

The Return Period Analysis of Beijing Heavy Rainfall Based on Copula

Weihua Cao

Institute of Urban Meteorology, China Meteorological Administration, Beijing 100089, China

基于 Copula 函数的北京强降水重现期分析

曹伟华

中国气象局北京城市气象研究所, 北京, 100089

Abstract

Heavy rainfall is a serious hazard, which lead to losses of urban property and cause casualties. In this paper, based on hourly precipitation data of Guanxiangtai meteorological station of Beijing, bivariate joint distribution of heavy rainfall is established using the observed data of accumulated precipitation and duration. The joint return periods of heavy rainfall are calculated. The results show there is a significant linear relationship between factors of accumulated precipitation and duration. GEV distribution is appropriate for accumulated precipitation and duration, and the joint distribution of the two variables can be characterized by Gumbel Copula, which is more comprehensive in analyzing the frequency characteristics of heavy rainfall events. The research can provide a foundation for further applications in disaster prevention and reduction.

Key words: Heavy rainfall, Copula, Return period, Beijing

摘要 强降水灾害可导致城市财产损失和人员伤亡, 具有严重的致灾风险。本文以北京南郊观象台逐小时观测资料为基础, 针对强降水的两个主要致灾因素——过程降水量和降水持续时间, 建立过程降水量和持续时间的二元联合分布, 实现对强降水的联合重现期计算。结果表明: 过程降水量和持续时间存在良好的线性相关关系, 广义极值分布适用于作为过程降

水量和降水持续时间的边缘分布, Gumbel Copula 函数能够较好的刻画北京南郊观象台强降水过程降水量和持续时间的联合分布。通过过程降水量和降水持续时间作为二元变量分析强降水事件的联合重现期, 能够更加全面的反应强降水事件的频率特征。研究结果为城市强降水过程防灾减灾提供依据。

关键词: 强降水, Copula 函数, 重现期, 北京

1 引言

随着全球气候变化的影响, 城市强降水事件频发, 引起城市内涝、山体滑坡等灾害, 造成了严重的财产损失和人员伤亡。近年来中国的许多城市遭受强降水灾害危害, 北京受强降水灾害影响尤为严重, 2012年7月21日, 北京特大暴雨灾害导致79人死亡, 经济损失近百亿元, 在全国范围内产生了广泛的社会影响。因此开展强降水风险分析与管理具有重要的现实意义。

强降水的重现期分析是其灾害风险管理的一个重要环节。目前, 国内外针对强降水重现期等已开展了许多研究工作, 以往更多利用降水量一个致灾要素分析其频率特征及重现期特性。近年来更多研究指出, 强降水的致灾特性往往与多种因素耦合作用有关, 降水持续时间和过程降水量将共同影响其致灾特性和后果, 而且致灾因素之间往往存在相互依存关系。相对于其他的多元联合分布, Copula 函数对于具有相互依存关系的多维指标的联合分布刻画具有特别的优势。研究表明, Copula 函数不需要变量服从相同的边缘分布, 随机变

量之间不需要满足独立性假设,可以灵活建立多变量联合分布^[1]。目前针对 Copula 函数边缘分布的选择、参数估计及模型检验等问题已展开了诸多分析^[2],其在水文、气象等领域 Copula 函数得到了广泛应用^[3-5],这些研究结果为认识当地的干旱、暴雨等危险性特征提供了参考。国内将基于二元 Copula 函数在干旱、环境等方面开展了一些应用分析^[6-7],然而,对于强降水重现期分析则更多地针对单个致灾因素,整体来看,缺乏从强降水的多个致灾变量耦合的角度展开分析,这并不足以全面的认识强降水的重现期特性。

鉴于目前强降水重现期分析中对于多个致灾指标考虑的不足,本文将以北京南郊观象台站为例,通过考虑强降水持续时间(D)和过程降水量(W)两个典型的致灾指标构建二元 Copula 分布,开展基于 Copula 函数的强降水重现期分析,从而揭示两个致灾因素耦合作用下北京地区强降水重现期特征,为认识城市强降水致灾危险性制定强降水防灾减灾规划提供科学依据。

2 研究方法

2.1 强降水的二元 Copula 分布

Copula 理论由 Sklar^[8]于 1959 年提出,设随机变量 X, Y 的联合分布函数为 $F(x, y)$, 其边缘分布函数分别为 $u = F_1(x)$ 、 $v = F_2(y)$ 且都是连续函数,则存在唯一的二维函数 $C(u, v)$ $u, v \in [0, 1]$, 使得:

$$F(x, y) = C(F_1(x), F_2(y)) \quad (1)$$

Copula 函数 $C(u, v)$ 本质上是一个连接函数,能够全面反映 X, Y 之间的相关性,将多个边缘分布连接在一起构成联合函数,且 Copula 函数与边缘分布的形式无关。Copula 族的类型较多,由于 Archimedean Copula 族参数形式简单、构建方便,已在水文、气象等自然领域被广泛应用。本文主要通过最常用 Archimedean Copula 函数——Gumbel、Clayton 和 Frank,研究强降水事件指标变量的联合分布,其二元 Copula 分布表达形式以及参数范围见表 1 所示,Copula 函数参数估计采用最大

似然估计法实现^[9]。

表 1. 二元 Archimedean Copula 函数及其参数

Copula	$C_\theta(u, v)$
Gumbel	$\exp(-[(-\ln u)^{1/\theta} + (-\ln v)^{1/\theta}]^\theta)$, $\theta \in (0, 1]$
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$, $\theta \in (0, \infty)$
Frank	$-\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right)$, $\theta \in (-\infty, \infty) \setminus \{0\}$

边缘分布的确定是构建强降水变量二元 Copula 函数的重要环节。本文将比较分析气象和水文领域广泛应用的 5 种概率分布拟合强降水的边缘分布,分别为广义极值分布 (Generalized Extreme Value Distribution, GEV)、伽玛分布 (Gamma)、对数正态分布 (Lognormal)、韦伯分布 (Weibull) 和指数分布 (Exponential)。采用最大似然估计法^[9]估计边缘分布参数,并通过 Kolmogorov-Smirnov (KS) 检验^[10]比较不同概率分布的拟合效果,从而确定强降水指标合适的边缘概率分布。

1) 参数估计

基于 n 对强降水观测数据

$\{(d_1, w_1), L, (d_n, w_n)\}$, 本文采用最大似然法

估计模型参数。记 $f(d; \alpha_1)$ 为选定的强降水持

续时间分布密度模型, $g(w; \alpha_2)$ 为选定的过程

降水量分布密度模型,其中 α_1 、 α_2 为模型参

数。记 $C(u, v; \theta)$ 为选定的一个 Copula 函数,其中 θ 为对应参数,记其密度函数为

$c(u, v; \theta) = \partial^2 C(u, v; \theta) / \partial u \partial v$ 。于是,观测数据

$\{(d_1, w_1), L, (d_n, w_n)\}$ 的对数似然函数可表

示为:

$$\sum_{i=1}^n \ln(c(F(d_i; \alpha_1), G(w_i; \alpha_2); \theta)) + \sum_{i=1}^n \ln(f(d_i; \alpha_1)) + \sum_{i=1}^n \ln(g(w_i; \alpha_2))$$

分两步求解对数似然函数的最大值点。第一步：基于降水持续时间数据，通过极大化

$\sum_{i=1}^n \ln(f(d_i; \alpha_1))$ 给出参数 α_1 的估计值 $\hat{\alpha}_1$ ；类

似地，使用过程降水量数据，通过极大化

$\sum_{i=1}^n \ln(g(w_i; \alpha_2))$ 给出参数 α_2 的估计值 $\hat{\alpha}_2$ 。第

第二步：通过极大化

$\sum_{i=1}^n \ln(c(F(d_i; \alpha_1), G(w_i; \alpha_2); \theta))$ 得到参数 θ 的估

计值 $\hat{\theta}$ 。

2) 拟合优度检验

AIC 准则 (Akaike Information Criterion) 是评价选取合适的概率分布的一种依据^[11]。本文根据 AIC 最小准则法判断二维降水变量拟合优劣，选择 AIC 值最小的 Copula 函数作为过程降水量和持续时间的连接函数，从而比较三种 Archimedean Copula 函数对过程降水量和持续时间联合分布的刻画能力。AIC 包括函数的拟合偏差以及模型参数数量带来的模型不可靠性两部分，其表达式如下：

$$MSE = \frac{1}{n-1} \sum_{i=1}^n [(F_e(d_i, w_i) - C(u_i, v_i))^2] \quad (2)$$

$$RMSE = \sqrt{MSE} \quad (3)$$

$$AIC = n \log(MSE) + 2m \quad (4)$$

其中， $F_e(d_i, w_i)$ 为二维强降水变量的经验概率， $C(u_i, v_i)$ 为 Copula 概率， n 是样本个数， m 是参数个数， MSE 是残差平方和。

2.2 重现期估计

重现期对气象、水文等自然灾害的风险量化具有实际意义。重现期定义为大于等于某强度的降水事件平均出现的时间间隔。对于强降水事件而言，令 D 表示降水持续时间， W 表

示过程降水量，基于单变量的降水重现期表示为：

$$T_D = \frac{\omega}{P[D \geq d]} \quad (5)$$

$$T_W = \frac{\omega}{P[W \geq w]} \quad (6)$$

强降水作为一种致灾因子，强降水的致灾特性不仅与单一因素有关，同时降水持续时间与降水量的耦合作用将共同影响了降水事件的致灾强度及致灾特性。因此本文将综合考虑降水持续时间和过程降水量对强降水重现期的作用。对于降水持续时间和过程降水量的双变量重现期，一是讨论两者同时超过某特定值的同现重现期，记为 T_{AND} ，二是讨论两者之一超过某一特定值的联合重现期，记为 T_{OR} ，表达式定义如下^[12]：

$$T_{AND} = \frac{\omega}{P[D \geq d \text{ and } W \geq w]} = \frac{\omega}{1 - u - v + C_{uv}(u, v)} \quad (7)$$

$$T_{OR} = \frac{\omega}{P[D \geq d \text{ or } W \geq w]} = \frac{\omega}{1 - C_{uv}(u, v)} \quad (8)$$

式 (5) — (8) 中， ω 是降水事件的平均时间间隔。

3 实例分析

3.1 数据

本文降水资料采用 2004—2014 年 5 月 1 日—9 月 30 日北京南郊观象台逐小时降水观测资料。首先利用逐小时降水数据识别独立的降水事件。本文主要通过无雨持续时间来判定区分两次连续降水事件，而对无雨持续时间间隔的定义根据地区的特点而有所不同。结合北京地区降水事件的研究结果^[13]，把一次降水之后连续 2h 无降水且降水强度大于 1mm/h 时判定为一次降水过程结束；把一次降水事件开始至结束期间的累积小时数定义为降水持续时间 D (h)，降水时间内的累积雨量作为过程降水量 M (mm)。在降水事件识别的基础上，把过程降水量大于等于 15mm 且平均降水强度大于等于 5mm/h 的降水事件定义为强降水事件。最终提取北京南郊观象台站 2004—2014 年 5 至 9 月期间共 37 次强降水事件 (表 2)，可以看出，最长持续时间为 16h，平均 4.59h，过程降水量最大为 160.2mm，平均为 38.09mm。

表 2. 北京南郊观象台站强降水基本统计特征

变量	最小值	最大值	均值
持续时间(h)	1.00	16.00	4.59
过程降水量(mm)	15.70	160.20	38.09

3.2 变量之间依存关系

从降水持续时间与过程降水量的散点分布(图 1)可以看出,过程降水量随着持续时间增加而增加,两者的 Kendall 秩相关系数是 0.629,达到 0.01 显著性水平。结果表明,过程降水量与降水持续时间两者之间存在着较强的正相关性,适用于应用 Copula 函数建立二元变量的联合分布。

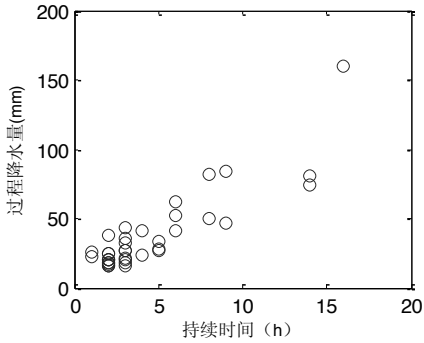


图 1. 降水持续时间与过程降水量的散点分布

3.3 边缘分布分析

表 3 给出了强降水变量在不同分布函数拟合的 KS 检验结果, P 值大于 0.05 时代表通过检验。可以看出,对于持续时间变量,GEV、Lognormal 和 Gamma 分布均可以通过 KS 检验,而 Weibull、Exponential 分布未达到 0.05 显著性水平检验。对于过程降水量变量,GEV、Lognormal 和 Weibull 分布通过 KS 检验,而 Gamma、Exponential 分布未通过检验。对比而言,GEV 分布对降水持续时间和过程降水量的样本数据都拟合较好,该分布 KS 检验的 P 值最大且统计量最小,说明相比而言 GEV 分布最适用于拟合北京地区强降水变量的边缘概率分布。因此,本文选用 GEV 分布作为降水持续时间和过程降水量的边缘分布函数,其表达式为:

$$F(x; \mu, \sigma, \kappa) = \exp \left\{ - \left(1 + \kappa \frac{x - \mu}{\sigma} \right)^{-1/\kappa} \right\} \quad (9)$$

式(9)中, $F(x)$ 是 GEV 分布的累积分布函数, κ, σ, μ 分别代表形状参数,尺度参数和位置参数,且 $\kappa \neq 0$ 。我们进一步采用最大似然法进行参数估计,获得降水持续时间 GEV 分布的参数估计值: $k=0.4650$, $\sigma=1.5630$, $\mu=2.6678$; 过程降水量 GEV 分布的参数估计值: $k=0.7279$, $\sigma=8.8961$, $\mu=23.1483$ 。

表 3. 强降水变量 KS 检验结果

边缘分布	持续时间		过程降水量	
	统计量	P 值	统计量	P 值
GEV	0.166	0.234	0.085	0.931
Lognormal	0.195	0.105	0.148	0.360
Gamma	0.215	0.055	0.221	0.046
Weibull	0.219	0.048	0.171	0.207
Exponential	0.299	0.002	0.338	0.000

3.4 二维 Copula 联合分布

为确定强降水二元 Copula 分布,我们对降水持续时间和过程降水量构成的二维随机样本,采用三种 Copula 函数进行二元联合分布的拟合,参数采用最大似然估计法实现。表 4 给出了 Copula 参数估计值以及拟合优度检验的 AIC 指标值。相比而言, Gumbel Copula 的 AIC 值最小,其次为 Frank Copula,而 Clayton Copula 的 AIC 值最大,这意味着 Gumbel Copula 函数的拟合效果较好。通过二元变量的经验频率与理论频率的散点分布图同样反映出 Gumbel Copula 的拟合效果最好(图 2),其联合累积概率的三维分布如图 3 所示。通过以上分析,本文采用 Gumbel Copula 描述观象台站降水持续时间和过程降水量的二元联合分布。

表 4. 二元 Copula 分布拟合优度检验结果

Copula	Gumbel	Frank	Clayton
AIC	-35.558	-29.401	-8.942
θ	2.291	6.539	0.770

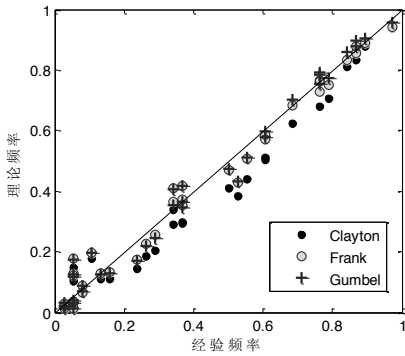


图2. 降水持续时间与过程降水量经验频率与理论频率散点图

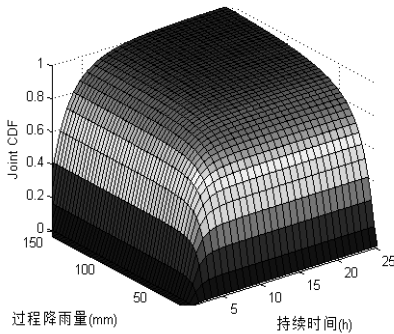


图3. 降水持续时间与过程降水量累积概率三维分布

3.5 重现期计算结果

图4给出降水持续时间和过程降水量的联合重现期 T_{OR} 以及两者的同现重现期 T_{AND} 等值线。为便于比较,表5同时给出了典型强降雨事件的双变量重现期及单变量重现期的计算结果。对比发现,不同的重现期对于降水频率的度量结果并不相同,对于特定事件而言,单变量重现期比双变量同现重现期 T_{AND} 偏小,比联合重现期 T_{OR} 偏大,说明对于两个致灾因素都具有关键作用的一类强降雨事件,采用降水持续时间或者过程降水量之一来刻画其频率特征或揭示其致灾因子的危险性特征将低估或高估其危险性。由于双变量重现期的含义不同, T_{OR} 表示两个变量之一超过某一特定值的重现时间, T_{AND} 表示两个变量同时超过某值的重现时间,而联合重现期 T_{OR} 值低于同现重现期 T_{AND} 值,意味着前者发生的频率比后者

更加频繁。总体而言,在降水持续时间和降水量相互作用并影响强降雨致灾强度时,采用双变量重现期刻画强降雨的致灾因子强度与频率特征将更加客观。

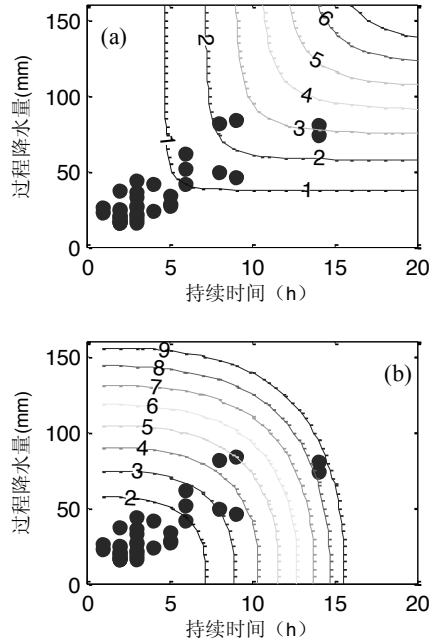


图4. 降水持续时间与过程降水量联合重现期(a)及同现重现期(b)等值线分布

表5. 强降雨重现期计算结果

持续 时间(h)	过程 降水量(mm)	单变量(年)		双变量(年)	
		持续时间	过程降水量	T_{AND}	T_{OR}
14	74	7.24	2.98	8.21	2.84
6	41.2	1.46	1.19	1.89	1.00
8	49.5	2.44	1.59	3.01	1.42
14	80.3	7.24	3.38	8.42	3.17
9	84.2	3.05	3.63	5.01	2.48
16	160.2	9.48	9.39	14.42	7.01
9	46.6	3.05	1.45	3.47	1.37
8	81.7	2.44	3.47	4.45	2.11

4 结论与讨论

本文以北京2004—2014年逐时观测资料为基础提取强降雨案例事件,针对强降雨的过程降水量和降水持续时间两个主要致灾因素,

Risk Analysis and Crisis Response in Big Data Era (RAC-16)

采用 Copula 函数建立强降水的二元联合分布, 并分析强降水事件的双变量重现期特征。本文结论如下:

1) 比较北京地区强降水的二元 Copula 函数发现, Gumbel Copula 函数能够很好的刻画北京观象台的强降水事件的过程降水量和持续时间的二维联合分布特征。

2) KS 检验结果表明, 广义极值分布对北京南郊观象台的过程降水量和降水持续时间的拟合效果较好, 适合作为该站点强降水变量的边缘分布。

3) 强降水案例的重现期计算结果表明, 基于单变量的重现期比双变量同现重现期偏小而比联合重现期偏大。二元 Copula 分布由于刻画了强降水变量之间的相互依存结构, 综合反映变量之间的耦合关系, 能够更加全面的反映强降水事件的重现期特征, 为进一步认识城市强降水的致灾因子危险性提供了重要分析手段。

同时, 由于强降水事件与天气系统、地形等有关, 且具有较强的局地性差异, 下一步将结合更高空间分辨率的资料, 研究北京地区强降水频率的时空分布特征。

Acknowledgements

This study was supported by Beijing Natural Science Foundation (8144046) .

致谢

本研究得到了北京市自然科学基金资助项目 (8144046) 的资助。

参考文献

- [1] Nelsen R B. An introduction to Copulas. 2nd ed. New York: Springer, 2006
- [2] Genest C, Favre A C. 2007. Everything you always wanted to know about copula modeling but were afraid to ask. J Hydrol Eng, 12(4), 347-368
- [3] De Michele C, Salvadori G. 2003. A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. J Geophys Res, 108 (D2): 4067, doi:10.1029/2002JD002534
- [4] Shiau J T. 2006. Fitting Drought Duration and Severity with Two-Dimensional Copulas. Water Resources Management, 20: 795-815
- [5] Zhang L, Singh V P. 2007. Bivariate rainfall frequency distributions using Archimedean copulas. J Hydrol, 332(1): 93-109
- [6] Liu D F, Wang D, Wang Y K. 2016. Eutrophication Hazard Evaluation Using Copula-Cloud. Journal of Risk Analysis and Crisis Response, 6(1):10-14
- [7] 刘晓云, 王劲松, 李耀辉, 等. 2015. 基于 Copula 函数的中国南方干旱风险特征研究. 气象学报, 73(6): 1080-1091
- [8] Sklar A. Fonctions de repartition an dimensions et leurs marges. Publications de l'Institut de Statistique de L'Universit'e de Paris, 1959, 8(1): 229-231
- [9] Zhang L, Singh V P. 2012. Bivariate rainfall and runoff analysis using entropy and copula theories. Entropy, 14: 1784-1812
- [10] Massey F J. 1951. The Kolmogorov-Smirnov test for goodness of fit. J Am Stat Assoc, 46(253): 68-78
- [11] Akaike H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control AC-19 (6), 716-722
- [12] Salvadori G, De Michele C. 2007. On the use of copulas in hydrology: Theory and practice, J Hydrol Eng, 12(4), 369-380
- [13] 李建, 宇如聪, 王建捷. 北京夏季降水的日变化特征. 科学通报, 2008, 59 (7): 829-832