

DOI: 10.19797/j.cnki.1000-0852.20210411

基于 Copula 函数的二维极端降水 联合概率特征分析

刘萍, 李帆, 潘奇鑫, 李春秀

(扬州大学 水利科学与工程学院, 江苏 扬州 225009)

摘要:根据多年逐日降水资料,采用第95个百分位值作为极端降水阈值,利用 Archimedean Copula 函数构建极端降水量和年最大日降水量(P_{95}, P_{\max})、极端降水强度和年最大日降水量(I_{95}, P_{\max})、极端降水量和极端降水贡献率(P_{95}, R_{95})之间的二维联合分布函数。以高邮站1968—2020年日降水资料为研究对象,进行实例分析。Frank Copula 能较好地描述(P_{95}, P_{\max})、(I_{95}, P_{\max})之间的联合分布,Clayton Copula 函数能较好地描述(P_{95}, R_{95})之间的联合分布。相比于联合重现期和同现重现期,Kendall 重现期能更准确地反映给定重现期情况下多变量极端降水要素的概率。

关键词:Copula 函数;极端降水;二维联合;重现期

中图分类号:TV125

文献标识码:A

文章编号:1000-0852(2023)01-0035-06

0 引言

目前,对极端降水的研究主要集中在极端降水时空特征变化、成因分析、气候条件下的降水极值概率分布模式拟合等方面^[1-3]。对极端降水的频率及重现期的研究仍侧重于单一要素^[4],无法反映各要素之间的相关性,不能有效反映极端降水的综合特征。因此有必要建立多要素联合概率模型进行分析计算。极端降水要素往往需要用不同的概率分布函数来表示,而 Copula 函数可采用任意边缘分布函数建立联合分布函数^[5],近年来在水文领域应用广泛,在水文事件多变量频率分析、遭遇组合分析、随机模拟、水文模型与预报等方面应用较多^[6-8]。

联合重现期和同现重现期是普遍认可的传统重现期^[9]。但这两种重现期对危险事件的判断存在一定的局限性,定义的危险区域存在着一些不合理性,并不一定能合理地描述多变量事件^[10]。SALVADORI 等^[11]根据

Kendall 函数得到 Copula 等值线,将超过给定阈值的所有事件的集合定义为危险域,继而提出 Kendall 重现期计算方法。相关原理在海岸工程中得到初步应用^[12]。陈子焱等^[13]以增江麒麟咀水文站为例,对比发现 Kendall 重现期可以更准确地反映洪峰洪量的联合风险率。

本研究以高邮站多年降水量为研究对象,通过 $P_{95}, P_{\max}, I_{95}, R_{95}$ 来定义极端降水事件,分别从极端降水事件发生的量级、强度和集中程度对极端降水事件进行描述,能较好地反映极端降水事件的主要特征及规律。基于 Archimedean Copula 函数构建 3 对二维联合分布,探讨极端降水要素的联合概率分布、相关性以及重现期情况,得到极端降水的重现期。

1 研究方法

1.1 变量选取与计算

$P_{95}, P_{\max}, I_{95}, R_{95}$ 是极端降水事件的重要特征变量^[7,14]。根据百分位阈值法^[15],将降水量按升序排序,得

收稿日期:2021-10-08

网络首发日期:2023-01-19

网络首发地址:<https://kns.cnki.net/kcms/detail//11.1814.P.20230117.1817.005.html>

基金项目:中国博士后科学基金面上项目(2019M651976);江苏省自然科学基金项目(BK20200942);江苏省高等学校自然科学研究项目(20KJB570003)。

作者简介:刘萍(1997—),女,江苏南通人,硕士研究生,主要研究方向为水文水资源。E-mail:1441327435@qq.com

通信作者:李帆(1982—),男,辽宁营口人,博士,副教授,主要研究方向为水文统计、风险分析等。E-mail:lifan@yzu.edu.cn

到 $x_1, x_2, \dots, x_j, \dots, x_n$, 某个值小于或等于 x_i 的概率, 计算第 95 个百分点的极端降水阈值, 具体计算方法可参考文献[16-17]。

1.2 边缘分布拟合

选用 P-III 分布、Gamma 分布、GEV 分布、Log-normal 分布作为初选分布函数, 分别对 P_{95} 、 I_{95} 、 P_{max} 、 R_{95} 进行拟合, 利用 K-S 检验对分布函数进行拟合优度检验, 并通过比较均方根误差, 选出最优的边缘分布函数。

1.3 二维联合概率分布

本文选用 Archimedean Copula 函数族中的 Gumbel-Hougaard Copula、Clayton Copula、Frank Copula 三种函数作为备选函数来构建极端降水要素二维联合分布, 其参数可采用相关性指标法^[18]进行估计。运用图形评价分析法可直观分析经验频率与理论频率之间的偏差^[19], 定性地得出拟合程度的优劣程度, 其中经验联合频率 F_{emp} 可表示为^[20]:

$$F_{emp}(x_{i1}, x_{i2}) = p(X_1 \leq x_{i1}, X_2 \leq x_{i2}) = \left(\sum_{g=1}^n \sum_{k=1}^n n_{g,k} - 0.44 \right) / (n + 0.12) \quad (1)$$

式中: $n_{g,k}$ 为同时满足 $X_1 \leq x_{i1}, X_2 \leq x_{i2}$ 时的联合观测值个数; n 为系列长度。理论联合频率由 Copula 函数得出。

采用 K-S 检验法进行拟合优度检验, p 值大于 0.05 即可通过检验。运用离差平方和最小准则 (OLS) 和 Akaike 信息准则 (AIC)^[21] 则可对 Copula 函数的拟合精度进行定量评价, 选择最优的 Copula 函数。

1.4 重现期

极端降水要素的单变量重现期 T 可表示为:

$$T(x) = \frac{\mu_T}{P(X \geq x)} \quad (2)$$

式中: $T(x)$ 为单变量重现期; μ_T 为极端降水事件间隔的期望值。

两变量的重现期可用联合重现期或同现重现期来表示^[22]。理论上, 重现期越大, 危险域越小, 较大重现期的危险域应包含在较小重现期的危险域之内。而上述两种重现期都会产生部分较大重现期危险域出现在较小重现期危险域之外的问题, 即大的重现期认定危险的事件, 小的重现期却认为安全的矛盾^[23]。针对这一问题, SALVADORI 等^[24]将 Kendall 测度 K_c 应用在重现期的定义中, 通过求累积概率小于或等于 t 的概率, 可以将多维极值事件投射为一维分布, 使变量组合在给定重现期下的危险域唯一。三种重现期的安全域如图 1 所示。联合重现期将左下角矩形部分 (OR) 定义为安全域。Kendall 重现期将 $C(u, v) = t$ 曲线

左下部分 (KEN) 和左下角矩形定义为安全域, 使产生的重现期比联合重现期大。同现重现期进一步添加 AND 部分, 定义为安全域, 得到最大的重现期。两变量 Kendall 测度 K_c 可表示为:

$$K_c(t) = P(C(u, v) \leq t) = t - \frac{\varphi(t)}{\varphi'(t^*)} \quad (3)$$

式中: $\varphi(t)$ 为生成元, $\varphi'(t^*)$ 为 $\varphi'(t)$ 的右导数。由 Kendall 测度确定的二次重现期可表示为:

$$T_k(x_1, x_2) = \frac{\mu_T}{P(C(u, v) \geq t)} = \frac{\mu_T}{1 - K_c(t)} \quad (4)$$

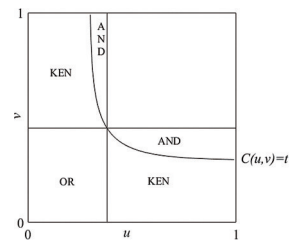


图1 三种重现期的安全域识别
Fig.1 Identification of three kinds of return period's security domain

本文基于百分位阈值法得到极端降水事件样本序列, 优选出单变量边缘分布函数, 判断要素间相关性, 构建二维联合分布函数并根据 OLS 准则和 AIC 信息准则进行优选, 继而根据优选出的 Copula 函数进行概率分析、重现期计算。

2 实例分析

2.1 研究区域

高邮属亚热带湿润气候区, 气候温和、雨量充沛、日光充足, 常年主导东南风, 年平均气温 15℃, 年均降雨量 1 030 mm。本次研究以高邮站 1968—2020 年逐日降水资料为基础, 验证方法有效性。

2.2 单变量分布函数拟合

由表 1、表 2 可知, K-S 检验 p 值均能通过检验; 均方根误差计算结果表明, 对于 P_{95} 、 P_{max} 、 I_{95} 使用 P-III 分布进行边缘分布拟合, 得到的均方根误差最小, 而对于 R_{95} , 则选取 GEV 分布进行边缘分布拟合, 得到的均方根误差最小。

表1 K-S 检验 p 值

Table1 p value of K-S Test				
	P_{95}	P_{max}	I_{95}	R_{95}
P-III	0.755	0.950	0.978	0.967
Gamma	0.654	0.797	0.980	0.816
GEV	0.560	0.987	0.989	0.982
Log-Normal	0.662	0.973	0.986	0.510

表2 各分布均方根误差值

	P-III	Gamma	GEV	Log-normal
P_{95}	41.818	42.397	42.786	50.897
P_{max}	3.605	6.282	4.555	5.032
P_{95}	1.173	1.539	1.411	1.408
R_{95}	0.012	0.020	0.012	0.034

注:加粗部分为相应均方根误差的最小值, R_{95} 对应的P-III的RMSE值为0.0122,GEV RMSE值为0.0118。

2.3 极端降水要素二维联合分布

利用 Spearman 相关系数 ρ 和 Kendall 相关系数 τ 分析变量之间的相关性, ρ 和 τ 均能通过 0.05 水平的显著性检验(见表3),变量间具有明显的相关性。选取 P_{95} 和 P_{max} , I_{95} 和 P_{max} , P_{95} 和 R_{95} 分别构建二维联合分布函数,运用图形评价分析法定性评价。理论值和经验值具

有较强的一致性(见图2),决定系数均大于0.97,但仅说明通过一致性检验,还需进一步定量评价。利用K-S检验进行拟合优度检验,根据OLS准则和AIC信息准则定量评价备选 Copula 函数,选出拟合程度最好的函数。由表4计算结果表明, (P_{95}, P_{max}) 和 (I_{95}, P_{max}) 联合概率拟合效果最佳的函数为 Frank Copula 函数, (P_{95}, R_{95}) 联合概率拟合效果最佳的函数为 Clayton Copula 函数,且通过了K-S检验。

表3 相关性指标计算结果

	(P_{95}, P_{max})	(I_{95}, P_{max})	(P_{95}, R_{95})
Spearman(ρ)	0.68	0.92	0.94
Kendall(τ)	0.49	0.77	0.80

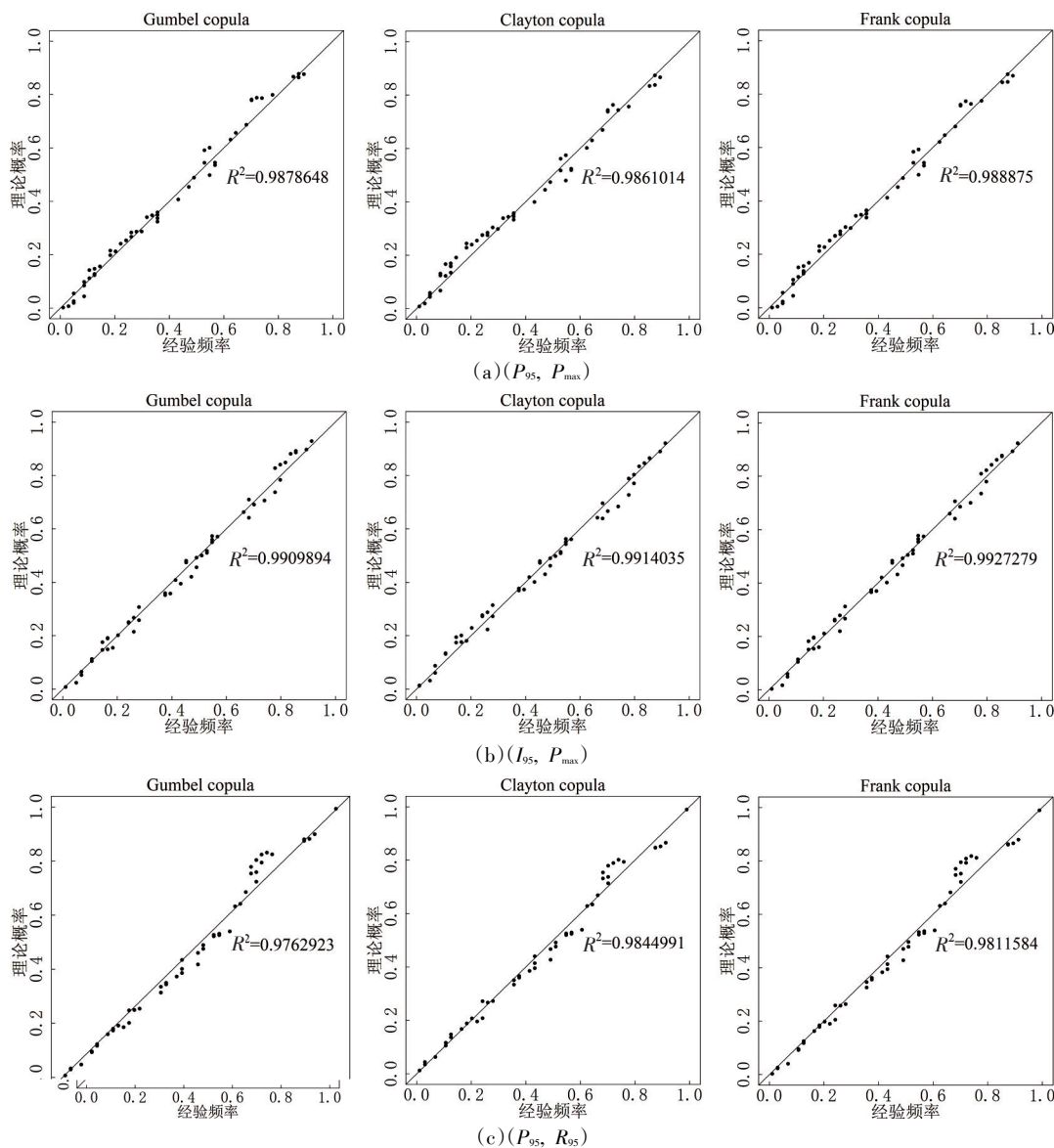


图2 各要素三种函数下P-P图

Fig.2 P-P diagram of three functions of each element

表4 Copula二维联合分布函数参数及检验值

Table4 Parameters and K-S test values of bivariate Copula

函数	参数及检验值	(P_{95}, P_{max})	(I_{95}, P_{max})	(P_{95}, R_{95})
Gumbel	θ	1.963	4.299	5.100
	OLS	0.029	0.025	0.041
	AIC	-367.208	-380.185	-329.822
Clayton	θ	1.926	6.597	8.200
	OLS	0.031	0.025	0.033
	AIC	-360.153	-382.631	-351.917
Frank	θ	5.567	15.352	18.595
	OLS	0.028	0.023	0.037
	AIC	-371.728	-391.331	-341.768
p值		0.970	1.000	0.970

注:加粗部分为相应检验值的最小值,P值为最优 Copula 函数对应的 K-S 检验结果。

图3为联合超越概率三维图和等值线图,立体直观地展现两变量同时大于等于某一定值的联合概率,随着变量值的增大,联合概率值减小,等值线主要分布在左下角,说明两个变量同时小于某一特定较小值的概率较大。图3第一张等值线图中, P_{95} 为550 mm的超过概率大于0.1, P_{max} 为130 mm的超过概率也大于0.1,但综合考虑两个变量的情况下,该事件发生的概率便小于0.1。因此,两变量联合比单变量分析更能真实全面

地反映极端降水事件的特征。造成这一现象的原因,一方面是暴雨事件各要素间有密切的物理联系,存在一定的因果关系,不能看成相互独立的变量;但另一方面,这种物理联系非常复杂,影响因素众多,无法精确计算,需要用到概率的相关方法来描述变量间的关系。

2.4 重现期分析

由表5可见,联合重现期均小于单变量重现期,同现重现期均大于单变量重现期,这与联合重现期、同现重现期的定义一致;各变量增幅相同的情况下,同现重现期比相应联合重现期的增幅大;联合重现期、同现重现期变化趋势是一致的,具有同步效应,一定程度上可以说明极端降水要素之间的关联性。Kendall 重现期介于联合重现期和同现重现期之间且大于单变量重现期($T_0 > T_k > T_a$),这是由危险域判断的不一致性造成的。联合重现期将不该纳入危险域的部分(图1中KEN部分)纳入了危险域,导致重现期偏小;同现重现期将本该纳入危险域的部分(图1中AND部分)未纳入危险域,导致重现期偏大。给定重现期越大,对危险域的误判越大,继而三种重现期之间的偏差越大。可以把联合分布的两种重现期看作是单变量重现期的两种极端情况,作为实际重现期的估计区间。联合重现期的危险率偏高,同现重现期的危险率偏低,而Kendall 重现期能较准确地反映特定频率情况下多变量极端降水要素的风险概率。

联合重现期图与 Kendall 重现期图类似,其等值线走势方向与图3中的走势方向相反,但相同要素组合条件下,Kendall 重现期大于联合重现期,而同现重现

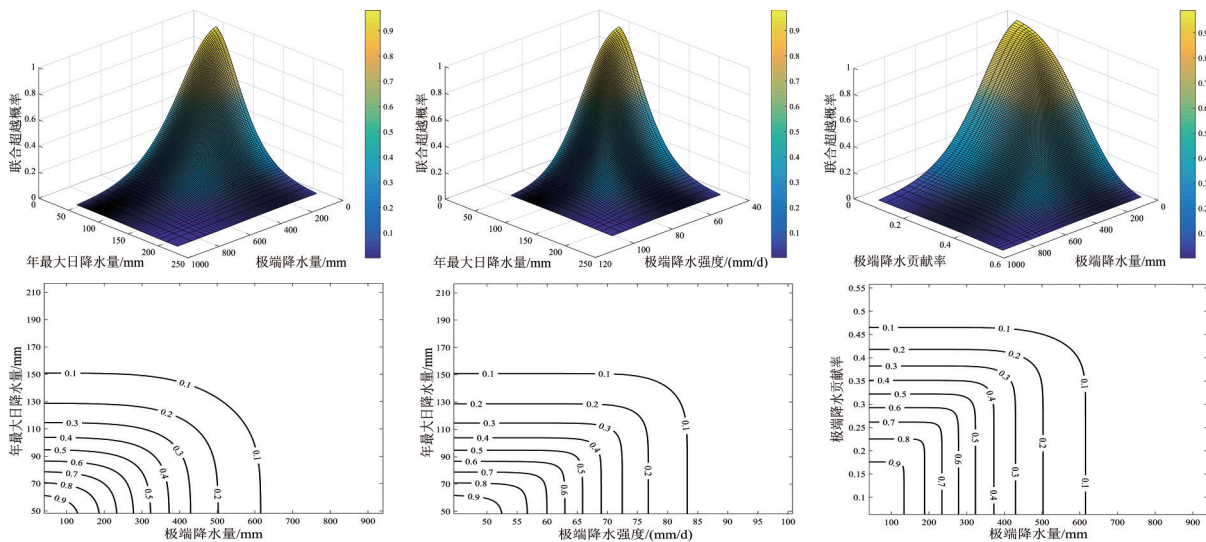


图3 二维联合超越概率图像

Fig.3 Two-dimensional joint transience probability based on the best fitted copulas

表5 极端降水要素重现期

Table5 Return period of extreme precipitation elements

重现期T/年	联合重现期T ₀ /年			同现重现期T ₀ /年			Kendall 重现期T ₀ /年		
	(P ₉₅ , P _{max})	(I ₉₅ , P _{max})	(P ₉₅ , R ₉₅)	(P ₉₅ , P _{max})	(I ₉₅ , P _{max})	(P ₉₅ , R ₉₅)	(P ₉₅ , P _{max})	(I ₉₅ , P _{max})	(P ₉₅ , R ₉₅)
2	1.63	1.83	1.85	2.59	2.20	2.18	2.22	2.08	2.06
5	3.43	4.10	3.89	9.23	6.39	7.01	6.73	5.55	5.73
10	6.11	7.26	6.72	27.55	16.06	19.51	17.67	12.43	13.77
20	11.23	12.83	11.96	91.12	45.35	60.93	52.87	30.77	37.79
50	26.32	28.34	27.15	496.71	211.87	315.85	265.71	124.38	174.84
100	51.36	53.57	52.22	1888.47	750.25	1175.59	978.25	410.40	620.52

期图与两者有较大差别,等值线走势方向与图3中方向一致。从重现期等值线图中可查询各种组合下的重现期,也可查询不同重现期下各要素的组合情况,为农业生产、工程建设、水资源管理等提供理论支持。根据 Kendall 重现期做出的决策,更能兼顾安全与经济。

分析结果与已有研究的结论一致。本文分析的三种重现期,与 Xu H S 等^[25]在降雨和风暴潮的风险联合分析中得出的重现期情况一致, Kendall 重现期都介于联合重现期和同现重现期之间,随着重现期的增大,这三种重现期之间的偏差越大,说明 Kendall 重现期的分析结果更加科学合理。在暴雨事件方面,冯介玲等^[8]分析了暴雨综合指标(暴雨量、年暴雨强度和年暴雨贡献率)的联合分布情况,结果表明, Archimedean Copula 可以用来构建暴雨要素的联合概率分布,其在联合重现期和同现重现期的分析结果与本研究的结论相似。

单变量重现期反映单变量重现特征,联合后的重现期综合考虑了极端降水要素,多变量联合后呈现出来的重现期信息是单变量重现期无法给出的,因此二维联合分布函数对于探索具有多要素的极端降水事件具有重要的实践意义。

3 结论与展望

(1)研究区 P_{95} 、 P_{max} 、 I_{95} 的边缘分布为 P-III 分布, R_{95} 的边缘分布为 GEV 分布。

(2)研究区的极端降水事件可采用 Frank Copula 函数构建 P_{95} 和 P_{max} 、 I_{95} 和 P_{max} 之间的二维联合分布函数,采用 Clayton Copula 函数构建 P_{95} 和 R_{95} 之间的二维联合分布函数。

(3)联合分布考虑了 (P_{95}, P_{max}) 、 (I_{95}, P_{max}) 、 (P_{95}, R_{95}) 的不同组合,求得不同组合下的联合超越概率和重现期,能够更加全面地考虑极端降水的特征,使得降水的重现期分析计算更加准确、合理。通过计算分析发现,联合重现期的危险率偏高,同现重现期的危险率偏低,而 Kendall 重现期能较准确地反映特定频率情况下多变量极端降水要素的风险概率,可为防灾减灾和水利基础设施的设计提供更加合理的理论依据。

本文仅选用了 Archimedean Copula 函数作为极端降水要素之间的联结函数,其他 Copula 函数的拟合效果,及其相应重现期问题有待进一步研究。此外,对于区域尺度的面雨量重现期的分析,可参考本文方法,进行多站点极端降水的重现期计算,并根据计算结果得到一定重现期条件下的区域面雨量,从而实现从点到面的扩展。

参考文献:

- [1] 鲁菁,张玉虎,高峰,等.近40年三江平原极端降水时空变化特征分析[J].水土保持研究,2019,26(2):272-282.
- [2] 关铁生,姚惠明,许钦,等.辽河区极端暴雨特性及其天气成因分析[J].水利水运工程学报,2015(2):18-25.
- [3] 张俊,杨文发,卓思佳,等.金沙江流域极端强降水分布拟合[J].人民长江,2017,48(17):58-61.
- [4] YIN Z E, YIN J, XU S Y, et al. Community-based scenario modelling and disaster risk assessment of urban rainstorm waterlogging [J]. Journal of Geographical Sciences, 2011, 21(2): 274-284.
- [5] 熊立华,郭生练,肖义,等. Copula 联结函数在多变量水文频率分析中的应用[J].武汉大学学报(工学版),2005,38(6):16-19.
- [6] 许月萍,童杨斌,富强,等.几种 Copulas 模拟不同历时降雨量的影响分析[J].浙江大学学报(工学版),2009,43(6):1107-1111.
- [7] LI H S, WANG D, SINGH V P, et al. Non-stationary frequency analy-

- sis of annual extreme rainfall volume and intensity using Archimedean copulas: A case study in eastern China [J]. *Journal of Hydrology*, 2019, 571(1): 114–131.
- [8] 冯介玲, 祝好. 基于 Copula 模型的暴雨综合指标分析——以万州为例[J]. *气象科技进展*, 2021, 11(2): 58–62.
- [9] SHIAU J T. Fitting drought duration and severity with two-dimensional Copulas [J]. *Water Resources Management*, 2006, 20(5): 795–815.
- [10] 史黎翔, 宋松柏. 基于 Copula 函数的两变量洪水重现期与设计值计算研究[J]. *水力发电学报*, 2015, 34(10): 27–34.
- [11] SALVADORI G, DE M C, DURANTE F. On the return period and design in a multivariate framework [J]. *Hydrology and Earth System Sciences*, 2011, 15(11): 3293–3305.
- [12] CORBELLA S, STRETCH D D. Multivariate return periods of sea storms for coastal erosion risk assessment [J]. *Natural Hazards and Earth System Sciences*, 2012, 12(8): 2699–2708.
- [13] 陈子燊, 刘占明, 赵青. 洪水峰量联合分布的4种重现水平对比[J]. *中山大学学报(自然科学版)*, 2018, 57(1): 130–135.
- [14] QIANG Z, LI J, SINGH V P, et al. Copula-based spatio-temporal patterns of precipitation extremes in China [J]. *International Journal of Climatology*, 2013, 33(5): 1140–1152.
- [15] 翟盘茂, 潘晓华. 中国北方近50年温度和降水极端事件变化[J]. *地理学报*, 2003, 058(0z1): 1–10.
- [16] BONSAI B R, ZHANG X, VINCENT L A, et al. Characteristics of daily and extreme temperatures over Canada [J]. *Journal of Climate*, 2010, 14(9): 1959–1976.
- [17] 罗梦森, 熊世为, 梁宇飞. 区域极端降水事件阈值计算方法比较分析[J]. *气象科学*, 2013, 33(5): 549–554.
- [18] GENEST C, RIVEST L P. Statistical inference procedures for bivariate Archimedean Copulas [J]. *Journal of the American Statistical Association*, 1993, 88(423): 1034–1043.
- [19] 方彬, 郭生练, 肖义, 等. 年最大洪水两变量联合分布研究[J]. *水科学进展*, 2008, 19(4): 505–511.
- [20] YUE S, OUARDA T B M J, BOBÉE, et al. The Gumbel mixed model for flood frequency analysis [J]. *Journal of Hydrology*, 1999, 226(3–4): 88–100.
- [21] AKAIKE H T. A new look at the statistical model identification [J]. *Automatic Control IEEE Transactions on*, 1974, 19(6): 716–723.
- [22] SALVADORI G, MICHELE C D. Frequency analysis via copulas: theoretical aspects and applications to hydrological events [J]. *Water Resources Research*, 2004, 40(12): 1–17.
- [23] 陈子燊, 路剑飞, 于吉涛. 基于非对称 Archimedean Copula 的三变量风浪重现水平分析[J]. *海洋通报*, 2017, 36(6): 631–637.
- [24] SALVADORI G, MICHELE C D. Multivariate multiparameter extreme value models and return periods: a copula approach [J]. *Water Resources Research*, 2010, 46(10): 219–233.
- [25] XU H S, XU K, LIAN J J, et al. Compound effects of rainfall and storm tides on coastal flooding risk [J]. *Stochastic Environmental Research and Risk Assessment*, 2019, 33(7): 1249–1261.

Analysis of the Joint Probability Characteristics of Bivariate Extreme Precipitation Based on Copula Function

LIU Ping, LI Fan, PAN Qixin, LI Chunxiu

(College of Hydraulic Science and Engineering of Yangzhou University, Yangzhou 225009, China)

Abstract: Based on the daily rainfall data for many years, the 95th percentile value of precipitation was computed and used as the threshold value of extreme precipitation events. Archimedean Copulas were used to construct the two-dimensional joint distribution function between extreme precipitation volume and annual maximum daily precipitation (P_{95} , P_{\max}), extreme precipitation intensity and annual maximum daily precipitation (I_{95} , P_{\max}), extreme precipitation volume and extreme precipitation contribution rate (P_{95} , R_{95}). The daily precipitation data of Gaoyou Station from 1968 to 2020 was used as a case study to test the effectiveness of the method. Frank Copula is the optimal bivariate joint distribution for (P_{95} , P_{\max}), (I_{95} , P_{\max}), while Clayton Copula function can describe the joint distribution of (P_{95} , R_{95}) better than the other Copula functions within the Archimedean copula family. Compared with the joint return period and co-occurrence return period, the Kendall return period is more accurate in risk analysis for multivariate extreme precipitation elements with a given return period.

Keywords: Copula function; extreme precipitation; two-dimensional joint; return period