Nordic Hydrology 3, 1972, 199–213
Published by Munksgaard, Copenhagen, Denmark

STATISTICAL INTERPRETATION OF HYDROMETEOROLOGICAL EXTREME VALUES

B. SAMUELSSON

United Nations Development Program,
Food and Agriculture Organization of the United Nations

The application of Jenkinson's method to extremal distributions for low probability annual extremes of rainfall and stream flow is studied and discussed.
A statistical method devised by Jenkinson has been examined and compared with other methods of fitting extreme value distributions to observed data. The Jenkinson method, being strictly objective, has the particular advantage of taking into account the extreme part of the extreme value distribution. The author shows, by applying the Jenkinson method to extreme values which significantly belong to several different kinds of frequency distributions, that this method could be applied as a standard one. Finally, the author indicates the possibility of using the Jenkinson method to extrapolate statistical characteristics from a series of statistically unstable short-term data.

The rapid introduction of machine processing of data in the field of hydrometeorology creates a demand for objective computation routines and for less subjective decisions, including, also, the question of how to deal with extreme events.

The aim of this study is, therefore, to try to find an objective standard procedure by using a statistical model based on the concepts of composite frequency distributions and excluding any "fitting by eye" or personal judgement.

What Causes Extreme Events

This is a question which should be asked before dealing with any kind of statistical evaluation of extreme values. In the field of hydrometeorology, there are two generally different causes of observed annual maxima or minima: one is the seasonal variation of the climate with a general circulation pattern, which is "normal"; the other consists of abnormalities in this pattern, such as extreme low or high index circulation, blocking, and unusual direction of the jetstream at certain locations.

It is strongly felt that the statistical treatment of hydrometeorological extreme events should start from a statistical evaluation of the variations of the general atmospheric circulation pattern. Little seems to have been done in this field so far, but there is a physical reason to believe that the rarer of the extreme values belong to a statistical distribution different from that which describes the occurrence of more common events.

A statistical method which combines the emphasising of the low probability part of the frequency distribution with an objective computation procedure has recently been devised by Jenkinson (1955), also in WMO Technical Note No. 98 (1969). As this seems to be the most realistic approach made so far to the statistical treatment of extreme values, a brief outline of the method will be given below. For a detailed description the reader is referred to the listed publications.

Extreme Value Distributions

The most commonly applied extreme value distributions are those introduced by Fréchet, Fisher and Tippett. They make a distinction between three different types of distribution: Type I, sometimes known as the Gumbel distribution, Type II, and Type III, which is known as the Weibull distribution. The Type I distribution is unlimited, whereas Type II and III distributions are limited in one direction, the direction of the less extreme values for Type II and the direction of the more extreme values for Type III.

As Jenkinson points out, one would not expect, on physical grounds, to find other than the Type III extremal distribution in nature, with the Type I as a limit.

Much hydrometeorological data, such as annual maximum values of daily rainfall and, usually, annual peak discharges, appear, however, to belong to the Type II distribution.

This leads Jenkinson to the conclusion that these data belong to different distributions: Type II for the low-valued annual maxima and Type III or Type

I as a limit for the high-valued annual maxima. Mathematically, it can also be shown that the lower part (37 %) of the data may not belong to the extreme value distribution as it is defined.

Extremal Distribution for Low Probability Annual Extremes

By studying a large number of sets of 5-year maxima for data whose annual maxima appear to belong to the Type II distribution, Jenkinson found that none of these sets of 5-year maxima showed any noticeable difference from the Type I distribution, whereas those sets of annual maxima which showed the Type III characteristics (extreme temperatures, wind, etc.) retained these characteristics when forming sets of 5-year maxima.

This is all right for long series of extreme values, but what about short-term data series?

Here Jenkinson indicates a procedure which is applicable to records covering any length of time down to, say, 8–10 years (as far as annual extremes are concerned).

Assuming that the actual order in which the annual extremes occur is random, a set of 5-year extremes can be simulated by taking all possible randomization of the data. The result is the assignment of frequencies to the annual extremes in such a way that the frequencies increase with decreasing probability.

As an illustration, let us take a series of observed annual extremes, x_1, $x_2, \ldots . x_n$, arranged in order of decreasing probability (increasing order of magnitude for annual maxima). From the first five x-values we get one 5-year extreme value which is equal to x_5. Next, we take the first six x-values and find that these can be combined to give a 5-year series in five different ways, each having the extreme value x_6. If we continue, we will find that for n values the number of combinations of five values, one of which is always x_n, the extreme of the n x-values, is:

$$F_n = \binom{n-1}{4} = \frac{(n-1)!}{4!\,(n-5)!} \tag{1}$$

We can now simulate a series of 5-year extreme values which will contain our original annual extreme values, except the first four, with a frequency equal to F_n assigned to each value. Consequently, the simulated series will contain $\sum_{5}^{n} F_i$ values.

Mathematical Dress for Extremal Distributions

If we consider the daily values of an n year series of a hydrometeorological element, let g_x be the average annual number of independent daily values greater than x. Then the probability that a daily value is less than x is $(1 - g_x/n)$. The probability (P(x)) that the annual maximum is less than x is then given by

$$P(x) = \left(1 - \frac{g_x}{n}\right)^n \approx e^{-g_x} \tag{2}$$

By introducing the so-called reduced variate, which is defined as

$$y \equiv -\log g_x \tag{3}$$

we get

$$P(x) \equiv e^{-e^{-y}} \tag{4}$$

Since by definition of return period an extreme value with return period T has a probability of 1/T of being exceeded in any one year

$$\frac{1}{T} \equiv 1 - P(x) \tag{5}$$

or

$$y \equiv -\log\log\frac{T}{T-1} \tag{6}$$

For $T > 10$ the relationship between T and y is closely given by

$$y = \log(T - 1/2) \tag{7}$$

In order to permit objective evaluation of extreme values with return periods exceeding the period of records, the relationship between x and P(x) or, according to (4), x and y, has to be expressed explicitly.

Jenkinson derived a general solution to the functional equation

$$P^z(x) \equiv P(a_z x + b_z) \tag{8}$$

where a_z and b_z are functions of the sample size, z. This general solution was obtained in the form

$$x = x_0 + \propto \frac{1 - e^{-ky}}{k} \tag{9}$$

x_0 is the value of x at $y \equiv 0$, $\propto$ is the slope of the x,y curve at $x \equiv x_0$, $y \equiv 0$, and k is a curvature parameter.

When $k = 0$ the expression (9) simplifies to the linear relation

$$x \equiv x_0 + \propto y \tag{10}$$

which is the mathematical expression for the Type I (Gumbel) distribution. The Type II and III distributions are pictured by expression (9) when k is negative and positive, respectively.

Fitting the Observed Data to the Extremal Distribution for Low Probability Annual Extremes

The most common method of curve fitting is the method of Least Squares. Although this method gives a good overall fit, it may not lead to a proper determination of the parameters of the theoretical distribution. If one is not deterred by further lengthy calculations by hand, or if one has access to a computer, the method of Maximum Likelihood may be chosen, by which the true parameter values can be approached more closely than by any other method.

For a full understanding of the method of Maximum Likelihood, it is necessary to make quite an extensive excursion into the field of mathematics. It may be sufficient here, however, to give only those expressions which are needed for practical application and which apply to the Type I distribution.

The Likelihood for a given sample is equal to the product of the values of the probability density function

$$p(x) \equiv \frac{|\,dP(x)\,|}{dx} \qquad (11)$$

for the actual values $x_1, x_2, \ldots x_n$. If we indicate by L the logarithm of the Likelihood, then

$$L \equiv \Sigma \log p(x) \equiv \Sigma \log \frac{|\,dP(x)\,|}{dx} \qquad (12)$$

By combining (4), 10) and (12) we get

$$-L \equiv n \log |\propto| + \Sigma y + \Sigma e^{-y} \qquad (13)$$

where

$$y \equiv \frac{x - x_0}{\propto} \qquad (14)$$

The Maximum Likelihood estimates for the statistical parameters x_0 and $\propto$ are those which maximize L, i. e. minimize $-$ L. At these values of x_0 and $\propto$ we have

$$\frac{-\delta L}{\delta x_0} = 0; \frac{-\delta L}{\delta \propto} = 0 \qquad (15)$$

From (13) and (14)

$$\frac{-\delta L}{\delta x_0} = -\frac{P}{\propto} \qquad (16)$$

$$\frac{-\delta L}{\delta \propto} = \frac{R}{\propto} \qquad (17)$$

where

$$P = n - \Sigma e^{-y} \qquad (18)$$

$$R = n - \Sigma y + \Sigma ye^{-y} \qquad (19)$$

Now, if the solutions of (15), (16) and (17) are

$$x_0'' = x_0 + x_0'$$

$$\propto \equiv \propto + \propto'$$

where x_0, $\propto$ are our initial estimates, and x_0', $\propto'$ are differences from the Maximum Likelihood, then we can, after substitution of x_0 by $x_0 - x_0'$ and of $\propto$ by $\propto - \propto'$, expand, $\frac{-\delta L}{\delta x}$ and $\frac{-\delta L}{\delta \propto}$ in a Taylor series, which eventually, combined with (15), will give us the tool for making successive approximations:

$$x_0' = \frac{\propto}{n}(1.11P - 0.26R) \qquad (20)$$

$$\propto' = \frac{\propto}{n}(0.20P - 0.61R) \qquad (21)$$

Practical Application

The following procedure is suggested:

(a) From the series of annual extreme values, $x_1, x_2, \ldots . x_n$, take a value near $x_{0.75n}$ and use this as a first estimate of x_0.

(b) Compute the standard deviation of the original extreme value series and use this as a first estimate of $\propto$.

(c) For each x-value compute $y = (x-x_0)/\propto$, e^{-y} and ye^{-y}.

(d) Form the sums Σy, Σe^{-y} and Σye^{-y}, but do not forget that each value of y, e^{-y} and ye^{-y} has to be counted F_n times.

(e) Form the sum ΣF_n.

(f) Compute x_0' and $\propto'$ from (2) and (21) where $n = \Sigma F_n$.

(g) Use $(x_0 + x_0')$ and $(\propto + \propto')$ as new estimates for x_0 and $\propto$ and repeat points (a) to (c).

(The calculating task may appear overwhelming, but fortunately the process is rapidly converging towards the Maximum Likelihood solution, so that in order to obtain the true parameter values within a reasonable margin of error, say, 1 %, it will be sufficient, in most cases, to repeat the first calculation twice.)

(h) Keeping in mind that the procedure so far has been dealing with 5-year extreme values, the parameters have to be adjusted to 1-year extreme values according to:

$$x_0(1\text{–year}) = x_0(5\text{–year}) - \propto_e \log 5$$

$$\propto(1\text{–year}) = \propto(5\text{–year})$$

(i) Compute x for any desired return period T, using the expression

$$x = x_0 - \propto \log \log \frac{T}{T-1}$$

Confidence Limits

The derived frequency equation can be considered only to represent the mean of the data at a given return period. Assuming that the data around the mean are normally distributed, it is possible to calculate for a given return period the limits of error for a certain probability. Usually, the probability of 68 % is chosen which corresponds to a deviation of plus or minus one standard deviation, S_x. As the mathematical procedure is cumbersome and involved, we give here only the final expression for the variance:

$$S_x{}^2 = \frac{\propto^2}{n}(0.61^2 + 2 \times 0.26y + 1.11^2y^2) \tag{22}$$

Testing the Method

As a first step, the Jenkinson procedure was applied to a series of annual maximum 24-hour rainfalls observed at three stations, Akoumia, Moni Asomaton and Gortys in Crete, Greece. Table 1 gives the values in chronological order, together with their mean M, standard deviation S, and coefficient of variation C_v for each station.

The 5, 20, 100 and 1,000 years' values were computed and listed in Tables 2 and 3, which also include corresponding values obtained by other methods.

As can be seen from Tables 2 and 3, there is close agreement between the maximum rainfall figures obtained by the Jenkinson method and those obtained assuming lognormal distribution, whereas the other methods tested give more or less deviating results.

Table 1.
Observed annual maximum 24-hour rainfall (mm)

Year	Akoumia	Moni Asomaton	Gortys
1938/39	–	–	55
39/40	–	–	108
40/41	–	–	118
41/42	–	–	39
42/43	–	–	49
43/44	–	–	42
44/45	–	–	47
45/46	–	–	67
46/47	–	–	–
47/48	–	–	95
48/49	–	–	47
49/50	–	–	55
50/51	–	–	39
51/52	–	–	32
52/53	–	–	41
53/54	–	–	45
54/55	82	170	96
55/56	49	68	34
56/57	94	161	32
57/58	84	91	33
58/59	59	126	60
59/60	183	163	46
60/61	100	120	44
61/62	109	92	97
62/63	98	82	98
63/64	56	79	55
64/65	79	127	66
65/66	71	83	35
66/67	50	75	70
67/68	112	109	46
68/69	–	–	27
Mean	876	110.4	57.3
S	34.6	34.9	25.3
C_v	0.40	0.32	0.44

Table 2.
Maximum 24-hour rainfalls (mm) at Gortys for indicated return periods computed by different methods

Return period (years)	Jenkinson (WMO, 1969)	Gumbel (WMO, 1969)	Brakensiek (Agric. Handbook, 1964)	Lognormal (Chow, 1964)	Modified lognormal (Agric. Handbook, 1964)
5	77 ± 7	70 ± 18	80	75	69
20	107 ± 13	108 ± 20	117	105	109
100	141 ± 20	172 ± 58	156	140	164
1,000	189 ± 30	323 ± 194	214	193	264

But which method gives the most reliable result?

In order to answer this question with reasonable confidence, we have to apply some kind of significance test. Keeping in mind, also, the meteorological background, which makes the more extreme values probably follow a distribution of their own, the test has to be applied to classes of values in the upper part of the records, when arranged with the data in decreasing order.

The last requirement, together with the need for at least 5 values in each class, makes the records of Table 1 seem a little too short. Instead, two sets of average monthly river flows, August and November, of River Limnitis in Cyprus, covering a period of 51 years, were selected and the χ^2-test was applied; see Table 4.

Table 3.
Maximum 24-hour rainfalls (mm) at Akoumia and Moni Asomaton computed by the Jenkinson method and derived from lognormal distribution

Return period (years)	Akoumia		Moni Asomaton	
	Jenkinson	Lognormal	Jenkinson	Lognormal
5	113 ± 13	112	143 ± 12	137
20	153 ± 26	153	177 ± 22	176
100	199 ± 40	199	215 ± 36	217
1,000	263 ± 59	268	269 ± 49	276

Table 4.

Significance tests by χ^2-method of different distributions (N = normal, Γ = gamma, J = Jenkinson) applied on 51-year records of average river flow in Limnitis River, Cyprus, during August and November, respectively. M = long-term mean flow, S = standard deviation

M m^3/s	S m^3/s	Interval	Observed no. of values	Theor. no. of values N	Theor. no. of values Γ	Theor. no. of values J	χ^2 N	χ^2 Γ	χ^2 J	Significance level % N	Significance level % Γ	Significance level % J
August												
117.0	57.3	$x > M+S$	10	8.2		8.2						
		$M+S \geq x > M+0.5S$	6	7.7		8.2						
		$M+0.5S \geq x > M$	10	9.7		13.3						
		$M \geq x > M-0.5S$	9	9.7		14.3						
		$M-0.5S \geq x > M-S$	7	7.7		6.6						
		$M-S \geq x$	9	8.2		0.5	1.0		13.4	80		1
		$x > 198.7$	5	4.1		4.1						
		$198.7 \geq x > 172.4$	5	4.6		4.1						
		$172.4 \geq x > 148.6$	5	6.1		7.1						
		$148.6 \geq x > 133.9$	5	4.6		6.1						
		$133.9 \geq x > 119.3$	5	5.1		7.1						
		$119.3 \geq x$	26	26.5		22.4	0.5		2.8	90		44

November									
18.8	25.8	x > 37.95	8	9.7	9.2				
		37.95 ≧ x > 19.1	9	7.1	8.7				
		19.1 ≧ x > 7.9	8	10.7	7.1				
		7.9 ≧ x > 2.95	8	8.7	4.1				
		2.95 ≧ x > 1.75	9	3.6	0.5				
		1.75 ≧ x	9	11.2	21.4	5.5	27.2	14	0
		x > 50.45	5	7.7	5.6				
		50.45 ≧ x > 36.25	5	2.6	4.1				
		36.25 ≧ x > 21.4	5	5.6	6.1				
		21.4 ≧ x > 13.65	5	5.1	5.1				
		13.65 ≧ x > 7.9	5	6.1	3.6				
		7.9 ≧ x	26	24.0	26.0	3.1	0.9	38	83

We have here extended our extreme-value distribution concept to average monthly river flows. This might seem too bold a measure, but as we particularly want to study the high-value observations to find out whether they form a frequency distribution different from the lower part of the values, we are still concerned with a kind of extreme-value distribution.

The general distributions of the two series of river flow were previously found to be normal and gamma, respectively (Samuelsson 1968). This is shown in Table 4 in the first of the two sets of classes for the two data series, where the number of values is approximately equally distributed among the classes. As expected, in this case, the Jenkinson distribution does not fit at all.

If, however, we divide the number of values in such a way that the lower half of the values is put together into one class and the remainder divided into 5 classes, the Jenkinson distribution becomes highly significant (a significance level of $\geqq 5\ \%$ is usually regarded as sufficient to prove a fit to the assumed theoretical distribution).

Particularly interesting is the applicability to data which are normally distributed. The very close fit to normal distribution which appears for the August data in Table 4 (80–90 % significance level) does not exclude the Jenkinson method from being a useful statistical standard procedure even in this case.

In Table 5 the average monthly river flows for various return periods have been computed from the normal distribution, as well as by the Jenkinson method, and although the latter values come out higher, the difference is significant (greater than the Jenkinson confidence limits) only for return periods of 1,000 years or more.

Returning to the question which of the methods used in Tables 2 and 3 gives the most reliable result, we can feel reasonably confident in giving preference

Table 5.
Average river flows of Limnitis River in Cyprus during August for indicated return periods computed from normal distribution and by the Jenkinson method

Return period (years)	Normal distribution	Jenkinson method
5	165	165 ± 10
20	211	216 ± 17
100	251	274 ± 26
1,000	294	354 ± 39

Table 6.
Values of parameters x and $\propto$ of the frequency equation for indicated periods; annual maximum of 24-hour rainfall at Gortys

Period	x_0	$\propto$
1938/39–68/69	46.0	20.7
1949/50–68/69	43.0	18.8
1954/55–68/69	51.9	17.6
1957/58–68/69	53.4	16.2
1959/60–68/69	58.5	15.6
1962/63–68/69	75.6	10.4

to the Jenkinson method. Furthermore, it seems that this method could be applied to any extreme-value distribution and, possibly, to other distributions of hydrometeorological data as far as the extreme part of the data is concerned.

Could the Jenkinson Method Be Used as a Tool for Extrapolating Statistical Characteristics from Short-Term Records?

From the annual maximum 24-hour rainfalls at Gortys, five different periods were selected, the last 20–, 15–, 12–, 10– and 7–year periods, respectively. To each period the Jenkinson method was applied and the parameters x_0 and $\propto$ of the frequency equation (10) were evaluated. The result of these calculations is summarized in Table 6, which lists the parameter values for each period.

As a comparison, the maximum spring discharges of the river Don at Listri

Table 7.
Values of parameters x and $\propto$ of the frequency equation for indicated periods; maximum spring discharges of the river Don at Liski

Period	x_0	$\propto$
28 years	3,000	1,910
last 20 years	3,780	1,260
last 15 years	4,100	1,220
last 12 years	4,000	1,360
last 10 years	3,790	1,370
last 7 years	5,160	1,040

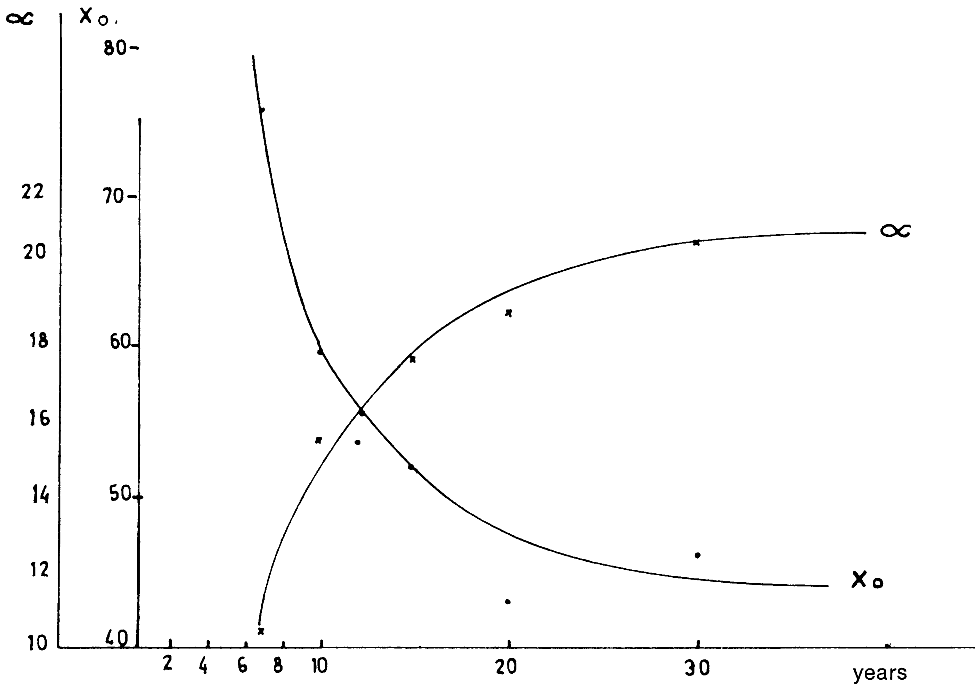

Fig. 1.
Variation of the frequency parameters x_0 and ∞ with length of records, maximum 24-hour rainfall at Gortys.

(see Table 4, Chapter 6, WMO Technical Note No. 98) were selected for a similar study, the result of which is summarised in Table 7.

The values have been plotted against length of period in Figures 1 and 2 and there appears to be a rather well defined decrease of x_0 and a corresponding increase of ∝ with increasing length of record. This could be interpreted to mean that for short-term records the number of values caused by unusual atmospheric circulation conditions are too few, if any, to affect the extreme-value distribution. As the period of records increases, more and more of these unusual events will be included until a statistical balance has possibly been reached between values representing normal atmospheric circulation conditions and values representing abnormalities of these conditions.

No general conclusion could be drawn from these two cases, but if for some hydrometeorological properties the variation of x_0 and ∞ with the length of records could be shown to follow a pattern similar to the one in Figure 1, where a point of statistical balance could possibly have been reached, and if

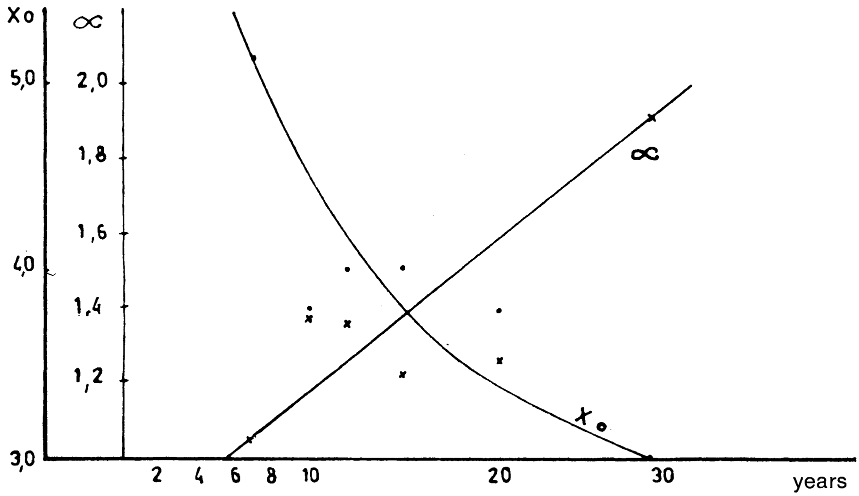

Fig. 2.
Variation of the frequency parameters x_0 and ∞ with length of records, maximum spring discharges of the Don at Liski (from Table 4, Chapter 6, WMO, Technical Note No. 98).

this pattern repeats over a certain area, then there would be a possibility, when dealing with short-term records, to use the Jenkinson method to extrapolate their statistical behaviour.

REFERENCES

Samuelsson, B. (1968) Statistical characteristics of some monthly river flows in Cyprus. Technical Note, FAO/UNDP CYP/6 Project.

Chow, V. T. The log-probability law and its engineering applications. *Amer. Soc. Civil Engr. Proc. 80,* Separate 536.

Chow, V. T. (1964) *Handbook of Applied Hydrology.* McGraw Hill.

Jenkinson, A. F. (1955) The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quart. J. roy. meteorol. Soc. 87,* 158.

US Dept. of Agriculture (1964) Simplified Techniques for Fitting Frequency Distributions to Hydrologic Data. Agric. Handbook No. 259, Washington, DC.

WMO Tech. Note No. 98 (1969) Estimation of Maximum Floods. Geneva.

Address

Bengt Samuelsson, Hydrologist,
UNDP-FAO, Casier ONU,
Rabat-Chellah, Maroc.

Received 1 July 1972.