on a fuzzy clustering iteration model with combined weight and an immune grey wolf optimizer algorithm[J]. Water. 2019, 11(1): 80.

[10] 甘甜,郑英,蒋云钟,等. 基于 DBSCAN-RF 洪水分类的洪水预报应用研究[J]. 水利水电技术(中英文), 2024, 55 (3): 77-89. (Gan Tian, Zheng Ying, Jiang Yunzhong, et al. Application research on flood classification-prediction based on DBSCAN-RF[J]. Water Resources and Hydropower Engineering, 2024, 55(3): 77-89. (in Chinese))

[11] Huang Minmin, Wang Ying, Jin Shuanggen, et al. Spatial-temporal variations and severity of the 2020 catastrophic floods in the Yangtze River Basin from Sentinel-1 SAR data[J]. Water, 2024, 16(23): 3445.

[12] 丁启,杨汉波,卢韦伟,等. 长江上游支流区域洪水频率分析[J]. 水利水电技术(中英文), 2022, 53 (11): 50-59. (Ding Qi, Yang Hanbo, Lu Weiwei, et al. Regional flood frequency analysis in tributaries of the upper reaches of the Yangtze River [J]. Water Resources and Hydropower Engineering, 2022, 53 (11): 50-59. (in Chinese))

[13] 赵圆元,张行南,黄钰凯,等. 基于动态权重的水文时间序列相似性分析方法[J]. 水电能源科学, 2021, 39 (2): 9-12. (Zhao Yuanyuan, Zhang Xingnan, Huang

Yukai, et al. Similarity analysis method of hydrological time series based on dynamic weight[J]. Water Resources and Power, 2021, 39(2): 9-12. (in Chinese))

[14] Zhang Yongyong, Zhang Yongqiang, Zhai Xiaoyan, et al. Predicting flood event class using a novel class membership function and hydrological modeling [J] Earth's Future, 2024, 12(6): e2023EF004081.

[15] Jia Benjun, Zhou Jianzhong, Zhang Yongchuan, et al. System dynamics model for the coevolution of coupled water supply-power generation-environment systems: Upper Yangtze River Basin, China [J]. Journal of Hydrology, 2021, 593: 125892.

[16] 张为,李昕,翟少军,等. 三峡水库次洪期间降雨特征对洪水过程的影响[J]. 水利学报, 2025, 56(5): 576-586. (Zhang Wei, Li Xin, Zhai Shaojun, et al. Influence of rainfall characteristics on the flooding process during flood events in the Three Gorges Reservoir [J]. Shuili Xuebao, 2025, 56(5): 576-586. (in Chinese))

[17] Sam'an M, Farikhin, Munsarif M. An improved decision tree model through hyperparameter optimization using a modified gray wolf optimization for diabetes classification [J]. Computer Methods in Biomechanics and Biomedical Engineering, 2026, 29(8): 1843-1859.

[18] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning[M]. New York: Springer, 2001.

(收稿日期: 2025 - 10 - 27 编辑: 王芳)

(上接第 138 页)

[29] 高玉琴,王辰,谢梦瑶,等. 洪涝灾害人口生命损失风险动态评估[J]. 水资源保护, 2025, 41 (5): 79-88. (Gao Yuqin, Wang Chen, Xie Mengyao, et al. Dynamic assessment of population life loss risk due to flood disasters[J]. Water Resources Protection, 2025, 41 (5): 79-88. (in Chinese))

[30] 卢兴超,徐宗学,李永坤,等. 基于多智能体模型的城市洪涝灾害动态风险评估[J]. 水资源保护, 2024, 40 (4): 36-47. (Lu Xingchao, Xu Zongxue, Li Yongkun, et al. Assessment on dynamic risk of urban flooding and waterlogging disaster based on multi-agent models [J]. Water Resources Protection, 2024, 40 (4): 36-47. (in Chinese))

[31] Lee H K, Hong W H, Lee Y H. Experimental study on the influence of water depth on the evacuation speed of elderly people in flood conditions [J]. International Journal of Disaster Risk Reduction, 2019, 39: 101198.

[32] Xue Wanjie, Wu Zening, Xu Hongshi, et al. A framework for amplification flood risk assessment and threshold determination of combined rainfall and river level in an inland city[J]. Journal of Hydrology, 2024, 640: 131725.

[33] 薛万杰. 城市暴雨-外河水位联合致涝规律及阈值研究 [D]. 郑州: 郑州大学, 2023.

(收稿日期: 2025 - 10 - 04 编辑: 王芳)

