# LIMITED TYPE OF PRIMARY PROBABILITY DISTRIBUTION APPLIED TO ANNUAL MAXIMUM FLOOD FLOWS

By Bradford F. Kimball

*Port Washington, N. Y.*

**1. Theoretical statement of problem.** There is no doubt that Gumbel's recent paper "The Return Period of Flood Flows" [1] has supplied an admirably simple technique for engineers to use in approximating the trend of *return periods* of annual maximum flood flows for purposes of extrapolation. This treatment is scientifically of great interest because it introduces for the first time into a subject already treated at considerable length by engineers, the theory of the probability distribution of maximum values as developed by Fisher and Tippett, von Mises, and others.[1] However, certain further observations should be made concerning the approach used by Gumbel.

Let $x$ represent the measure of daily stream flow having a probability distribution $w(x)$. Let the probability distribution of the associated annual maximum stream flows be denoted by $V(x)$ with

$$W(x) = \int_0^x V(s)\, ds, \tag{1}$$

denoting probability that annual maxima be less than or equal to $x$. The *return period* $T(x)$ of an annual maximum flow of measure $x$ is then defined by

$$T(x) = \frac{1}{1 - W(x)}. \tag{2}$$

In this paper the probability distribution $w(x)$ will be called the *primary* probability distribution associated with the probability distribution of maximum values $V(x)$ and its *cumulative* distribution $W(x)$.

Gumbel argues that for the type of primary probability distribution that might reasonably be expected to apply, $W(x)$ will be of the type introduced by R. A. Fisher:

$$W(x) = \exp\,[-\exp - \alpha(x - u)]. \tag{3}$$

It is further implied that a primary probability distribution involving an upper limit would lead to a probability distribution of maximum values of the type

$$W_1(x) = \frac{k}{u}\left(\frac{u}{x}\right)^{k+1} \cdot e^{-(u/x)^k}, \tag{4}$$

for which moments of order $k$ or higher do not exist. The inference is then drawn that a primary probability distribution leading to such a cumulative distribution of maximum values would seem to be less likely to be the correct

[1] See references at end of Gumbel's paper, loc. cit.

one than one leading to the distribution (3). To this argument we do not object; but we question the implied conclusion that *hence the use of a limited type of primary distribution is to be disallowed.*

If the primary probability distribution be of the *limited* Galton type

$$w(x) = K \exp\left(-\tfrac{1}{2}u^2\right), \tag{5}$$

where $K$ is a constant and

$$u = k[b - \log(a - x)], \qquad 0 \leqq x \leqq a, \tag{6}$$

it can be shown that the limiting form of the cumulative distribution of maxima of $n$ values takes the same type form (3) where $x$ is replaced by $u$. This can be seen by observing that the transformed variate $u$ becomes infinite as $x$ approaches $a$, and hence has infinite range to the right, which places (5) in the category of distributions which are known to lead to cumulative distribution of maximum values of form (3). More explicitly, considering $w(x)$ as a finite distribution in $x$, if one traces the reasoning as set forth in von Mises' derivation [2] of the limiting distribution (3), one finds that since the cumulative primary probability $\int_0^x w(s)\,ds$ does *not* have a non-vanishing derivative of finite order at $x = a$, that what von Mises terms *the case of a limited distribution* does *not* apply, while the argument for a cumulative distribution of maxima of form (3) *does* carry through, in spite of the fact that $x$ has limited range to the right. This fact was not mentioned by Gumbel.

One is thus led to the conclusion that there is no logical exclusion of the assumption of a primary probability distribution of the form (5).

One might well argue for a first approximation of the actual primary probability distribution of stream flows—using any regular time interval such as a day or an hour—of the form (5). Differentiating $u$ with respect to $x$, one obtains

$$k\,dx = (a - x)\,du, \tag{7}$$

which means that to a constant probability increment $\Delta u$ there corresponds a maximum increment $\Delta x$ in measure of stream flow equal to $(a/k)\Delta u$ when $x$ is at the lower limit zero. This corresponding increment in stream flow decreases linearly to zero as $x$ approaches its upper bound $a$, imposed because of the existence of a finite watershed.

**2. Technique of fitting probability distribution of maximum values in case primary probability distribution is of the limited type (5)–(6).** Write the cumulative maximum distribution (3) in the form:

$$\begin{aligned} W(x) &= \exp(-\exp -y), \qquad y = \alpha(u(x) - u_1), \\ u(x) &= k[b - \log(a - x)], \qquad 0 \leqq x \leqq a. \end{aligned} \tag{8}$$

Now it is known that for the distribution

$$dW = e^{-e^{-y}} e^{-y} \, dy, \tag{9}$$

the mean value and standard deviation of $y$ are given by

$$\begin{aligned} \bar{y} &= .577215 \text{ (Euler's constant } C) \\ \sigma^2(y) &= \pi^2/6. \end{aligned} \tag{10}$$

Hence

$$\bar{y} = \alpha[\bar{u}(x) - u_1] = \alpha k[(b - u_1/k) - \bar{L}] = C$$

where $\bar{L}$ denotes the mean value of $\log(a - x)$, with $x$ representing the observed *maximum flood flows*. Also

$$\sigma(y) = \alpha k \, \sigma(L) = \pi/\sqrt{6}$$

where $\sigma(L)$ denotes the standard deviation of $\log(a - x)$. Hence

$$\alpha k = (\pi/\sqrt{6})/\sigma(L), \qquad b - u_1/k = C/\alpha k + \bar{L}, \tag{11}$$

and $y$ is determined as a function of $x$ by the relation

$$y = \alpha k[(b - u_1/k) - \log(a - x)]. \tag{12}$$

It is interesting to observe that it has not been necessary to determine the constants $k$ and $b$ of the primary probability distribution. Only the upper bound $a$ and observed flood flows are used in this process. From the relation (12) the theoretical curve in terms of $x$ may easily be computed from tables relating $y$ to $W$ (See Gumbel, loc. cit., Table II, page 173).

The difficulty of determining what the upper bound $a$ should be in a specific case is a practical one and does not concern the objective theoretical problem of choosing the *type* of curve which most nearly describes the behavior of annual maximum flood flows. The point to be made in this paper is that the use of what seems to be a reasonable value of $a$, *will* materially alter forecasts of future annual flood flows relative to forecasts made on the assumption that such an upper limit may be neglected. It is also ventured that the resulting theoretical probability distribution of maxima will in general give a better fit to the series of observed floods than one based on the latter premise. Techniques for determination of upper bound $a$ will not be discussed in this paper.

**3. Examples.** In order to demonstrate the point in question the two methods have been applied to a 57 year record of the annual flood flows of the Tennessee River at Chattanooga for the years 1875 to 1931.[2]

[2] The author has already used this series in a previous article [3] and for this reason has found it convenient to use it here.

TABLE I

*Series of observed annual flood flows*

(Tennessee River at Chattanooga, 1875–1931)

| (1) Observed Flood $x$ | (2) Ratio to Mean | (3) Per cent of Time | (4) Return Period, $T(x)$ |
|---|---|---|---|
| 85.9 | .412 | 0.88 | 1.007 |
| 108 | .518 | 2.63 | 1.027 |
| 123 | .590 | 4.39 | 1.043 |
| | | | ..... |
| 310 | 1.487 | 95.61 | 22.8 |
| 349 | 1.674 | 97.37 | 38.0 |
| 361 | 1.731 | 99.12 | 114. |

In Table I, col. (1) is shown the incomplete series of observed annual floods in units of 1,000 c.f.s. arranged in order of magnitude. The complete series may be referred to in *Water-Supply Paper 771* entitled "Floods in the United States," *U. S. Geological Survey*, 1936, p. 401. The mean annual maximum flood of this series is 208.56. The ratio of each annual maximum to the mean is shown in Col. (2). In Col. (4) is shown the observed return period which is taken here as the harmonic mean between what has been called the *exceedance interval* and the *recurrence interval* (see Gumbel, loc. cit., Table I, p. 167). Thinking of the 57 year record as a span of 57 years, the above procedure is equivalent to taking the observed probability $W(x)$ that a given annual flood will not be exceeded as the mid-point of the part of this time-span covered by the observed flood in question. Thus the lowest flood-peak 85,900 c.f.s. corresponds to the span from zero to 1.754 per cent of the whole time-span, and hence $W(x)$ is taken at the mid-point, $-0.877$ per cent. Similarly the greatest flood, 361,000 c.f.s. corresponds to interval from 98.246 to 100 per cent and is taken at 99.12 per cent. These arithmetic means correspond to harmonic means of the "recurrence" and "exceedance" intervals referred to above. This is the procedure which Hazen [4] originally followed.

Data from Cols. (1) and (4) of this table determined position of dots on Fig. 1. Data from Cols. (2) and (3) gave the points indicated by dots on Fig. 2, with $1 - W(x)$ recorded on the chart rather than $W(x)$.

The two theoretical distributions fitted to these annual flood maxima will be referred to as distributions A and B.

*Distribution A*. In this case the limited type of primary probability distribution (5) – (6) is assumed. From previous studies of this data series made by the author [3], an upper limit of annual floods of some 609,000 c.f.s. was found to be reasonable, and for purposes of this example the same upper limit will be assumed for the primary probability distribution. Thus the transformation (6) becomes:

$$u = k[b - \log (609 - x)],$$

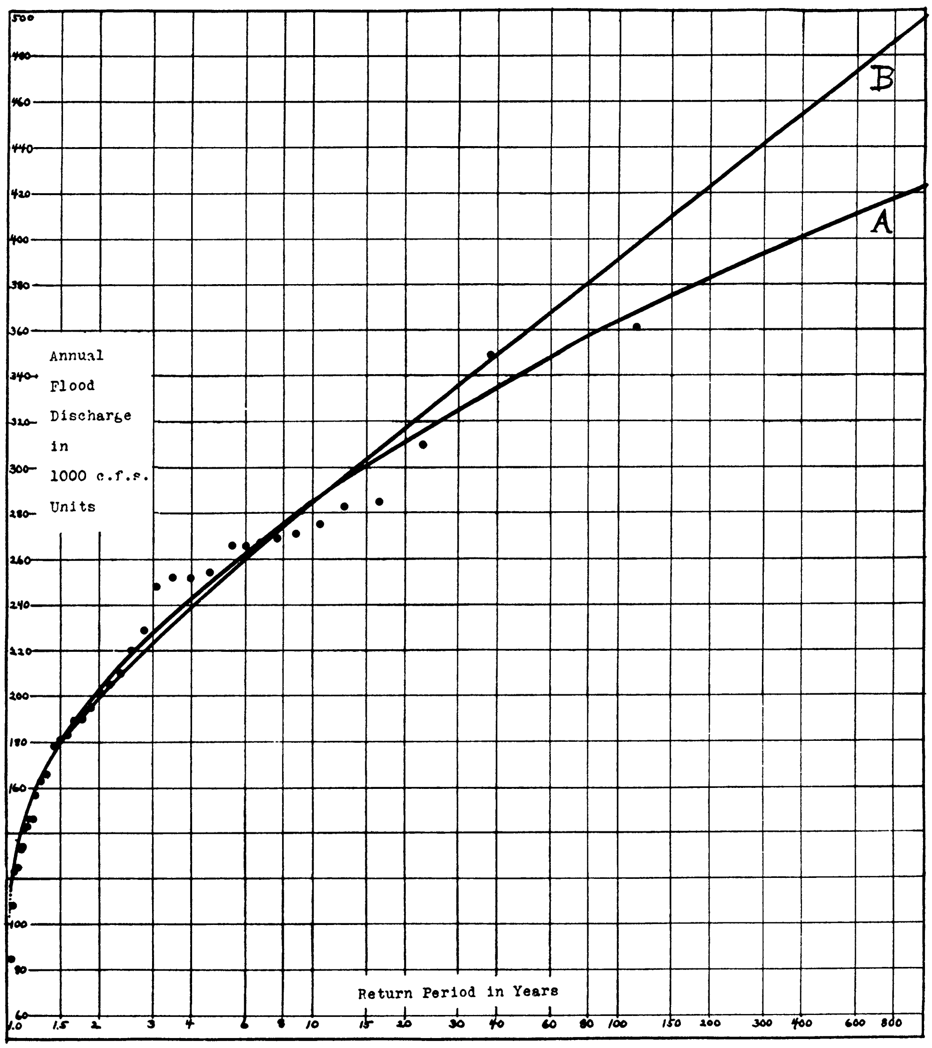


FIG. 1. Comparison of methods of fitting annual flood peaks, (Tennessee River at Chattanooga, 1875–1931)—return periods plotted against annual flood discharges on semi-logarithmic chart.

where the logarithm to base 10 can be used without loss of generality since the constant $k$ will absorb the conversion factor. The mean value of the logarithm, and its standard deviation come to

$$\bar{L} = 2.59772, \qquad \sigma(L) = .06576$$

The constants of the transformation (12) are thus determined by

$$\alpha k = (\pi/\sqrt{6})/(.06576), \qquad b - u_1/k = C/(\alpha k) + 2.59772$$

Thus

$$1/(\alpha k) = .05127, \qquad b - u_1/k = 2.6273$$

and solving (12) for log $(609 - x)$,

$$\log (609 - x) = 2.6273 - (.05127)\, y \tag{13}$$

Using a table for the known relations between $y$, $W(x)$, and $T(x)$ for the Fisher-Tippett distribution of maximum values similar to Table II of Gumbel's article

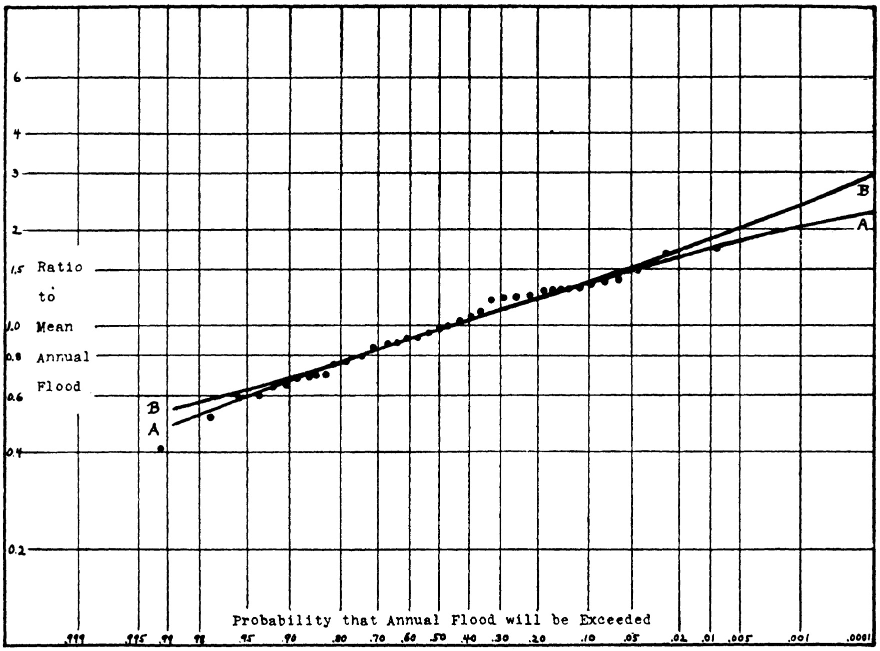


FIG. 2. Comparison of methods of fitting annual flood peaks, (Tennessee River at Chattanooga, 1875–1931)—Data plotted on logarithmic probability chart designed by Hazen, Whipple and Fuller.

(loc. cit.) the corresponding values $x$ of the annual floods are easily determined. Thus a theoretical relation between $x$ and $W(x)$ is set up. This is indicated as Curve A on the two charts exhibited here.

*Distribution B.* The primary probability distribution in this case is taken as unlimited to the right, and in general is assumed to have the character of an exponentially decreasing function of the measure of stream flow $x$ (see Gumbel, loc. cit.). The parameter $y$ of the distribution of annual maxima is given directly by

$$y = \alpha(x - x_1)$$

and

$$1/\alpha = (\sqrt{6}/\pi) \text{ (stand. dev. of annual floods)} = (.77970)\ (58.26) = 45.425$$

$$x_1 = \text{(mean annual flood)} - C/\alpha = 208.6 - (.57722)\ (45.425) = 182.4$$

Hence

$$x = 182.4 - (45.425)\, y \tag{14}$$

and using the table of corresponding values of $y$, $W(x)$ and $T(x)$ for the Fisher-Tippett distribution referred to above, a theoretical relation between $x$ and $W(x)$ is easily set up. This is plotted as Curve B on the accompanying charts.

**4. Discussion of examples.** In Fig. 1 it is to be noted that if theoretical curves are continued to the right to give readings for a return period of 1,000 years, the divergence of Curve A from Curve B is large enough to be of significance, numerically. Visual inspection does not indicate which curve is the better fit to the observation points.

In Fig. 2 the curves are plotted on "logarithmic probability" graph paper. This paper was designed by Hazen and Fuller [4] specifically for the purpose of plotting annual maxima of stream-flows. A significant divergence in trend is to be noted at the right hand end.

These charts indicate that the use of an upper limit may materially affect extrapolation of fitted theoretical curves, for purposes of estimating floods with a return period, say of 1,000 years.

If the trends of observed floods in Gumbel's recent paper in the *Transactions of the American Geophysical Union* [5] are examined, it will be observed that in the case of the Connecticut, Mississippi and Rhone rivers, there is a decided tendency for the curve of observed floods to turn downwards, away from the theoretical curves, which correspond to Curve B exhibited in Figure 1. In the case of the Tennessee, Cumberland and Columbia rivers the tendency is not decisive, while in the case of the Rhine river at Basel (Switzerland) the tendency of the observed curve is upwards rather than downwards. As the writer has observed elsewhere [6], this last data series seems to be rather unique in character and is possibly the result of a watershed greatly influenced by all year around snow deposits. Possibly a radically different primary probability distribution should be used in this case.

**5. Conclusion.** The writer has demonstrated in this paper that in fitting a theoretical probability distribution of maximum values to annual maxima of stream flows, the use of an upper bound for measures of stream flow by assumption of a primary probability distribution of the type (5)–(6)

(1) is not inconsistent with the use of the Fisher-Tippett distribution of maxima,

(2) has a reasonable logical basis from the point of view of the hydrologist,

(3) may materially affect the estimation of return periods when extrapolation is involved, relative to results obtained when no upper bound is assumed.

It has not been within the scope of this paper to discuss techniques for determining such an upper bound, nor to apply the theory to enough data series to draw conclusions concerning goodness of fit.

## REFERENCES

[1] E. J. GUMBEL, "The return period of flood flows," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 163–190.

[2] RICHARD VON MISES, "La distribution de la plus grande de $n$ valeurs," *Revue Mathematique de l'Union Interbalkanique*, Vol. 1, Athens (1936).

[3] BRADFORD F. KIMBALL, "Probability-distribution-curve for flood-control studies," *Trans. Am. Geographical Union*, 1938, pp. 466–475.

[4] ALLEN HAZEN, *Flood-flows*, John Wiley & Sons, New York, 1930.

[5] E. J. GUMBEL, "Probability-interpretation of the observed return-periods of floods," *Trans. Am. Geophysical Union*, 1941, pp. 836–850.

[6] BRADFORD F. KIMBALL, *Discussion* of paper entitled "Statistical control curves for flood discharges" by E. J. Gumbel, *Trans. Am. Geophysical Union*, 1942.