

DOI:10.3969/j.issn.1672-1144.2019.02.002

水文频率计算研究面临的挑战与建议

宋松柏^{1,2}

(1. 西北农林科技大学 水利与建筑工程学院, 陕西 杨凌 712100;

2. 西北农林科技大学 旱区农业水土工程教育部重点实验室, 陕西 杨凌 712100)

摘 要: 水文频率计算是水文分析计算的主要内容之一,是各类涉水工程规划、设计确定工程规模和管理决策的主要依据。经过 140 多年的研究和实践,水文频率计算已经形成较为完整的理论体系与方法。包括:水文样本选择和数据检验、经验频率计算、频率分布函数选取、分布函数参数估计、分布模型检验、设计值计算和设计值不确定性分析。所有这些支撑了涉水工程的规划、设计和管理。根据国内外有关研究进展,总结了水文序列频率计算的前提条件、频率分布函数和参数计算方法。从水文频率分布函数选择、水文数据、分布函数参数计算、经验频率等方面分析了水文频率计算研究面临的挑战。在此基础上,提出了水文频率计算研究的一些建议,以期促进水文频率计算理论的进一步发展。

关键词: 水文频率;分布函数;参数估计;水文设计值

中图分类号: TV122+.3

文献标识码: A

文章编号: 1672—1144(2019)02—0012—07

Research challenges and suggestions of hydrological frequency calculation

SONG Songbai^{1,2}

(1. College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China;

2. Key Laboratory of Agricultural Water and Soil Engineering, Ministry of Education, Northwest A&F University, Yangling, Shaanxi 712100, China)

Abstract: Hydrological frequency calculation is one of main contents of hydrological analysis and calculation. It also can provide scientific basis the planning, design and management for all types of water projects. During 140 years researches and applications, hydrological frequency calculation has been formed a complete theories and methods. Including: hydrological data sampling and data validation, empirical probability calculation, parameters estimation of probability distribution function, goodness - of - fit testing, hydrological design values calculation and their uncertainty analysis. These achievements have supported the planning, design and management for all types of water projects. In this paper, on the basis of the systematic summary of the existing research results, such as prerequisite of hydrological frequency calculation, probability distribution function and its parameters estimation methods, empirical probability. The author proposed some suggestions of future hydrological frequency calculation, and expect to promote the development of hydrological frequency calculation theory.

Keywords: hydrological frequency; probability distribution function; parameters estimation; hydrological design values

水文频率计算是综合运用水文学、水文统计学和其它数学原理,利用计算区的水文资料,分析水文事件的统计规律,定量表征水文变量设计值与设计标准(频率或重现期)之间的关系,是各类涉水工程规划、设计确定工程规模和管理决策的主要依据^[1-2]。Foster H A 认为水文频率计算始于 1880 年—1890 年期间^[3],Herschel C 和 Freeman G H 应用历

时曲线(现称为频率曲线)第一次进行径流序列频率分析^[3]。1896 年,Horton R E 根据 Rafter G W 的建议,采用正态分布在半对数格纸上进行径流序列适线研究。1913 年,Fuller W E 采用半对数格纸进行重现期与径流序列设计值的适线研究,一些学者认为这是首次进行综合性水文频率计算。1921 年,Hazen A 采用正态分布概率对数格纸进行适线^[3]。1924

收稿日期:2018-12-24

修稿日期:2019-01-21

基金项目:国家自然科学基金项目(51479171,51579059,51179160)

作者简介:宋松柏(1965—),男,陕西永寿人,博士,教授,主要从事水文水资源研究工作。E-mail: ssb6533@nwsuaf.edu.cn

年,Foster 提出了 P-III 分布分析方法,并计算出离均系数表格供计算者使用。1935 年,前苏联克里茨基和门克尔提出组合概率近似分析法,是最早的水文多元概率分布研究^[3]。按照金光炎^[3]的文献,1932 年,周镇纶^[4]应用美国雨量站资料,绘制了正态分布和 P-III 分布曲线,发表“全年雨量之常率线及常率积分线”研究论文。1933 年,我国学者须恺^[5]绘制了宜昌、汉口、九江、芜湖和镇江 5 站的年最大月雨量、日雨量频率曲线,并通过径流系数等参数转化为洪水值,1935 年著有“淮河洪水之频率”研究论文。1937 年,陈椿庭^[6]发表“中国五大河洪水量频率曲线之研究”论文,系统地介绍了 Grassberger、Hazen 和 Foster 经验概率计算和概率格纸,并进行长江、黄河、永定河和淮河年最大洪水的频率计算。1947 年,他的这篇论文获国民政府教育部优秀论文奖^[7]。此外,陈椿庭^[8]著有“绘制洪水流量频率曲线的简便新法”、“水文频率曲线点绘方法比较”等论文。这些反映了我国早期的水文频率计算研究成果。新中国成立后,随着水利工程建设的发展,我国水文学者在吸收国外水文统计理论的基础上,广泛地开展工程水文分析计算问题研究和实践。20 世纪 80 年代,丁晶、宋德敦、马秀峰、刘光文、郭生练等学者先后提出了 P-III 分布参数估计的概率权重法、单权函数法、双权函数法和非参数估计法,他们的研究成果至今被广泛地应用于水文频率分布的参数估计或被许多计算手册、教科书引用。

目前,主要代表性的研究成果和理论体系总结有《Statistical Methods in Hydrology》^[9]《Fundamentals of Statistical Hydrology》^[10]《Statistical Methods in Hydrology and Hydroclimatology》^[11]《Flood frequency analysis》^[12]《Predictive Hydrology: A Frequency Analysis Approach》^[13]《Entropy - Based Parameter Estimation in Hydrology》^[14]《水文统计学》^[15]《水文学的概率统计基础》^[16]《设计洪水研究进展与评价》^[17]《水科学技术中的概率统计方法》^[18]《水文气象统计通用模型》^[19]《统计试验方法及应用》^[20]《水文水资源应用数理统计》^[21]《水文统计的原理与方法》^[22]《实用水文统计法》^[23]《水文统计计算》^[24]《水文水资源随机分析》^[25]《工程数据统计分析》^[26]《水文水资源分析研究》^[27]《水文水资源计算务实》^[28]《统计水文学》^[29]《水文风险分析的理论和方法》^[30]《Copulas 函数及其在水文中的应用》^[31]《单变量水文序列频率计算原理与应用》^[32]《非一致性水文概率分布估计理论和方法》^[33]《变化环境下地表水资源评价方法》^[34]《变化环境下区域水资源变异问题研究》^[35]

《变化环境下的水文频率分析方法及应用》^[36]《淮河复合河道洪水概率预报方法与应用》^[37]《Copula 函数理论在多变量水文分析计算中的应用研究》^[38]等。其中,金光炎的《水文统计的原理与方法》^[22]专著推动了我国水文统计理论的普及和应用,黄振平等^[2]主编的《水文统计学》至今仍是我国水文与水资源工程专业的通用教材。从 1880 年—1890 年 Herschel 和 Freeman 的径流序列频率分析算起,水文频率计算距今已有 140 多年的研究和实践历史。目前,水文频率计算方法是水文学最为活跃的研究领域之一,受到国外学者和水文计算者的高度重视,经过几代水文科学工作者的不懈努力,他们的研究成果为涉水工程规划建设和管理提供了坚实的支撑。

Singh V P 等^[39]认为水文频率分析方法大致可以分为 4 类:(1) 经验法;(2) 现象法;(3) 动力法;(4) 随机模型结合蒙特卡洛模拟法。按照采用测站的多少,水文频率分析分为单站频率分析和区域频率分析。如果按照变量的数目来说,水文频率分析又可分为单变量频率分析和多变量频率分析。按照事件个数,也可以分为单事件频率分析和游程事件频率分析(如干旱历时频率分析)。经验法是上述 4 类方法中应用较广泛的方法,采用经验法进行单站频率分析是工程规划设计使用最多的方法。水文频率分析推求设计值主要有参数统计和非参数统计两种途径。参数统计方法是国内外研究和应用较多的方法,需事先假定水文序列的分布模型,利用参数估计方法估算样本参数,根据样本统计特征值与分布参数的关系,求出分布参数。这种方法涉及 6 个步骤^[13,39]:(1) 水文样本数据选择和数据检验;(2) 选择经验公式(绘点位置公式)计算样本经验概率;(3) 选择概率分布函数,采用合适的参数估算技术拟合水文样本;(4) 水文分布模型检验;(5) 给定设计频率,进行水文设计值计算;(6) 水文设计值不确定性分析。根据国内外研究文献的报道^[9-43],本文根据现有水文频率计算研究进展,总结了以下有关面临的挑战和研究建议,以期促进水文频率计算理论的进一步发展。

1 水文序列频率计算的前提条件

水文样本数据一般根据实际需要选取,并形成某类特征值数据序列,如一定时空尺度的极值、月值、枯水值和年值数据等。上述序列选择不同,其频率和重现期计算有所差异。一般来说,选取水文数据序列必须满足下述计算前提条件^[13]:(1) 数据正确地揭示水文变化规律;(2) 形成数据序列的物理

机制没有发生变化(一致性),满足平稳性和同质性;(3) 数据序列满足随机简单样本特性。随机性是指样本数据服从同一概率分布,而简单样本则指一个样本数据不影响后续值的发生,即数据间满足独立性;(4) 数据序列应具有足够的长度。数据检验分为参数检验和非参数检验两大类,包括样本特征参数与分布参数的一致性检验;两个样本分布的同一性检验;样本服从某一概率分布的检验;样本数据间的相依性检验。水文分布模型拟合检验方法有图形法、Chi-square 检验、Kolmogorov-Smirnov 检验、GPD 检验、Anderson-Darling 检验、矩图法、线性矩图法以及分布函数模型选优比较法。

2 水文序列频率分布函数与参数计算方法

Rao 等在他们的《Flood frequency analysis》^[12]专著中,系统地总结了目前常用的水文序列频率分布函数与参数计算方法,给出了大量详细的应用实例。常用的单变量水文频率线型有 20 多种,主要分为以下四大类:(1) Γ 分布类包括指数分布、两参数 Γ 分布、P-III 分布(三参数 Γ 分布)、对数 P-III 分布和四参数指数 Γ 分布、克里茨基-门克尔分布等;(2) 极值分布类包括极值 I 型、II 型、III 型分布以及广义极值分布;(3) 正态分布类包括正态、两参数对数正态和三参数对数正态分布;(4) Wakeby 分布类主要包括五参数 Wakeby 分布、四参数 Wakeby 分布和广义 Pareto 分布。除上述分布外,还有一些常用的偏态非对称分布模型。这些模型的主要特点是引入了一个新的参数 λ 控制分布模型中的偏态(skew)和峰度(kurtosis)系数,包含不完全 Beta 函数、第一类完全椭圆函数积分、第二类完全椭圆函数积分、广义超几何函数、普西函数、 $\delta(\cdot)$ 函数、Riemann's zeta 函数、补余误差函数、不完全 Γ 函数、第二类修正 Bessel 函数积分等特殊函数,概率密度函数形式复杂,相应的概率分布函数计算及其困难。偏态非对称分布主要有偏正态分布、混合偏正态分布、偏态 t 分布、混合偏态 t 分布、偏态拉普拉斯分布、偏态 Logistic 分布、偏态均匀分布、偏态指数幂分布、偏态贝塞尔函数分布、偏态皮尔逊 II 分布、偏态皮尔逊 VII 分布、偏态广义 t 分布等。四参数指数 Γ 分布是中国水利水电科学研究院孙济良先生根据 Γ 分布经过指数变换推得的分布,当其参数取某些特定值时,四参数指数 Γ 分布可转化为两参数 Γ 分布、P-III 分布、克里茨基-门克尔分布、极值 III 型分布、卡方分布、指数分

布、正态分布、对数正态和极值 I 型分布。因而也称为水文频率通用模型^[22]。对于删失或截取水文序列样本,通常也采用删失分布或截取分布,他们是完全不同上述的概率分布。

水文频率分布参数计算涉及高等数学、概率论与数理统计、水文学、数值计算、优化计算等学科的交叉和渗透。单变量水文频率分布参数估计方法主要有:(1) 矩法;(2) 极大似然法;(3) 概率权重和线性矩法;(4) 最小二乘法;(5) 最大熵原理;(6) 混合矩法;(7) 广义矩法;(8) 不完整均值法;(9) 单位脉冲响应函数法;(10) 部分概率权重矩法和高阶概率权重矩法。其中,矩法、极大似然法和概率权重法是最广泛的参数估计方法。由于 P-III 分布的分位数不能表示为相应概率分布值的显函数,概率权重法求解参数困难。20 世纪 80 年代,四川大学丁晶教授和南京水利科学研究院宋德敦研究员通过积分变换,率先提出了 P-III 分布参数估计的概率权重法计算公式,他们的研究成果至今被各国广泛地应用于水文频率分布的参数估计,并写入计算手册。美国地质勘查局也使用指数洪水分析和回归分析进行区域频率分析。除经验法外,现象法、动力法和随机模型结合 Monte Carlo 模拟法目前还处于学术研究层面,工程实际应用较少。非参数统计方法主要指在所处理对象总体分布族的数学形式未知情况下,对其进行统计研究的方法。而非参数估计就是在没有参数形式的密度函数可以表达时,直接使用独立同分布的观测值,对总体的密度函数进行估计的方法。主要包括概率密度核估计和非参数回归估计模型。

对于多变量联合概率分布来说,常用的分布模型有二维 Gamma 分布、Izawa bigamma 模型、Moran 模型、Smith-Adelfang-Tubbs (SAT)模型、Farlie-Gumbel-Morgwntern (FGM)模型、两变量乘积 XY 的二维 gamma 分布、二维 Gumbel Mixed (GM)分布、二维 Gumbel Logistic (GL)分布、二维 Nagao-Kadoya 指数 (BVE)分布、 d 维正态分布 $N(\mu, \Sigma)$ 、二维对数正态分布、 d 维 student t 分布 $N(\mu, \Sigma)$ 等。近年来,为了克服传统多变量概率分布计算的不足, Copulas 函数被引入多变量水文变量联合概率分布计算。主要的 Copulas 函数有对称 Archimedean Copulas、非对称 Archimedean Copulas、Plackette Copula、Metaelliptical Copulas 和混合 Copulas 等。常见的对称 Archimedean copulas 函数见表 1。copulas 函数参数估计方法有精确极大似然法、边际函数推断法和半参数法。

表 1 对称 Archimedean copulas 函数

名称	$C(u_1, u_2)$	$\varphi(t)$	θ
Nelsen No 1	$\max[(u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}, 0]$	$\frac{1}{\theta}(t^{-\theta} - 1)$	$[-1, \infty] \setminus \{0\}$
Nelsen No 2	$\max(1 - [(1 - u_1)^\theta + (1 - u_2)^\theta]^{\frac{1}{\theta}}, 0)$	$(1 - t)^\theta$	$[-1, \infty)$
Nelsen No 3	$\frac{u_1 u_2}{1 - \theta(1 - u_1)(1 - u_2)}$	$\ln \frac{1 - \theta(1 - t)}{t}$	$[-1, 1)$
Nelsen No 4	$e^{-[(- \ln u_1)^\theta + (- \ln u_2)^\theta]^{\frac{1}{\theta}}}$	$(- \ln t)^\theta$	$[-1, \infty)$
Nelsen No 5	$-\frac{1}{\theta} \ln \left[1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right]$	$-\ln \frac{e^{-\theta t} - 1}{e^{-\theta} - 1}$	$[-\infty, \infty) \setminus \{0\}$
Nelsen No 6	$1 - [(1 - u_1)^\theta + (1 - u_2)^\theta - (1 - u_1)^\theta(1 - u_2)^\theta]^{\frac{1}{\theta}}$	$-\ln[1 - (1 - t)^\theta]$	$[-1, \infty)$
Nelsen No 7	$\max[\theta u_1 u_2 - (1 - \theta)(u_1 + u_2 - 1), 0]$	$-\ln[\theta \cdot t - (1 - \theta)]$	$(0, 1]$
Nelsen No 8	$\max \left[\frac{\theta^2 u_1 u_2 - (1 - u_1)(1 - u_2)}{\theta^2 - (\theta - 1)^2(1 - u_1)(1 - u_2)}, 0 \right]$	$\frac{1 - t}{1 + (\theta - 1)t}$	$[-1, \infty)$
Nelsen No 9	$u_1 u_2 e^{-\theta \ln u_1 \ln u_2}$	$\ln(1 - \theta \ln t)$	$(0, 1]$
Nelsen No 10	$\frac{u_1 u_2}{[1 + (1 - u_1^\theta)(1 - u_2^\theta)]^{\frac{1}{\theta}}}$	$-\ln(2t^{-\theta} - 1)$	$(0, 1]$
Nelsen No 11	$\max([u_1^\theta u_2^\theta - 2(1 - u_1)^\theta(1 - u_2)^\theta]^{\frac{1}{\theta}}, 0)$	$\ln(2 - t^\theta)$	$(0, \frac{1}{2}]$
Nelsen No 12	$\{1 + [(u_1^{-1} - 1)^\theta + (u_2^{-1} - 1)^\theta]^{\frac{1}{\theta}}\}^{-1}$	$(\frac{1}{t} - 1)^\theta$	$[-1, \infty)$
Nelsen No 13	$e^{1 - [(1 - \ln u_1)^\theta + (1 - \ln u_2)^\theta - 1]^{\frac{1}{\theta}}}$	$(1 - \ln t)^\theta - 1$	$(0, \infty)$
Nelsen No 14	$\{1 + [(u_1^{-\frac{1}{\theta}} - 1)^\theta + (u_2^{-\frac{1}{\theta}} - 1)^\theta]^{\frac{1}{\theta}}\}^{-\theta}$	$(t^{-\frac{1}{\theta}} - 1)^\theta$	$[1, \infty)$
Nelsen No 15	$\max(\{1 - [(1 - u_1^{\frac{1}{\theta}})^\theta + (1 - u_2^{\frac{1}{\theta}})^\theta]^{\frac{1}{\theta}}\}^\theta, 0)$	$(1 - t^{\frac{1}{\theta}})^\theta$	$[1, \infty)$
Nelsen No 16	$\frac{1}{2}(S + \sqrt{S^2 + 4\theta})$ $S = u_1 + u_2 - 1 - \theta(\frac{1}{u_1} + \frac{1}{u_2} - 1)$	$(\frac{\theta}{t} + 1)(1 - t)$	$[0, \infty)$
Nelsen No 17	$\left\{ \frac{[(1 - u_1)^{-\theta} - 1][(1 - u_2)^{-\theta} - 1]}{2^{-\theta} - 1} \right\}^{-\frac{1}{\theta}} - 1$	$-\ln \frac{(1 + t)^{-\theta} - 1}{2^{-\theta} - 1}$	$(-\infty, \infty) \setminus \{0\}$
Nelsen No 18	$\max \left\{ 1 + \frac{\theta}{\ln(e^{u_1 - 1} + e^{u_2 - 1})}, 0 \right\}$	$\frac{\theta}{e^t - 1}$	$[2, \infty)$
Nelsen No 19	$\frac{\theta}{\ln(e^{u_1} + e^{u_2} - e^\theta)}$	$e^{\frac{\theta}{t}} - e^\theta$	$(0, \infty)$
Nelsen No 20	$[\ln(e^{u_1^{-\theta}} + e^{u_2^{-\theta}} - e)]^{\frac{1}{\theta}}$	$e^{t^{-\theta}} - e$	$(0, \infty)$

3 水文频率计算面临的挑战

水文频率计算虽然经历了 140 多年的研究和实践,由于水文序列组成和特性具有高度的复杂性,其理论体系与方法仍然需要不断地完善和发展,面临着许多挑战。

3.1 水文频率分布函数选择

目前,人们还无法从水文机理上证明水文事件的概率(频率)分布函数。实际中,计算者选用现有的随机变量分布函数,根据收集的水文数据进行审查,通过选用合理的参数估计方法进行分布函数参数计算,经分布函数拟合度检验后,依据水文数据的经验频率与选用分布函数理论频率之间的拟合效果来评估水文序列频率分布函数的拟合效果。实际上,这种处理是一种近似水文频率分布函数的选择

方法,缺乏相应的理论支撑。另外,现有分布函数随机变量的取值范围一般为 $(-\infty, \infty)$ 或 (a, ∞) 、 $(-\infty, a)$, a 为某一常数,而实际水文取值可在某个可能最小与最大之间,不可能取无穷大值,这种取值范围不符合水文值的实际取值范围,其计算的结果必然出现偏差。因此,现有随机单变量的概率分布实际上是对水文变量分布的逼近,频率拟合在 25% ~ 75% 的频率段一般能够取得较好的拟合效果,而频率曲线的上尾或下尾部(频率曲线的两端外延部分)拟合出现较大的偏差,无法得到满意的拟合效果。另外,在分析计算中,计算者往往需要在频率曲线的两端外延部分进行设计值推求。

3.2 水文数据

现有水文频率的计算方法必须满足概率论随机试验要求,依据随机事件原理计算水文事件概率,即

要求水文序列满足独立、同分布(一致性或平稳性)要求。实际中,水文序列不同程度地难以完全满足随机试验条件要求。

(1) 径流可能是由多个气候机制形成的。对于季节性径流选样(如分期洪水选样),其值可能是暴雨、融雪、冰川融化、飓风、台风等形成的径流。对于年值径流选样来说,同样暴雨、融雪、冰川融化、飓风、台风等参与径流的形成过程。显然,按照随机试验条件,假定流域下垫面不变、流域气候变化稳定和不受人类活动影响下,上述的序列也不满足平稳性,它们可能服从混合分布。

(2) 观测较早或历史推算洪水难以获取水文值发生期内流域下垫面和气候变化情况。洪水频率计算中,通常实测洪水序列与历史特大洪水共同组成不连序系列。在洪涝灾害的记载中,清代以前和清代初期的史籍中记载非常简单,洪水记录难以使用。因而,实际计算主要根据清代中后期的记载资料,选用距今 200 多年的洪水作为特大洪水。这些洪水形成的下垫面条件和气候条件与实测期的下垫面条件和气候条件有很大的差异。因此,这种不连序系列也难以满足平稳性要求。

(3) 在气候变化和高强度的人类活动影响的流域和区域,水文序列呈现出非平稳性特性,不能满足现有水文频率计算的前提和条件。

3.3 分布函数参数计算

按照随机过程原理,水文频率计算假定水文序列具有平稳各态历经性。即一个平稳序列的各种时间平均值依概率 1 收敛于相应的集合平均(大量样本函数在特定时刻取值通过统计方法计算平均值所得的数字特征值)。平稳各态历经性说明了随机序列发生的各个样本函数都同样地经历了随机序列的各种可能状态,任一样本函数的统计特性都可充分代表整个随机序列的统计特性,也可以用这个样本函数的时间平均来代替整个随机序列的集合平均值^[44-45]。显然,当样本函数序列长度无限大时,样本函数的时间平均等于随机序列的集合平均值,可满足水文频率分布函数的参数估计要求。但是,水文序列是一个有限长度的序列,特别是发展中国家,水文站建站较晚,其观测资料长度较短,无法满足这种计算条件。因此,无论采用任何参数估计方法,有限水文序列估计水文频率分布函数的参数不可避免地带来计算偏差。

3.4 经验频率

经验频率,也称绘点位置或秩次概率,是根据样本按递减(或递增)顺序排列,采用一定的计算方法估计样本每项值的频率,这个估计频率值称为经验

频率。目前,经验频率公式种类很多,代表性的计算公式有 California (1923) 公式、Hazen (1930) 公式、Weibull (1939) 公式、Leivikov (1955) 公式、Blom (1958) 公式、Tukey (1962) 公式、Gringorten (1963) 公式、Cunnane (1978) 和 Hosking. (1985) 公式。我国《水利水电工程设计洪水规范》推荐采用 P-III 型曲线和图解适线法推求设计频率对应的设计洪水值。在选定线型和适线准则的前提下,经验频率是评价参数估计优劣的依据。经验频率与理论频率偏差越小,则分布参数估计越好。因此,采用适线法确定水文频率分布参数或以经验频率作为参数方法评价依据,经验频率计算尤为重要,其计算精度主要取决于经验频率。严格来说,经验频率依赖于水文序列的频率分布函数。由于水文序列真正的频率分布函数无法确定,在实际水文分析中,广泛采用数学期望(Weibull, 1939)公式,这是一种近似的做法。

上述单变量频率计算的问题同样存在于多变量频率计算中:(1) 人们也无法获得机理性的联合概率分布函数,一般在变量组合取值的两端外延部分拟合出现偏差;(2) 按照 Copulas 函数计算原理,单变量频率计算精度直接地影响 Copulas 函数概率分布计算;(3) 单变量非平稳性导致相应的 Copulas 函数联合概率的非平稳性,其计算方法尤为复杂,计算方法带来的偏差难以估计;(4) 高维 Copulas 函数较少,而 Pair Copulas 构建高维联合概率密度函数容易,但是,其高维联合概率分布值需要通过高维数值积分进行计算,计算精度依赖于高维数值积分方法;(5) 高维的经验概率分布计算公式仍缺乏相应的理论支撑。

4 水文频率计算研究的建议

根据现有水文频率计算面临的挑战,本文建议深入开展以下研究工作,以期为我国水文频率计算提供参考和支撑。

(1) 删失或截取分布。按照时间序列分析观点,现有水文序列频率计算一般要求水文序列为完整序列。实际中,由于仪器分辨率、观测或选样要求,水文序列会出现由超过某一门限值或低于某一门限值的数据组成。这种序列是完整序列的一个子集,称为部分序列,超出了时间序列分析领域。而现有的水文频率分布是假定序列为完整序列,部分序列采用完整序列的频率计算方法,显然是合理的。另外,水文序列分布为有界取值概率分布函数,应用 $(-\infty, \infty)$ 或 (a, ∞) 、 $(-\infty, a)$ 取值的频率分布进行拟合,实际上是一种近似计算。统计学中删失或截取分布随机变量取值范围恰好具备部分序列的取

值区间。因此,删失或截取分布可能是提供上述部分序列频率的计算途径。目前,删失或截取分布应用于工业质量、产品寿命和金融等领域,取得了许多成功案例。因此,有必要进一步研究删失或截取分布在水文中的应用。

(2) 提高频率分布参数估计精度。如上所述,水文频率分参数估计方法很多。从水文频率计算研究结果来看,最大熵原理取得了较好的拟合效果。其原因是,许多分布函数经过复杂的数学推导,最终频率分布参数计算归结频率分布参数函数的数学期望方程或方程组,属于函数的一阶矩或二阶矩计算,避免了高阶函数矩计算。因此,低阶矩约束条件下的最大熵原理推求水文频率分参数估计有待深入研究。另外,如果计算站资料长度较短,可选用计算站临近或相似流域具有长系列的测站资料,通过序列的相依性分析,进行 Copulas 联合分布计算,进而推求计算站的设计频率和设计值。这也是一种提高频率分布参数估计精度的一个途径,有待于深入研究。

(3) 实用的非平稳水文序列频率计算方法。由于气候变化和高强度的人类活动的影响,改变了流域的降雨特性、产汇流和河道水流的天然时空分配规律,使长序列的统计特性发生变化。另外,径流可能是由多个气候机制形成的。因而,这些不可避免地造成计算期内资料序列的非平稳性。目前,非平稳水文序列频率计算方法主要有水文极值系列重构、混合分布函数、水文物理机制洪水频率和时变参数概率分布等^[41-42]。这些计算方法仍旧处在研究层面,各种方法计算结果差异较大,一般计算过程较为复杂,且重点关注了非平稳水文序列的理论频率计算,相应的经验频率和评估依据缺乏充分的理论依据支撑。因此,亟待深入研究实用的非平稳水文序列频率计算方法,以满足实际水文计算的需要。

(4) 基于 Copulas 随机变量和差积商分布解析计算。流域设计断面各部分洪水的地区组成、干旱分析中干旱历时与非干旱历时变量和(干旱间隔)分布、干旱历时与干旱间隔的比值(干旱事件比例)分布、干旱强度与干旱历时的乘积(干旱烈度)分布、水资源评价和配置中研究断面来水与区间耗水量的组合分布分析等都可以归结为水文变量和、差、积、商的分布计算。《水利水电工程设计洪水计算手册》推荐使用地区组合法、频率组合法及随机模拟法进行设计洪水的地区组成计算。这些方法在积分计算、信息失真、模拟序列统计特征保持性和人为因素等存在不足。因此,提高水文变量和、差、积、商的分布计算精度是流域水资源梯级开发和水资源管理中亟待解决的重要科学问题。

(5) 水文设计值置信区间估计。假定水文序列数据满足计算要求的前提下,水文样本长度、分布函数和参数估计方法等均会对水文设计值产生不确定性。分布函数给定下,参数估计方法不同,其水文设计值置信区间估计方法不同。目前,矩法和极大似然估计参数的水文设计值置信区间一般有相应的计算公式。但是其他估计方法和许多新型参数估计方法的水文设计值置信区间估计研究较少。另外,依据经验概率和理论概率的 P-P 图、实测数据和计算数据的 Q-Q 图或其他误差评定基本上属于拟合度的定性评价,拟合度必须通过分布函数的拟合度检验进行分析。因此,有必要开展量化水文设计值置信区间估计的不确定性研究。

(6) 多变量联合概率分布。对称 Archimedean Copulas 也称为可交换。其最大的特点是有一个参数,计算简单,应用较多。实际中,水文变量间可能是正、负相依性或独立的,他们也可能是相依性不对称的样本序列组合。对于二维变量,其联合分布可用对称 Archimedean Copulas 描述,但是,由于对称 Archimedean Copulas 要求变量为对称相依,仅用一个生成函数描述正的相依性。因而,三维以上对称的 Archimedean Copulas 不是描述高维变量联合概率分布最好的 Copulas 函数选择。因此,嵌套 Archimedean 构造、层次 Archimedean Copulas 和配对 Copula 等非对称 Archimedean Copulas 的构建方法和应用,以及非平稳下的多变量联合概率分布计算有待于进一步深入研究。

参考文献:

- [1] 水利部长江水利委员会水文局.水利部南京水文水资源研究所.水利水电工程设计洪水计算手册[M].北京:中国水利水电出版社,2001.
- [2] 黄振平,陈元芳.水文统计学[M].北京:中国水利水电出版社,2017.
- [3] 金光炎.水文统计理论与实践[M].南京:东南大学出版社,2012.
- [4] 周镇伦.全年雨量之常率线及常率积分线[J].水利,1932,5(5):15-63.
- [5] 须 恺.淮河洪水之频率[J].水利,1935,5(2):39-46.
- [6] 陈椿庭.中国五大河洪水量频率曲线之研究[J].水利,1947,14(6):240-287.
- [7] 陈椿庭.七十五年水工科技忆述[M].北京:中国水利水电出版社,2012.
- [8] 陈椿庭.水工水力学及水文论文集[C].北京:水利电力出版社,1993.
- [9] Charles Thomas Haan. Statistical Methods in Hydrology[M]. Second Edition. Iowa State Press,2002.
- [10] Mauro Nighettini. Fundamentals of Statistical Hydrology

- [M]. Springer International Publishing Switzerland, 2017.
- [11] Rajib Maity. Statistical Methods in Hydrology and Hydroclimatology[M]. Springer Nature Singapore Pte Ltd, 2018.
- [12] Rao A R, Hamed K H. Flood Frequency Analysis[M]. CRC Press LLC.: Washington D. C., USA, 2000.
- [13] Meylan Paul, Favre Anne Catherine, Musy Andre. Predictive Hydrology: A Frequency Analysis Approach[M]. Science Publishers. Jersey, British Isles, 2012.
- [14] Singh V P. Entropy - Based Parameter Estimation in Hydrology[M]. Kluwer Academic Publishers: Dordrecht, The Netherlands, 1998.
- [15] 王俊德. 水文统计学[M]. 北京: 中国水利水电出版社, 1992.
- [16] 丛树铮. 水文学的概率统计基础[M]. 北京: 水利水电出版社, 1980.
- [17] 郭生练. 设计洪水研究进展与评价[M]. 北京: 中国水利水电出版社, 2005.
- [18] 丛树铮. 水科学技术中的概率统计方法[M]. 北京: 科学出版社, 2010.
- [19] 孙济良, 秦大庸, 孙翰光. 水文气象统计通用模型[M]. 北京: 中国水利水电出版社, 2001.
- [20] 陈元芳. 统计试验方法及应用[M]. 哈尔滨: 黑龙江人民出版社, 2000.
- [21] 秦毅, 张德生. 水文水资源应用数理统计[M]. 西安: 陕西科学技术出版社, 2006.
- [22] 金光炎. 水文统计的原理与方法[M]. 北京: 水利电力出版社, 1958.
- [23] 金光炎. 实用水文统计法[M]. 北京: 水利电力出版社, 1958.
- [24] 金光炎. 水文统计计算[M]. 北京: 水力出版社, 1980.
- [25] 金光炎. 水文水资源随机分析[M]. 北京: 中国科学技术出版社, 1993.
- [26] 金光炎. 工程数据统计分析[M]. 南京: 东南大学出版社, 2002.
- [27] 金光炎. 水文水资源分析研究[M]. 南京: 东南大学出版社, 2003.
- [29] 张济世, 刘立昱, 程中山, 等. 统计水文学[M]. 郑州: 黄河水利出版社, 2006.
- [34] 谢平, 陈广才, 雷红富. 变化环境下地表水水资源评价方法[M]. 北京: 科学出版社, 2009.
- [28] 金光炎. 水文水资源计算务实[M]. 南京: 东南大学出版社, 2010.
- [30] 程根伟, 黄振平. 水文风险分析的理论与方法[M]. 北京: 科学出版社, 2010.
- [31] 宋松柏, 蔡焕杰, 金菊良, 等. Copulas 函数及其在水文中的应用[M]. 北京: 科学出版社, 2012.
- [32] 宋松柏, 康艳, 宋小燕, 等. 单变量水文序列频率计算原理与应用[M]. 北京: 科学出版社, 2018.
- [33] 熊立华, 郭生练, 江聪. 非一致性水文概率分布估计理论和方法[M]. 北京: 科学出版社, 2018.
- [35] 谢平, 许斌, 章树安, 等. 变化环境下区域水资源变异问题研究[M]. 北京: 科学出版社, 2012.
- [36] 胡义明, 梁忠民. 变化环境下的水文频率分析方法及应用[M]. 北京: 中国水利水电出版社, 2017.
- [37] 王凯, 梁忠民, 胡友兵. 淮河复合河道洪水概率预报方法与应用[M]. 北京: 中国水利水电出版社, 2017.
- [38] 陈璐. Copula 函数理论在多变量水文分析计算中的应用研究[M]. 武汉: 武汉大学出版社, 2013.
- [39] Singh V P, Strupczewski W G. On the status of flood frequency analysis[J]. Hydrological Processes, 2002, 16: 3737-3740.
- [40] 郭生练, 闫宝伟, 肖义, 等. Copula 函数在多变量水文分析计算中的应用及研究进展[J]. 水文, 2008, 28(3): 1-7.
- [41] 梁忠民, 胡义明, 王军. 非一致性水文频率分析的研究进展[J]. 水科学进展, 2011, 22(6): 864-871.
- [42] 熊立华, 江聪, 杜涛, 等. 变化环境下非一致性水文频率分析研究综述[J]. 水资源研究, 2015, 4(4): 310-319.
- [43] 郭生练, 刘章君, 熊立华. 设计洪水计算方法研究进展与评价[J]. 水利学报, 2016, 47(3): 302-314.
- [44] 王文圣, 丁晶, 金菊良. 随机水文学[M]. 3 版. 北京: 中国水利电力出版社, 2016.
- [45] 卜雄洙, 吴健, 牛杰. 现代信号分析与处理[M]. 北京: 清华大学出版社, 2018.

特邀作者简介

宋松柏, 工学博士, 西北农林科技大学教授、水文学及水资源专业博士生导师; 中国农业工程学会农业系统工程专业委员会副主任委员, 江西省鄱阳湖水资源与环境重点实验室学术委员会委员; 中国统计水文学会会员, 高等学校水利类水文与水资源工程教学研究会委员, 中国自然学会水资源专业委员会常务委员, 中国水力发电工程学会水水泥沙专业委员会委员; 美国 Texas A & M 大学访问学者; 《水利与建筑工程学报》、《南水北调与水利科技》和《水资源与水工程学报》编委。

主要从事水文不确定性、水文模拟与预报、水文干旱、水资源可持续利用等教学与科研工作。主持国家自然科学基金项目、参加国家重点研发计划项目和省部级项目等 20 余项。在《Stochastic Environmental Research and Risk Assessment》、《Hydrological Processes》、《Journal of Hydrology》、《水利学报》、《水科学进展》、《水力发电学报》、《生态学报》、《应用生态学报》、《农业工程学报》和《自然资源学报》等期刊发表论文 180 余篇; 在科学出版社出版《Copulas 函数理论及其在水文中的应用》、《单变量水文序列频率计算原理与应用》专著。