研究论文

广州短历时降水极值概率分布模型研究①

毛慧琴　杜尧东　宋丽莉

（广东省气候与农业气象中心，广州　510080）

提　　要

利用1959～2000年广州9个降雨历时（5、10、15、20、30、45、60、90、120分钟）暴雨资料，分别用皮尔逊Ⅲ型分布、对数正态分布、指数分布和耿贝尔-Ⅰ分布函数进行拟合，并按柯尔莫哥洛夫方法进行拟合优度检验。结果表明，广州短历时暴雨概率分布遵循皮尔逊Ⅲ型分布。

关键词： 短历时降水　概率分布　参数估计　统计检验

引　言

随着城镇化进程的加快和城市人口的急剧增加，城市短历时降水极值对城市的防洪排水能力、城市生态环境乃至居民生命财产的影响越来越重要。短历时降水极值作为随机变量具有一定的不确定性，但它仍表现出明显的统计规律[1]。虽然样本容量足够大时，可以用经验分布曲线近似地估计它的概率分布。但实际工作中，往往由于样本容量有限，难于求得可靠的分布曲线，特别是经验分布不能给出数学模型，不便于进行理论分析。为了将经验上升为理论，需要确定短历时降水的理论概率分布模型。关于城市短历时降水极值的理论概率分布，目前尚无公认的统一模型[2]，如美国采用耿贝尔分布和对数正态分布[3]，我国许多城市直接采用《室外排水设计规范》[4]建议的指数分布或皮尔逊-Ⅲ型分布模型，并没有用多种概率模型进行对比分析研究。广州地处低纬和珠江三角洲城市群的中心，海洋、陆地、大气之间的作用强烈，受低纬度热带天气系统和中高纬度天气系统的交替影响，其降水复杂多变，其短历时降水极值究竟符合何种分布，目前尚无人研究。本文利用广州42年（1959～2000年）的9个短历时（5、10、15、20、30、45、60、90、120分钟）降水资料，根据概率分析和数理统计原理，拟合其概率分布模型，以期为当地城市防洪排水排涝工程设计提供依据。

1　资　料

本文原始资料来源于广州市气象局降水自记曲线，资料的选取依据中华人民共和国国家标准《室外排水设计规范》GBJ 14－87，采用5、10、15、20、30、45、60、90、120分钟等9个降雨历时的资料，先逐年分别求算出9个降雨历时的最大值。然后不论年次按从大到小顺序分别对9个降雨历时进行排序，得到9个序列，并对各序列按从大到小选择年数4倍（42×4＝168）的资料，从而得到9个统计样本。

2　分布函数模型的选择及模型参数的估计方法

本文选择了皮尔逊Ⅲ型分布、指数分布、对数正态分布、指数分布和耿贝尔-Ⅰ分布。各分布模型的概率密度函数如下：

2.1　皮尔逊Ⅲ型分布

皮尔逊Ⅲ型分布的概率密度函数为：

$$f(x \mid a,b) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}(x-x_0)^{\alpha-1}e^{-\frac{x-x_0}{\beta}}$$

① 广东省“十五”科技攻关项目（编号：2002C20707）和广东省气象局重点项目（编号：200202）共同资助。

$$x \geqslant x_0 \tag{1}$$

其中，x_0 为随机变量 x 可能取到的最小值。α 为形状参数，α 愈小，则愈为正偏。β 为尺度参数。本文采用矩法估计参数 α、β、x_0[5]。

2.2 指数分布

指数分布的概率密度函数为：

$$f(x \mid \mu) = \frac{1}{\mu} e^{-\frac{x}{\mu}} \tag{2}$$

μ 为参数，同样采用矩法估计参数 μ。

2.3 对数正态分布

假定气候极值的对数服从正态分布，是气候极值统计中常用的分析极值的方法之一。设 X 为一极值变量，x 为它的取值。它的对数 $y = \ln x$ 服从正态分布，即

$$f(x) = \frac{1}{\sigma_y \sqrt{2\pi}} e^{\frac{-(y-m_y)^2}{2\sigma_y^2}}$$

$$-\infty < y < \infty \tag{3}$$

它的参数可由下式估计

$$\hat{m}_y = \frac{1}{n}\sum_{i=1}^{n} \ln x_i,$$

$$\hat{\sigma}_y^2 = \frac{1}{n}\sum_{i=1}^{n}(\ln x_i - m_y)^2 \tag{4}$$

2.4 Gumbel-Ⅰ型分布

Gumbel-Ⅰ型分布的概率密度函数为

$$f(x \mid a, b) = a e^{-a(x-b)^b - e^{-a(x-b)}} \tag{5}$$

本文采用 Gumbel 法估计参数 a，b[6]。

3 检 验

本文采用柯尔莫哥洛夫检验[6]进行拟合优度检验。以确定选择哪一种分布函数较好。另外，还计算了理论分布函数与经验分布函数的线性相关系数，作为选择分布函数的参考。

3.1 柯尔莫哥洛夫检验(K-G 检验)

设待选的分布函数为 $f(x)$，经验分布函数为 $F_n^*(x)$：

$$F_N^*(x) = \begin{cases} 0 & x \leqslant x_1^* \\ \dfrac{m}{n} & x_m^* < x \leqslant x_{m+1}^* \\ 1 & x > x_n^* \end{cases} \tag{6}$$

其中 $x_1^* \leqslant x_2^* \leqslant \cdots \leqslant x_m^* \cdots \leqslant x_n^*$ 为按由小到大顺序排列的样本。

柯尔莫哥洛夫检验不是用区间来考虑经验分布 $F_n^*(x)$ 与理论分布 $F(x)$ 的差异，而是逐点考虑它们的偏差，即考虑

$$D_n = \max\left| F_n^*(x) - F(x) \right| \tag{7}$$

柯尔莫哥洛夫检验是检验统计量：$r_x = D_n \cdot n^{\frac{1}{2}}$，柯尔莫哥洛夫已证明，对任意的 $r > 0$，都有

$$\lim_{n\to\infty} p(r_x < r) = Q(r)$$

$$= \sum_{-\infty}^{+\infty} (-1)^k e^{-2k^2 r^2} \tag{8}$$

若样本 n 很大，则认为 r 近似服从分布 $Q(r)$，这样根据信度 α，找到满足 $Q(r_\alpha) = 1 - \alpha$ 的临界值 r_α，提出原假设，认为样本服从分布；若 $r_x < r_\alpha$ 则通过检验，原假设正确；反之，若 $r_x \geqslant r_\alpha$ 则不能通过检验。取 $r = 0.05$，查表(柯尔莫哥洛夫检验中函数值表)，可得 $r_\alpha = 1.35$。

3.2 拟合相关系数

设待选的分布函数为 $F(x)$，将样本从小到大排列，重新记为 $x_k(k = 1,2,\cdots n)$，样本值 x_k 对应的理论分布函数值记为 $F(k)$，经验分布函数值 $F_n^*(k)$ 由式(6)求出，计算 $F(k)$ 和 $F_n^*(k)$ 的线性相关系数 R

$$R = \frac{\sum(F_n^*(k) \cdot F(k)) - \frac{1}{n}\sum F_n^*(k)\sum F(k)}{\sqrt{\left[F_n^{*2}(k) - \frac{1}{n}\left(\sum F_n^*(k)\right)^2\right] \cdot \left[\sum F^2(k) - \frac{1}{n}\left(\sum F(k)\right)^2\right]}} \tag{9}$$

R 表示分布函数值 $F(k)$ 与经验分布函数值 $F_n^*(k)$ 的线性相关程度，R 越大，说明这种分布函数拟合得越好。

4 计算结果

对9个统计样本(5、10、15、20、30、45、60、90、120分钟)分别用四种概率分布进行拟合，求其参数，并按上述方法进行拟合优度检验，检验结果见表1，表中以0接受原假设，通过检验；以1表示拒绝原假设，没有通过检验。

从表1可看出，9个降雨历时都不服从指数分布；服从耿贝尔分布的只有10、30分钟降水极值，服从对数正态分布的有10、15、20、30、45分钟降水极值，这5种降雨历时的经验分布和理论分布函数相关系数都超过0.98。从皮尔逊Ⅲ型分布拟合情况看，9个降水极值都通过了柯尔莫哥洛夫检验，降雨历时的经验分布和理论分布函数相关系数都超过0.99。可见广州短历时降水极值较好服从皮尔逊Ⅲ型分布，10～45分钟的短历时降水极值也可以用对数正态分布拟合，只是精度没皮尔逊Ⅲ型分布高。

表1　9历时降水极值拟合优度检验结果

历时 /分钟	D_n				K-G检验结果				相关系数			
	P-Ⅲ	对数正态	指数	Gumbel-Ⅰ	P-Ⅲ	对数正态	指数	Gumbel-Ⅰ	P-Ⅲ	对数正态	指数	Gumbel-Ⅰ
5	0.079	0.112	0.539	0.115	0	1	1	1	0.994	0.978	0.945	0.986
10	0.036	0.088	0.530	0.079	0	0	1	0	0.999	0.989	0.958	0.995
15	0.028	0.078	0.537	0.105	0	0	1	1	0.999	0.985	0.946	0.992
20	0.045	0.091	0.525	0.098	0	0	1	1	0.998	0.988	0.959	0.993
30	0.054	0.089	0.513	0.081	0	0	1	0	0.998	0.992	0.961	0.999
45	0.037	0.091	0.513	0.108	0	0	1	1	0.999	0.988	0.961	0.993
60	0.038	0.099	0.500	0.095	0	1	1	1	0.998	0.987	0.957	0.991
90	0.045	0.110	0.498	0.111	0	1	1	1	0.998	0.983	0.954	0.987
120	0.038	0.100	0.483	0.138	0	1	1	1	0.999	0.984	0.950	0.984

5　概率密度图的比较

为了比较皮尔逊Ⅲ型分布和对数正态分布的拟合差异，我们把估计得到的参数代入相应的概率密度函数，绘出皮尔逊Ⅲ型分布和对数正态分布的理论概率密度曲线，同时，绘出了实测值的频率密度直方图。图1～图5分别给出了皮尔逊Ⅲ型分布和对数正态分布理论概率密度曲线与实测10、15、20、30、45分钟降水极值频率密度直方图。从图中可看出两种分布理论概率密度曲线都较好地拟合出实测频率密度的趋势。但是对峰值的大小和位置的拟合皮尔逊Ⅲ型分布明显要比对数正态分布好。

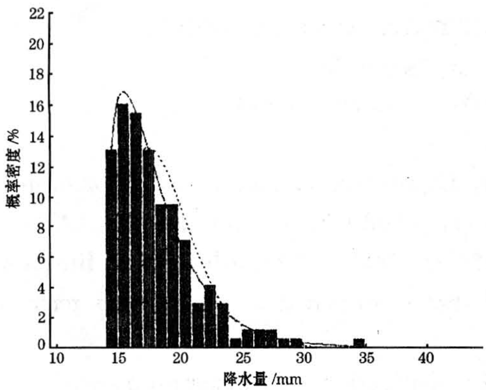

图1　10分钟降水极值频率密度直方图及概率密度曲线

实线为皮尔逊Ⅲ型分布，虚线为对数正态分布

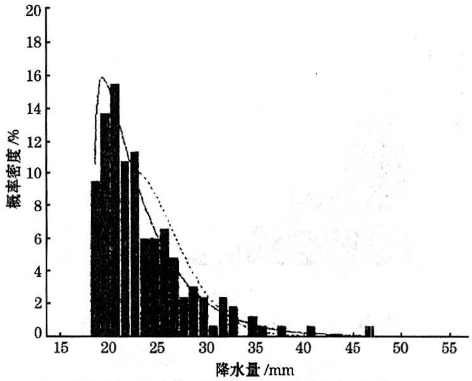

图2　15分钟降水极值频率密度直方图及概率密度曲线

实线为皮尔逊Ⅲ型分布，虚线为对数正态分布

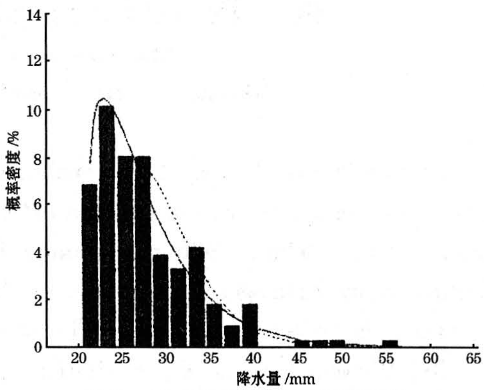

图3　15分钟降水极值频率密度直方图及概率密度曲线

实线为皮尔逊Ⅲ型分布，虚线为对数正态分布

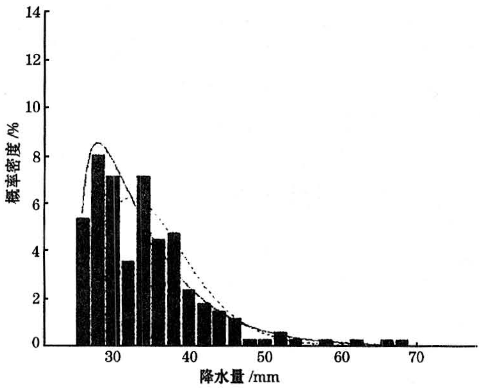

图4 20分钟降水极值频率密度直方图及概率密度曲线

实线为皮尔逊Ⅲ型分布,虚线为对数正态分布

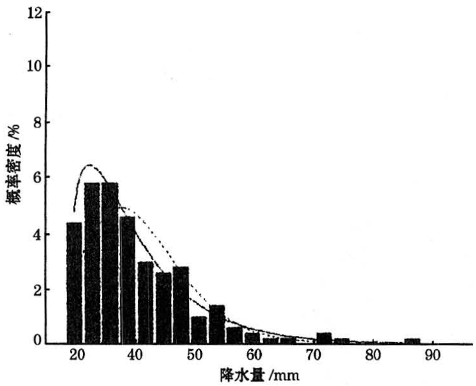

图5 45分钟降水极值频率密度直方图及概率密度曲线

实线为皮尔逊Ⅲ型分布,虚线为对数正态分布

6 结 论

由于城市短历时降水极值的理论概率分布目前尚无公认的统一模型,为了找到适合于广州的理论概率模型,本文采用国内外应用较多的四种概率模型对广州的短历时降水极值实况资料进行拟合对比分析研究,得出以下结论:

(1) 广州短历时降水极值不服从指数分布和耿贝尔分布;对历时为10～45分钟降雨极值可以用对数正态模型拟合,但较长历时(60～120分钟)不合适。

(2) 9个短历时降水极值都服从皮尔逊Ⅲ型分布,其经验分布和理论分布函数相关系数都超过0.99,拟合精度高。

参考文献

1 马玉开.气候统计原理与方法(第一版).北京:气象出版社,1993,123～129.

2 朱颖元.城市短历时暴雨强度公式参数.福州大学学报,2000,28(1):48

3 霍尔 M J.城市水文学,詹道江等译.南京:河海大学出版社,1989.

4 中华人民共和国标准.室外排水设计规范 GBJ14-87.北京:中国计划出版社,1990.

5 屠其璞,王俊德,丁裕国等.气象应用概率统计学,北京:气象出版社,1984:130～147.

6 高绍凤,陈万隆等.应用气候学,北京:气象出版社,2001:124～129.

Research on Probability Distribution Models of Short-Period Precipitation Extreme in Guangzhou

Mao Huiqin Du Yaodong Song Lili

(Guangdong Climate and Agrometeorology Center, Guangzhou 510080)

Abstract

Pearson-Ⅲ distribution, Lognormal distribution, Exponential distribution, and Gumbel-Ⅰ distribution are used to fit precipitation extreme of short period (5min, 10min, 15min, 20min, 30min, 45min, 60min, 90min, 120min) from 1959 to 2000 in Guangzhou. The fitness is checked up by Kolmogorov test. The results show that short-period precipitation extreme of Guangzhou is subject to the Pearson-Ⅲ distribution.

Key Words: short-period precipitation probability distribution parameter estimation