

关于合理解决水文频率计算方法的问题

中国科学院
水利电力部 水利学研究院 謝家澤

提要: 我国近期水文频率计算方法的研究工作在選擇方法, 經驗频率公式, 参数的誤差和利用我国水文資料檢驗各种频率綫型等方面有了一定的成果和实用的結論〔1〕, 但是下面两个問題还没有获得解决:

(1) 能否从机率理論証明水文频率属于何种分布律?

(2) 水文频率曲綫应当采用什么綫型? 如何确定参数? 本文試图解答以上两个問題。本文分析了水文系列的时序性質和区域性質, 把机率問題按先驗、极限和后驗三种基本性質对水文频率問題进行了分析, 利用开布屯推导偏态分布的方法分析了水文系列的分布性質, 并从而批判了有关水文频率肯定属于对数正态律, 耿贝尔极限律或二項式定理等等說法。認為属于后驗机率性質的水文频率, 不能从机率理論証明它属于何种分布律。

最后提出联合利用各站水文資料来選擇綫型和确定参数的方法, 并建議在兩端有限对数正态和皮尔遜Ⅲ型两种綫型中選擇一种作为統一采用的綫型, 对兩端有限曲綫提出了簡易可行的确定上下极限的方法, 对皮尔遜Ⅲ型曲綫認為應該把 C_s 作 C_v 的倚变参数。

一、从水文系列的性質看水文频率计算方法的基础

一定的天气形势和条件必然产生一定的降水情况, 从而在一定的流域自然地理条件下必然产生一定的洪水情况。从这个意义来看, 水文現象無論如何錯綜复杂, 却都有其严格的因果关系。同时, 任一水文現象都是很多成因因素作用的结果, 例如任何地点任何一次洪水都是海洋蒸发, 气团活动, 水气凝結, 地面截留和吸收, 地面及河槽集流等很多成因因素作用的结果。这許多成因因素作用的情况和大小都不是固定的。而是随机变量, 因而其作用的结果也仍是随机变量, 从这个意义上来说, 水文現象可以認為是一个随机变量, 亦即此时此地恰好产生这样大的洪水虽然是此次降雨和流域条件的必然结果, 却应当認為是偶然的。

但是, 水文現象与一般的随机变量还有不同之处。

(一) 水文系列的时序性質 水文現象与多数地球物理現象一样, 都是属于时序系列的范围。一般的时序系列的变化都包括下面四种組成部分: (一)、長时期稳定趋势的变化, (二)、循环性的变化; (三)、季节性的变化和(四)、偶然性的变化〔11〕。在水利工程的设计中, 由于水工建筑的使用年限較短, 長时期稳定趋势的变化可以不考虑。季节性的变化可以用按年按季統計的方法来避免它的影响。偶然性的变化是我們研究的对象, 但是水文的循环性变化, 象丰水年和枯水年的成組出現的情况, 假如也不加考虑, 就会造成很大的誤差。我国各大河流的丰水年和枯水年的連續出現是很普遍的。例如松

花江哈尔滨站的1957年平均流量(1899~1955)为1130秒公方, 1899~1927的29年枯水期平均流量为896秒公方(其中只有6年超过1957年平均值); 而1928~1955的28年丰水期的平均流量竟达1370秒公方(其中只有7年低于1957年平均值)。黄河、长江、淮河、辽河等也有类似的情况。哈尔滨的年径流的离差系数 $C_v = 0.42$, 据此如有28~29年的系列, 其均值的误差按误差理论计算约为 $\pm 8\%$, 而在哈尔滨的实例中却达到21%。29年的系列均值误差到达 $\pm 21\%$, 是十分可观的, 并且29年枯水年之后继之以28年的丰水年。这不能认为是单纯偶然现象, 而可能是水文多年循环性变化的反映[7]。由此可见: 水文现象是一种带有一些隐蔽周期性的时序系列, 也可以叫做随机平稳变化过程[2]。假如我们把水文现象当作单纯的独立随机事件处理, 严格讲来是不正确的。并且在统计计算中有时会产生严重的错误。虽然把水文现象当作时序系列来分析还只是今后的一个发展方向, 但是注意到水文系列的时序性质就一方面能帮助我们减少水文统计上的错误, 另一方面也可以清楚的看出: 既然现行水文统计方法的基础还有问题, 把统计结果(稀遇洪水或暴雨)绝对化的看法是完全没有根据的了。

(二)水文系列的区域性质 在降雨成因因素和条件完全相同的甲乙两个雨量站上的降水观测成果必然并不一样, 尤其短期暴雨量, 其差异更大。这是由于暴雨途径, 暴雨位置和暴雨分布的随机变化所造成的。因此, 由两个相邻站的短期资料所得的雨量系列和分布参数 ($\bar{X} \cdot C_v$ 等)的相互差异是完全可以理解的。但是, 既然两站的降雨成因因素和条件都完全相同, 那么两站雨量系列的分布性质和参数就应当一样, 上述可以理解的差异是由于系列短而产生的。系列愈长, 这一差异就会趋于很小。此外, 甲站的系列也完全有理由在乙站发生, 反之亦然。这就是说, 把甲、乙两站的资料联合使用不但可以解决了甲乙两站间差异的矛盾, 并且从暴雨途径, 暴雨位置和暴雨分布的情况来看也起了延长系列的作用。我们目前盛行的按单站进行统计的方法。从工程要求对固定地点计算稀遇洪水来看是必要的, 但从水文系列的区域性质来看, 等于把较大的统计场(例如暴雨成因因素和条件相同的区域)人为地缩小到一个小面积或一个点, 因而减小了样品容量和降低了统计成果的精度。当然象站年法那样机械的联合使用方法也有问题, 我们必须对于各站资料如何联合使用进行具体分析。但是, 水文系列的区域性质要求我们必须把各站资料加以联合使用。例如, 对点雨量频率计算来说。在暴雨条件相同的气候区内, 把区内各单站的统计参数 ($\bar{X} \cdot C_v$) 的加权平均值或算数平均值作为全区的标准值就可能是较为合理的方法。

二、先验机率问题

二项式、多项式、超几何级数以及普阿松等分布律都是先验机率的分布律, 假如事件 A 在每一次试验中的发生机率(P)为已知(所以叫做先验机率)。问在 n 次试验中, 事件 A 出现 m 次 ($n \geq m \geq 0$) 的机率如何? 假如只求事件 A 的出现和不出现的机率, 则可用二项分布律计算。如有多项事件 $A_1, A_2, \dots, A_n$ 其每一次试验的发生机率各为 $P_1, P_2, \dots, P_n$, 则各种事件同时出现的复合事件的机率可用多项式分布律计算。假如前

后事件的出現机率不是完全独立的,亦即后面事件出現机率將受前面事件出現情况的影响,則屬於超几何級数分布律的問題。在二項式分布律中,假如 P 值很小,并且 $P \cdot n = \lambda = \text{常数}$,則服从普阿松分布律。总之,先驗机率的問題大都可用机率理論进行精確計算,所以这些分布律也称作理論的分布律。

由此可見,已知頻率的洪水(暴雨)的发生机率的計算正是典型的先驗机率問題,但是这个300年前已被发现和証明的定理,却在实际問題中应用得很不够[12]。

(一)設計重現期、使用年限和保証率的关系:建筑物的使用年限,保証率和設計重現期三者之間的关系是水工設計中一个很重要的問題,但是一直沒有获得很好的闡明,并且在概念和名詞上都有所混淆[6]。

茲令 n = 建筑物使用年限

P = 建筑物設計洪水的頻率(%)

T_d = 建筑物設計洪水的重現期(簡称設計重現期);亦即

$$T_d = \frac{1}{P}$$

S = 建筑物在使用年限 n 年內的保証率

U = 建筑物在使用年限 n 年內的破坏率 $= 1 - S$

$Q_P\%$ = 建筑物的設計洪水流量

在建筑物使用年限 n 年內至少出現 1 次等于或大于設計洪水 $Q_P\%$ 的机率 U (即为建筑物的破坏率)可以按照机率相乘定理計算如下:

$$U = 1 - (1 - P)^n$$

下表列出不同設計洪水頻率(P)和不同使用年限(n)的建筑物破坏率 $U(\%)$ [6]。

表 1 建筑物各种設計頻率 $P=1/T_d$ 及各种使用年限 n 的破坏率 $U(\%)$

建筑物使用年限 n		5	10	20	30	50	100	200
設計洪水頻率 P (%)	0.01	0.05	0.10	0.20	0.30	0.50	1.00	1.82
	0.1	0.50	1.00	1.98	2.96	4.88	9.52	18.0
	0.5	2.48	4.89	9.54	14.0	22.2	39.4	63.4
	1	4.9	9.50	18.2	26.0	39.5	63.4	86.6
	2	6.6	18.3	33.2	45.5	63.6	86.7	98.2
	3	14.1	26.3	45.6	59.9	78.2	95.2	99.8
	5	22.7	40.1	64.1	78.5	92.3	99.4	100
	10	41.0	65.0	87.8	95.8	99.5	100	100

上述的問題也可以用另一种提法:假如已知建筑物使用年限 n 和破坏率 U , 求設計洪水重現期 T_d 应为建筑物使用年限 n 的若干倍,亦即求 $T_d = rn$ 中的 r 值。下表列出各种不同 n 和 U 值情况下的 r 值[12]

表 2 不同 n 值和 U 值情况下的 r 值 ($r = Td/n$)

建筑物破坏率 U (%)		63.2	50.0	40.0	33.3	30.0	25.0	20.0	10.0	5.0
建 年 限 物 使 用	2	1.27	1.71	2.22	2.73	3.06	3.73	4.74	9.75	19.76
	10	1.05	1.49	2.01	2.52	2.85	3.52	4.52	9.52	19.57
	∞	1.00	1.44	1.96	2.47	2.80	3.45	4.48	9.49	19.50

由上表可见,如建筑物使用年限 n 超过 10~15 年,则 r 值即可采用 $n = \infty$ 的情况。由上表可知,使用年限为 100 年的建筑物,如按 100 年洪水设计(即 $r = 1.0$)。则破坏率竟达 63%,亦即保证率不过 37%;如按 1,000 年洪水设计(即 $r = 10$),则破坏率约为 10%,而保证率为 90%。上述计算建筑物保证率和破坏率的方法在制订水工建筑物等级和标准时很有用。

由上可知,建筑物的设计洪水频率 P 乃指建筑物每年遭受 $Q \geq Q_p$ 的机率,亦即每年的破坏率为 P ,而每年的保证率应为 $1 - P$ 。建筑物使用年限 n 年内遭受 $Q \geq Q_p$ 的机率,亦即建筑物在 n 年内的破坏率为 $U = 1 - (1 - P)^n$,而保证率为 $S = 1 - U = (1 - P)^n$ 。通常把频率曲线称作保证率曲线,等于把含义几乎相反的频率和保证率混同起来,是很不妥当的。

(二)系列首项的经验频率的置信限:系列首尾两项的经验频率误差很大,其置信限也可以利用二项式分布律进行估算[1][14]。

已知 n 项系列的首项的每次出现机率为 P_1 ,则 n 次中至少出现一次的机率为

$$P\{P_1\} = 1 - (1 - P_1)^n$$

假定经验频率摆动在 P_a 与 P_b 之间,即 $P_a < P_1 < P_b$,根据机率分布的性质和

$$\begin{aligned} P &= P\{P_a < P_1 < P_b\} = [1 - (1 - P_b)^n] - [1 - (1 - P_a)^n] \\ &\equiv (1 - P_a)^n - (1 - P_b)^n \end{aligned}$$

令 P_a 、 P_b 及 P_1 有下列简单形式:

$$P_a = \frac{1}{Kn}, P_b = \frac{K}{n}, \text{ 及 } P_1 = \frac{1}{n}$$

其中 K 为置信限的常数,故

$$P = \left(1 - \frac{1}{Kn}\right)^n - \left(1 - \frac{K}{n}\right)^n$$

当 n 较大时,上式趋近于

$$P = e^{-\frac{1}{K}} - e^{-K}$$

上式与 n 无关。某些 K 与 P 的关系如下表[1]

如果取 P_1 落在 P_a 及 P_b 间的可能性为 50% 时,则 P 的可能出现范围为

$$\frac{P_1}{2.1} < P < 2.1P_1$$

表 3

置信限常数 K 及置信机率 $P(\%)$ 关系表

K	2.1	3.13	5.0	10.0	100	333.3
$P(\%)$	50	68.3	81.2	90.5	99.0	99.7

或 $0.48P_1 < P < 2.1P_1$

如取 P_1 落在 P_a 及 P_b 間的可能性为 99% 时, 則

$$0.01P_1 < P < 100P_1$$

由此可見, 利用先驗机率分布律可以很簡單地估計系列首尾兩項的經驗頻率的置信限。同时, 由此可見适綫时如机械地要求通过首尾項的經驗頻率点, 就是把偶然性当作必然性处理, 就难免造成很大的誤差。

此外, 先驗机率分布律也可以很方便地用来对小集水面积的調查洪水的頻率和不同取样方法(年最大值法及超定量法)的頻率进行換算[6][1]。总之, 先驗机率的分布律不能用以直接解决后驗机率的問題。但是在經驗机率可以認為是已知数之后。那么先驗机率分布律就是頻率分析和驗算中的一个很有力的武器。

(三)水文頻率是后驗机率問題: 有些水文学者說二項式分布律可以用来描述水文系列的分布[5]。又有人認為稀遇洪水的分布服从普阿松分布律[9]。前面已經提到, 二項式、多項式、超几何級数以及普阿松等分布律都是先驗机率問題, 亦即事件在每一次試驗中的出現机率(P)是先驗的(即已知的)。問題在于求 n 次試驗中事件出現 m 次的机率如何。洪水頻率計算問題中洪水的发生机率(P)是事先不知道的(亦即是后驗的或經驗的), 正是需要推算的对象, 所以是屬於后驗机率問題的范疇。那些把先驗机率分布律来解答后驗机率問題的提法很容易引起誤会和錯覺, 好象水文頻率可以利用先驗机率分布律从理論上加以計算似的。認為水文系列分布可以采用二項式分布律的說法的不妥当还可以

从下面的情况看出: 二項式分布律的离差系数 $C_v = \sqrt{\frac{q}{nP}}$, 偏差系数 $C_s = \sqrt{\frac{q-P}{nPq}}$ [3],

亦即 C_v 和 C_s 都随着試驗次数的增長而減小, 假如洪水頻率也服从二項式分布律。亦即洪水頻率的 C_v 和 C_s 都随着观测年数的增長而減小, 那么 C_v 和 C_s 都將失去統計参数的意义。也就沒有計算的必要了。許多書上把二項式偏态分布与皮尔遜Ⅲ型等同起来[10], 这就无异于把先驗机率与后驗机率問題等同起来。很不妥当。

三、极限机率問題

在机率問題中, 当試驗系列的項数很大时, 并在很一般的限制条件下, 根据机率論的极限理論推論, 未知的随机变数的分布函数就与某些极限分布律相趋近。这种把問題由有限轉为极限的办法对于后驗机率問題(例如水文統計)具有特別的意义。有很長一个时期, 大家把尋求使独立随机变数和的分布函数趋于正态率的最一般条件看作是机率論的中心問題。現在已經判明: 极限律远不尽于正态律。于是問題在于: 在諸項上应加什

么条件, 乃能使和的分布函数收敛于某极限率[2]。可惜在解决具体问题时, 我们很难判明那些最一般的条件和假定是否符合, 因而限制了极限律的预测未知分布函数的作用[11]。

我们现在只就与水文统计关系很密切的中心极限定理和极限值分布律来加以讨论。

(一)中心极限定理的应用: 中心极限定理证明: 在最一般的条件下。独立随机变数和 Z_n 当试验项数 n 很大时趋近于正态分布。所谓最一般的条件是: 和的方差趋向无限而各变数的方差与和的方差的比高趋近于零。其中一个特例是: 当各变数具有同样分布时。则只要满足一个条件: 即这个共同的分布只要具有均值和方差就行了[11]。

(1)开布屯(*kaptyen*)对偏态分布的推导: 开布屯曾经十分强调: 他不满足于去描述观测资料的分布函数。而而努力去寻求形成分布函数的理由。所以他对形成偏态分布的理由作了如下的推导[11]:

如令具有初值 x_0 的构件, 经过很多道工序。将它从其初值 x_0 依次变为预期的数值 x_1, x_2 等。在第1次工序中的变量 $x_i - x_{i-1}$ 可以假设其与工序成因作用的强度 g_i 和构件本身的数值 x_{i-1} 有下列关系:

$$x_i - x_{i-1} = g_i h(x_{i-1})$$

其中: g_i = 成因作用的强度, 及 $h(x)$ = 反应函数。

n 次工序中构件数值的变化可用下列诸式表示:

$$x_1 = x_0 + g_1 h(x_0)$$

$$x_2 = x_1 + g_2 h(x_1)$$

.....

$$x_n = x_{n-1} + g_n h(x_{n-1})$$

末了把变量 $x_n - x_0$ 作为 $g_1, g_2, \dots, g_n$ 的函数

$$g_i = \frac{x_i - x_{i-1}}{h(x_{i-1})}$$

于是

$$\sum_{i=1}^n g_i = \sum_{i=1}^n \frac{x_i - x_{i-1}}{h(x_{i-1})}$$

设成因作用的工序次数很多, 每一工序的变量可以认为相对的, 则

$$\sum_{i=1}^n g_i = \sum_{i=1}^n \frac{x_i - x_{i-1}}{h(x_{i-1})} \simeq \int_{x_0}^x \frac{dx}{h(x)}$$

令

$$Z_n = \sum_{i=1}^n g_i, \text{ 及 } g(x) = \int_{x_0}^x \frac{dx}{h(x)} \quad \text{则}$$

$$Z_n = \sum_{i=1}^n g_i = g(x_n)$$

照中心极限定理, 在前述的最一般条件下, 当 $n \rightarrow \infty$ 时, 独随机变数和 $Z_n = \sum_{i=1}^n g_i$ 将趋近于正态分布。于是上式说明: 不是构件数值 x 作正态分布, 而是数值的函数 $g(x)$ 作正态分布。

如 $h(x) = 1$, 则 $g(x) = x - x_0$, 于是 x 作正态分布。也就是说, 如果反应函数是常数, 亦即数值变量与成因作用开始时数值的大小无关, 则数值大小本身的分布是正态的。

如反应函数 $h(x)=x$, 則

$$g(x) = \int_{x_0}^x \frac{dx}{x} = \ln x - \ln x_0$$

于是 $\ln x$ 作正态分布。亦即, 成因作用所产生的数值变量如与成因作用强度和数值本身的大小都成正比, 則数值的对数的分布是正态的。

(2) 周文德对水文系列的对数正态分布的理論解釋: 周文德認為[13]: 由于水文特征值 x 是許多气象的, 地理的成因因素的綜合作用結果, 用公式表示:

$$x = x_1 x_2 \cdots x_r$$

其中 $x_1, x_2, \cdots, x_r$ 即代表各成因因素的作用, 并且成因因素的数目 r 很大。將上述兩端取自然对数, 即得

$$\ln x = \sum_{r=1}^r \ln x_r$$

于是根据中心极限定理。假如 r 很大, 則 $\ln x$ 作正态分布。这就是水文系列之所以是对正数正态分布的理論解釋。

必須指出, 虽然認為水文特征是許多气象的, 地理的成因因素的綜合作用的結果是完全正确的, 但是把这个錯綜复杂的过程写成为

$x = x_1, x_2, \cdots, x_r$ 的簡單数字关系是缺乏根据的。一則 $x_1, x_2, \cdots, x_r$ 的含义和指标都不明; 其次从前节开布屯的偏态分布推导中已很清楚: 成因因素的作用强度与反应函数是必須加以分別考虑, 否則就必然得出这样武断的錯誤結論, 任何綜合作用結果的現象不是正态分布便是对数正态分布。

(3) 利用开布屯推导偏态分布的方法分析水文系列的分布性質, 可以把空气含水量比作开布屯推导中的構件的数值, 空气含水量的初值 x_0 可以假定其为湿润空气源地的正常水气量, 例如我国东南季风在源地夏威夷港面的低层气温約为 20°C , 水汽含量約每公斤空气12克, 空气的上升、下降、膨胀、收缩、受热、去热、蒸发、凝結等等就是直接或間接地改变空气含水量的各种成因作用。从降雨是空气含水量剧烈改变的结果来看是与中心极限定理的条件之一, 即各次变量較之总变量相对的小。不相符合的。

显然, 空气含水量变化的大小不仅与成因作用的强度成比例, 并且也与空气含水量大小成比例, 亦即开布屯推导公式中的 $h(x)$ 并非常数所以降雨系列的分布显然不是正态的。其次, 各种成因作用都是直接地或間接地通过改变空气温度来改变空气含水量, 而气温和饱和含水量不是正比关系, 而是指数关系; 亦即开布屯推导公式中的 $h(x)$ 并不恰好等于 x , 所以降雨系列并不恰好是对数正态分布, 虽然可能与对数正态分布比較相近。

决定水文系列分布的偏态程度到底是那些因素, 在以往是不清楚的, 从开布屯的偏态分布的推导来看, 如果 $h(x)=1$, 則系正态分布, 亦即 $C_s=0$; 如若 $h(x)=x$, 則系对数正态分布, 亦即 $C_s=3C_v+C_v^3[8][13]$ 。由此看来, 反应函数 $h(x)$ 是偏差系数的决定因素。这一推論对于 C_s 的研究和确定很有用处, 值得进一步研究。

由以上的討論可知。中心极限定理在水文統計方面的应用, 可以說明一些問題, 但

是却还不能肯定的、直接地解答水文频率服从什么分布律的问题。

(二)极限值分布律的应用: 如有容量为 n 的 N 个样本, 这些样本的极限值(即最大值和最小值)的分布, 当 N 和 n 都很大时, 将向一定的极限分布函数收敛。研究指出, 最大值的分布函数只可能有下列三种类型[4][14][15]。

第一类型(即耿贝尔定律)。原始分布是指数型函数(如正态分布、对数正态分布等属之), 极限值分布有 $e^{-e^{-x}}$ 的形式; 变数的范围是从 $-\infty$ 到 $+\infty$ 。

第二类型(即伏瑞谢定律), 原始分布为 *Cauchy* 型分布, 极限值分布具有 $e^{-x^{-k}}$ (k 为正值)的形式, 变数的范围由 0 到 $+\infty$ 。

第三类型的极限值分布具有 $e^{-(x)^k}$ (k 为正值)的形式, 变数范围由 $-\infty$ 至 0。

耿贝尔认为一般资料适合于第一类型, 他对于第一类型的极限值分布律作了详尽的研究, 所以第一类型也称作耿贝尔定律。据耿贝尔研究。所有属于指数型的原始分布函数, 它的极限值的分布都是第一类型,

第一类型的频率曲线的方程式为

$$P = 1 - e^{-e^{-\alpha(x-u)}}$$

或使

$$y = \alpha(x-u)$$

则

$$P = 1 - e^{-e^{-y}}$$

其中 α 为常数, u 为众值。

上述频率曲线的偏差系数($C_s=1.139$)及峰度系数($C_e=2.4$)都是常数, 所以只要求二阶矩便能确定参数 α 及 u 。耿贝尔特别推荐用以计算洪水频率[14]。

耿贝尔频率曲线可以认为是一个理论分布律, 但是它是以下列三个条件为基础:

(1) 原始分布函数属于指数型

(2) 变数是独立随机的

(3) N 和 n 都很大

耿贝尔对于洪水的原始分布是否属于指数型的函数未加讨论, 他认为把逐日平均流量作为原始变数, 每年365天各日流量虽然不是相互独立的, 但其中的100或200天的各日流量却可认为是独立的。亦即年最大洪水是100至200个变数中的最大值, 所以 n 值已经很大了。他本人没有提出 N 和 n 的最小限度。有人认为不能小于20, 最好不小于30至50[12]。一年中能有20次以上的独立并且同性质的洪水也不是常有的事。加之, 洪水的原始分布属于指数型函数也没有根据, 并且变数范围由 $-\infty$ 至 $+\infty$ 也与洪水的物理现象不相符合。由此可见。耿贝尔极限分布律虽然是理论分布律, 但是这个理论分布律的主要条件在洪水现象中既不确立, 那么也就失去它的理论基础了。

如把伏瑞谢分布律和耿贝尔分布律在洪水频率计算中应用实例进行对比, 上述的结论就更加明显[15]图1. 和图1. 说明 Black canon Colorado 河的洪水资料与耿贝尔曲线偏离很大, 而与伏瑞谢曲线却符合很好。Bernier 在该文末尾还举这样一个例子: 即 Durance 河在 Archidiaacre 站的调查春汛洪水 $Q_{max}=2,000$ 秒公方, 如按伏瑞谢曲线其重现期约为2,000年左右尚属可能; 如按耿贝尔曲线, 则重现期将为 10^8 年显然不合

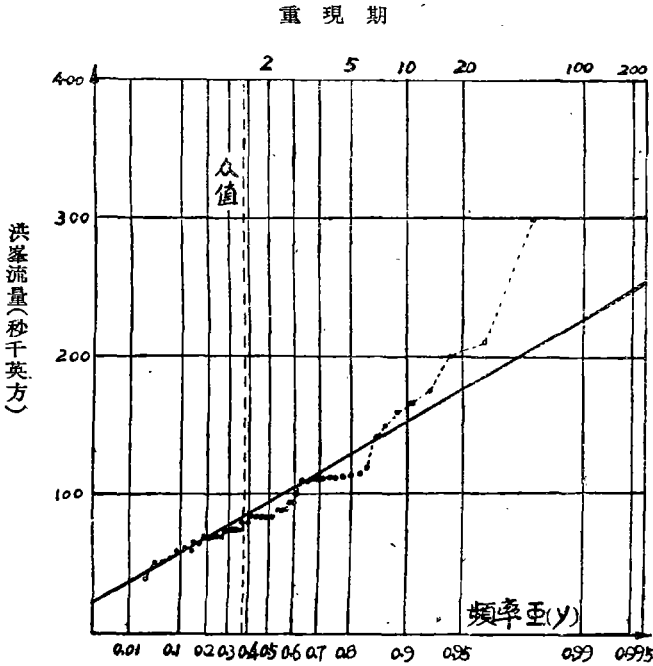


图 1. 用耿貝尔定律适线

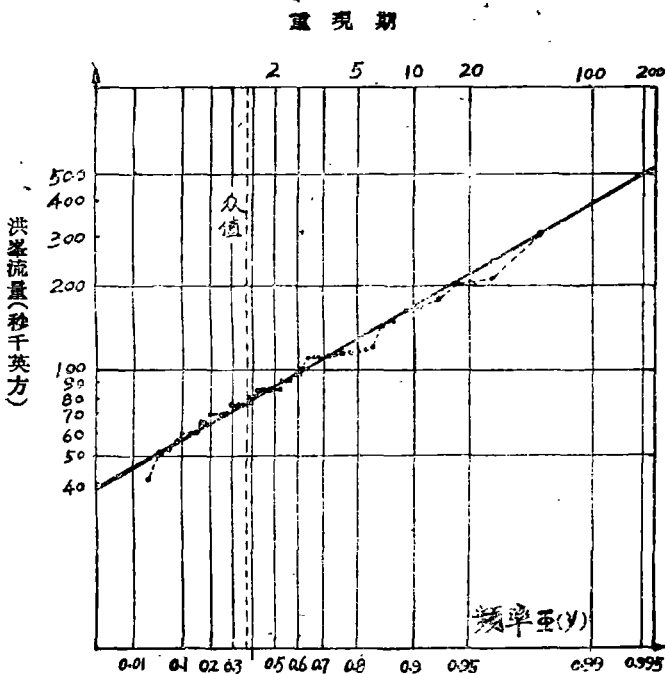


图 1. 用伏瑞謝定律适线

理。此外，伏瑞謝分布律的变数范围是 0 至 $+\infty$ ，也与洪水现象的物理性质较为接近。我們虽然不能由此得出伏瑞謝曲线适合洪水频率计算的结论，但是却可以明显地看出，洪水频率肯定属于耿貝尔分布律的说法无论在理论上和应用上都是根据不足的。

四、后驗机率問題：

一般情况下不但随机变数发生机率是未知的，并且随机变数的成因和分布的一般性质也不清楚。这就是所谓后驗机率的问题，它们不能利用先驗机率分布律或极限定理来解决。后驗机率的分布只能通过分析它的随机样本（即资料系列）的经验频率，并利用一定数学形式来加以描述和决定。从这个意义来看，后驗机率的分布律是经验性质的。描述后驗机率的分布的数学方法很多，但是按其数学处理方法的性质可以分为下列三类：

（一）概括机率分布某些

通性的方法：这个类型中最著名并且最常用的是皮尔遜曲线族。皮尔遜发现。一般机率密度曲线都有两个通性，即众值处切线的斜率为零和曲线两端或一端与横轴相切。他利用下列微分方程

式概括了这两个通性:

$$\frac{dy}{dx} = \frac{(x+d)y}{b_0 + b_1x + b_2x^2}$$

解: $b_0 + b_1x + b_2x^2 = 0$ 中的 x 的两根有各种情况, 例如等根, 实根, 或虚根以及根的符号相同与否等等。根据这些根的不同组合, 皮尔逊族可以分为13型。

皮尔逊曲线的最大优点是概括性强, 从分布通性的概括上看。它能约略符合二项式分布、普阿松分布、超几何分布及正态分布等。但是却不能说它们是同等的, 因为皮尔逊曲线是描述后验机率的经验曲线, 而二项式分布等是理论推导的先验机率分布律。此外皮尔逊族完全根据 $b_0 + b_1x + b_2x^2 = 0$ 的解法来分类。物理意义不明, 这也是它的一个缺陷[16]。

另一个例子是古德力区曲线。即 $P = 1 - e^{-k \frac{(x-a)^d}{(b-a)^d}}$ 。通常都把它叫做纯经验性的频率曲线。但是它利用上列的数学形式概括了分布的三个特性: 即随机变数的最大最小极限为 b 和 a , 而分布函数为指数函数, 从数学处理方法来看, 古德力区曲线与皮尔逊族具有相似的性质。

(二)利用级数和多项式修正基本分布律的方法: 格拉姆——夏里埃以正态分布为基础。利用台劳级数和爱尔密特多项式展开为无穷级数。可以描述任何偏态程度的分布。勃罗夫柯维奇曲线则以 $C_s = 2C_v$ 的皮尔逊Ⅲ型为基础, 利用拉盖尔多项式为扰动函数去修正。上述两种曲线虽然所用基本分布律不同。但从数学处理方法的性质看, 属于同一类型。它们的优点也是概括性强, 但是计算复杂, 并且可能出现负的机率, 因此在水文统计中应用很少。

(三)函数转换的方法、论证: 随机变数 X 的分布函数 $F(x)$ 可以通过适当的函数转换关系 $x = x(z)$ 把它转变为任何指定的已知的分布函数 $F_z(z)$, 只要 $x = x(z)$ 是连续函数, 就可通过下式转换:

$$F(x)dx = F_z(z)dz$$

水文统计中利用函数转换的分布族很多, 例如海森(Hazen)格拉斯伯格(Grassberger)史莱特(Slade)等频率曲线都是假定水文资料的对数值服从正态律, 所以它们都属于对数正态律。克里茨基—閔凯里假定 $Z = ax^b$ 服从 $C_s = 2C_v$ 的皮尔逊Ⅲ型分布律, 由此得出弹性很大的所谓克里茨基—閔凯里曲线, 由于正态律已经研究得很充分。所以通过函数转换把偏态分布函数变为正态分布之后就大大地便利了进一步的分析检验。因此对数正态律虽然也是经验机率的性质, 但是却有其方法上的优点。

把机率问题分为先验机率, 极限机率和后验机率三种类型虽然是有意义的和必要的。但是必须注意同一个分布律, 应用在不同的问题中便具有不同的性质, 正态律和对数正态律便是很好的例子。当 $P = \frac{1}{2}$ 或者 n 值很大时二项式分布律都表现为正态律。由此可见, 同是一个正态律, 当 P 为已知并且等于 $\frac{1}{2}$ 时, 便属于先验机率分布率。假如事件发

生机率 P 的数值虽然并不确知(即不是先验机率),但是确知 P 值是稳定的,则当 n 值很大时,也表现为正态律,这种情况下的正态律却不是先验机率而是极限机率问题。又如在随机变数的形成过程中,假如我们确知变数在每次成因作用中的变量是与作用强度和变数的数量都成正比例,并且成因作用的程序很多而且每次的变量都较总变量小得多,那么根据中心极限定理的推断,随机变数的分布服从对数正态律。在此,对数正态律属于极限机率分布律的范围。但如我们对于随机变数的成因和分布的一般性质都不清楚,只是知道资料的经验频率大致符合对数正态律。在此对数正态律就是后验机率的性质。

五、水文频率曲线的綫型选择和参数确定问题

(一)问题在于破除迷信和解放思想:合理解决水文频率计算的綫型选择和参数确定问题的首要关键在于破除迷信和解放思想。有人说水文频率应当服从某种理论分布律[9],有人说已经证明水文频率属于对数正态律[13]、耿贝尔律[14],还有不少人正在钻研水文频率属于何种分布律的答案[15],有更多的人正在渴待上述理论探讨的完成,找出理想的频率曲线,可以彻底解决问题。本文前几节的分析已经说明:水文频率属于后验机率问题,并且水文现象的成因过程很复杂,不可能从机率理论直接证明水文频率属于何种分布律。因此,我们必须首先从迷信和等待理论频率曲线的束缚中解放出来,以便从实际需要出发,利用资料综合分析检验的方法,求得合理解决的途径。

其次,必须从根据单站短期资料选择綫型的错误作法中解放出来。本文从水文现象的区域性质已经说明,根据单站短期资料选择綫型和确定参数是不合理的,是产生矛盾和混乱的主要根源。从图2的官厅洪水频率曲线可以看出:三种綫型(皮尔逊Ⅲ型,克-閔曲线和对数正态曲线)对于实测资料的适合性相差很少,但在外延部分则三者差别很大,可见根据单站短期资料既不能据以判明曲线内插部分的适合性,又不能据以判明曲线外延部分的合理性。所以让设计人根据单站短期资料去选择綫型带有很大的任意性,将会造成分歧和混乱。合理的途径是:确定一些选择綫型的原則,分析各种綫型的性质和异同,然后利用各站的和各种的水文资料加以综合分析和校验之后,选择一种比较合适的綫型,规定下来,统一采用,只要能起到统一外延的方法和使得外延成果大致合理的两个作用,便应当认为基本上满足了水文频率计算的实际要求。

第三,要从只凭单站短期资料确定参数的不合理的作法中解放出来,图3代表上海最大一日暴雨频率曲线按年期分段计算的结果[1]。75年(1877~1951)的均值 $\bar{x}=87.7$ 公厘, $Cv=0.37$; 15年小雨期(1892~1906)的均值 $\bar{x}=76.6$ 公厘, $Cv=0.30$; 15年多雨期(1877~1891)的均值 $\bar{x}=90$ 公厘, $Cv=0.43$ 。由此可见,即使有15年的记录来计算 $\bar{x}$ 和 Cv ,误差仍然可能很大。单站长期资料的分期计算已经说明:不能单凭单站短期资料来确定参数;同时也不能把个别单站短期资料对水文频率曲线的适合程度作为曲线取舍的标准,例如图3的[1]綫虽对15年小雨期和15年多雨期的点据都脱离很远,但却能更好地代表上海的暴雨频率的规律性。另一个例子是河北省太行山前沿京汉铁路一带地区的暴雨频率[1]。各单站最大一日暴雨量的均值变化在73至93公厘之间, Cv 变

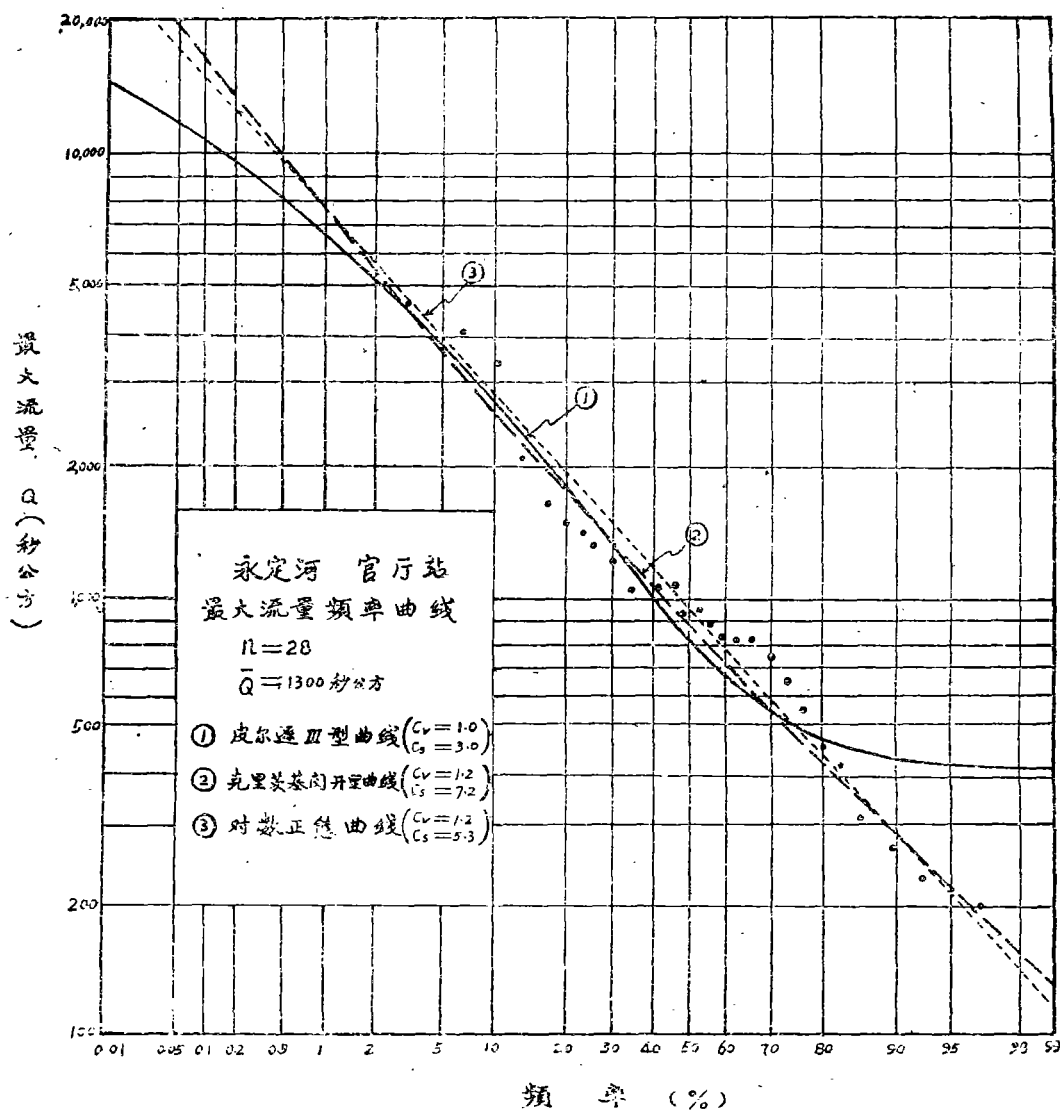


图 2

化在0.33至1.24之間；全区平均情况（按站年法計算）均值为75公厘， $C_v=0.68$ 。显然利用站年法（尽管站年法还有缺点）計算的全区平均情况比較單站計算的結果可靠，而全区平均情况的频率曲线（ $\bar{x}=75$ 公厘， $C_v=0.68$ ）必然与很多單站的短期資料偏离很远，这也說明單站短期資料不能作为确定参数和曲线取舍的主要依据。同时，如以單站短期資料作为主要依据，势必矛盾百出。例如，暴雨成因条件相同的区内各站的同频率暴雨如按各單站資料单独計算，就可能相差悬殊，这显然很不合理。又如，无论那个站只要增加一次特大暴雨，就必然要大大地修改原来的频率曲线（如按站年法統計就影响

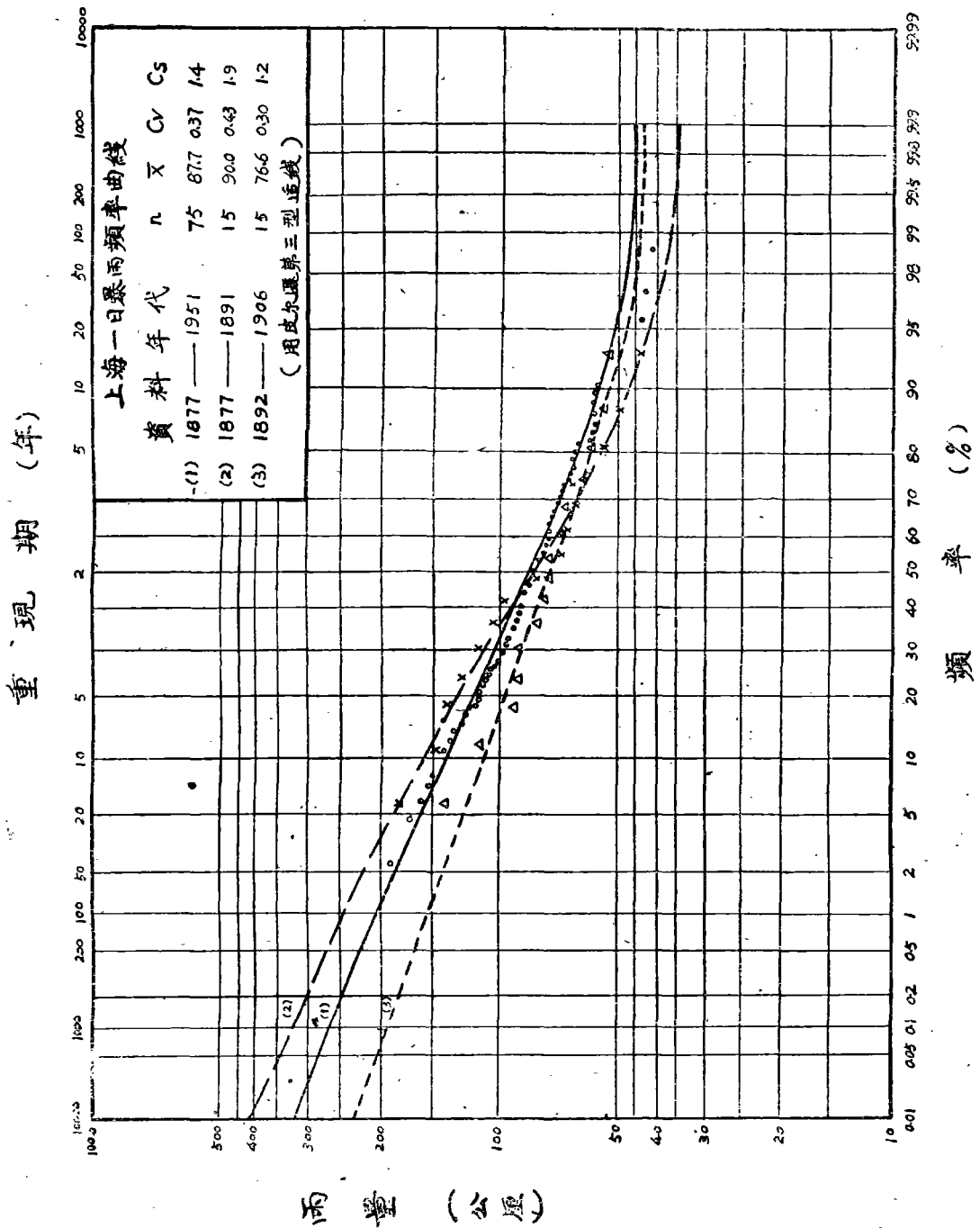


图 3

很小)。这样的频率曲线如此多变,还有什么规律性和使用价值?至于硬性要求频率曲线必须通过单站短期资料的首尾两端的经验频率点并据以选定 C_s 值,则是更加错误的方法。总之,迷信单站短期资料的实质就是把偶然性当作必然性处理,不从这种错误的方法中解放出来,就很难找到合理解决水文频率计算方法的途径。

(二)选择綫型的原則:

如上所述,綫型是統一外延的工具,要使外延的成果大致合理,并且計算簡便,这也就是选择綫型的基本原則,这些原則的进一步具体化,就是統一綫型。統一綫型的曲线形式必須大致符合水文現象的分布性質,亦即曲线能与实际資料的分布基本符合;其次曲线的形式应当大致符合水文現象的物理性質,例如曲线应有最大的和最小的极限值,不能出現負值等。統一的綫型一方面必須有較大的偏态范围,亦即 C_s 要有較大的变化范围;例如耿貝尔曲线的 $C_s=1.14$ 便不能符合水文現象偏态程度变化很大的要求,因而不宜采用。另方面确定綫型所需要的独立統計参数不能超过2个,亦即只能把均值 $\bar{x}$ 和变差系数 C_v ,作为独立参数,而偏差系数 C_s 只能作为倚变参数,因为水文系列都很短,二阶矩以上的計算誤差太大,沒有实用的意义。

(三)建議在兩端有限对数正态曲线和皮尔遜Ⅲ型曲线中选择一种作为統一采用的綫型。

(1)兩端有限对数正态曲线:如令 $K = \frac{x}{\bar{x}}$ 代表变数的比值, K' 代表上限比值, K'' 代表下限比值,則所謂兩端有限对数正态曲线就是 $y = \ln \frac{K' - K}{K - K''}$ 的分布是正态的,亦即 y 值的频率分布在对数机率格紙上成直綫。許多学者研究过并且推荐过兩端有限对数正态曲线,〔16,17,18〕,但是一直沒有解决如何确定上下极限的問題,所以在实际应用中沒有采用这种曲线。这种曲线有很多优点:兩端有限基本符合水文現象的物理性質,假如有了确定上下限的方法,則只剩下2个独立統計参数 $\bar{x}$ 及 C_v ;偏态范围很大;对数正态型曲线一般都能符合水文現象的分布性質。由比可見,兩端有限对数正态曲线基本上符合选择綫型的全部原則,而問題在于如何确定上下限。作者建議:选择一定的极限重现期(例如 10^4 , 10^5 或 10^6 年),从普遍对数正态曲线上讀取与选定的极限重现期相应的 K 值作为极限值。因为一般水工建筑物的使用年限約为100至200年,重现期大于 10^4 年的洪水,其出現机率很小,可以認為是几乎不可能事件,因而可以作为极限值。至于极限重现期究竟选用 10^4 年, 10^5 年或 10^6 年,要用資料比較分析后选定。由此可見,上述确定上下限的方法不但簡便、明确,并且基本符合机率极限的概念。采用上述确定上下极限值的方法后,兩端有限对数正态曲线就具备了作为統一綫型选择对象的全部条件。

(2)皮尔遜Ⅲ型曲线:在水文频率計算中应用已久,为大家所熟悉,具有能够适合一般水文資料的分布性質,偏态范围很大,适綫彈性很大等优点。但是也有一些缺点:一端无限,不符合水文現象的物理性質。有三个独立参数,确定 C_s 很困难;当 C_s 值很大时,(例如 $C_s > 2$ 时),曲线右端平翹,与資料显著脱离。上述缺点中除兩端无限外,

其余都可设法补救。例如使 C_s 与 C_v 建立一定的函数关系,象苏联规定 $C_s=2C_v$ (雪成洪水)和 $C_s=4C_v$ (雨成洪水)是解决确定 C_s 的困难的一种方法。假如 C_s 和 C_v 的关数选择适当也能克服曲线右端平翘的缺点。因此,皮尔逊Ⅲ型也应当是綫型选择对象之一。

(3)其他的綫型:有些与对数正态曲线相接近,有些与皮尔逊Ⅲ型曲线相接近,就可以不必再加考虑。还有一些綫型很复杂,具有3个以上的独立参数,那就根本没有考虑的必要,克-閔曲线在我国试用的结果,对北方河流的洪水频率,适合性似较皮尔逊Ⅲ型为好,但是它也具有一些另外的缺点[1,4],所以不宜作为主要选择对象来考虑。

(四)选择和检验綫型的方法:

(1)检验有关测资料部分(即频率曲线中部)的适合情况:假如各站实测资料在对数机率格纸上绝大多数都呈弯曲的形状,则对数正态曲线便不适合,而皮尔逊Ⅲ型曲线可能较好;反之,如绝大多数都呈直线,则对数正态曲线较好。除单站资料分别点绘比较外,还须要作联合比较。雨量资料可以把同一雨区内的资料按各单站计算的频率都点在一张机率格纸上,这样可以看出资料分布的总趋势,不致于受个别的误差和偶然性的影响。洪水,径流等资料可以把 C_v 值极为相近的资料组(例如 $C_v=0.8-0.9$)画在同一张机率格纸上(当然要用 K 值,而不是用真值 x 填绘)进行比较。

(2)检验曲线两端(即外延部分)的合理性:这项检查,不能按单站进行,必须把各站资料联合起来检验,亦即雨量按地区联合起来,洪水和径流按 C_v 值分组联合起来,要检查已发生过的特大值和特小值在频率曲线上的重现期是否大致合理,例如重现期超过 10^4 或 10^5 年就值得怀疑。从频率曲线上读取的稀遇洪水或暴雨,例如千年的,万年的数值是否合理和可能。同一地区内的稀遇洪水径流深假如等于或大于同频率的点雨深,就不合理。北方的稀遇年雨量和年径流假如等于或大于同频率的南方的数值也是不合理的。

以上两种检验可以对于两端有限对数正态曲线连同不同的极限重现期和皮尔逊Ⅲ型曲线连同不同的 C_s/C_v 关系的适合性和合理性进行比较,从而选择一种比较满意的作为统一綫型。

(五)确定统计参数的方法:按照上述綫型选择的原则,统一綫型规定之后,剩下的问题在于确定 $\bar{x}$ 及 C_v 。水文系列短,按单站资料计算 $\bar{x}$ 及 C_v 仍然可能有很大的误差,不能作为确定频率曲线的最后依据。均值 $\bar{x}$ 要考虑资料年份的丰枯水期。雨量的均值可以绘制全国的可分区的等值线图。雨量的 C_v 值可以利用各分区内各单站 C_v 值的算术平均值作为校核, C_v 值也可以绘制等值线图,或分区绘出 C_v 值,洪水和径流的 $\bar{x}$ 及 C_v 除与相似河流比较对照外,洪峰流量要利用洪水调查确定特大洪水的近似重现期,再计算 $\bar{x}$ 及 C_v 。

六、簡 結

(一)水文现象属于带有隐蔽因期性的时序系列,严格说来,不能按独立随机变数处理。因此,把现行水文统计方法的计算成果绝对化的看法是不妥当的。

(二)从水文现象的区域性质来看,现行以单站统计作根据的方法,等于把原来较大的统计场(水文区)人为地缩小成一个小面积或一个点。因而降低了统计成果的精度。因此,同一水文区内各站资料的联合使用是提高统计精度的重要方法。

(三)认为洪水频率服从二项式定理或普阿松定理是把后验机率与先验机率混同起来的提法,不妥当。二项式定理是分析先验机率问题的有力武器,在水文频率分析中没有得到应有的和正确的运用。

(四)本文利用开布屯推导偏态分布的方法分析水文系列分布性质的结果表明:

(1)周文德对水文系列的 $\ln$ 正态分布的理论解释是据根不足的。

(2)水文现象的成因过程很复杂,不能从机率理论证明水文频率属于何种分布律。

(五)洪水频率都服从耿贝尔分布律的说法无论在理论上或在实际检验中都是根据不足的。

(六)描述后验机率分布律的方法,按其数学处理方法的性质可以分为三类:

(1)概括分布通性的方法:如皮尔逊曲线,古德力区曲线等。

(2)利用级数和多项式修正基本分布律的方法:如格拉姆夏里埃曲线,勃罗夫柯维奇曲线等。

(3)函数转换的方法,如 $\ln$ 正态曲线,克里茨基——閔凯里曲线等。

值得提出注意的是:不但皮尔逊,古德力区等曲线是属于经验性质的频率曲线,就是原属理论性质的 $\ln$ 正态律和 $\ln$ 正态律在用以作为描述后验机率分布的模型时,也就成为经验频率曲线的性质了。

(七)不应该把水文频率计算神秘化和绝对化。水文频率曲线的綫型不应当让计算者去自由选择,而应当经过综合分析后统一规定下来。不能只靠单站资料去确定统计参数,而应当综合利用各站的资料。

(八)建设在两端有限 $\ln$ 正态曲线和皮尔逊Ⅲ型两种綫型中选择一种作为统一采用的綫型。建议从普通 $\ln$ 正态曲线上读取与选定的极限重现期(例如 10^4 年, 10^5 或 10^6 年)相应的 K 值作为两端有限曲线的上下限。如采用皮尔逊Ⅲ型曲线则必须把 C_s 作为 C_v 的倚变参数。

(九)选择和检验綫型的方法在于联合利用各站的资料(雨量按地区联合,径流按 C_v 值分组联合)来检验有资料部分的适合情况和外延部分的合理情况。

(十)在确定参数 $\bar{x}$ 及 C_v 方面建议,点雨量的 $\bar{x}$ 利用等值綫圈, C_v 采用分区平均值,以减少单站计算的误差和避免同区成果悬殊的矛盾。洪水频率计算,应尽量利用洪水调查等方法确定特大洪水的大小及其重现期。洪水和径流的 $\bar{x}$ 及 C_v ,除与相似河流比较对照外,利用同频率的径流与点雨量作比较,也能检查采用参数和计算成果是否合理。

参 考 文 献

1. 水利科学研究院水文研究所: 暴雨洪水频率计算方法的研究 1957, 8.
2. Б. Б. 格涅坚科: 概率论教程.
3. А. З. 龙西斯基: 概率论及数理统计学要义, 1956.
4. 金光炎编: 水文统计的原理与方法.
5. Д. И. 索科洛夫斯基, 河川径流.
6. Г. А. 阿列克塞也夫: 水文计算中所用的保证率的概念和标准的修正, 水利译丛, 1952, 2.
7. 谢家泽: 松花江流域水文特性的初步分析. 黑龙江流域综合考察学术报告(第一集).
8. С. Н. 克里茨基和 М. Ф. 閔凯里: 河川径流计算所用的机率曲线型式的选择. 水利译丛, 1956, 6.
9. А. Д. 薩瓦林斯基等, 苏联水文界关于水文学发展方向的学术争论, 水利译丛, 1957, 6.
10. Г. А. 阿列克塞夫: 苏联河流洪水径流的计算, 1956.
11. A. Hald, Statistical Theory with Engineering applications 1952.
12. Arnold Court: Some new statistical Techniques in geophysics.
13. Ven Te Chow: The log-probability Law and its Engineering Applications Proceeding A.S.C.E. Nov, 1954.
14. E. J. Gumbel: Statistical Theory of Extreme Values and Some Practical application 1954.
15. M. Jecques Bernier: Sur L'application des Diverses Lois Limites des Valeurs Extrêmes u Problème des Débits de Cure, La Houille Blanche Nov, 1956.
16. J. J. Slade: An Asymmetric Probability Eunction, Trans, A.S.C.E. Vol. 101, 1936.
17. Shigehisa Iwai: Duration Curves of Logrithmic normal Distribution Type and Their applications 1950.
18. "Floods in the United States" Waler-Supply Paper 771, 1936.

ON THE METHOD OF HYDROLOGICAL FREQUENCY ANALYSIS

Hsieh Chia-tseh

Institute of Hydrotechnical Research

Academia Sinica

and

Ministry of Water Conservancy and Power

Abstract

This paper attempts to answer the following two questions:

- (1) Is it possible to derive the law of distribution of hydrological frequency theoretically?
- (2) What type of distribution curve should be adopted as the model of hydrological frequency curve and how to determine their parameters?

The results obtained may be summarized as follows:

1. Hydrological phenomena are time series with concealed periodic fluctuations. The results from statistical analysis based upon the current assumption that hydrological phenomena are independent stochastic variables should be accepted with due considerations.
2. In view of the regional nature of hydrological phenomena, the current practice of analyzing samples taking from a single station only is, in effect, to narrow the sampling field arbitrarily from a larger area to a point, thus reducing the accuracy of the statistical results. Hence, the synthetic utilization of the data of all stations within the hydrologically homo-

geneous region is an important measure to increase the accuracy of statistical analysis.

3. The belief that the flood frequency obeys the binomial theorem or Poisson's theorem is but to mix up the priori with the empirical probability problem. The binomial theorem, being a powerful weapon to deal with the problems of priori probability, has not been adequately and properly utilized in the hydrological frequency analysis.

4. Analyses have been made of the nature of distribution of hydrological series on the basis of Kaptzen's derivation of the skew distribution, which indicate:

- (1) That the theoretical interpretation of the log-probability law of the hydrologic phenomena by V. T. Chow is not sound;
- (2) that hydrologic phenomena being results of very complicated meteorological and hydrological processes, it is impossible to derive theoretically the law of distribution for the hydrological series.

5. The view that the flood frequency obeys the Gumbel's distribution is theoretically not sound and also not verified by actual data.

6. According to the nature of the mathematical treatments applied, the method of description of the empirical probability can be classified into three systems:

- (1) The methods of the generalization of the characteristic factors of the distributions, such as Pearson's curves, Goodrich's curves, etc.;
- (2) The methods of the modification of a fundamental distribution by series and polynomials, such as Gram-Charlier curves, Бровкович curves, etc.;
- (3) The methods of transformed functions, such as the log-probability law, Крицкий-Менкель curves, etc.

It should be remarked that not only Pearson's and Goodrich's curves are frequency curves of empirical nature, but even the theoretical laws, such as the normal law and the log-probability law, will be accepted as curves of empirical nature, when used as models for empirical probability problem.

7. Hydrological frequency analysis should not be mystified and made absolute. Instead of free selections, the models of hydrological frequency curve should be uniquely selected and specified. Statistical parameters should be determined not solely by the short period data of single station, but also by the synthetic utilization of the data of possible more stations.

8. It is recommended that one of the two types of distribution, i.e. the log-normal frequency curve with both sides limited and the Pearson's type III curve, may be selected as unified models. The author suggests that the K-value corresponding to recurrence intervals of say 10^4 , 10^5 , or 10^6 years may be selected as the upper and lower limits for the log-normal curve. For Pearson's type III curves, C_s should be treated not as independent but as dependent variables of C_v .

9. The proper way to select and determine the model frequency curve is to see whether it fits well with the actual data of grouped stations (stations to be grouped by regions for rainfall data and by C_v for runoff data) and the reasonableness of the extrapolating part.

10. Suggestions on the method of determination of $\bar{x}$ and C_v : For point rainfall, iso- $\bar{x}$ map may be utilized, and the mean C_v for each hydrologic region may be adopted in order to minimize the errors from single stations and to avoid the discrepancies in results obtained from the same region. With regard to flood frequency analysis, flood mark reconnaissance must be utilized to determine the magnitude and the recurrence interval of the unusual flood. The $\bar{x}$ and C_v values of the floods and runoffs of hydrologically similar river basins may be compared. Besides, the reasonableness of the results of frequency calculations as well as of the statistical parameters adopted therein may be checked by comparing runoffs and point-rainfall values of the same frequency.