

水文频率分析中抽样误差的统计试验研究

唐 林¹,任智慧²

(1.黄河勘测规划设计有限公司,河南 郑州 450003;2.黄河水利科学研究院,河南 郑州 450003)

摘 要:基于统计试验方法,探讨水文频率分析中抽样误差的分布问题。试验结果表明:小样本容量时,设计值抽样误差存在“系统性”偏小,误差均值小于零,随样本容量的增加,偏差逐渐缩小,趋于零均值正态分布;当 C_v 、 C_s 值较大时,抽样误差明显增大,小样本情况下还须额外考虑抽样误差均值的修正。

关键词:水文频率分析;抽样误差;统计试验

中图分类号:P333

文献标识码:A

文章编号:1000-0852(2013)06-0006-05

1 引言

水文频率分析是研究设计洪水的主要途径之一。19 世纪中后期,国内外众多水文学者、专家针对频率分析中线型的选择做了大量的工作,在不同的地区得到不同的适应线型^[1-9]。与此同时,针对频率分析中参数的估计方法也做了大量研究,并取得了丰硕的成果^[10-16]。样本抽样问题同样也是频率分析的重要内容之一,但相关的研究却很少。这主要是由于水文实测系列相对过短,现有资料无法得到可靠的结论。当前很多水文站点资料较短,大江、大河上重要站点中最长系列也仅有 100 多年,相对于江河数百万年的发展历程而言着实太短。基于实际需要,又必须依据仅有的短系列资料估算设计值,其风险是可观的。

据刘光文教授的观点^[17],为保证水文设计成果质量与精度,从水文特征的安全因数取 1.5 出发,假定代表性不足造成的误差约 20%,则容许设计值误差需要控制在 8%~10%以内。众所周知,样本的代表性与样本容量有关,即抽样误差与样本容量关系密切。以往经验表明,样本容量越大,系列代表性越好,抽样误差越小。反之,抽样误差过大,难以保证设计成果质量。因此,研究样本容量与其抽样误差分布之间的关系具有重要的现实意义。

现行中国水利行业标准《水利水电工程设计洪水

计算规范》^[18]中,针对设计洪水估计值的抽样误差,建立有 P-Ⅲ 型分布设计洪水估计值均方差估算公式对其进行安全修正。该方法前提是假定估计值没有系统的偏差,因此仅对均方差进行估值修正。实际上,设计洪水估计值的抽样误差可能存在这种系统性的偏差,即抽样误差为非零均值分布。此外,不同的抽样样本容量,误差大小也有所差异。因此,有必要针对上述问题进行探索研究。

受实测资料长度限制,无法直接采取实测值做较为准确的抽样误差分析。随机分析学科的发展成为解决该问题的重要方法,并在水文工程科学中得到了广泛的应用^[19]。

因此,本文基于随机试验方法,假定水文系列满足 P-Ⅲ 分布线型,统计不同样本容量下设计值的误差分布,分析抽样误差的影响因素。

2 统计试验原理

统计试验方法(即 Monte-Carlo 法),是 20 世纪 40 年代中期,随着电子计算机诞生和科技的快速发展,而提出的一种基于概率统计理论的数值计算方法,是通过人工生成随机数来解决实际问题近似解方法的总称^[20]。现实中有很多问题本身相当复杂,很难用解析的方法进行分析计算,而 Monte-Carlo 法具有很好的数值计算和数值模拟功能,在很多复杂领域中起到相当

重要的作用,成为解决问题的一种新途径^[21]。

Monte-Carlo 法本质上是通过大量的、满足一定分布的随机离散点模拟相应特征的分布过程,其首先需要解决的是均匀分布随机数的生成,其他类型分布的随机数可在均匀分布随机数的基础上通过变换方法或舍选抽样得到。

因舍选抽样方法灵活、使用方便、计算简单,得到广泛应用。本文即采用舍选抽样方法来生成 P- 型分布随机数。

3 试验方法

3.1 线型选择

根据前人经验及研究成果,并参照规范^[18]推荐成果,采用 P- 型曲线作为模拟水文系列的理论分布曲线。

3.2 参数估计

根据水文统计学原理^[22],对于 P- 型分布,矩法的参数估计中均值 Ex 为无偏估计量,离势系数 Cv 、偏态系数 Cs 为近似无偏估计量,从理论上来说矩法的参数估计公式具有无偏性,即没有“系统误差”,因此,可认为该方法的估值误差主要源于样本抽样代表性不足所致。

据此,本文采用矩法做参数估计(均值 Ex 、离势系数 Cv 、偏态系数 Cs),计算公式如下:

$$Ex = \frac{1}{n} \sum_{i=1}^n X_i \quad (1)$$

$$Cv = \sqrt{\frac{1}{n-1} \sum_{i=1}^n \left(\frac{X_i}{Ex} - 1 \right)^2} \quad (2)$$

$$Cs = \frac{\sum_{i=1}^n \left(\frac{X_i}{Ex} - 1 \right)^3}{(n-3)Cv^3} \quad (3)$$

3.3 样本抽样

Monte Carlo 方法生成的样本基本上能够很好的满足独立、随机、同分布要求,因此,可认为该方法对水文系列具有较好的模拟性能。样本抽样具体步骤如下:

(1) 假定一组 P- 型理论分布曲线特征参数,计算特定频率 $P(\%)$ 对应的理论设计值 X_p 。

(2) 依据随机试验原理,确定抽样次数 M 以及样本容量 n ,随机生成 n 个数据,得到样本 U_1^n ;再随机生成 n 个数据,得到样本 U_2^n ;以此类推,直至最后,得到样本 U_M^n 。

(3) 对样本 U_j^n ,采用矩法估计 P- 型分布特征

参数 (Ex_j^n, Cv_j^n, Cs_j^n) , 并计算频率 P 对应的样本估计值 X_{pj}^n 。

(4) 统计样本 X_{pj}^n 估计值相对于理论值 X_p 的抽样误差分布情况。

(5) 变换抽样个数 n ,重复(1)~(4),得到不同抽样个数下抽样误差的分布。

4 评价标准

不失一般性,将抽样误差表达为相对值(即抽样相对误差,以下同)。抽样误差的分布可通过误差的均值、均方差来表示。

(1) 特征均值

$$\Delta_{X_p}^n = \frac{1}{M} \sum_{j=1}^M \left(\frac{X_{pj}^n - X_p}{X_p} \right) \quad (4)$$

(2) 特征均方差

$$\sigma_{X_p}^n = \sqrt{\frac{1}{M-1} \sum_{j=1}^M \left(\frac{X_{pj}^n - X_p}{X_p} - \Delta_{X_p}^n \right)^2} \quad (5)$$

式中: M 为随机生成样本的次数; n 为样本容量; X_{pj}^n 为样本容量为 n 时第 j 次抽样相应设计频率为 P 的设计值; X_p 为设计频率为 P 的理论设计值。

5 结果分析

5.1 试验方案

P- 型分布总体参数设定为:均值 $Ex=1$,离势系数 $Cv=0.2, 0.4, 0.6, 0.8, 1.0$, 偏态系数 Cs 分别取 2 倍及 3 倍的 Cv ,构造 10 组试验方案。按前述试验方法,抽样个数 n 分别取 10、20、30、40、50、100、150、200,对每组抽样方案,进行 $M=10^5$ 次独立重复抽取,从而保证统计结果具有较好的稳定性和代表性。参数估计采用矩法,推求设计频率 $P=1\%$ 的设计值,直接统计设计值的抽样误差,结果见表 1。依据统计结果的离散点值,采用克里希插值法,建立 $Cv \sim n$ 抽样误差特征值分布图,如图 1~4 所示。

5.2 结果分析

(1) 从图表中可以看出:抽样误差均值和均方差随样本个数的增加而减小,说明增加样本数量是降低抽样误差的有效方式。

(2) 不同样本容量下,抽样误差均值小于零,说明由样本推求的设计值总体上较理论设计值偏低。

(3) 抽样误差均值和均方差随样本容量的增加而

表1 统计试验分析成果表
Table 1 The statistical test analysis results

C_v	C_s/C_v	抽样个数 n	10	20	30	40	50	100	150	200
0.2	2.0	期望	-0.8%	-0.5%	-0.4%	-0.4%	-0.3%	-0.2%	-0.2%	-0.1%
		均方差	11.9%	8.8%	7.4%	6.5%	5.9%	4.4%	3.6%	3.1%
	3.0	期望	-1.80%	-1.20%	-0.90%	-0.80%	-0.60%	-0.40%	-0.20%	-0.20%
		均方差	13.10%	9.80%	8.30%	7.40%	6.70%	4.90%	4.10%	3.50%
0.4	2.0	期望	-4.0%	-2.6%	-2.0%	-1.6%	-1.3%	-0.7%	-0.5%	-0.3%
		均方差	20.8%	15.9%	13.4%	11.9%	10.8%	7.9%	6.5%	5.7%
	3.0	期望	-6.40%	-4.10%	-3.00%	-2.40%	-2.00%	-1.10%	-0.70%	-0.50%
		均方差	23.60%	18.20%	15.50%	13.80%	12.60%	9.40%	7.80%	6.80%
0.6	2.0	期望	-7.4%	-4.8%	-3.5%	-2.8%	-2.3%	-1.2%	-0.8%	-0.6%
		均方差	27.6%	21.3%	18.2%	16.2%	14.8%	10.9%	9.1%	8.0%
	3.0	期望	-11.30%	-7.20%	-5.50%	-4.50%	-3.80%	-2.20%	-1.50%	-1.2%
		均方差	31.3%	24.7%	21.3%	19.1%	17.5%	13.3%	11.2%	9.9%
0.8	2.0	期望	-10.9%	-7.0%	-5.3%	-4.2%	-3.6%	-2.0%	-1.4%	-1.1%
		均方差	32.8%	25.8%	22.1%	19.7%	18.0%	13.6%	11.4%	10.1%
	3.0	期望	-15.8%	-10.5%	-8.1%	-6.7%	-5.8%	-3.4%	-2.50%	-2.0%
		均方差	37.6%	30.0%	26.0%	23.5%	21.7%	16.5%	14.0%	12.4%
1.0	2.0	期望	-14.0%	-9.3%	-7.0%	-5.7%	-4.9%	-2.9%	-2.0%	-1.6%
		均方差	37.3%	29.4%	25.6%	22.8%	21.1%	15.9%	13.5%	11.9%
	3.0	期望	-20.5%	-13.6%	-10.8%	-8.90%	-7.70%	-4.80%	-3.50%	-2.80%
		均方差	42.3%	34.6%	30.2%	27.5%	25.3%	19.4%	16.5%	14.6%

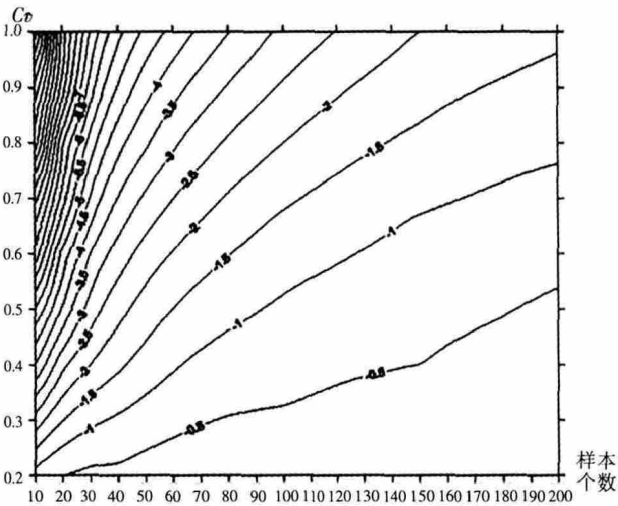


图1 $C_s/C_v=2$, 1%设计值抽样误差均值(%)分布
Fig.1 Mean value (%) distribution of SRE for designed value with 1% frequency and $C_s/C_v=2$

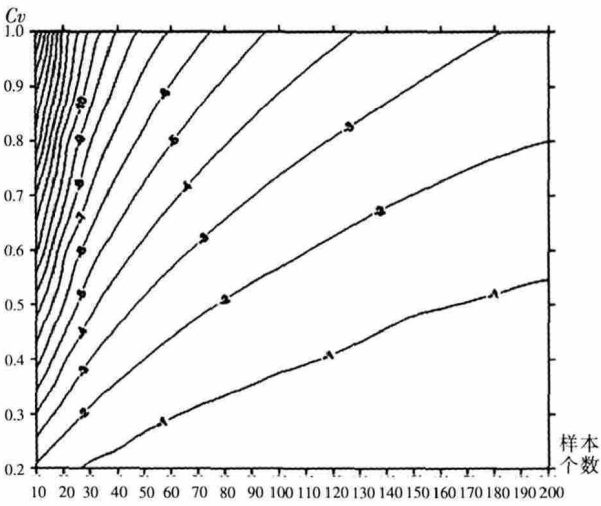


图2 $C_s/C_v=3$, 1%设计值抽样误差均值(%)分布
Fig.2 Mean value (%) distribution of SRE for designed value with 1% frequency and $C_s/C_v=3$

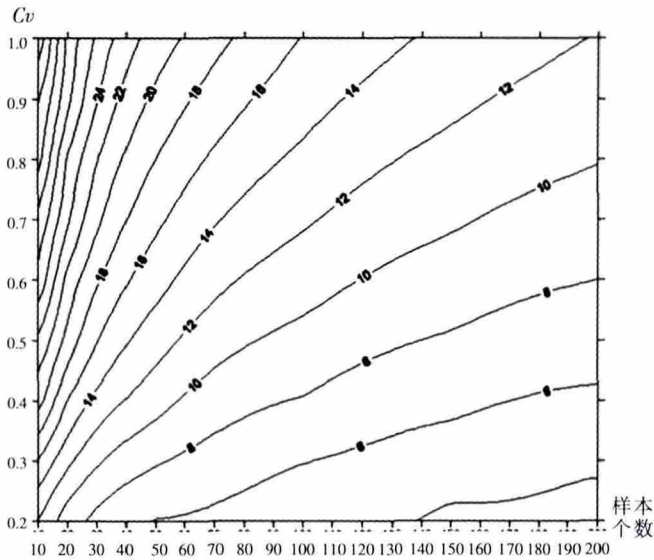


图3 $C_s/C_v=2$, 1%设计值抽样误差均方差(%)分布
Fig.3 Mean square deviation (%) distribution of SRE for designed value with 1% frequency and $C_s/C_v=2$

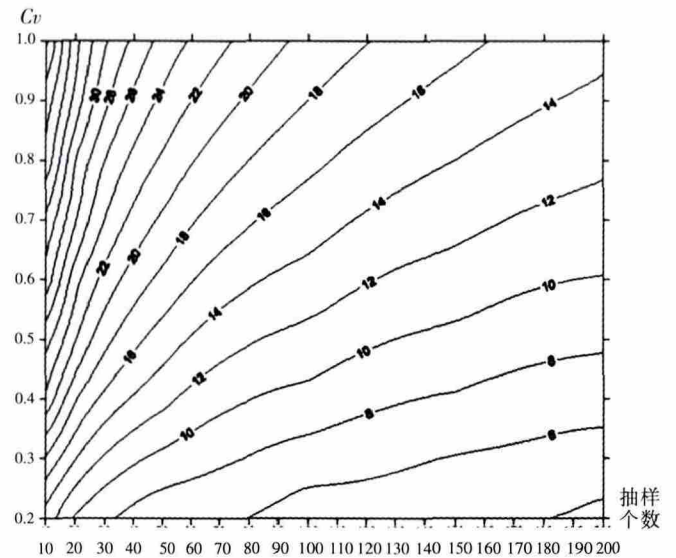


图4 $C_s/C_v=3$, 1%设计值抽样误差均方差(%)分布
Fig.4 Mean square deviation (%) distribution of SRE or designed value with 1% frequency and $C_s/C_v=3$

减小;随 C_v 值的增加而增大;随 C_s 值的增加而增大。由 P- 型频率曲线的变化规律可知: C_v 值越大,曲线越陡, C_s 值越大,曲线越弯曲。当 C_v 值、 C_s 值较大时, P- 频率曲线低频位置斜率大,参数估计中微小变动将会导致估计值的较大变化,增大抽样误差。因此,试验结论与 P- 型频率曲线的物理含义是一致的。

6 误差分布

以 $C_v=0.6$, $C_s/C_v=3.0$, 样本容量取 30, 推求 1% 设计值为例来说明抽样误差的分布情况。由 10^5 个随机样本, 建立抽样误差的频率密度散点图, 如图 5 所示。从图中可以看出:

(1) 抽样误差密度分布是一种正偏分布, 这种偏态性在小样本容量时尤为明显;

(2) 抽样误差小于零的概率较大, 即由样本估计设计值较理论值偏小的概率大; 本例中, 抽样误差有大约 66% 的概率小于零。

由此可知, 抽样误差正偏分布及负均值特性, 将导致规范规定方法修正估计值不一定能达到既定的安全程度。以图 5 分布为例, 原假定抽样误差为零均值正态分布, 按 1 倍均方差修正设计值, 其安全保证率应为 $P^*(x-0>-\sigma)=84.1\%$, 但实际上, 本例中 $P(x-0>-\sigma)$ 仅能达到 76.3%, 降低了预定安全保证率。不利于工程设计安全。若扣除负均值的影响, $P(x-\Delta>-\sigma)$ 可达到 85.1%, 与 P^* 基本一致, 可满足安全要求。

因此, 当样本数量较少、估计参数 C_v 、 C_s 值较大时, 除了需要按规范要求对设计值进行方差修正外, 还应额外考虑抽样误差均值的修正, 即在图 1~2 中查得对应相对误差, 直接修正设计值。

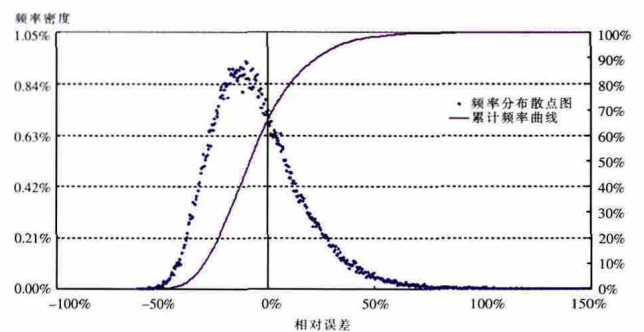


图5 抽样误差密度分布散点图
Fig.5 Scatter diagram of SRE density distribution

7 结论

本文基于统计试验原理, 对水文频率分析中不同样本容量的抽样误差分布进行探讨研究, 结果表明:

(1) 小样本容量时, 水文频率分析成果抽样误差“系统性”偏小, 误差均值小于零, 随样本容量的增加, 偏差逐渐减小, 趋于零均值正态分布;

(2) 当 C_v 、 C_s 值较大时, 抽样误差明显增大, 小样本情况下还须额外考虑抽样误差均值的修正。

参考文献:

- [1] Gumbel, E. J. Statistical Theory of Extreme Values and Some Practical Applications [R]. U. S. National Bureau of Standards, Applied

- Mathematics Series 33, 1954.
- [2] 谢家泽. 关于合理解决水文频率计算方法的问題[J]. 水利学报, 1958 (4). (XIE Jiase, Study on rational solving problems of hydrological frequency calculation method [J]. Journal of Hydraulic Engineering, 1958,(4). (in Chinese))
- [3] 陈志恺. 论皮尔逊 III 型及克里茨基-闵克里曲线对设计洪水的适用性 [A]. 水文计算经验汇编 (第二集)[C]. 北京: 中国工业出版社, 1964. (CHEN Zhikai. Applicability study on design flood with the methods of Person-III and K-M Curve [A]. Proceedings of Hydrological Calculation Experience (2nd) [C]. Beijing: China Industrial Press, 1964. (in Chinese))
- [4] Benson, M. A.. Uniform flood-frequency estimation methods for federal agencies[J]. Water Resources Research, 1968,4 (5): 891-908.
- [5] Reich, M.. Log-Pearson Type III and Gumbel Analyses of Flood[J]. Floods and Droughts, 1973.
- [6] Bobee, B.. The Log Pearson type 3 distribution and its application in hydrology[J]. Water Resources Research, 1975,(11).
- [7] 丛树铮, 谭维炎, 等. 水文频率计算中参数估计方法的统计试验研究 [J]. 水利学报, 1980,(3):1-14. (CONG Shuzheng, TAN Weiyan, et al. Research on statistical test of parameter estimation method in hydrological frequency calculation[J]. Journal of Hydraulic Engineering, 1980,(3):1-14. (in Chinese))
- [8] 李松仕. 对数皮尔逊 III 型频率分布统计特性分析 [J]. 水利学报, 1985,(9):43-48. (LI Songshi. Analysis on statistical characters of logarithm Person-III distribution [J]. Journal of Hydraulic Engineering, 1985,(9):43-48 (in Chinese))
- [9] 李松仕. 指数 Γ 分布及其在水文中的应用[J]. 水利学报, 1990,(5):30-37. (LI Songshi, Application of exponent Γ distribution in hydrology[J]. Journal of Hydraulic Engineering, 1990,(5):30-37 (in Chinese))
- [10] Greenwood, J.A., et al. Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form [J]. Water Resources Res., 1979,15(5).
- [11] 宋德敦, 丁晶. 概率权重矩法及其在 P-III 分布中的应用[J]. 水利学报, 1988,(3):1-11. (SONG Dedun, DING Jing. Application of probability weighted moments in Person-III distribution [J]. Journal of Hydraulic Engineering, 1988,(3):1-11. (in Chinese))
- [12] 秦大庸, 孙济良. 概率权重矩法在指数 Γ 分布中的应用[J]. 水利学报, 1989,(11):1-9. (QIN Dayong, SUN Jiliang. Application of probability weighted moments in exponent Γ distribution [J]. Journal of Hydraulic Engineering, 1989,(11):1-9 (in Chinese))
- [13] Hosking J.R.M.. L-moments Analysis and estimation of distributions using linear combination of order statistics [J]. J.R. Stat Soc, SerB, 1990, 52(2):105-124.
- [14] 金光炎. 线性矩法的特点评析和应用问题[J]. 水文, 2007,27(6):16-21. (JIN Guangyan, Characteristics and application of linear moment method[J]. Journal of China Hydrology, 2007,27(6): 16-21. (in Chinese))
- [15] 马秀峰. 计算水文频率参数的权函数法[J]. 水文, 1984,21(3):1-8. (MA Xiufeng. Study on weighted function method for calculation parameters of hydrological frequency[J]. Journal of China Hydrology, 1984,21(3):1-8. (in Chinese))
- [16] 刘光文. 皮尔逊 型分布参数估计 [J]. 水文, 1990,(4):1-15. (LIU Guangwen. Parameter estimation for Person-III distribution [J]. Journal of China Hydrology, 1990,4: 1-15.(in Chinese))
- [17] 刘光文. 皮尔逊 型分布参数估计 [J]. 水文, 1990,(5):1-14. (LIU Guangwen. Parameter estimation for Person-III distribution [J]. Journal of China Hydrology, 1990,(5):1-14. (in Chinese))
- [18] SL 44-2006, 水利水电工程设计洪水计算规范[S]. (SL 44-2006, Regulation for Calculating Design Flood of Water Resources and Hydropower Projects [S]. (in Chinese))
- [19] 王文圣, 金菊良. 水文随机模拟进展 [J]. 水科学进展, 2007,18(5): 768-775. (WANG Wensheng, JIN Juliang. Advances in stochastic simulation of hydrology [J]. Advances in Water Science, 2007,18 (5):768-775. (in Chinese))
- [20] Dibi A..蒙特卡洛方法在系统工程中的应用[M]. 西安: 西安交通大学出版社, 2007. (Dibi A.. Monte Carlo Applications in System Engineering [M]. Xi'an: Xi'an Jiaotong University Press, 2007. (in Chinese))
- [21] Christian P. Robert, George Casella. Monte Carlo Statistical Methods (2nd Edition) [M]. Springer Verlag, 2004.
- [22] 黄振平. 水文统计学 [M]. 南京: 河海大学出版社, 2003. (HUANG Zhenping, Hydrological Statistics [M]. Nanjing: Hohai University Press, 2003. (in Chinese))

Statistical Test Study on Sampling Error in Hydrological Frequency Analysis

TANG Lin¹, REN Zhihui²

(1. Yellow River Engineering Consulting Co., Ltd, Zhengzhou 450003, China;

2. Yellow River Institute of Hydraulic Research, Zhengzhou 450003, China)

Abstract: Based on statistical test method, sampling error distribution in hydrological frequency analysis has been studied. The test results show that: a) Design value is estimated "systemic" smaller than actual value according to smaller sample size. The sampling relative error (SRE) is below zero. With the sample size increasing, deviation gradually reduce, and SRE is characteristics of zero mean normal distribution. b) SRE would significantly increase with large value of C_v and C_s . So additional revision of SER should be considered in small sample size condition.

Key words: hydrological frequency analysis; sampling error; statistical test