



# 年最大值抽样估算的暴雨频率设计值的低估问题探讨

## 摘要

由年最大值抽样(AMS)和年起大值抽样(AES)的基本理论及重现期(RP)的定义可知,AMS并不符合以“事件”为基础的重现期的定义.以美国西南半干旱区1438个雨量站和太湖流域96个雨量站的降雨资料为例,通过经验频率与超过概率的比较,发现AMS估算的暴雨频率设计值偏小,尤其是对常遇频率降雨设计值的影响更加显著.美国的降雨量资料站点多、系列长,实际资料验证与理论分析一致.通过对太湖流域AMS资料的分布形态进行分析的结果表明:太湖流域的站点不多,资料长度不够,且大部分站点在雨量最大值区数据稀少,使得频率直方图不连续,是造成我国太湖流域的资料验证效果不理想的可能原因.

## 关键词

水文频率计算;地区线性矩法;年最大值抽样;超过概率

中图分类号 P333

文献标志码 A

收稿日期 2014-06-05

资助项目 水利部公益性行业科研专项(201001047);江苏省气象局青年科研基金(Q201602);江苏省自然科学基金(BK20141001);淮河流域气象开放研究基金(HRM201502)

作者简介

吴俊梅,女,硕士,研究方向为水文气象.  
wjmqingkong@126.com

1 江苏省昆山市气象局,昆山,215337

2 南京信息工程大学 水文气象学院,南京,210044

3 94857 部队气象台,芜湖,241000

## 0 引言

我国幅员辽阔,地处北半球中纬度地带,是洪涝灾害频发的国家.无论是工程防洪设计,或是地区和都市的防洪规划,都需要一个科学可靠的防洪设计标准.在实际应用中,水文频率计算是制定防洪设计标准的核心理论和关键技术,其核心内容包括参数估计、分布线型和抽样方法.长期以来,我国水文学家对参数估计和分布线型做了大量工作,却对抽样方法的研究甚少<sup>[1]</sup>.

我国大多数水文站点的观测资料系列短缺,一般仅有几十年,却要求估计百年一遇甚至千年一遇的设计值,估计误差较大<sup>[2]</sup>.通常,洪水频率分析被归结为极值统计.在中国,水文设计工程师往往采用年最大值抽样(Annual Maximum Series, AMS)方法,即在一年的观测资料中只选取最大一次洪水作为样本进行统计分析,该方法选样简单、资料易得,但会遗漏一些数值较大,而在年内排位第二或第三的洪水,尽管这些洪水可能大于其他年份的最大值,AMS抽样方法却未能从有限的水文观测资料中获得尽可能多的水文信息来增加样本系列长度,这就使得“资料系列短”问题变得更为突出<sup>[3]</sup>.因此,正确使用信息量(抽样方法的选择)一直是提高设计洪水计算可靠性和促进洪水频率分析技术发展的关键之一<sup>[4]</sup>.

为了更大限度利用实测资料的信息量,20世纪60年代以来,部分时段系列或年多次抽样(Partial Duration Series, PDS)方法应运而生<sup>[5]</sup>.实际上,PDS与我国水文界在20世纪60年代初提出的超定量系列<sup>[4]</sup>或欧洲水文统计学家使用的超门槛抽样(Peak Over Threshold, POT)并无实质性区别.PDS模型通过设定阈值选取样本的方法能够避免AMS模型的缺陷,获取更多的信息<sup>[5]</sup>.而PDS模型的缺陷是由于一年中可能选取多场洪水,样本系列的独立性无法保证,使用前必须严格论证资料的独立性而后取舍.此外,PDS的资料收集、处理也较AMS繁琐.

近年来,常遇频率的洪涝事件越来越频繁,主要存在以下几方面原因:1)全球气候变暖大背景的影响,加剧了极端水文气象事件的发生;2)城市人口剧增,城市化进程加剧,人类活动使得下垫面发生显著变化,各种工程建设和改造规划使得下垫面“固化(不透水)”面积比例加大,降雨径流情势改变,径流系数显著增大;3)在水文频率计算中,抽样方法的选择存在问题,常用的年最大值抽样不符合重现期

的含义,低估了 2~25 a 一遇的常遇频率暴雨设计值.上述第一和第二项原因超出了本文探讨的范畴.本文在地区线性矩法<sup>[3,6-8]</sup>的基础上,探讨年最大值抽样估算的暴雨频率设计值的低估问题,针对第三方面原因展开深入研究,从理论和计算技术上分析常遇频率暴雨事件频繁发生的原因.

## 1 基本理论

### 1.1 抽样方法

常用的 3 种抽样方法分别为年最大值系列 AMS、部分时段系列 PDS 和年超大值系列(Annual Exceedance Series, AES)<sup>[8]</sup>.

1) 年最大值系列 AMS 由每一年的最大事件组成,不考虑该年中是否存在次大事件,尽管它可能超过其他年的最大事件.AMS 包含了水文气象事件诸如洪峰径流或降雨的重要信息<sup>[9]</sup>,概念清楚、选样简单、独立性强,但是,每年只抽取最大值,信息的利用有缺陷.

2) 部分时段系列 PDS,或称年多次系列,是指所选系列的量值超过某一设定阈值,其样本长度往往比有限资料年数大很多.与 AMS 相比,该方法利用了更多的水文信息.

3) 年超大值系列 AES 被认为是 PDS 的一种特殊情况,即选取 PDS 中最大的前  $n$  个事件<sup>[10]</sup>,此处  $n$  是有效的资料年数.AES 是与重现期含义相一致的抽样方法,依据 AES 进行频率计算得到的频率设计值比较合理可靠<sup>[7]</sup>.在本文介绍的研究中,统一采用 AES 方法.

在我国的水文频率计算中,因简单直观,几乎都采用 AMS,很少采用 PDS 和 AES<sup>[10]</sup>.

### 1.2 重现期和超过概率

重现期  $T$  (Return Period, RP) 定义为某一随机事件(诸如洪水、暴雨、地震等自然现象)重复出现的时间间隔的平均数,即平均的重现间隔期<sup>[11]</sup>.根据定义,重现期具有统计平均概念,是以“事件(event)”而不是以“年(year)”为对象,并不意味着每个时间间隔内都发生一次水文事件.重现期的含义在工程水文设计和防洪规划中,更多的是指“量级”或“标准”,而不是指时间或时段.然而,AMS 在每一年中仅选择最大值,可能忽略丰水年份的次大值,而这个次大值甚至第三大值可能大于某一年的最大值,这样估算的暴雨频率设计值一般会偏低.其矛盾在于:AMS 选择的是“一年中的最大事件”,和

重现期“平均时间间隔”的意义不一致.因此,AMS 并不符合以“事件”为基础的重现期的含义.

超过概率  $P_T(X \geq x_T)$  定义为一个大于等于  $x_T$  的水文事件发生的概率.根据定义,在一段相当长的时间内,例如百年内,2 a 一遇事件的平均发生次数应为 50 次,准确地说,在某一年中 2 a 一遇事件出现的概率为 0.5.在数据量有限的情况下,事件发生的概率  $P_T$  通常可以用有限样本的频率  $f_T$  来进行估算.超过概率与重现期的关系有 2 种表示法<sup>[11]</sup>: 1) 当研究暴雨洪水时,一般  $P_T < 50\%$ ,采用  $T = \frac{1}{P_T}$ ; 2) 当研究枯水问题时,一般  $P_T > 50\%$ ,采用  $T = \frac{1}{1 - P_T}$ .

## 2 实例计算及分析

从以上分析可以看出,AMS 不符合以“事件”为基础的重现期的含义,我国的水文频率计算常用 AMS,是导致常遇频率暴雨设计值偏小的原因之一.以下研究采用中国太湖流域实测资料进行验证,并与美国西南半干旱区的例子做对比分析,探讨年最大值抽样对暴雨频率设计值的影响.

### 2.1 资料选取

我国太湖流域(Taihu Lake Basin, TLB)位于长江三角洲南缘,三面濒江临海,以平原为主,属湿润的北亚热带气候区.本文选取太湖流域具有 20 a 以上资料的 96 个雨量站的年最大日雨量系列和年超大日雨量系列,站点空间分布近于均匀,各站资料长度均不相同,其平均有 37 a,最长达 70 a.按地区线性矩法,划分为 8 个水文气象一致区,如图 1 所示.

为更好地验证,本文引用美国近年来重新编制的降雨频率估算图集的例子做对比分析,选取美国西南半干旱区(Southwest Semiarid U.S., SSUS) 1 438 个雨量站的年最大值系列和年超大值系列,将其划分为 59 个水文气象一致区,如图 2 所示<sup>[12]</sup>.

### 2.2 实测资料分析

#### 2.2.1 资料验证

以美国西南半干旱研究区 59 个一致区 1 438 个雨量站的 AMS 为例<sup>[7]</sup>.首先,采用地区线性矩法计算各雨量站相应重现期的频率设计值,然后计算各站实测资料中大于等于相应重现期的频率设计值的经验频率(Exceedance Frequency, EF),最后所有站平均,得出该区平均出现的经验频率及相应的重现年数,结果如表 1 所示.其中阴影部分为 2~100 a 一

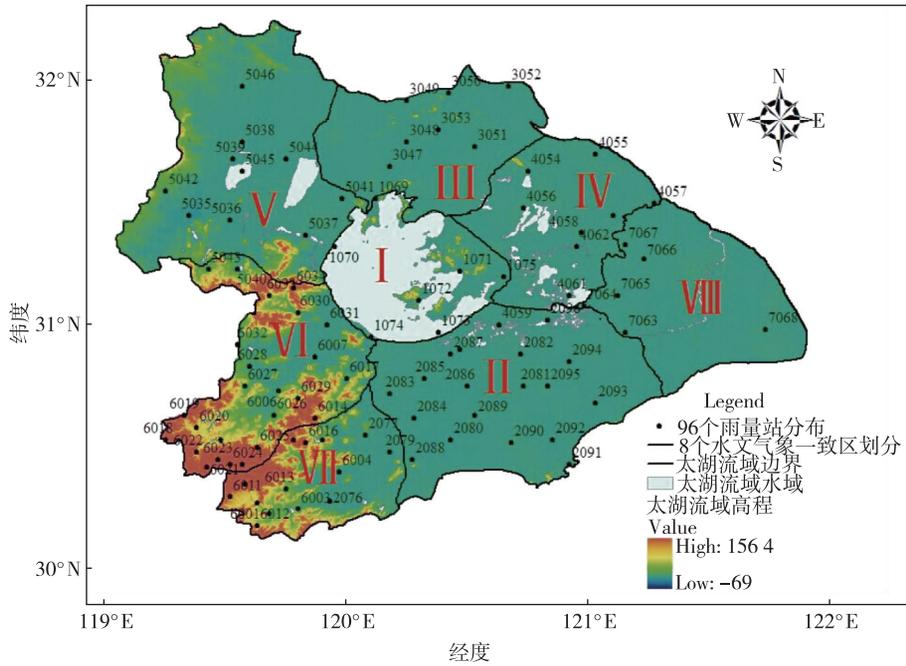


图1 太湖流域8个水文气象一致区的雨量站分布

Fig. 1 Spatial distribution of the rainfall stations over 8 regions with regionalization in the TLB

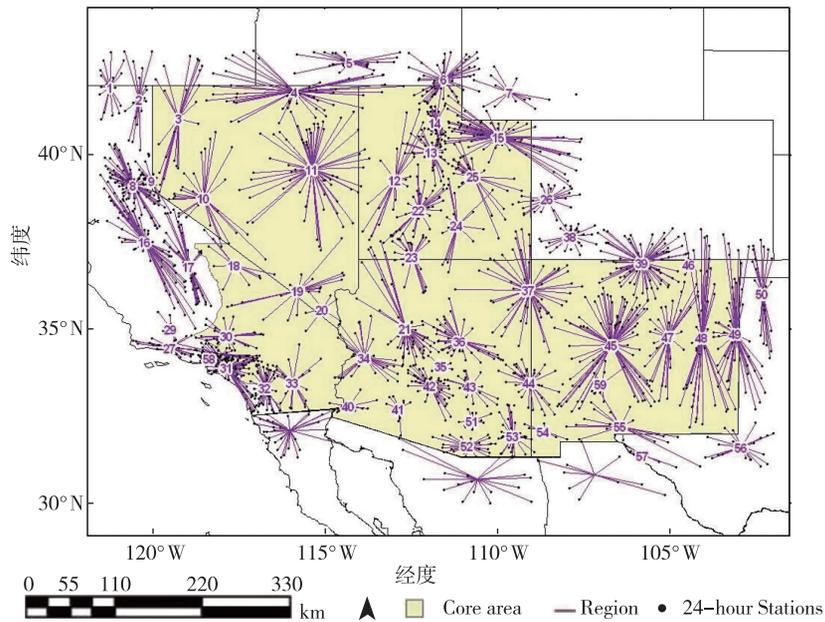


图2 美国西南半干旱区59个水文气象一致区的雨量站分布<sup>[12]</sup>

Fig. 2 Spatial distribution of the rainfall stations over 59 regions with regionalization in the SSUS

遇的重现期及其超过概率 (Exceedance Probability, EP).由表1可以看出,美国西南半干旱区2 a一遇的暴雨平均经验频率是0.706(相应的重现期是1.42 a),远大于其超过概率0.5,也就是说,当采用AMS时,2 a一遇的暴雨其实平均1.42 a就出现一

次,相比于设计频率发生得频繁,其余相应重现期下的暴雨也有类似情况,且对于2~25 a一遇的常遇事件这种现象更为显著.

采用上述方法,对太湖流域的雨量资料做相同验证,结果如表1所示.选用太湖流域8个一致区96

个雨量站的 AMS,比较其 EF 与 EP.从区域平均结果看出,5-a(表示 5 a 一遇,下类推)、10-a、25-a 重现期下的经验频率略高于超过概率,AMS 对 5~25 a 一遇的常遇频率暴雨有影响,而 2-a、50-a 和 100-a 的经验频率略低于超过概率.总体而言,效果不如美国资料明显.

表 1 SSUS 59 个一致区与 TLB 8 个一致区分别采用 AMS 时平均的 EF 与 EP 的比较

Table 1 Comparison of average EF and EP based on AMS data of 59 regions in the SSUS and 8 regions in the TLB

RP	EP	SSUS		TLB	
		EF	RP	EF	RP
2-a	0.50	0.706	1.42-a	0.496	2.02-a
5-a	0.20	0.229	4.37-a	0.203	4.93-a
10-a	0.10	0.107	9.35-a	0.103	9.71-a
25-a	0.04	0.042	23.81-a	0.045	22.22-a
50-a	0.02	0.021	47.62-a	0.019	52.63-a
100-a	0.01	0.010	100-a	0.009	111.11-a

图 3a 是美国西南半干旱区 1 438 个雨量站采用 AMS 时平均的 EF 结果.明显看出,常遇频率(一般 25 a 一遇以下)事件出现的经验频率明显大于超过概率,即 AMS 估算的常遇频率设计值明显偏小.这也解释了为何近年来常遇频率的降雨越来越频繁的问题.

图 3b 是太湖流域资料的比较结果.理论上,AMS 经验频率应远大于 AMS 超过概率,但实测资料验证存在偏差,两者很接近.

### 2.2.2 可能原因分析

理论上分析,AMS 的概念与重现期的含义不一

致,估算的常遇频率暴雨设计值偏小,美国不同研究区几千个雨量站的历史资料验证了这一理论,而太湖流域的资料验证存在偏差.因此选取太湖流域各一致区内雨量站的 AMS 直方图(Histogram)作偏态形状分析,寻找可能原因.分别以二区和八区内最大雨量值所在站点(2090 和 7066 站)以及大部分雨量间隔都有频数的站点(2087 和 7063 站)为例,制作直方图,结果如图 4 所示.

4 个雨量站的实际资料均呈正偏分布,符合统计特征.尽管二区中 2090 站有最大日雨量值 323.80 mm,但图 4a 指出其余大部分雨量值集中在小值区,大值区数据稀少,即在大值区的大部分间隔没有频数;而 2087 站最大日雨量值 194.90 mm,图 4c 指出大部分雨量间隔都有频数,比较两者的经验频率 EF,结果如表 2 所示.从 2 a 一遇至 10 a 一遇,2087 站的 EF 高于 2090 站;由于最大雨量值在 2090 站,远大于 2087 站,因而从 25 a 一遇至 100 a 一遇 2090 站的 EF 要高.同样地分析八区,7066 站有最大日雨量值 282.50 mm,7063 站的最大值是 179.80 mm,由表 2 看出,从 2 a 一遇至 25 a 一遇,7063 站的 EF 都高于 7066 站.类似分析其余一致区,大部分站点的实际资料在雨量大值区数据稀少,造成直方图不连续,可以得到相似结论.

总结太湖流域的研究结果与理论分析存在偏差的可能原因是,此次研究中的雨量站不够多,资料不够长,并且大部分站点在雨量大值区数据稀少,造成直方图不连续,使得计算经验频率时缺少这一段区间的数,从而比较经验频率和超过概率时数据量不对称,结论就难定.理论上的结论需要大量的

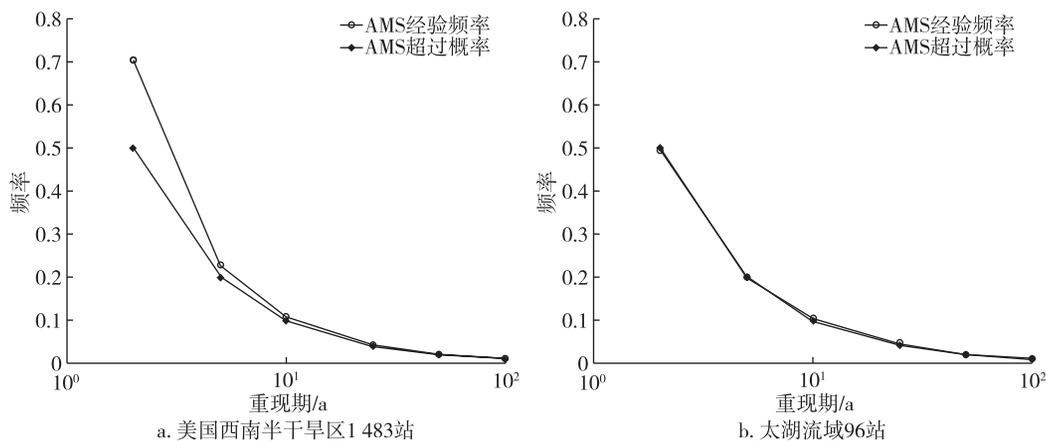


图 3 采用 AMS 时 EF 与 EP 的比较

Fig. 3 Comparison of average EF and EP based on AMS data

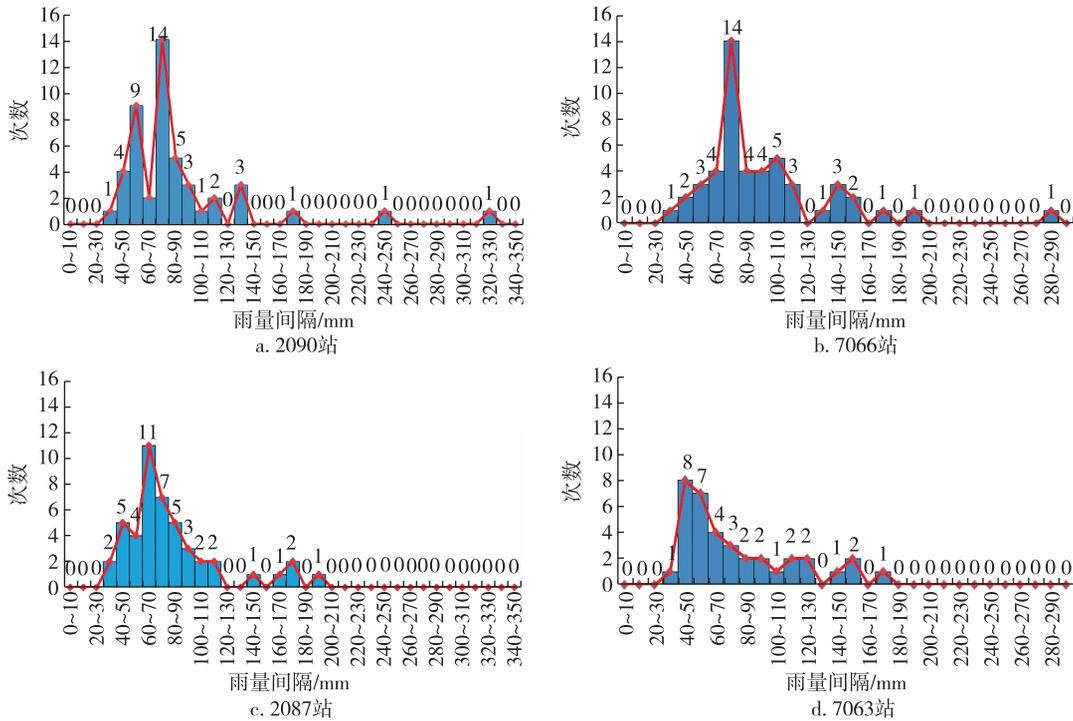


图4 AMS 实测资料偏态分析  
Fig. 4 Skewness analysis of AMS data

表2 2090、7066、2087 和 7063 站采用 AMS 时 EF 的比较  
Table 2 Comparison of EF based on AMS data among station  
2090, 7066, 2087 and 7063

RP	EP	2090(EF)	2087(EF)	7066(EF)	7063(EF)
		47 a	46 a	49 a	36 a
2-a	0.50	0.276 6	0.369 6	0.387 8	0.388 9
5-a	0.20	0.127 7	0.130 4	0.163 3	0.222 2
10-a	0.10	0.063 8	0.108 7	0.040 8	0.111 1
25-a	0.04	0.042 6	0.021 7	0.020 4	0.027 8
50-a	0.02	0.021 3	0.000 0	0.020 4	0.000 0
100-a	0.01	0.021 3	0.000 0	0.000 0	0.000 0

研究站点和足够长的研究资料的验证,因而在今后的研究中需要收集我国其他更大范围、更长资料系列做进一步分析.

### 3 小结

本文从理论上介绍了年最大值抽样 AMS 和年超大值抽样 AES 的概念及差异,同时分别采用美国西南半干旱区和中国太湖流域的实测数据分析,得到以下结论:

1) 美国实际数据分析方面,AMS 的经验频率远高于 AMS 超过概率,即用 AMS 估算的常遇频率暴雨事件出现的机会相比于设计频率要频繁,或者说

AMS 估算的常遇频率暴雨设计值偏小,验证了理论分析的结论.

2) 太湖流域实际数据分析方面,AMS 的经验频率与 AMS 超过概率相接近,即 AMS 估算的暴雨频率设计值的低估现象不明显,与理论分析存在偏差.

3) 针对理论分析,太湖流域实测资料的研究成果不如美国资料的验证效果明显.究其原因,美国提供了大量的研究站点和足够长的数据资料供研究,而本文中,太湖流域的站点不够多,资料长度不够,且大部分站点在雨量大值区数据稀少,造成直方图不连续.因而在今后的研究中,需要收集我国大量的资料来验证本文的理论推断.

### 参考文献

#### References

[ 1 ] 唐林,任智慧.水文频率分析中抽样误差的统计试验研究[J].水文,2013,33(6):6-10  
TANG Lin,REN Zhihui.Statistical test study on sampling error in hydrological frequency analysis [J]. Journal of China Hydrology, 2013, 33(6):6-10

[ 2 ] 王剑峰,宋松柏.广义 Pareto 分布在超定量洪水序列频率分析中的应用[J].西北农林科技大学学报(自然科学版),2010,38(2):191-196  
WANG Jianfeng, SONG Songbai. Application of generalized Pareto distribution in POT flood series fre-

- quency analysis [ J ]. Journal of Northwest A & F University ( Natural Science Edition ), 2010, 38 ( 2 ): 191-196
- [ 3 ] 朱惠玲.区域线性矩法在黄河下游洪水频率分析中的应用研究[ D ].上海:同济大学土木工程学院,2006  
ZHU Huiling. Flood frequency design in the lower of Yellow River by regional L-moments method[ D ].Shanghai: School of Civil Engineering, Tongji University, 2006
- [ 4 ] 王善序.洪水超定量系列频率分析[ J ].人民长江, 1999, 30( 8 ): 23-25  
WANG Shanxu. Frequency analysis on over quota flood series[ J ]. Yangtze River, 1999, 30( 8 ): 23-25
- [ 5 ] 戴昌军,梁忠民,栾承梅,等.洪水频率分析中 PDS 模型研究进展[ J ].水科学进展, 2006, 17( 1 ): 136-140  
DAI Changjun, LIANG Zhongmin, LUAN Chengmei, et al. Advance in flood frequency analysis for partial duration series[ J ]. Advances in Water Science, 2006, 17 ( 1 ): 136-140
- [ 6 ] Hosking J R M. L-moments: Analysis and estimation of distributions using linear combinations of order statistics [ J ]. Journal of the Royal Statistical Society, Series B: Methodological, 1990, 52( 1 ): 105-124
- [ 7 ] Lin B, Bonnin G M, Martin D L, et al. Regional frequency studies of annual extreme precipitation in the United States based on regional L-moments analysis[ C ] // World Environmental and Water Resources Proceedings, 2006, DOI: 10. 1061/40856( 200) 219
- [ 8 ] 林炳章,邵月红,闫桂霞,等.水文气象促进工程水文计算核心课题研究的发展[ C ] // 邓坚.中国水文科技新发展: 2012 中国水文学学术讨论会论文集.南京: 河海大学出版社, 2012: 50-63
- LIN Bingzhang, SHAO Yuehong, YAN Guixia, et al. Hydrometeorology promotes the development of core research topics in engineering hydrology[ C ] // DENG Jian. New Development of Hydrological Science and Technology in China: Proceedings of Hydrological Academic Seminar. Nanjing: Hohai University Press, 2012: 50-63
- [ 9 ] 方彬,郭生练,柴晓玲,等.FPOT 方法在洪水频率分析中的应用研究[ J ].水力发电, 2005, 31( 2 ): 9-12  
FANG Bin, GUO Shenglian, CHAI Xiaoling, et al. Flood frequency analysis based on FPOT method [ J ]. Water Power, 2005, 31( 2 ): 9-12
- [ 10 ] 许洁,李文杰,周则凯.暴雨频率计算中的选样方法探讨[ J ].浙江水利科技, 2007( 3 ): 51-52  
XU Jie, LI Wenjie, ZHOU Zekai. Discussion on sampling in computing storm frequency [ J ]. Zhejiang Hydraulics, 2007( 3 ): 51-52
- [ 11 ] 詹道江,徐向阳,陈元芳.工程水文学[ M ]. 4 版.北京: 中国水利水电出版社, 2010: 147-153  
ZHAN Daojiang, XU Xiangyang, CHEN Yuanfang. Engineering Hydrology [ M ]. 4th ed. Beijing: China Water & Power Press, 2010: 147-153
- [ 12 ] Lin B. Evaluation of exceedance probabilities [ R ]. Maryland US: NOAA/NWS/Office of Hydrologic Development: HDSC, 2008
- [ 13 ] 施能.气象科研与预报中的多元分析方法[ M ]. 2 版.北京: 气象出版社, 2002: 4-5  
SHI Neng. Multivariate analysis methods of weather research and forecast [ M ]. 2nd ed. Beijing: China Meteorological Press, 2002: 4-5

## Underestimation of precipitation quantile estimates based on AMS data

WU Junmei<sup>1</sup> LIN Bingzhang<sup>2</sup> PU Jian<sup>3</sup>

<sup>1</sup> Kunshan Meteorological Bureau of Jiangsu Province, Kunshan 215337

<sup>2</sup> School of Hydrometeorology, Nanjing University of Information Science & Technology, Nanjing 210044

<sup>3</sup> Meteorological Station of 94857 Army, Wuhu 241000

**Abstract** This paper briefed the concept of annual maximum series(AMS) and annual exceedance series(AES), indicating that the use of AMS contradicts the definition of return period. Having taken the rainfall data of 1 438 stations in the Southwest Semiarid area of the U.S.A(SSUS) and 96 stations in the Taihu Lake Basin(TLB) of China as examples for frequency analysis, we find out that the quantiles estimated based on AMS under current conventional computation are underestimated, especially for frequent events, by comparison of the empirical frequency of data with the theoretical exceedance probability. However, the findings for the SSUS data are more justified than those for the data in the TLB. The study suggests the possible causes for the difference could be, via a skewness analysis of AMS data in the TLB, that the stations used for the study in the TLB are very limited and data series are short in comparison with the SSUS data. Furthermore, only few high values are available in the high value interval, resulting in discontinuous AMS histograms for most stations in the TLB.

**Key words** hydrologic frequency analysis; regional L-moments analysis method; annual maximum series; exceedance probability