THE RETURN PERIOD OF FLOOD FLOWS

By E. J. Gumbel

New School for Social Research

Introduction. Engineers have used various interpolation formulas to represent the observed distribution of flood discharges. These formulas are sometimes constructed *ad hoc* for a given stream, and have no general meaning. Most of them are rather complicated.[1] Some authors have tried to introduce upper and lower limits to the discharges, even though it is doubtful that such limits exist. Others have introduced the third and fourth moments of the distribution, in spite of the fact that these numerical values are subject to large errors. For some formulas it is impossible to give a meaning to the constants; different formulas applied to the same stream give rather contradictory results; and consequently there is considerable confusion. For example, Slade [20] has stated that "the statistical method in whatever form employed is an entirely inadequate tool in the determination of flood frequencies." According to Saville [19] "the engineer should satisfy himself that he has used an adequate number of methods, whether mathematical, graphic or otherwise, which have real support from either theory or experience, and then form his own judgement."

The main reason for this situation is that these studies have little or no theoretical basis. The author believes it possible to give exact solutions, exactitude being interpreted from the standpoint of the calculus of probabilities [10]. Our solutions are simply the consequences of a truism: "The flood discharges are the largest values of the discharges." The present study is but an explanation of this statement.

Many American authors start with a statistical function, which we call the return period of floods. Therefore we shall first analyse the notion of return period and show how it can be derived as a consequence of the concept of distribution. We then give a short résumé of the theory of largest values. The discharge, and in consequence the flood discharge, is considered as an unlimited statistical variable; it is not necessary to determine its distribution. We are justified in representing the observed distribution of flows by one of the theoretical distributions of largest values. The distribution we choose contains only two constants, and both have a clear hydrological meaning. The numerical values are calculated by the method of moments.

---

[1] In recent years many articles discussing this topic have been published by the American Society of Civil Engineers and the American Geophysical Union [8]. A review of some of the proposed formulas is given in the Water Supply Paper 771 [17].

The application of the notion of return period to the largest values leads to a simple formula for the return period of the floods. In the last part of this paper we represent the flood flows of the Rhône and Mississippi Rivers by our formula.

1. The return period. Let us consider a continuous statistical variable x, having a theoretical distribution $w(x)$. The probability $W(x)$ of a value less than or equal to x, and the probability $P(x)$ of a value greater than or equal to x, are

$$(1) \qquad W(x) = \int_{-\infty}^{x} w(z)\,dz, \qquad P(x) = \int_{x}^{\infty} w(z)\,dz,$$

where z denotes the variable of integration. Clearly

$$(1') \qquad W(x) + P(x) = 1.$$

Let n be the number of observations. Let x_m ($m = 1, 2, \cdots, n$) be the observed values arranged in increasing magnitude, where m is the serial number beginning with the lowest ("from below"). The lowest observation has the serial number $m = 1$, the highest has the serial number $m = n$. These observed values will be written x_1, and x_n respectively. The number of observations below or equal to x_m is $m = n'W\,(x_m)$ where $'W(x_m)$ is the observed relative number corresponding to the probability $W(x)$. The graphic representation of this series is called a cumulative histogram.

In hydraulics many authors arrange the observations in decreasing magnitude. Let ${}_mx$ ($m = 1, 2, \cdots, n$) be these observed values. The serial number m is counted in a descending scale ("from above"). For the largest value $m = 1$, for the lowest value $m = n$. The number of observations above or equal to ${}_mx$ is $m = n'P({}_mx)$ where $'P({}_mx)$ corresponds to $P(x)$. The numbers $'W(x_m)$ will never decrease; the number $'P({}_mx)$ will never increase. The mth value on a descending scale is the $n - m + 1$th value on an ascending scale. Therefore

$$(2) \qquad n'P({}_mx) = n - n'W(x_m) + 1,$$

and

$$(2') \qquad nP(x) = n - nW(x).$$

The difference between formulas (2) and (2') will play a certain rôle later.

Different methods are used in statistics in comparing the theoretical values $W(x)$ or $P(x)$ and $w(x)$ with the corresponding observations $'W(x_m)$, or $'P({}_mx)$ (cumulative frequencies) and $\Delta'W(x_m)$ (frequency distribution). They all have in common an arrangement of observed values according to magnitude.

For the purpose of considering the observations in chronological order, we introduce a statistical criterion which at first glance may appear to have a new logical structure. It is assumed here that the observations are made at constant time intervals, and this interval is considered the unit of time. We suppose that the observations are homogeneous, i.e., subject to a common set of forces.

Furthermore, we suppose that the events are independent of one another: the occurrence of a high or low value for x has no influence on the value of any succeeding observation. Let us choose a low value x, and ask the following: After what number of observations does this or a greater value return? We calculate the mean of these chronological intervals between every two consecutive values, equal to or greater than x. We repeat these operations for a second, third, . . . till the penultimate value of x.

These means are called the *observed return periods*. The criterion consists of the comparison of the observed, and the theoretical return period for increasing values of x. For a discontinuous variable we could obtain the return period for a value equal to x, (not equal to or greater than x). This average time, which is sometimes used in physics, does not interest us, as our variable, the discharge, is continuous. We limit our consideration to the return period of a value equal to or greater than x, called: value greater than x.

The determination of the theoretical return period is a classical problem: How many trials must, on the average, be made, in order that an event of a given probability should happen? Our event, the realization of a value, equal to or greater than x, has the probability $P(x) = 1 - W(x)$.

The mean number of trials $T(x)$ which are necessary to obtain our event once, is evidently

$$T(x) = \frac{1}{1 - W(x)}, \tag{3}$$

or

$$T(x) = \frac{1}{P(x)}. \tag{3'}$$

This value $T(x)$ is the mean chronological interval between two values, equal to or greater than x. If we start at the time when such a value has been observed for the first time, we can interpret $T(x)$ as the theoretical return period of a value equal to or greater than x. We designate it as the *theoretical return period*. This concept has not been used in statistics. It is a well-known concept in hydraulics which was introduced by Fuller [6]. To every theoretical distribution $w(x)$ there is a corresponding return period $T(x)$ and conversely, to every theoretical return period $T(x)$ there is a corresponding distribution

$$w(x) = \frac{T'(x)}{T^2(x)}, \tag{4}$$

obtained by differentiating (3).

If the variable is without limit to the left, the return period will start with $T = 1$. If the variable is limited to the left by $x \geqq \epsilon$ the corresponding return period will be

$$T(\epsilon) \geqq 1 \qquad \text{if } W(\epsilon) \geqq 0 \tag{5}$$

In the graphic representation, the return period $T(x)$ which has a time dimension, will be the abcissa and x the ordinate. Therefore we consider x as a function of $T(x)$; from (4) we obtain

$$\frac{dx}{d \ln T} = \frac{1}{w(x)T(x)} \tag{6}$$

where ln signifies the natural logarithm. The increase of x as a function of $\ln T(x)$ will be very rapid for small values of T. For a limited distribution the same result is obtained, provided the probability $W(\epsilon)$ and the density of probability $w(\epsilon)$ are sufficiently small. Clearly, the return periods of the three quartiles are respectively $1\frac{1}{3}$, 2, 4. The return period will always increase with x. It will tend towards infinity even if the variable is limited to the right.

Let us now consider the calculus of the observed return periods. Instead of values equal to or greater than x_m we will only speak of values greater than x_m. The observed return period is the interval between the first and the last observation greater than x_m, divided by the number of intervals between all observations greater than x_m. The number of observations greater than x_m is $n - n'W(x_m)$. Between these observations there are $n - n'W(x_m) - 1$ intervals. This denominator is independent of the chronological order of the observed values. We can calculate the mean of the observed intervals up to a value x_m so that $n - n'W(x_m) = 2$. For this value of x_m there are only two observations, i.e., only one interval. In that case no mean can be calculated.

The numerator, the interval between the first and the last observation greater than x_m will be $n - 1$, provided that the first and the last value in chronological order are greater than x_m. But in general the first value greater than x_m will be the $('k + 1)$th in chronological order. The first value greater than x_m found in the reverse chronological order, will be the $(k' + 1)$th. Let $'k + k' = l$, then the interval between the last and the first value greater than x_m is $n - 1 - l$. The mean observed interval is thus

$${}_1T(x_m) = (n - 1 - l)/(n - 1 - n'W(x_m)),$$

or

$${}_1T(x_m) = \left(1 - \frac{l}{n-1}\right) \Big/ \left(1 - \frac{m}{n-1}\right). \tag{7}$$

This magnitude depends only on the chronological order of the first and the last value greater than x_m. It is independent of the chronological order of all other observations. Even in the case $l = 0$ this value differs from the theoretical value (3). The observed value surpasses the theoretical value, even if the frequency $'W(x_m)$ is identical with the probability $W(x)$.

In the general case, $l > 0$, this difference is a function of l. The number l depends upon the times at which the observations begin and cease; but it is not a characteristic of the chronological order. As a result of these disadvantages of formula (7) we prefer to introduce other definitions, in which the chronological order does not enter. These definitions have an added advantage in that they are constructed in a manner analogous to the theoretical formula.

The observed value which corresponds to (3) is

$$'T(x_m) = \frac{n}{n - n'W(x_m)}, \tag{8}$$

or

$$'T(x_m) = n/(n - m). \tag{9}$$

But this definition of the observed return period is not the only one which corresponds to (3). Starting with the serial number m, in a descending scale, Fuller [6] puts

$$''T(x_m) = \frac{n}{m}. \tag{8'}$$

According to this definition, the return period of the mth value from below is

$$''T(x_m) = n/(n - m + 1). \tag{9'}$$

TABLE I

Two definitions of the observed return periods

observed variable	serial number from below	serial number from above	exceedance interval formula (9)	recurrence interval formula (9')
x_1	1	n	$n/(n-1)$	1
x_2	2	$n-1$	$n/(n-2)$	$n/(n-1)$
x_m	m	$n-m+1$	$n/(n-m)$	$n/(n-m+1)$
x_{n-1}	$n-1$	2	$n/1$	$n/2$
x_n	n	1	—	$n/1$

This observed return period corresponds to the theoretical return period (3'). The difference between (9) and (9') results from the fact that the relation (2) between the observed cumulative frequencies $'W(x_m)$ and $'P({}_mx)$ differs from the relation (2') between the probabilities $W(x)$ and $P(x)$. The two definitions of the observed return periods are related by

$$''T(x_{m+1}) = 'T(x_m) < 'T(x_{m+1}). \tag{10}$$

From a purely logical standpoint the first definition is as justifiable as the second one. Both are used in hydraulics. In order to avoid confusion between formulas (9) and (9') Horton [16] calls $'T(x_m)$ the *exceedance* interval, i.e., "the average interval at which an event of given magnitude is exceeded," whereas he defines $''T(x_m)$, the recurrence interval as "the average interval of occurrence of values equalling or exceeding a given magnitude." Of course, the exceedance interval surpasses the recurrence interval. Since both observed intervals correspond to a common theoretical return period we designate both of them as observed return periods.

The difference between formulas (9) and (9') is made clear in Table I.

Each of the definitions (9) and (9′) and the theoretical expression $T(x)$ has different properties. For the lowest observation

$$n'W(x_1) = 1; \qquad n'P({}_nx) = n.$$

Therefore

$$'T(x_1) = 1 + \frac{1}{n-1}; \qquad ''T(x_1) = 1,$$

whereas for an unlimited distribution $\lim_{x \to -\infty} T(x) = 1.$

If the number of observations is sufficiently large the numerical differences between the two observed periods are rather small, except for very large values of the variable. For the last observation

$$n'W(x_n) = n; \qquad n'P({}_1x) = 1.$$

Therefore the return period $'T(x_n)$ for the last observation does not exist. According to the second definition the return period for the last value is equal to the total number of observations. But in general there is only one observation of the last value.

The preference given formula (9) over (9′) corresponds with the preference given to $W(x)$ over $P(x)$ when comparing the theoretical with the observed values. Therefore it is natural to count m from below. Since both definitions are equally applicable and since they lead to different results for large values of the variable, one should not calculate the return period for a small number of observations.

The observed return periods (9) and (9′) differ from the theoretical return period (3) in the same way that the frequencies $'W(x_m)$ or $'P({}_mx)$ differ from the probabilities $W(x)$ or $P(x)$. The chronological order enters neither into formula (7) nor into (9) or (9′). We need not take it into consideration, since the theoretical return period is obtained from the probability and the observed return period from the cumulative histogram. Therefore the usual statistical methods can be used for making the comparison between observed and theoretical return periods.

The return period is a statistical function like the distribution, $w(x)$ or the probability $W(x)$. No formula for $T(x)$ that contradicts the properties of $w(x)$ can be accepted. The return period $T(x)$ will contain the same number of independent constants as the distribution $w(x)$. Consequently the fit of the theoretical curve $T(x)$ to the observations $'T(x_m)$ or $''T(x_m)$ cannot be improved by introducing a new constant without also changing the distribution $w(x)$. The theoretical curve $x = f(T)$ will fit the observed curves $(x_m, 'T(x_m))$ and $(x_m, ''T(x_m))$ in a way that depends upon the fit of $W(x)$ and $P(x)$ to $'W(x_m)$ and $'P({}_mx)$.

Let us suppose that $w(x)$ contains k constants; that they are determined by the method of moments which conserves the arithmetic mean $\bar{x}$, the mean of the squares $\overline{x^2}$ etc. of the observed distribution. For the return period these moments have a meaning. Let us consider for the sake of simplicity a positive variable. The kth moment M_k

$$M_k = \int_0^\infty x^k \, dW(x)$$
$$= -\int_0^\infty x^k \, d(1 - W(x))$$
$$= k \int_0^\infty (1 - W(x)) x^{k-1} \, dx$$

is according to (3)

$$M_k = k \int_0^\infty \frac{x^{k-1} \, dx}{T(x)}, \tag{11}$$

whence for $k = 1$ and $k = 2$

$$\bar{x} = \int_0^\infty \frac{dx}{T(x)}; \qquad E(x^2) = 2 \int_0^\infty \frac{x \, dx}{T(x)}. \tag{11'}$$

For a given distribution containing two constants, the method of moments conserves the area and the center of gravity of the reciprocal of the return period. Even if the method of methods gives the best determination of the constants, for the distribution, it need not give the best determination for the return period. But if the observed return periods were used for the determination of the constants we would get two sets, since there are two observed curves having equal validity, but different values for large x. We will get one and only one set if the constants are calculated from the observed distribution, for here the difference between $'T(x_m)$ and $''T(x_m)$ does not matter. The fact that we do not take the constants from the observed return periods, but from another statistical function, might be a cause for deviations between the observed and the theoretical return periods.

Once the constants have been found, we compare the observed curves $(x_m, 'T(x_m))$ and $(x_m, ''T(x_m))$ with the theoretical curve $x = f(T)$. To avoid discontinuity the observed return period will be established for all values of x_m arranged in increasing order.

If the observed return periods for small values of x are systematically smaller (greater) than the theoretical period, it is reasonable to conclude that there exists an attraction (repulsion) for small values of the variable and a repulsion (attraction) for the large values. But it must be remembered that the observed values have different weights in that the return periods for small values of x are based on many observations. This number diminishes as x increases. The last observed return period is based only on two observations. Therefore the divergence between theory and observation will increase with the variable. With this precaution the criterion of the return period suggests one cause of difference between theory and observation. In order to apply this method to the largest values we must first establish the corresponding distribution.

2. Theory of the largest value. Let x be a statistical variable unlimited to the right having the distribution $w(x)$. Among the N observed values, one will be larger than the others. We wish to determine its theoretical value.

According to the principle of multiplication the probability $\mathfrak{W}_N(x)$ that N values are inferior to x is

$$\mathfrak{W}_N(x) = W^N(x). \tag{12}$$

This is the probability of x being the largest value. The largest value is a new statistical variable which possesses a mode, a mean $\bar{u}$, a standard deviation s and higher moments. To get the mean the distribution $\mathfrak{w}_N(x)$ of the largest value is needed. From (12) by differentiation

$$\mathfrak{w}_N(x) = NW^{N-1}(x)w(x). \tag{13}$$

The mode will be the solution of

$$\frac{N-1}{W(x)}\, w(x) + \frac{w'(x)}{w(x)} = 0. \tag{13'}$$

For a given initial distribution $w(x)$ and for small N we have to solve this equation. But the mean and the moments cannot be obtained in a general way by the use of the exact distribution (13). However we can reach general solutions if N is large, provided we limit ourselves to certain classes of initial distributions. We have studied this problem in previous publications [11–13]. For our present purpose it is sufficient to give the results in a form due to R. von Mises [18].

We define a large value u of the variable x by

$$N(1 - W(u)) = 1. \tag{14}$$

This means that the expected number of observations equal to or greater than u is one. Equation (14) is but another form of definition (3). The mean number of trials is used in (3) whereas the original variable x is used in (14).

The probability $\alpha\, du$ that a value greater than u will be contained between u and $u + du$ is given by

$$\alpha = \frac{w(u)}{1 - W(u)}. \tag{15}$$

Obviously α and u are functions of N and the constants in the initial distribution $w(x)$. There are two limiting forms of the probability (12)

$$\lim_{N\to\infty} W^N(x) = F(x); \qquad \lim_{N\to\infty} W^N(x) = \mathfrak{W}(x).$$

If

$$\lim_{u\to\infty} \alpha u = k > 0, \tag{16}$$

we obtain

$$F(x) = e^{-(u/x)^k}. \tag{17}$$

This probability function was first established by Fréchet [5]. If

$$\lim_{u \to \infty} \frac{d}{du}\left(\frac{1}{\alpha}\right) = 0, \tag{18}$$

we obtain

$$\mathfrak{W}(x) = e^{-e^{-\alpha(x-u)}}. \tag{19}$$

This probability function is due to R. A. Fisher [4]. Let us consider the first limit. The initial distributions which lead to it belong to the *Pareto type.* For this distribution

$$w(x) = \frac{k}{x^{k+1}}; \qquad W(x) = 1 - \frac{1}{x^k}; \qquad x \geqq 1$$

and condition (16) holds; for *any* value of x

$$\frac{xw(x)}{1 - W(x)} = k.$$

The distribution $f(x)$ of the largest value, which corresponds to (17), is

$$f(x) = \frac{k}{u}\left(\frac{u}{x}\right)^{k+1} e^{-(u/x)^k}. \tag{20}$$

The mode $\tilde{x}_N$ of the largest value is the solution of

$$\frac{d}{dx}\left[(k+1)\ln\frac{u}{x} - \left(\frac{u}{x}\right)^k\right] = 0,$$

hence

$$\frac{k+1}{x} = \frac{ku^k}{x^{k+1}},$$

or

$$\tilde{x}_N = u\left(\frac{k}{k+1}\right)^{1/k}. \tag{21}$$

According to the definition (14) the mode of the largest value will increase with N. For a finite number of observations, which is always the case, the mode will be limited. But the moments of order k or higher will not exist. For $k < 1$, no moment will exist. For $k < 2$, only the first moment, the mean, exists, and so on.

Let us consider now the second limit (19). The initial distributions which lead to it belong to the *exponential type.* For this distribution [14]

$$w(x) = e^{-x}; \qquad W(x) = 1 - e^{-x}; \qquad x \geqq 0,$$

and for *any* value of x

$$\frac{d}{dx}\left(\frac{1 - W(x)}{w(x)}\right) = 0,$$

which means that condition (18) is fulfilled. Most of the distributions used in statistics belong to this type. According to (19) the distribution of the largest value is

$$\mathfrak{w}(x) = \alpha e^{-\alpha(x-u)-e^{-\alpha(x-u)}}. \tag{22}$$

If we introduce a reduced variable y without dimension by the linear transformation

$$y = \alpha(x - u), \tag{23}$$

we get the reduced probability $\mathfrak{B}(y)$

$$\begin{aligned} \mathfrak{B}(y) &= \mathfrak{W}(x) \\ &= e^{-e^{-y}}. \end{aligned} \tag{24}$$

The numerical values of this function, calculated by means of Becker's tables [1], are given in Table II, col. 1 and 2. The reduced distribution

$$\mathfrak{v}(y) = e^{-y-e^{-y}}, \tag{25}$$

makes clear the meaning of u: the distribution has one and only one maximum which occurs for the reduced value $y = 0$. Therefore u is the mode of the largest value for a given set of N observations. For an initial distribution $w(x)$ satisfying (18), and for large N, definition (3) of the return period as a function of x becomes identical with relation (14) which involves the number of observations N and the corresponding most probable value u.

We wish to decide which distribution of the largest value is to be used to represent the given observations. This decision depends, according to (16) and (18), on the nature of the initial distribution at the extreme values of the variable. If the law of the observed initial variable is known, a precise answer can be given. But generally speaking, a distribution chosen to represent given observations is nothing but an interpolation formula. Formulas having different analytical properties may all give satisfactory results. One might fulfill condition (16), and another (18). The conditions apply to the differential coefficient, whereas the initial observations are always discontinuous. Therefore they will not enable us to decide which, if any, of the conditions is met. For extreme values of the variable x the observed differences are large and nonuniform, and there is therefore no way to replace the differentiation by a finite difference. Consequently we have to use the observations of the largest values to control the two competing theories and not the conditions. The fact that distribution (20) has higher moments only under certain conditions, is a strong practical argument in favor of distribution (22). Therefore the following development will be based on this distribution.

It can be shown that the mean error θ of distribution (22) is related to the constant α by

$$\theta = 0.98/\alpha. \tag{26}$$

Therefore the constant u is the most probable largest value for N observations and $1/\alpha$ a multiple of the mean error.

TABLE II

Probabilities and return periods of largest values

reduced variable y	probability $\mathfrak{W}(x)$	return period $\log T(x)$	Flood discharges per second	
			in cubic meter x Rhône R.	in 1000 cubic feet x Mississippi R.
−2.00	0.00062	0.000		
−1.75	0.00317	0.001		
−1.50	0.01131	0.005	1355	803
−1.25	0.03049	0.013	1492	869
−1.00	0.06599	0.030	1629	936
−0.75	0.12039	0.056	1766	1002
−0.50	0.19230	0.093	1903	1069
−0.25	0.27693	0.141	2040	1135
0.00	0.36788	0.199	2177	1202
0.25	0.45896	0.267	2314	1268
0.50	0.54524	0.342	2451	1335
0.75	0.62352	0.424	2588	1401
1.00	0.69220	0.512	2725	1468
1.25	0.75088	0.604	2862	1534
1.50	0.80001	0.699	2999	1601
1.75	0.84048	0.797	3136	1667
2.00	0.87342	0.899	3273	1734
2.25	0.89996	1.000	3410	1800
2.50	0.92119	1.103	3547	1867
2.75	0.93807	1.208	3686	1933
3.00	0.95143	1.314	3822	2000
3.25	0.96197	1.420	3959	2066
3.50	0.97025	1.527	4096	2133
3.75	0.97675	1.634	4233	2199
4.00	0.98185	1.741	4370	2266
4.25	0.98584			
4.50	0.98895			
4.75	0.99138			
5.00	0.99329			
5.25	0.99477			
5.50	0.99592			
5.75	0.99682			
6.00	0.99752			

TABLE III

Observed return periods

Rhône, Lyon (France) (1826–1936)

Flood discharge x_m	Serial number m	Return period log $'T(x_m)$	Flood discharge x_m	Serial number m	Return period log $'T(x_m)$
899	1	.004	2475	57	.313
1172	2	.008	2475	58	.321
1231	3	.012	2475	59	.329
1272	4	.016	2491	60	.338
1272	5	.020	2514	61	.346
1432	6	.024	2514	62	.355
1432	7	.028	2514	63	.364
1439	8	.032	2514	64	.373
1444	9	.037	2538	65	.382
1502	10	.041	2554	66	.392
1541	11	.045	2586	67	.402
1560	12	.050	2594	68	.412
1639	13	.054	2594	69	.422
1706	14	.058	2594	70	.432
1780	15	.063	2602	71	.443
1829	16	.068	2626	72	.454
1850	17	.072	2627	73	.465
1857	18	.077	2643	74	.477
1913	19	.081	2675	75	.489
1913	20	.086	2675	76	.501
1934	21	.091	2773	77	.514
1955	22	.096	2773	78	.527
1992	23	.101	2773	79	.540
1992	24	.106	2839	80	.554
2006	25	.111	2856	81	.568
2006	26	.116	2881	82	.583
2013	27	.121	2881	83	.598
2050	28	.126	2965	84	.614
2050	29	.131	3007	85	.630
2072	30	.137	3050	86	.647
2094	31	.142	3058	87	.665
2101	32	.148	3067	88	.684
2115	33	.153	3067	89	.703
2145	34	.159	3126	90	.723
2145	35	.164	3179	91	.744
2153	36	.170	3214	92	.766

TABLE III—*Concluded*

Flood discharge x_m	Serial number m	Return period log $'T(x_m)$	Flood discharge x_m	Serial number m	Return period log $'T(x_m)$
2160	37	.176	3250	93	.790
2168	38	.182	3266	94	.825
2175	39	.188	3293	95	.841
2206	40	.194	3310	96	.869
2206	41	.200	3310	97	.899
2206	42	.206	3354	98	.931
2221	43	.213	3426	99	.966
2236	44	.219	3444	100	1.004
2240	45	.226	3444	101	1.045
2258	46	.232	3480	102	1.091
2281	47	.239	3606	103	1.142
2296	48	.246	3625	104	1.200
2327	49	.253	3708	105	1.267
2342	50	.260	3801	106	1.346
2358	51	.267	3810	107	1.443
2381	52	.274	3905	108	1.568
2420	53	.282	4096	109	1.744
2444	54	.289	4105	110	2.045
2452	55	.297	4390	111	
2467	56	.305			

$\Sigma x_m = 276{,}773.$ $\Sigma x_m^2 = 744{,}538{,}565.$

The arithmetic mean $\bar{u}$ of distribution (22) is [4]

$$\bar{u} = u + \frac{c}{\alpha}, \tag{27}$$

where $c = 0.5772157$ is Euler's constant. The standard deviation s is

$$s = \pi/\alpha\sqrt{6}. \tag{28}$$

Therefore

$$\bar{u} = u + 0.45005s. \tag{29}$$

The reduced variable y introduced by (23) is related to the reduced variable

$$z = \frac{x - \bar{u}}{s} \tag{30}$$

by

$$z = \frac{\alpha\sqrt{6}}{\pi}(x - u) - \frac{c\sqrt{6}}{\pi}.$$

The substitution of the numerical values leads to

$$z = 0.77970y - 0.45005. \tag{30'}$$

Conversely,

$$y = 1.28255z + 0.57722. \tag{31}$$

The value (32) $v = s/\bar{u}$, the coefficient of variation, is related to the product αu. By (27) $\alpha u = \alpha\bar{u} - c$ and by (28)

$$\alpha u = \frac{\pi}{\sqrt{6}} \cdot \frac{1}{v} - c. \tag{33}$$

Therefore the numerical value of αu can also be considered as a characteristic of an observed distribution of largest values.

For the two constants we calculate for the observed distribution of largest values the two first moments

$$\bar{u} = \frac{1}{n} \sum_{m=1}^{n} x_m, \tag{34}$$

and

$$\overline{u^2} = \frac{1}{n} \sum_{m=1}^{n} x_m^2. \tag{35}$$

To get the observed standard deviation we use the Gaussian formula

$$s = \sqrt{\left(1 + \frac{1}{n-1}\right)(\overline{u^2} - \bar{u}^2)}. \tag{36}$$

According to (28) and (27)

$$\frac{1}{\alpha} = 0.7796968s, \tag{37}$$

and

$$u = \bar{u} - \frac{0.5772157}{\alpha}. \tag{38}$$

These formulas give the two constants in the distribution of largest values.

3. Flood flows interpreted as largest values. We will now apply the theory of largest values to flood flows. Let us consider the daily flow as a statistical variable, unlimited to the right. This idea is not new. The formulas proposed by Fuller [7], Hazen [15], and numerous other authors all incorporate this assumption. Gibrat [9] supposes that the daily flows vary according to Galton's distribution. Instead of postulating a specific formula for the distribution of flows we shall only suppose that it belongs to the usual exponential type, which means that condition (18) is fulfilled.

We define a flood as being the largest value of the $N = 365$ daily flows. The flood flows are therefore the largest values of flows. This commonplace implies the distinction between floods and inundations. For each year there exists one or more floods of the same magnitude, but there might exist several different inundations or none at all. If there are several inundations in a year the greatest one will be a flood; but a flood need not to be an inundation: even a dry year has a flood. We limit ourselves to floods, assume that $N = 365$ is a large number, and represent the distribution of annual floods by the distribution (22) of largest values.

There have been objections to the concept that the daily flow is an unlimited variable. Horton [16] believes that this implies the absurd idea of unlimited floods. This opinion is shared by Slade [20], who claims that there is a definite upper limit to the magnitude of the floods for a given stream. The theory of largest values confirms only partially Horton's opinion. If we should choose distribution (20), the most probable annual flood will be limited. For this distribution, however, it might happen that the mean annual flood has no meaning. To avoid this we have chosen distribution (22), for which the mean annual flood and all the moments will be finite. A further justification of the use of (22) might be derived from the fact that Galton's distribution belongs to the exponential type. As a final argument, numerical calculations show that formula (22) gives a better fit to the observed distributions of flows.

The variable x is the annual flood flow measured in cubic meters or cubic feet per second. The mean $\bar{u}$ is the annual mean flood, whereas u is the most probable annual flood. The value s is the standard deviation of the distribution of annual floods. Finally y is called the reduced flood.

The distribution (22) possesses the properties of the observed distribution of flood flows. It is asymmetrical; rising rather quickly but falling rather slowly. The modal value is to the left of the mean (see Fig. 3).

To apply the theory of return periods let us consider the event of the highest annual discharge being greater than x. We have to replace in formula (3) the general probability $W(x)$ by the probability of flood discharges (19). The number of observations n is the number of years for which observations exist.

To use formula (3) we have to suppose that the intervals between the successive floods are all equal to one year. This assumption conforms more or less to the seasonal nature of floods.

The return period of a flood greater than x

$$(39) \qquad T(x) = \frac{1}{1 - e^{-e^{-y}}}$$

is the arithmetic mean of the intervals between two years, which have a flood discharge greater than x; the discharges for the intervening years are all less than x. Therefore $T(x)$ is the mean of the number of years for which x will be surpassed once. Formula (39) gives the meaning of u from the standpoint of the return period. For $y = 0$

$$T(u) = \frac{e}{e - 1}.$$

The return period $T(u)$ of the most probable annual flood is 1.58198 years. In other words, the constant u is the flood discharge with return period

$$\log T(u) = 0.19920 \tag{40}$$

where log signifies the common logarithm. The return period of the mean annual flood is by (27) and (39) equal to 2.32762 years.

Let us now consider the relation between the flood discharge x and its return period for small and large values of x. To small values of x correspond large negative values of y and therefore return periods T approximating 1. The distribution (25) of the largest values being unlimited, the flood discharge considered as a function of log T will by (6) increase rapidly at first. To large values of x correspond large values of y and $T(x)$. If we introduce the natural logarithm, (39) gives

$$-\ln\left(1 - \frac{1}{T(x)}\right) = e^{-y}.$$

For large values of x, viz., $T(x) \geqq 10$, it is sufficiently accurate to use

$$\frac{1}{T(x)} = e^{-y},$$

so that

$$y = \ln T(x). \tag{41}$$

If the common logarithm is used,

$$\log T(x) = 0.434294\alpha(x - u). \tag{42}$$

The logarithm of the mean number of years for which the flood discharge will once be exceeded, converges towards a linear function of x. This property of the distribution of largest values was established by M. Coutagne [2]. Let us write

$$x = u + \frac{2.30258}{\alpha} \log T(x). \tag{43}$$

Then $1/\alpha$ can be considered as a measure of the increase of a flood discharge with respect to the logarithm of time.

According to the general formulas (6) and (42) the shape of the return period as a function of the flood discharge x is as follows: at the beginning i.e., for small flood discharge, the return periods are close to 1 and increase very slowly. At the end, i.e., for large flood discharges, the logarithm of the return period converges to a linear function of x.

Another form of (43) is

$$\frac{x}{u} = 1 + \frac{2.30258}{\alpha u} \log T(x). \tag{44}$$

The ratio of the flood discharge which will be exceeded in the mean once in T years to the modal annual flood converges to a linear function of the logarithm of the return period. The constant $1/\alpha u$ of dimension zero depends, by (33), on the coefficient of variation. Its value is a characteristic of the stream. If we introduce the arithmetic mean $\bar{u}$ and the standard deviation s we obtain by (42), (27), and (28)

$$x = \bar{u} - 0.45005s + (0.77970)\,(2.30258)s \log T(x).$$

Therefore, approximately,

$$\frac{x}{\bar{u}} = 1 - \frac{9}{20}v + 1.796v \log T(x). \tag{45}$$

The right hand member of this linear equation contains only one constant, the coefficient of variation of the floods. Finally by (42) and (31)

$$\log T(x) = 0.25068 + 0.55700\frac{x - \bar{u}}{s}. \tag{46}$$

There is still another way of interpreting these asymptotic formulas. Let $T(2x)$ be the return period of the value $2x$, then by (43)

$$2x = u + \frac{\ln T(2x)}{\alpha},$$

therefore

$$2 = \frac{\alpha u + \ln T(2x)}{\alpha u + \ln T(x)},$$

and finally

$$T(2x) = T^2(x)e^{\alpha u}. \tag{47}$$

The return period of a flood of magnitude $2x$ is equal to the square of the return period of x multiplied by a factor which depends only upon the coefficient of variation.

All these asymptotic formulas are good approximations only for return periods above ten years, which means according to Table II, $y \geqq 2.25$ or according to (23), (30) and (31) $x \geqq \bar{u} + 1.3s$. The corresponding value of the flood probability is by (3) $\mathfrak{W}(x) \geqq 0.9$. The consequences of (41) can be applied to only 10% of the observations, i.e. to the large flood discharges. Their observed return periods are based on a few observations and may therefore differ considerably from the theoretical values. In spite of the above restrictions the linear formula (43) has a meaning for values of T equal to or greater than unity. We now ask: How will the most probable largest value increase with the number of observations? This number of years can again be called T. The answer to the above question requires the solution of (13′) where the distribution (25) of largest values $\mathfrak{w}(y)$ must be introduced as the initial distribution $w(x)$.
From (24)

$$\frac{T-1}{e^{-e^{-y}}}\,e^{-y-e^{-y}} - 1 + e^{-y} = 0,$$

or

$$Te^{-y} = 1,$$

which is identical with (41). For $T = 1$ the most probable annual flood is of course u. Therefore the relation (41), valid for $T \geqq 1$, means: The most probable flood $u(T)$ to be reached within T years is a linear function of the logarithm of T

$$u(T) = u + \frac{2.30258 \log T}{\alpha}. \tag{41'}$$

The constant $1/\alpha$ is the slope of this straight line. The results (41–46) are related to Fuller's well-known formula [6]. This author, the first to investigate flood flows systematically, proposed a linear relation between the logarithm of the return period and the arithmetic mean of the flood discharges greater than the mth value (m taken from above). A similar empirical formula has been stated by Lane [7] and has been applied by Saville [19]. The similarities and differences between these interpolation formulas and our theory can be stated in the following way: If we start from the theory of largest values we reach these formulas as asymptotic expressions for the return period of large floods. Considered this way, our theory gives a certain justification to Fuller's hypothesis. But Fuller's and similar formulas were intended to apply to all flood discharges. Now, the distribution of the flood discharges (4) corresponding to these return periods does not fit the observations. It can be shown that these formulas involve the assumption of a simple exponential distribution $\varphi(x)$ for the flood discharges

$$\varphi(x) = \frac{1}{\bar{u} - \epsilon} e^{-(x-\epsilon)/(\bar{u}-\epsilon)}; \tag{48}$$

and the existence of a lower limit ϵ of the flood discharges given by $\epsilon = \bar{u} - s$. In Fuller's formula all flood discharges must be greater than 2/3 of the mean annual flood. The density of probability always diminishes with increasing magnitude of the flood. This neglects the ascending branch (about one third) of the distribution of floods (see Fig. 3) and is incompatible with the observed facts. We therefore prefer our formula which takes account of the total variation, but we do not minimize the importance of Fuller's work which has led to much valuable research.

Formula (39) gives the theoretical return periods $T(x)$ as a function of the reduced flood discharge y, and holds for the entire range of observations. The general numerical values are given in Table II, cols. 1 and 3. For a given stream, the return period of a flood discharge greater than x depends by (23) upon the two constants α and u. If these values have been calculated by (37) and (38) the theoretical flood discharge x corresponding to $T(x)$ is obtained by the linear transformation

$$x = u + y/\alpha. \tag{49}$$

The asymptotic formula (42) suggests the coordination of the flood discharges x and the logarithm of the return periods.

4. Rhône and Mississippi Rivers. We think that our system of formulas is simple, logically consistent and free of artificial assumptions. Now it remains to be shown that the arithmetic involved is simple and that the results fit the observations. For the Rhône we shall analyze the observed cumulative frequency, the distribution, and the return periods. For the Mississippi River we shall limit ourselves to the return periods.

For each year we choose the maximum of the daily discharges (we do not use momentary peaks). The 111 values x_m for the Rhône 1826–1936 published by Coutagne [3] and arranged in order of increasing magnitude are given in Table III (col. 1). The supposition that the intervals between consecutive floods are all equal to one year is not always true. Only 77 of the 111 floods occurred between October and March, whereas 34 were scattered throughout the year. But the differences in the lengths of the intervals compensate each other. The second column of Table III contains the serial number m. According to (9) we calculate for the mth observed flood discharge x_m, taken in ascending magnitude, the logarithm of the observed return period $\log n/(n - m)$ (col. 3), where $n = 111$ and $m = 1, 2, \cdots, 110$, and obtain the exceedance intervals. The other observed curve, the recurrence interval, is obtained by (10) through the coordination of x_{m+1} and $\log n/(n - m)$. Both curves are plotted in Fig. 1. The recurrence and exceedance intervals differ for the large flood discharges. The observed flood discharges arranged in increasing magnitude are plotted in the cumulative histogram, Fig. 2.

TABLE IV

Calculation of constants

Stream observation station	Rhône Lyon (France) 1826–1936	Mississippi River Vicksburg (Miss.) 1890–1939
Number of observations n	111	50
Annual mean flood $\bar{u}$	2,493.5	1,355.6
Mean squared flood $\overline{u^2}$	6,707,555.0	1,951,828.8
Standard deviation s	703.1	341.3
Constant $1/\alpha$	548.2	266.1
Most probable annual flood u	2,177.0	1,201.9

To compare these observations with our theory, we calculate the two constants $1/\alpha$ and u according to the formulas (34)–(38). The values Σx_m and Σx_m^2 are given at the end of Table III. Division by $n = 111$ gives the mean flood $\bar{u}$ and the mean squared flood $\overline{u^2}$ (Table IV). The Gaussian correction being $1 + 1/110$ we obtain from formula (36) the standard deviation s (Table IV)

TABLE V

Observed and theoretical distributions of flood discharges

Rhône

Reduced variable y	Variable x	Midpoints $x+\frac{\Delta x}{2}$	Observed distribution $111\Delta'\mathfrak{W}(x)$	Theoretical distribution $111\Delta\mathfrak{W}(x)$	Cumulative frequency $111\mathfrak{W}(x)$
−2.75	670				
−2.50		807	1		0.00
−2.25	944			0.01	0.01
−2.00		1081	1	0.34	0.07
−1.75	1218			1.19	0.35
−1.50		1355	7	3.03	1.26
−1.25	1492			6.07	3.38
−1.00		1629	5	9.98	7.33
−0.75	1766			14.02	13.36
−0.50		1903	13	17.38	21.35
−0.25	2040			19.49	30.74
0.00		2177	21	20.21	40.84
0.25	2314			19.68	50.95
0.50		2451	19	18.26	60.52
0.75	2588			16.31	69.21
1.00		2725	14	14.14	76.83
1.25	2862			11.97	83.35
1.50		2999	9	9.94	88.80
1.75	3136			8.15	93.29
2.00		3273	8	6.61	96.95
2.25	3410			5.30	99.90
2.50		3547	6	4.23	102.25
2.75	3686			3.45	104.13
3.00		3822	4	2.65	105.70
3.25	3959			2.00	106.78
3.50		4096	2	1.64	107.70
3.75	4233			1.28	108.42
4.00		4370	1	1.01	108.98
4.25	4507			0.79	109.43
4.50		4644	0	0.61	109.77
4.75	4781			0.48	110.04
5.00		4918		0.38	110.25
5.25	5055			0.30	110.42
5.50		5192		0.23	110.55
5.75	5329			0.18	110.65
6.00		5466		0.27	110.73
			111	111.00	

and finally from (37) and (38) the constant $1/\alpha$ and the most probable annual flood u. From the numerical values in Table IV the linear transformation (49) for the Rhône is

$$x = 2177.03 + 548.19y.$$

TABLE VI

Observed return periods

Mississippi River, Vicksburg, (Miss.) (1890–1939)

Flood discharge x_m	Serial number m	Return period $\log' T(x_m)$	Flood discharge x_m	Serial number m	Return period $\log' T(x_m)$
760	1	0.0088	1357	26	.3188
866	2	.0178	1457	27	.3273
870	3	.0269	1397	28	.3566
912	4	.0362	1397	29	.3768
923	5	.0458	1402	30	.3980
945	6	.0555	1406	31	.4202
990	7	.0655	1410	32	.4437
994	8	.0758	1410	33	.4686
1018	9	.0862	1426	34	.4949
1021	10	.0969	1453	35	.5229
1043	11	.1079	1475	36	.5529
1057	12	.1192	1480	37	.5851
1060	13	.1308	1516	38	.6198
1073	14	.1427	1516	39	.6576
1185	15	.1549	1536	40	.6990
1190	16	.1675	1578	41	.7448
1194	17	.1805	1681	42	.7959
1212	18	.1939	1721	43	.8539
1230	19	.2076	1813	44	.9208
1260	20	.2219	1822	45	1.0000
1285	21	.2366	1893	46	1.0969
1305	22	.2518	1893	47	1.2219
1332	23	.2676	2040	48	1.3980
1342	24	.2840	2056	49	1.6990
1353	25	.3011	2334	50	

$\Sigma x_m = 67{,}780. \quad \Sigma x_m^2 = 97{,}591{,}440.$

This leads to the determination of the theoretical flood discharges. The theoretical return periods $\log T(x)$ are given in Table II, col. 3 as a function of the reduced variable y and of x (col. 4). The discharges x obtained by letting y take on the values -2.75 to 6.00 in the linear transformation, are given in

Table V, cols. 2 and 3 and plotted in Fig. 1. The distances Δx used in the calculations of the theoretical discharges are $1/4\alpha = 137.05$.

Along the abscissa are plotted the logarithm of the return periods and the return periods in years; along the ordinate are plotted the corresponding flood discharges and the modal annual flood u. The straight line from the point $(u, 0)$ to the asymptote gives the most probable flood as a function of time. The theoretical curve corresponds quite closely with the general course of the observations. For small floods the theoretical return periods are practically iden-

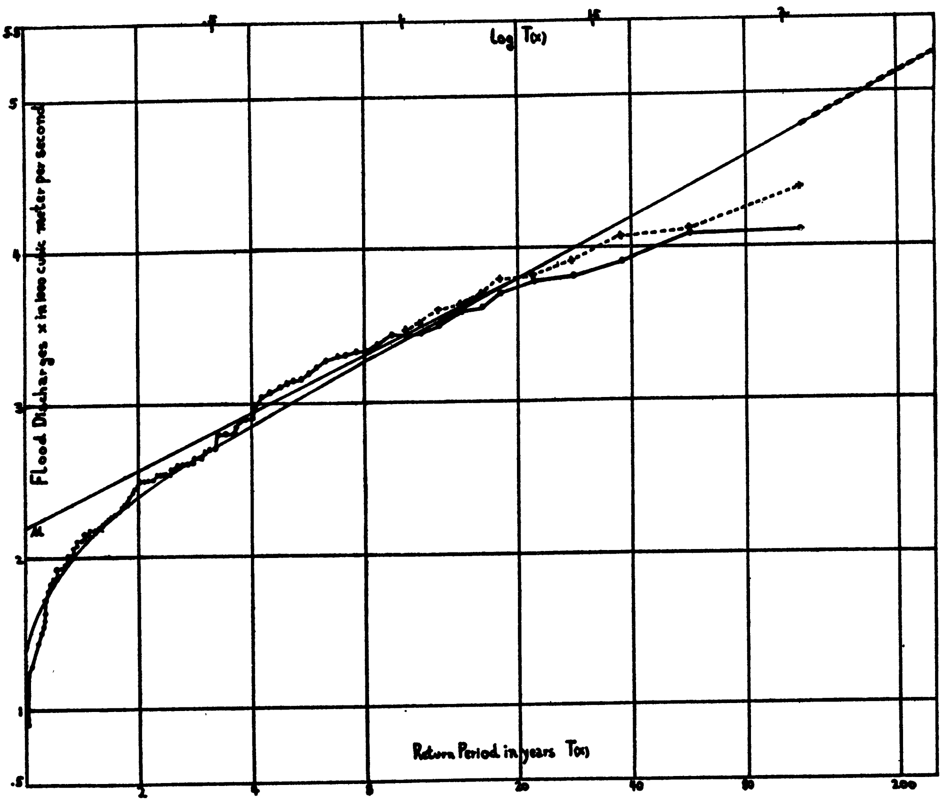

FIG. 1. RHÔNE AT LYON (FRANCE) 1826–1936

Observations Table III: Recurrence intervals, + — — +; Exceedance intervals, •——•; Return periods, ———; Theory Table II, cols. 3 and 4: Extrapolation, — —.

tical with the observed values. But for the very large floods the theoretical curve surpassed both the exceedance and recurrence intervals.

The observed cumulative histogram is shown in Fig. 2. We calculate from Table II, col. 2, the frequencies $111\mathfrak{W}(x)$ (Table V, col. 6). These theoretical values $(x, 111\mathfrak{W}(x))$ are also plotted in Fig. 2. The agreement between theory and observations is very good.

For the comparison of the observed and theoretical distributions of the flood discharges we use what might be called the natural classification. For the observations, the length of the class intervals and the beginning of the first class interval are arbitrary. In order to obtain the observed distribution of the flood discharges, it is natural to use the theoretical class intervals set forth in Table V, col. 2. The data of the third column can be interpreted as the midpoints of the class intervals given in col. 2. The frequencies for these class intervals are ob-

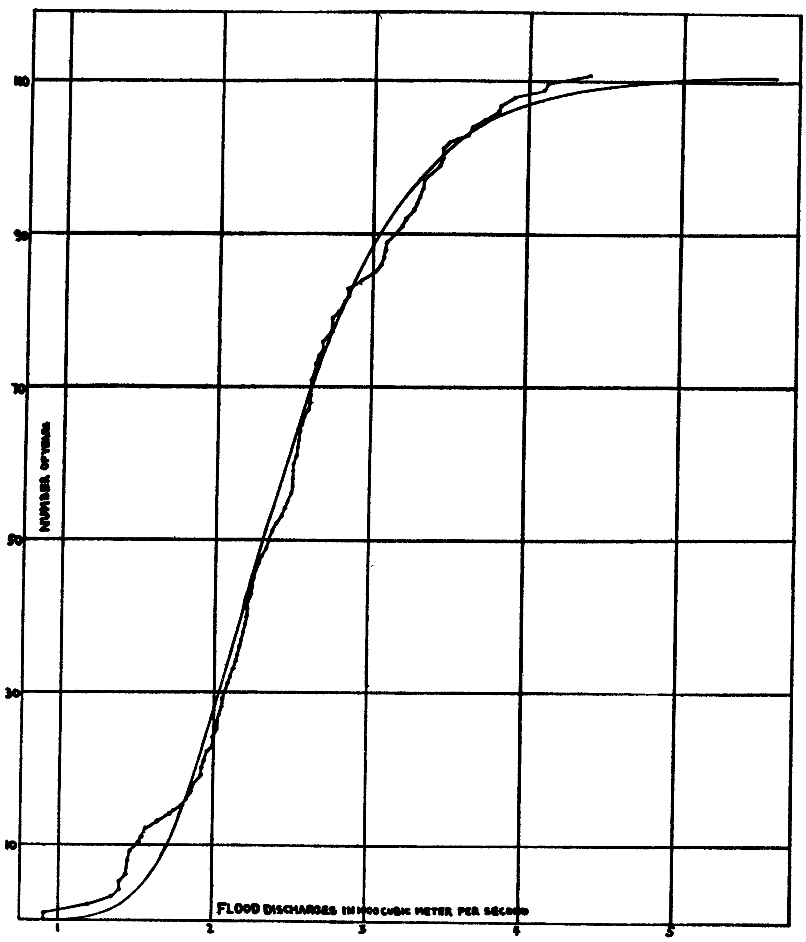

FIG. 2. CUMULATIVE FREQUENCY OF THE FLOOD DISCHARGES. RHÔNE, LYON (FRANCE) 1826–1936

Observations Table III cols. 1 and 2, •—•; Theory Table V cols. 2, 3 and 6, /

tained from Table III, and are given in Table V, col. 4. The observed distribution is shown in Fig. 3. To obtain the corresponding theoretical distribution we calculate from Table V, col. 6, the difference between two cumulative frequencies disjoined by one, i.e., we pair consecutively the first and third, the second and fourth items and so on. This theoretical distribution given in col. 5 and the observed distribution are based on class intervals of the same length. Fig. 3

shows that the theoretical distribution $\Delta\mathfrak{W}(x)$ of the largest values agrees in a satisfactory way with the observed distribution $\Delta'\mathfrak{W}(x)$ of the flood discharges. Table VI, col. 1, gives the corrected[2] flood discharges x_m, measured in units of 1000 cubic feet per second, for the Mississippi River at Vicksburg (1890–1939), ($n = 50$), arranged according to increasing magnitude; col. 2 gives the serial number m. We calculate the logarithm of the observed return periods log $n/(n - m)$, (col. 3). The observations (x_m, log $'T(x_m)$) and (x_{m+1}, log $'T(x_m)$) are plotted in Fig. 4. The constants obtained by formulas (34)–(38) are shown

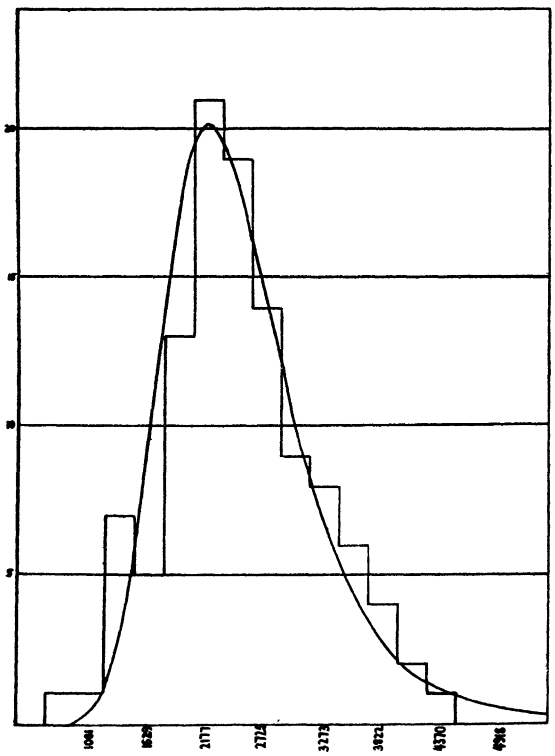

FIG. 3. DISTRIBUTION OF THE FLOOD DISCHARGES. RHÔNE, LYON (FRANCE) 1826–1936
Observations Table V cols. 2, 3 and 4, ⊓; Theory Table V cols. 2, 3 and 5, ⌒

in Table IV. By (49) the theoretical floods x corresponding to the return periods $T(x)$ presented in Table II, col. 3, are

$$x = 1201.98 + 266.14y.$$

These floods are given in Table II, col. 5. The class interval used is

$$1/4\alpha = 66.5.$$

[2] These data have been put at my disposal through the courtesy of Mr. A. E. Brandt of the U. S. Department of Agriculture.

The theoretical curve $(x, \log T(x))$, plotted in Fig. 4, agrees in a very satisfactory way with the observations. For the large floods the theoretical return periods are between the exceedance and recurrence intervals.

The calculations of the theoretical return periods for other streams, e.g. the Columbia, Connecticut, Cumberland, Rhine, and Tennessee Rivers, for which reliable observations exist for more than 60 years, also show a good agreement with the observations. The goodness of fit diminishes for streams for which the number of observations is smaller and for which the data are not very reliable.

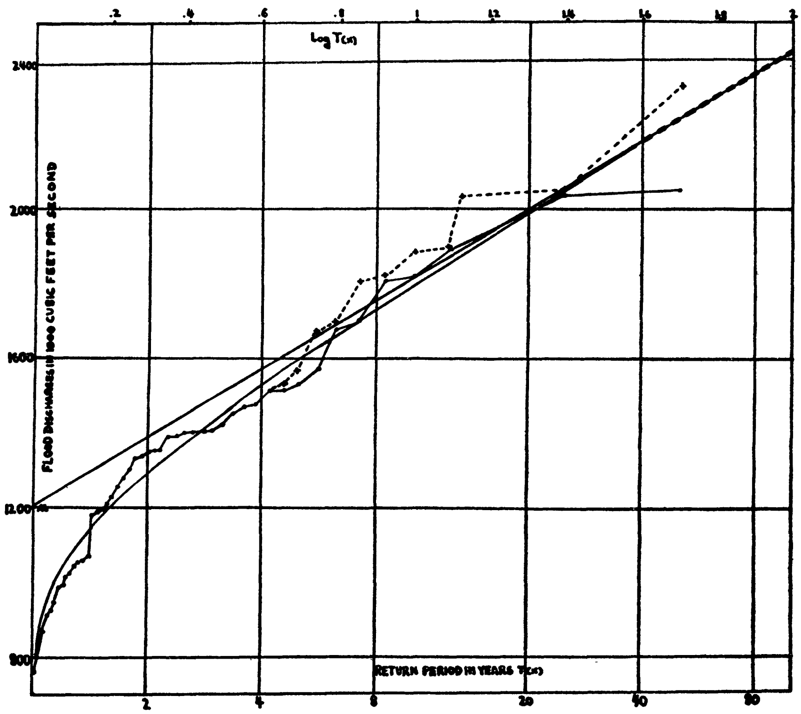

FIG. 4. MISSISSIPPI RIVER AT VICKSBURG, (MISS.) 1890–1939

Observations Table VI: Recurrence intervals, + — — +; Exceedance intervals, •——•; Return periods, ———; Theory Table II, cols. 3 and 5; Extrapolation, — —.

5. Summary and conclusions. In order to apply any theory we have to suppose that the data are homogeneous, i.e. that no systematical change of climate and no important change in the basin have occurred within the observation period and that no such changes will take place in the period for which extrapolations are made. It is only under these obvious conditions that forecasts can be made.

The theoretical return period $T(x)$, the mean number of years between two annual flood discharges greater than or equal to x, is a statistical function such as the distribution $w(x)$ or the probabilities $W(x)$ and $P(x)$. There are two sets of observed values corresponding to the theoretical set. The exceedance interval $'T(x_m)$ formula (9), and the recurrence interval $''T(x_m)$ formula (9'); x_m being the mth flood discharge, where m is counted from below. As any theory must include both notions, no separate theory for exceedance or recurrence intervals is possible.

The return period $T(x)$ of a flood discharge x is found by formula (39). For large values of x the flood discharge converges toward a linear function (42) of the logarithm of the return period. This is the scientific basis of Fuller's empirical formula. The two constants of our formula u and $1/\alpha$, are, respectively, the most probable annual flood discharge and a multiple of the standard deviation (28). Their values depend upon the drainage basin and known geological and meteorological factors. It is beyond our present task to consider the influence of these factors. Our method can be summarized by the following rules:

1) For each year find the maximum daily discharge x_m (do not use momentary peaks) and arrange these n data in increasing magnitudes.

2) Calculate for each discharge x_m ($m = 1, 2, \cdots, n - 1$), the values $\log {}'T(x_m) = \log n - \log (n - m)$ and plot the curves x_m, $\log n/(n - m)$, and x_{m+1}, $\log n/(n - m)$. These are the observed exceedance and recurrence intervals.

3) Calculate the annual mean flood $\bar{u}$ and the annual mean squared flood $\overline{u^2}$; determine according to (36)–(38) the standard deviation

$$s = \sqrt{\left(1 + \frac{1}{n-1}\right)(\overline{u^2} - \bar{u}^2)}$$

and the two constants

$$1/\alpha = 0.77970s,$$

$$u = \bar{u} - \frac{0.57722}{\alpha}.$$

4) The theoretical flood discharges x corresponding to the logarithm of the return period $T(x)$ given in Table II, col. 3, are obtained by the linear transformation

$$x = u + y/\alpha$$

where y is taken from Table II, col. 1. Plot x as a function of $\log T(x)$. For large values of x and for extrapolation it is sufficient to use the linear asymptote obtained graphically.

The linear part of the theoretical curve (x, $\log T$) permits of two interpretations: First, T is the theoretical return period of a flood greater than or equal to x; second, x is the most probable flood to be reached within T years. The second interpretation holds for the straight line through the point (u, 0).

The figures show a close agreement between observed and theoretical values.

The observed curvature of the return periods is brought out by the theoretical graph.

The agreement between theory and observation is excellent for floods which correspond to reduced values of $y \leqq 3$. For the two or three extreme floods, the return periods are based on a few observations and, consequently, the agreement is not very good. No theory can be verified by two or three observations. Generally speaking, the theory fits the observations as closely as could be expected for such a complicated phenomenon.

In order to make a further test of our results, we need a numerical measure for the weights to be given to the theoretical points. Therefore, for a given probability we must find the corresponding theoretical limits for the observed return periods. The theory of positional values will give these control curves. Since it was the purpose of this article to develop and make clear the basic method, we have refrained from introducing this subject.

It is our claim that the calculus of probabilities and especially the theory of largest values, is an efficient tool for the solution of certain hydrological problems.

REFERENCES

[1] G. E. Becker and C. E. van Orstrand, *Hyperbolic Functions*, Smithsonian Mathematical Tables, Washington, 1931.

[2] A. Coutagne, "Etude statistique des débits de crue," *Revue Générale de l'Hydraulique* Paris (1937).

[3] A. Coutagne, "Etude statistique et analytique des crues du Rhône à Lyon," *Comptes Rendus du Congrès pour l'Utilisation des Eaux*, Lyon, (1938).

[4] R. A. Fisher and L. H. C. Tippett, "Limiting forms of the frequency distribution of the smallest and the largest member of a sample," *Proc. Camb. Phil. Soc.*, Vol. 24 (1928).

[5] M. Fréchet, "Sur la loi de probabilité de l'écart maximum," *Annales Soc. Polon. Math.*, Vol. 6, (1927).

[6] Weston E. Fuller, "Flood flows," *Trans. Am. Soc. Civil Eng.*, Vol. 77 (1914).

[7] Weston E. Fuller, E. Lane and others, "Discussion on flood flow characteristics," *Trans. Am. Soc. Civil Eng.*, Vol. 89 (1926).

[8] John C. Geyer, "New curve fitting method for analysis of flood-records," *Trans. Am. Geophy. Union*, Part II (1940), pp. 660–668.

[9] Robert Gibrat, "Aménagement hydro-électrique des cours d'eau," "Statistique mathématique et calcul des probabilités, "*Revue Générale de l'Electricité*, Vol. 32, No. 15, 16, Paris (1932).

[10] Eugene L. Grant, "The probability-viewpoint in hydrology," *Trans. Am. Geophy. Union*, Part 1 (1940), pp. 7–12.

[11] E. J. Gumbel, "Les valeurs extrêmes des distributions statistiques," *Annales de l'Institut Henri Poincare*, Vol. 4 (1935), p. 115.

[12] E. J. Gumbel, "La plus grande valeur," *Aktuarske Vedy*, Vol. 5, No. 2, p. 83, No. 3, p. 133, No. 4, p. 146, Prague (1935–36).

[13] E. J. Gumbel, *La Durée Extrême de la Vie Humaine, Actualités Scientifiques et Industrielles*, Hermann et Cie, Paris, 1937.

[14] E. J. Gumbel, "Les intervalles extrêmes entre les émissions radioactives," *Jour. de Phys.*, Serie 7, Vol. 8, No. 8, No. 11 (1937).

[15] Allen Hazen, *Flood Flows, A Study of Frequencies and Magnitudes*, John Wiley and Sons, Inc., New York, 1930.

[16] ROBERT E. HORTON, "Hydrologic conditions as affecting the results of the application of methods of frequency analysis to flood records," *Geological Survey Water-Supply Paper* 771, Washington (1936).

[17] CLARENCE S. JARVIS, "Floods in the United States, Magnitude and Frequency," *Geological Survey Water Supply Paper* 771, Washington (1936).

[18] R. VON MISES, "La distribution de la plus grande de *n* valeurs," *Revue Math. de l'Union Interbalkanique,* Vol. 1, Athens (1936).

[19] T. SAVILLE, "A study of methods of estimating flood flows applied to the Tennessee River," *Publications from College of Engineering,* Nr. 6, New York (1935–36).

[20] J. J. SLADE, "The reliability of statistical methods in the determination of flood frequencies," *Geological Survey Water-Supply Paper* 771, Washington (1936).