

doi: 10.11918/j.issn.0367-6234.2015.04.015

3 维 Copula 函数在降雨特征多变量频率分析中的应用

刘成林^{1,2,3}, 周玉文^{1,3}, 隋 军², 高 琳^{1,3}

(1.北京工业大学 建筑工程学院,100124 北京;2.广州市市政工程设计研究院,510060 广州;
3.北京市水质科学与水环境恢复重点实验室,100124 北京)

摘要: 鉴于径流数据缺乏且难以长期监测而降雨数据相对完整,通常假定降雨和径流同频率,采用设计降雨进行水文分析计算,但此方法很难真实全面地反映降雨变化特征.为此,提出一种基于3维 Copula 函数的降雨特征多变量频率分析方法.首先利用降雨强度法将连续的降雨时间序列分割成若干个降雨事件,采用年最大值法取样,统计出表征雨量的特征变量,然后引入3维 Copula 函数构建降雨特征3变量联合概率模型,并以广州1961~2012年历史降雨数据为例进行分析.结果表明,基于3维 Copula 函数的多变量分析方法计算简单、可靠性高,可以进行3种不同降雨特征变量的组合分析,得到各种不同量级变量的遭遇概率和条件概率,能够更全面地反映降雨特征并更好地满足水文分析计算需求.

关键词: 降雨特征;3维 Copula 函数;多变量分析;水文

中图分类号: TU992 **文献标志码:** A **文章编号:** 0367-6234(2015)04-0087-06

Research of methodology of multivariate analysis of design storm based on 3-copula function

LIU Chenglin^{1,2,3}, ZHOU Yuwen^{1,3}, SUI Jun², GAO Lin^{1,3}

(1.College of Architecture and Civil Engineering, Beijing University of Technology, 100124 Beijing, China;
2. Guangzhou Municipal Engineering Design & Research Institute, 510060 Guangzhou, China;
3.Key Laboratory of Beijing for Water Quality Science and Water Environment Recovery Engineering, 100124 Beijing, China)

Abstract: There is an underlying assumption that run-off and rainfall in a given urban catchment are equivalent and, further, to use design rainfall depth as a proxy for run-off in hydrological analyses and calculations. However, when employing this approach, it is difficult to accurately and fully reflect the variability in rainfall characteristics. To address this issue, a method for the copula-based multivariate frequency analysis of rainfall characteristics was proposed by using historical rainfall data (1961-2012) from Guangzhou city. First, continuous rainfall time series were divided into individual rainfall events using the rainfall intensity method. Then the characteristic variables of rainfall were calculated by sampling using the annual maximum method. Finally, a three-dimensional copula was introduced to build a multivariate joint probability distribution model of rainfall characteristics. The results show that the copula-based multivariate analysis is easy to implement and provides reliable results. This approach can be used to analyse the conditional probabilities of variables for different orders of magnitude. It can fully reflect rainfall characteristics, which serve an important reference for urban flood control and drainage planning.

Keywords: rainfall character; 3-copula function; 3-variate analysis; hydrology

我国大部分城市的洪水主要由暴雨形成,与

通过流量资料估算设计洪水相比,根据暴雨资料推求设计洪水是一种间接方法^[1],但在实际工作中经常遇到所在地点流量系列太短无法直接推求设计洪水的现象.降雨监测较为系统,数据序列长且完整,因此,通常基于降雨开展水文分析,进而确定城市洪涝系统设施规模.一般地,一场降

收稿日期: 2014-04-19.

基金项目: 国家重大水专项(2013ZX07304-001).

作者简介: 刘成林(1978—),男,博士研究生;

周玉文(1952—),男,教授,博士生导师.

通信作者: 刘成林, Liucl@gzmedri.com.

雨通常可用雨量、历时和雨强等特征量以及雨量在时空上的变化来反映,设计降雨是地表径流计算的重要参数,主要包括以下特征要素:频率(重现期)、雨量(设计降雨量、次降雨量、降雨峰值)、历时(设计历时、总历时、峰值位置)。目前,暴雨强度公式是普遍采用的设计降雨计算方法,是基于降雨量的单变量分析,不具备计算影响洪涝设施过流和调蓄能力的“降雨峰值”和“总降雨量”等特征量的功能。近几十年来,人们已经认识到采用单变量极值分布进行频率分析研究暴雨事件存在一定的局限性,开始探讨利用两变量联合分布描述暴雨不同特征值之间的相互关系,以更加全面地描述整个暴雨事件^[2],但用 3 变量及以上的联合分布描述暴雨事件的研究还鲜有报道。

Copula^[3-4]是求解多变量概率问题优良的数学工具,该函数可将多个随机变量的边际分布连接起来得到其联合分布,是研究与变量尺度无关的相关性度量的一种途径,同时可以基于给定的边际分布构造联合分布。近几年,Copula 函数理论广泛应用于降雨^[2,5-7]、洪水^[8-9]、干旱^[10]和枯

水^[11]的多特征属性频率分析等问题。为此,以降低降雨特征量(设计降雨量、次降雨量和降雨峰值)频率研究为背景,引入 3 维 Copula 函数构建降雨特征变量的联合概率分布模型,研究不同量级降雨特征变量的遭遇概率及条件概率,以期在城市防洪排涝提供科学参考。

1 研究方法

1.1 Copula 函数简介

Copula 函数是用来描述变量间相依关系的函数,不限定变量的边际分布。通过 Copula 模型可以将 k 个任意形式的边际分布连接起来,生成一个多变量联合概率分布模型。Copula 函数的原理可参考文献^[3-4,8]。Copula 函数主要有 3 种类型,即椭圆型、阿基米德型和二次型。目前,在水文领域应用最广泛的是 Archimedena Copula 函数,其常见的 3 维 Copula 函数主要有 Clayton copula、Gumbel-Hougaard (GH) copula、Ali-Mikhail-Haq (AMH) copula 和 Frank copula (见表 1)。

表 1 3 维 Copula 函数

Copula 函数	表达式
Clayton Copula	$C(u_1, u_2, u_3) = (u_1^{-\theta} + u_2^{-\theta} + u_3^{-\theta} - 2)^{-1/\theta}, \theta \geq 0$
GH Copula	$C(u_1, u_2, u_3) = \exp\{-[(-\ln u_1)^\theta + (-\ln u_2)^\theta + (-\ln u_3)^\theta]^{1/\theta}\}, \theta \in [1, \infty)$
AMH Copula	$C(u_1, u_2, u_3) = u_1 u_2 u_3 / [1 - \theta(1 - u_1)(1 - u_2)(1 - u_3)], \theta \in [-1, 1)$
Frank Copula	$C(u_1, u_2, u_3) = -\frac{1}{\theta} \ln\left[1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)(e^{-\theta u_3} - 1)}{(e^{-\theta} - 1)^2}\right], \theta \in R$

θ 为 Copula 函数的参数; u_1, u_2, u_3 为边际分布函数。

1.2 参数估计及拟合检验

Copula 函数不限定变量的边际分布,确定适宜的边际分布是应用 Copula 函数的第一步。选用 3 种在水文频率分析中常用的分布线型(P-III 型分布曲线(P3)、对数正态分布曲线(LN2)和广义极值分布曲线(GEV))分别对“设计降雨量”、“降雨峰值”和“次降雨量”进行曲线拟合,采用概率点据相关系数检验法(PPCC)、拟优平方和准则法(RMSE)和拟优绝对值准则法(MAE)3 种拟合优度检验方法确定出与各变量数据系列拟合效果最好的边际分布线型^[12]。

Copula 函数的参数估计主要有非参数法和参数法,其中非参数法要求有明确的 Kendall's τ 与 Copula 参数 θ 的表达式,适用于单参数 2 维 Copula 函数的参数确定。而参数法相对比较灵活,本文采用参数法中的 Inference of Functions for Margins (IFM)^[13]方法确定 3 维 Copula 函数的参数。

IFM 方法是将对边缘分布的参数与 Copula 函数的参数分别进行估计,该过程由以下步骤完成。

采用极大似然法(ML 方法)^[14]估计边缘分布中的参数,即

$$\begin{aligned} \hat{\varphi}_1 &= \arg \max \sum_{i=1}^n \ln f_1(x_{1i}; \varphi_1), \\ \hat{\varphi}_2 &= \arg \max \sum_{i=1}^n \ln f_2(x_{2i}; \varphi_2), \\ \hat{\varphi}_3 &= \arg \max \sum_{i=1}^n \ln f_3(x_{3i}; \varphi_3). \end{aligned} \quad (1)$$

采用 ML 方法估计 Copula 中的参数

$$\hat{\theta} = \arg \max \sum_{i=1}^n \ln c(F_1(x_{1i}, \hat{\varphi}_1), F_2(x_{2i}, \hat{\varphi}_2), F_3(x_{3i}, \hat{\varphi}_3); \theta). \quad (2)$$

Copula 函数在实际应用中的一个主要问题就是函数形式的选择。Embrechts 等^[15]对不同的 Copula 函数模型进行了比较,发现采用不同形式的 Copula 函数可能导致不同的分析结果,因此,

选择合适的 Copula 函数尤为重要. 检验与评价 Copula 函数的方法较多,本文采用 AIC 信息准则法(AIC)及离差平方和最小(OLS)准则法^[9]对 Copula 函数的拟合优度进行评价.

1.3 风险概率

给定不同的变量条件可以得到不同的条件概率分布. 主要考虑以下两种条件风险概率.

1) 给定 $X_3 = x_3$, 条件概率分布函数可表示为

$$F(x_1, x_2 | X_3 = x_3) = \frac{\partial F(x_1, x_2, x_3) / \partial x_3}{f_{x_3}(x_3)} = \frac{C(U_1 \leq u_1, U_2 \leq u_2 | U_3 = u_3)}{\frac{\partial C(u_1, u_2, u_3)}{\partial u_3}} \quad (3)$$

式中 $f_{x_3}(x_3)$ 为 x_3 的概率密度函数,其他符号意义同前.

2) 给定 $X_3 \leq x_3$, 条件概率分布函数可表示为

$$F(x_1, x_2 | X_3 \leq x_3) = \frac{F(x_1, x_2, x_3)}{F(x_3)} = \frac{C(u_1, u_2, u_3)}{u_3} \quad (4)$$

2 研究区域概况

广州地处广东省东南部,珠江三角洲北缘,范围在东经 $112^{\circ}57' \sim 114^{\circ}3'$,北纬 $22^{\circ}26' \sim 23^{\circ}56'$,濒临南海,毗邻港澳. 广州地处亚热带沿海,属海洋性亚热带季风气候,全年平均气温 21.97°C ,平均相对湿度 68%,无霜期 300~341 d,日照时数为 1 571~2 053 h,市区年降雨量在 1 700 mm 以上. 全年中,4~6 月为雨季,7~9 月天气炎热,多台风.

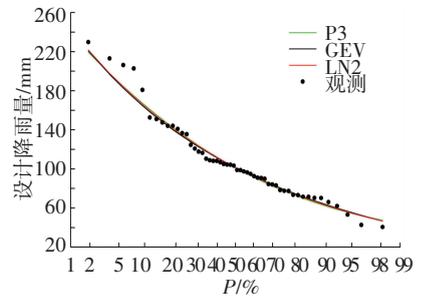
以广州市天河区五山站 1961~2012 年逐分钟自记数据为例,采用降雨强度法^[16]将连续的降雨时间序列分割成降雨事件,按照“降雨强度小于 0.1 mm/h ,且持续时间超过 10 h”的标准进行降雨数据分割,将 1961~2012 年的降雨时间序列分割为 7 833 场降雨事件,然后采用年最大值法进行取样. 以设计历时 720 min (即 12 h) 为例,取样样本 52 个,进而统计“设计降雨量”、“次降雨量”和“降雨峰值”3 个表征雨型的特征变量,计算 3 者的联合概率分布和各条件下的遭遇概率分布.

3 结果分析

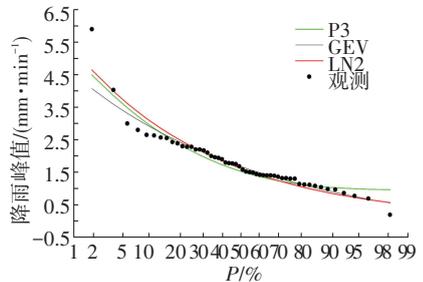
3.1 边际分布确定

图 1 为“设计降雨量”、“降雨峰值”和“次降雨量”边际分布的拟合结果. 可以看出,所采用的 3 种分布线型与实测点据吻合较好,不同线型稍有差距,但总体上看均具有较好的拟合效果.

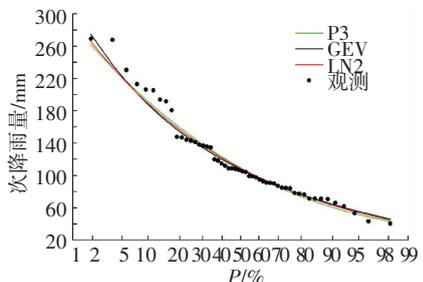
为了定量评价各分布线型的拟合效果,分别计算了各分布线型对于不同观测数据的 RMSE、MAE 及 PPCC,结果见表 2. 对设计降雨而言, LN2 分布具有最小的 RMSE 值,而 GEV 分布具有最小的 MAE 值和最大的 PPCC 值. 综合考虑,选定 GEV 分布作为数据的分布线型;对降雨峰值来说,GEV 分布具有最小的 MAE 值,而 LN2 分布具有最小的 RMSE 值和最大的 PPCC 值,因此,选定 LN2 分布作为降雨峰值数据的分布线型; GEV 分布对次降雨量具有最小的 MAE、RMSE 值和最大的 PPCC 值,所以,选定 GEV 分布作为次降雨量数据的分布线型.



(a) 设计降雨量



(b) 降雨峰值



(c) 次降雨量

图 1 设计降雨量、降雨峰值和次降雨量边际分布拟合

表 2 边际分布拟合优度检验

变量	边际分布	RMSE	MAE	PPCC
设计降雨量	P3	6.996	4.621	0.988
	GEV	7.095	4.208	0.990
	LN2	6.876	4.306	0.989
降雨峰值	P3	0.273	0.162	0.955
	GEV	0.282	0.114	0.952
	LN2	0.250	0.140	0.960
次降雨量	P3	8.873	6.431	0.988
	GEV	8.786	5.582	0.990
	LN2	8.930	5.954	0.990

3.2 Copula 函数参数估计及拟合检验分析

采用阿基米德型 Copula 家族中应用较广泛的 Frank Copula 和 AMH Copula 函数分别构建降雨特征变量的联合分布. 一般地, 不同形式的 Copula 函数对于变量的相关性有不同的要求, Frank Copula 函数对正、负相关关系的随机变量均适用; AMH Copula 函数适用于弱相关关系的随机变量^[13].

为检验 Copula 函数的拟合精度, 首先比较各个观测点对 (x_i, y_i, z_i) 的经验累积概率和理论累积概率的一致性^[9,13]. 图 2 为经验频率与不同 Copula 函数理论频率的拟合图, 可以看出, 经验频率点和理论频率点均分布在 45° 线附近, 总体上看, 经验频率点和理论频率点拟合较好. Frank Copula 函数的拟合精度(图 2(b)) 优于 AMH Copula 函数(图 2(a)).

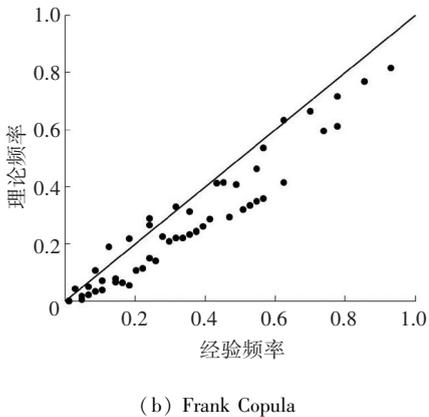
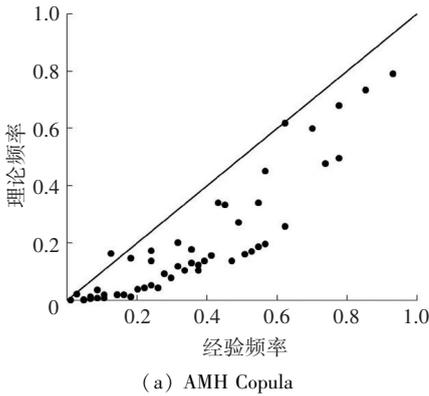


图 2 经验累积概率与理论累积概率比较

为了进一步确定“设计降雨量”、“降雨峰值”和“次降雨量”拟合最好的 Copula 函数, 利用 MSE、AIC 和 OLS 进行拟合优度评价. 从表 3 可以看出, Frank Copula 函数的 3 个评价指标值均小于 AMH Copula 函数, 表明 Frank Copula 函数对 3 变量的拟合效果最佳. 所以, 选取 Frank Copula 函数作为“设计降雨量”、“降雨峰值”和“次降雨量”的联合分布函数.

表 3 3 维 Copula 函数拟合检验结果

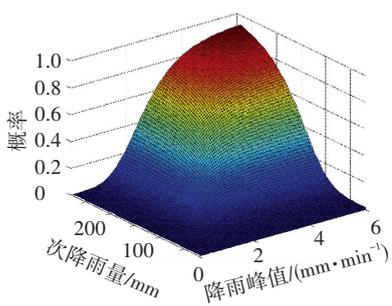
Copula 函数	MSE	AIC	OLS
AMH Copula	0.037	-169.347	0.193
Frank Copula	0.011	-234.102	0.103

3.3 风险概率分析

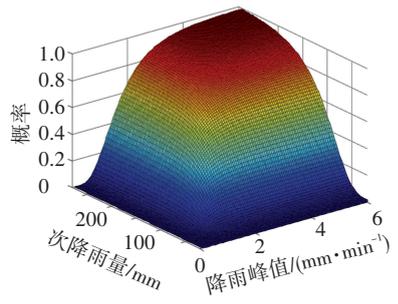
运用 Frank Copula 函数计算“设计降雨量”、“降雨峰值”和“次降雨量”的联合分布函数, 根据 1.3 节公式计算条件风险概率. 分别以 100 a 重现期的“设计降雨量”、“降雨峰值”和“次降雨量”为条件, 计算其他两两变量组合的风险概率. 从图 3 可以看出, 与单变量等于某一设计值条件下的遭遇概率分布(图 3(a)、(c)、(e))相比, 对应条件下单变量小于等于某一设计值下的遭遇概率分布图(图 3(b)、(d)、(f)) 显得“高”、“胖”, 说明小于等于某一设计值条件下的风险概率出现的可能性更大.

以设计降雨条件为例, 给出了几种重现期下降雨峰值和次降雨量遭遇的风险概率. 表 4 为设计降雨量等于某一设计值(100, 50, 20 和 10 a)的条件下, 降雨峰值和次降雨量同时小于等于某一重现期设计值的概率. 可以看出, 在设计降雨不变的情况下, 降雨峰值和次降雨量遭遇的风险概率随着重现期的减小而降低, 当降雨峰值和次降雨量设计值不变时, 其遭遇的风险概率随着设计降雨重现期的减小而上升. 具体来说, 当设计降雨等于 100 一遇时, 小于等于 100 a 一遇降雨峰值和次降雨量同时遭遇的概率为 92.18%, 而小于等于 10 a 一遇同时遭遇的概率仅为 48.44%. 表 5 为设计降雨小于等于某一设计值(100, 50, 20 和 10 a)的条件下, 降雨峰值和次降雨量同时小于等于某一重现期设计值的概率. 对比表 4 和表 5 可知, 两种条件下的风险概率变化具有相似性, 但设计降雨小于等于某一设计值条件下的风险概率出现的可能性更大.

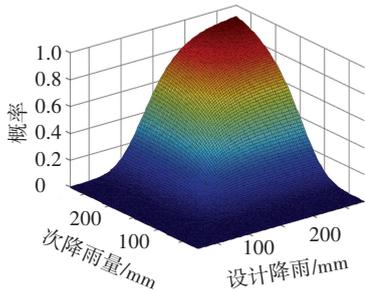
从以上分析可以看出, 设计降雨量、降雨峰值和次降雨量 3 变量之间具有较强的关联性, 用多变量联合分布描述暴雨不同特征值之间的相互关系可以更全面地描述整个暴雨事件, 细化水文分析前端输入, 提高分析精度. 具体地说, 如进行 50 a 一遇水文分析时, 根据表 4, 除给定设计降雨量值(218.9 mm)外, 可根据城市等级及对安全度的需要, 结合自身经济条件, 同时选取某重现期的降雨峰值以及次降雨量作为输入, 如 50 a 一遇的降雨峰值 4.6 mm/min 或 50 a 一遇次降雨量 271.6 mm 来确定洪涝设施规模.



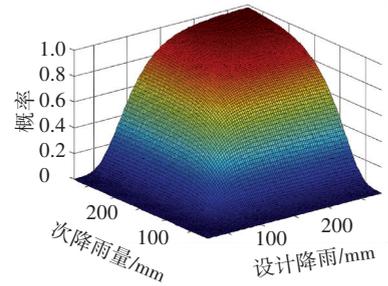
(a) 设计降雨量 = 243.15 mm (100 a)



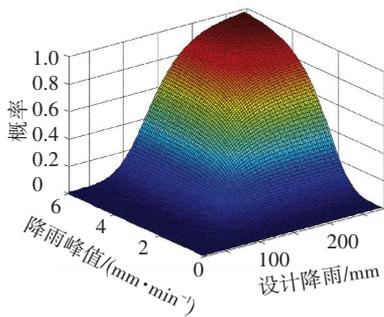
(b) 设计降雨量 \leq 243.15 mm (100 a)



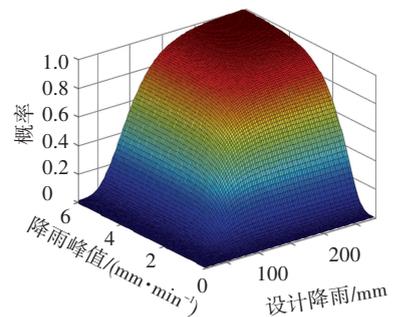
(c) 降雨峰值 = 5.28 mm/min (100 a)



(d) 降雨峰值 \leq 5.28 mm/min (100 a)



(e) 次降雨量 = 311.40 mm (100 a)



(f) 次降雨量 \leq 311.40 mm (100 a)

图 3 降雨特征变量组合的条件风险概率

表 4 设计降雨量等于某一设计值的条件风险概率

设计降雨量/mm	重现期/a	降雨峰值/(mm·min ⁻¹)	重现期/a	次降雨量/mm	重现期/a	风险概率/%
243.2	100	5.3	100	311.4	100	92.18
243.2	100	4.6	50	271.6	50	85.10
243.2	100	3.7	20	222.6	20	67.55
243.2	100	3.1	10	187.8	10	48.44
218.9	50	5.3	100	311.4	100	92.48
218.9	50	4.6	50	271.6	50	85.64
218.9	50	3.7	20	222.6	20	68.49
218.9	50	3.1	10	187.8	10	49.52
187.0	20	5.3	100	311.4	100	93.34
187.0	20	4.6	50	271.6	50	87.16
187.0	20	3.7	20	222.6	20	71.22
187.0	20	3.1	10	187.8	10	52.77
162.8	10	5.3	100	311.4	100	94.57
162.8	10	4.6	50	271.6	50	89.40
162.8	10	3.7	20	222.6	20	75.45
162.8	10	3.1	10	187.8	10	58.11

表 5 设计降雨量小于等于某一设计值的条件风险概率

设计降雨量/mm	重现期/a	降雨峰值/(mm·min ⁻¹)	重现期/a	次降雨量/mm	重现期/a	风险概率/%
243.2	100	5.3	100	311.4	100	98.13
243.2	100	4.6	50	271.6	50	96.29
243.2	100	3.7	20	222.6	20	90.99
243.2	100	3.1	10	187.8	10	83.41
218.9	50	5.3	100	311.4	100	98.18
218.9	50	4.6	50	271.6	50	96.40
218.9	50	3.7	20	222.6	20	91.23
218.9	50	3.1	10	187.8	10	83.76
187.0	20	5.3	100	311.4	100	98.35
187.0	20	4.6	50	271.6	50	96.71
187.0	20	3.7	20	222.6	20	91.90
187.0	20	3.1	10	187.8	10	84.79
162.8	10	5.3	100	311.4	100	98.59
162.8	10	4.6	50	271.6	50	97.18
162.8	10	3.7	20	222.6	20	92.93
162.8	10	3.1	10	187.8	10	86.43

4 结 论

1) 基于 3 维 Copula 函数的降雨多变量分析方法, 根据历史降雨数据, 利用 Copula 联结函数, 实现了 3 变量的联合分析, 有效解决了单变量分析方法的不足, 实现了对降雨事件特征真实全面的反映。

2) 实际应用中, 可基于设计降雨量确定设施标准, 利用三者的组合风险概率复核次降雨量和降雨峰值的失效概率, 为更科学合理地分析、评估和确定城市防洪排涝设施能力提供有效支撑。

3) 如何基于地区特征、安全性和经济性的定量分析, 优化降雨变量组合, 科学合理地确定防洪排涝设施标准, 是该方法需进一步研究的重点。

参 考 文 献

- [1] ZHANG N, GUO S L, XIAO Y, et al. Design storm method based on bivariate joint distribution [J]. *Water Power*, 2008, 34(1): 18-21.
- [2] FONTANAZZA C M, FRENI G, LA LOGGIA G, et al. Uncertainty evaluation of design rainfall for urban flood risk analysis [J]. *Water Science & Technology*, 2011, 63(11): 2641-2650.
- [3] SALVADORI G, DE MICHELE C. Frequency analysis via copulas: theoretical aspects and applications to hydrological events [J]. *Water Resources Research*, 2004, 40(12): W12511.
- [4] NELSEN R. An introduction to copulas [M]. New York: Springer Verlag, 2006.
- [5] ZHANG L, SINGH V P. Gumbel-Hougaard copula for trivariate rainfall frequency analysis [J]. *Journal of Hydrologic Engineering*, 2007, 12(4): 409-419.
- [6] 武传号, 黄国如, 吴思远. 基于 Copula 函数的广州市短

历时暴雨与潮位组合风险分析 [J]. *水力发电学报*, 2014, 33(2): 33-41.

- [7] ZHANG Q, 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.
- [8] ZHANG L, SINGH V P. Bivariate flood frequency analysis using the copula method [J]. *Journal of Hydrologic Engineering*, 2006, 11(2): 150-164.
- [9] 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.
- [10] SHIAU J T. Fitting drought duration and severity with two-dimensional copulas [J]. *Water Resources Management*, 2006, 20(5): 795-815.
- [11] YU K X, XIONG L, GOTTSCHALK L. Derivation of low flow distribution functions using copulas [J]. *Journal of Hydrology*, 2014, 508: 273-288.
- [12] 李兴凯, 陈元芳. 暴雨频率分布线型优选方法的研究 [J]. *水文*, 2010, 30(2): 50-53.
- [13] 谢华, 罗强, 黄介生. 基于三维 copula 函数的多水文区丰枯遭遇分析 [J]. *水科学进展*, 2012, 23(2): 186-193.
- [14] MYUNG I J. Tutorial on maximum likelihood estimation [J]. *Journal of Mathematical Psychology*, 2003, 47: 90-100.
- [15] EMBRECHTS P, HÖING A, JURI A. Using copulae to bound the value-at-risk for functions of dependent risks [J]. *Finance and Stochastics*, 2003, 7(2): 145-167.
- [16] POWELL D N, AZIZ N M, KHAN A A. Impact of new rainfall patterns on the design of hydraulic structures [C]// *Design of Hydraulic Structures*. World Environmental and Water Resource Congress. Omaha: [s.n.], 2006: 1-10. (编辑 刘 彤)