THE STATISTICAL TREATMENT OF FLOOD PEAKS

by

J. Reimann

Department of Civil Engineering Mathematics, Technical University, Budapest

Received: September 15, 1979

1. Theoretical probability distribution for flood peaks

According to methods by Todorović [1] and Zelenhasić [2], first of all, probability distributions of the number and magnitude of peaks x_1, x_2, $\ldots$, x_ν exceeding a given base level c in a hydrograph related to a section of a river within a determined time interval $(0, t)$ are considered.

For the time interval $(0, t)$ a quarter of a year was chosen (e.g. first three months in every year). Namely this investigation always referred to the same quarter of a year since experience showed the water regime of rivers in the *Carpathian* basin (e.g. *Danube* and *Tisza*) to be homogeneous in spring (or summer) of every year.

Flood peaks above a certain level c — base level — are dealt with. The elevation of level c has been selected high enough, the number of flood peak exceedances occurring in a season is relatively small.

The exceedances $X_1, X_2, \ldots, X_\nu$ are continuous, independent, identically distributed random variables. The number of exceedances ν in itself is a random variable. The durations of exceedances (up to the time when the water level returns to below the level c) are Y_1, Y_2, $\ldots$, Y_ν. These are also independent random variables. (In the subsequent examples this fact will be proven statistically.)

The distribution of the random variable ν is generally an inhomogeneous *Poisson*'s distribution, as proven by Zelenhasić.

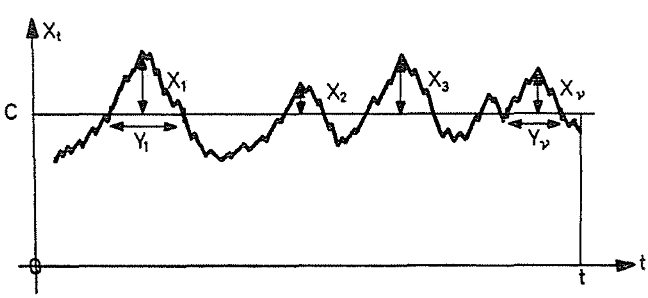

Fig. 1

According to the present investigations, choosing the interval $[0, t)$ the first three months (or the second quarter) of every year, the distribution of ν can be approximated by a homogeneous *Poisson*'s distribution:

$$P(\nu = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \qquad (k = 0, 1, 2, \dots)$$

where λt is the average number of exceedances in the time interval $(0, t)$.

Many authors have found that the probability distribution of exceedances $X_1, X_2, \dots, X_\nu$ generally follows the gamma distribution, or in special cases the exponential distribution.

In the case of the river *Tisza* the distribution of the exceedances was found to be the exponential distribution

$$P(X_i < x) = H(x) = 1 - e^{-\beta x}.$$

Let

$$Z(t) = \sup_{1 \le i \le \nu} \{X_1, X_2, \dots, X_\nu\} \tag{1.0}$$

be the greatest exceedance occurring in the time interval $[0, t)$. The base level c appoints the water level values, therefore the set $X_1, X_2, \dots, X_\nu$ of exceedances is a sample of random size.

Let the corresponding ordered sample of random size be

$$X_1^* < X_2^* < \dots < X_\nu^*.$$

It is easy to see that if $H(x)$ is the common cumulative distribution function of the random variables X_i then

$$P(X_\nu^* < x) = [H(x)]^\nu.$$

Consequently

$$P[Z(t) < x \mid \nu = k] = [H(x)]^k.$$

The distribution of the yearly greatest exceedance is only meaningful if there is any exceedance. Let $F_t(x)$ be the conditional cumulative distribution function of the random variable $Z(t)$ [in (1.0)] under the condition that there is at least one exceedance above the base level c:

$$F_t(x) = P[Z(t) < x \mid \nu > 0) = \frac{\sum_{k=1}^{\infty} [H(x)]^k P(\nu = k)}{P(\nu > 0)}. \tag{1.1}$$

If

$$H(x) = 1 - e^{-\beta x} \qquad \text{and} \qquad P(\nu = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

then

$$F(x) = \frac{e^{-\lambda t e^{-\beta x}} - e^{-\lambda t}}{1 - e^{-\lambda t}}. \tag{1.2}$$

In case of the river *Tisza*, this conditional distribution function is seen to fit very well the empirical cumulative distribution function of the observed largest exceedances.)

It is well known that if the common probability distribution of the random variables $X_1, X_2, \ldots, X_\nu$ is the exponential distribution with the c.d.f. $H(x) = 1 - e^{-\beta x}$ then the random variables

$$\delta_1 = \nu X_1^*, \quad \delta_2 = (\nu - 1)(X_2^* - X_1^*), \ldots, \delta_k = (\nu - kH)(X_k^* - X_{k-1H}^*) \quad (1.3)$$

are independent and of exponential distribution with mean $1/\beta$ and variance $1/\beta^2$ since

$$X_k^* = \frac{\delta_1}{\nu} + \frac{\delta_2}{\nu - 1} + \ldots + \frac{\delta_k}{\nu - k + 1} \quad (1.4)$$

it follows that

$$E(X_k^*) = \frac{1}{\beta}\left(\frac{1}{\nu} + \frac{1}{\nu - 1} + \ldots + \frac{1}{\nu - k + 1}\right) \quad (1.5)$$

hence:

$$E(Z(t) \mid \nu = k) = \frac{1}{\beta} \sum_{i=1}^{k} \frac{1}{i} \quad (1.6)$$

$$D^2(Z(t) \mid \nu = k) = \frac{1}{\beta^2} \sum_{i=1}^{k} \frac{1}{i^2} \,. \quad (1.7)$$

Since the mean value of the conditional mean value equals the unconditional mean value,

$$E(Z(t)) = \frac{e^{-\lambda t}}{\beta} \sum_{k=1}^{\infty} \frac{(\lambda t)^k}{k!} \left(\sum_{i=1}^{k} \frac{1}{i}\right) . \quad (1.8)$$

It is easy to give a lower bound for the mean value $E[Z(t)]$. Considering that the mean value of the yearly greatest exceedances is not smaller than the mean value of all the exceedances, i.e., if the exceedances have the expenential distribution with mean $1/\beta$, then

$$E[Z(t)] \geq \frac{1}{\beta} \,. \quad (1.9)$$

In giving an upper bound for the mean value, remind that:

$$1 + \frac{1}{2} + \ldots + \frac{1}{k} - \ln(kH) \to p = \text{const}\ (p = 0.577 \ldots) . \quad (1.10)$$

In every case it is true that

$$1 + \frac{1}{2} + \ldots + \frac{1}{k} \leq \ln(kH) + 2$$

and

$$\ln(kH) \leq k \qquad (k = 0, 1, 2, \ldots)$$

therefore

$$E[Z(t)] \leq \frac{e^{-\lambda t}}{\beta}\left[\sum_{k=1}^{\infty}\frac{(\lambda t)^k}{k!}k + 2\sum_{k=1}^{\infty}\frac{(\lambda t)^k}{k!}\right] \leq \frac{\lambda t + 2(1-e^{-\lambda t})}{\beta}. \qquad (1.11)$$

The calculation of the variance of the largest exceedance $Z(t)$ is quite complicated. On the contrary, it is easy to calculate the conditional variance of the random variable $Z(t)$ provided the number of exceedances is given, and the exceedances are exponentially distributed with mean $1/\beta$.
In this case:

$$D^2(Z(t) \mid \nu = k) = \frac{1}{\beta^2}\sum_{i=1}^{k}\frac{1}{i^2}. \qquad (1.12)$$

The unconditional variance of the largest exceedance $Z(t)$ of exponential distribution can be calculated from the relationship:

$$D^2[Z(t)] = E[D^2(Z(t) \mid \nu)] + D^2[E(Z(t) \mid \nu)] = \qquad (1.13)$$

$$= \frac{e^{-\lambda t}}{\beta^2}\left\{\sum_{k=0}^{\infty}\left[\sum_{i=1}^{k}\frac{1}{i^2} + \left(\sum_{i=1}^{k}\frac{1}{i^2} - e^{-\lambda t}\sum_{k=0}^{\infty}\frac{(\lambda t)^k}{k!}\sum_{i=1}^{k}\frac{1}{i^2}\right)^2\right]\frac{(\lambda t)^k}{k!}\right\}.$$

Generally, this complicated formula can be avoided, because in the above case an upper bound can be given for the variance:

$$D^2[Z(t)] \leq \frac{\pi^2}{6\beta^2} \qquad \text{i.e.} \qquad D[Z(t)] \leq \frac{\pi}{\beta\sqrt{6}}. \qquad (1.14)$$

(Note: in practice the level c — of importance for flood control — is chosen high enough to have only a few years where more than one exceedances occur and therefore

$$D[Z(t)] \approx \frac{1}{\beta}$$

is a good estimation of the variance of the random variable.)

2. Investigation of the stochastic connection between random variables

For deriving the probability distribution of the largest exceedance $Z(t)$, the exceedances X_i $(i = 1, 2, \ldots, \nu)$ were supposed to be independent and identically distributed random variables. This fact of course must be proven by statistical analyses. One of the possible methods for this study is the *Wald-Wolfowitz* test based on serial correlation. Calculating the arithmetical mean $\overline{X} = \frac{\sum_{i=1}^{\nu} X_i}{\nu}$ of the observed variables $X_1, X_2, \ldots, X_\nu$ and then forming the sequence:

$$X_i' = X_i - \overline{x} \qquad (i = 1, 2, \ldots, \nu)$$

lead to the following statistics:

$$R = \sum_{i=1}^{\nu-1} X_i' X_{i+1}' - X_\nu' X_1' . \tag{2.1}$$

It can be shown that

$$E(R) = \frac{S_2}{\nu - 1} \tag{2.2}$$

$$D^2(R) = \frac{S_2^2 - S_4}{\nu - 1} + \frac{S_2^2 - 2S_4}{(\nu-1)(\nu-2)} + \frac{S_2^2}{(\nu-1)^2} \tag{2.3}$$

where

$$S_r = \sum_{i=1}^{\nu} X_i'^r .$$

Wald and *Wolfowitz* have shown that the asymptotic distribution of the standard random variable

$$R^* = \frac{R - E(R)}{D(R)} \tag{2.4}$$

is the standard Gaussian distribution

$$P(R^* < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

and if $|R^*| > 2$, the hypothesis of independence cannot be accepted. The calculation of R^* is quite tedious. A more convenient method is suggested for the investigation of the stochastic connection between the random variables.

Let X and Y be continuous random variables with c.d.f. $F(x)$ and $G(y)$, respectively and with joint distribution function $H(x, y)$. The indicator variables $\tilde{X}(x)$ and $\tilde{Y}(y)$ are introduced in the following forms:

$$\tilde{X}(x) = \begin{cases} 1 & \text{if} \quad X < x \\ 0 & \text{if} \quad X \geq x \end{cases} \tag{2.5}$$

$$\tilde{Y}(y) = \begin{cases} 1 & \text{if} \quad Y < y \\ 0 & \text{if} \quad Y \geq y \end{cases} \tag{2.6}$$

for a given pair of data (x, y). The coefficient of correlation of variables $\tilde{X}$ and $\tilde{Y}$ is:

$$\varrho(\tilde{X}(x), \tilde{Y}(y)) = \frac{E[\tilde{X}(x) \cdot \tilde{Y}(y)] - E[\tilde{X}(x)]E[\tilde{Y}(y)]}{D[\tilde{X}(x)]D[\tilde{Y}(y)]} = \tag{2.7}$$

$$= \frac{H(x, y) - F(x)G(y)}{\sqrt{F(x)[1 - F(x)]G(y)[1 - G(y)]}} .$$

Denote the α-quantiles of the distribution of the variable X by $\tilde{x}_\alpha$, i.e. let x be the number for which

$$F(\tilde{x}_\alpha) = \alpha$$

and y_α for which

$$G(\tilde{y}_\alpha) = \alpha\,. \tag{2.8}$$

In this case

$$-1 \leq \varrho_\alpha = \frac{H(\tilde{x}_\alpha, \tilde{y}_\alpha) - \alpha^2}{\alpha(1-\alpha)} \leq 1\,.$$

For $\varrho_\alpha > 0$ a positive association exists between the random variables X and Y. In this case

$$\alpha^2 \leq H(\tilde{x}_\alpha, \tilde{y}_\alpha) \leq \alpha\,. \tag{2.9}$$

Evidently

$$H(\tilde{x}_\alpha, \tilde{y}_\alpha) = \varrho_\alpha \cdot \alpha + (1 - \varrho_\alpha)\alpha^2\,. \tag{2.10}$$

It is directly seen from Eq. (2.7) that if X and Y are independent then $\varrho_\alpha = 0$. If there is a monotonous functional dependence between X and Y; $Y = \varphi(X)$ then $\varrho_\alpha = 1$. Namely if $Y = \varphi(X)$ (where $y = \varphi(x)$ is a monotonous increasing function) then

$$\alpha = P(Y < \tilde{y}_\alpha) = P(\varphi(x) < \tilde{y}_\alpha) = P(X < \varphi^{-1}(\tilde{y}_\alpha)) = P(X < \tilde{x}_\alpha) \tag{2.11}$$

consequently $\quad \varphi^{-1}(\tilde{y}_\alpha) = \tilde{x}_\alpha\,; \quad$ i.e. $\quad \tilde{y}_\alpha = \varphi(\tilde{x}_\alpha)\,.$

As a consequence, if there is a monotonous functional dependence between X and Y then

$$H(\tilde{x}_\alpha, \tilde{y}_\alpha) = P(X < \tilde{x}_\alpha, Y < \tilde{y}_\alpha) = P(X < \tilde{x}_\alpha, \varphi(x) < \varphi(\tilde{x}_\alpha)) = P(X < \tilde{x}_\alpha) =$$
$$= F(\tilde{x}_\alpha) = \alpha\,.$$

Consequently, in this case

$$\varrho_\alpha = \frac{H(\tilde{x}_\alpha, \tilde{y}_\alpha) - \alpha^2}{\alpha - \alpha^2} = \frac{\alpha - \alpha^2}{\alpha - \alpha^2} = 1$$

(if $y = \varphi(x)$ is a monotonous decreasing function) then

$$\alpha = P(y < \tilde{y}_\alpha) = P(\varphi(x) < \tilde{y}_\alpha) = P(X > \varphi^{-1}(\tilde{y}_\alpha)) = 1 - P(X < \varphi^{-1}(\tilde{y}_\alpha))\,,$$

i.e.

$$P(X < \varphi^{-1}(\tilde{y}_\alpha)) = 1 - \alpha \qquad \text{therefore}$$
$$\varphi^{-1}(\tilde{y}_\alpha) = \tilde{x}_{1-\alpha} \qquad \text{i.e.} \qquad \varphi(\tilde{x}_{1-\alpha}) = \tilde{y}_\alpha\,.$$

If the joint distribution function of the random variables X and Y is not known, then the value of $H(x_\alpha, y_\alpha)$ is approximated by the relative frequency of the points in the lower left-hand side quadrant, defined by the point $(\tilde{x}_\alpha, \tilde{y}_\alpha)$. These points originate from the two-dimensional statistical sample $(X_1,$

$Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$. $\tilde{x}_\alpha$ and $\tilde{y}_\alpha$ are the α quantiles of the one-dimensional probability distribution of X and Y, respectively.

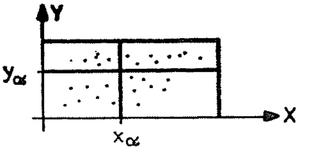

Fig. 2

If the number of points in the lower left-hand side quadrant is k, then

$$\varrho_\alpha \approx \frac{\frac{k}{n} - \alpha^2}{\alpha - \alpha^2}.$$

If X and Y are independent, then according to the *Moivre—Laplace*'s limit theorem:

$$P\left(\left|\frac{k}{n} - \alpha^2\right| > 2\sqrt{\frac{\alpha^2(1-\alpha^2)}{n}}\right) \approx 0.05 \tag{2.12}$$

and if the left-hand side relation is satisfied, then

$$|\varrho_\alpha| > 2\sqrt{\frac{1-\alpha^2}{n(1-\alpha)^2}} = \frac{2}{\sqrt{n}}\sqrt{\frac{1+\alpha}{1-\alpha}}. \tag{2.13}$$

If (2.13) is valid, the hypothesis of independence can be rejected. If the indicator correlation parameter ϱ_α is to be used to decide whether the exceedances $X_1, X_2, \ldots, X_\nu$ can be considered as independent, then the set of points

$$(X_1, X_2), (X_2, X_3), \ldots, (X_{\nu-2}, X_{\nu-1}), (X_{\nu-1}, X_\nu)$$

is composed (this set of points is a statistical sample of random vectors (X, Y) where the first coordinate is always the value of the previous, and the second coordinate the value of the forthcoming exceedance. Afterwards the value of the indicator correlation ϱ_α will be calculated with regard to point $(\tilde{x}_\alpha, \tilde{y}_\alpha)$ (e.g. in the case of $\alpha = 1/2$) and investigated whether the relationship (2.13) is valid.

It should be noted that according to (2.11), if there is a monotonous functional dependence $Y = \varphi(X)$ between random variables X and Y then, if $y = \varphi(x)$ is monotonous increasing, the curve $y = \varphi(x)$ fits the points $(\tilde{x}_\alpha, \tilde{y}_\alpha)$ and if $y = \varphi(x)$ is monotonous decreasing, it fits the points $(\tilde{x}_\alpha, y_{1-\alpha})$. If the distribution functions $F(x)$ and $G(y)$ of the variables X and Y, respectively, are known, then the values of $\tilde{x}_\alpha$ and $\tilde{y}_\alpha$ can be calculated at least for a few α values, and the curve fitted to the points $(\tilde{x}_\alpha, \tilde{y}_\alpha)$ approximately reflects the character of the common tendency of both variables. Let this curve be called quantile curve. As a first approximation a curve through the three points $(\tilde{x}_{1/4}, \tilde{y}_{1/4})$, $(\tilde{x}_{1/2}, \tilde{y}_{1/2})$, $(\tilde{x}_{3/4}, \tilde{y}_{3/4})$ is acceptable for representing the common tendency between the variables X and Y, as it will be proven further on.

3. Application of the theoretical results for the flood waves of river Tisza

As an illustrative example the results of calculations based on the gauge readings on the river *Tisza* at *Szeged* are presented. The investigations were based on daily data of 70 years.

For the interval $[0, t)$ always the second quarter of every year was chosen and 6370 data were obtained (see Table 1). The gauge reading of 650 was selected as the base level equal to the readings warning to the first stage of flood protection. Exceedances higher than this level were observed on 740 days. The empirical mean value of the number of exceedances was found to be

$$\lambda t = 0.44 .$$

To show that the exceedance values $X_1, X_2, \ldots, X_{30}$ are independent, identically distributed random variables, the *Wald—Wolfowitz*'s test was executed, based on serial correlation.

This provided the following values:

$$\overline{X} = \frac{\sum_{i}^{n} X_i}{n} = 100.35 \approx 100$$

$$R = \sum_{t=1}^{n-1} X_i' X_i'4 + X_n' X_1' = 1256 ; \qquad E(R) = \frac{S_2}{n-1} = 7181$$

$$D(R) = \sqrt{\frac{S_2^2 - S_4}{n-1} + \frac{S_2^2 - 2S_4}{(n-1)(n-2)} - \frac{S_2^2}{(n-1)^2}} - 37\,297 ;$$

$$(S_r = \sum_{i=1}^{n} X_i'^r)$$

$$R^* = \frac{R - E(R)}{D(R)} = -0.1588 .$$

Since R^* is of asymptotically standard-normal distribution, there is no reason to reject the hypothesis of independence in the case of $|R^*| < 2$. The same conclusion is obtained using the indicator correlation statistics. Plotting the two-dimensional set of points $(X_1, X_2), (X_2, X_3), \ldots, (X_{29}, X_{30})$ and indicating the value $\tilde{X}_{1/2} = 69.3$ of the median, Fig. 3 was obtained.

Table 1

Magnitudes and durations of flood peak exceedances

X = magnitude of a given exceedance
Y = duration of the exceedance
$Z(t)$ = the largest exceedance in the time interval $[0, t)$ = (April 1st to June 30th) in each year at Szeged. Base level $x = 650$ cm.

Year	X [cm]	Y [Days]	Z(t) [cm]
1901	29	5	29
1902	14	3	14
1907	108	42	108
1912	{ 72	19	72
	{ 34	10	
1914	128	22	128
1915	110	35	110
1916	73	13	73
1919	266	49	266
1920	16	21	16
1922	124	36	124
1924	220	51	220
1932	273	42	273
1937	53	11	53
1940	{ 197	38	197
	{ 40	8	
	{ 28	5	
1941	204	68	204
1942	{ 38	7	60
	{ 51	11	
	{ 60	14	
1944	4	3	4
1952	2	5	2
1956	39	10	39
1958	{ 37	7	66
	{ 66	25	
1962	170	33	170
1964	114	19	114
1965	98	15	98
1967	134	41	134
1970	309	91	309

$E(X) = 100.35$; $E(Y) = 20.94$; $E[Z(t)] = 115.32$.

The calculated

$$\varrho_{1/2} \approx \frac{\frac{k}{n} - \frac{1}{4}}{\frac{1}{2} - \frac{1}{4}} = 4\frac{k}{n} - 1 = 4 \cdot \frac{9}{30} - 1 = 0.2$$

value is seen to be considerably smaller than the critical one:

$$\frac{2}{\sqrt{30}} \sqrt{\frac{1 + \frac{1}{2}}{1 - \frac{1}{2}}} = \frac{2\sqrt{3}}{\sqrt{3}\sqrt{10}} = \frac{2\sqrt{10}}{10} \approx 0.63 .$$

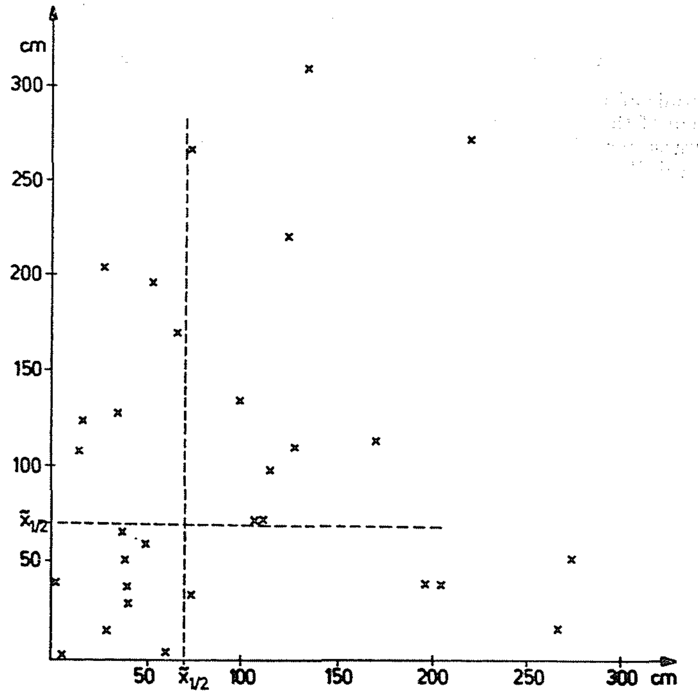

Fig. 3. Testing hypotheses of independence of the magnitudes of exceedances, Tisza river at Szeged

Thus the exceedance values may be considered as independent, consequently the theory of order statistics can be applied.

To show that in this case the number of exceedances is of *Poisson*'s distribution, a further table has been compiled, where ν_k is the number of years featuring k exceedances; NP_k is the mean value of the number of such years in case of *Poisson*'s distribution ($\lambda t = 0.45$).

k:	ν_k:	NP_k
0	45	44.8
1	21	19.6
2 or more	4	5.6

Calculation:

$$\chi^2 = \sum_{k=0}^{2} \frac{(\nu_k - NP_k)^2}{NP_k} = 0.3046\,.$$

The tabulated value of χ^2 distribution of two degrees of freedom shows that if the hypothesis is true, the number of exceedances in the given case follows *Poisson*'s distribution with a parameter $\lambda t = 0.45$, in 85% of the cases the random character of the variables causes a higher exceedance than the given value and therefore the fitting to the *Poisson*'s distribution is acceptable.

Let the hypothesis H_0 be that the values of all the exceedances form an exponentially distributed set with a distribution function $H(x) = 1 - e^{-\beta x}$, the estimated value of parameter β is $\beta = 0.01$.

Constructing the empirical distribution function $H_n(x)$ of the exceedances and calculating the statistics:

$$D_n = \sqrt{n} \sup_x |H_n(x) - H(x)| = 0.58 \, .$$

Kolmogorov's test shows that

$$P(D_n > 0.58) = 0.85$$

i.e. the fitting to the exponential distribution is good, and the hypothesis H_0 is acceptable. Considering the yearly largest exceedances, and supposing that their cumulative distribution function is

$$F_t(x) = \frac{e^{-\lambda t e^{-\beta x}} - e^{-\lambda t}}{1 - e^{-\lambda t}}$$

where in our case $\lambda t = 0.44$, $\beta = 0.01$, and composing the empirical distribution function of the largest exceedances:

$$D_n = \sqrt{n} \sup_x |F_t^*(x) - F_t(x)| = 0.66 \, .$$

The table of *Kolmogorov* shows that

$$P(D_n > 0.66) \approx 0.74$$

i.e. the fitting to the supposed distribution is surprisingly good. It is worth while to show the following table, on the basis of which the graph of the distribution function $F_t(x)$ was plotted (Fig. 4).

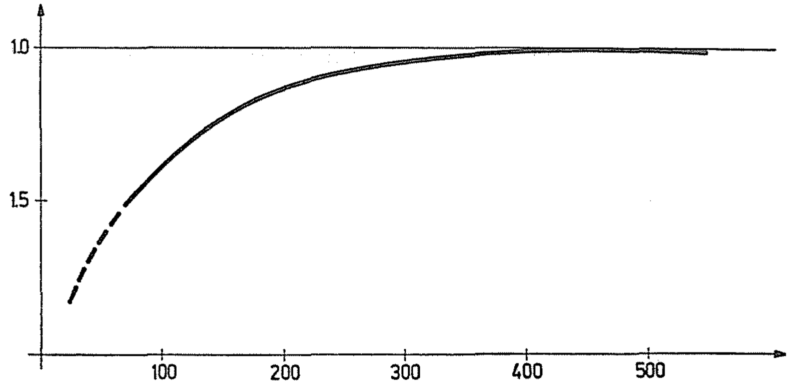

Fig. 4. Conditional cumulative distribution function of the yearly largest exceedances

$F_t(x)$	$Z(t)$ [cm]	$650 + Z(t)$ [cm]
0, 5000	73, 18	723, 17
0, 5500	83, 64	733, 64
0, 6000	95, 22	745, 21
0, 6499	108, 19	758, 19
0, 6999	123, 02	773, 02
0, 7500	140, 36	790, 38
0, 7999	161, 42	811, 42
0, 8499	188, 27	838, 27
0, 8999	225, 76	875, 76
0, 9499	289, 84	939, 24
0, 9899	435, 37	1085, 30
0, 9970	544, 30	1194, 30

Gauge readings where an excess means warning to various stages of flood protection are presently on the river *Tisza* at *Szeged* the following:

Degree I c = 650 cm
Degree II c = 750 cm
Degree III c = 800 cm.

According to our results $P[Z(t) > c_2 \mid Z(t) > c_1] \approx 0.35$
$P[Z(t) > c_3 \mid Z(t) > c_1] \approx 0.20.$

The same results were obtained in the case of readings observed at *Tokaj*, *Szolnok*, *Tiszafüred* and *Tiszaug* both in the first and the second quarters of the year (of course having different c levels and parameters λt and β). The stochastic connection between the magnitude of exceedance X and the duration of the exceedance Y (days) should be investigated.

The form of set of points $(X_1, Y_1), (X_2, Y_2), \ldots (X_{30}, Y_{30})$ is shown in Fig. 5.

Calculating the value of the indicator correlation ϱ_α in the case $\alpha = 1/2$ (using $\tilde{x}_{1/2}, \tilde{y}_{1/2}$ as median values):

$$\varrho_{1/2} = \frac{\frac{k}{n} - \frac{1}{4}}{\frac{1}{2} - \frac{1}{4}} = 4\frac{k}{n} - 1 = 4 \cdot \frac{14}{30} - 1 = 0.86\,,$$

a very close stochastic connection.

According to the location of two-dimensional points and the curve crossing the quantiles pair of points

$$(\tilde{x}_{1/4}, \tilde{y}_{1/4}), (\tilde{x}_{1/2}, \tilde{y}_{1/2}), (\tilde{x}_{3/4}, \tilde{y}_{3/4})$$

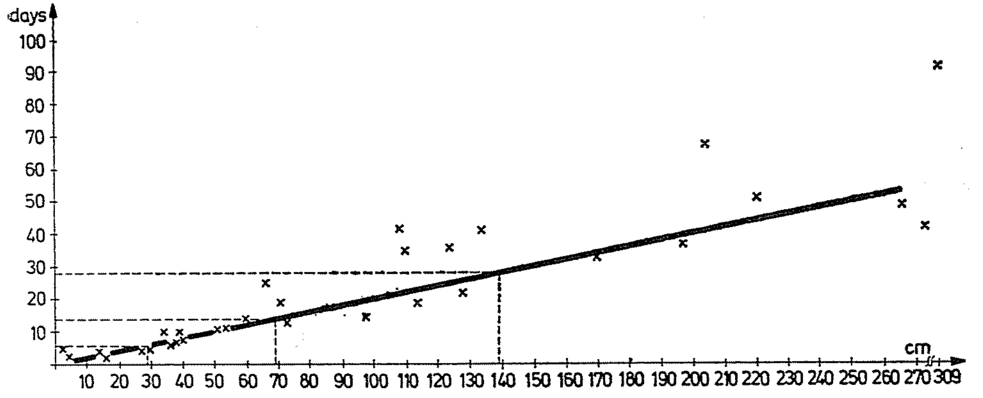

Fig. 5. Stochastic dependence between the magnitude and the duration of exceedances; Tisza river at Szeged

(which in the present case is linear) the functional relationship in this case can be stated to be linear (Fig. 5).

Supposing a linear functional relationship $Y = aX + b$ the constants can be calculated:

$$a = \frac{\tilde{y}_{3/4} - \tilde{y}_{1/4}}{\tilde{x}_{3/4} - \tilde{x}_{1/4}} \approx \frac{27.7 - 5.6}{138.6 - 29} = \frac{22.1}{109.6} \approx \frac{22}{110} = 0.2.$$

The value of the constant b is estimated considering the median values

$$\tilde{x}_{1/2} = 69.3\,; \qquad \tilde{y}_{1/2} = 13.8\,; \qquad \frac{y - 13.8}{x - 69.3} = 0.2$$

$$Y = 0.2X - 0.04 \approx 0.2X.$$

This result shows that if the value of the exceedance X in cm is known, then (at least at *Szeged* and in the investigated time interval) the duration of the exceedance is approximately

$$Y = \frac{X}{5}\ \text{days.}$$

If X is of exponential distribution with cumulative distribution function $F(x) = 1 - e^{-\beta x}$ (where $\beta = 0.01$) then naturally the random variable $Y = aX$ is also of exponential distribution with distribution function $G(x)$:

$$G(x) = P(y < x) = P(aX < x) = P\left(X < \frac{x}{a}\right) = 1 - e^{-\frac{\beta}{\alpha} x}$$

and

$$E(Y) = \frac{a}{\beta} = \frac{0.2}{0.01} = 20\,.$$

It may be noted that the arithmetical mean of the duration of exceedances (from Table 1) is

$$\overline{Y} \approx E(Y) = 20.93.$$

The same result is obtained, if instead of the theoretical quantile values $\tilde{x}_\alpha$, $\tilde{y}_\alpha$, the sample quantiles are used.

Summary

The Department of Civil Engineering Mathematics, Technical University, Budapest performed mathematical-statistical analysis of the extreme flood peaks of the river Tisza, of importance from the point of view of flood control, on commission by the Computing Bureau of the National Water Authority. The statistical regularities of the number of flood peaks above a certain base level c — significance level for flood control — and the magnitude and duration of the exceedance were to be established, to make some important forecasting possible. The stochastic connection between the magnitude and the duration of the exceedance has been investigated, and some simple methods are presented for calculating the closeness of the relationship and constructing the regression curve.

The water levels of 71 years on the river Tisza observed at the gauging stations of Tokaj, Szolnok, Tiszafüred, Tiszaug and Szeged were chosen as basic data set.

References

1. TODOROVIĆ, P. 1970. On some problems involving random number of random variables. Annals of Mathematical Statistics, Vol. 41, No. 3.
2. ZELENHASIĆ, E. 1970. Theoretical probability distributions for flood peaks. Hydrology Papers Colorado State Univ. Fort Collins, Colorado No. 42.

Prof. Dr. József REIMANN, H-1521, Budapest