

年最大洪水两变量联合分布研究

方 彬^{1,2}, 郭生练¹, 肖 义¹, 刘 攀¹, 武 见³

(1. 武汉大学水资源与水电工程科学国家重点实验室, 湖北 武汉 430072; 2. 中国南方电网电力调度通信中心, 广东 广州 510623;
3. 黄河勘测规划设计有限公司, 河南 郑州 450000)

摘要: 采用 Von Mises 分布拟合年最大洪水发生时间的概率分布, 采用皮尔逊 III 型分布拟合年最大洪水量级的概率分布, 选用能够较好反映年最大洪水发生时间和量级之间相关结构的 Gumbel Archimedean Copula 函数, 建立两变量联合分布, 并定义和分析条件频率、联合频率和两变量重现期。实例分析表明年最大洪水的两变量分布拟合较好, 可挖掘更多信息, 为洪水设计分析提供了一条新的途径。

关 键 词: 年最大洪水; 两变量联合分布; 发生时间; 洪水量级; Copula 函数

中图分类号: TV122.1 文献标识码: A 文章编号: 1001-6791(2008)04-0505-07

洪水资源化或者洪水资源利用通过合理配置洪水资源, 在保障防洪安全的同时, 努力增加水资源的有效供给, 维系良好生态, 为全面建设小康社会提供有力的防洪抗旱支撑^[1]。利用汛限水位调控洪水资源是洪水资源利用的重要手段之一。研究年最大洪水在一年中的发生时间和量级的分布, 可为汛限水位的制定和优化提供依据。但是现行洪水频率分析都只涉及年最大洪水系列量级的分布类型和参数估计方法, 未曾考虑年最大洪水发生时间。因此, 如何能够同时考虑年最大洪水发生时间和量级这两个因子, 进行联合频率分析, 是一个值得探讨的新课题。目前, 水文多变量分布研究已经逐渐引起人们越来越多的兴趣^[2~5]。Copula 函数可以用来描述水文变量之间的相关性结构, 能够灵活地构造边缘分布为任意分布的水文变量联合分布^[6~8]。本文基于 Gumbel Archimedean Copula^[9,10]建立了年最大洪水发生时间和量级的两变量分布, 首次采用 Von Mises 分布^[11]拟合年最大洪水发生时间的概率分布, 采用皮尔逊 II 型(P-III)分布^[12]拟合年最大洪水量级的概率分布, 并定义和分析两变量分布的重现期。通过实例分析, 研究探讨年最大洪水发生时间和量级的两变量联合分布的实用性。

1 年最大洪水分布

1.1 年最大洪水发生时间分布

将洪水发生时间看作具有周期性变化的矢量, 通过三角函数变换, 使原始数据成为线性资料, 然后计算矢量特征值即平均角和集中度, 并设洪水发生时间服从 Von Mises 分布^[11]。

(1) 平均角和集中度

设计算期内总天数为 L , 第 i 个洪水样本的发生时间为 $D_i(d)$, 令

$$\bar{a} = \sum_{i=1}^N \cos x_i / N \quad b = \sum_{i=1}^N \sin x_i / N \quad (1)$$

式中 N 为样本容量; $x_i = D_i \frac{2\pi}{L}$ 为第 i 个洪水的发生时间(弧度), $0 \leq x_i \leq 2\pi$ 。

计算期发生洪水的平均角 μ 和集中度 r 分别为

收稿日期: 2007-07-24

基金项目: 国家自然科学基金资助项目(50609017); 国际科技合作重点项目(2005DFA20520)

作者简介: 方 彬(1981—), 女, 福建福州人, 博士, 主要从事水文水资源规划设计研究。E-mail: fangbin800@sohu.com

$$\mu = \begin{cases} \arctan \bar{b} / \bar{a} & \bar{a} > 0, \bar{b} > 0 \\ 2\pi + \arctan \bar{b} / \bar{a} & \bar{a} > 0, \bar{b} < 0 \\ \pi + \arctan \bar{b} / \bar{a} & \bar{a} < 0 \\ \pi/2 & \bar{a} = 0, \bar{b} > 0 \\ 3\pi/2 & \bar{a} = 0, \bar{b} < 0 \\ \text{不定} & \bar{a} = 0, \bar{b} = 0 \end{cases} \quad (2)$$

$$r = \sqrt{\bar{a}^2 + \bar{b}^2} \quad 0 \leq r \leq 1 \quad (3)$$

(2) Von Mises 分布

设洪水发生时间随机变量 X 服从 Von Mises 分布, 则其密度函数为

$$f(x) = \frac{\exp[k \cos(x - \mu)]}{2\pi I_0(k)} \quad 0 \leq x \leq 2\pi, 0 \leq \mu \leq 2\pi, k > 0 \quad (4)$$

式中 μ 为位置参数(平均角); 集中度 $r = A(k)$, 尺度参数 $k = A^{-1}(r)$, 可根据函数 A 的逆函数查表求得^[11]。 $I_0(k)$ 为第一类 0 阶变型 Bessel 函数。

1.2 年最大洪水量级分布

P-III 分布^[12] 广泛应用于水文频率分析中。设年最大洪水(如洪峰、几日洪量)量级随机变量 Y 服从 P-III 分布, 其概率密度函数为

$$f(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} (y - \delta)^{\alpha-1} \exp[-\beta(y - \delta)] \quad \alpha > 0, \beta > 0, \delta \leq y < +\infty \quad (5)$$

式中 α 、 β 和 δ 分别为 P-III 分布的形状、尺度和位置参数; $\Gamma(\cdot)$ 为伽玛函数。

1.3 年最大洪水两变量分布

Gumbel Archimedean Copula 是极值型的 Copula 函数, 适合描述水文极值变量的相关结构^[8~10], 数学表达式为

$$C(u, v) = \exp\{-[(-\ln u)^\theta + (-\ln v)^\theta]\}^{1/\theta} \quad \theta \geq 1 \quad (6)$$

式中 u 、 v 为随机变量; $\theta = 1/(1 - \tau)$, τ 为变量 u 、 v 的 Kendall 相关系数; 当 $\theta = 1$ 时, 变量 u 、 v 相互独立, $C(u, v) = uv$ 。

设 $F(x, y)$ 为变量 X 和 Y 的联合分布函数, 则

$$F(x, y) = C(F_X(x), F_Y(y)) = \exp\{-[(-\ln F_X(x))^\theta + (-\ln F_Y(y))^\theta]\}^{1/\theta} \quad \theta \geq 1 \quad (7)$$

式中 $F_X(x) = \int_0^x f_X(x) dx$; $F_Y(y) = \int_0^y f_Y(y) dy$ 。 $f_X(x)$ 和 $f_Y(y)$ 分别为 X 和 Y 的概率密度函数, $F_X(x)$ 为 Von Mises 分布函数, $F_Y(y)$ 为 P-III 分布函数。

$F(x, y) = P(X \leq x, Y \leq y)$ 的物理意义为年最大洪水发生在时间 x 前, 且其量级小于 y 的概率, 其重现期为年。

给定 $X > x$ 时, Y 的条件分布函数为

$$F(y | X > x) = P(Y \leq y | X > x) = \frac{P(X > x, Y \leq y)}{P(X > x)} = \frac{F_Y(y) - F(x, y)}{1 - F_X(x)} \quad (8)$$

同理可得 $Y > y$ 时, X 的条件分布函数为

$$F(x | Y > y) = P(X \leq x | Y > y) = \frac{P(X \leq x, Y > y)}{P(Y > y)} = \frac{F_X(x) - F(x, y)}{1 - F_Y(y)} \quad (9)$$

2 两变量分布的重现期

定义变量 X 和 Y 的联合重现期 T_o 为

$$T_o(x, y) = \frac{1}{P(X > x \text{ or } Y > y)} = \frac{1}{1 - F(x, y)} \quad (10)$$

根据联合分布函数 $F(x, y)$ 、边缘分布函数 $F_X(x)$ 和 $F_Y(y)$, 定义另一种联合重现期 T_a 为

$$T_a(x, y) = \frac{1}{P(X > x, Y > y)} = \frac{1}{1 - F_X(x) - F_Y(y) + F(x, y)}$$

(11)

单变量重现期为

$$T(x) = \frac{1}{1 - F_X(x)} \qquad T(y) = \frac{1}{1 - F_Y(y)}$$

(12)

则有

$$T_o(x, y) \leq \min(T(x), T(y)) \leq \max(T(x), T(y)) \leq T_a(x, y)$$

(13)

3 实例计算分析

清江隔河岩水库流域控制面积 17 000 km², 多年(1951—2005 年)日平均流量 393 m³/s, 资料长度 $N=55$ 年, 选取汛期 5 月 1 日~9 月 30 日(总天数 $L=153$ d)的日流量资料进行年最大洪水取样(洪峰及其发生日期)两变量分布计算, 年最大洪水洪峰及其发生时间的分布参数列于表 1。

表 1 各分布函数的参数

Table 1 Estimated parameters of different distribution functions

Von Mises 分布			P-III分布			两变量联合分布	
μ	r	k	α	β	δ	Kendall τ	θ
2.535	0.443	1.000	3.916	0.001	1.029	0.228	1.296

3.1 拟合效果

采用 Gringorten 公式分别计算年最大洪水发生时间 X 、量级 Y 的经验频率^[13]。设 $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ 为联合观测值组合, 将这些组合按照变量 X 值的升序排列, 通过计算 $x_j \leq x_i, y_j \leq y_i, i, j=1, 2, \dots, N$ 的 (x_j, y_j) 的个数, 按下式得到 (x_i, y_i) 的联合经验频率^[14]:

$$H(x_i, y_i) = P(X \leq x_i, Y \leq y_i) = \frac{m_i - 0.44}{N + 0.12}$$

(14)

式中 m_i 为联合观测值样本中满足 $(x_j \leq x_i, y_j \leq y_i)$ 的联合观测值的个数; N 为样本容量。

图 1 给出了年最大洪水发生时间的经验点据和边缘分布的理论频率曲线, 可见曲线与经验点据拟合较好。从图 1 可知当纵坐标 $x=0$ (对应于汛期开始日期 5 月 1 日)时, 横坐标 $P=P_X(X > x)=1$, 说明年最大洪水发生在 5 月 1 日之后的概率为 1; 当纵坐标 $x=2\pi \approx 6.283$ (对应于汛期结束日期 9 月 30 日)时, 横坐标 $P=P_X(X > x)=0$, 说明年最大洪水发生在 9 月 30 日之后的概率为 0, 频率曲线呈现两段有限的走势。图 2 给出了年最大洪水发生量级的经验点据和边缘分布的理论频率曲线, 可以看出两者拟合较好。

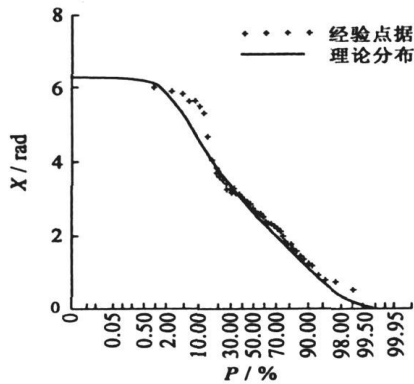


图 1 年最大洪水发生时间 X 分布曲线

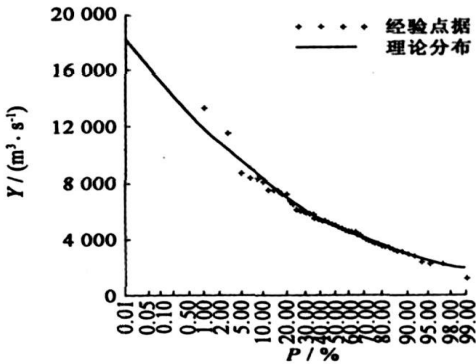


图 2 年最大洪水量级 Y 分布曲线

Fig 1 Frequency curve of AM flood occurrence dates

Fig 2 Frequency curve of AM flood magnitudes

在求出联合经验频率和理论分布值后, 将理论分布值按升序排列, 相对应的联合经验频率值的排序也随之

改变。图 3 给出了顺序统计量 X 和 Y 的联合经验点据和理论分布曲线的拟合图，可以看出两者拟合得较好。将联合经验频率和理论分布值绘成散点图如图 4 所示，可见点据都落在 45° 对角线附近，相关系数为 0.995，因此所采用的理论联合分布是合理的。

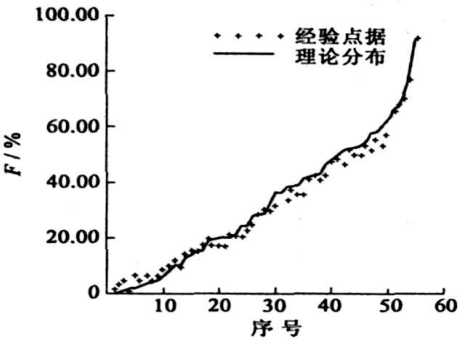


图 3 X 和 Y 的联合经验点据和理论分布曲线

Fig 3 Comparison of empirical and theoretical probabilities of the observed combinations of X and Y

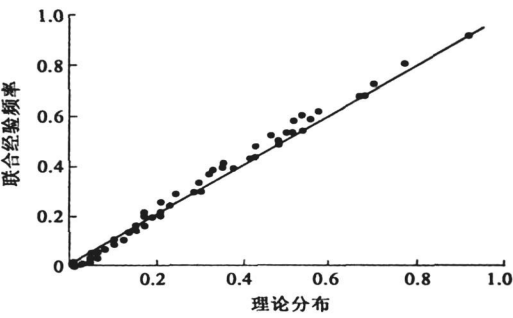


图 4 X 和 Y 的联合经验频率和理论分布的 P-P 图

Fig 4 P-P plot of empirical and theoretical probabilities of the observed combinations of X and Y

为检验洪水频率曲线的拟合效果，采用平均偏差 $Bias$ 和均方误差 $RMSE$ 作为拟合优度指标，计算公式如下

$$Bias = \frac{1}{N} \sum_{i=1}^N (z_i - \hat{z}_i) / z_i \tag{15}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{\hat{z}_i - z_i}{z_i} \right)^2} \tag{16}$$

式中 \hat{z}_i 和 z_i 分别为变量的估计值和实测值， $i=1, 2, \dots, N$ 。

采用 χ^2 检验统计量检验分布的可行性，计算公式如下

$$\chi^2 = \sum_{i=1}^N (o_i - e_i)^2 / e_i \tag{17}$$

式中 o_i 和 e_i 分别为变量的经验频率和估计频率， $i=1, 2, \dots, N$ 。

拟合优度指标和 χ^2 检验统计量计算结果见表 2。从表中可以看出， $Bias$ 和 $RMSE$ 值都较小，说明单变量分布对经验点据的拟合效果较好。在显著性水平 0.005 下，Von Mises 分布、P-III 分布和两变量联合分布都通过了 χ^2 检验。例如 Von Mises 分布的 $\chi^2=4.400 < \chi^2_{1-0.005}(55-6-1) = \chi^2_{0.995}(48) = 76.969$ 。

表 2 拟合优度指标和 χ^2 检验统计量

Table 2 Goodness of fit and χ^2 test statistics

指 标	$Bias$	$RMSE$	χ^2	参数个数 c	$\chi^2_{0.995}(N-c-1)$
Von Mises 分布	-4.378	0.982	0.253	2	82.001
P-III 分布	0.254	0.327	0.903	3	80.747
两变量分布			4.400	6	76.969

3.2 条件频率

表 3 列出了各频率 p 下 X 的条件频率 $P(X > x_p | Y > y_{1\%})$ ，可知在年最大洪水发生量级 $Y > y_{1\%}$ 的条件下，此年最大洪水发生日期在 5 月 27 日之后的概率为 98.45%，进一步可推知此年最大洪水发生日期在 5 月 27 日 ~ 9 月 25 日之间的概率为 $98.45\% - 29.86\% = 68.59\%$ ，发生在 7 月 18 日 ~ 29 日之间的概率为 $81.16\% - 75.29\% = 5.87\%$ 。图 5 和图 6 分别给出了 X 的条件频率 $P(X > x | Y > y)$ 曲线、 Y 的条件频率 $P(Y > y | X > x)$ 曲线，其中虚线分别是 X 、 Y 的单变量频率曲线。

表3 各频率 p 下 X 的条件频率 $P(X > x_p | Y > y_{1\%})$

Table 3 Conditional probabilities of $P(X > x_p | Y > y_{1\%})$

频率 $p/\%$	0.01	0.1	1	10	20	30	40	50	70	90	99
x_p/rad	6.28	6.27	6.09	4.48	3.71	3.27	2.92	2.62	2.02	1.11	0.17
x_p 对应的日期	9月29日	9月28日	9月25日	8月17日	7月29日	7月18日	7月10日	7月2日	6月18日	5月27日	5月4日
$P(X > x_p Y > y_{1\%})/\%$	0.84	6.17	29.86	65.42	75.29	81.16	85.47	88.93	94.36	98.45	99.87

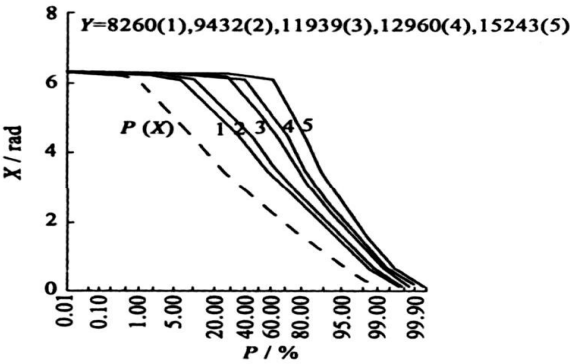


图5 X 的条件频率 $P(X > x | Y > y)$ 曲线
Fig. 5 Conditional probability of X , $P(X > x | Y > y)$

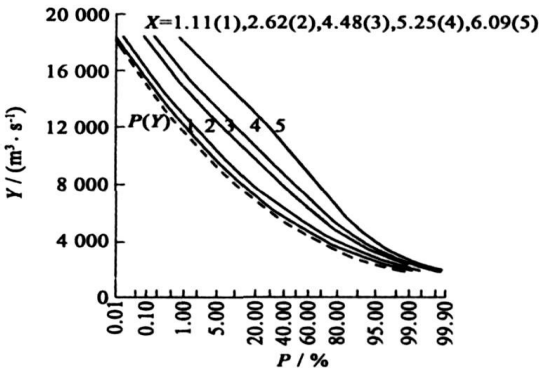


图6 Y 的条件频率 $P(Y > y | X > x)$ 曲线
Fig. 6 Conditional probability of Y , $P(Y > y | X > x)$

3.3 联合重现期

根据年最大洪水发生时间 X 和量级 Y 的边缘分布, 可以推求各种 X 和 Y 组合的 T_o (或 T_a) 以及给定 T_o (或 T_a) 下可能的 X 和 Y 组合。本文计算给定频率 p 的设计值 x_p 、 y_p , 然后计算 (x_p, y_p) 的联合重现期 $T_o(x_p, y_p)$ 和 $T_a(x_p, y_p)$ 。表4给出了计算结果以及 x_p 对应的日期, 可知 $T_o(x_{10\%}, y_{10\%})=6$ 年, 即在8月17日后发生年最大洪水或者发生量级 $Y > y_{10\%}=8260 \text{ m}^3/\text{s}$ 的年最大洪水的联合重现期是6年; $T_a(x_{10\%}, y_{10\%})=28$ 年, 即在8月17日后发生量级 $Y > y_{10\%}=8260 \text{ m}^3/\text{s}$ 的年最大洪水的联合重现期是28年。

表4 联合重现期 $T_o(x_p, y_p)$ 和 $T_a(x_p, y_p)$

Table 4 Joint return period $T_o(x_p, y_p)$ and $T_a(x_p, y_p)$

频率 $p/\%$	$T=\frac{1}{p}/\text{a}$	x_p/rad	x_p 对应的日期	$y_p/(\text{m}^3\cdot\text{s}^{-1})$	T_o/a	T_a/a
1	100	6.09	9月25日	11939	59	335
10	10	4.48	8月17日	8260	6	28
20	5	3.71	7月29日	6987	3	12
50	2	2.62	7月2日	4979	1	3

4 结 论

本文假设年最大洪水发生时间服从 Von Mises 分布、年最大洪水量级服从 P-II 分布, 选用能够较好反映年最大洪水发生时间和量级之间相关结构的 Gumbel Archimedean Copula 函数, 建立了两变量联合分布。两变量联合分布的物理意义为年最大洪水发生在某时间前且其量级小于某值的概率, 其重现期为年。通过实例分析, 研究探讨了两变量联合分布的实用性。结论如下:

- (1) 拟合优度指标和 χ^2 检验统计量表明 Von Mises 分布能较好地拟合年最大洪水发生时间。
- (2) P-P 图和 χ^2 检验都表明年最大洪水发生时间和量级服从两变量联合分布。
- (3) 从两变量联合分布可以推知年最大洪水发生时间(或量级)的条件分布, 可以获得年最大洪水发生时间和量级的各种组合的联合概率和联合重现期。

从两变量联合分布可以知道年最大洪水发生时间和量级的概率分布, 可为水库洪水调度提供信息, 比如在水库汛限水位优化设计中, 可将汛限水位作为年最大洪水发生时间和量级的函数, 这不同于以往的研究思维, 为洪水设计分析开辟了一条新的途径。

参考文献:

- [1] 程殿龙, 尚全民, 万海斌, 等. 以科学精神和积极态度对待洪水资源化[J]. 中国水利, 2004(15): 25—27. (CHENG Dian-long, SHANG Quan-min, WAN Hai-bin, *et al.* Perceive flood water utilization with scientific and positive attitude[J]. China Water Resources 2004 (15): 25—27. (in Chinese))
- [2] 肖 义, 郭生练, 熊立华, 等. 一种新的洪水过程随机模拟方法研究[J]. 四川大学学报(工程科学版), 2007, 39(2): 55—61. (XIAO Yi, GUO Sheng-lian, XIONG Li-hua, *et al.* A new random method for constructing synthetic flood hydrograph[J]. Journal of Sichuan University (Engineering Science Edition), 2007, 39(2): 55—61. (in Chinese))
- [3] 肖 义, 郭生练, 刘 攀, 等. 基于两变量分布的峰量联合分析[J]. 长江科学院院报, 2007, 24(2): 13—16. (XIAO Yi, GUO Sheng-lian, LIU Pan, *et al.* Joint analysis of peak and volume based on bivariate distribution[J]. Journal of Yangtze River Scientific Research Institute, 2007, 24(2): 13—16. (in Chinese))
- [4] 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(工程科学版), 2005, 38(6): 16—19. (XIONG Li-hua, GUO Sheng-lian, XIAO Yi, *et al.* Application of Copulas to multivariate hydrological frequency analysis[J]. Engineering Journal of Wuhan University, 2005, 38(6): 16—19. (in Chinese))
- [5] 肖 义, 郭生练, 刘 攀, 等. 分期设计洪水频率与防洪标准关系研究[J]. 水科学进展, 2008, 19(1): 54—60. (XIAO Yi, GUO Sheng-lian, LIU Pan, *et al.* Seasonal flood frequency analysis and flood prevention standard[J]. Advances in Water Science, 2008, 19(1): 54—60. (in Chinese))
- [6] MICHELE D C, SALVADORI G, CANOSSIM, *et al.* Bivariate statistical approach to check adequacy of dam spillway[J]. Journal of Hydrologic Engineering, 2005, 10(1): 50—57.
- [7] SALVADORI G, MICHELE D C. Frequency analysis via Copulas: Theoretical aspects and applications to hydrological events[J]. Water Resources Research, 2004, 40(12), W12511, doi: 10.1029/2004WR003133.
- [8] FAVRE A C, ADLOUNIS E, PERREAULT L, *et al.* Multivariate hydrological frequency analysis using Copulas[J]. Water Resources Research, 2004, 40(1), W01101, doi: 10.1029/2004WR002456.
- [9] NELSON R B. An introduction to Copulas[M]. New York: Springer, 1999.
- [10] NELSON R B. Dependence and order in families of Archimedean Copulas[J]. Journal of Multivariate Analysis, 1997, 60(1): 111—122.
- [11] 李元生. 方向数据统计[M]. 北京: 中国科学技术出版社, 1998. (LI Yuan-sheng Directional Statistics[M]. Beijing: Science and Technology Press, 1998. (in Chinese))
- [12] 水利部, 能源部. 水利水电工程设计洪水计算规范 SL44-93[M]. 北京: 水利电力出版社, 1993. (Ministry of Water Resources, Ministry of Energy. Regulation for calculating design flood of water resources and hydropower projects SL44-93[M]. Beijing: Hydropower and Electrical Press, 1993. (in Chinese))
- [13] GUO S L. A discussion on unbiased plotting positions for the general extreme value distribution[J]. Journal of Hydrology, 1990, 121(1—4): 33—44.
- [14] YUE S, OUARDA T B M J, BOBÉE B, *et al.* The Gumbel mixed model for flood frequency analysis[J]. Journal of Hydrology, 1999, 226(1—2): 88—100.

Annual maximum flood occurrence dates and magnitudes frequency analysis based on bivariate joint distribution^{*}

FANG Bin^{1, 2}, GUO Sheng-lian¹, XIAO Yi¹, LIU Pan¹, WU Jian³

(1. State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China;

2. Power Dispatching and Communication Center of China Southern Power Grid, Guangzhou 510623, China;

3. Yellow River Engineering Consulting Co., Ltd., Zhengzhou 450000, China)

Abstract: Annual maximum flood occurrence dates and magnitudes both can provide important information for the hydraulic engineering design and the reservoir operation. The existing literatures only consider the distribution of flood magnitudes, but ignores the flood occurrence dates. In this paper, Von Mises distribution and Pearson Type III distribution are used to describe the occurrence dates and magnitudes of annual maximum flood respectively. A bivariate joint distribution with Von Mises distribution and Pearson Type III distribution margins is developed based on the Gumbel Archimedean Copula and used to describe the annual maximum flood series. The approaches for calculating conditional probability, joint probability and bivariate return period are presented. Case study shows that the bivariate joint distribution can fit both occurrence dates and magnitudes of annual maximum flood series well. It can mine more flood information and provide a new way for flood frequency analysis.

Key words: annual maximum flood; bivariate joint distribution; occurrence dates; magnitudes; Copula function

《水科学进展》编辑部启事

《水科学进展》编辑部原地址为: 南京市西康路1号南京水利科学研究院水文水资源研究所《水科学进展》编辑部; 邮政编码: 210024。现地址变为: 南京市广州路225号南京水利科学研究院《水科学进展》编辑部; 邮政编码: 210029。联系电话: 025—85829770/85828537; 传真: 025—85829770; E-mail: skxj@chinajournal.net.cn; 本刊网址: <http://skxjz.nhi.cn>。

^{*} The study is financially supported by the National Natural Science Foundation of China (No. 50609017) and the Chinese International Cooperation Research Fund (No. 2005DFA20520).