Jour. Ind. Soc. Ag. Statistics
Vol. XLII, No. 1 (1990), pp. 140-159

ISAS YOUNG SCIENTISTS AWARD PAPER

BIVARIATE EXTREMES*

N. P. SINGH
Haryana Agricultural University, Hisar-125 004

SUMMARY

The application of Gumbel's extreme value theory in the analysis and prediction of excessive water discharges/floods which are also known as exceedances has been subject of numerous investigations by many authors. Todorovic and Zelenhasic [16] developed more reasonable theory in the form of distribution of maximum and minimum exceedances. The application of these distributions to flood phenomena represent an improvement over Gumbel's classical theory in univariate case for partial duration data. In the present investigation the distribution for maximum and minimum exceedances are derived in bivariate case which will be more suitable in practical situation in comparison to distributions given by Sibuya [8] Gumbel and Goldstein [3], Gumbel and Mustafi [4], Nair [7] and Tiago de Oliveira [13, 14]. The suitability of bivariate distribution is illustrated by using the data of water discharges for a partial duration series from 1948 to 1977 of river Narmada at two stations, i.e. Mortakka and Gardeshwar. A fairly good agreement between theoretical and observed results is seen.

Key Words :— Extremes, Floods, Exceedances, Fourier series, Return period, Medial correlation.

Introduction

The analysis and prediction of excessive water discharges (floods) have been subject of numerous investigations by many authors. The common approach to these problems consists of either applying Gumbel's distribution or selecting a flood frequency distribution function based on the

*Presented at the 43rd annual conference of the Society held at HAU, Hisar, December, 1989.

criterion of 'best curve fit' to the observed largest water discharge values at a lapes of time. The criterion of 'best curve fit' seems somewhat adhoc on physical grounds and more cumbersome. Also the application of Gumbel's extreme value distribution is based on the following two assumptions.

(i) The sequence of daily water discharge values of whole year forms a sequence of independent and identically distributed random variables;
(ii) These random variables are assumed to have a distribution with an exponential tail.

For the limitations of physical basis of these assumptions the Gumbel [1] statement is quoted as :

"It must be admitted that the good fit cannot be foreseen from the theory which is based on three assumptions :

(i) the distribution of daily discharges is of exponential type;
(ii) $n = 365$ is sufficiently large; (iii) the daily observations are independent.

Assumption (i) cannot be checked since the analytic form of the distribution of daily water discharge values is unknown. $n = 365$ is not sufficiently large number for the convergence of a distribution function to the Gumbel extreme value distribution. The third assumption definitely does not hold as the daily observations are in general not independent.

On the basis of this criticism by Gumbel one can say that mathematical assumptions underlying the classical extreme value theory may not always be applicable to a water discharge series. If the time interval of interest is less than a year which is often the case in most countries under the influence of monsoon, then the use of an aspmptotic distribution function can hardly be justified in the light of above criticism of the application of asymptotic theory. This criticism points out the need to develop a theory that is physically more meaningful for flood frequency analysis than the classical extreme value theory.

The first attempt to develop such a theory from the properties of stream flow (rather than to explain the properties of stream flow from such a theory) is that of Todorovic and his coworkers. Their formulism is based on the partial duration series of water discharges. These results were successfully applied for the analysis of partial duration water discharge series by Zelenhasic [20], Todorovic and Rouselle [18], Todorovic and Woolhiser [19]. Gupta *et al.* [5] and Singh [10, 11].

Whereas, lot of work has been done in theory of extreme values for one dimensional random variables, not much work is reported in literature for two dimensional random variables. However, in problems of applied nature, it is always possible to get information on some other variables which are dependent on the variable of interest without much efforts and cost. Keeping in view these points, the distribution of number of exceedances which is a bivariate Poisson distribution with time dependent intensity functions of minimum and maximum exceedances is presented in Section 3. The distribution functions of maximum and minimum exceedances are presented in Section 4. The results of Todorovic and his cowoikers are reviewed in Section 2. The results are applied to the water discharge series of river Narmada at two points (stations) i.e., Mortakka and Gardeshwar and discussed in Section 5. The data refer to the period 1948 to 1977 for a partial duration series (starting from 1st July to October 28) for every year. Tho expected numbers of these exceedances in a fixed time interval for both the stations were expressed by a Fourier expansion.

2. Distribution Function of Exceedances in Univariable case

Todorovic [15, 17] utilizes a stochastic model to describe and predict the behaviour of floods. He starts with a stochastic process $X(t)$ defined as the highest magnitude of random variable (maximum water discharge 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 magnitude of these peaks are random variables, the foregoing stochastic model seems to conform well to rhe flood phenomenon.

Let Q_i's be the the flood peak values that exceed a suitable chosen base level x_0 without being less than x_0 and $Z_i = Q_i - x_0$. Call Z_i as the magnitude of the i-th exceedance.

The distribution function of supremum and infimum of Z_i's are given by

$$F_{t_s}(x) = EP\left[\underset{\tau(k)=t}{\text{Sup}}\, Z_k \leqslant x/\eta(t)\right] \tag{1}$$

$$F_{t_s}(x) = EP\left[\underset{\tau(k)=t}{\text{Inf}}\, Z_k \leqslant x/\eta(t)\right] \tag{2}$$

Where $\eta(t)$ is tha number of exceedances over x_0 in the time interval $(0, t)$. It is a non-decreasing function of time and assumes valuəs

0, 1, 2, ... And $\tau(k)$ is the time of k-th exceedance. Todorovic and Zelenhasic [16] assumed that these exceedances are governed by a Poisson process which has a time dependent intensity function. Expectation of (1) and (2) is given by

$$F_{t_s}(s) = P\ [\sup_{0<k\leqslant n} Z_k \leqslant x \cap E_n^t] \qquad (3)$$

and

$$F_{t_l}(x) = 1 - P\ [\inf_{0\leqslant k\leqslant n} Z_k \leqslant x \cap E_n^t] \qquad (4)$$

Under the assumptions that (i) the sequence of Z_k's is a sequence of independent random variables with common distribution function $F(x)$ and (ii) the number of exceedances $\eta(t)$ are independent of their magnitude Z_k, the distribution functions (3) and (4) can be written as

$$F_{t_s}(x) = \sum_{n=0}^{\infty} [F(x)]^n P(E_n^t) \qquad (5)$$

$$F_{t_l}(x) = 1 - \sum_{n=1}^{\infty} [1 - F(x)]^n P(E_n^t) \qquad (6)$$

Where $P(E_n^t)$ is the probability that n event will occur in $(0, t)$. Substituting the probability of Poisson process for $P(E_n^t)$ one can get

$$F_{t_s}(x) = \exp(-V(t))(1 - F(x)) \qquad (7)$$

$$F_{t_l}(x) = 1 + \exp(V(t) - \exp(V(t))F(x)) \qquad (8)$$

If the magnitudes of exceedances are exponentially distributed with distribution function.

$$F(x) = 1 - e^{-\alpha x}; \quad \alpha = 0,\ x \geqslant 0 \qquad (9)$$

then the distribution function Sup Z_k is given by

$$F_{t_s}(x) = \exp(V(t)\ e^{-\alpha x}); \quad \alpha > 0,\ x \geqslant 0 \qquad (10)$$

It may be noted that the functional form of (10) is similar to Gumbel's extreme value distribution function. However, Gumbel's distribution is an asymptotic expression whereas the expression (10) represents an exact (non asymptotic) expression governed by Poisson process of independent identically distributed exponential random variables. This express on is also used implicitly by Sibuya [9] to generate exponential random numbers.

3. Distribution Function of Number of Exceedances in Bivariate Case

Let $X(t)$ and $Y(t)$ be defined as the highest magnitude of two random variables (maximum water discharge) in an interval of time $(0, t)$. We denote by $\tau(k)$ and $\tau(l)$ the times that elapsed before the k-th and l-th exceedance occurred in $X(t)$ and $Y(t)$ variables respectively. It is assumed that $\tau(0) = 0$ for both the events.

The time $\tau(k)$ and $\tau(l)$ in $(0, t)$ are random variables. In addition these are continuous and dependent on each other. The time $\tau(k)$ and $\tau(l)$ satisfy the conditions.

$$0 \leqslant \tau(k) \leqslant \tau(k+1),\ 0 < \tau(l) \leqslant \tau(l+1) \tag{11}$$

Let us denote by Z_k and Q_l, the magnitude of the kth and l-th exceedances in $X(t)$ and $Y(t)$ respectively which exceed suitably chosen fixed base levels. And also assume $Z_0 = 0$, $\theta_0 = 0$. The expression for the distribution function of number of exceedances for $X(t)$ and $Y(t)$ is derived as follows.

Let us again define an event

$$E_{k,l}^{t,t} = [\tau(k) \leqslant t \leqslant \tau(k+1),\ \tau(l) \leqslant t < \tau(l+1)] \tag{12}$$

Obviously $E_{k,l}^{t,t} \cap E_{m,n}^{t,t} = \phi$ for $k, l \neq m, n$.

and $\bigcup\limits_{k,l} E_{k,l}^{t,t} = \Omega$ where $E_{k,l}^{t,t}$ defines k exceedances in $X(t)$ and l exceedances in $Y(t)$ in time interval $(0, t)$.

By virtue of (12) we have

$$\begin{aligned} P(E_{k,l}^{t,t}) = {} & P(\tau(k) \leqslant t, \tau(l) \leqslant t) - P(\tau(k) \leqslant t, \tau(l+1) \leqslant t) \\ & - P(\tau(k+1) \leqslant t, \tau(l) \leqslant t) \\ & + P(\tau(k+1) \leqslant t, \tau(l+1) \leqslant t) \end{aligned} \tag{13}$$

Now in order to derive an expression for the probabilities $P\ (E^{t,t}_{n,m})$, let $\eta(t)$ and $\beta(t)$ be the number of exceedances in the time interval $(0, t)$ for x and y respectively. Then for all $t > 0$ and $\Delta t > 0$

$$\eta(t) \leqslant \eta(t + \Delta t), \qquad \beta(t) \leqslant \beta(t + \Delta t) \tag{14}$$

The possibilities accounting for exceedances in two variables are enumerated below.

$$\text{(i) } [X(t) > x_0,\ Y(t) < Y_0], \text{ (ii) } [X(t) \geqslant x_0,\ Y(t) \geqslant Y_0],$$
$$\text{(iii) } [X(t) < x_0,\ Y(t) > Y_0] \tag{15}$$

where x_0 and Y_0 are suitably chosen constants.

The corresponding probabilities can be derived as follows. Suppose that in time interval $(0, t)$, k and l exceedances had already occurred in $X(t)$ and $Y(t)$ respectively, we denote by $\mu_{k,l}(t)$, $\nu_{k,l}(t)$ and $\lambda_{k,l}(t)$ the probabilities corresponding to three events of (15) that is in short interval $(t, t + \Delta t)$ an exceedance will occur either in $X(t)$ or in $Y(t)$ or in both.

$$\mu_{k,l}(t) = \lim_{\Delta t \to 0} \frac{P\left[E^{t,t+\Delta t\,;\,t,t+\Delta t}_{k+l,t} \;\middle|\; E^{tt}_{k,l} \right]}{\Delta t}$$

$$\text{where } E^{t,t+\Delta t\,;\,t,t+\Delta t}_{k+l,l} = \begin{bmatrix} \eta(t+\Delta t) - \eta(t) = 1 \\ \beta(t+\Delta t) - \beta(t) = 0 \end{bmatrix} \tag{16}$$

i.e. one exceedance occurs in $X(t)$ with no exceedance in $Y(t)$. Similariy for the second event we have probability

$$\nu_{kl}(t) = \lim_{\Delta t \to 0} \frac{P\left[E^{t,t+\Delta t\,;\,t,t+\Delta t}_{k,l+1} \;\middle|\; E^{t,t}_{k,l} \right]}{\Delta t}$$

$$\text{where } E^{t,t+\Delta t\,;\,t,+\Delta t}_{k,l+1} = \begin{bmatrix} \eta(t+\Delta t) - \eta(t) = 0, \\ \beta(t+\Delta t) - \beta(t) = 1 \end{bmatrix} \tag{17}$$

and finally the probability for the third event is

$$\lambda_{k,l}(t) = \lim_{\Delta t \to 0} \frac{P\left[E^{t,t+\Delta t\,;\,t,t+\Delta t}_{k+1,\,l+1} \;\middle|\; E^{t,t}_{k,l} \right]}{\Delta t}$$

$$\text{where } E^{t,t+\Delta t\,;\,t,t+\Delta t}_{k+1,\,l+1} = \begin{bmatrix} \eta(t+\Delta t) - \eta(t) = 1 \\ \beta(t+\Delta t) - \beta(t) = 1 \end{bmatrix} \tag{18}$$

The probabilities $P(E^{t,t}_{k,l})$ satisfy a system of differential equations which may be derived by considering the event and their associated conditional probabilities. Using the arguments of a stochastic process we get the following differential equation.

$$dP\,(E^{t,t}_{k,l}) = \lambda(t)\,P\,[E^{t,t}_{k-1,\,l-1}] + \mu(t)\,P\,[E^{t,t}_{k-1,\,l}] + \nu(t)\,P\,[E^{t,t}_{k,\,l-1}]$$

$$-\,[\lambda(t) + \mu(t) + \nu(t)]\,P\,[E^{t,t}_{k,l}]$$

$$\text{where, } \mu_{k,l}(t) = \mu(t),\ \nu_{k,l}(t) = v(t),\ \lambda_{k,l}(t) = \lambda(t). \tag{19}$$

It is assumed that the intensity functions are independent of K and l, in some relevant cases of the application of this theory for example water discharge/flood analysis. The generating function of equation (19) is given by

$$P(s_1, s_2, t) = \exp\,(s_1 s_2 - 1)\ W(t) + (s_1 - 1)\ U(t) + (s_2 - 1)\ V(t)$$

$$\text{where } W(t) = \int_0^t \lambda(t)\,dt;\ U(t) = \int_0^t \mu(t)\,dt;\ V(t) = \int_0^t \nu(t)\,dt. \tag{20}$$

The expression (20) is identified by Martiz [6] as the probability generating function of bivariate Poisson distribution with parameters $W(t)$, $U(t)$ and $V(t)$. Now the $P(E^{t,t}_{k,l})$ corresponding to generating function 2(0) can be wri:ten as

$$P\,(E^{t,t}_{k,l}) = \exp\,(-\,(U(t) + V(t) + W(t))\ \frac{a^k}{k!}\ \frac{b^l}{l!}$$

$$\sum_{r=0}^{s} \frac{(k)^{(r)}\,(l)^{(r)}}{a^r\;b^r}\;\frac{[W(t)]^r}{r!}$$

where $a = (U(t) - W(t)) > 0$, $b = (V(t) - W(t)) > 0$

$$(k)^{(r)} = K(k-l),\ldots,(k-r+l) \text{ and } s = \min(k, l) \tag{21}$$

4. Distribution Function of (Sup Z_k, Sup θ_l) and (Inf Z_k, Inf θ_l)

In this section the joint distribution function of two extremes of two dependent sequences of exceedances is derived. The number of these exceedances for the sequences Z_k's and θ_l's will be taken n and m respectively. These exceedances can occur in three ways as given in (15). The bivariate distribution function of (Sup Z_k, Sup θ_l) and (Inf Z_k. Inf θ_l) are given in the following theorems.

THEOREM 4.1. *If the sequences of random variables satisfy the two assumptions of Section 2, then the joint distribution function of maximum of two dependent sequences is given by*

$$F_{ts}(x,y) = \text{Exp}\,(-U(t) - V(t) + W(t) + W(t)\,F(x,y)$$

$$+ (U(t) - W(t))\,F(x) + (V(t) - W(t))\,F(y)) \tag{22}$$

Proof: The joint distribution function $F_{ts}(x, y)$ of the maximum of two dependent sequences is derived as the mathematical expectation of the following conditional probability

$$P[\sup_{\tau(k) \leqslant t} Z_k = x, \sup_{\tau(l) < t} \theta_l \leqslant y \,/\, \eta(t), \beta(t)] \tag{23}$$

The expression for $F_{ts}(x, y)$ is

$$F_{ts}(x, y) = E\,[P\,[\sup_{\tau(k) \leqslant t} Z_k < x, \sup_{\tau(k) < t} \theta_l \leqslant y]/\eta(t), \beta(t)]$$

$$\sum_{n=0}^{\infty} \sum_{m=0}^{\infty} P[\sup_{0 < k \leqslant n} Z_k \leqslant x, \sup_{0 < l \leqslant m} \theta_l \leqslant y] \cap P(E^{t,t}_{n,m}) \tag{24}$$

Under the assumptions of Section 2 the expression (24) can be written as

$$F_{t_s}(x, y) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} P\left[\sup_{0<k<n} Z_k \leqslant \bar{x}, \sup_{\sigma<l\leqslant m} \theta_l \leqslant \dot{y}\right] P\left[E_{n,m}^{t,t}\right] \tag{25}$$

Taking into consideration the three situations in which three exceedances in two variables occur, the expression for

$$P\left[\sup_{\min(m,n)} Z_k \leqslant x, \sup \theta_l \leqslant y\right] = \sum_{r=0}^{\min(m,n)} [F(x, y)]^r [F(x)]^{n-r} [F(y)]^{m-r} \tag{26}$$

where r is the number of exceedances when the exceedance in both the sequences occurs. Substituting (26) and (21) in (25) we obtain

$$F_{t_s}(x, y) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \sum_{r=0}^{\min(m,n)} [F(x, y)]^r [F(x)]^{n-r} [F(y)]^{m-r} \exp[-U(t) - V(t) + W(t)]$$

$$\frac{[U(t) - W(t)]^n}{n!} \frac{[V(t) - W(t)]^m}{m!} \sum_{r=0}^{\min(m,n)} \frac{[W(t)]^r}{r!}$$

$$\frac{n^{(r)}}{[U(t) - W(t)]^r} \frac{m^{(r)}}{[V(t) - W(t)]^r} \tag{27}$$

Interchanging and re-arranging the summations and the limits in (27) we obtain

$$F_{t_s}(x, y) = \exp\left[-U(t) - V(t) + W(t)\right] \sum_{r=0}^{\infty} \frac{[W(t) F(x, y)]^r}{r!}$$

$$\sum_{n=r}^{\infty} \frac{[U(t) - W(t)) F(x)]^{n-r}}{(n-r)!} \sum_{m=r}^{\infty} \frac{[V(t) - W(t)) F(y)]^{m-r}}{(m-r)!} \tag{28}$$

$$= \exp(-U(t) - V(t) + W(t)) \exp(W(t) F(x, y))$$

$$\exp((U(t) - W(t)) F(x)). \exp((V(t) - W(t)) F(y))$$

$$= \exp(-U(t) - V(t) + W(t) + W(t) F(x, y)$$

$$+ (U(t) - W(t)) F(x) + (V(t) - W(t)) F(y))) \tag{29}$$

The distribution function (29) satisfies the properties of distribution function and we can get marginals from it as follows

$$F_{ts}(x, \infty) = \exp(-U(t)(1 - F(x))$$

and

$$F_{ts}(\infty, y) = \exp(-V(t)(1 - F(y))$$

THEOREM 4.2. *The joint distribution function* $F_{tt}(x, y)$ *of minimum observations of two dependent sequences is given by*

$$F_{tl}(x, y) = 1 + \exp(-U(t)) - \exp(-U(t)(1 - P(x))$$
$$+ \exp(-V(t)) - \exp(-V(t)(1 - P(y)) - \exp(-U(t)$$
$$- V(t) + W(t)) + \exp((-U(t) - V(t) + W(t)$$
$$+ W(t) P(x, y)) + P(Y)(V(t) - W(t)) + P(x)(U(t)$$
$$- W(t)) \tag{30}$$

where $P(x) = 1 - F(x)$, $P(y) = 1 - F(y)$,

$$P(x, y) = 1 - F(x) - F(y) + F(x, y)$$

Proof:— Same as in Theorem 4.1.

As Z_k and θ_l are exponentially distributed, therefore, their joint bivariate distribution function should be a bivariate exponential distribution. As it is well known that infinite number of distribution functions can exist corresponding to given marginals then question arises, what should be the form of bivariate exponential distribution function $F(x, y)$. The selection of bivariate distribution function depends much upon the estimation of its parameters with the help of given data. Due to complicacy of calculations it is not easy to find out the estimate for dependence and other parameters for all the bivariate distribution functions. It can be done only by relating (22) and (30) to different correlation coefficients or by the method of quadrants. On the basis of calculations for the estimation of dependence parameters, it is observed that Gumbel type II distribution, due to Gumbel [2], is an appropriate form for $F(x, y)$ in (30). Substituting $F(x)$, $F(y)$ and $F(x, y)$ in (30) $F_{t_s}(x,y)$ is

$$F_{t_s}(x, y) = \exp(-U(t) \exp(-\alpha_1 x) - V(t) \exp(-\alpha_2 y) + W(t) \exp(-((\alpha_1 X)^m + (\alpha_2 y)^m)^{1/m})) \quad (31)$$

and its marginals are

$$F_{t_s}(x) = \exp(-U(t) \exp(-\alpha_1 X))$$

$$F_{t_s}(y) = \exp(-V(t) \exp(-\alpha_2 y))$$

where m is the dependence parameter and α_1 and α_2 are the exponential distribution parameters.

It is worthwhile to mention that the functional form of the distribution function (31) is similar to the asymptotic bivariate extreme value distribution given by Gumbel and Mustafi [4].

5. Illustration

The application of the distribution function (31) for the data of water discharge of river Narmada at Mortakka and Gardeshwar is considered. Data are taken only for a partial duration series starting from July 1st to October 28 for every year at both the stations during the period 1948 to 1977. The expression (31) indicates that the estimation and prediction of floods is possible through it, if one has the estimate of $U(t)$, the average number of exceedances in $(0, t)$ at Mortakka (x), $V(t)$ the average number of exceedances in $(0, t)$ at Gardeshwar (y) and $W(t)$ the average number of exceedances in $(0, t)$ at both Mortakka (x) and Gardeshwar (y). The estimation of parameters of exponential distribution functions $F(x)$, $F(y)$ and $F(x, y)$ is also required.

5.1 Estimation of U (t), V (t) and W (t)

The partial duration series of water discharges values is divided into eight equal intervals of fifteen days period for every year. The base levels x_0 and y_0 are taken as 150,000 c.f.s. (cubic feet per second) for both the stations and only isolated water discharges greater than 150,000 c.f.s. were taken into consideration. Zelenhasic [2] had expressed the average number of exceedances as a finite Fourier series based on the analysis of data. Using the same method, the expressions for $V(t)$, $U(t)$ and $W(t)$, the expected number of exceedances at Mortakka and Gardeshwar and at both are as follows :

$$\begin{aligned} U(t) = & - 5.2142856 + 14,0396824\, t + 6.5357143\, t^2 \\ & - 0.6944444\, t^3 - 0.0181075 \cos 45\, t \\ & - 0.0106729 \sin 45\, t + 0.0238095 \cos 90\, t \\ & + 0.0583746 \sin 135\, t + 0.0111111\, (-1)^t \\ & + 0.0738095 \sin 90\, t + 0.0554091 \cos 135\, t \end{aligned} \quad (32)$$

$$\begin{aligned} V(t) = & - 1.5000000 + 5.7871573\, t + 7.5259740\, t^2 \\ & - 0.7373737\, t^3 - 0.0157369 \cos 45\, t \\ & - 0.0110213 \sin 45\, t + 0.0113997 \cos 90\, t \\ & + 0.0838383 \sin 90\, t + 0.0648711 \cos 135\, t \\ & + 0.0288776 \sin 135\, t + 0.0086580\, (-1)^t \end{aligned} \quad (33)$$

$$\begin{aligned} W(t) = & - 5.2142856\, t + 14.0396824\, t \\ & + 6.5957143\, t^2 - 0.6944444\, t^3 \\ & - 0.0181075 \sin 45\, t + 0.0106728 \cos 45\, t \\ & + 0.0238095 \sin 90\, t + 0.0738095 \cos 90\, t \\ & + 0.0583746 \sin 135\, t + 0.0554091 \cos 135\, t \\ & + 0.0106301\, (-1)^t. \end{aligned} \quad (34)$$

5.2 Estimate of α_1, α_2

The next step is to evaluate the parameter α_1, α_2 of (31). The parameters α_1 and α_2 are estimated by the well known method of maximum likelihood method and presented in Table. 3.1.

TABLE 5.1

Station	$E(Z_k)$	$\alpha = (E(Z_k))^{-1}$
Mortakka (α_1)	241037.93	4.1487246×10^{-6}
Gardeshwar (α_2)	343391.88	2.9121247×10^{-6}

The marginal distribution functions (31) can be calculated now for the different values of x and y. The observed and corresponding theoretical distribution functions of the flood peak exceedances are presented in Fig. 1 and 2. The distribution functions of the largest flood exceedances for both stations are presented in Fig. 3 and 4. A fairly good

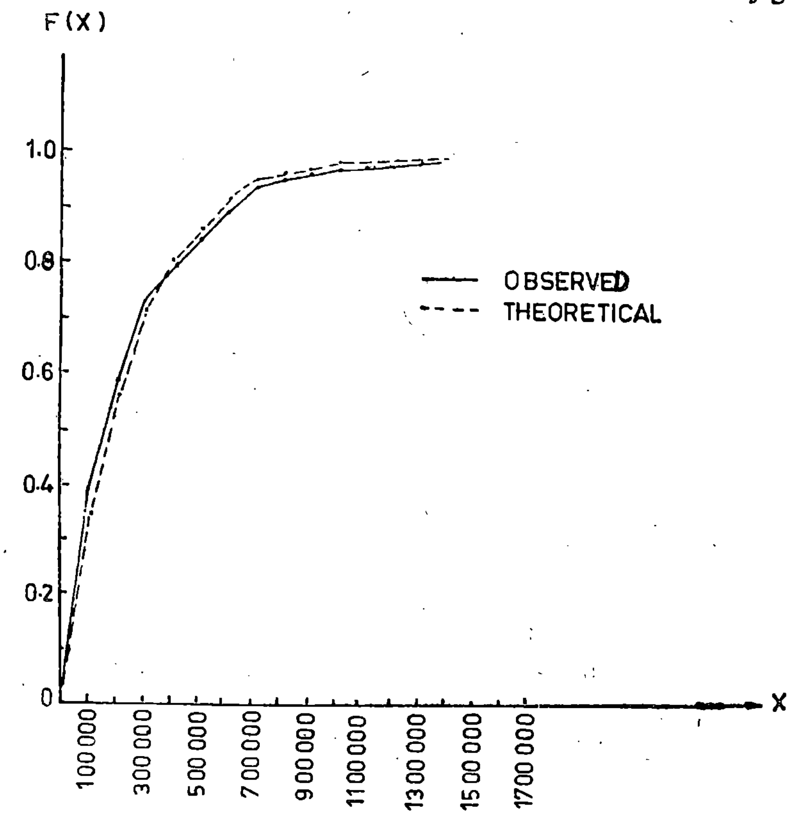

Fig. 1. Observed and Theoretical Distribution Functions of Exceedances for the Narmada River at Mortakka for 120 Days Period

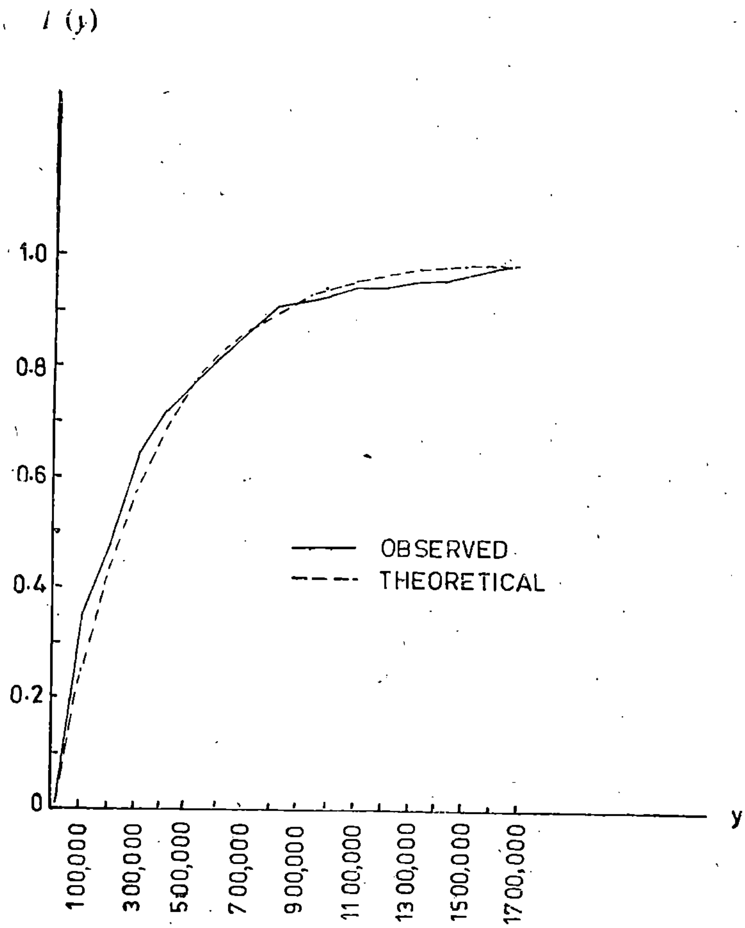

Fig. 2. Observed and Theoretical Distribution Functions of Exceedances for the Narmada River at Gardeshwar for 120 days Period

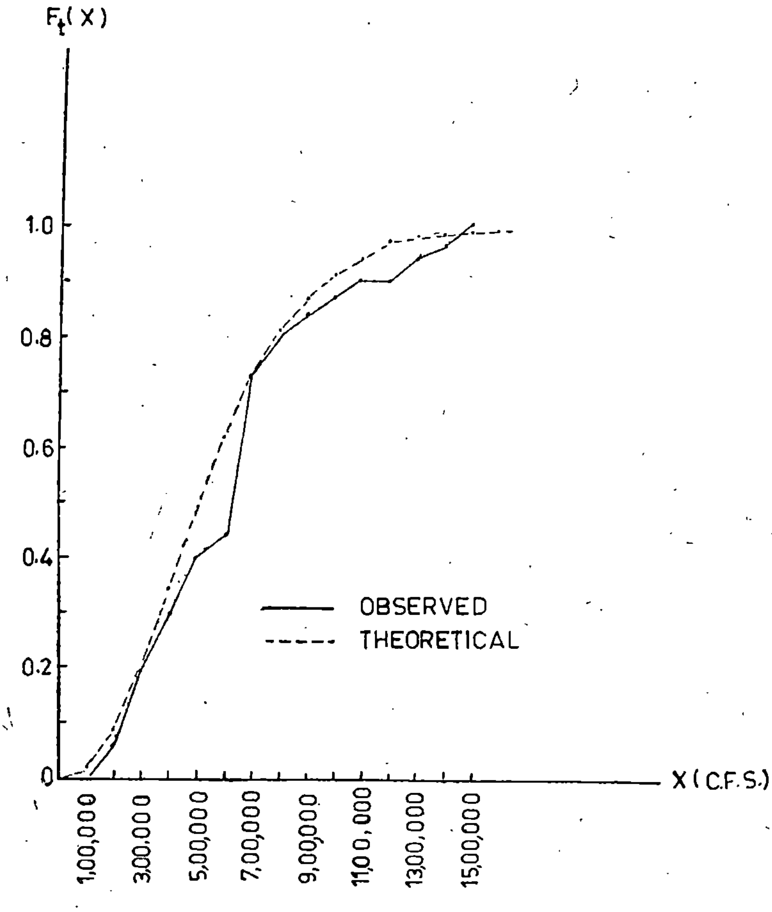

Fig. 3. Observed and Theoretical Distribution Functions of the Maximum Flood Peak Exceedances for the Narmada River at Mortakka

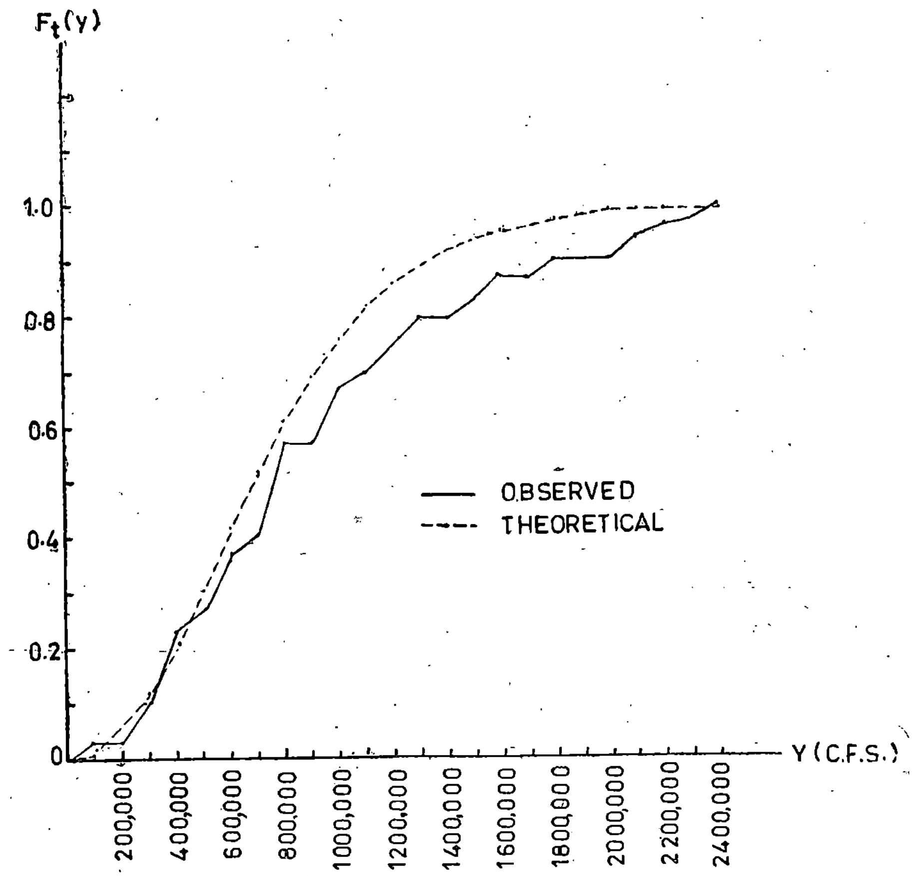

Fig. 4. Observed and Theoretical Distribution Functions of the Maximum Flood Peak Exceedances for Narmada River at Gardeshwar

agreement between theoretical and observed results is seen. However, return period of a particular value can be calculated by the formula $T(x) = (1 - F(x))^{-1}$.

5.3 *Estimate of Dependence Parameter (m)*

The dependence parameter '*m*' may be estimated by relating it to the medial correlation coefficient. The medial correlation coefficient (ν) can be expressed in the form of bivariate distribution function $F(x, y)$ and its marginals $F(x)$ and $F(y)$ as follows :

$$\nu = (F(\bar{x}, \bar{y}) - F(\bar{x}) F(\bar{y}) \quad (35)$$

where x and y are the medians. After substituting the value of distribution functions in the expression (5.4) we have

$$F_{ts}(\bar{x}, \bar{y}) = 4[1 - e^{-\alpha_1 \bar{x}} - e^{-\alpha_2 \bar{y}} + \exp - ((\alpha_1 \bar{x})^m + (\alpha_2 \bar{y})^m)^{1/m}]^{-1} \quad (36)$$

One can find out the value of ν corresponding to each value of $b = (1/m)$ after substituting the values of α_1, α_2, x and y in (36). The values of x and y are 173004.50 and 243313.00 c.f.s. (cubic feet/second) respectively and the values of α_1 and α_2 are substituted from Table 5.1.

After estimating all the constants in (36) the values of ν corresponding to different values of b are calculated and presented in Table 5.2.

TABLE 5.2.

The value of dependence parameter (b) = 1/m	*Medial correlation coefficient*
1.0	0.039966
0.9	0.113625
0.8	0.234780
0.7	0.334973
0.6	0.436276
0.5	0.538144
0.4	0.640075
0.3	0.741542
0.2	0.842173
0.1	0.941486

The next step is to calculate the value of ν, the medial correlation coefficient which is given by $\nu = 2d/n - 1$ by using graphical procedure. Here n = total no. of points which are plotted in a scatter diagram as given in Fig. 5 and d = the total number of points in positive quadrant and its opposite quardant. From Fig. 5 we note that $n = 118$ and $d = 86$

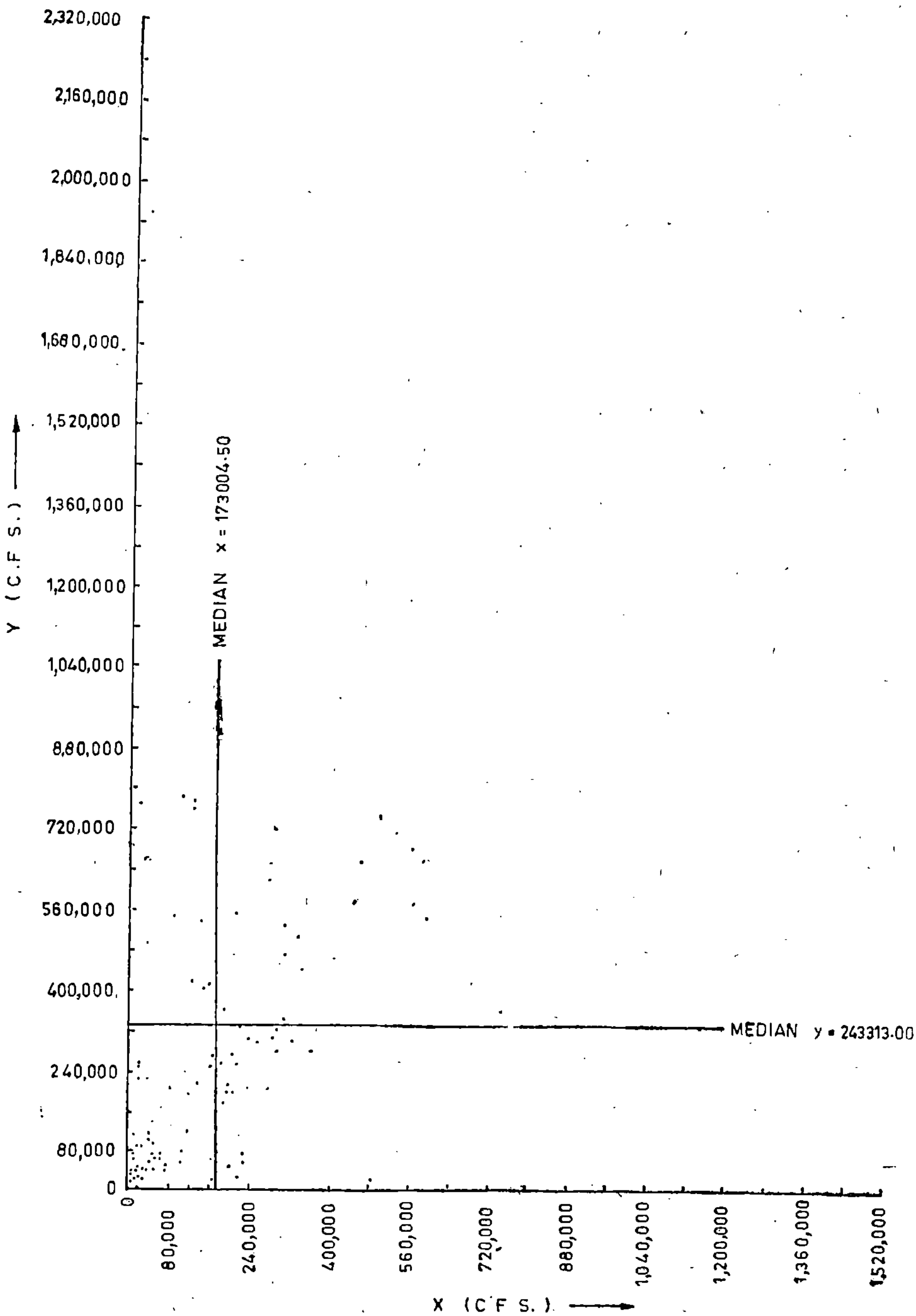

Fig. 5. Scatter Diagram of Paired (*X*, *Y*)

hence $\nu = (171/118) - 1 = 1.457627 - 1 = 0.457627$. Now Table 5.2 can be used to interpolate b for the above value of ν. The corresponding 'b' comes to be 0.557 (approximately).

5.4 Return Period in Bivariate Case

The return period in bivariate case is defined by

$$T(x, y) = 1/P(X \geqslant x, Y \geqslant y) \tag{37}$$

Using the estimates of α_1, α_2, $U(t)$, $V(t)$, $W(t)$, m and $F_{t_1}(x)$, $F_{t_2}(y)$ and $F_{t_3}(x, y)$ in (37) the return period is calculated for some value of x and y and entered in Table 5.3.

TABLE 5.3

Water discharge (c.f.s.)	*Return period (years)*	*Water discharge (c.f.s.)*	*Return period (years)*
100,000	1.0417	800,000	8.9438
200,000	1.1373	900,000	14.7245
300,000	1.3453	1000,000	24.6090
400,000	1.7340	1100,000	41.5037
500 000	2.4143	1200,000	70.3229
600,000	3.5785	1300,000	119.2733
700,000	5.5614	1400,000	205.9858

6. Conclusions

A fairly good agreement between theoretical and observed results is observed in case of estimation of exceedances and fitting of distribution functions of exceedances. The distribution function (31) is an exact expression and can be used for prediction of water discharges at one location if the water discharge of other is known. If two variables of study are water discharge and water level, then the knowledge of one can be utilized to predict the other.

ACKNOWLEDGEMENT

The author acknowledges gratefully Central Water Commission for providing the required data and Dr. O. P. Srivastava for his valuable suggestions.

REFERENCES

[1] Gumbel, E.J. (1958). *Statistics of Extremes.* Columbia University Press, New York.

[2] Gumbel, E.J. (1960). Bivariate exponential distributions. *J. Am, Statist. Ass.* **55** : 698-707.

[3] Gumbel, E.J. and N. Goldstein (1964). Analysis of empirical bivariate extremal distributions. *J. Amer. Statist. Assoc.* 59 : 794-816.

[4] Gumbel, E.J. and C.K. Mustafi (1967). Some analytical properties of bivariate extreme value distributions, *J. Amer. Statist. Assoc.* 62 : 569-588.

[5] Gupta, V.K., Duckstein, L. and Peebles. R.W. (1976). On the joint distribution of the largest flood and its time of occurrence. *Wat. Resour. Res.* 12 (2):295-304.

[6] Martiz, J.S. (1952). Note on a certain family of discrete distributions. *Biometrika* 39 : 196-198.

[7] Nair, K.A. (1976). Bivariate extreme value dittributions. *Comm. in Statist.* **5** : 575-581.

[8] Sibuya, M. (1960). Bivariate extreme statistics. *Ann. Inst. Statist. Math.* **11**: 195.210.

[9] Sibuya, M. (1962). On exponential and other random variable generators. *Ann. Inst. Statist. Math.* **13** : 231-237.

[10] Singh, N.P. (1982). Theory of Extreme Values and its Application to Flood Control. *Unpublished Ph. D. Thesis,* HAU, Hisar, (India).

[11] Singh, N.P. (1985). Stochastic analysis of floods. *Proc. Indian Natn. Sci. Acad.,* **51** : 351-357.

[12] Tiago de Oliveira, J. (1962/63). Structure theory of bivariate extremes : Extensions: *Estudos de Math. Estat. Econom.* **7** : 165-175.

[13] Tiago de Oliveira, J. (1971). A new model of bivariate extremes : Statistical decision. In : *Studi Prob. Stat. ricerca oper. in onoredi G. Pampiji.* Oderisi, Gubbi. pp-1-13.

[14] Tiago de Oliveira, J. (1976). Bivarlate extremes. Extensions. *Proc. 40th Session ISI.* Warsaw.

[15] Todorovic, P. (1970). On some problems involving random number of random variables. *Ann. Math. Statist.* 41 : 1059-1063,

[16] Todorovic, P. and Zelenhasic, E. (1970). A stochastic model flood analysis. *Wat. Resour. Res.* 6 (6) : 1641-1658.

[17] Todorovic, P. (1971). On extreme problems in hydrology. Paper presented at joint statistics meeting. *An. Statist. Ass. and Inst. of Math. Statist.* Colo. State Univ. Foot Collins.

[18] Todorovic, P. and Rousselle, J. (1971). Some problems of flood analysis. *Wat. Resour. Res.* 7 (5) : 1144-1150.

[19] Todorovic, P. and Woolhiser, D. A. (1972). On the time when the extreme flood occurs. *Wat. Resour. Res.* 8 (8) : 1433-1438.

[20] Zenlenhasic, E. (1970). Theoretical probability distribution for flood peaks, Hydrol. Rep. 42, Colo. Stat. Univ. Fert. Collins.