DOI:10.13476/j.cnki.nsbdqk.2020.0091
潘汀超,戚蓝,田福昌,等.组合赋权-模糊聚类算法的改进及其在洪灾风险评价的应用[J].南水北调与水利科技(中英文),2020,18(5):38-56. PAN T C,QI L,TIAN F C,et al. Improvement of combination weighting-fuzzy clustering algorithm and its application in flood risk assessment[J]. South-to-North Water Transfers and Water Science & Technology,2020,18(5):38-56.(in Chinese)

组合赋权-模糊聚类算法的改进及其在洪灾风险评价的应用

潘汀超,戚蓝,田福昌,苑希民

(天津大学 水利工程仿真与安全国家重点实验室,天津 300350)

摘要:针对洪水灾害风险评价中单一赋权法存在一定局限性和风险等级难以客观划分的问题,提出改进组合赋权-模糊聚类算法,开展洪水灾害风险评价研究。考虑洪灾危险性、敏感性和易损性等3个方面筛选构建评价指标体系,并分别采用直觉模糊层次分析法和VC-CRITIC法赋予主、客观权重,通过改进博弈论组合赋权法计算最优组合权重,加权计算不同评价单元洪灾风险度,利用高斯混合模型模糊聚类算法划分区域洪水灾害风险等级。以茨南淝左片防洪保护区遭遇淮河干流百年一遇洪水为例进行洪灾风险评价算法应用研究,结果表明:极高风险区和高风险区共占保护区总面积的24.87 %,基本为淹没水深较大、地形位指数较低和社会经济价值较高的区域。评价结果较为合理可靠,所提改进组合赋权-模糊聚类算法可为防洪保护区洪水灾害风险评价和防灾减灾决策提供技术支持。

关键词:洪灾风险评价;组合赋权-模糊聚类算法;直觉模糊层次分析法;VC-CRITIC;博弈论;茨南淝左片防洪保护区;淮河干流

中图分类号:TV122.4 **文献标志码:**A **开放科学(资源服务)标志码(OSID):**

洪水灾害是威胁人民生命财产安全的主要自然灾害之一,开展洪水灾害风险评价对防洪减灾应急决策具有重要意义[1-2],其中确定洪灾风险等级标准和评价指标权重是洪灾风险评价的关键环节。

洪灾风险等级的划分具有模糊性和高维性的特点。从前人研究成果分析,在自然断裂法、云模型[3-4]、可变模糊集[5]、信息扩散[6]等诸多评价等级划分方法中,模糊聚类算法可以很好地考虑评价过程中的模糊性与高维性,已经应用于地下水水质评价[7]、变压器运行状态评估[8]等领域,将模糊聚类算法应用于洪水风险评估可以为洪灾风险等级划定提供客观分析结果。

在以往的研究中,洪灾风险评价指标赋权通常使用单一赋权法,如层次分析法[9-11]、专家打分法和德尔菲法等主观赋权法,或使用熵权法[12]、TOPSIS法、主成分分析[13-14]、因子分析[15]等客观赋权法。主观赋权法依赖于决策者的主观经验,存在主观性和随意性;客观赋权法中决策者参与程度低、方法通用性较差,可能会出现指标权重和实际情况相悖的现象;博弈论是一种通过多个决策主体的决策均衡实现各方面利益最大化的方法,结合了主客观赋权法各自的优点,既考虑了实际数据特点又参考了决策者意见,使指标赋权实现了主观与客观的统一,得到较为客观合理的指标权重,在多属性决策问题上

收稿日期:2019-10-09 **修回日期:**2020-03-07 **网络出版时间:**2020-03-26
网络出版地址:http://kns.cnki.net/kcms/detail/13.1334.TV.20200326.1357.002.html
基金项目:国家重点研发计划(2017YFC0405601);高等学校学科创新引智计划(B14012)
作者简介:潘汀超(1995—),男,福建惠安人,主要从事防洪减灾及水利信息化研究。E-mail:pantingchao@126.com
通信作者:苑希民(1968—),男,江苏徐州人,教授,博士,主要从事防洪减灾及水利信息化研究。E-mail:yxm@tju.edu.cn

得到广泛应用[16-18]。

本文在深入研究以往洪水灾害风险评价方法基础上，提出基于组合赋权与模糊聚类的洪灾风险评价改进方法，评价指标主客观赋权分别采用考虑犹豫和弃权情况的直觉模糊层次分析法以及基于客观数据信息量和变异程度的 VC-CRITIC 法，通过改进博弈论组合赋权法进行寻优组合，确定出最优的组合权重；同时，考虑到评价过程中隶属度区间的模糊性，利用高斯混合模型模糊聚类算法划分区域洪水灾害风险等级，并将该方法应用于茨南淝左片防洪保护区洪水灾害风险评价中，研究成果可为防洪保护区洪水灾害风险评价和减灾决策提供技术支持。

1　基于改进组合赋权-模糊聚类算法的洪灾风险评价方法

1.1　基本流程

基于改进组合赋权-模糊聚类算法的洪灾风险评价方法基本流程见图 1。

图 1　基于组合赋权-模糊聚类算法的洪灾风险评价流程

(1)结合评价区域实际情况确定洪灾风险评价指标，并将指标进行归一化处理。

(2)分别采用直觉模糊层次分析法与 VC-CRITIC 法计算洪水风险评价指标主观和客观权重，兼顾主观经验与客观数据所蕴藏的信息。直觉模糊层次分析法相较于模糊层次分析法能够更加准确地反映赋权犹豫或弃权的情况，并且能够使定性和定量因素的集结方法达成一致，因此在表达决策者评价不确定性方面会更加全面、准确。VC-CRITIC 法考虑了客观数据的信息量和变异程度对指标权重的影响，较常用的 CRITIC 法具有显著的优越性。

(3)基于改进博弈论组合赋权法计算最优主客观组合权重。改进博弈论组合赋权法通过借鉴离差最大化组合赋权法的约束条件，相较博弈论组合赋权法，线性组合系数不会出现负数的情况。

(4)根据得到的指标组合权重值，利用 ArcGIS 平台栅格计算器功能叠加各指标标准化数据图层，加权计算不同栅格单元对应的洪灾风险度。

(5)采用戴维森堡丁指数(Davies Bouldin index DBI，表征聚类效果优劣的指数)得出最佳聚类中心数目，根据得到的最佳聚类中心数目，按最大隶属度原则将所有洪灾风险评价单元的风险度聚类，按洪灾风险高低赋予洪灾风险等级。

(6)最后运用 GIS 技术成图，直观展示研究区域内洪灾风险的空间分布。

1.2　关键方法

1.2.1　数据归一化

针对各评价指标量级与量纲间的差异，赋权计算前需要对各评价指标数据进行归一化处理。构建评价矩阵 $\boldsymbol{X}_{m\times n}=(x_{ij})_{m\times n}$，其中 x_{ij} 为第 i 个洪灾风险评价单元的第 j 个评价指标值，洪灾风险评价单元数为 m，评价指标个数为 n。采用极差变换标准化处理方法，构造样本归一化矩阵

$$\boldsymbol{X}^{*}_{m\times n}=(x^{*}_{ij})_{m\times n} \tag{1}$$

对于与风险度呈正相关的指标

$$x^{*}_{ij}=\frac{x_{ij}-x_{\min}(i)}{x_{\max}(i)-x_{\min}(i)} \tag{2}$$

对于与风险度呈负相关的指标

$$x^{*}_{ij}=\frac{x_{\max}(i)-x_{ij}}{x_{\max}(i)-x_{\min}(i)} \tag{3}$$

式中：x^{*}_{ij} 为第 j 个样本第 i 项指标的标准化值，且 $0\leqslant x^{*}_{ij}\leqslant 1$；$x_{\max}(i)$和 $x_{\min}(i)$分别为第 i 项指标对应的所有样本数据中的最大值和最小值。

1.2.2　主客观赋权

1.2.2.1　直觉模糊层次分析法主观赋权

直觉模糊层次分析法(intuitionistic fuzzy analytic hierarchy process，IFAHP)[19]是一种基于模糊层次分析法改进的主观赋权法，计算步骤如下。

(1)直觉模糊判断矩阵构建。两两相互比较指标体系中一、二级指标的重要性，构建直觉模糊判断矩阵 $\boldsymbol{W}=(w_{ij})_{n\times n}$，其中：$i$ 和 j 分别表示直觉模糊判断矩阵中的行和列，$w_{ij}=(u_{ij},v_{ij})$，u_{ij} 表示隶属度，即第 i 个指标比第 j 个指标重要的程度；v_{ij} 表示非隶属度，即第 j 个指标比第 i 个指标重要的程度；π_{ij} 表示犹豫度，$\pi_{ij}=1-u_{ij}-v_{ij}$。

(2)一致性检验及自动迭代调整。为了保证各指标之间重要度的协调一致性，采用直觉模糊信息的距离测度 d 对直觉模糊判断矩阵进行一致性检验，不通过检验则自动进行迭代调整。

(3)权重计算。根据通过一致性检验的直觉模糊判断矩阵，可计算出各个评价指标针对上一层的

权重为

$$w_i=\left[\frac{\sum_{j=1}^{n}u_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n}(1-v_{ij})},1-\frac{\sum_{j=1}^{n}(1-u_{ij})}{\sum_{i=1}^{n}\sum_{j=1}^{n}v_{ij}}\right]$$

$$i=1,2,\cdots,n \tag{4}$$

(4)信息集结。运用直觉模糊集的乘法运算法则可以得到任意二级指标 t 的绝对权重

$$w_k\otimes w_t=(u_k\cdot u_t,v_k+v_t-v_k\cdot v_t) \tag{5}$$

式中:w_k 表示二级指标 t 所属的一级指标 k 的权重;w_t 表示二级指标 t 的权重;u_k、u_t 分别表示 w_k 和 w_t 的隶属度;v_k、v_t 分别是 w_k 和 w_t 的非隶属度。

基于直觉模糊集的排序函数计算任意二级指标 t 的综合权重

$$\rho(W_t)=1-\frac{(1+\pi_t)(1-u_1)}{2} \tag{6}$$

式中:π_t 表示 w_t 的非隶属度;u_1 表示 w_1 的隶属度。

1.2.2.2 VC-CRITIC 法客观赋权

VC-CRITIC 法[20]是一种基于 CRITIC 法的客观赋权方法,可以减少主观因素的影响。该方法考虑了客观数据的信息量和变异程度对指标权重的影响,计算步骤如下。

(1)计算指标的变异系数 v_j。计算公式为

$$v_j=\frac{s_j}{\bar{x}_j}\qquad j=1,2,\cdots,n \tag{7}$$

式中:$\bar{x}_j$ 为第 j 个指标的平均值;s_j 为第 j 个指标的均方差。

(2)计算指标的独立性系数 η_j。计算公式为

$$\eta_j=\sum_{k=1}^{n}(1-r_{kj})\qquad j=1,2,\cdots,n \tag{8}$$

式中:r_{kj} 为第 j 个和第 k 个指标之间的相关性系数。

(3)计算指标的综合性系数 C_j。基于各评价指标的变异系数和独立性系数,可以计算各评价指标的综合性系数,公式为

$$C_j=v_j\eta_j\qquad j=1,2,\cdots,n \tag{9}$$

1.2.3 改进博弈论组合赋权

博弈论组合赋权的基本思想为通过使各个指标权重与最优线性组合指标权重之间的离差极小化,在不同赋权方法求解得到的权重之间寻找一致或妥协。设评价指标个数为 n,计算评价指标权重的方法有 P 种,则可得 P 个基本权重集,其中第 p 个基本权重集可表示为 $\omega_p=\{\omega_{p1},\omega_{p2},\cdots,\omega_{pn}\}$,$p=1,2,\cdots,P$,则 P 个基本权重集任一线性组合可表示为

$$\omega=\sum_{p=1}^{P}\alpha_p\omega_p^{\mathrm{T}} \tag{10}$$

式中:α_p 为线性组合系数。

基于博弈论组合赋权的基本思想,优化的对策模型为

$$\min_{t=1,2,\cdots,P}\left\|\sum_{p=1}^{P}\alpha_p\omega_p^{\mathrm{T}}-\omega_t\right\|_2 \tag{11}$$

式中:ω_t 为第 t 种赋权方法得到的基本权重集。根据矩阵微分性质,为防止组合系数为负,改进博弈论组合赋权[21]最优条件为

$$\min_{\alpha_1,\alpha_2,\cdots,\alpha_P}f=\sum_{t=1}^{P}\left|\left(\sum_{t=1}^{P}\alpha_p\omega_t\omega_p^{\mathrm{T}}\right)-\omega_t\omega_t^{\mathrm{T}}\right| \tag{12}$$

式中:$\alpha_p>0$,$p=1,2,\cdots,P$ 且 $\sum_{p=1}^{P}\alpha_p^2=1$,构建拉格朗日函数求解有

$$L(\alpha_p,\lambda)=\sum_{t=1}^{P}\left|\left(\sum_{p=1}^{P}\alpha_p\omega_t\omega_p^{\mathrm{T}}\right)-\omega_t\omega_p^{\mathrm{T}}\right|+\frac{\lambda}{2}\left(\sum_{p=1}^{P}\alpha_p^2-1\right) \tag{13}$$

组合系数解为

$$\alpha_p=\frac{\sum_{t=1}^{p}\omega_t\omega_p^{\mathrm{T}}}{\sqrt{\sum_{p=1}^{P}\left(\sum_{t=1}^{P}\omega_t\omega_p^{\mathrm{T}}\right)^2}} \tag{14}$$

1.2.4 高斯混合模型模糊聚类

高斯混合模型(Gaussian mixture model,GMM)属于无监督机器学习中的模糊聚类算法,能够从所有评价单元的风险度聚类划分保护区的洪灾风险等级。高斯混合模型本质上是一个描述混合密度分布的经典模型,其概率密度函数可表示为

$$f(x)=\sum_{k=1}^{n}\omega_k\times f_N(x|\mu_k,\sigma_k^2) \tag{15}$$

式中:n 为组成高斯混合模型的正态分布个数;$f_N(x|\mu_k,\sigma_k^2)$为第 k 个正态分布的概率密度函数;ω_k、μ_k 和σ_k^2 分别为第 k 个正态分布的权重、期望和方差。其中,权重 ω_k 满足

$$\begin{cases}0\leqslant\omega_k\leqslant1\\ \sum_{k=1}^{n}\omega_k=1\end{cases} \tag{16}$$

对于高斯混合模型中的参数 ω_k、μ_k 和 σ_k^2,通常采用期望最大化(expectation maximization,EM)算法进行求解。EM 算法的基本思想是通过引入隐含变量,求解模型分布参数的极大似然估计,然后对隐含变量期望公式和模型分布参数重估公式进行反复迭代,直至似然函数值收敛。

采用戴维森堡丁指数(Davies Bouldin index,DBI)分析不同聚类中心数目下的聚类效果[22-23],DBI 越小代表着类自身越紧密,类与类之间越分散,聚类效果更好。DBI 表示洪灾风险等级划分的效果优劣,具体计算方法可见文献[22]。计算公式为

$$\gamma_{\mathrm{DBI}}=\frac{1}{G}\sum_{i=1}^{G}\max\nolimits_{j\neq i}R_{i,j} \tag{17}$$

式中:G 为聚类中心个数;$R_{i,j}$ 为类间相似度,计算公

式为

$$R_{i,j}=(S_i+S_j)/M_{i,j} \tag{18}$$

式中：S_i 和 S_j 分别为 i 类和 j 类内数据到簇质心的平均距离；$M_{i,j}$ 表示 i 类质心和 j 类质心的距离。

2　实例应用

2.1　研究区域概况

以茨南淝左片防洪保护区遭遇淮河干流百年一遇洪水情景为例，进行洪灾风险评价算法及其应用研究。茨南淝左片防洪保护区属淮河中下游平原地区，位于安徽省境内西淝河左堤、茨淮新河右堤与淮北大堤所包围的区域，在东经 116°13′22″～117°12′17″和北纬 32°38′21″～33°2′34″，面积约 1 816.0 km^2，见图 2。

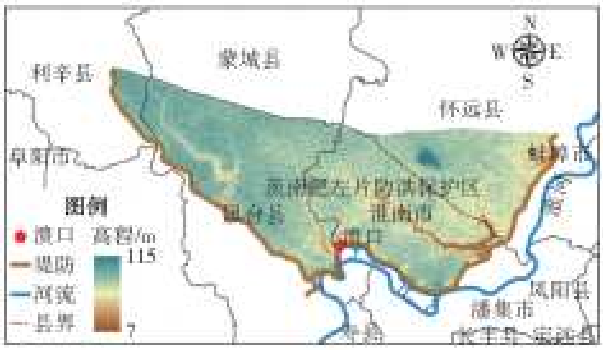


图 2　研究区域地理位置

2.2　指标体系构建及量化

根据灾害系统理论[24]，并参考相关研究[25-26]，结合保护区实际情况，从洪水灾害致灾因子危险性、孕灾环境敏感性、承灾体易损性等 3 个方面对评价指标进行筛选，兼顾指标数据的重要性和获取难易程度，建立代表性高、操作性强的防洪保护区洪水灾害风险评价指标体系。该指标体系共 3 个一级指标和 14 个二级指标，见表 1。评价指标具体说明如下。

（1）致灾因子危险性指标。该指标包括最大淹没水深、前锋到达时间、最大洪水流速和最大淹没历时 4 个洪水风险要素，4 个指标均可通过水力学模型模拟获得，且均为 2010 年水利部发布的《洪水风险图编制导则》中指出的洪水风险要素，是洪水灾害的直接影响因素。

（2）孕灾环境敏感性指标。该指标包括归一化地形位指数、干流堤防缓冲区、区域支流缓冲区、土壤透水能力和植被覆盖率等 5 个指标。高程、植被覆盖率来源自地理空间数据云（http://www.gscloud.cn/），坡度是利用 ArcGIS 平台从高程数据中提取得到的，土壤透水能力来源自世界土壤数据库（harmonized world soil database）。地形位指数指标能综合反映区域高程和坡度：地形位指数越小，区域高程越低；地形坡度越小，洪水越容易汇集，洪灾风险越高。地形位指数

$$T=\log_{10}\left[\left(\frac{H}{H_{\mathrm{mean}}}+1\right)\times\left(\frac{S}{S_{\mathrm{mean}}}+1\right)\right] \tag{19}$$

式中：H 与 H_{mean} 分别为栅格的高程和平均高程；S 与 S_{mean} 分别为坡度与平均坡度。堤防、河流缓冲区级别指标能充分体现河流、湖泊和堤防周边区域受漫溢、溃堤洪水等威胁的洪灾风险。根据防洪保护区内实际情况，堤防和河流缓冲区设为 5 级缓冲区，堤防缓冲区缓冲宽度分别为 500 m、1 km、2 km、3 km 和>3 km，河流缓冲区缓冲宽度分别为 250 m、500 m、750 m、1 km 和>1 km。植被和渗透性较好的土壤对暴雨洪水灾害具有一定的缓冲作用。

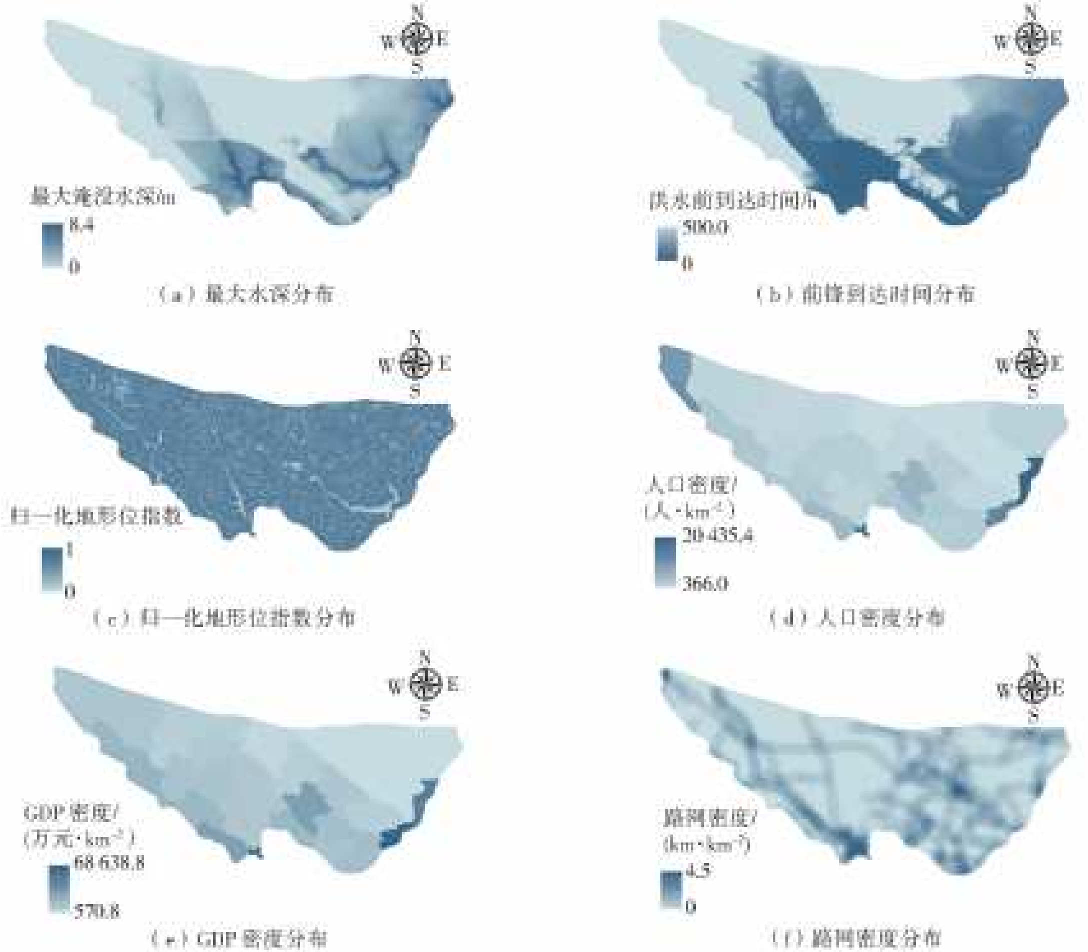
表 1　茨南淝左片防洪保护区洪灾风险评价指标体系

目标层	准则层	指标层
洪水灾害风险评价系统 A	致灾因子危险性 B_1	最大淹没水深 C_{11}/m
		洪水前锋到达时间 C_{12}/h
		最大洪水流速 C_{13}/(m·s^{-1})
		最大淹没历时 C_{14}/h
	孕灾环境敏感性 B_2	地形位指数 C_{21}
		干流堤防缓冲区 C_{22}
		区域支流缓冲区 C_{23}
		土壤透水能力 C_{24}
		植被覆盖率 C_{25}
	承灾体易损性 B_3	人口密度 C_{31}/(人·km^{-2})
		GDP 密度 C_{32}/(万元·km^{-2})
		交通路网密度 C_{33}/(km·km^{-2})
		固定资产密度 C_{34}/(万元·km^{-2})
		土地利用类型 C_{35}

（3）承灾体易损性指标。该指标由承灾体的人口、财产和社会经济等因素共同决定，可将其分为经济易损性指标和社会易损性指标。本研究选取的社会易损性评价指标包括人口密度和土地利用类型，经济易损性评价指标包括 GDP 密度、交通路网密度和固定资产密度。地区易损性指标越高，承灾体潜在威胁越高，洪水灾害风险越大。人口、土地利用及社会经济指标数据来源国家统计资料，具有一定的可靠性和易获取性。

栅格单元具有载体信息分布特征明显、空间位置精确的特点。在 GIS 空间分析技术的支持下，根据防洪保护区面积大小、指标数据分布、灾害特征、乡镇大小、地形地貌特征等，将 150 m×150 m 栅格

作为茨南淝左片防洪保护区评价的基本单元，把防洪保护区划分为64 592个栅格，并将各项指标按照1.1 节陈述方法进行归一化处理。茨南淝左片防洪保护区部分评价指标分布见图 3。对各指标进行相关性分析，结果表明各评价指标间的相关性系数均小于 0.9，指标间的相互影响在合理范围内。



(a) 最大水深分布　(b) 前锋到达时间分布

(c) 归一化地形位指数分布　(d) 人口密度分布

(e) GDP 密度分布　(f) 路网密度分布

图 3　评价指标空间分布

2.3　指标计算评价

按照 1.2.1 的 IFAHP 方法计算主观权重 ω_1，运用 1.2.2 的 VC-CRITIC 法得到客观权重 ω_2。两种方法得到的权重赋值存在差异，基于得到的主观权重集和客观权重集，运用改进博弈论组合赋权法计算组合系数，归一化得到 $\{\alpha_1^*, \alpha_2^*\}=(0.513\ 9, 0.486\ 1)$，组合赋权得出综合权重为 $\omega=(0.094\ 3, 0.091\ 3, 0.074\ 9, 0.058\ 5, 0.085\ 2, 0.076\ 7, 0.075\ 2, 0.064\ 1, 0.072\ 1, 0.068\ 9, 0.059\ 6, 0.060\ 1, 0.063\ 9, 0.055\ 2)^{\mathrm{T}}$，分别对应最大淹没水深、洪水到达时间、最大洪水流速、最大淹没历时、地形位指数、干流堤防缓冲区、区域支流缓冲区、土壤类型、植被覆盖、人口密度、GDP 密度、交通路网密度和固定资产密度和土地利用类型的权重，见图 4。

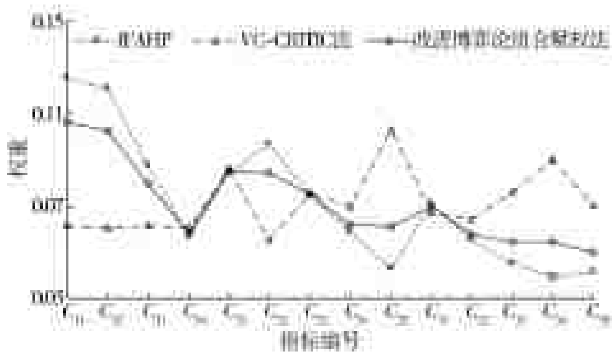


图 4　采用不同方法的指标权重变化曲线

2.4　评价结果分析

按照 1.4 的方法对各评价单元风险度进行 GMM 模糊聚类分析，取模糊聚类中心数目 $G\in[2,10]$，DBI 随聚类中心数目 G 变化的曲线见图 5。

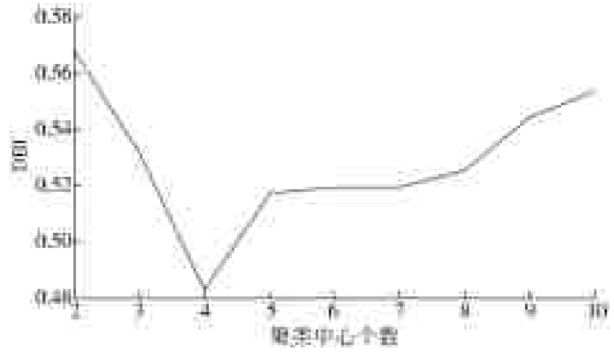


图 5　DBI 与聚类中心数目 G 的关系

由图 5 可知，聚类中心个数为 4 时 DBI 最小，因此确定评价单元最佳聚类数目 $G=4$。按最大隶属度原则，将各评价单元按洪灾风险高低分为高风险区、中高风险区、中等风险区和低风险区等 4 类。为了更直观地展示防洪保护区内洪灾风险空间分布，利用 ArcGIS 制图，得到茨南淝左片防洪保护区洪灾风险等级分布，见图 6。

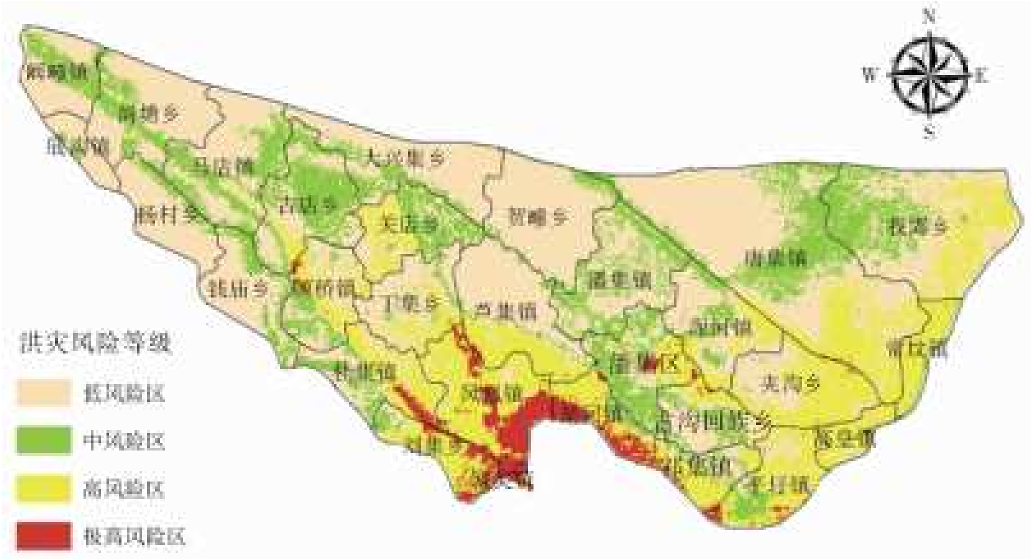


图 6　基于改进组合赋权-模糊聚类算法的茨南淝左片防洪保护区洪灾风险等级分布

运用 ArcGIS 中的区域分析工具对茨南淝左片防洪保护区风险等级分布特征进行分析，得到研究区域不同洪灾风险等级所占面积比例和分布情况。

(1)极高风险区占保护区总面积的 3.29 %，主要位于茨南淝左片防洪保护区南部，分布于城关镇大部、桂集镇中南部、城北乡中部和南部、架河乡南部、平圩镇西南部、田集镇中部和顾桥镇西北部等区域。该区域属于极高风险区的主要原因是最大淹没水深较大、地形位指数较小且比邻淮河干流，易受到淮河干流洪水的影响，城关镇大部的洪灾风险等级较高主要因为路网密度、河网密度、人口和 GDP 密度较高，桂集镇中南部洪灾风险等级较高还受到路网和支流河网较为密集的影响。

(2)高风险区占比为 21.58%，主要位于刘集镇中部、桂集镇东部、城关镇北部、架河乡北部、田集乡东部、古沟回族乡北部、祁集乡南部、平圩镇大部、夹沟乡东部、高皇镇大部、万福镇南部、常坟镇大部和找郢乡东部等区域。该区域属于高风险区主要是因为其处于河流周边区域，洪水前锋到达时间较短且具有较高的社会价值。

(3)中风险区和低风险区分别占 19.20% 和 55.93%，主要位于研究区域的北部。洪灾风险等级较低的原因是地势较高、河网密度较低、植被覆盖率较高、人口和工矿企业分布较少而且土壤透水性较好。

由此可见，茨南淝左片防洪保护区洪灾风险评价结果与洪水威胁程度、地形地貌、社会经济等因素关联紧密。

为了与实际洪涝灾害情况进行比较，选取情景较为接近的淮河干流 1954、1991 年典型极端洪水灾害，以洪水淹没范围为基础，并结合灾情记录，洪涝灾害较严重的区域，即凤台县南部区域和潘集区，与图 6 中的极高风险区较为一致，从而验证了将改进组合赋权-模糊聚类算法应用于洪灾风险评价的评估得到的风险区分布较为合理可靠，研究成果可为该区域防洪评价及洪灾风险防御提供科学依据。

为了进一步验证改进组合赋权-模糊聚类算法的可靠性，选取风险评估中常用的 TOPSIS 进行研究区域洪涝风险评估，将评估结果与改进组合赋权-模糊聚类算法对比。TOPSIS(technique for order preference by similarity to an ideal solution)是基于评价事物与理想化目标的接近程度排序从而进行评估的方法。采用自然断裂法分级的 TOPSIS 法洪涝风险评估结果见图 7，与改进组合赋权-模糊聚类算法评估结果相比，识别的极高风险区和高风险区所占比例极低，分别为 0.19 %及 2.86 %，呈现出一定的不合理性，如实际洪涝灾害较严重的潘集区大部和河流缓冲区在图 7 中仅被识别为低风险区域，评估结果与实际有一定的偏差。改进组合赋权-模糊聚类算法能够较好识别致灾因子、孕灾环境和承灾体在灾害系统的作用，评价结果客观可信，能够为类似区域洪灾风险评价提供技术支持与参考。

图 7　基于 TOPSIS 的茨南淝左片防洪保护区洪灾风险等级分布

3　结　论

(1)构建了包含致灾因子危险性、孕灾环境敏感性和承灾体易损性的洪灾风险评价指标体系,提出了基于直觉模糊层次分析法、VC-CRITIC 法和改进博弈论的主客观组合赋权改进算法,实现了单位约束条件下评估指标组合赋权的寻优计算,使评价指标赋权更具科学性和合理性。

(2)利用高斯混合模型模糊聚类算法对所有评价单元的风险度进行聚类,根据 DBI 指标分析确定最佳聚类中心数目,并按最大隶属度原则将各个洪灾风险评价单元按洪灾风险高低赋予洪灾风险等级,弥补了传统评价方法对评价区间模糊性考虑的不足,能够较为客观准确地提供洪灾风险等级划分结果。

(3)以茨南淝左片防洪保护区遭遇淮河干流百年一遇洪水为研究工况,构建改进组合赋权与模糊聚类洪水风险评价模型,并开展洪灾风险评价。评价结果表明,极高风险区和高风险区共占保护区总面积的 24.87%,基本为淹没水深较大、地形位指数较低和社会经济价值较高的区域,评价结果与区域实际情况基本相符,所提评价方法较为合理可靠,可为区域防洪风险评价与灾害防御提供参考依据。

参考文献:

[1]　苑希民,王亚东,田福昌.溃漫堤洪水多维耦联数值模型及应用[J].天津大学学报,2018,51(7):675-683. DOI:10.11784/tdxbz201707017.

[2]　KOMOLAFE A A ,HERATH S ,AVTAR R. Establishment of detailed loss functions for the urban flood risk assessment in Chao Phraya River basin,Thailand[J]. Geomatics Natural Hazards and Risk,2019,10(1):633-650. DOI:10.1080/19475705.2018.1539038.

[3]　石晓静,查小春,刘嘉慧,等.基于云模型的汉江上游安康市洪水灾害风险评价[J].水利水电科技进展,2017,37(3):29-34. DOI:10.3880/j.issn.10067647.2017.03.005.

[4]　YANG S, HAN X, BO C, et al. Cloud-model-based method for risk assessment of mountain torrent disasters[J]. Water, 2018, 10(7): 830-. DOI: 10.3390/w10070830.

[5]　任玉峰,刘国东,周理,等 基于证据理论和可变模糊集的成都市洪灾风险评估[J].农业工程学报,2014,30(21): 147-156. DOI: 10.3969/j.issn.1002-6819.2014.21.018.

[6]　LONGJIAN Z. Using information diffusion to analyze the membership degree in risk assessment of flash flood[J]. Journal of Water Resources Research,2016,5(6):598-604. DOI:10.12677/JWRR.2016.56069.

[7]　LEE K J, YUN S T, YU S, et al. The combined use of self-organizing map technique and fuzzy c-means clustering to evaluate urban groundwater quality in Seoul metropolitan city, south Korea[J]. Journal of Hydrology, 2019, 569: 685-697. DOI: 10.1016/j.jhydrol.2018.12.031.

[8]　EKE S, CLERC G, AKA-NGNUI T, et al. Transformercondition assessment using fuzzy c-means clustering techniques[J]. IEEE Electrical Insulation Magazine, 2019, 35(2): 47-55. DOI: 10.1109/MEI.2019.8636165.

[9]　蒋雯京,程春梅,张艳蓓,等.基于 GIS/AHP 集成的浙江省洪涝灾害风险评估[J].测绘通报,2019,503(2):134-139. DOI:CNKI:SUN:CHTB.0.2019-02-027.

[10] GHOSH A，KAR S K. Application of analytical hierarchy process（AHP） for flood risk assessment：A case study in Malda district of west Bengal，India[J]. Natural Hazards，2018，94（1）：349-368. DOI：10. 1007/s11069-018-3392-y.

[11] SEEJATA K，YODYING A，WONGTHADAM T，et al. Assessment of flood hazard areas using analytical hierarchy process over the Lower Yom basin，Sukhothai Province[J]. Procedia Engineering，2018(212)：340-347. DOI：10. 1016/j. proeng. 2018. 01. 044.

[12] HONGSHI X，CHAO M，JIJIAN L，et al. Urban flooding risk assessment based on an integrated k-means cluster algorithm and improved entropy weight method in the region of Haikou，China[J]. Journal of Hydrology，2018，563：975-986. DOI：10. 1016/j. jhydrol. 2018. 06. 060.

[13] 王一新，苑希民. 基于主成分分析的城市防洪减灾能力综合评价[J]. 自然灾害学报，2016，25(6)：3-10. DOI：CNKI：SUN：ZRZH. 0. 2016-06-001.

[14] KOLLI Z，JOUANNIC G，MARCHETTI M ，et al. Flood risk vulnerability assessment：hierarchization of the main factors at a regional scale[A]. 3rd European Conference on Flood Risk Management[C]. Lyon，France，2016. DOI：10. 1051/e3sconf/20160708014.

[15] 高玉琴，吴靖靖，王宗志，等. 基于模糊聚类因子分析的区域洪灾风险评价[J]. 自然灾害学报，2018，27(2)：143-150. DOI：CNKI：SUN：ZRZH. 0. 2018-02-017.

[16] 路遥，徐林荣，陈舒阳，等. 博弈论组合赋权的泥石流危险度评价[J]. 灾害学，2014(1)：194-200. DOI：10. 3969/j. issn. 1000-811X. 2014. 01. 035.

[17] LAI C，CHEN X，CHEN X，et al. A fuzzy comprehensive evaluation model for flood risk based on the combination weight of game theory[J]. Natural Hazards，2015，77(2)：1243-1259. DOI：10. 1007/s11069-015-1645-6.

[18] 张力澜，赖成光，陈晓宏，等. 基于博弈论组合赋权的洪灾风险评价及其年代际演变[J]. 热带地理，2018，38(2)：217-225. DOI：10. 13284/j. cnki. rddl. 003018.

[19] XU Z，LIAO H. Intuitionistic fuzzy analytic hierarchy process[J]. IEEE Transactions on Fuzzy Systems，2014，22(4)：749-761. DOI：10. 1109/TFUZZ. 2013. 2272585.

[20] YIN Q，NIU K，LI N. Using CV-CRITIC to determine weights for smart city evaluation[C]//IEEE International Conference on Tools with Artificial Intelligence. IEEE Computer Society，2017(1)：996-1000. DOI：10. 1109/ICTAI. 2017. 00153.

[21] 李爱华. 岸边集装箱起重机安全评价方法研究[D]. 武汉：武汉理工大学，2017.

[22] DAVIES D L，BOULDIN D W. A cluster separation measure[J]. IEEE transactions on pattern analysis and machine intelligence，1979，PAMI-1(2)：224-227. DOI：10. 1109/TPAMI. 1979. 4766909.

[23] BANDYOPADHYAY S，MAULIK U. Genetic clustering for automatic evolution of clusters and application to image classification[J]. Pattern Recognition，2002，35(6)：1197-1208. DOI：10. 1016/s0031-3203(01)00108-x.

[24] 史培军. 四论灾害系统研究的理论与实践[J]. 自然灾害学报，2005，14(6)：1-7. DOI：10. 13577/j. jnd. 2005. 0601.

[25] 李绍飞，冯平，孙书洪. 突变理论在蓄滞洪区洪灾风险评价中的应用[J]. 自然灾害学报，2010，19(3)：132-138. DOI：10. 13577/j. jnd. 2010. 0322.

[26] 吴泽宁，申言霞，王慧亮. 基于能值理论的洪涝灾害脆弱性评估[J]. 南水北调与水利科技，2018，16(6)：13-18，36. DOI：10. 13476/. cnki. nsbdqk. 2018. 0146.

·译文(Translation)·

DOI:10.13476/j.cnki.nsbdqk.2020.0091

Improvement of combination weighting-fuzzy clustering algorithm and its application in flood risk assessmen

PAN Tingchao, QI Lan, TIAN Fuchang, YUAN Ximin

(State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, Tianjin 300350, China)

Abstract: Given the limitation of the single weighting method in flood risk assessment and the difficulty of the objective division of risk level, an improved combination weighting-fuzzy clustering algorithm is proposed to carry out the research of flood risk assessment. Considering the three aspects of flood risk, sensitivity and vulnerability, the evaluation index system is constructed. The subjective and objective weights are given by the intuitionistic fuzzy analytic hierarchy process and VC-CRITIC method. The optimal combination is calculated by the game-theory-based combination weighting method. Combination weights are used to calculate the flood risk of different evaluation units, and the Gaussian-mixture-model-based fuzzy clustering algorithm is used to classify the regional flood risk levels. Taking the 100-year flood of the mainstream of the Huai River as an example, the application of the flood risk assessment algorithm was carried out in the Cinanfeizuopian flood control protected zone. The results show that the extremely high-risk area and high-risk area which account for 24.87% of the total area of the protected area submerge. In areas with large water depth, low topographic index and high socio-economic value, the evaluation results are more reasonable and reliable. The improved combination weighting-fuzzy clustering algorithm may provide flood disaster risk assessment and disaster prevention and mitigation decision-making in flood control protected zones with technical support.

Key words: flood risk assessment; combination weighting-fuzzy clustering algorithm; intuitionistic fuzzy analytic hierarchy process; VC-CRITIC; game theory; Cinanfeizuopian flood control protected zone

Floods are one of the major natural disasters that threaten people's lives and property. Conducting flood risk assessment is of great significance for emergency decision-making on flood control[1-2]. Determining the criteria of flood risk levels and the weights of assessment indexes is a key aspect of flood risk assessment.

The classification of flood risk levels is characterized by fuzziness and high dimensionality. From the previous studies, it can be obtained that among many classification methods of assessment levels such as the natural breaks method, the cloud model-based method[3-4], the variable fuzzy set method[5] and the information diffusion method[6], the fuzzy clustering algorithm can well take into account the fuzziness and high dimensionality in the assessment process. It has been applied in fields such as groundwater quality evaluation[7] and oper-

Received: 2019-10-09 **Revised**: 2020-03-07 **Online publishing**: 2020-03-26

Online publishing address: http://kns.cnki.net/kcms/detail/13.1334.TV.20200326.1357.002.html

Funds: National Key Research and Development Program of China (2017YFC0405601); Talent Attracting Project Foundation for the Discipline Innovation of Colleges and Universities (B14102)

Author brief: PAN Tingchao (1995-), male, Huian Fujian Province, mainly engaged in the research on flood control and hydraulic informatization. E-mail: pantingchao@126.com

Corresponding author: YUAN Ximin (1968-), male, professor, PhD, Xuzhou Jiangsu Province, mainly engaged in the research on flood control and hydraulic informatization. E-mail: yxm@tju.edu.cn

ating condition assessment of transformers[8]. If the fuzzy clustering algorithm is applied to flood risk assessment, it can provide objective analysis results for the classification of flood risk levels.

In the previous studies, the weighting of flood-risk assessment indexes usually used the single weighting method. It includes the subjective weighting methods such as the analytic hierarchy process[9-11], the expert scoring method and the Delphi method, and the objective weighting methods such as the entropy weight method[12], the technique for order preference by similarity to an ideal solution (TOPSIS) method, the principal component analysis method[13-14] and the factor analysis method[15]. The subjective weighting methods depend on the experience of decision makers, which have certain subjectivity and arbitrariness. The decision makers participate less in the objective weighting methods, and the methods are not very common. The weights of indexes may be contrary to the actual situation. The game theory is a method for maximizing the interests of all parties through the decisional balance of multiple decision makers. It combines the advantages of subjective and objective weighting methods. With considering the characteristics of actual data and referring to decision makers' opinions, the game theory realizes the unification of subjective and objective for index weighting. It can obtain more objective and reasonable weights of indexes, so it has been widely used in multi-attribute decision-making problems[16-18].

Based on deeply studying the previous flood risk assessment methods, this paper proposes an improved method of flood risk assessment based on combination weighting and fuzzy clustering. The intuitionistic fuzzy analytic hierarchy process (IFAHP), which considers hesitation and abstention, and the VC-CRITIC method, which is based on the information amount and the variation degree of objective data, are used for subjective and objective weighting of assessment indexes. We search optimal combinations by improving the combination weighting method based on game theory, so as to confirm the optimal combination weights. Meanwhile, the fuzzy clustering algorithm based on Gaussian mixture model (GMM) is used to classify regional flood risk levels given the fuzzy nature of membership intervals in the assessment process. This method is applied to the flood risk assessment in the Cinanfeizuopian flood control protected zone. The results can provide technical support for flood risk assessment and decision making on disaster reduction in the flood control protected zone.

1　Flood risk assessment method based on improved combination weighting-fuzzy clustering algorithm

1.1　Basic process

The basic process of the flood risk assessment method based on the improved combination weighting-fuzzy clustering algorithm is shown in Fig. 1.

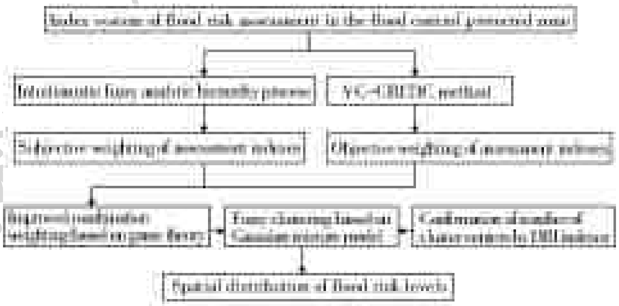


Fig. 1　Process of flood risk assessment based on combination weighting and fuzzy clustering algorithm

(1) The indexes of flood risk assessment are determined combining the actual situation of the evaluated zone and then normalized.

(2) The subjective and objective weights of flood risk assessment indexes are respectively calculated by the IFAHP and the VC-CRITIC method. This method takes both the subjective experience and the information contained in the objective data into consideration. Compared with the fuzzy analytic hierarchy process, the IFAHP can more accurately reflect the situations of weighting hesitation or abstention, and can reach agreement on the aggregated method of qualitative and quantitative factors. Thus, it is more comprehensive and accurate in describing decision makers' assessment uncertainties. The VC-CRITIC method considers the influences of the information amount and the variation degree of objective data on weights of in-

dexes, so it has significant advantages compared with the common CRITIC method.

(3) The optimal subjective and objective combination weights are calculated by the improved combination weighting method based on game theory. The improved combination weighting method based on game theory draws on the constraint condition of the deviation maximization-based combination weighting method. Thus, unlike the combination weighting method based on game theory, the coefficients of linear combinations in the improved combination weighting method based on game theory will not be negative.

(4) According to the obtained combination weights of indexes, the standard data layers of indexes are overlaid through the calculator function of grids at the ArcGIS platform, so as to weight the flood risk levels corresponding to different grid units.

(5) The optimal number of cluster centers is obtained by the Davies Bouldin index (DBI) which can evaluate the quality of clustering. According to the obtained optimal number of cluster centers, the risk levels of all flood risk assessment units are clustered based on the maximum membership principle. Then flood risk levels are assigned according to risk degrees.

(6) At last, the map is made by the GIS technology, and the spatial distribution of flood risks in the studied zone is intuitively demonstrated.

1.2 Key methods

1.2.1 Data normalization

As the assessment indexes vary in the dimension and the order, the assessment indexes are normalized before weighting calculation. The judgement matrix $\boldsymbol{X}_{m\times n}=(x_{ij})_{m\times n}$ is established, where: x_{ij} is the No. j assessment index of No. i flood risk assessment unit; m is the number of flood risk assessment units; n is the number of assessment indexes. By the standardized processing method of range transformation, the normalized matrix of samples is constructed

$$\boldsymbol{X}_{m\times n}^{*}=(x_{ij}^{*})_{m\times n} \tag{1}$$

The indexes which are positively correlated with the risk level are

$$x_{ij}^{*}=\frac{x_{ij}-x_{\min}(i)}{x_{\max}(i)-x_{\min}(i)} \tag{2}$$

The indexes which are negatively correlated with the risk level are

$$x_{ij}^{*}=\frac{x_{\max}(i)-x_{ij}}{x_{\max}(i)-x_{\min}(i)} \tag{3}$$

where: x_{ij}^{*} is the standard value of No. i index of No. j sample, and $0\leqslant x_{ij}^{*}\leqslant 1$; $x_{\max}(i)$ and $x_{\min}(i)$ are maximum and minimum of all the sample data corresponding to No. i index.

1.2.2 Subjective and objective weighting

1.2.2.1 Subjective weighting by IFAHP

The IFAHP[19] is a subjective weighting method which is improved based on the fuzzy analytic hierarchy process. The calculation process is as follows.

(1) Establishment of the intuitionistic fuzzy judgement matrix. The importance degrees of primary and second-level indexes in the index system are pairwisely compared, so the intuitionistic fuzzy judgement matrix $W=(w_{ij})_{n\times n}$ is established, where: i and j indicate the row and the column of the intuitionistic fuzzy judgement matrix; $w_{ij}=(u_{ij}, v_{ij})$, u_{ij} indicates the membership, namely that the degree to which No. i index is more important than No. j index; v_{ij} indicates the non-membership, namely that the degree to which No. j index is more important than No. i index; π_{ij} indicates the hesitancy degree, and $\pi_{ij}=1-u_{ij}-v_{ij}$.

(2) Consistency check and automatic iteration adjustment. To assure the coordinated consistency of indexes on the importance degree, the method uses the distance measure d of intuitionistic fuzzy information to conduct consistency check for the intuitionistic fuzzy judgement matrix. If the matrix does not pass the check, iteration adjustment is performed automatically.

(3) Weight calculation. According to the intuitionistic fuzzy judgement matrix which has passed the consistency check, the weight of each assessment index for the upper level can be obtained as

$$w_{i}=\left(\frac{\sum_{j=1}^{n}u_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n}(1-v_{ij})},1-\frac{\sum_{j=1}^{n}(1-u_{ij})}{\sum_{i=1}^{n}\sum_{j=1}^{n}v_{ij}}\right)\quad i=1,2,\cdots,n \tag{4}$$

(4) Information aggregation. The absolute weight of any second-level index t can be obtained by the multiplication of the intuitionistic fuzzy set

$$w_k \otimes w_t = (u_k \cdot u_t, v_k + v_t - v_k \cdot v_t) \tag{5}$$

where: w_k indicates the weight of the primary index k to which the second-level index t belongs; w_t indicates the weight of the second-level index t; u_k and u_t indicate the memberships of w_k and w_t; v_k and v_t indicate the non-memberships of w_k and w_t.

The comprehensive weight of any second-level index t is calculated by the ranking function of the intuitionistic fuzzy set

$$\rho(W_t) = 1 - \frac{(1+\pi_t)(1-u_1)}{2} \tag{6}$$

where: π_t indicates the non-membership of w_t; u_1 indicates the membership of w_1.

1.2.2.2 Objective weighting by VC-CRITIC method

The VC-CRITIC method[20] is an objective weighting method based on the CRITIC method, which can reduce the influences of subjective factors. The method takes the influences of the information amount and the variation degree of objective data on weights of indexes into consideration. The calculation steps are as follows.

(1) Calculate the coefficient of variation v_j of indexes. The formula is

$$v_j = \frac{s_j}{\bar{x}_j} \qquad j = 1, 2, \cdots, n \tag{7}$$

where: $\bar{x}_j$ is the average of No. j index; s_j is the mean square deviation of No. j index.

(2) Calculate the dependence coefficient η_j of indexes. The formula is

$$\eta_j = \sum_{k=1}^{n} (1 - r_{kj}) \qquad j = 1, 2, \cdots, n \tag{8}$$

where: r_{kj} is the correlation coefficient between No. j index and No. k index.

(3) Calculate the comprehensive coefficient C_j of indexes. Based on the coefficient of variation and dependence coefficient of each index, the comprehensive coefficient of each assessment index can be calculated by the following formula

$$C_j = v_j \eta_j \qquad j = 1, 2, \cdots, n \tag{9}$$

1.2.3 Improved combination weighting method based on game theory

The basic principle of the combination weighting method based on game theory is to search consistency or compromise among weights obtained through different weighting methods by minimizing the deviation between the weights of each index and the optimal linear combination weights. It is set that the number of assessment indexes is n and the number of methods for calculating weights of assessment indexes is P. Thus, P basic weight sets can be obtained, and No. p basic weight set can be indicated as $\omega_p = \{\omega_{p1}, \omega_{p2}, \cdots, \omega_{pn}\}$, $p = 1, 2, \cdots, P$. Any linear combination of P basic weight sets can be indicated as

$$\omega = \sum_{p=1}^{P} \alpha_p \omega_p^{\mathrm{T}} \tag{10}$$

where: α_p is the coefficient of linear combinations.

Based on the basic principle of the combination weighting method based on game theory, the optimized gaming model is

$$\min_{t=1,2,\cdots,P} \left\| \sum_{p=1}^{P} \alpha_p \omega_p^{\mathrm{T}} - \omega_t \right\|_2 \tag{11}$$

where ω_t is the basic weight set obtained through No. t weighting method. According to the differential property of matrices, to avoid negative combination coefficients, the optimal condition of the improved combination weighting method based on game theory[21] is

$$\min_{a_1, a_2, \cdots, a_P} f = \sum_{t=1}^{P} \left| \left(\sum_{t=1}^{P} \alpha_P \omega_t \omega_p^{\mathrm{T}} \right) - \omega_t \omega_t^{\mathrm{T}} \right| \tag{12}$$

where $\alpha_P > 0$, $p = 1, 2, \cdots, P$ and $\sum_{p=1}^{P} \alpha_p^2 = 1$. The Lagrange function is established as

$$L(\alpha_p, \lambda) = \sum_{t=1}^{P} \left| \left(\sum_{p=1}^{P} \alpha_p \omega_t \omega_p^{\mathrm{T}} \right) - \omega_t \omega_p^{\mathrm{T}} \right| + \frac{\lambda}{2} \left(\sum_{p=1}^{P} \alpha_p^2 - 1 \right) \tag{13}$$

The solution to the combination coefficient is

$$\alpha_p = \frac{\sum_{t=1}^{p} \omega_t \omega_p^{\mathrm{T}}}{\sqrt{\sum_{p=1}^{P} \left(\sum_{t=1}^{P} \omega_t \omega_p^{\mathrm{T}} \right)^2}} \tag{14}$$

1.2.4 Fuzzy clustering based on GMM

The Gaussian mixture model (GMM) is a fuzzy clustering algorithm in non-supervised machine learning, which can classify the flood risk levels of a protected zone from the risk clustering of all assessment units. The essence of the GMM is a classical model of describing the mixture density distribution. Its probability density function can be indicated as

$$f(x) = \sum_{k=1}^{n} \omega_k \times f_N(x \mid \mu_k, \sigma_k^2) \tag{15}$$

where: n is the number of normal distributions constituting the GMM; $f_N(x|\mu_k,\sigma_k^2)$ is the probability density function of No. k normal distribution; ω_k, μ_k and σ_k^2 are weight, expectation and variance of No. k normal distribution. The weight ω_k satisfies

$$\begin{cases} 0\leqslant\omega_k\leqslant 1 \\ \sum_{k=1}^{n}\omega_k=1 \end{cases} \tag{16}$$

The expectation maximization (EM) algorithm is usually used to solve the parameters (ω_k、μ_k and σ_k^2) in the GMM. The basic principle of the EM algorithm is to solve the maximum likelihood estimation of distribution parameters of models by introducing latent variables, and then to iterate expectation formulas of latent variables and re-estimating formulas of distribution parameters of models until the likelihood function values are convergent.

The Davies Bouldin index (DBI) is used to analyze the clustering effects with different numbers of cluster centers[22-23]. With the decrease of the value of DBI, the class itself is tighter; the classes are more dispersed, and the clustering effect is better. DBI indicates the quality of the classification of flood risk levels, and the specific calculation method can be referred to Reference [22]. The calculation formula is as follows

$$\gamma_{\mathrm{DBI}}=\frac{1}{G}\sum_{i=1}^{G}\max\nolimits_{j\neq 1}R_{i,j} \tag{17}$$

where: G is the number of cluster centers; $R_{i,j}$ is the similarity between classes, which is calculated by

$$R_{i,j}=(S_i+S_j)/M_{i,j} \tag{18}$$

where: S_i and S_j are average distances of data in No. i class and No. j class from the cluster centroid; $M_{i,j}$ indicates the distance between the centroid of No. i class and the centroid of No. j class.

2 Case application

2.1 Introduction to the studied region

With the 100-year flood of the mainstream of the Huaihe River in the Cinanfeizuopian flood control protected zone as an example, the flood risk assessment algorithm and its application are studied. The Cinanfeizuopian flood control protected zone is located in the area surrounded by the left embankment of Xifei River, the right embankment of Cihuaixin River and the Huaibei embankment in Anhui Province, which belongs to the plain area in the middle and lower reaches of the Huaihe River. This flood control protected zone is located at 116°13′22″E-117°12′17″E and 32°38′21″N-33°2′34″N, with an area of about 1 816.0 km², as shown in Fig. 2.

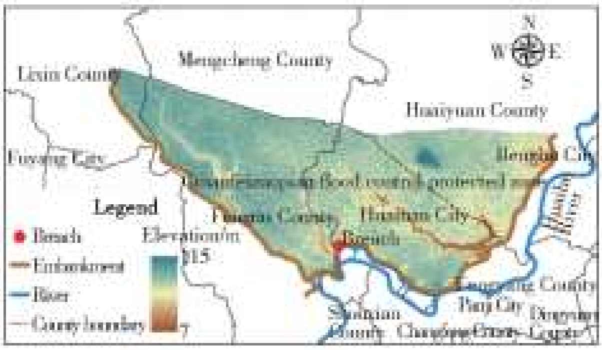


Fig. 2　Geographic position of studied zone

2.2 Establishment and quantization of index

According to the disaster system theory[24], by referring to related studies[25-26], we choose the assessment indexes from three perspectives of hazard of flood disaster-causing factors, sensitivity of disaster-inducing environment and vulnerability of disaster-bearing subjects combining with the actual situation of the protected zone. The importance and available degrees of index data are both taken into account. The index system of flood risk assessment of flood control protected zones is established, which is representative and operative. This index system includes three primary indexes and fourteen second-level indexes, as shown in Tab. 1. The assessment indexes are described specifically in the following.

(1) The index of hazard of disaster-causing factors. The index contains four flood risk factors: the maximum inundation depth, the flood arrival time, the maximum flood flow velocity and the maximum inundation duration, which can be obtained by simulation of hydraulic models. These four indexes are the flood risk factors in the *Flood Risk Mapping Guidelines* published by the Ministry of Water Resources of the People's Republic of China in 2010, which are the direct influencing factors of flood.

(2) The index of sensitivity of disaster-inducing environment. The index contains five indexes

of the normalized terrain niche index, the buffer zone of mainstream embankment, the buffer zone of regional tributaries, the permeation capacity of soil and vegetation coverage. The elevation and vegetation coverage come from the Geospatial Data Cloud (http://www.gscloud.cn/); the slope is extracted from the elevation data based on the ArcGIS platform; the permeation capacity of soil comes from the Harmonized World Soil Database. The terrain niche index can reflect the elevation and slope of the region. With the reduction of the terrain niche index, the regional elevation reduces; as the terrain slope decreases, the flood is easier to pool, and the flood risk is higher. The terrain niche index is

$$T=\log_{10}\left[\left(\frac{H}{H_{\mathrm{mean}}}+1\right)\times\left(\frac{S}{S_{\mathrm{mean}}}+1\right)\right] \tag{19}$$

where: H and H_{mean} are elevation and average elevation of grids; S and S_{mean} are slope and average slope. The indexes at the levels of embankment and river buffer zones adequately reflect the flood risks such as overflow and dam-break flood in the surrounding areas of rivers, lakes and embankments. According to the actual situation in the flood control protected zone, the embankment and river buffer zones are set as five-level buffer zones. The buffer width of the embankment buffer zone is 500 m, 1 km, 2 km, 3 km and >3 km, respectively. The buffer width of the river buffer zone is 250 m, 500 m, 750 m, 1 km and >1 km, respectively. Vegetation and permeable soil have a certain buffering effect against heavy rainfall and flood disasters.

Tab. 1　Index system of flood risk assessment of Cinanfeizuopian flood control protected zone

Objective layer	Criterion layer	Index layer
Flood risk assessment system A	Hazard of flood disaster-causing factors B_1	Maximum inundation depth C_{11}/m
		Flood arrival time C_{12}/h
		Maximum flood flow velocity C_{13}/(m · s^{-1})
		Maximum inundation duration C_{14}/h
	Sensitivity of disaster-inducing environment B_2	Terrain niche index C_{21}
		Buffer zone of mainstream embankment C_{22}
		Buffer zone of regional tributary C_{23}
		Permeation capacity of soil C_{24}
		Vegetation coverage C_{25}
	Vulnerability of disaster-bearing subject B_3	Population density C_{31}/(person · km^{-2})
		GDP density C_{32}/(CNY 10 000 · km^{-2})
		ensity of transportation networks C_{33}/(km · km^{-2})
		Density of fixed assetsC_{34}/(CNY 10 000 · km^{-2})
		Land-use type C_{35}

(3) The index of vulnerability of disaster-bearing subjects. The index is jointly determined by the population, property and socio-economic factors of the disaster-bearing subject. This index includes economic vulnerability indexes and social vulnerability indexes. In this study, the selected social vulnerability indexes include the population density and the land-use type. The economic vulnerability indexes contain the GDP density, the density of transportation networks and the density of fixed assets. As the regional vulnerability indexes increase, the potential threat to the disaster-bearing subject grows higher, and the flood risk enlarges. The data of the population, land-use type and socio-economic indexes come from national statistics, which are reliable and easily accessible.

The grid units have obvious distribution characteristics of carrier information and accurate spatial positions. With the support of the GIS spatial analysis technology, the 150 m×150 m grid is used as the basic unit for the assessment of the Cinanfeizuopian flood control protected zone, and the flood control protected zone is divided into 64 592 grids according to the area of the flood control protected zone, the distribution of indexes, disaster characteristics, sizes of townships, and the topographic feature. The indexes are normalized according to the method stated in Section 1.1. The distribution of some assessment indexes of the Cinanfeizuopian flood control protected zone is shown in Fig. 3. Correlation analysis is conducted for the indexes. The results show that the correlation coefficients between the assessment indexes are all smaller than 0.9, and the interactions between the indexes are within a reasonable range.

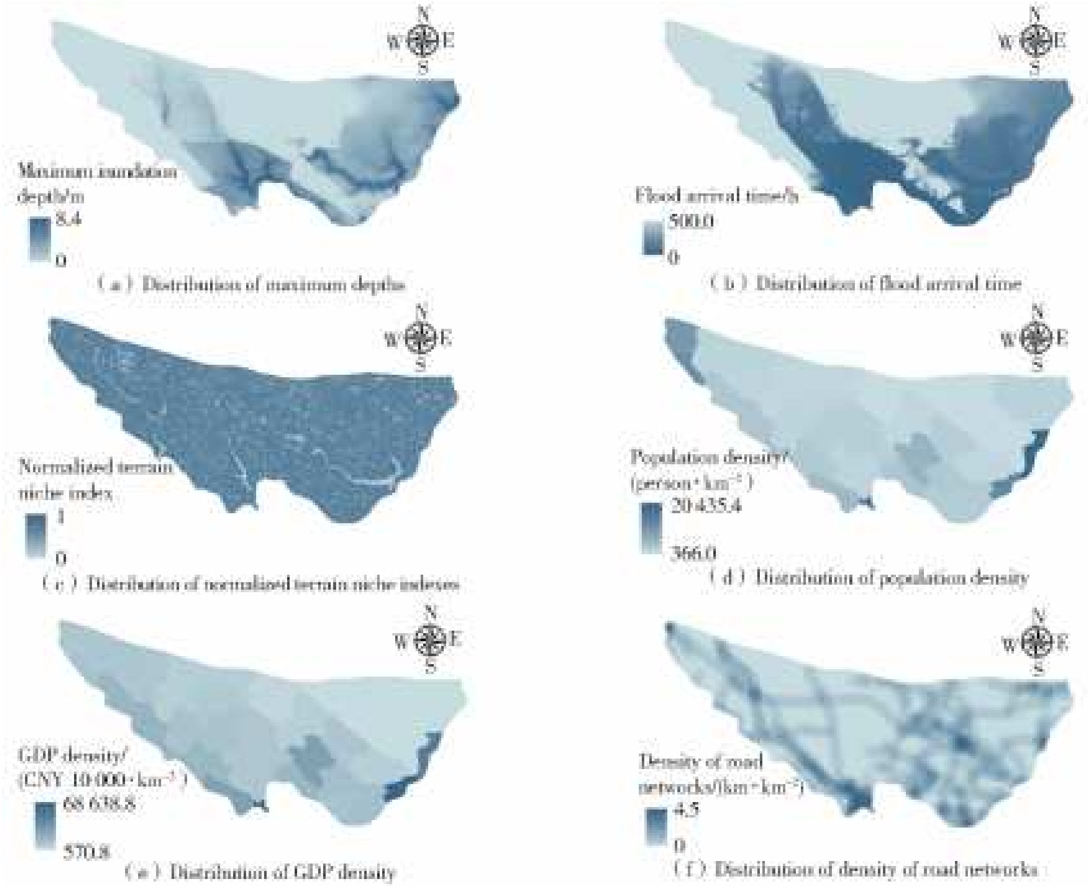


Fig. 3　Spatial distribution of assessment indexes

2.3　Assessment of index calculation

The subjective weight ω_1 is calculated according to the IFAHP method stated in Section 1.2.1, and the objective weight ω_2 is calculated according to the VC-CRITIC method described in Section 1.2.2. The weights obtained through the two methods are different. Based on the obtained subjective weight set and objective weight set, the combination coefficients are calculated by the improved combination weighting method based on game theory. Through normalization, it can be obtained that $\{\alpha_1^*, \alpha_2^*\} = (0.513\,9, 0.486\,1)$. The comprehensive weights obtained through the combination weighting method are $\omega = (0.094\,3, 0.091\,3, 0.074\,9, 0.058\,5, 0.085\,2, 0.076\,7, 0.075\,2, 0.064\,1, 0.072\,1, 0.068\,9, 0.059\,6, 0.060\,1, 0.063\,9, 0.055\,2)^{\mathrm{T}}$, which correspond to the maximum inundation depth, the flood arrival time, the maximum flood flow velocity, the maximum inundation duration, the terrain niche index, the buffer zone of mainstream embankment, the buffer zone of regional tributary, the land type, the vegetation coverage, the population density, the GDP density, the density of transportation networks, the density of fixed assets and the land-use type, as shown in Fig. 4.

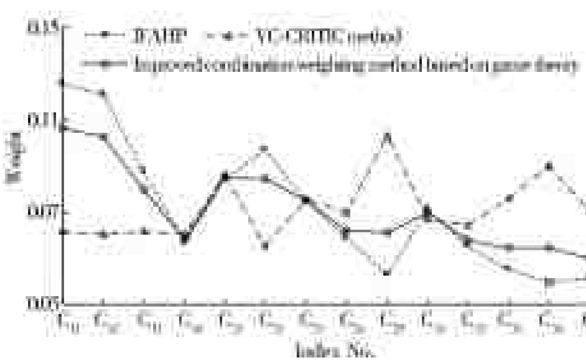


Fig. 4　Change curves of weights of indexes obtained through different methods

2.4 Analysis of assessment results

According to the method stated in Section 1.4, the GMM-based fuzzy clustering analysis is performed on the risk levels of assessment units. The number of fuzzy cluster centers is set as $G \in [2, 10]$. The curve of DBI varying with the number G of cluster centers is shown in Fig. 5.

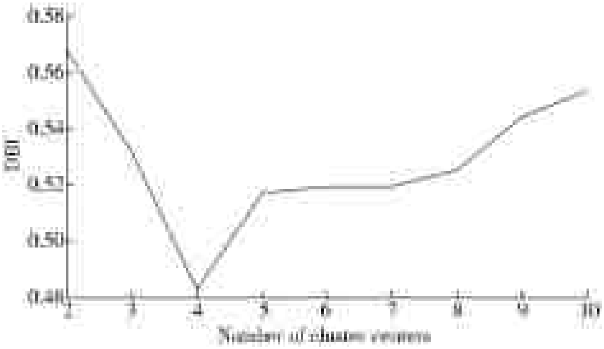


Fig. 5　Relationship between DBI and number of cluster centers G of cluster centers

It can be seen from Fig. 5 that DBI is the minimum when the number of cluster centers is four. Thus, the optimal number G of cluster centers of the assessment units is confirmed as four. Based on the maximum membership principle, the assessment units are divided into four classes according to the flood risk levels: the extremely high-risk area, the high-risk area, the medium-risk area and the low-risk area. To more intuitively display the spatial distribution of flood risks in the flood control protected zone, this paper uses the ArcGIS to make the map, so the distribution of flood risk levels in the Cinanfeizuopian flood control protected zone can be obtained, as shown in Fig. 6.

The regional analysis tool in ArcGIS is used to analyze the distribution characteristics of risk levels in the Cinanfeizuopian flood control protected zone. Thus, the proportions of areas and the distribution situations of different flood risk levels in the studied zone can be obtained.

Fig. 6　Distribution of flood risk level in the Cinanfeizuopian flood protected zone based on improved combination weighting-fuzzy clustering algorithm

(1) The extremely high-risk area accounts for 3.29% of the total area of the protected zone, which is mainly located in the southern area of the Cinanfeizuopian flood control protected zone. The area distributes in the most area of Chengguan Town, the middle and southern area of Guiji Town, the middle and southern areas of Chengbei Township, the southern area of Jiahe Township, the southwestern area of Pingwei Town, the middle area of Tianji Town, the northwestern area of Guqiao Town and other areas. The major reason of these areas belonging to the extremely high-risk area is that they have big maximum inundation depths and small terrain niche indexes. In addition, they are close to the mainstream of the Huaihe River, which are subject to the flood in the mainstream of the Huaihe River. Due to high density of road networks, density of river networks population density, and GDP density, the most area of Chenguan Town has a high flood risk level. The high flood risk level in the middle and southern area of Guiji Town is also influenced by the dense road networks and tributary river networks.

(2) The high-risk area accounts for 21.58%,

which is mainly located in the middle area of Liuji Town, the eastern area of Guiji Town, the northern area of Chenguan Town, the northern area of Jiahe Township, the eastern area of Tianji Township, the northern area of Gugou Hui Ethnic Township, the southern area of Qiji Township, the most area of Pingwei Town, the eastern area of Jiagou Township, the most area of Gaohuang Town, the southern area of Wanfu Town, the most area of Changfen Town, the eastern area of Zhaoying Township and other areas. The main reason of these areas belonging to the high-risk area is that they are located in the neighboring areas of rivers, which have short flood arrival time and high social values.

(3) The medium-risk area and the low-risk area account for 19.20% and 55.93%, respectively, which are mainly located in the northern area of the studied zone. The low flood risk level in these areas is because these areas have high terrains, low densities of river networks, high vegetation coverages, small distributions of population and industrial and mining enterprises, and good permeation of soil.

It can be seen that the flood risk assessment result of the Cinanfeizuopian flood control protected zone is closely related to the threat degree of flood, topography, and socio-economic factors.

In order to compare with the actual flood situations, this paper selects the typical extreme flood disasters in the mainstream of the Huaihe River in 1954 and 1991 as an example. Based on the flood inundation areas and the disaster records, the areas with serious flood disasters, namely, the southern area of Fengtai County and Panji District, are consistent with the extremely high-risk areas shown in Fig. 6. This verifies that the distribution of risk areas obtained by applying the improved combination weighting-fuzzy clustering algorithm to the assessment of flood risk is more reasonable and reliable. The research results can provide a scientific basis for the assessment of flood control and the prevention of flood risk in this zone.

To further verify the reliability of the improved combination weighting-fuzzy clustering algorithm, this paper selects the TOPSIS commonly used in the risk assessment to assess the flood risk of the studied zone. The assessment results are compared with those of the improved combination weighting-fuzzy clustering algorithm. TOPSIS is an assessment method according to the order of the similarities between the evaluated subject and the ideal objective. The flood risk assessment results of the TOPSIS method, which classifies by the natural break method, are shown in Fig. 7. Compared to the assessment results of the improved combination weighting-fuzzy clustering algorithm, the proportions of the recognized extremely high-risk area and high-risk area are very low, which are 0.19% and 2.86%, respectively, showing certain unreasonableness. For example, the most area of Panji District and the river buffer zone, where the floods are serious, are only identified as low-risk areas in Fig. 7. There is a certain deviation between the assessment result and the reality. However, the improved combination weighting-fuzzy clustering algorithm can better identify the roles of the disaster-causing factors, the disaster-inducing environment and the disaster-bearing subject in the disaster system. The assessment results are objective and reliable, which can provide technical support and reference for the assessment of flood risk in similar areas.

3 Conclusions

(1) The index system of flood risk assessment including the hazard of flood disaster-causing factors, sensitivity of disaster-inducing environment and vulnerability of disaster-bearing subject was established. The improved subjective and objective combination weighting algorithm was proposed, which was based on the IFAHP, the VC-CRITIC method and the improved weighting method based on game theory. The optimized calculation of combination weighting of assessment indexes under the unit constraint was realized, so the weighting of assessment indexes was more scientific and reasonable.

(2) The risk levels of all assessment units were clustered by the fuzzy clustering algorithm based on GMM. The optimal number of cluster centers was confirmed according to the analysis of DBI. Based on the maximum membership principle, the flood risk level was assigned to each flood risk

assessment unit according to the size of the flood risk. This compensated for the shortcoming that traditional assessment methods failed to pay enough attention to the fuzziness of assessment intervals, and can provide the classification results of flood risk levels objectively and accurately.

Fig. 7 Distribution of flood risk levels in Cinanfeizuopian flood control protected zone based on TOPSIS

(3) With the 100-year flood of the mainstream of the Huaihe River in the Cinanfeizuopian flood control protected zone as an example, the flood risk assessment model based on improved combination weighting and fuzzy clustering was established, and the flood risk assessment was developed. The assessment results showed that the extremely high-risk area and the high-risk area accounted for 24.87% of the total area of the protected zone, which were basically the areas with big inundation depth, small terrain niche index and high socio-economic values. The assessment results were basically consistent with the actual situation of the zone. The proposed assessment method was reasonable and reliable, which could provide reference for regional flood risk assessment and prevention of disasters.

References:

[1] YUAN X M, WANG Y D, TIAN F CH. Application of multidimensional coupling numerical model for collapse flood[J]. Journal of Tianjin University, 2018, 51(7): 675-683. (in Chinese) DOI: 10.11784/tdxbz201707017.

[2] SHI X J, ZHA X CH, LIU J H, et al. Cloud model-based risk assessment of flood disasters in Ankang City on upper reaches of Hanjiang River[J]. Advances in Science and Technology of Water Resources, 2017, 37(3): 29-34. DOI: 10.3880/j.issn.10067647.2017.03.005.

[3] SHI X J, ZHA X CH, LIU J H, et al. Cloud model-based risk assessment of flood disasters in Ankang City on upper reaches of Hanjiang River[J]. Advances in Science and Technology of Water Resources, 2017, 37(3): 29-34. (in Chinese) DOI: 10.3880/j.issn.10067647.2017.03.005.

[4] YANG S, HAN X, BO C, et al. Cloud-model-based method for risk assessment of mountain torrent disasters[J]. Water, 2018, 10(7): 830. DOI: 10.3390/w10070830.

[5] REN Y F, LIU G D, ZHOU L, et al. Risk evaluation of Chengdu's flood hazard based on evidence theory and variable fuzzy sets theory[J]. Transactions of the Chinese Society of Agricultural Engineering, 2014, 30(21): 147-156. (in Chinese) DOI: 10.3969/j.issn.1002-6819.2014.21.018.

[6] LONGJIAN Z. Using information diffusion to analyze the membership degree in risk assessment of flash flood[J]. Journal of Water Resources Research, 2016, 5(6): 598-604. (in Chinese) DOI: 10.12677/JWRR.2016.56069.

[7] LEE K J, YUN S T, YU S, et al. The combined use of self-organizing map technique and fuzzy c-means clustering to evaluate urban groundwater quality in Seoul Metropolitan city, south Korea[J]. Journal of Hydrology, 2019(569): 685-697. DOI: 10.1016/j.jhydrol.2018.12.031.

[8] EKE S, CLERC G, AKA-NGNUI T, et al. Transformer condition assessment using fuzzy c-means clustering techniques[J]. IEEE Electrical Insulation Magazine, 2019, 35(2): 47-55. DOI: 10.1109/MEI.2019.8636165.

[9] JIANG W J, CHENG CH M, ZHANG Y B, et al. Risk assessment of flood disasters in Zhejiang Province based on GIS/AHP integration method[J]. Bulletin of Surveying and Mapping, 2019, 503(2): 134-139. (in

Chinese) DOI:CNKI:SUN:CHTB. 0. 2019-02-027.

[10] GHOSH A,KAR S K. Application of analytical hierarchy process (AHP) for flood risk assessment: A case study in Malda district of West Bengal,India[J]. Natural Hazards, 2018, 94 (1): 349-368. DOI: 10. 1007/s11069-018-3392-y.

[11] SEEJATA K,YODYING A,WONGTHADAM T,et al. Assessment of flood hazard areas using analytical hierarchy process over the Lower Yom basin, Sukhothai Province[J]. Procedia Engineering,2018,212: 340-347. DOI:10. 1016/j. proeng. 2018. 01. 044.

[12] HONGSHI X,CHAO M,JIJIAN L,et al. Urban flooding risk assessment based on an integrated k-means cluster algorithm and improved entropy weight method in the region of Haikou,China[J]. Journal of Hydrology, 2018, 563: 975-986. DOI: 10. 1016/j. jhydrol. 2018. 06. 060.

[13] WANG X Y,YUAN X M. Comprehensive evaluation of urban flood control and disaster reduction ability based on principal component analysis[J]. Journal of Natural Disasters, 2016, 25 (6): 3-10. (in Chinese) DOI:CNKI:SUN:ZRZH. 0. 2016-06-001.

[14] KOLLI Z, JOUANNIC G, MARCHETTI M, et al. Flood risk vulnerability assessment: hierarchization of the main factors at a regional scale[A]. 3rd European Conference on Flood Risk Management[C]. Lyon, France,2016. DOI:10. 1051/e3sconf/20160708014.

[15] GAO Y Q,WU J J,WANG Z ZH,et al. Regional flood risk assessment based on fuzzy cluster factor analysis [J]. Journal of Natural Disasters,2018,27(2):143-150. (in Chinese) DOI:CNKI:SUN:ZRZH. 0. 2018-02-017.

[16] LU Y,XU L R,CHEN SH Y,et al. Combined weight method based on game theory for debris flow hazard risk assessment[J]. Journal of Catastrophology,2014 (1): 194-200. (in Chinese) DOI: 10. 3969/j. issn. 1000-811X. 2014. 01. 035.

[17] LAI C,CHEN X,CHEN X,et al. A fuzzy comprehensive evaluation model for flood risk based on the combination weight of game theory[J]. Natural Hazards, 2015, 77(2): 1243-1259. DOI: 10. 1007/s11069-015-1645-6.

[18] ZHANG L L,LAI CH G,CHEN X H,et al. Decadal variation and evaluation of flood risk based on the combination weight of game theory[J]. Tropical Geography,2018,38(2):217-225. (in Chinese) DOI:10. 13284/j. cnki. rddl. 003018.

[19] XU Z,LIAO H. Intuitionistic fuzzy analytic hierarchy process[J]. IEEE Transactions on Fuzzy Systems, 2014,22(4):749-761. DOI:10. 1109/TFUZZ. 2013. 2272585.

[20] YIN Q,NIU K,LI N. Using CV-CRITIC to determine weights for smart city evaluation[C]//IEEE International Conference on Tools with Artificial Intelligence. IEEE Computer Society, 2017 (1): 996-1000. DOI:10. 1109/ICTAI. 2017. 00153.

[21] LI A H. Research on safety assessment methods of quayside container crane[D]. Wuhan:Wuhan University of Technology,2017. (in Chinese)

[22] DAVIES D L,BOULDIN D W. A cluster separation measure[J]. IEEE transactions on pattern analysis and machine intelligence,1979,PAMI-1(2):224-227. DOI:10. 1109/TPAMI. 1979. 4766909.

[23] BANDYOPADHYAY S,MAULIK U. Genetic clustering for automatic evolution of clusters and application to image classification[J]. Pattern Recognition,2002,35 (6): 1197-1208. DOI: 10. 1016/s0031-3203 (01) 00108-x.

[24] SHI P J. Theory and practice on disaster system research in a fourth time[J]. Journal of Natural Disasters,2005,14(6):1-7. (in Chinese) DOI:10. 13577/j. jnd. 2005. 0601.

[25] LI SH F,FENG P,SUN SH H. Application of catastrophe theory to evaluation of flood risk in detention basin[J]. Journal of Natural Disasters,2020,19(3): 132-138. (in Chinese) DOI: 10. 13577/j. jnd. 2020. 0322.

[26] WU Z N,SHEN Y X,WANG H L. Vulnerability assessment of flood disasters based on emergy theory [J]. South-to-North Water Transfers and Water Science & Technology,2018,16(6):13-18,36. (in Chinese) DOI:10. 13476/j. cnki. nsbdqk. 2018. 0146.