

基于 G - H copula 函数的秦淮河流域洪水风险分析

高玉琴, 叶柳, 赖丽娟

(河海大学 水利水电学院, 江苏 南京 210098)

摘要: Copula 函数构造灵活, 适应性强, 被广泛应用于多变量水文频率分析计算中。选取 G - H copula 函数描述洪量、洪峰、洪水位间的相依结构, 边际分布模拟考虑广义极值 (GEV), 对数正态 (ln2) 和皮尔逊 III 型分布 (P - III)。以相关性最强的洪峰和洪水位为例进行二维变量洪水风险分析, 比较分析其传统重现期和 Kendall 重现期性质, 并计算条件风险率。构建 3 变量联合风险评价模型, 并与二变量情况进行对比。以秦淮河流域为例进行了应用, 表明多维联合重现期小于同现重现期, Kendall 重现期在两者之间且与单因素重现期误差最小; 流域洪峰和洪水位同现风险率最大, 洪峰量级越小, 洪水位超越风险率越小; 基于三维 G - H copula 函数计算的各变量同频率设计值大于相应单变量设计值, 并且相对于二维情况, 各重现期与单因素重现期之间的差值明显增大。研究结论可为秦淮河地区水利工程规划设计和洪灾风险评估提供有益参考。

关键词: 洪水风险分析; 超越概率; Kendall 重现期; 联合重现期; 秦淮河流域

中图分类号: TV122

文献标识码: A

文章编号: 1672-643X(2018)01-0172-06

Analysis of flood risk in Qinhuai River Basin based on G - H copula

GAO Yuqin, YE Liu, LAI Lijuan

(College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China)

Abstract: Copula function is commonly used in the flood frequency calculation of multidimensional variables because of its flexible construction and wide adaptability. G - H copula is selected to describe the dependence structure of three flood variables, namely flood volume, peak flow and water level. GEV, Logistical-normal and Pearson III distribution are considered for fitting the marginal distribution. A two-dimensional G - H copula flood risk assessment model based on peak flow and water level, which are of the highest dependence, is developed to investigate and analyze the characteristics of joint return period (JRP), co-recurrence return period and Kendall return period (KRP). The conditional risk of different combination of flood events is also computed. A three-dimensional G - H copula model is also developed to compare with the bivariate scenario. Qinhuai river basin is chosen as the study area. The result shows that the multi-dimensional joint return period is quite smaller than the co-recurrence return period. Kendall return period is in between of joint and co-recurrence return period but has the minimum error with the single variable return period. The co-recurrence probability of peak flow and water level turns out to be the highest; the smaller the magnitude of peak flow, the less possible for water level to exceed certain threshold. The design value of flood variables in same frequency computed by trivariate G - H copula is larger than the corresponding single variable design value and the difference of other return period with single variable return period is obviously larger than that of bivariate case. The conclusion can be useful for water resource project planning and design and risk assessment of flood disaster in the studied area.

Key words: flood risk analysis; exceedance probability; Kendall return period; joint return period; Qinhuai river basin

收稿日期: 2017-08-10; 修回日期: 2017-10-13

基金项目: 国家自然科学基金青年项目 (51309076); 中央高校基金前瞻性项目 (2014B05814); 江苏省优势创新平台项目

作者简介: 高玉琴 (1978-), 女, 四川成都人, 博士, 副教授, 主要从事水利规划与水利经济、洪水控制方面研究。

1 研究背景

随着洪水风险问题研究越来越深入,从多角度对洪水进行定义和描述的多变量联合风险分析方法已经越来越受到关注。Copula 函数因为其构造灵活,计算简单,且不限定边缘分布类型等优点而备受青睐,已被大量应用于降雨、干旱等多变量联合概率分析以及洪水遭遇、多站点多水文区的洪水联合风险分析等水文领域。

二维 copula 研究理论及应用已经相对成熟,三维及更高维数 copula 函数由于维数的拓展和各变量间相依性不对等,理论更为复杂。涂新军等^[1]应用 Archimedean copula 进行了滨海城市降雨和潮位联合分布的模拟和设计,并对比基于二次重现期的设计值和传统联合重现期设计值,证明二次重现期设计更为安全。黄锦林等^[2]基于最大熵 - copula 方法分析降雨和潮位关联性,得出多种雨潮设计值遭遇组合风险概率。闫宝伟等^[3]采用 clayton copula 构建长江和清江峰现时间与量级二维分布,进而探索洪水遭遇风险特性。Chowdhary 等^[4]详细论述了 copula 函数类型优选方法并介绍其在二元洪水频率计算中的应用。Zhang 等^[5]利用 Archimedean copula 函数研究了降雨强度、径流深、降雨历时两两变量间的联合分布,并分析相依性和边缘分布对结果的影响;Reddy 等^[6]具体阐述了二维洪水变量频率分析的步骤流程以及用 Monte Carlo 数值模拟法判断 copula 拟合程度的方法。在三维 copula 方面,陈子燊等^[7]分析洪量、历时、洪峰三变量联合分布及各种重现期的设计分位数;侯芸芸等^[8]考虑洪峰、洪量和历时之间的相依性,构造了三维对称性 Archimedean copula 函数,探讨洪水联合概率分布和条件概率分布;Ganguli 等^[9]利用多种联结函数类型描述洪峰流量、洪量、洪水历时的两变量、三变量联合概率分布,并讨论了两种重现期特性和尾部相关性。

基于次洪变量,如时段洪量、洪峰、峰现时间、洪水历时等进行的联合分布研究以及降雨和洪水、降雨和潮位相关性研究较多,但是目前从流域洪水风险角度将影响风险大小的洪量、洪峰流量和洪水位三者进行联合分析的研究较少。本文以秦淮河流域为研究区域,选取 P-III、GEV 分布和两参数对数正态分布 Ln2 描述洪水风险变量洪量、洪峰和洪水位的边缘分布类型,利用 G-H copula 构建二维和三维风险评价模型,计算不同组合洪水事件的联合重现期、条件重现期以及二次重现期,为区域工程设

计和风险评估工作提供依据。

2 数据与方法

2.1 研究区域与数据来源

秦淮河流域地处长江下游地区,面积为 2 631 km²,干流长度为 34 km,属于亚热带湿润、半湿润季风气候区,汛期雨量充沛,洪水灾害频繁。根据 1986-2006 年流域出口处秦淮新河和武定门闸水文站的日流量资料和 1986-2006 年东山站站前水位资料,采用年最大值选样法(AM),基于袁玉等^[10]根据秦淮河流域水文、气象、降雨等资料建立的 HEC-HMS 水文模型及 HEC-RAS 水力模型,模拟年最大洪水事件演进到东山站前的水位。洪量、洪峰均为秦淮新河闸与武定闸两处叠加值,以全流域径流深的形式表示洪量,洪水位为东山站点前水位。

2.2 G-H copula 函数

G-H Copula 是 Archimedean copula 函数的一种,通常用于描述正相关和上尾相关性随机变量^[11],当维数为 2 和 3 时,G-H Copula 表达式分别为:

$$C(u_1, u_2) = \exp\{-[(-\ln(u_1))^\theta + (-\ln(u_2))^\theta]^{1/\theta}\} \quad \theta \in (1, \infty) \quad (1)$$

$$C(u_1, u_2, u_3) = \exp\{-[(-\ln(u_1))^\theta + (-\ln(u_2))^\theta + (-\ln(u_3))^\theta]^{1/\theta}\} \quad \theta \in (1, \infty) \quad (2)$$

2.3 参数估计与拟合优度检验

核密度估计、相关性指标法、IFM 法、EML 法等都可进行 copula 参数估计。对于二维 Archimedean Copula 常用相关性指标法,根据 τ 和 θ 的关系由 τ 来估计。但三维 copula 显然不能使用相关性指标法估算参数,本研究采用 IFM 法。

为验证 Copula 和实测样本经验累积概率之间的拟合程度,需对构建的模型进行拟合优度检验。常见检验方法有 RMSE 法、AIC 法和 BIC 法。本文采用均方根误差法:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (F_{emp} - C)^2} \quad (3)$$

式中: F_{emp} 为经验概率值; C 为理论概率值; n 为系列长度。RMSE 值越小,拟合情况越好。三维经验频率的计算公式为:

$$F_{emp}(x_{i1}, x_{i2}, x_{i3}) = P(X_{i1} \leq x_{i1}, X_{i2} \leq x_{i2}, X_{i3} \leq x_{i3}) \\ = \left(\sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n n_{jkl} - 0.44 \right) / (n + 0.12) \quad (4)$$

式中: n_{jkl} 为同时满足 $X_{i1} \leq x_{i1}, X_{i2} \leq x_{i2}, X_{i3} \leq x_{i3}$ 时的联合观测个数。

2.4 重现期

洪水特征变量的联合概率分布、条件概率分布以及相应的重现期计算是水文频率分析的重要成果,超过某一重现期设计阈值的洪水事件被定义为危险事件。传统同现重现期和联合重现期对洪水事件安全和危险域识别存在局限性,有可能发生识别错误的情况,造成风险管理和工程设计的成本增加或风险抵抗能力不足,而 Kendall 重现期能在同一临界水平下对任意洪水组合的安全和危险域识别保持一致性^[12-13]。本文将 Kendall 重现期应用到多因素风险分析中,并传统重现期进行对比。传统重现期定义见参考文献[7],在此不再赘述。

Kendall 重现期,又称二次重现期,根据 Genest 和 Rivest 定义的 Kendall 分布函数,对于给定的 t ,通过求解特征量联合分布 C 小于或等于 t 的累积概率将多维变量信息投射为一维。Kendall 重现期定义为:

$$T_K = \frac{1}{P(C(u_1, u_2) \geq t)} = \frac{1}{1 - K_C(t)} \quad (5)$$

式中: K_C 为 Kendall 分布函数, $K_C(t) = P(C(u_1, u_2) \leq t)$ 。对于 Archimedean copula 函数族中的二维 G-H copula, K_C 可以表示为:

$$K_C(t) = t - \frac{\varphi(t)}{\varphi'(t)} = t - t\theta^{-1} \ln t \quad (6)$$

对于三维 G-H copula, 可以表示为:

$$K_C(t) = t - \frac{t(3\theta - \ln t - 1) \ln t}{2\theta^2} \quad (7)$$

3 结果与讨论

3.1 相依性度量和边际分布

采用 Pearson' γ , Spearman' ρ 和 Kendall' τ 进行洪量和洪峰、洪量和洪水位、洪峰和洪水位两两变量间的相依性度量,并在括号中给出显著性水平为 0.05 时对应的 P 值,计算结果如表 1。由表 1 可以看出,变量之间的相关性均较大,表明洪量(W)、洪峰(Q)、洪水位(Z)之间存在较强相关性,因而可进行洪水变量间的联合概率特性分析。

表 1 变量之间的相依性

变量	Pearson' γ	Spearman' ρ	Kendall' τ
$W \& Q$	0.865 (4.2×10^{-7})	0.921 (4.5×10^{-6})	0.762 (5.8×10^{-8})
$W \& Z$	0.798 (1.46×10^{-5})	0.893 (5.1×10^{-8})	0.730 (4.4×10^{-6})
$Q \& Z$	0.975 (6.2×10^{-14})	0.974 (1.2×10^{-13})	0.921 (6.6×10^{-9})

选取 P-III、GEV 和 ln2 描述变量的边际分布,边际分布概率密度表达式见文献[14]。采用最大似然法估计参数并通过 KS 检验判断拟合优度,KS 统计量值越小,拟合情况越好; P 值越大,则有越大可能接受该分布。表 2 为洪水特征变量边际分布参数和拟合优度检验结果。由表 2 可知,3 种边际分布均能被接受,基于最小 KS 统计量分别选择 GEV, ln2 和 P-III 分布为洪量、洪峰、洪水位的边际分布。绘制变量边际分布经验频率与理论频率关系图,如图 1 所示,由图 1 可以看出,点据分布在 1:1 斜线附近,证明拟合程度较好。

表 2 边际分布参数估计和拟合优度检验

变量	分布	位置	尺度	形状	KS	P 值
W	P-III	28.38	0.004	0.44	0.23	0.20
	GEV	72.60	53.86	0.31	0.12	0.88
	ln2	4.51	0.87		0.14	0.79
Q	P3	-247.93	0.006	4.63	0.15	0.71
	GEV	424.66	263.84	0.05	0.16	0.61
	ln2	6.18	0.70		0.14	0.73
Z	P-III	0.56	2.49	16.00	0.14	0.79
	GEV	6.42	1.29	-0.17	0.16	0.63
	ln2	1.92	0.21		0.16	0.63

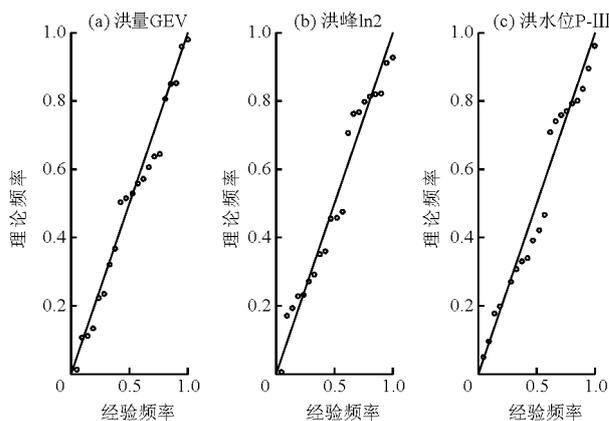


图 1 边际分布拟合图

3.2 联合分布函数

采用 G-H Copula 描述二维变量间相依结构,用 IFM 法估计参数,计算得洪量洪峰、洪量洪水位和洪峰洪水位二维联合分布函数的参数分别为 3.1231、2.8655、6.7281,对应的均方根误差 RMSE 分别为 0.0090、0.0106、0.0149。以相关性最强的洪峰和洪水位为主进行二维变量洪水风险分析,探讨其联合重现期、同现重现期和 Kendall 重现期性质,并分析洪峰不超过一定重现期时洪水位风险率。分别计算洪峰和洪水位在 5~200 a 重现期内

的联合 T^U , 同现 T^N 和二次重现期 T^K 值, 并给出单变量条件下的设计值, 见表 3。 $Q-Z$ 二维分布条件下, 相较于单因素重现期, T^U 更小, T^N 更大, 而 Kendall 重现期在 T^U 和 T^N 之间且与单因素重现期误差最小, 大小顺序为 $T^U < T < T^K < T^N$ 。应用 T^N 安全裕度大, 但增加经济投入, 工程设计中可以考虑采用给定二次重现期的最大概率设计组合来达到较低成本承受较大风险的目标。

绘制洪峰和洪水位的概率分布图和相应的重现期等值线图, 并加入实测数据进行对比, 如图 2 和图 3。由图 2、3 可知, $Q-Z$ 最大联合重现期约为 15 a, 最大同现重现期则超过 20 a。可计算任意 $Q-Z$ 组合事件的重现期, 如发生 50 年一遇洪峰和 20 年一遇洪水位的联合重现期和同现重现期分别为 19.99 a 和 50.03 a。绘制 $W-Q$ 和 $W-Z$ 的同现重现期等值线图, 如图 4。 $Q-Z$ 的同现重现期较 $W-Q$ 和 $W-Z$ 的

同现重现期小, 说明秦淮河流域洪峰和洪水位同现风险率最大。

绘制洪峰不超过 10、20、50 年一遇时洪水位的超越概率 $F_{Z|Q} = P(Z > z | Q \leq q)$, 如图 5。洪峰不超过 10、20、50 年一遇设计值时, 洪水位超过 5 年一遇 (8.29 m) 的概率分别为 0.1113、0.1579、0.1837。给定某一洪水位阈值, 洪峰量级越小, 洪水位风险率越小, 即洪峰量级小, 洪水位也越不容易超过该阈值。

表 3 单变量设计值及各种重现期

T	W	Q	Z	T^U	T^N	T^K
200	807.99	2956.98	11.81	180.47	224.27	211.87
100	629.68	2481.93	11.27	90.26	112.10	105.91
50	485.98	2049.66	10.69	45.15	56.01	52.93
20	337.38	1538.24	9.84	18.09	22.36	21.15
10	248.99	1191.97	9.12	9.07	11.14	10.55
5	175.87	875.27	8.29	4.56	5.53	5.25

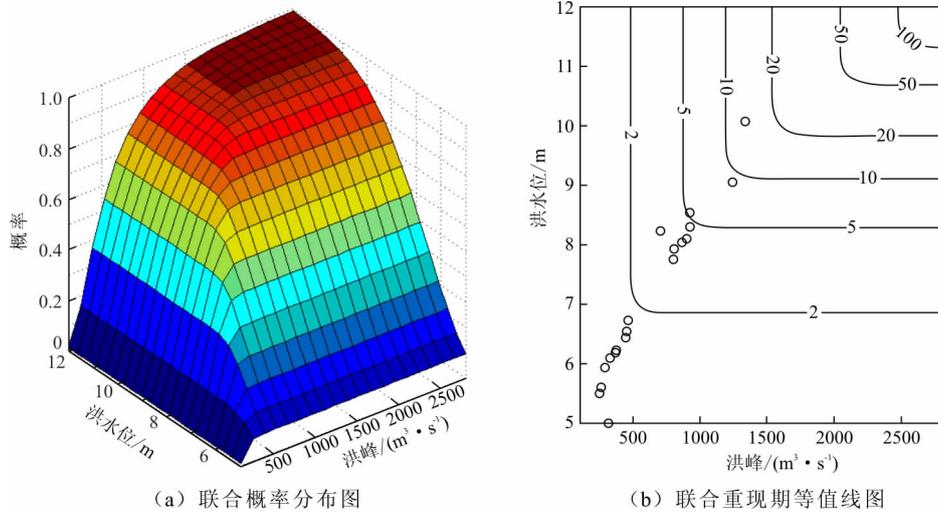


图 2 $Q-Z$ 联合概率分布图及联合重现期等值线图

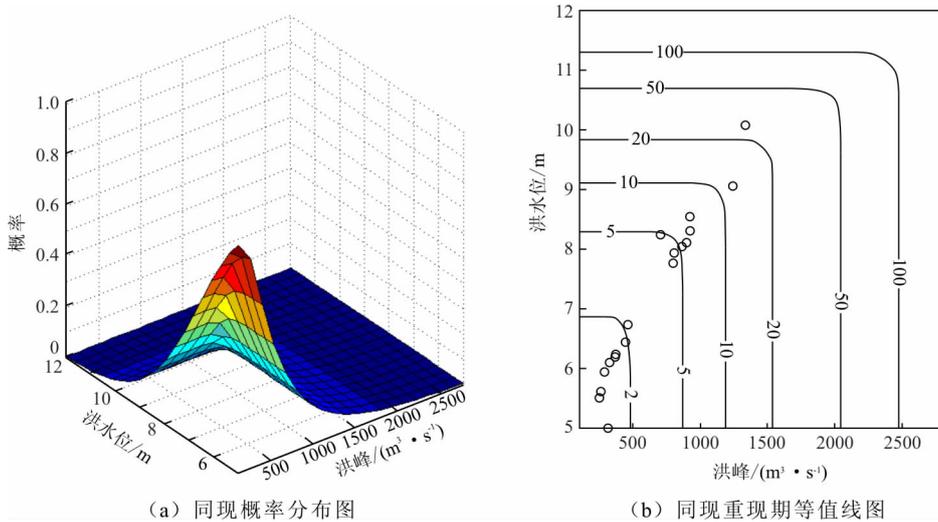


图 3 $Q-Z$ 同现概率分布图及同现重现期等值线图

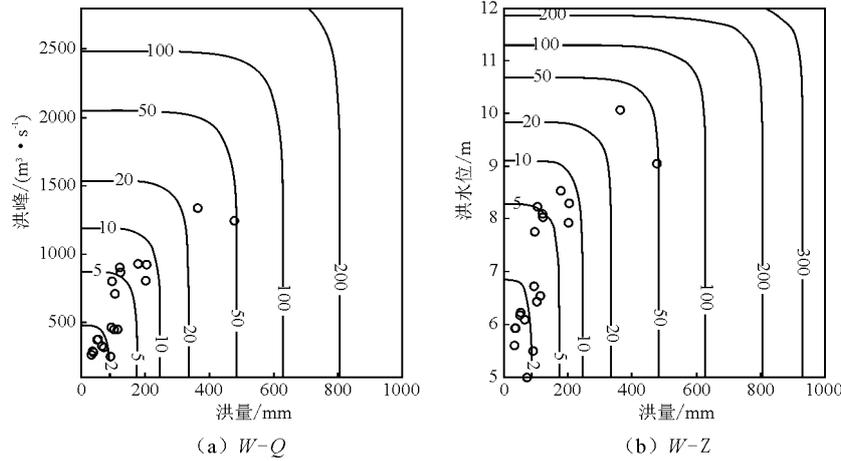


图4 W-Q、W-Z 同现重现期等值线图

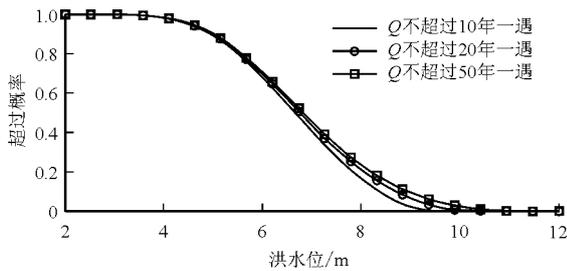


图5 洪峰不超过某一重现期时洪水位风险率

按照公式(3),通过各变量边缘分布构建三维G-H copula模型,用IFM法由洪水资料估计参数,计算得出 $\theta = 3.4472$,用均方根误差法对G-H copula函数进行拟合优度检验, $RMSE = 0.0147$,误差值较小,满足拟合要求。分别计算洪量、洪峰和洪水位在5~200 a 重现期内的 T^U 、 T^N 、 T^K 值和3变量情况下的同频率设计值,如表4。由表4可以看出,3变量同频率设计值大于相应单变量设计值,洪量洪峰洪水位三维分布条件下,各重现期大小排列依然是 $T^U < T < T^K < T^N$,但是在二维联合分布基础上再考虑洪水风险

因子后,传统重现期、二次重现期与单因素重现期之间的差值明显增大,风险率变化幅度增大。

由《南京城市防洪规划报告(2011-2020年)》可知,东山站20年一遇规划水位为10.8 m,绘制洪水位不超过10.8 m条件下洪量洪峰的条件概率 $F_{W|Z} = P(W \leq w, Q \leq q | Z \leq 10.8)$ 及相应的重现期等值线图,如图6。洪水位不超过10.8 m条件下,实测资料中大部分W-Q联合重现期小于或等于5 a,说明较高洪水位条件下,洪量和洪峰的任意一个超过风险率较大。

表4 3变量同频率设计值及各种重现期

T	W	Q	Z	T^U	T^N	T^K
200	907.00	3197.99	12.03	145.56	268.34	239.40
100	705.88	2692.12	11.52	72.85	134.05	119.60
50	547.42	2240.83	10.96	36.49	66.91	59.70
20	383.42	1705.12	10.14	14.68	26.62	23.76
10	285.74	1340.43	9.44	7.41	13.19	11.78
5	204.80	1004.15	8.65	3.78	6.47	5.79

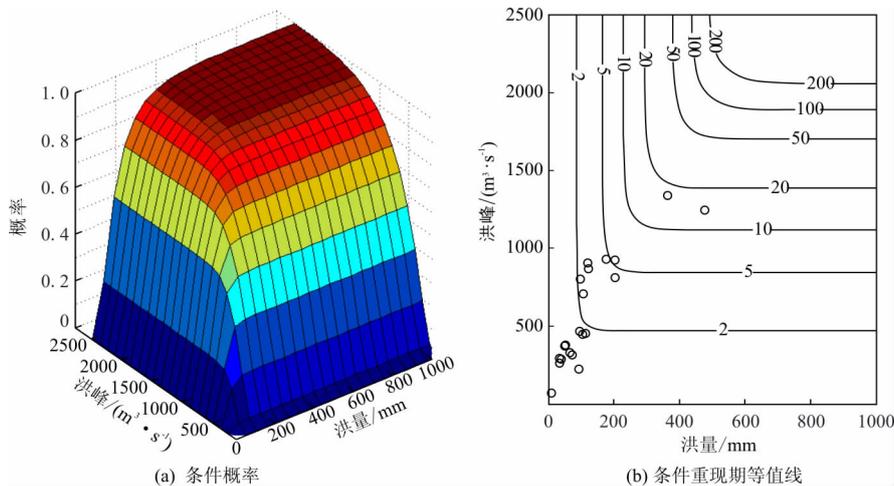


图6 Z ≤ 10.8 m 时 W-Q 条件概率分布及条件重现期等值线

4 结 论

以秦淮河流域为例,基于边际分布拟合,选择广义极值分布GEV,对数正态分布ln2和P-III分布作为洪量(W)、洪峰(Q)、洪水位(Z)的边际分布类型,构建二维和三维G-H copula洪水风险评价模型,并计算多变量联合概率、条件概率以及传统重现期和Kendall重现期,得出了以下结论:

(1)相较于单因素重现期, $Q-Z$ 二维 T^U 更小, T^D 更大,而Kendall重现期在两者之间且与单因素重现期误差最小。

(2) $Q-Z$ 的 T^U 小于 $W-Q$ 和 $W-Z$ 的 T^D ,说明秦淮河流域洪峰和洪水位同现风险率最大; $Q-Z$ 在二维分布条件下,洪峰量级越小,则洪水位风险率越小。

(3)计算3变量联合分布和限制洪水位条件下 $W-Q$ 联合概率分布,3变量情况下的同频率设计值大于相应单变量设计值,3变量传统重现期、Kendall重现期与单因素重现期之间的差值明显增大。

参考文献:

- [1] 涂新军,杜奕良,陈晓宏,等. 滨海城市雨潮遭遇联合分布模拟与设计[J]. 水科学进展,2017,28(1):49-58.
- [2] 黄锦林,范嘉炜,唐造造. 基于最大熵-Copula方法的降雨潮位关联性分析——以广州为例[J]. 灾害学,2017,32(1):65-71.
- [3] 闫宝伟,郭生练,陈璐,等. 长江和清江洪水遭遇风险分析[J]. 水利学报,2010,41(5):553-559.
- [4] CHOWDHARY H, ESCOBAR L A, SINGH V P. Identification of suitable copulas for bivariate frequency analysis of

flood peak and flood volume data[J]. Hydrology Research, 2011, 42(2-3):193.

- [5] ZHANG L, SINGH V P. Bivariate rainfall frequency distributions using Archimedean copulas[J]. Journal of Hydrology, 2007, 332(1):93-109.
- [6] REDDY M J, GANGULI P. Bivariate flood frequency analysis of upper godavari river flows using archimedean copulas[J]. Water Resources Management, 2012, 26(14):3995-4018.
- [7] 陈子燊,黄强,刘曾美. 基于非对称 Archimedean Copula 的三变量洪水风险评估[J]. 水科学进展,2016,27(5):763-771.
- [8] 侯芸芸,宋松柏,赵丽娜,等. 基于 Copula 函数的 3 变量洪水频率研究[J]. 西北农林科技大学学报(自然科学版),2010,38(2):219-228.
- [9] GANGULI P, REDDY M J. Probabilistic assessment of flood risks using trivariate copulas[J]. Theoretical & Applied Climatology, 2013,111(1-2):341-360.
- [10] 袁玉,高玉琴,吴锡. 基于 HEC_HMS 水文模型的秦淮河流域圩垸式防洪模式洪水模拟[J]. 三峡大学学报(自然科学版),2015,37(5):34-39.
- [11] 潘瑾林,陈子燊. 基于 GH Copula 的韩江水文干旱联合概率分布研究[J]. 中山大学学报(自然科学版),2015,54(1):110-115.
- [12] 黄强,陈子燊. 基于二次重现期的多变量洪水风险评估[J]. 湖泊科学,2015,27(2):352-360.
- [13] 范嘉炜,黄锦林. 基于 Kendall 重现期的降雨潮位风险分析[J]. 水电能源科学,2017,35(5):21-24+20.
- [14] 高玉琴,陈钊西,赵立梅,等. 秦淮河流域不同频率降雨联合概率分析[J]. 水电能源科学,2016,34(3):1-5+23.