第 9 卷◆第 9 期◆版本 1.0◆2025 年
文章类型：论文|刊号（ISSN）：2529-7821 /（中图刊号）：868GL002

基于贝叶斯推断的暴雨洪水频率曲线参数不确定性研究

郎喜白
广东省水文局湛江水文分局
DOI:10.12238/hwr.v9i9.6560

[摘　要] 本文基于贝叶斯MCMC方法,对雷州半岛三站年最大24h暴雨和洪水极值序列进行频率分析,采用P-III和GEV分布拟合参数,通过先验设置与后验抽样估计参数及不同重现期设计值。结果表明：GEV模型在描述极端尾部特征时优于P-III,其后验不确定性更小；P-III偏度参数对特大值敏感但整体稳健。贝叶斯方法所得置信区间覆盖传统频率结果,为热带季风区暴雨洪涝风险分析和防洪设计提供了可靠的不确定性量化依据。
[关键词] 贝叶斯推断；MCMC；广义极值分布；Pearson-III分布；暴雨频率分析；雷州半岛
中图分类号：TV122+.1　文献标识码：A

A Bayesian Inference Approach to Quantifying Parameter Uncertainty in Storm-Flood Frequency Curves

Xibai Lang
Guangdong Provincial Bureau of Hydrology,Zhanjiang Hydrological Branch

[Abstract] This study applies a Bayesian MCMC approach to conduct frequency analysis of annual maximum 24-hour rainfall and flood extremes recorded at three stations on the Leizhou Peninsula. The Pearson type III (P-III) and generalized extreme value (GEV) distributions were adopted to model the data, with parameters and design values for different return periods estimated through prior specification and posterior sampling. The results indicate that the GEV model outperforms the P-III in characterizing extreme tail behavior, exhibiting smaller posterior uncertainty. While the skewness parameter of the P-III distribution is sensitive to extreme values, the overall model remains robust. The confidence intervals derived from the Bayesian framework encompass those obtained from conventional frequency methods, providing a quantitative basis for uncertainty assessment in rainstorm and flood risk analysis, as well as flood control design in tropical monsoon regions.
[Key words] Bayesian inference; MCMC;Pearson type III distribution; GEV distribution; flood frequency analysis

引言

暴雨作为一种典型的极端气象现象,其发生频率与强度的变化直接关系到区域防洪减灾、城市排水、基础设施设计、水资源管理等多方面的安全与可持续发展。暴雨频率曲线参数的准确估计在水资源管理、防洪减灾、城市排水规划以及应急响应策略中具有至关重要的意义。雷州半岛由于地处热带季风气候区,降雨具有强烈的不均匀性和时空分布特点,暴雨事件频发且具有极端性,这使得该区域的暴雨重现期及极端降雨量评估成为亟待解决的关键问题。本研究旨在基于贝叶斯推断方法,对雷州半岛多个降雨站点的长历时降雨数据进行极值分析,并采用Pearson-III与GEV两种常用分布模型,探讨各模型参数的后验推断结果及其联合分布情况,为暴雨重现期计算和极值评估提供科学依据。

1 研究方法

1.1贝叶斯分析的基本原理

贝叶斯定理是关于随机事件A和B的条件概率：

$$P(\theta|X)=\frac{P(X|\theta)P(\theta)}{P(X)} \tag{1}$$

其中,P(θ)是θ的先验概率。P(θ|X)是已知X发生后θ的条件概率,称作θ的后验概率。P(X|θ)是已知θ发生后X的条件概率,称作X的后验概率。P(X)是X的先验概率,并且

$$P(X)=\int d\theta^{*}P(X|\theta^{*})P(\theta^{*}) \tag{2}$$

在这里,P(X|θ*)是似然函数,P(θ*)是先验分布,P(X)是归一化常数,也称为证据或边缘似然。计算上的问题是难以评估

文章类型：论文|刊号（ISSN）：2529-7821 /（中图刊号）：868GL002

分母中的积分。可以使用蒙特卡洛方法,其中最重要的是马尔可夫链蒙特卡洛(MCMC)。对于贝叶斯模型,积分中的分布p(x)是后验分布。使用MCMC,从(简单的)提议分布中抽取样本,以便每次抽样仅依赖于前一次抽样的状态(即样本形成马尔可夫链)。在一定条件下,马尔可夫链将具有唯一的稳定分布。为每次抽样设置接受准则,基于比较连续状态与目标分布的差异,以确保稳定分布是所关注的后验分布。经过一段时间,被接受的抽样的马尔可夫链将收敛到稳定分布,并且将这些样本用作来自后验分布的(相关的)抽样,并以与普通蒙特卡洛积分相同的方式找到后验分布的函数。

1.2后验分布的MCMC模拟

定义π_i代表分布π下,状态i出现的概率。如果π是连续分布,则π_i为分布的PDF(概率密度函数)在i处的取值,即$\pi(i)$。

定义q_{ij}为已知状态转移矩阵Q中状态i转移到j的概率,即

$$q_{ij}=P(Y=j|X=i) \tag{3}$$

假设已经采样到$X_n=i$,那么下一个采样X_{n+1}的计算公式如下。采样Y=j,其中Y=j的PDF为$f(j)=q_{ij}$,计算提议概率：

$$a_{ij}=\min(1,\frac{\pi_j q_{ji}}{\pi_i q_{ij}}) \tag{4}$$

采样u～U(0,1),则X_{n+1}为：

$$X_{n+1}=\begin{cases} Y, & u<a_{ij} \\ X_n, & u\geq a_{ij} \end{cases} \tag{5}$$

通过迭代重复上述步骤。这样对足够大的n值,序列X_{M+1},X_{M+2},…,X_{M+N}可以认为是后验分布的抽样,和独立样本一样,可以用这个序列来估计一些数字特征,比如后验均值。这种方法可以完全不考虑所选的q(当然必须服从某些正则条件)。从极限意义上讲,式(5)给出的拒绝步骤确保了模拟序列的平稳分布具有所需的边缘分布。此抽样方法称为Metropolis-Hastings算法,是较早出现且最有名的MCMC方法,它由Metropolis等人提出,之后由Hastings 对其加以推广而形成。

1.3对Pearson-III分布的模拟

Pearson-III分布是Gamma函数的移位变形,其概率密度函数为

$$f(x,\kappa)=\frac{|\beta|}{\Gamma(\alpha)}(\beta(x-\zeta))^{\alpha-1}\exp(-\beta(x-\zeta)) \tag{6}$$

其中κ为偏度Cs,式(6)也称为密度函数的标准化形式定义,并且其中

$$\beta=\frac{2}{\kappa} \tag{7}$$

$$\alpha=\beta^2=\frac{4}{\kappa^2} \tag{8}$$

$$\zeta=-\frac{\alpha}{\beta}=-\beta \tag{9}$$

更进一步使用位置参数loc,形状参数shape,尺度参数scale表示,则函数的参数传递可写为：

pearson3(loc=ex,shape=Cs, scale=ex*Cv) (10)

对于绘制区域暴雨等值线的目的,显然应锁定均值ex,如按照常用的取Cs/Cv=3.5的情况,则只需要以Cv为自变量；更广泛的情况是以Cv、Cs两个自变量,进而探讨Cv、Cs的联合分布关系。

1.4 GEV分布的模拟

$$F(x)=\begin{cases} \exp\{-\left[1+\xi\frac{x-\mu}{\sigma}\right]^{-1/\xi}\} & \xi\neq 0 \\ \exp\{-\exp[-\frac{x-\mu}{\sigma}]\} & \xi=0 \end{cases} \tag{11}$$

其中：μ、σ、ξ分别为GEV分布函数的位置参数、尺度参数和形状参数,并满足$\mu\in R$, $\sigma>0$, $\xi\in R$, $1+\xi(x-\mu)/\sigma>0$,这三个参数并不能直接由序列均值、方差计算,即GEV的分布函数不能表示成genextreme(ex,Cv,Cs)的形式。

2 实例分析

雷州半岛地处北回归线附近,受海陆季风与地形抬升共同影响,暴雨频次高、强度大,对当地农业、基础设施和生态环境造成严重威胁。半岛东部为鉴江入海口,西部北部为九洲江流域,地形以沿海台地为主,河流多为流域面积200km²左右独流入海。选取东部序列较长(1951～2024)、代表性较好的梅菉站各历时暴雨序列、中部序列长度较短(2007～2024)但有特大值的幸福农场站各历时暴雨序列,以及九洲江控制站缸瓦窑站年最大洪水流量序列,其中缸瓦窑站序列包括1955～2023的实测值和1906、1914、1942的调查值,实测序列中,1959～1969因故停测。以此检验贝叶斯推断方法的适应性。为验证不同分布函数的适应性,采用均值固定的以Cv、Cs为参数的皮尔逊III型(P-III)分布和三参数的GEV分布进行拟合。

2.1分布参数的贝叶斯估计

分别对梅菉、幸福农场2个站的年最大24h雨量、缸瓦窑水文站实测年最大流量建立分布模型。采用Metropolis-Hastings算法产生随机样本,用去除前k个不稳定样本后的序列对分布参数进行统计推断。其中,P-III分布的先验密度函数为$\pi(Cv,Cs)=\pi_{Cv}(Cv)\pi_{Cs}(Cs)$,参数Cv、Cs相互独立;GEV分布的先验密度函数为$\pi(\mu,\sigma,\xi)=\pi_{\mu}(\mu)\pi_{\sigma}(\sigma)\pi_{\xi}(\xi)$,参数$\mu$,$\sigma$,$\xi$相互独立。分布的参数$\pi_{Cv}(Cv),\ldots,\pi_{\xi}(\xi)$等为不同方差的正态分布概率密度函数。通过调整每个参数的方差,可得到不同的接受率。

以梅菉站为例。梅菉站年最大24h雨量P-III模型经过20000次采样产生的序列链图。其中,前2000次采样称为燃烧期(burn-in period),主要为采样收敛过程,剩余模拟值即为后验分布观测值。表1为3个站点基于贝叶斯理论MCMC法的P-III分布和GEV分布参数后验分布的统计特征(置信水平为90%)。与传统参数的估计方法相比,贝叶斯法不仅给出了参数的估计 值,还给出了参数估计值的置信区间,从而充分论证了能量化模型参数估计的不确定性。

2.2拟合检验

以梅菉站为例,根据梅菉站年最大24h雨量P-III分布模型贝叶斯估计分位数拟合图。图中,不同分布的样本点大体集中在

文章类型：论文|刊号（ISSN）：2529-7821 /（中图刊号）：868GL002

45°线,但特大值偏离明显不同,表明对特大值的估计仍需进一步分析。根据生成参数后验值,可计算模拟后验观测估计值 $\hat{x}_N$,以置信水平90%计算,可以得到估计值的90%置信区间上下限,也可以得到均值、中位数曲线。

表1 各站点P-III、GEV分布参数估计

站名	分布模型	参数	贝叶斯估计		
			均值	5%下限	95%上限
缸瓦窑	P-III	Cs	1.385	1.014	1.7
		Cv	0.619	0.531	0.715
	GEV	μ	678.8	675.2	684.2
		σ	635.8	630.2	641.1
		ξ	0.126	-0.040	0.341
梅菉	P-III	Cs	0.959	0.573	1.3
		Cv	0.402	0.35	0.463
	GEV	μ	137.5	132.3	142.0
		σ	53.3	46.1	57.9
		ξ	0.034	-0.104	0.191
幸福农场	P-III	Cs	2.826	2.056	3.846
		Cv	0.752	0.561	1.012
	GEV	μ	100.0	88.7	113.4
		σ	190.2	185.0	195.8
		ξ	0.170	-0.056	0.446

2.3重现期水平的估计

概率分布函数(PDF)的逆函数是累积分布函数(CDF),P-III分布的累积分布函数一般用不完全Gamma函数表示

$$F(X\leq x;\alpha,\mu,\sigma)=\frac{\gamma\left(\frac{4}{\alpha^2},\frac{x-\mu_0}{\theta}\right)}{\Gamma\left(\frac{4}{\alpha^2}\right)} \tag{12}$$

其中,α,μ,σ分别为偏度,位置参数,形状参数,$\theta=(\alpha\sigma)/2$,$\mu_0=\mu-2\sigma/\alpha$；对于GEV分布,其CDF表达式为：

$$F(X\leq x;\xi,\mu,\sigma)=\begin{cases}\exp\left[-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{\frac{1}{\xi}}\right],\xi\neq0,1+\xi\frac{x-\mu}{\sigma}>0\\ \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right],\xi=0\end{cases} \tag{13}$$

利用分布的累积分布函数,可计算某暴雨洪水值该分布下的非超过概率p,其总体为参数后验长度的集合,可计算其均值、中位数、众数,相应重现期T=1/p。表2给出了三个站最大值的不同方法的重现期估计,其中P-III方法未进行特大值考证。由表可以看出,贝叶斯估计的置信区间总可以覆盖P-III方法的重现期。对于特大值偏差较大的序列,P-III贝叶斯估计的参数后验均值计算的设计值与GEV估计的设计值众数较为接近,差不多可以认为是特大值的考证年限。如幸福农场站,按照暴雨图集查得的参数,其最大24h雨量1188.2mm计算的重现期要在2000a以上,远大于贝叶斯两种估计的值。

表2 各站点年最大值的重现期估计

站名	特大值	P-III	贝叶斯估计					
			方法	90%	推断参数	平均值	中位数	众数
幸福农场	1188.2	74.4	P-III	79.8-2996.3	292.3	237.2	422.3	499.7
			GEV	19.3-595.4	54.9	47.1	59.5	285.6
梅菉	384.5	158.9	P-III	67.7-597.3	160.6	147.5	167.6	199.9
			GEV	25.7-557.5	74.6	66.6	81.3	181.7
缸瓦窑	3740	87.3	P-III	35.5-147.2	61.3	60	62.5	74
			GEV	18.2-273.8	43.6	40.8	47.5	80

3 结论

本文基于P-III分布、GEV分布与Metropolis-Hastings算法进行贝叶斯MCMC暴雨洪水频率分析研究,并应用于雷州半岛区域降雨与水文站点。通过对比模型的参数估计结果,可以发现：(1)P-III型分布在描述暴雨的整体分布特性方面较为稳定,但其偏度参数容易受到极少数异常值的影响,从而导致后验分布的扩散较大。这一特点提示在应用P-III型分布时,需要特别注意数据极值的处理和先验分布的合理设定。(2)GEV分布在暴雨极值事件分析中表现出较好的适应性,其形状参数能够捕捉暴雨尾部的变化特性。基于贝叶斯MCMC方法得到的后验分布显示出较小的不确定性,且参数间相关性符合预期,表明GEV分布对于区域暴雨分析具有较高的可靠性。(3)特大值的偏离可能揭示了序列的代表性不足,贝叶斯估计的成果可以作为特大值考证年限的重要参考。

[参考文献]

[1]水利水电工程设计洪水计算规范:SL 44-2006[S],2006.

[2]胡静,王帅人,覃光华,等.基于地区线性矩法的四川省24h极值降雨频率分析[J].人民长江,2024,55(01).

[3]史道济.实用极值统计方法[M].天津:天津科学技术出版社,2006.

作者简介：

郎喜白(1979--),男,汉族,河南邓州人,大学本科,高级工程师。研究方向为水文水资源。