VOL. 6, NO. 6 WATER RESOURCES RESEARCH DECEMBER 1970

A Stochastic Model for Flood Analysis

P. TODOROVIC AND E. ZELENHASIC

Colorado State University, Fort Collins, Colorado 80521

Abstract. A stochastic model, based on the recent developments in the theory of extreme values, is presented to describe and analyze excessive streamflows. The model is a particular stochastic process $\chi(t)$ defined as the maximum term among a random number of random observations in an interval of time $[0, t]$. Since the number of hydrograph peaks in $[0, t]$ that exceed a certain level x_0 and the magnitudes of these peaks are random variables, the foregoing model seems to conform well to the flood phenomenon. The passage time $T(x)$ of the process $\chi(t)$ relevant to the risk evaluation in the design of hydraulic structures is also considered. The results obtained are applied on the 72-year record of the Susquehanna River at Wilkes-Barre, Pennsylvania. Theoretical and observed results agree reasonably well.

INTRODUCTION

In various problems of flood analysis it is of interest to determine the distribution of the number of flood occurrences in a specific interval of time. This problem has already been studied by several authors [*Shane and Lynn*, 1964; *Benson,* 1968; *Kirby,* 1969]. For a complete description of the flood phenomenon, however, it is necessary to consider not only the frequency of flood events but also the magnitudes of the corresponding hydrograph peaks all simultaneously.

This paper is concerned with a new theoretical approach to the problem of flood analysis. The approach is based on the recent developments in the theory of extreme values [*Todorovic,* 1970] and represents an attempt to develop a more general stochastic model to describe and predict behavior of floods. The model is a stochastic process $\chi(t)$ defined as the maximum term among a random number of random variables in an interval of time $[0, t]$. Since the number of flood peak discharges in $[0, t]$ exceeding a certain level x_0 and the magnitudes of these peaks are random variables, the foregoing model seems to conform well to the flood phenomenon.

Practical aspects suggest that the entire process of instantaneous discharge not be considered but rather the sequence of the hydrograph peaks. In this way a sequence of random variables $Q_1, Q_2, \cdots$ is obtained (Figure 1). To investigate the maximum peak, it might seem logical to apply the classical extreme values theory. Unfortunately this method cannot be applied because the number of Q_i in $[0, t]$ and the time when Q_i emerges are random variables.

The data used to illustrate and to test the new approach are obtained from the partial duration series of flood peak discharges for the Susquehanna River at Wilkes-Barre, Pennsylvania. These data refer to the period 1891–1964. Two years, 1898 and 1899, are omitted because of the nonhomogeneity of data. The already existing partial duration data on floods have been checked out using the flood hydrographs.

PHENOMENOLOGICAL CONSIDERATIONS

Following *Kirby* [1969], any streamflow hydrograph can be interpreted as a sequence of nearly instantaneous hydrograph peaks separated by relatively long periods of low flow. Because of the nature of the phenomenon, the number of these peaks in a certain interval of time $[0, t]$ and their magnitudes are random variables. Since the number of peaks in $[0, t]$ is random, the times when these peaks emerge are random variables too.

If we consider a certain level x_0 and if we consider only those peaks Q_i in $[0, t]$ that exceed x_0, we can define

$$\xi_i = Q_i - x_0$$

where $\xi_i > 0$ is a random variable for all $i =$

1, 2, $\cdots$. With each ξ_i we associate the time $\tau(i)$ when the corresponding peak occurred (Figure 1). When a flood hydrograph is a multiple peaked hydrograph [*Chow*, 1964, p. 14–18], only the largest peak is taken into consideration. For simplicity, flood peak exceedance flows ξ_i's from now on will be called exceedances only.

Consider an interval of time $[0, t]$ and denote by $\chi(t)$ the largest ξ_ν in this interval. Since the number of ξ_ν in $[0, t]$ is a random variable that depends on time t, $\chi(t)$ is defined as follows:

$$\chi(t) = \sup_{\tau(\nu) \leq t} \xi_\nu \tag{1}$$

By virtue of definition it follows that for every $t \geq 0$ and $\Delta t > 0$

$$\chi(t) \leq \chi(t + \Delta t)$$

This means that $\chi(t)$ is a stochastic process of nondecreasing sample functions (Figure 2).

In the following an attempt is made to determine a one-dimensional distribution function $F_t(x)$ of the stochastic process $\chi(t)$:

$$F_t(x) = P[\chi(t) \leq x] \tag{2}$$

DISTRIBUTION OF THE NUMBER OF THE EXCEEDANCES

In this section the distribution function of the number of exceedances of the level x_0 is determined. Denote by $\eta(t)$ the number of exceedances in the interval of time $[0, t]$. By definition, $\eta(t)$ may be 0, 1, 2, $\cdots$, and for all $t \geq 0$ and $\Delta t > 0$, $\eta(t) \leq \eta(t + \Delta t)$. In addition, $\eta(t)$ depends on x_0 (for fixed t, $\eta(t)$ is a nonincreasing function of x_0). However, in what is to follow it is assumed that x_0 is a fixed number.

Denote by $E_\nu^t = [\eta(t) = \nu]$ then it follows that

$$E_i^t \cap E_j^t = \Theta \text{ for all } i \neq j \text{ and } \bigcup_{\nu=0}^{\infty} E_\nu^t = \Omega \tag{3}$$

where Θ stands for the impossible and Ω stands for a certain event. Let $\Lambda(t)$ stand for $E[\eta(t)]$, i.e.,

$$\Lambda(t) = \sum_{\nu=1}^{\infty} \nu P(E_\nu^t) \tag{4}$$

then because of the seasonal variation $\Lambda(t)$ is a nonlinear function of time.

Write $F_n(t) = P[\tau(n) \leq t]$. Then from *Todorovic* [1970]

$$P(E_n^t) = F_n(t) - F_{n+1}(t) \tag{5}$$

From equation 5 we obtain

$$F_n(t) = \sum_{i=n}^{\infty} P(E_i^t) \tag{6}$$

Under certain very general assumptions one may

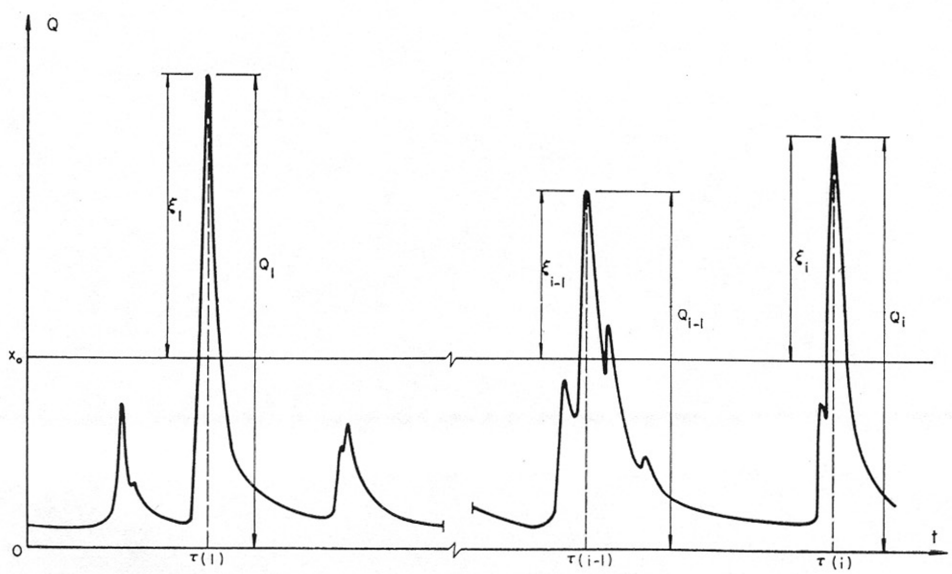

Fig. 1. Schematic representation of a streamflow hydrograph.

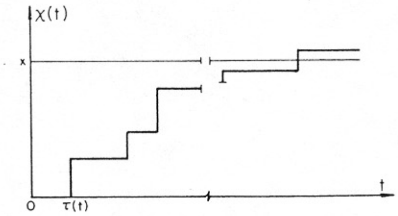

Fig. 2. A sample function of the process $\chi(t)$.

show that probabilities $P(E_k^t)$ satisfy the following system of differential equations [*Todorovic*, 1970]:

$$\frac{dP(E_k^t)}{dt} = \lambda_{k-1}(t)P(E_{k-1}^t) - \lambda_k(t)P(E_k^t) \qquad k = 1, 2, \cdots$$

$$\frac{dP(E_0^t)}{dt} = -\lambda_0(t)P(E_0^t) \qquad (7)$$

where

$$\lambda_k(t) = \lim_{\Delta t \to 0} \frac{P(E_1^{t,t+\Delta t} \mid E_k^t)}{\Delta t} \qquad (8)$$

and

$$E_1^{t,t+\Delta t} = [\eta(t + \Delta t) - \eta(t) = 1]$$

It is not difficult to verify that system 7 has the following solution:

$$P(E_0^t) = \exp\left[-\int_0^t \lambda_0(s)\, ds\right]$$

$$P(E_k^t) = \exp\left[-\int_0^t \lambda_k(s)\, ds\right] \int_0^t \lambda_{k-1}(t_1) \cdot \exp\left\{\int_0^{t_1} [\lambda_k(s) - \lambda_{k-1}(s)]\, ds\right\} \cdot \int_0^{t_1} \cdots \int_0^{t_{k-1}} \lambda_0(t_k) \cdot \exp\left\{\int_0^{t_k} [\lambda_1(s) - \lambda_0(s)]\, ds\right\} dt_k\, dt_{k-1} \cdots dt_1 \qquad (9)$$

Evidently simple expression for each $P(E_k^t)$ in terms of $[\lambda_k(t)]$ is not possible in general; however, several special cases have been solved [*Todorovic and Yevjevich*, 1969].

The case we believe is of relevance in flood analysis is when

$$\lambda_k(t) \equiv \lambda(t) \text{ (independent of } k)$$

Under this condition one can easily check that

$$P(E_k^t) = \exp\left[-\int_0^t \lambda(s)\, ds\right]\left[\int_0^t \lambda(s)\, ds\right]^k / k! \qquad (10)$$

This is a time dependent Poisson process. From the mathematical expectation given by equation 4, $\Lambda(t)$ becomes

$$\Lambda(t) = \int_0^t \lambda(s)\, ds \qquad (11)$$

Equation 6 represents the distribution function of the time of the nth exceedance. Denote by $f_n(t)$ the corresponding density function since $F_n(t)$ can be written as follows:

$$F_n(t) = 1 - \sum_{j=0}^{n-1} P(E_j^t)$$

Taking into account equation 10, after differentiation, it follows that

$$f_n(t) = \frac{\lambda(t)}{\Gamma(n)} \exp\left[-\int_0^t \lambda(s)\, ds\right]\left[\int_0^t \lambda(s)\, ds\right]^{n-1} \qquad (12)$$

DISTRIBUTIONS OF $\chi(t)$ AND $T(x)$

According to the definition given in a previous section, $\chi(t)$ represents the maximum flood peak exceedance flow in an interval of time $[0, t]$, and, as shown, $\chi(t)$ is a stochastic process of nondecreasing (step) sample functions. In connection with $\chi(t)$, another process $T(x)$ is defined as the (random) instant when for the first time (counting from zero) an observation ξ_ν exceeds a given value x. In mathematical terms $T(x)$ is defined as follows:

$$T(x) = \inf\, [t; \chi(t) > x] \qquad (13)$$

i.e., $T(x)$ is the smallest t for which $\chi(t) > x$.

Consider the process $\chi(t)$ and denote by $F_t(x)$ the corresponding distribution function; i.e.,

$$F_t(x) = P[\chi(t) \leq x] \qquad t \geq 0 \qquad x \geq 0$$

Then according to theorem 1 [*Todorovic*, 1970]

$$F_t(x) = P(E_0^t) + \sum_{k=1}^{\infty} P\left[\bigcap_{\nu=1}^{k} (\xi_\nu \leq x) \cap E_k^t\right] \qquad (14)$$

Distribution function 14 may be interpreted as the probability that all exceedances ξ_ν in $[0, t]$

will be less than or equal to x. If $x = 0$ it follows from equation 14 that

$$F_t(0) = P(E_0^t) \tag{15}$$

Identity 15 represents the probability that there will be no exceedances in the interval $[0, t]$. From the foregoing discussion it follows that $F_t(x)$ is not differentiable at the point $x = 0$ (Figure 3).

The mathematical expectation and variance of the stochastic process $\chi(t)$ can be determined as follows. Denote by I_A the indicator of the set A. Then with regard to relation 3 it follows that

$$I_{\bigcup_{\nu=0}^{\infty} E_\nu^t} = \sum_{\nu=0}^{\infty} I_{E_\nu^t} = 1$$

From this, assuming $E[\chi(t)]$ exists, we have

$$E[\chi(t)] = E\left[\chi(t) I_{\bigcup_{\nu=0}^{\infty} E_\nu^t}\right] \sum_{\nu=0}^{\infty} E[\chi(t) \cdot I_{E_\nu^t}] \tag{16}$$

By virtue of definition of $\chi(t)$ it turns out that

$$\chi(t) I_{E_\nu^t} = \sup_{0 \le k \le \eta(t)} \xi_k I_{E_\nu^t} = \sup_{0 \le k \le \nu} \xi_k \cdot I_{E_\nu^t}$$

because on the set E_ν^t the random variable $\eta(t) = \nu$. Therefore, mathematical expectation 16 becomes

$$E[\chi(t)] = \sum_{\nu=1}^{\infty} E(\sup_{1 \le k \le \nu} \xi_k I_{E_\nu^t})$$

$$= \sum_{\nu=1}^{\infty} E(\sup_{1 \le k \le \nu} \xi_k \mid E_\nu^t) P(E_\nu^t) \tag{17}$$

In a similar way one can prove that

$$E[\chi(t)]^2 = \sum_{\nu=1}^{\infty} E[(\sup_{1 \le k \le \nu} \xi_k)^2 \mid E_\nu^t] P(E_\nu^t) \tag{18}$$

Consider now the stochastic process $T(x)$ and write

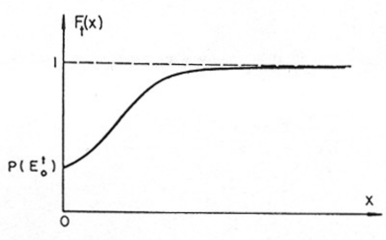

Fig. 3. Distribution function $F_t(x)$.

$$\Phi_x(t) = P[T(x) \le t]$$

Since the sample functions of the process $\chi(t)$ are nondecreasing step functions, the following relation is obvious (Figure 2):

$$P[\chi(t) \le x] = P[T(x) > x] \tag{19}$$

From this it follows that

$$\Phi_x(t) = 1 - F_t(x)$$

or

$$\Phi_x(t) = 1 - P(E_0^t) - \sum_{k=1}^{\infty} P\left[\bigcap_{\nu=1}^{k} (\xi_\nu \le x) \cap E_k^t\right] \tag{20}$$

It is apparent that for all $x > 0$

$$\Phi_x(0) \equiv 0$$

RISK EVALUATION OF EXCEEDING THE DESIGN FLOW

One of the most important problems in the design of a hydraulic structure is the selection of its capacity, which must be accommodated to the existing conditions. It is extremely uneconomical to build the structure of large enough capacity to carry all possible flows during its life. Instead, a structure of smaller capacity is usually constructed. This of course implies a certain risk; therefore it is necessary to find a way to determine whether the risk is a reasonable one.

The passage time $T(x)$ of the stochastic process $\chi(t)$ represents a natural measure of the risk. However, certain phenomenological reasons suggest not using the passage time directly, but rather proceeding as follows: Denote by

$$\chi_1, \chi_2, \cdots \tag{21}$$

the sequence of the annual maximum values and define the new random variable N_x in the following way:

$$N_x = \inf(\nu; \chi_\nu > x) \tag{22}$$

where $x > 0$. It is apparent that for all $x > 0$, N_x may assume only nonnegative integer values. In addition, by virtue of definition it follows that for every $n = 1, 2, \cdots$,

$$P(N_x = n) = P(\chi_1 \le x, \cdots, \chi_{n-1} \le x, \chi_n > x) \tag{23}$$

and for $n = 0$

$$P(N_x = 0) = P(E_0^{t^*})$$

where t^* is equal to the 1-year period.

Because of the nature of the phenomenon, it seems reasonable to assume that χ_ν represents a sequence of independent random variables with the common distribution function $F_{t^*}(x)$; i.e.,

$$P(\chi_\nu \leq x) = F_{t^*}(x)$$

Hence

$$P(N_x = n) = [F_{t^*}(x)]^{n-1}[1 - F_{t^*}(x)]$$

From this it follows that the mathematical expectation of the random variable N_x for all $x > 0$ is equal to

$$E(N_x) = \sum_{n=1}^{\infty} n[F_{t^*}(x)]^{n-1}[1 - F_{t^*}(x)] = \frac{1}{1 - F_{t^*}(x)} \qquad (24)$$

$E(N_x)$ represents the average number of years when the first exceedance of the value x occurs; it has been used as a measure of risk in the sense that for a given hazardous flow x, equation 24 provides the average passage time of the level x. Taking into consideration the lifetime of the structure, the designer may decide what the capacity of the structure should be.

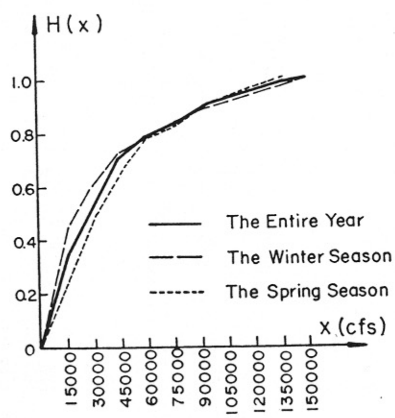

Fig. 4. Observed distributions of the magnitude of flood peak exceedance flows for the winter season, spring season, and the entire year for the Susquehanna River at Wilkes-Barre, Pennsylvania.

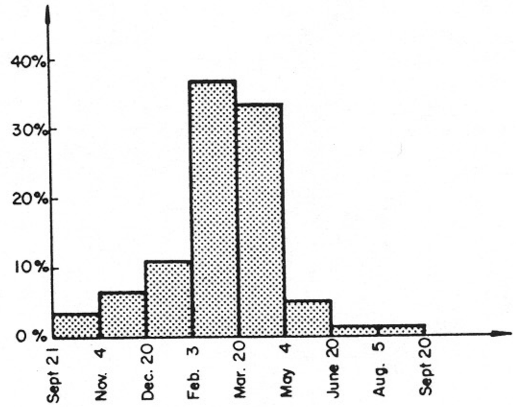

Fig. 5. Percentage of the total number of floods of the Susquehanna River at Wilkes-Barre, Pennsylvania, for the period of 72 years.

A better measure of the risk can be obtained in the following way. For the given lifetime of the structure (the number of years n) and the risk α, the level x can be determined so that

$$\sum_{\nu=n}^{\infty} [F_{t^*}(x)]^{\nu-1}[1 - F_{t^*}(x)] = \alpha$$

or

$$F_{t^*}(x) = \alpha^{1/(n-1)}$$

APPLICATIONS

In this section an application of the foregoing results on the Susquehanna River at Wilkes-Barre, Pennsylvania, is considered. The data available refer to the period 1891–1964; the base x_0 for the partial duration series is 82,000 cfs (cubic feet per second).

Before going further it is necessary to say that all isolated peaks larger than the level x_0 are taken into consideration. When a flood hydrograph is a multiple peaked hydrograph [*Chow*, 1964, p. 14–18] only the largest peak is taken into consideration (Figure 1). This procedure was adopted to make obtained values for ξ_ν independent. The other way to make ξ_ν independent random variables is to use the method of hydrograph separation. However, this method has a serious shortcoming because the actual maximum peak and the one obtained by the hydrograph separation procedure may not be the same.

In the sequel the following are the basic assumptions:

1. (ξ_ν) is a sequence of independent identically distributed random variables with $H(x) = P(\xi_\nu \leq x)$.
2. (ξ_ν) and $(\tau(\nu))$ are mutually independent sequences.

Under these two hypotheses the distribution function 14 and the mathematical expectation $E[\chi(t)]$ become

$$F_t(x) = P(E_0{}^t) + \sum_{k=1}^{\infty} [H(x)]^k \cdot P(E_k{}^t) \qquad (25)$$

$$E[\chi(t)] = \sum_{k=1}^{\infty} \left\{ \int_0^{\infty} xd[H(x)]^k \right\} \cdot P(E_k{}^t) \qquad (26)$$

To justify the foregoing hypothesis one should be reminded that at least for practical purposes those ξ_ν observed for a season may be assumed independent and identically distributed [*Chow,* 1964]. As mentioned before, a streamflow hydrograph can be interpreted as a sequence of nearly instantaneous hydrograph peaks separated by relatively long periods of low flow. Regarding this and the foregoing procedure, physical intuition is not violated by treating the exceedances of a certain level x_0 as independent random variables, not only for a particular season but for the entire year.

The second part of the first hypothesis asserts that ξ_ν are identically distributed random variables. This claim does not seem justifiable; in fact it does not look realistic. To resolve this problem for the particular case of the Susquehanna River at Wilkes-Barre ($x_0 = 82{,}000$ cfs), three distribution functions of values of ξ_ν are made. The first distribution is related to the spring season, the second one to the winter season, and the third one to the entire year. The number of ξ_ν in the other two seasons was too small to be analyzed. In Figure 4 these three distributions are depicted. The Kolmogorov-Smirnov test has shown that these three distribution functions are not significantly different. This result, combined with the statement that inside seasons ξ_ν are identically distributed, supports the hypothesis that at least for the case under consideration (ξ_ν) can be regarded as a sequence of identically distributed random variables.

FREQUENCY ANALYSIS OF EXCEEDANCES

Contrary to Kirby's claim, the times between hydrograph peaks cannot be considered as identically distributed random variables (although the physical intuition is not violated by treating the times as independent). Because the distribution of the exceedances is not uniform throughout a year (Figure 5), probabilities

$$P[\eta(t_1 + \Delta t) - \eta(t_1) = k] \neq P[\eta(t_2 + \Delta t) - \eta(t_2) = k]$$

for different t_1 and t_2, are different.

In the following an attempt is made to estimate probabilities $P(E_\nu{}^t)$. According to equation 10, for this purpose it is necessary to evaluate the function $\Lambda(t)$ (the average number of exceedances in $[0, t]$). For the Susquehanna River the observed $\Lambda'(t)$ given in Figure 7 is determined in the following way. The interval of the water year October 1–September 30 was divided into seventeen 20-day and one 25-day periods. Then assuming that October 1 is the origin for periods of 0–20 days, 0–40 days, $\cdots$

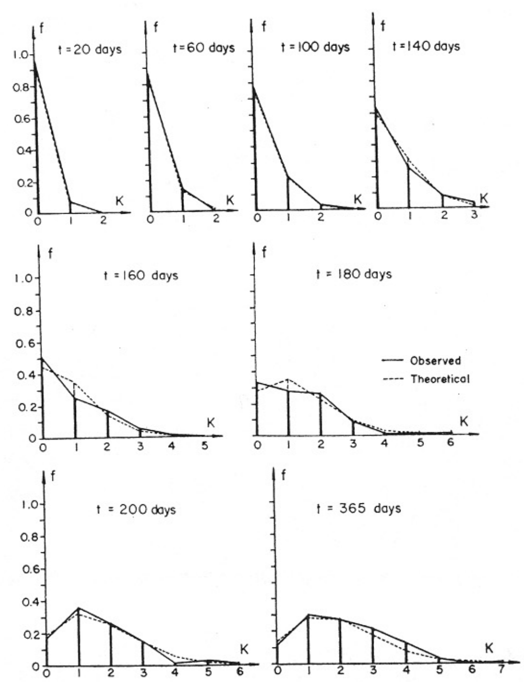

Fig. 6. Observed and the corresponding theoretical (Poisson) distributions of the number of exceedances for periods of 20, 60, 100, 140, 160, 180, 200, and 365 days.

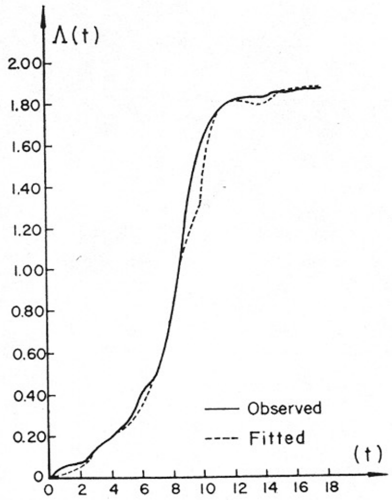

Fig. 7. Observed $\Lambda(t)$ and the fitting function. Unit interval on t axis stands for 20-day period.

0–365 days, the observed and the corresponding theoretical (Poisson) distributions of the number of exceedances were determined (Figure 6). From these distributions the values of $\Lambda'(t)$ at points $t = 20, 40, \cdots 365$ are obtained. The fitting function $\Lambda(t)$ has the following expression:

$$\begin{aligned}\Lambda(t) = {} & 0.1015 + 0.1050t \\ & + 0.3936 \cos\left(\frac{\pi t}{9} + 0.6032\pi\right) \\ & + 0.1280 \cos\left(\frac{2\pi t}{9} - 0.4074\pi\right) \\ & + 0.0604 \cos\left(\frac{2\pi t}{6} + 0.5892\pi\right) \\ & + 0.0130 \cos\left(\frac{2\pi t}{3} - 0.2041\right)\end{aligned} \quad (27)$$

As has been mentioned, Shane and Lynn studied this problem. Their result, that the distribution of the number of exceedances is a time independent Poisson process, is not general enough. In this case $\Lambda(t) = \lambda t$. However, from Figure 7 this is obviously not the case. Therefore in the equation

$$P(E_{\nu}^{t}) = e^{-\Lambda(t)} \frac{[\Lambda(t)]^{\nu}}{\nu!} \quad (28)$$

$\Lambda(t)$ is given by equation 27.

ANALYSIS OF DISTRIBUTION FUNCTION $F_t(x)$

Evaluation of the distribution function $F_t(x)$ represents the central problem of this section. On the basis of equation 28, distribution function 25 becomes

$$F_t(x) = \exp\{-\Lambda(t)[1 - H(x)]\} \quad (29)$$

where $\Lambda(t)$ is given by equation 27 and $H(x)$ will be determined. It can easily be seen from Figure 4 that the exponential distribution suits the observed distributions. Taking the annual observed distribution of exceedances as representative and determining from these data the corresponding theoretical (exponential) distribution (Figure 8) one may obtain from (29)

$$F_t(x) = \exp[-\Lambda(t) \exp(-2.628 \cdot 10^{-5} \cdot x)] \quad (30)$$

where x is measured in cfs. Figure 9 represents graphically distribution function 30 for $t = 160$, 200, and 365 days and the corresponding observed distribution.

COMPUTATION OF THE LARGEST ANNUAL EXCEEDANCE WHOSE RETURN PERIOD IS 100 YEARS

In this case $t^* = 365$ days and

$$E(N_x) = \frac{1}{1 - F_{t^*}(x)} = 100$$

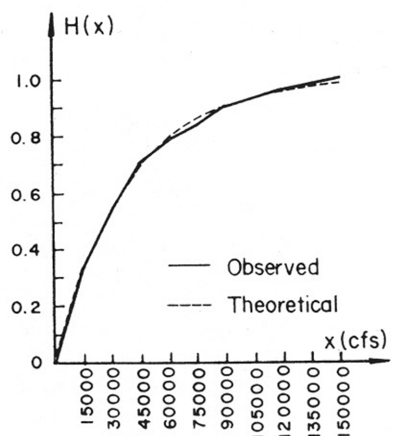

Fig. 8. Observed and theoretical distributions of the magnitude of exceedances for the Susquehanna River at Wilkes-Barre, Pennsylvania, for a 1-year period.

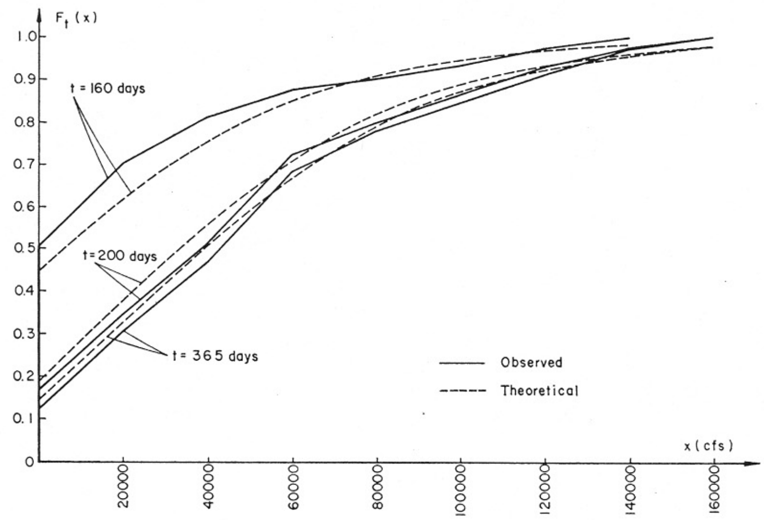

Fig. 9. Observed and theoretical distribution functions of the maximum flood peak exceedance flow for the Susquehanna River at Wilkes-Barre, Pennsylvania, for 160-, 200-, and 365-day periods.

from which we obtain

$$\exp\left[-\Lambda(t)\exp\left(-2.628 \times 10^{-5}x\right)\right] = 0.990 \quad (31)$$

The function $\Lambda(t)$ has the value 1.889 for $t^* =$ 365 days. Therefore, the solution of equation 31 is

$$x = 1.992 \times 10^5 \text{ cfs}$$

which is the value of the largest annual flood peak exceedance flow having a return period of 100 years.

In this way, the design flood against which a dam is to be built, for example, can be computed.

CONCLUSION

This paper, using some recent results in the theory of extreme values of a particular class of stochastic processes, discusses the development of a probabilistic model that describes the flood phenomenon. The model is sufficiently general to be applied to most cases with important practical applications. The simplest form of the model, when exceedances $\xi_1, \xi_2, \cdots$ represent a sequence of independent identically distributed random variables independent of $\tau(1), \tau(2), \cdots$, is applied to the 72-year record of the Susquehanna River at Wilkes-Barre, Pennsylvania. Observed and theoretical results seem to agree fairly well (application of the model on the Greenbrier River at Alderson, West Virginia, has also shown good agreement between theoretical and observed results). Certainly for some rivers the simplest form of the model is not suitable, and it will be necessary to develop new particular models from the general model given by equation 14.

Acknowledgment. This research was supported by National Science Foundation grants 11444 and 11499.

REFERENCES

Benson, M., Uniform flood-frequency estimating methods for federal agencies, *Water Resour. Res.*, *4*(5), 891–908, 1968.

Chow, V. T., *Handbook of Applied Hydrology*, McGraw-Hill, New York, 1964.

Gumbel, E. J., *Statistics of Extremes*, 2nd ed., Columbia University Press, New York, 1967.

Kirby, W., On the random occurrence of major floods, *Water Resour. Res.*, *5*(4), 778–784, 1969.

Shane, R., and W. Lynn, Mathematical model for flood risk evaluation, *J. Hydraul Div., Amer. Soc. Civil Eng.*, *90*(HY6), 1–20, 1964.

Todorovic, P., On some problems involving random number of random variables, *Ann. Math. Statist.*, *41*(3), 1059–1063, 1970.

Todorovic, P., and V. Yevjevich, Stochastic process of precipitation, *Hydrol. Pap. 35*, Colorado State University, Fort Collins, Colorado, September 1969.

(Manuscript received June 2, 1970;
revised July 24, 1970.)