

基于 POT 与 P-IOISM 的洪水风险信息挖掘*

郑炎辉^{1,2}, 何艳虎^{1,2}, 李深林^{1,2}, 陈晓宏^{1,2**}

(1: 中山大学水资源与环境研究中心, 广州 510275)

(2: 华南地区水循环和水安全广东普通高校重点实验室, 广州 510275)

摘要: 为了从有限的实测资料中挖掘更多洪水信息, 以武江流域为研究对象, 在考证历史洪水进行特大值处理基础上, 基于犁市(二)水文站 1956—2009 年实测日流量数据提取了 POT 样本, 构建基于 P-III 模型确定控制点而改进的非等步长内集-外集模型(P-IOISM)进行洪水风险信息挖掘。结果表明: 结合 POT 和 P-IOISM 模型的洪水风险评估能够从有限的实测资料中获取更多的洪水风险信息, 得到更贴近事实的风险评估结果, 能为决策者提供更多的依据, 从而使决策结果更加可靠实用。

关键词: 洪水风险; 信息挖掘; POT; P-III 分布; 信息分配; P-IOISM; 武江流域

The flood risk data mining based on POT and P-IOISM

ZHENG YanHui^{1,2}, HE Yanhu^{1,2}, LI Shenlin^{1,2} & CHEN Xiaohong^{1,2**}

(1: *Center for Water Resources and Environment, Sun Yat-sen University, Guangzhou 510275, P.R.China*)

(2: *Key Laboratory of Water Cycle and Water Security in Southern China of Guangdong Higher Education Institutes, Sun Yat-sen University, Guangzhou 510275, P.R.China*)

Abstract: In order to get more information about the flood risk, the Peak-over-threshold (POT) flood samples are extracted after extraordinary flood processing with historical information based on daily flow data of Lishi(2) station in the Wujiang River Basin from 1956–2009. Then the improved non-isochronous Interior-Outer-Set Model (IOISM) which uses P-III to determine the control point is applied to analyse the flood risk. The results show that: we can get more flood risk information from limited data and flood risk assessment result that is closer to fact by combining POT and P-IOISM models. It will provide more evidence for policy-makers which makes the decision-making more reliable and practical.

Keywords: Flood risk; datamining; POT; P-III distribution; information diffusion; P-IOISM; Wujiang River Basin

我国大多数水文监测站点的洪水观测数据时间序列较短, 观测系列通常只有几十年^[1], 同时洪水灾害风险具有随机、模糊、灰色等诸多的不确定性^[2]。在现有条件下洪水灾害的风险是估不准的, 传统概率计算方法(如概率论、数理统计等理论与方法)建立在大量连续时间序列统计资料基础上, 而实际中会常遇到“小样本”的不完备信息问题, 存在不精确和不确定的理论瓶颈^[3-4]。因此, 为了提高洪水灾害的风险分析结果的可靠性和准确性, 从有限的实测资料中挖掘更多洪水信息的研究显得尤为重要。

近年来, 许多研究者都通过 POT 抽样法扩大了洪水样本, 在此基础上进行洪水次数、量级、风险等研究, 弥补改善了传统 AMS 抽样样本较少的这一缺陷^[5-9]。与传统年最大值(Annual Maximum Series, AMS)抽样相比, 超定量(Peak-over-Threshold, POT)抽样扩大了样本容量, 不仅反映洪水量级, 还能反映洪水发生过程, 相比较而言, 更具物理相关性, 能更有效利用实测资料描述洪水特征, 能更充分利用洪水信息^[10-12]。

自从黄崇福等^[13-16]系统提出信息扩散、内集-外集模型(IOISM)等多种模糊风险评价方法用于洪水灾害风险分析、地震灾害风险分析等领域后, 邹强^[17]、Liu 等^[18]、Feng 等^[19]将 IOISM 模型应用于洪水灾害、火灾、

* 国家自然科学基金项目(51210013, 51479216, 51509127)和国家科技支撑计划项目(2012BAC21B0103)联合资助。
2016-09-19 收稿; 2016-11-12 收修改稿。郑炎辉(1990~), 男, 博士研究生; E-mail: zhengyh9@mail2.sysu.edu.cn.

** 通信作者; E-mail: eesexh@mail.sysu.edu.cn.

台风灾害等领域,在有限的观测样本中进行了较为完备、可靠地灾害风险分析.使用基于信息分配原理建立的内集-外集模型,可以针对信息不足而导致风险值估不准的缺陷,将样本进行模糊处理来充分利用样本的模糊过渡信息,使具有单值观测值的样本点转化为具有模糊集值的样本集,从而尽可能地填补小样本信息存在的不完备空隙,通过有限的知识来试图挖掘信息的内在联系和发生规律.通过这个途径,将所有样本的单值信息扩散到设定好的控制域中所有点上,得到较为丰富的信息资料^[20],可在保证期望值合理可靠的前提下体现风险估计的不精确性、不完备性,为快速而合理的风险判断、及时而有效的风险规避提供科学依据^[13-15].

在模糊处理技术领域的信息分配中,存在一个瓶颈问题——控制点的选取.传统内集-外集模型控制点的选取通常采用 Otness 等^[21]提出的控制点选择公式,但是该公式的前提是正态总体和样本量较大时,因此在选取内集-外集模型的控制点时,直接引用会出现一些问题,特别是信息较为集中且存在特大值的时候,部分控制区间会没有样本点.因而本文将对其进行改进,根据皮尔逊 III 型曲线的频率分析结果确定控制区间,将原来的等步长 IOSM 模型改进为根据皮尔逊 III 型曲线的频率分析结果确定控制区间的非等步长 P-IOSM 模型.

综上所述,本研究在 POT 抽样基础上,以南方湿润区武江流域为例,将基于信息分配原理改进的非等步长 P-IOSM 模型应用到洪灾风险分析中,从有限的实测资料中挖掘尽可能多的洪水信息,在现有信息不完备条件下得到了可能性概率风险值,可为决策者提供多层次、多值化的洪水风险信息.

1 研究区与数据来源

研究区域为武江流域(24°50′~25°31′N,112°50′~113°35′E),面积 7097 km²,干流全长 260 km,为珠江流域北江水系的一级支流,发源于湖南省临武县三峰岭,于韶关市汇入北江.

洪水超定量序列和年最大值序列的提取采用武江下游重要控制站犁市(二)水文站 1956—2009 年的日流量资料(其中“2006·07”洪水作特大洪水处理)^[6].

2 研究方法

2.1 POT 样本提取

POT 样本提取有两个关键,即洪峰样本独立性判别和超定量序列门限值的确定^[12].

2.1.1 洪峰样本独立性判别标准 进行洪水超定量频率分析的前提是超定量洪峰样本具有独立性.本文结合美国水资源协会(USWRC)提出的判别标准与王善序等提出的超定量抽样原则^[6-7],进行洪峰独立性判别.

同时选取两个连续洪峰的条件为^[12]:

$$\theta > 5 + \ln(A) \text{ 且 } X_{\min} < 0.75 \min[Q_1, Q_2] \quad (1)$$

式中, θ 为两个峰间的间隔时间(d); A 为流域面积(km²); Q_i 为第*i*场洪水的最大日流量.

不满足上述条件的连续洪峰中,只取其中最大一次洪峰^[1].

2.1.2 超定量序列门限值确定 超定量序列门限值 S 是确定超定量序列样本的重要参数.目前普遍认为,门限值 S 的选择并不唯一,关键在于使提取的超定量样本服从某些标准.门限值应根据超定量系列发生次数分布、超定量洪水频率分布以及独立同分布假设共同确定^[12].

(1)超定量样本均值法: Davison 等^[22]认为门限值应在以下范围内选取:超定量样本超过部分的均值($\bar{X}_s - S$)是门限值 S 的线性函数,其中 \bar{X}_s 是超定量样本均值.该方法是通过门限值的确定使 POT 模型的广义 Pareto 分布参数估计稳定性最高^[12].

(2)分散指数法: Ashkar 等^[23]认为门限值的选择应使样本分散指数在一个置信区间内,以确保超定量发生次数服从 Poisson 分布.

$$I = \text{var}(m) / \text{mean}(m) \quad (2)$$

$$h = \frac{\sum_{i=1}^{NY} [m(i) - \bar{m}]^2}{\bar{m}} = I(NY - 1) \quad (3)$$

式中, m 为年超定量发生次数序列, $m(i)$ 为第 i 年的超定量发生次数; NY 为实测资料的年数; I 为超定量序列发生次数的分散指数; h 服从自由度为 $(NY-1)$ 的卡方分布.

泊松过程的分散指数为 1, 取置信区间为 $[5\%, 95\%]$, 则若 $\chi^2(5\%)/(NY-1) < I < \chi^2(95\%)/(NY-1)$, 超定量次数服从泊松分布; 其中 $\chi^2(5\%)$ 的自由度为 $(NY-1)$.

(3) 年均超定量发生次数 n 法: $n > 1.65$ 时采用指数分布做 POT 模型能取得较好的效果^[24], n 过大易影响样本独立性; 国内关于洪水超定量频率分析多控制在 2~3 次^[5].

本文在满足条件(1)和(2)的门限值范围内, 选择满足 $n > 2$ 的较大门限值.

2.2 P-III 型频率曲线

我国一般采用 P-III 型频率曲线来计算设计洪水^[25], 分布参数采用适线法估计, 具体计算公式参考相关文献^[26-27].

2.3 改进的内集-外集模型 (P-IOISM)

2.3.1 内集-外集模型 (IOISM) 设洪水的观测样本为:

$$X = \{x_1, x_2, \dots, x_n\} \tag{4}$$

设 U 是样本 X 的论域, u_1, u_2, \dots, u_m 是给定步长为 Δ 的离散点, 即 Δ 为控制点步长, $\Delta = u_j - u_{j-1}, j = 2, 3, \dots, m$.

$$U = \{u_1, u_2, \dots, u_m\} \tag{5}$$

IOISM 用来计算不利事件发生在区间 I_j 的模糊概率, 这里:

$$I_j = \left[u_j - \frac{\Delta}{2}, u_j + \frac{\Delta}{2} \right], j = 1, 2, \dots, m \tag{6}$$

显然控制点 u_j 是 I_j 的中间点, 且对于任意样本点 x_i , 均只落入一个区间.

根据样本点分布的具体情况和 IOISM 的计算要求, 区间总数 m 的确定方法为^[21]:

$$m = 1.87(n - 1)^{0.4} \tag{7}$$

当样本点 x_i 受到随机扰动时, 样本点会与 I_j 发生位置的变化, 可能会离开或进入 I_j . 这里将离开定义为游离, 将进入定义为漂入, 对样本点 x_i 而言, 游离或漂入区间 I_j 的可能性分别记为 q_{ij}^- 和 q_{ij}^+ .

为了方便地计算, 首先给出样本集合 X 关于区间 I_j 的内集和外集定义:

(1) 内集 $X_{in-j}: X_{in-j} \triangleq X \cap I_j$, 由所有在区间 I_j 内元素构成.

(2) 外集 $X_{out-j}: X_{out-j} \triangleq X \setminus X_{in-j}$, 由所有不在区间 I_j 内元素构成. 在此基础上, 内指标集和外指标集定义如下:

如下:

(1) 内指标集 S_j : 如果 $\forall s \in S_j$, 有 $x_s \in X_{in-j}$, 且 $\{x_s | s \in S_j\} = X_{in-j}$.

(2) 外指标集 T_j : 如果 $\forall t \in T_j$, 有 $x_t \in X_{out-j}$, 且 $\{x_t | t \in T_j\} = X_{out-j}$.

如果 S_j 或 X_{in-j} 的容量为 n_j , 记为 $|S_j| = n_j$, 则 $|T_j| = n - n_j$. 应用信息分配公式:

$$q_{ij} = \begin{cases} 1 - \frac{|x_i - u_j|}{\Delta}, & |x_i - u_j| \leq \Delta \\ 0, & |x_i - u_j| \geq \Delta \end{cases} \quad x_i \in X, u_j \in U \tag{8}$$

样本点 x_i 游离或漂入区间 I_j 的可能性 q_{ij}^- 和 q_{ij}^+ 的计算公式分别如下:

$$q_{ij}^- = \begin{cases} 1 - q_{ij}, & x_i \in X_{in-j} \\ 0, & x_i \in X_{out-j} \end{cases} \tag{9}$$

$$q_{ij}^+ = \begin{cases} q_{ij}, & x_i \in X_{out-j} \\ 0, & x_i \in X_{in-j} \end{cases} \tag{10}$$

显然, 对于 I_j 内集中的样本点 x_i , 如果其 q_{ij} 值越小, 则其离开所在区间 I_j 并漂移到邻近区间的可能性越大, 反之亦然. 同样, 对于 I_j 外集中的样本点 x_i , 如果其 q_{ij} 值越小, 则其落入区间 I_j 的可能性越大, 反之亦然. 特别地, 当 x_i 是 I_j 内集中的点时, 称 x_i 已漂入 I_j , 其漂入的可能性定义为 0; 而当 x_i 是 I_j 外集中的点时, 称 x_i 已游离 I_j , 其游离的可能性定义为 0. 在区间论域 $I = \{I_1, I_2, \dots, I_m\}$ 和离散概率论域 $P = \{P_1, P_2, \dots, P_n\} = \{k/n | k=0, 1, 2, \dots, n\}$ 上, 当 $|S_j| = n_j$ 时, 根据式(10)的内集-外集模型, 可计算出一个关于随机不利事件

发生在区间 I_j 中的可能性-概率分布 (possibility-probability distribution, PPD):

$$\pi_{I_j}(p) = \begin{cases} \bigwedge_{s \in S} q_{sj}^-, p = p_0 \\ \dots\dots \\ \bigvee_{S_1, S_2, S_3 \in S, S_1 \neq S_2 \neq S_3} (q_{s_1j}^- \wedge q_{s_2j}^- \wedge q_{s_3j}^-), p = p_{n_j-3} \\ \bigvee_{S_1, S_2 \in S, S_1 \neq S_2} (q_{s_1j}^- \wedge q_{s_2j}^-), p = p_{n_j-2} \\ \bigvee_{s \in S} q_{sj}^-, p = p_{n_j-1} \\ 1, p = p_{n_j} \\ \bigvee_{t \in T} q_{tj}^+, p = p_{n_j+1} \\ \bigvee_{t_1, t_2 \in T, t_1 \neq t_2} (q_{t_1j}^+ \wedge q_{t_2j}^+), p = p_{n_j+2} \\ \bigvee_{t_1, t_2, t_3 \in T, t_1 \neq t_2 \neq t_3} (q_{t_1j}^+ \wedge q_{t_2j}^+ \wedge q_{t_3j}^+), p = p_{n_j+3} \\ \dots\dots \\ \bigwedge_{t \in T} q_{tj}^+, p = p_n \end{cases} \quad (11)$$

式中, $\pi_{I_j}(p)$ 表示不利事件属于区间 I_j 的概率为 p 的可能性, 可以看作是区间 I_j 对概率 p 的隶属度.

根据给定的样本集合 X , 可以得到在 $\{I_1, I_2, \dots, I_m\} \times \{p_0, p_1, \dots, p_n\}$ 下的 PPD:

$$\Pi = (\pi_{I_j}(p_i))_{m \times (n+1)} = \begin{pmatrix} \pi_{I_1}(p_0) & \pi_{I_1}(p_1) & \dots & \pi_{I_1}(p_n) \\ \pi_{I_2}(p_0) & \pi_{I_2}(p_1) & \dots & \pi_{I_2}(p_n) \\ \dots & \dots & \dots & \dots \\ \pi_{I_m}(p_0) & \pi_{I_m}(p_1) & \dots & \pi_{I_m}(p_n) \end{pmatrix} \quad (12)$$

2.3.2 改进的内集-外集模型 (P-IOSM) 当采用式(7)确定控制区间, 存在特大值的时候会出现部分区间可能性概率为 0, 导致流量较小区间可能性估计值小于流量较大区间的异常现象. 因而对其进行改进, 基于皮尔逊 III 型能较好拟合我国大多数河流的洪水系列, 因而将原来的等步长 IOSM 模型改进为根据皮尔逊 III 型曲线的频率分析结果确定控制区间的不等步长 P-IOSM 模型.

设 I_j 是控制区间, Δ_j 是控制区间步长, 由频率分析得到各控制区间最大值 $I_{jP_{max}}$ 和最小值 $I_{jP_{min}}$, 则可得各控制区间范围及控制区间步长:

$$I_j = [I_{jP_{min}}, I_{jP_{max}}], j = 1, 2, \dots, m \quad (13)$$

$$\Delta_j = I_{jP_{max}} - I_{jP_{min}}, j = 1, 2, \dots, m \quad (14)$$

同时, 式(8)改为:

$$q_{ij} = \begin{cases} 1 - 2 \frac{x_i - u_j}{\Delta_j + \Delta_{j+1}}, 0 \leq x_i - u_j \leq \Delta_j/2 + \Delta_{j+1}/2 \\ 1 - 2 \frac{u_j - x_i}{\Delta_{j-1} + \Delta_j}, 0 \leq u_j - x_i \leq \Delta_{j-1}/2 + \Delta_j/2 & x_i \in X, u_j \in U \\ 0, x_i - u_j \geq \Delta_j/2 + \Delta_{j+1}/2 \text{ 或 } u_j - x_i \geq \Delta_{j-1}/2 + \Delta_j/2 \end{cases} \quad (15)$$

2.4 基于 IOSM 模型的风险估计

由式(8)可知, 在一个概率-可能性分布里, 区间数量为 m , 可能性数量为 $n+1$. 可以这样认为, 任意一个区间 I_j , 对应有 $n+1$ 个可能性, 即 $\pi_{I_j}(p_i)$ ($i = 0, 1, \dots, n$), 相应的可能性可以看作是区间 I_j 对概率 p_i 的隶属度. 为了能够尽可能多地为风险管理提供决策消息, 这里用控制点 u_j 来代表 I_j , 可以根据式(12)~(15)计算出区间 I_j 上可能性的权重 ω_i 和期望值 $E[u_j]$:

$$E[u_j] = \sum_{i=0}^n \omega_i \cdot p_i \quad (16)$$

式中, ω_i ($i = 0, 1, \dots, n$) 根据式(17)~(19)得到^[28]:

$$\omega_0 = \frac{1}{2} (\mu_0 + \max_{0 \leq r \leq n} \mu_r - \max_{0 < r \leq n} \mu_r) \quad (17)$$

$$\omega_i = \frac{1}{2}(\max_{0 \leq r \leq i} \mu_r - \max_{0 \leq r < i} \mu_r + \max_{i \leq r \leq n} \mu_r - \max_{i < r \leq n} \mu_r), 1 \leq i \leq n - 1 \quad (18)$$

$$\omega_n = \frac{1}{2}(\max_{0 \leq r \leq n} \mu_r - \max_{0 \leq r < n} \mu_r + \mu_n) \quad (19)$$

式中, $\mu_r = \pi_{ij}(p_r), r=0, 1, \dots, n$. 显然, $\omega_i \geq 0$ 且 $\sum_{i=0}^n \omega_i = 1$.

这样, 根据式(16)就得到了各个区间的期望值. 在此基础上, 计算得到不利事件发生在区间 I_j 的概率:

$$p(u_j) = E[u_j] / \sum_{j=1}^m E[u_j], j = 1, 2, \dots, m \quad (20)$$

那么风险估计值即为可能性估计值的累积:

$$P(u_j) = p(u \geq u_j) = \sum_{k=j}^n p(u_k) \quad (21)$$

3 结果与讨论

3.1 超定量门限值提取

根据公式(1), 符合独立性判别标准的洪峰之间的时间间隔约为 14 天. 对梨市(二)站 54 年日流量资料进行洪峰独立性判别, 提取独立洪峰超定量样本. 在提取过程中, 部分洪峰虽然流量较大, 但因洪峰之间时间间隔小于 14 天或者洪峰之间最小流量过大, 不满足独立性判别标准而未入选.

根据所选独立洪峰, 综合考虑超定量样本均值法、分散指数法和年均超定量发生次数, 通过试算, 确定最大门限值取 $S = 1079 \text{ m}^3/\text{s}$, 年均超定量发生次数为 2.4, 样本数目为 130 个.

具体确定过程如下:

a. 对门限值进行试算, 由图 1 超定量样本超过部分均值 ($\bar{X}_s - S$) 与门限值 S 的关系曲线可知, 当门限值在区间 $[419, 1079] \text{ m}^3/\text{s}$ 时, ($\bar{X}_s - S$) 与 S 呈稳定的线性关系.

b. 计算超定量序列发生次数的分散指数, 绘制分散指数图, 由图 2 可知 $S > 324 \text{ m}^3/\text{s}$ 时分散指数位于 $[5\%, 95\%]$ 的置信区间内, 即年超定量个数服从泊松分布.

c. 根据 a 和 b 可知在 $324 \sim 1079 \text{ m}^3/\text{s}$ 范围内选择满足 $n > 2$ 的门限值, 选择最大门限值以加强样本独立性, 取 $S = 1079 \text{ m}^3/\text{s}, n = 2.4$. 由此提取 130 个洪峰构成 POT 样本.

3.2 P-III型频率分析

根据 POT 取样获取 130 个独立洪峰样本, 其中 2006 年 7 月洪水作特大洪水处理, 对 POT 序列进行 P-III 型频率分析, 拟合效果优良, 相关系数达到 0.98 (图 3).

各设计频率洪峰流量见表 1.

表 1 梨市(二)站不同设计频率对应洪峰流量

Tab.1 Corresponding peak discharge of different design frequencies in Lishi(2) station

设计频率	1%	2%	5%	10%	20%	50%
洪峰流量 (m^3/s)	5590	4812	3807	3073	2376	1567

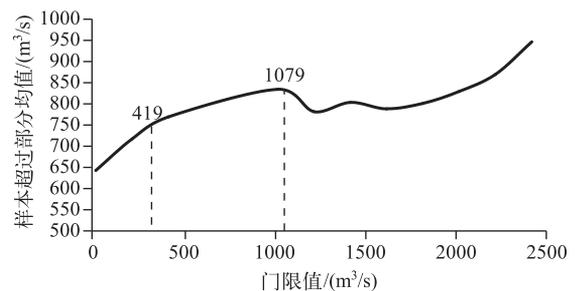


图 1 超定量样本超过部分均值与门限值的关系曲线

Fig.1 The relation curve between the values that average of the peak over threshold minus the corresponding threshold and the corresponding thresholds

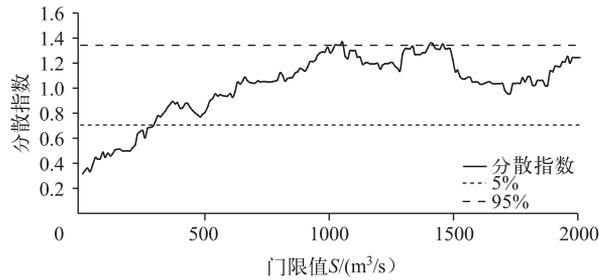


图 2 分散指数图

Fig.2 Figure of dispersion index

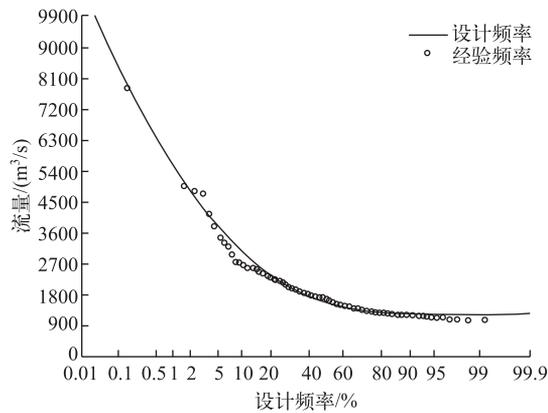


图 3 犁市(二)站 POT 洪水频率曲线

Fig.3 POT flood frequency curve of Lishi(2) station

3.3 基于 P-IOISM 的可能性-概率分布

如果采用式(7)确定控制区间,则区间数目为 14,根据 IOISM 模型计算 [7500,8000] 区间最大可能性概率为 0.008,而区间 [7000,7500]、[6500,7000]、[6000,6500]、[5500,6000] 可能性概率都为 0,这与现实不符,原因是观测样本有限,观测样本并未能覆盖所有区间,因而导致了部分区间可能性概率都为 0 的情况.因而皮尔逊 III 型曲线的频率分析结果确定控制区间,根据不同设计频率洪峰流量划分控制区间,得到控制区间与相应控制中心值(表 2).

表 2 控制区间与相应控制中心值(m³/s)

Tab.2 The intervals and the corresponding discrete domains

	灾害等级						
	七级	六级	五级	四级	三级	二级	一级
控制区间	1000<x<1567	1567<x<2376	2376<x<3073	3073<x<3807	3807<x<4812	4812<x<5590	5590<x<8000
控制中心	1283.5	1971.5	2724.5	3440	4309.5	5201	6795

根据一维线性信息分配函数即公式(8)计算扩散信息矩阵 q_{ij} , 然后根据公式(9)和公式(10)分别计算得到游离信息矩阵 q_{ij}^- 和漂入信息矩阵 q_{ij}^+ ; 在此基础上,根据公式(11)和公式(12)可得可能性-概率分布矩阵(PPD)(表 3 和图 4).

由表 3 和图 4 可以看出灾害等级越高的洪水,其最大可能性(可能性为 1)对应的概率值越小,与灾害等级越高的洪水发生的概率越小这一现实相符.

表 3 基于 P-IOSM 的可能性-概率分布
Tab.3 Probability-possibility distribution based on P-IOSM

概率	七级	六级	五级	四级	三级	二级	一级
0	0.012	0.004	0.013	0.191	0.298	0.517	0.884
0.008	0.012	0.028	0.099	0.327	0.897	0.877	1.000
0.015	0.023	0.053	0.159	0.518	0.914	1.000	0.510
0.023	0.023	0.103	0.159	1.000	1.000	0.866	0.394
0.031	0.023	0.103	0.185	0.883	0.904	0	0.311
0.038	0.023	0.120	0.329	0.801	0.626	0	0
0.046	0.048	0.120	0.357	0.202	0.409	0	0
0.054	0.058	0.120	0.386	0.202	0.031	0	0
0.062	0.058	0.169	0.501	0.038	0	0	0
0.069	0.083	0.177	0.558	0.038	0	0	0
0.077	0.083	0.194	0.730	0	0	0	0
0.085	0.083	0.201	0.791	0	0	0	0
0.092	0.083	0.201	0.816	0	0	0	0
1	0	0	0	0	0	0	0

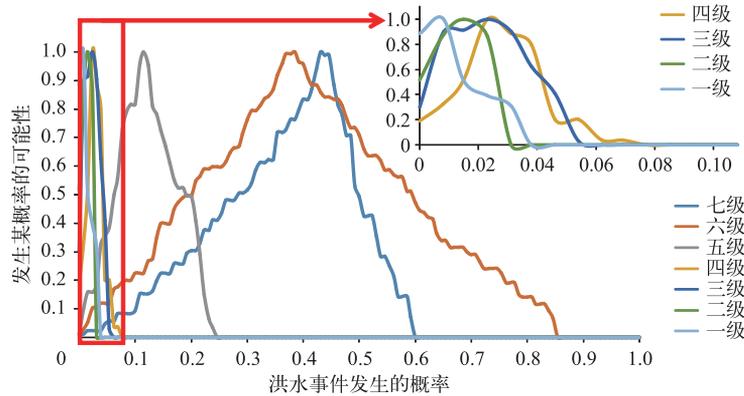


图 4 基于 P-IOSM 的可能性-概率分布

Fig.4 Figure of probability-possibility distribution based on P-IOSM

根据公式(16)~(21)来计算得到不同等级灾害概率估计值 $p(u_j)$ 、风险估计值 $P(u_j)$,从 PPD 可知,从不利事件样本 X 出发,不能确定不利事件发生的概率值,而是一个模糊集,即一个不利灾害事件对应着几个概率值,只是可能程度不同而已.它充分体现了在现有条件下自然灾害风险估不准这一特性,与实际情况更加接近.

例如对控制点 u_4 而言,虽然洪峰流量落在以 $u_4 = 3440$ 为控制中心的区间 $3073 < x < 3807$ 内以 $p = 0.023$ 的可能性最大(为 1),但是也不能忽视其他概率发生的可能性,即便是对于从不发生的 $p = 0$,仍然也有 0.191 的可能性.

不同级别灾害的可能性-概率分布具有相似的规律,只是其最大可能性对应的概率不同.

表 4 基于 P-ISOM 的风险评估值结果

Tab.4 Risk value estimation based on P-IOSM

灾害等级	区间	风险水平	概率估计值	风险估计值
七级	[1000,1567]	$u_1 = 1283.5$	0.396	1
六级	[1567,2376]	$u_2 = 1971.5$	0.408	0.604
五级	[2376,3073]	$u_3 = 2724.5$	0.122	0.196
四级	[3073,3807]	$u_4 = 3440.0$	0.028	0.073
三级	[3807,4812]	$u_5 = 4309.5$	0.023	0.046
二级	[4812,5590]	$u_6 = 5201.0$	0.014	0.023
一级	[5590,8000]	$u_7 = 6795.0$	0.009	0.009

如果用传统直方图计算频率,用以代替概率,可计算出七级、六级、五级、四级、三级、二级、一级的概率依次为 0.431、0.385、0.115、0.023、0.023、0.015 和 0.008,与基于 P-IOSM 计算的各等级灾害最大可能性一致,但若仅使用传统直方图计算灾害发生概率,则会出现三级灾害和四级灾害出现概率相等的结论,实际上三级灾害出现概率应小于四级灾害,这在 PPD 中有所体现,三级灾害发生概率为 0.015 有 0.914 可能性,四级灾害发生概率为 0.015 只有 0.518 可能性,说明三级灾害比四级灾害更大可能性概率为 0.015,即三级灾害比四级灾害发生概率小的可能性更大,同时从表 4 可知三级灾害出现概率为 0.023,小于四级灾害的出现概率 0.028,与现实相符。

可能性-概率分布本身保留了许多不确定信息,这些附加信息更贴近事实,能为决策者提供更多的依据,从而使决策结果更加可靠实用。

4 结论与展望

基于武江流域犁市(二)水文站 1956—2009 年实测日流量数据提取了 POT 样本,在考证历史洪水进行特大值处理的基础上,构建基于 P-III 模型确定控制点而改进的非等步长内集-外集模型(P-IOSM)进行洪水模糊风险分析。主要结论如下:

1) 与年最大值模型相比,采用 POT 模型进行洪水样本选取,能充分利用实测洪水信息,增大样本容量,降低抽样误差,从有限的观测数据中获取更多洪水样本信息。

2) 基于 P-III 模型确定控制点,能够使洪水信息更为合理地在研究区域上分配,解决了洪水信息集中且存在较大值时,无法合理分配区间的问题。

3) 结合 POT 与 P-IOSM 的洪水风险分析得到的可能性-概率分布,挖掘了更多的不确定信息,这些附加信息更贴近事实,能为决策者提供更多的依据,从而使决策结果更加可靠实用。

在将来的研究中,如果能获取流域内更多站点数据资料时,将可以制定出整个流域的洪水风险图,从而为决策者提供更直观更强有力的支持,具有良好的工程应用情景。

5 参考文献

- [1] Wang Shanxu. The flood frequency analysis based on POT flood series. *Yangtze River*, 1999, **30**(8): 23-25. [王善序. 洪水超定量系列频率分析. 人民长江, 1999, **30**(8): 23-25.]
- [2] Huang Chongfu. Fuzzy risk assessment of urban natural hazards. *Fuzzy Sets & Systems*, 1996, **83**(2): 271-282.
- [3] Zou Qiang. Study on the theory and method of comprehensive analysis and intelligent assessment of flood disaster risk [Dissertation]. Wuhan: Huazhong University of Science & Technology, 2013. [邹强. 洪灾风险综合分析与智能评价的理论与方法[学位论文]. 武汉: 华中科技大学, 2013.]
- [4] Huang Qiang, Chen Zishen. Multivariate flood risk assessment based on the secondary return period. *J Lake Sci*, 2015, **27**(2): 352-360. DOI: 10.18307/2015.0221. [黄强, 陈子森. 基于二次重现期的多变量洪水风险评估. 湖泊科学, 2015, **27**(2): 352-360.]
- [5] Fang Bin, Guo Shenglian, Cai Xiaoling et al. Flood frequency analysis based on FPOT method. *Water Power*, 2005, **31**(2): 9-12. [方彬, 郭生练, 柴晓玲等. FPOT 方法在洪水频率分析中的应用研究. 水力发电, 2005, **31**(2): 9-12.]
- [6] Zhang Lijuan, Chen Xiaohong, Ye Changqing et al. POT flood frequency analysis with historical floods in Wujiang River. *Journal of Hydraulic Engineering*, 2013, (3): 268-275. [张丽娟, 陈晓宏, 叶长青等. 考虑历史洪水的武江超定量洪水频率分析. 水利学报, 2013, (3): 268-275.]
- [7] Ye Changqing, Chen Xiaohong, Zhang Lijuan et al. Changing rule and influence of threshold for peak-over-threshold flood series under the changing environment in Wujiang, China. *Advances in Water Science*, 2013, (3): 392-401. [叶长青, 陈晓宏, 张丽娟等. 变化环境下武江超定量洪水门限值响应规律及影响. 水科学进展, 2013, (3): 392-401.]
- [8] Martins ES, Stedinger JR. Generalized Maximum likelihood pareto—Poisson estimators for partial duration series. *Water Resources Research*, 2001, **37**(10): 2551-2557.
- [9] Parent E, Bernier J. Bayesian POT modeling for historical data. *Journal of Hydrology*, 2003, **274**(1/2/3/4): 95-108.
- [10] Dai Changjun, Liang Zhongmin, Luan Chengmei et al. Advance in flood frequency analysis for partial duration series. *Advances in Water Science*, 2006, **17**(1): 136-140. [戴昌军, 梁忠民, 栾承梅等. 洪水频率分析中 PDS 模型研究进展.]

- 水科学进展, 2006, **17**(1): 136-140.]
- [11] Beguería S. Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value. *Journal of Hydrology*, 2005, **303**(1/2/3/4): 215-230.
- [12] Lang M, Ouarda TBMJ, Bobée B. Towards operational guidelines for over-threshold modeling. *Journal of Hydrology*, 1999, **225**(3/4): 103-117.
- [13] Huang Chongfu ed. The natural disaster risk assessment: Theory and practice. Beijing: Science Press, 2006. [黄崇福. 自然灾害风险评价: 理论与实践. 北京: 科学出版社, 2006.]
- [14] Huang Chongfu. An application of calculated fuzzy risk. *Information Sciences—Informatics & Computer Science an International Journal*, 2002, **142**(142): 37-56.
- [15] Huang Chongfu, Moraga C, Yuan X eds. Calculation vs. Subjective Assessment with Respect to Fuzzy Probability//Computational Intelligence. Theory and Applications. Berlin Heidelberg: Springer, 2001: 392-411.
- [16] Huang Chongfu ed. The natural disaster risk analysis and management. Beijing: Science Press, 2012. [黄崇福. 自然灾害风险分析与管理. 北京: 科学出版社, 2012.]
- [17] Zou Qiang, Zhou Jianzhong, Zhou Chao *et al.* Flood disaster risk analysis based on variable fuzzy sets theory. *Transactions of the Chinese Society of Agricultural Engineering*, 2012, **28**(5): 126-132. [邹强, 周建中, 周超等. 基于可变模糊集理论的洪水灾害风险分析. 农业工程学报, 2012, **28**(5): 126-132.]
- [18] Liu X, Zhang J, Cai W *et al.* Information diffusion-based spatio-temporal risk analysis of grassland fire disaster in northern China. *Knowledge-Based Systems*, 2010, **23**(1): 53-60.
- [19] Feng LH, Luo GY. Analysis on fuzzy risk of landfall typhoon in Zhejiang province of China. *Mathematics & Computers in Simulation*, 2009, **79**(11): 3258-3266.
- [20] Zou Q, Zhou J, Zhou C *et al.* The practical research on flood risk analysis based on IIOSM and fuzzy α -cut technique. *Applied Mathematical Modelling*, 2012, **36**(7): 3271-3282.
- [21] Otnes RK, Enochson L eds. Digital time series analysis. Taylor & Francis: Wiley, 1972.
- [22] Davison AC, Smith RL. Models for exceedances over high thresholds. *Journal Royal Statistical Society B*, 1990, **52**(3): 393-442.
- [23] Ashkar F, Ouarda TBMJ. On some methods of fitting the generalized Pareto distribution. *Journal of Hydrology*, 1996, **177**: 117-141.
- [24] Cunnane C. A note on the poisson assumption in partial duration series models. *Water Resource Research*, 1979, **15**: 489-494.
- [25] Jin Sheng, Lei Xiaoyun, Li Hui *et al.* Hydrological P-III frequency curve calculation software development research. *South-to-North Water Transfers and Water Science & Technology*, 2009, **7**(5): 70-72. [靳晟, 雷晓云, 李慧. 水文 P-III 频率曲线计算软件开发研究. 南水北调与水利科技, 2009, **7**(5): 70-72.]
- [26] Liu Guangwen. The parameter estimation of Pearson III distribution. *Journal of China Hydrology*, 1990, (4): 1-15. [刘光文. 皮尔逊 III 型分布参数估计. 水文, 1990, (4): 1-15.]
- [27] Song Dedun, Ding Jing. The application of probability weighted moments to estimating the parameters for Pearson type-III distribution. *Journal of Hydrology*, 1988, **101**(101): 47-61.
- [28] Shen Q, Zhao R. Risk assessment of serious crime with fuzzy random theory. *Information Sciences*, 2010, **180**(22): 4401-4411.