强降水发生的概率估计与比较*

邹 波
（民航飞行学院交通分院空管系，广汉 618307）

摘 要 年以下时间尺度的降水量分布常在极值部分比高斯分布有着更大的发生值，为厚尾分布。作者根据在极值理论中处于重要地位的GP分布模型（Generalized Pareto distribution），讨论了分布模型的一些性质．运用该模型对月和日降水量的强降水概率进行了分布拟合，得到各站强降水发生的直观年数；此外，重点分析了模型中--个重要的参数——尾指数的算法和意义，并使用尾指数来比较各地强降水分布尾部的大小，对预防洪涝灾害提供了一个科学的依据。

关键词：厚尾分布；参数估计；尾指数；强降水；概率估计

1 引言

1991 年和 1998 年长江中下游出现了大水。一个值得注意的问题是：这样的大水能否从统计中确定其发生概率，而不是只从观测样本中来谈所谓的“百年一遇、十年一遇”。多年来许多气候学家或统计学家曾致力于这方面的研究，提出各种解析函数模式来拟合长年累积的气候资料。研究成果表明，就降水而言，不同的降水指标，其概率分布模式是不一样的，即使同一种降水指标，在不同的地区，其概率分布模式也有差异[1]；其次，降水的涨落可以有很宽的强度范围，年以下时间尺度降水量的分布往往是非常偏斜的“厚尾”分布[2]。所谓厚尾分布，粗略地说，就是它的极值实现值要比正态分布的时候大，并且出现更频繁。衡量分布尾部厚薄程度的参数为尾指数[3]，正态分布的尾指数为零（其尾部呈指数函数衰减）。尾指数大于零时，分布尾部呈幂函数衰减，为厚尾分布，并且尾指数越大，其尾部越厚。因此尾指数可以作为强降水情况的指示量。本文采用 1951 年 1 月至 1999 年 2 月重庆、九江、南京的月平均降水量资料和 1964 年 2 月至 1991 年 6 月常州的日、候降水量资料为例，运用极值统计理论中处于重要地位的 GP 分布模型（Generalized Pareto distribution）来拟合它们的强降水事件，并说明了尾指数的意义。

2 厚尾分布

为了说明这部分强降水的多发性，把它和高斯分布做比较。由于大的降水量是一种小概率事件，为了说明大于某降水量事件的发生率，文中所例相关图形都采用了超

2002-03-27 收到，2003-05-29 收到再改稿
* 国家重点基础研究发展规划项目 G1998040900 和国家自然科学基金资助项目 40035010 资助

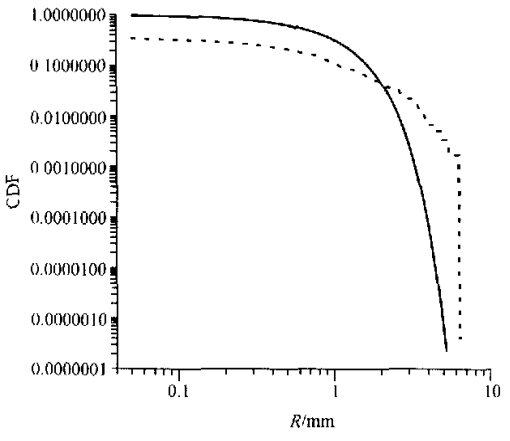

图 1 高斯分布（实线）和南京站实际的超出累积分布（虚线）

出累积概率分布的概念。如图 1 所示，图中实线是标准正态分布的超出累积概率曲线，虚线是经过标准化处理后的南京站月平均降水量的超出累积分布曲线。图中纵坐标是分布函数值，横坐标是标准化的月降水量，为了突出表现和比较月降水的高值部分（即分布的尾部），已对横纵坐标分别取了以 10 为底的对数。从图中可见，随着降水量的增长，高斯分布很快趋近于 0，而实际降水还有很高的分布值。例如当标准化月降水量为 5.0 时，高斯分布的累积分布函数值已为 6.5×10^{-7}，而实际月降水在此处的累积分布值还有 3.5×10^{-3}，几乎是高斯分布的 5 000 多倍；标准化月降水量为 6.0 时，高斯分布的累积分布函数值已为 2.3×10^{-9}，而实际月降水在此处的累积分布值还有 1.7×10^{-3}，几乎是高斯分布的 10^5 倍。可见，对相对大量随机事件所满足的正态分布来说，月降水量的概率统计分布在远离平均值处仍有较高的几率，是一种厚尾分布。目前常用的一些统计函数如指数分布、泊淞分布等都很难拟合出强降水部分，往往使人们低估了这部分强降水发生的可能性，给生产生活造成巨大的损失。

3 GP 分布模型

3.1 GP 分布的定义

记

$$W_\gamma(x)=1-(1+\gamma x)^{-1/\gamma},\ x\in\left[0,\frac{1}{\max(0,-\gamma)}\right] \tag{1}$$

式中 γ 为尾指数。对 $\gamma=0$，$W_\gamma(x)$ 为 $W_\gamma(x)$ 当 $\gamma\to0$ 时的极限，即 $W_0(x)=1-e^{-x}$，$x>0$。称 $W_\gamma(x)\kappa$ 为标准的 GP 分布，如果加进位置参数 μ、尺度参数 σ，则有

$$W_{\gamma,\mu,\sigma}(x)=1-\left(1+\gamma\frac{x-\mu}{\sigma}\right)^{-1/\gamma},\ \sigma>0,\ x-\mu\in\left[0,\frac{\sigma}{\max(0,-\gamma)}\right] \tag{2}$$

所谓 GP 分布模型，指的是分布族

$$G_p=\{W_{\gamma,\mu,\sigma}(x):-\infty<\gamma<\infty,-\infty<\mu<\infty,\sigma>0\},$$

事实上，G_p 是由以下 3 种类型的分布组成：

指数分布：$W_0(x)=1-e^{-x}$，$x>0$，

Pareto 分布：$W_{1,\gamma}(x)=1-x^{-1/\gamma}$，$\gamma>0$，$x\geqslant1$，

Beta 分布：$W_{2,\gamma}(x)=1-(-x)^{-1/\gamma}$，$\gamma<0$，$-1\leqslant x\leqslant0$，

即 GP 分布可分成上述 3 个不相交的子类。

3.2 GP分布可作为高门限[1]（High threshold）超出的参数模型

设数据 x_i，$i=1$，2，… 来自一个未知的分布 F，记

$$\omega(F):=\sup\{x:F(X)<1\},$$

称 $\omega(F)$为 F 的上端点。设 $u<\omega(F)$是一个门限，记

$$F^{[u]}(x)=\frac{F(x)-F(u)}{1-F(u)}=P(X\leqslant x\mid X>u),\ x\geqslant u, \tag{3}$$

称 $F^{[u]}$为门限 u 超出的分布。

Balkema 和 de Haan[4]，以及 Pickands[5]证明了如下定理：

定理　若存在常数 $\alpha_u>0$，b_u 使当 $u\to\omega(F)$时 $F^{[u]}(\alpha_u x+b_u)$有连续的极限分布，则

$$F^{[u]}(x)-W_{\gamma,u,\sigma_u}(x)\to 0,\ u\to\omega(F), \tag{4}$$

对某个 γ 和 σ_u（与 u 有关）成立。此定理表明在很广泛的情况下，GP 分布可作为高门限超出分布的近似分布。

3.3 用GP分布拟和数据的尾部

设样本 x_1，…，x_n 的经验分布函数为 $\hat{F}_n$，因为

$$\sup_{x\in R}|\hat{F}_n(x)-F(x)|\to 0\text{a.s.},\ n\to\infty,$$

故对 $x\geqslant u$，

$$\hat{F}_n^{[u]}(x)-F^{[u]}(x)\to 0,\ n\to\infty. \tag{5}$$

由定理知，当 $u<\omega(F)$时，对某 γ 和 σ_u 有

$$\frac{F_n(X)-\hat{F}_n(u)}{1-\hat{F}_n(u)}\approx W_{\gamma,u,\sigma_u}(x),\ x\geqslant u, \tag{6}$$

从而

$$\hat{F}_n(x)\approx[1-F_n(u)]W_{\gamma,u,\sigma_u}(x)+F_n(u)=W_{\gamma,u,\sigma}(x),\ x\geqslant u, \tag{7}$$

其中

$$\sigma=\frac{\sigma_u}{1+\gamma W_\gamma^{-1}\hat{F}(u)}=\sigma_u\left(\frac{\kappa}{n}\right)^\gamma \tag{8}$$

$$\mu=u-\sigma W_\gamma^{-1}\hat{F}_n(u)=u-\frac{\sigma_u}{\gamma}\left[1-\left(\frac{\kappa}{n}\right)^\gamma\right], \tag{9}$$

这里的 κ 为 x_1，…，x_n 中超出门限 u 的个数，即 $\kappa=\sum_{i=1}^n I(X_i\geqslant u)$，而 σ_u 可由 u、γ 和方程 $\hat{F}_n^{[u]}(x)=W_{\gamma,u,\sigma_u}(x)$，$(x>u)$决定，可解得

$$\sigma_u=\frac{\gamma(x-u)}{[1-\hat{F}_n^{[u]}(x)]^{-\gamma}-1} \tag{10}$$

由上述推导可见，GP 分布 $w_{\gamma,u,\sigma}$可拟和经验分布函数 $\hat{F}_n$（或数据）的尾部。

$W_{\gamma,\mu,\sigma}(x)$中的参数 γ 为形状参数，通常称为 F 的尾指数。它决定了 F 尾部的厚度，如，$\gamma=0$ 表明 F 的上尾部与指数分布的尾部相仿；$\gamma>0$ 表明 F 的上端点为$+\infty$，$1-F(x)$衰减的速度为 $x^{-1/\gamma}(x\to\infty)$；$\gamma<0$ 表明 F 的上端点 $\omega(F)$为有限的，当 $x\to\omega(F)$时，$1-F(x)$的衰减速度为$[\omega(F)-x]^{-1/\gamma}$。因此，$\gamma$ 的估计值是数据分布尾部厚度的一个指示量。尾指数越大（小），分布尾部越厚（薄）。

1）称 u 为高门限，意指 $u<\omega(F)$且与 $\omega(F)$很靠近。

3.4 GP分布的参数估计

由3.3节的讨论可见，对选定的门限 u，用来拟和数据尾部的GP分布 $W_{\gamma,\mu,\sigma}(x)$的参数 γ、μ、σ 的估计值就可由（8）、（9）和（10）式得出。因此，对于一组来自于分布 F 的样本 X_1，…，X_n，最基本的一步是给出尾指数 γ 的估计量。

设 $X_{1,n}\leqslant\cdots\leqslant X_{n,n}$ 是 X_1，…，X_n 的顺序统计量。取一个随机的门限

$$u = X_{n-\kappa,n},\ \kappa\to\infty,\ \kappa/n\to 0, \tag{11}$$

（11）式能保证 $u\to\omega(F)(n\to\infty)$，故当 n 较大时，$u=X_{n-\kappa,n}$ 是一个高门限。对尾指数 γ 采用 Moment 估计量[6]：

$$\hat{M}_n = m_{n1} + 1 - \frac{1}{2}\left(1 - \frac{m_{n1}^2}{m_{n2}}\right)^{-1},$$

其中

$$m_{nj} = \frac{1}{\kappa}\sum_{i=0}^{\kappa-1}(\lg X_{n-i,n} - \lg X_{n-\kappa,n})^j,\ j=1,2,\ \kappa/n\to 0. \tag{12}$$

有了 γ 的估计量 $\hat{\gamma}$，由（8）、（9）和（10）式即可得 μ 和 σ 的估计量 $\hat{\mu}$ 和 $\hat{\sigma}$。对（11）式中的门限，选取（10）式中的 $x=X_{n-i,n}$，$0\leqslant i\leqslant\kappa$，那么

$$\hat{F}_n^{[u]}(x) = \frac{\hat{F}_n(x) - F_n(u)}{1 - \hat{F}_n(u)} = \frac{\frac{n-i}{n} - \frac{n-\kappa}{n}}{1 - \frac{n-\kappa}{n}} = 1 - \frac{i}{\kappa}.$$

从而 σ_μ 的估计量为

$$\hat{\sigma}_\mu = \frac{\hat{\gamma}(x_{n-1,n} - X_{n-\kappa,n})}{(\frac{i}{\kappa})^{-\hat{\gamma}} - 1}. \tag{13}$$

这里 i 在1，…，$\kappa-1$ 之间适当选取一数即可。这样，由（8）、（9）式得 μ 和 σ 的估计量为

$$\hat{\mu} = X_{n-\kappa,n} - \frac{\hat{\sigma}_u}{\hat{\gamma}}\left[1 - \left(\frac{\kappa}{n}\right)^{\hat{\gamma}}\right], \tag{14}$$

$$\hat{\sigma} = \sigma_u \cdot \left(\frac{\kappa}{n}^{\hat{\gamma}}\right) \tag{15}$$

一个值得注意的问题是上述估计量中 κ 的选取问题。一种直观而粗糙的方法是[7]：画出估计量随 κ 变化的图（κ，$\hat{\gamma}(\kappa)$），$\kappa=1$，2，…$[n/4]$ 选取使图像较平稳的一段对应的 κ。

4 月、日降水量序列的直观描述

4.1 尾指数的估计

对南京、九江月降水序列和常州日降水、候降水序列进行排序后[8]，按上述方法先得到 $\hat{\gamma}$ 随 κ 的变化图（图2）。由图2可见，不论是月降水还是日降水量资料，κ 在很宽的数值范围内 $\hat{\gamma}$ 都有稳定的值。由此得到南京、九江、常州候和日降水4个序列的 Moment 估计量分别为0.1859、0.0796、0.27733、0.27042。其尾指数与正态分布的尾指数 $\gamma=0$ 有明显差别，这表明无论是月降水还是日、候降水序列均服从厚尾分布，

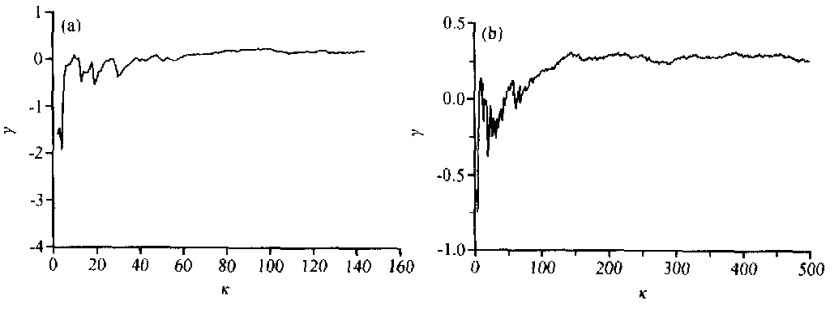

图 2　$\hat{\gamma}$ 随 κ 的变化

(*a*) 南京月降水；(*b*) 常州候降水

从而强降水的波动程度（波幅和频率）要比正态分布情形大。有了 $\hat{\gamma}$ 估计值，便可得到其他的估计量，进行降水序列的分布拟合。

4.2　用 GP 分布拟合强降水

超出累积经验分布 CDF（即气象要素值取等于或大于某数值的概率分布函数）和 GP 分布相比较（见图 3）。图中为了比较强降水部分的分布拟合，已分别对横坐标（降水量的大小）和纵坐标（GP 函数值或经验分布频数值）取了对数。可见，无论是日还是月降水量，GP 分布对强降水部分拟合的都很好。我们还分别做了我国西北地区宁夏、阿勒泰等地月降水量的 GP 分布，同样拟合的很好。这说明 GP 分布可能是可以拟合任何地区、任何时间尺度降水分布的一个函数。

有了强降水部分的概率分布函数，我们就能够知道并比较各地强降水发生的情况。人们常说的“百年一遇、50 年一遇”降水往往都是从现有的有限时间段的观测资料中比较而言，实际上这种说法并不科学。表 1 是根据 GP 分布函数值，分别例举了南京站超过 305、405、505、605 mm 强降水量各自对应的概率以及发生的年数。可见，其强降水发生的概率是远远大于人们想象的[9]。

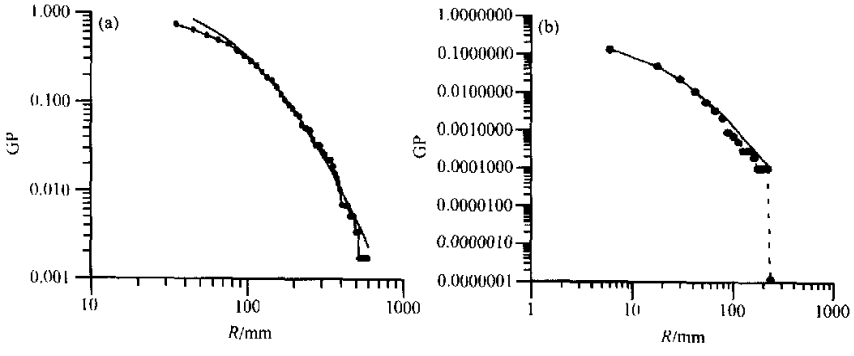

图 3　GP 分布拟合强降水

(a) 南京月降水；(b) 常州候降水

实线为 GP 超出累积分布，点线为经验超出累积分布

表 1 南京站强降水发生的概率及年数

月降水量/ mm	305	405	505	605
GP 概率	0.02489	0.00999	0.00458	0.00232
年数	3.3	8.3	18	36

对于日降水量来说，大多数时候降水为零，有降水的概率相对较小，大降水的概率就更小，且各地日降水情况差别很大，因此日降水的概率分布很难找到一个好的拟合函数。GP 分布很好地拟合了日降水的分布，对于常州站的日降水量来说，超过 102 mm 的大约是 65 年一遇，而超过 150 mm 的约 200 年一遇。

5 尾指数与样本经验分布尾部的比较

尾指数的大小反映了分布函数尾部的厚薄程度。尾指数越大，尾部越厚，强降水发生的概率就越大。我们看到，南京和九江两站的尾指数分别为 0.186 和 0.0796，九江的尾指数小于南京，说明南京的尾部较厚，即南京强降水发生的概率比九江大得多（见图 4）。为了能相互比较，图 4 中横坐标已经除以了各自的平均降水量。比较结果可见，南京的分布尾部确实比九江的厚，说明强降水的发生概率南京比九江大，两者分布都呈幂函数衰减，但快慢不同。尾指数越小，衰减越快。

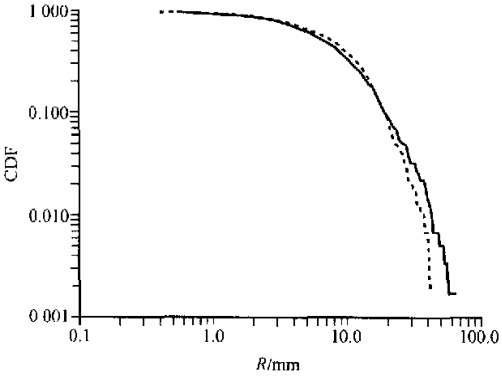

图 4 九江（虚线）和南京（实线）经验分布尾部的比较

6 结论

（1）对年以下尺度的降水量统计序列，其样本分布是非正态的，且是厚尾分布，强降水发生的值和频率都比高斯分布为大。

（2）可用 GP 分布拟合年以下尺度的降水量统计序列，对日、候降水量拟合效果很好，且方法简便；对月降水量序列在接近极值降水部分拟合也很好。

（3）GP 分布函数可能是拟合任何地区呈偏斜状态降水量普遍适用的一种方法。

（4）尾指数反映了降水分布尾部的厚薄程度，用尾指数的大小可以比较两地强降水发生概率的大小。

参 考 文 献

1　么枕生、丁裕国编著，气候统计，北京：气象出版社，1990，161～180.

2　刘式达、刘式适，地球物理中的混沌，沈阳：东北师范大学出版社，1999，66～68.

3　DuMouchel, W., and H. Estimating. The stable index α in order to measure tail thickness: A Critique, *Ann. Statist*, 1983, **11**, 1 019～1 031.

4　Balkema, A. A., and L. de Haan, Residual life time at Great Age, *Ann. Probab*, 1974, **2**, 792～804.

5　Pickands, J., Statistic inference using extreme value order statistics, *Ann. Statist*, 1975, **3**, 119～131.

6　Dekkers, A. L. M., J. H. J. Einmahl, and L. de Hann, A moment estimator for the index of an extreme value distribution, *Ann. Statist*, 1989, **17**, 1 833～1 855.

7　潘家柱、丁美春，GP 分布模型与股票收益率分析，北京大学学报（自然科学版），2000，**36**（3），295～306.

8　张美根、韩志伟、雷孝恩，数值与统计方法在广东核事故应急系统中的应用，气候与环境研究，2000，**5**（2），189～195.

9　李吉顺、王昂生，1998 年长江流域洪涝灾害分析，气候与环境研究，1998，**3**（1），390～396.

Statistical Analysis of Strong Rainfall

Zou Bo

(*Department of Air Transport Control, Civil Aviation Flight School, Guanghan* 618307)

Abstract　Climate of month and daily rainfall data do not distribute like Gause Distribution Function. Some properties of the generalized Pareto distribution are discussed. Then GP model is used to analyze the month and daily rainfall data. A quantitative indicator of strong rainfall is mentioned.

Key words: thick tail distribution; parameter estimate; tail index; strong rainfall; probability estimate