Stochastic Environmental Research and Risk Assessment 17 (2003) 42–57 © Springer-Verlag 2003
DOI 10.1007/s00477-003-0125-9

Return period of bivariate distributed extreme hydrological events

J. T. Shiau

Abstract. Extreme hydrological events are inevitable and stochastic in nature. Characterized by multiple properties, the multivariate distribution is a better approach to represent this complex phenomenon than the univariate frequency analysis. However, it requires considerably more data and more sophisticated mathematical analysis. Therefore, a bivariate distribution is the most common method for modeling these extreme events. The return periods for a bivariate distribution can be defined using either separate single random variables or two joint random variables. In the latter case, the return periods can be defined using one random variable equaling or exceeding a certain magnitude and/or another random variable equaling or exceeding another magnitude or the conditional return periods of one random variable given another random variable equaling or exceeding a certain magnitude. In this study, the bivariate extreme value distribution with the Gumbel marginal distributions is used to model extreme flood events characterized by flood volume and flood peak. The proposed methodology is applied to the recorded daily streamflow from Ichu of the Pachang River located in Southern Taiwan. The results show a good agreement between the theoretical models and observed flood data.

Keywords: Return period, Bivariate extreme value distribution, Gumbel distribution

1 Introduction

The return period for extreme hydrological events, such as floods and droughts, is a common criterion employed in the design of hydraulic structures and water supply systems. Traditionally, a univariate distribution is used to describe the extreme hydrological phenomena, such as flood peak or rainfall intensity. However, a complex phenomenon is often characterized by multiple aspects. For example, Goel et al. (1998) and Yue et al. (1999) indicated that flood flows appear as multivariate events described by peak, volume and duration. Hence, the best approach for analyzing such complex events is through the joint distribution of several random variables, considering the correlations among them. Anderson and Nadarajah (1993) and Anderson et al. (1994) used the multivariate extreme

J. T. Shiau
Department of Water Resources and Environmental Engineering,
Tamkang University, 151 Ying-chuan Road,
Tamsui, 251, Taiwan, ROC
e-mail: jtshiau@mail.tku.edu.tw

The author wishes to thank the two anonymous reviewers for their constructive comments that improving the quality of this work.

model to study dependent structure between the maximum reservoir water level and environmental factors including rainfall, catchment wetness, snowmelt, wind, and reservoir fullness. Although a multivariate modeling approach offers improved understanding and application, it requires considerably more data and more sophisticated mathematical analysis. A bivariate distribution is the more common and easier model for describing complex extreme hydrological events. Gupta et al. (1976) developed an expression for the joint distribution function of the largest flood peak and its time of occurrence. Singh and Singh (1991) derived a bivariate probability density function with exponential marginal distributions to study rainfall intensity and the corresponding depth. Bacchi et al. (1994) used the bivariate distribution with marginal exponential distributions to model extreme rainfall duration and severity. Goel et al. (1998) used the bivariate normal distribution to represent the joint distribution of flood peaks and volumes based on a partial duration series. Yue et al. (1999) used the Gumbel mixed model, the bivariate extreme value distribution with the standard Gumbel marginal distributions, to represent the joint probability distribution of flood peaks and volumes and the joint probability distribution of flood volumes and durations. Yue (2001) used bivariate logistic distributions with the standard Gumbel marginal distributions to model the El Niño maximum intensity and magnitude, El Niño magnitude and duration, and El Niño maximum intensity and duration. Nadarajah and Withers (2001) studied the joint behavior of the climate annual maxima for New Zealand by the bivariate extreme value model.

The bivariate extreme hydrological event distributions and corresponding return periods have been extensively studied. However, little attempt has been made to interpret the return period characteristics defined using bivariate distribution and the relationships between the return periods defined using univariate and bivariate distributions. Fitting a specific bivariate distribution to a specific extreme hydrological phenomenon was not the major purpose in this study. Instead, after selecting a bivariate distribution to model an extreme hydrological event, the return period and related probability distributions are derived and the relationships between the univariate and bivariate distribution return periods are investigated.

In this study, a theoretical derivation of the univariate and bivariate distribution return periods is introduced based on the stochastic process concept. Under the assumption that the bivariate extreme value distribution with the Gumbel marginal distributions can be used to model the flood peak and volume, the flood peak and volume return periods and the joint return period for the flood peak and volume are derived. The associated properties and the relationships between the return periods defined by the univariate and bivariate distributions are also discussed. A recorded daily streamflow from Pachang River located in Southern Taiwan is used as a case study to illustrate the proposed methodology.

2
Derivation of return period

2.1
Return period for a univariate distributed annual maximum series

According to the definition and notation used by Chow et al. (1988), the return period for a univariate distributed annual maximum series is described below. Suppose that an extreme event occurs if the random variable X is greater than or equal to some magnitude x_T. The recurrence interval T_X is defined as the time period between occurrences for the event $X \geq x_T$. The return period for the event $X \geq x_T$ is the expected value of T_X, denoted by $E(T_X)$ in this study. The return period for an event of a given magnitude is thus defined as the average recurrence interval between events equaling or exceeding a specific magnitude.

The return period for the event $X \geq x_T$ can be related to the probability of occurrence for such events in the following way. It is assumed that the probability of occurrence for the event $X \geq x_T$ in any year is $P(X \geq x_T)$. Because the annual maximum series observation in each year is independent, the probability of a recurrence interval T_X is the product of the probabilities for $T_X - 1X < x_T$ events followed by one $X \geq x_T$ event, that is $P(X < x_T)^{T_X-1}P(X \geq x_T)$. The expected value for T_X is then given by

$$E(T_X) = \sum_{T_X=1}^{\infty} T_X P(X < x_T)^{T_X-1} P(X \geq x_T) = \frac{1}{P(X \geq x_T)} = \frac{1}{1 - P(X < x_T)} \quad (1)$$

The above equation is the return period for an extreme event described by a single random variable and based on the annual maximum series. Obviously, the return period depends on the distribution of the selected random variables, that is, the longer the return period, $E(T_X)$, the less the frequency, $P(X \geq x_T)$, and the larger the magnitude of the random variable, x_T.

2.2
Return period for a univariate distributed partial duration series

The return period for a univariate distributed partial duration series can be derived based on the stochastic process concept under the assumption that the events are independently and identically distributed. The occurrences of independent event X for different magnitudes x are shown in Fig. 1. Let L denote the time period between any two successive events without consideration of the magnitude, called the interarrival time in this study. The events with a magnitude equal to or greater than any value $x, X \geq x$, are denoted by • in Fig. 1, while the events with a magnitude less than any value $x, X < x$, are denoted by ∘ in Fig. 1. Hence, the time period between two events with magnitudes equaling or exceeding any magnitude x, namely the recurrence interval, denoted by T_X, is equal to the summation of the interarrival time for all events between them. This relationship is expressed as

$$T_X = \sum_{i=1}^{N_X} L_i \quad (2)$$

where L_i is the interarrival time between any two successive events and N_X is the number of events until the occurrence of the next event $X \geq x$.

Obviously, the recurrence interval T_X is also a random variable and its expected value is called the return period for $X \geq x$. Hence

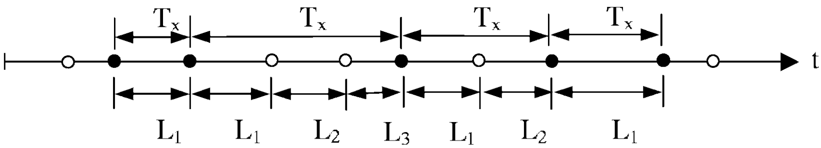

Fig. 1. The occurrences of event X, ∘ denoting $X < x$, and • denoting $X \geq x$

$$E(T_X) = E\left(\sum_{i=1}^{N_X} L_i\right) = E(N_X)E(L_i) \tag{3}$$

If the interarrival time, L_i, is assumed to have an identical and independent distribution, then the above equation can be simplified to

$$E(T_X) = E(N_X)E(L) \tag{4}$$

The magnitude of an event is inversely related to its frequency of occurrence, with very severe events occurring less frequently than more moderate events. N_X is therefore a random variable that depends on the distribution of X. Let $F(x)$ denote the cumulative distribution function of X, i.e. $F(x) = P(X \leq x)$. Because X is considered as a continuous random variable, therefore, $P(X \leq x) = P(X < x)$ (Stone, 1996). The probability of an event with a magnitude equaling or exceeding x, namely the probability of $X \geq x$, is

$$P(X \geq x) = 1 - P(X < x) = 1 - F(x) \tag{5}$$

The recurrence interval for event $X \geq x$ is $N_X - 1 \cdot X < x$ followed by one $X \geq x$. That is, N_X has a geometric distribution with parameter $1 - F(x)$ and its probability mass function is given by

$$P(N_X = n) = P(X < x)^{n-1}P(X \geq x) = F(x)^{n-1}[1 - F(x)], \; n = 1, 2, 3, \ldots \tag{6}$$

The expected value of N_X is

$$E(N_X) = \frac{1}{P(X \geq x)} = \frac{1}{1 - F(x)} \tag{7}$$

The return period for events with a magnitude equal to or greater than x therefore becomes

$$E(T_X) = E(N_X)E(L) = \frac{E(L)}{1 - F(x)} \tag{8}$$

Shiau and Shen (2001) applied this method to derive the return period of hydrological droughts of differing severity. The above procedure can be applied to the partial duration series as well as the annual series. When applying this to the annual maximum series, $E(L) = 1$ year because each event is selected from one year. Hence, Eq. (8) and Eq. (1) are identical.

2.3
Return period for a bivariate distributed partial duration series

Without a doubt, the above procedure can be directly applied to a bivariate distributed event. However, the correlation structure between random variables must be considered. It is assumed that an extreme hydrological event can be characterized using two random variables X and Y and the correlation coefficient ρ between them. The return periods for a bivariate distributed event can be derived in two ways. The first method treats each random variable separately, namely this method derives the return period for random variable X and the return period for random variable Y. The correlation between X and Y is not considered if X and Y are independent or treated in other ways. The occurrence of $X \geq x$ and the occurrence of $Y \geq y$ are shown in Fig. 2(a) and (b), respectively. Apparently, the condition $X \geq x$ and $Y \geq y$ may not occur simultaneously in one event. The return periods for $X \geq x$ and $Y \geq y$ are therefore derived separately and given by

$$E(T_X) = \frac{E(L)}{1 - F(x)} \tag{9a}$$

$$E(T_Y) = \frac{E(L)}{1 - F(y)} \tag{9b}$$

where $E(L)$ is the expected value of interarrival time; $E(T_X)$ and $E(T_Y)$ are the return periods of $X \geq x$ and $Y \geq y$, respectively; $F(x)$ and $F(y)$ are the cumulative distribution functions of X and Y, respectively.

The above univariate frequency analysis is useful when only one extreme random variable events is significant in the design criterion or these two random variables are less dependent. However, a separate analysis of random variables X and Y cannot reveal the significant correlation relationship between them if the correlation is important information in the design criterion. The second method, on the other hand, considers the random variables X and Y jointly. This can be done either by defining the joint return periods for X and Y or by defining the conditional return period for X given Y or vice versa. In this study, the joint return periods for X and Y were defined in two cases: the return period for $X \geq x$ or $Y \geq y$ and the return period for $X \geq x$ and $Y \geq y$. Both methods are given below

$$E(T_{XY}) = \frac{E(L)}{P(X \geq x \text{ or } Y \geq y)} = \frac{E(L)}{1 - F(x,y)} \tag{10}$$

$$E(T'_{XY}) = \frac{E(L)}{P(X \geq x \text{ and } Y \geq y)} = \frac{E(L)}{1 - F(x) - F(y) + F(x,y)} \tag{11}$$

Both of the above equations are defined using the magnitudes of X and Y simultaneously. These relationships imply that various combinations of values, x and y, can result in the same return period. In addition, the return period can also be defined by the event for X given $Y \geq y$ or event for Y given $X \geq x$ that are called the conditional return period for X given $Y \geq y$ and the conditional return period for Y given $X \geq x$, respectively.

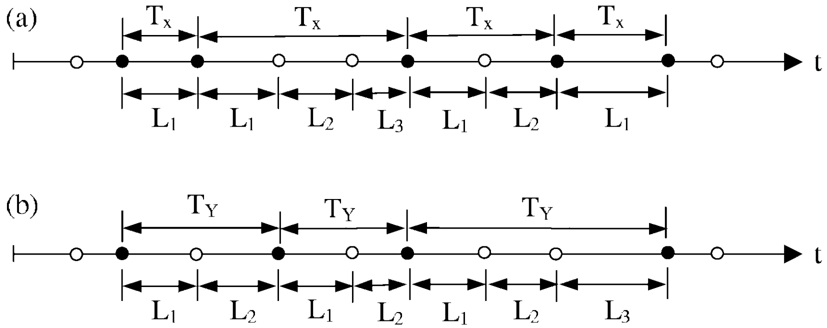

Fig. 2. The occurrences of events, (a) ○ denoting $X < x$, ● denoting $X \geq x$, (b) ○ denoting $Y < y$, ● denoting $Y \geq y$

$$E(T_{X|Y\geq y}) = \frac{E(T_Y)}{P(X \geq x, Y \geq y)} = \frac{E(L)}{1-F(y)}\frac{1}{1-F(x)-F(y)+F(x,y)} = \frac{E(L)}{[1-F(y)][1-F(x)-F(y)+F(x,y)]} \tag{12a}$$

$$E(T_{Y|X\geq x}) = \frac{E(T_X)}{P(X \geq x, Y \geq y)} = \frac{E(L)}{1-F(x)}\frac{1}{1-F(x)-F(y)+F(x,y)} = \frac{E(L)}{[1-F(x)][1-F(x)-F(y)+F(x,y)]} \tag{12b}$$

where $E(T_X)$ and $E(T_Y)$ are the return periods for $X \geq x$ and $Y \geq y$, respectively.

Because the return period is the expected period for the recurrence interval, the conditional distribution of Y given $X \geq x$ and the conditional distribution of X given $Y \geq y$ can also be derived in the following.

$$F(y|X \geq x) = \frac{P(X \geq x, Y < y)}{P(X \geq x)} = \frac{F(y) - F(x,y)}{1-F(x)} \tag{13a}$$

$$F(x|Y \geq y) = \frac{P(X < x, Y \geq y)}{P(Y \geq y)} = \frac{F(x) - F(x,y)}{1-F(y)} \tag{13b}$$

3
Return period for a bivariate extreme value distribution

Kotz and Nadarajah (2000) provided a comprehensive survey of the bivariate and multivariate extreme value models. However, not every bivariate distribution can be applied to hydrological processes. Gumbel and Mustafi (1967) proposed two general forms for bivariate extreme value distributions in term of the marginal distributions. The cumulative distribution functions for the bivariate extreme value distribution for Type A and Type B have the following forms, respectively.

$$F(x,y) = F(x)F(y)\mathrm{e}^{\left\{-\theta\left[\frac{1}{\ln F(x)}+\frac{1}{\ln F(y)}\right]^{-1}\right\}} \tag{14}$$

where *F(x)* and *F(y)* are the marginal distributions for random variables X and Y, respectively; and θ is a parameter, $0 \leq \theta < 1$.

$$F(x,y) = \mathrm{e}^{-[(-\ln F(x))^m+(-\ln F(y))^m]^{\frac{1}{m}}} \tag{15}$$

where m is a parameter, $m \geq 1$.

Because the marginal distributions of the above two bivariate distributions are the extreme value distributions, i.e. the Gumbel distribution, many researchers used them to represent the extreme hydrological events. For instance, Yue et al. (1999) used the Type A distribution, called the Gumbel mixed model, to represent flood events. Yue (2001) used the Type B distribution, called the bivariate logistic distributions with standard Gumbel marginal distributions, to model El Niño events. The limitation of the Type A distribution is that the correlation of the random variables must lies between 0 and 2/3. Hence, the Type B distribution was used as an example to illustrate how to derive the return period and associated probability properties for a bivariate distribution in this study.

In Eq. (15), m is the parameter describing the correlation between X and Y. The estimators for m is given by

$$m = \frac{1}{\sqrt{1-\rho}} \tag{16}$$

where ρ is the correlation coefficient for X between Y.

The marginal distributions for X and Y take the Type I extreme value distribution, namely the Gumbel distribution, expressed as

$$F(x) = e^{-e^{-\frac{x-\beta_x}{\alpha_x}}} \tag{17a}$$

$$F(y) = e^{-e^{-\frac{y-\beta_y}{\alpha_y}}} \tag{17b}$$

where $\alpha_x, \beta_x, \alpha_y, \beta_y$ are parameters.

Hence, the cumulative distribution function for the bivariate extreme value distribution becomes

$$F(x,y) = e^{-\left[e^{-\frac{m(x-\beta_x)}{\alpha_x}}+e^{-\frac{m(y-\beta_y)}{\alpha_y}}\right]^{\frac{1}{m}}} \tag{18}$$

Kotz et al. (2000) stated that such a distribution is also called the logistic model, and Yue (2001) called it the bivariate logistic distributions with Gumbel marginal distributions.

The probability density function, therefore, is

$$\begin{aligned} f(x,y) = \frac{1}{\alpha_x\alpha_y} e^{-\left[e^{-\frac{m(x-\beta_x)}{\alpha_x}}+e^{-\frac{m(y-\beta_y)}{\alpha_y}}\right]^{\frac{1}{m}}} \left[e^{-\frac{m(x-\beta_x)}{\alpha_x}} + e^{-\frac{m(y-\beta_y)}{\alpha_y}}\right]^{\frac{1}{m}-2} \\ \left[\left(e^{-\frac{m(x-\beta_x)}{\alpha_x}} + e^{-\frac{m(y-\beta_y)}{\alpha_y}}\right)^{\frac{1}{m}} + m - 1\right] e^{-\frac{m(x-\beta_x)}{\alpha_x}-\frac{m(y-\beta_y)}{\alpha_y}} \end{aligned} \tag{19}$$

Hence, the return period of random variable X equal to or greater than x is

$$E(T_X) = \frac{E(L)}{1-e^{-e^{-\frac{x-\beta_x}{\alpha_x}}}} \tag{20}$$

where *E(L)* is the average interarrival time for the extreme hydrological events under consideration.

The conditional distribution for Y given $X \geq x$ is

$$F(y|X \geq x) = \frac{e^{-e^{-\frac{y-\beta_y}{\alpha_y}}} - e^{-\left[e^{-\frac{m(x-\beta_x)}{\alpha_x}}+e^{-\frac{m(y-\beta_y)}{\alpha_y}}\right]^{\frac{1}{m}}}}{1-e^{-e^{-\frac{x-\beta_x}{\alpha_x}}}} \tag{21}$$

The conditional return period for Y given $X \geq x$ is

$$E(T_{Y|X\geq x}) = \frac{E(L)}{\left[1 - e^{-e^{-\frac{x-\beta_x}{\alpha_x}}}\right]\left[1 - e^{-e^{-\frac{x-\beta_x}{\alpha_x}}} - e^{-e^{-\frac{y-\beta_y}{\alpha_y}}} + e^{-\left[e^{-\frac{m(x-\beta_x)}{\alpha_x}} + e^{-\frac{m(y-\beta_y)}{\alpha_y}}\right]^{\frac{1}{m}}}\right]} \tag{22}$$

Similarly, the return period for random variable Y, the conditional distribution for X given $Y \geq y$, and the conditional return period for X given $Y \geq y$ are given by

$$E(T_Y) = \frac{E(L)}{1 - e^{-e^{-\frac{y-\beta_y}{\alpha_y}}}} \tag{23}$$

$$F(x|Y \geq y) = \frac{e^{-e^{-\frac{x-\beta_x}{\alpha_x}}} - e^{-\left[e^{-\frac{m(x-\beta_x)}{\alpha_x}} + e^{-\frac{m(y-\beta_y)}{\alpha_y}}\right]^{\frac{1}{m}}}}{1 - e^{-e^{-\frac{y-\beta_y}{\alpha_y}}}} \tag{24}$$

$$E(T_{X|Y\geq x}) = \frac{E(L)}{\left[1 - e^{-e^{-\frac{y-\beta_y}{\alpha_y}}}\right]\left[1 - e^{-e^{-\frac{x-\beta_x}{\alpha_x}}} - e^{-e^{-\frac{y-\beta_y}{\alpha_y}}} + e^{-\left[e^{-\frac{m(x-\beta_x)}{\alpha_x}} + e^{-\frac{m(y-\beta_y)}{\alpha_y}}\right]^{\frac{1}{m}}}\right]} \tag{25}$$

The joint return periods for event $X \geq x$ or $Y \geq y$ and event $X \geq x$ and $Y \geq y$ are

$$E(T_{XY}) = \frac{E(L)}{1 - e^{-\left[e^{-\frac{m(x-\beta_x)}{\alpha_x}} + e^{-\frac{m(y-\beta_y)}{\alpha_y}}\right]^{\frac{1}{m}}}} \tag{26}$$

$$E(T'_{XY}) = \frac{E(L)}{1 - e^{-e^{-\frac{x-\beta_x}{\alpha_x}}} - e^{-e^{-\frac{y-\beta_y}{\alpha_y}}} + e^{-\left[e^{-\frac{m(x-\beta_x)}{\alpha_x}} + e^{-\frac{m(y-\beta_y)}{\alpha_y}}\right]^{\frac{1}{m}}}} \tag{27}$$

The parameters $\alpha_x, \beta_x, \alpha_y, \beta_y$ and the average interarrival time *E(L)* must be estimated from the data used, that is, these parameters are event-specific.

4 Applications

4.1 Data used

Flooding in Taiwan is a common phenomenon and often causes considerable economic and life losses. As flooding can be represented by an extreme value distribution (Gumbel, 1958), hence, flood flow is used as an example to illustrate the proposed methodology. The Pachang River, with an 81 Km in length and 475 Km^2 in drainage area, is an important river located in Southern Taiwan.

Thirty-nine yearly daily streamflow records, from 1961 to 1999, were employed to determine the return period for extreme flood events. In this study, the flood events, shown in Fig. 3, were defined as a daily streamflow equaling or exceeding a certain threshold. Two properties, flood volume (V) and flood peak (Q), were used to characterize the extreme flood events. The flood duration was not considered in characterizing the flood events because most of the rivers in Taiwan are so steep that only floods with a short duration occur. The flood peak is defined as the maximum daily flow during the flood period, while the flood volume is defined as the cumulative flow volume during the flood period.

The threshold was considered at 100 cms in this study, i.e. flood events are defined as a daily streamflow equal to or greater than 100 cms. Fifty events were abstracted from the recorded daily streamflow data. The correlation coefficient between the flood volume and flood peak for these fifty flood events was 0.403.

4.2 Bivariate extreme value distribution of flood events

It was assumed that the Gumbel distribution could be used to represent both the flood peak and flood volume for extreme flood events. Therefore, flood events characterized by flood volume and flood peak can be fitted into a bivariate extreme value distribution with the Gumbel marginal distributions. The correlation coefficient between flood volume and flood peak was 0.403, hence $m = 1.294$ according to Eq. (16). The parameters, α and β, of the Gumbel distribution were estimated using the method of moments suggested by Yue (2001) and given by

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

$$\beta = M - 0.577\alpha \tag{29}$$

where M and s are the mean and standard deviation of the sample data, respectively.

Estimated from the observed flood events, the flood volume parameters are $\alpha_V = 512.6, \beta_V = 1256.7$ and the flood peak parameters are $\alpha_Q = 269.3, \beta_Q = 479.6$. The flood volume and flood peak cumulative distribution function is expressed as

$$F(v, q) = e^{-\left[e^{-\frac{1.294(v-1256.7)}{512.6}} + e^{-\frac{1.294(q-479.6)}{269.3}}\right]^{0.773}} \tag{30}$$

Figure 4 shows the contour plot, with probabilities 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9, of cumulative distribution function of flood volume and flood peak. The observed flood events are also shown in the Fig. 4.

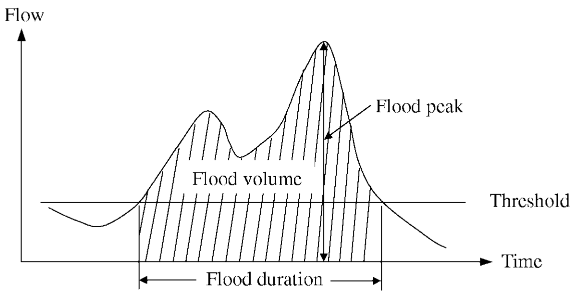

Fig. 3. Characteristics of a flood event

The theoretical marginal distributions for flood volume and flood peak are expressed in the following and shown in Figs. 5 and 6, respectively.

$$F(v) = e^{-e^{-\frac{v-1256.7}{512.6}}} \tag{31a}$$

$$F(q) = e^{-e^{-\frac{q-479.6}{269.3}}} \tag{31b}$$

The observed flood volume and flood peak for these fifty flood events are also shown in Figs. 5 and 6. Compared to the theoretical Gumbel distributions, Eqs. (31a) and (31b), the observed data are well fitted and accepted by the chi-square goodness of fit test.

4.3 Return periods

The average interarrival time for these fifty flood events, defined as streamflow equaling to or exceeding 100 cms, for the Ichu gauge station was 283.2 days, namely 0.776 years. Based on the information estimated from the observed floods, the return periods, expressed in years, for flood volume and flood peak are given by

$$E(T_V) = \frac{0.776}{1 - e^{-e^{-\frac{v-1256.7}{512.6}}}} \tag{32a}$$

$$E(T_Q) = \frac{0.776}{1 - e^{-e^{-\frac{q-479.6}{269.3}}}} \tag{32b}$$

Figure 7 illustrates the relationship between the return period and flood volume and peak defined by Eqs. (32a) and (32b), respectively. If the n-year flood events are defined as a return period equal to n years, the flood volume and flood peak for 2, 5, 10, 20, 50, and 100-year flood events are listed in Table 1.

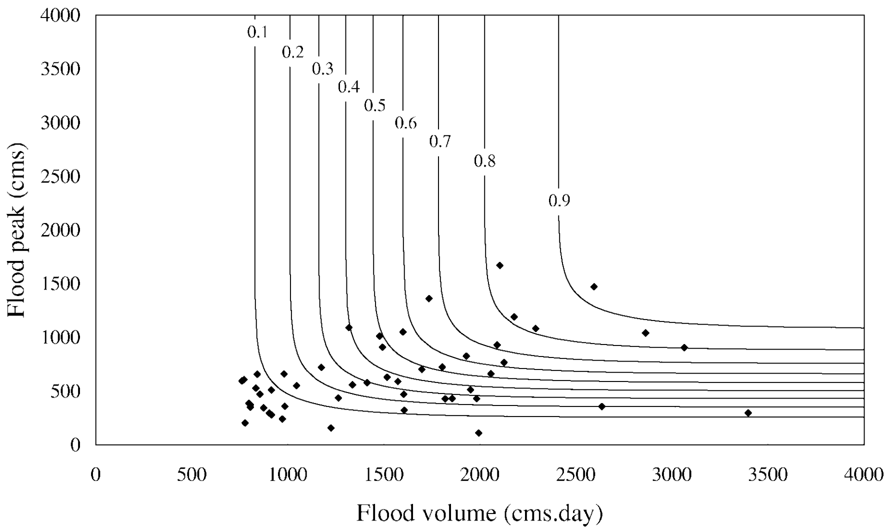

Fig. 4. The contour plots of bivariate distribution of flood volume and flood peak and observed flood events

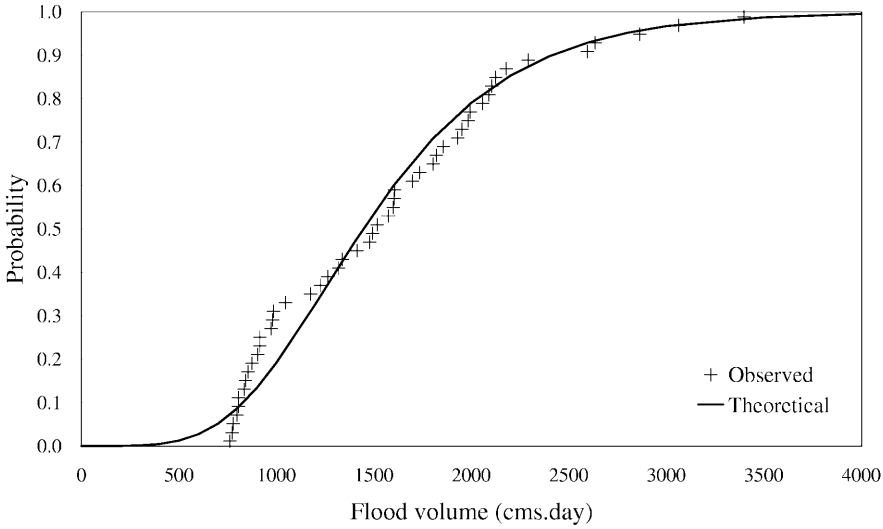

Fig. 5. The marginal distribution of flood volume

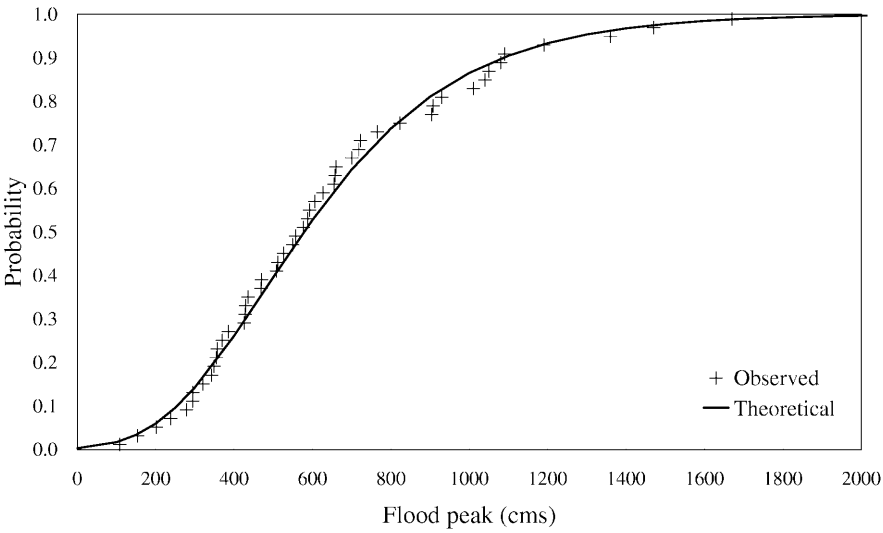

Fig. 6. The marginal distribution of flood peak

The conditional flood peak distribution given flood volume equaling or exceeding a specific threshold is

$$F(q|V \geq v) = \frac{e^{-e^{-\frac{q-479.6}{269.3}}} - e^{-\left[e^{-\frac{1.294(v-1256.7)}{512.6}} + e^{-\frac{1.294(q-479.6)}{269.3}}\right]^{0.773}}}{1 - e^{-e^{-\frac{v-1256.7}{512.6}}}} \tag{33}$$

If the specific threshold for flood volume in Eq. (33) is selected as the same values in column 2 of Table 1, then the conditional flood peak distribution given 2, 5, 10, 20, 50, and 100-year flood events defined solely by the flood volume is shown in Fig. 8. According to above information, one can estimate the return period for flood peak exceeding a certain magnitude given flood volume equaling or exceeding any specific value. For example, given a 2-year flood event defined solely by the flood volume, i.e., flood volume exceeding 1621 cms · day, the return period of flood peak equal to or greater than 704 cms is 2 years. Obviously, the above information cannot be obtained from the univariate frequency analysis.

Similarly, the conditional flood volume distribution given flood peak equaling or exceeding a certain magnitude is expressed in Eq. (34). The conditional distribution for flood volume given 2, 5, 10, 20, 50, and 100-year flood events defined solely by the flood peak is shown in Fig. 9.

$$F(v|Q \geq q) = \frac{e^{-e^{-\frac{v-1256.7}{512.6}}} - e^{-\left[e^{-\frac{1.294(v-1256.7)}{512.6}} + e^{-\frac{1.294(q-479.6)}{269.3}}\right]^{0.773}}}{1 - e^{-e^{-\frac{q-479.6}{269.3}}}} \tag{34}$$

The joint return period for flood volume equal to or greater than a certain value or flood peak equal to or greater than another certain value, i.e. $V \geq v$ or $Q \geq q$, is shown in Fig. 10 and expressed as

$$E(T_{VQ}) = \frac{0.776}{1 - e^{-\left[e^{-\frac{1.294(v-1256.7)}{512.6}} + e^{-\frac{1.294(q-479.6)}{269.3}}\right]^{0.773}}} \tag{35}$$

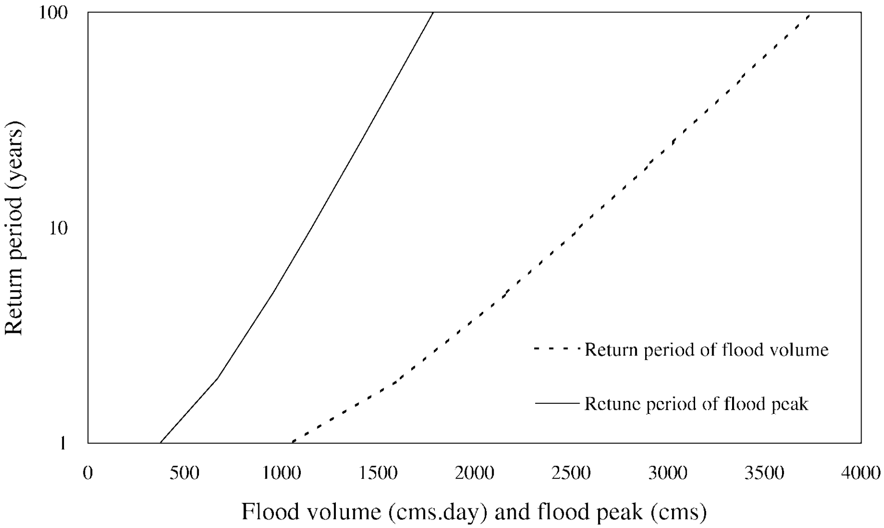

Fig. 7. Return periods for flood volume and flood peak

Table 1. The 2, 5, 10, 20, 50, and 100-year flood events defined separately by flood volume and flood peak

Return period (years)	Flood volume (cms · day)	Flood peak (cms)
2	1621	671
5	2169	959
10	2546	1157
20	2912	1349
50	3388	1599
100	3745	1787

The joint return period for flood volume equal to or greater than a certain value and flood peak equal to or greater than another certain value, $V \geq v$ and $Q \geq q$, is shown in Fig. 11 and expressed as

$$E(T'_{VQ}) = \frac{0.776}{1 - e^{-e^{-\frac{v-1256.7}{512.6}}} - e^{-e^{-\frac{q-479.6}{269.3}}} + e^{-\left[e^{-\frac{1.294(v-1256.7)}{512.6}} + e^{-\frac{1.294(q-479.6)}{269.3}}\right]^{0.773}}} \tag{36}$$

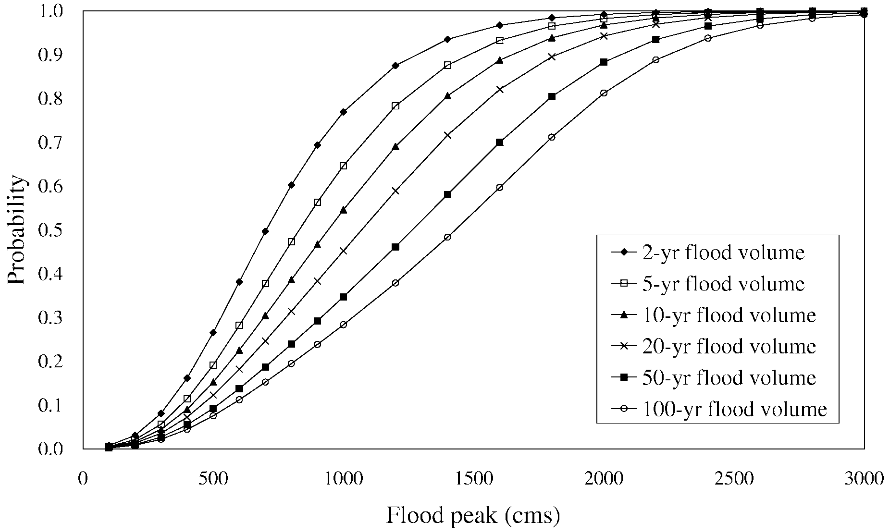

Fig. 8. Conditional distribution of flood peak given 2, 5, 10, 20, 50, and 100-year flood event defined by flood volume

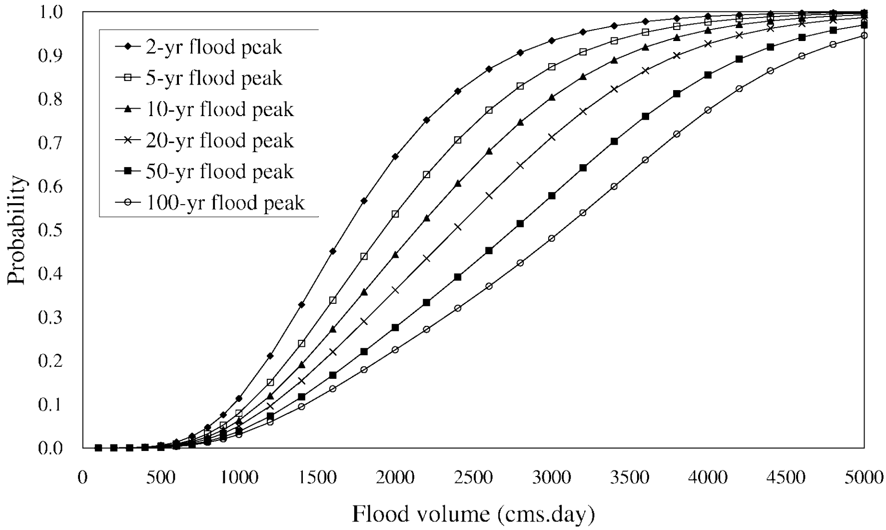

Fig. 9. Conditional distribution of flood volume given 2, 5, 10, 20, 50, and 100-year flood event defined by flood peak

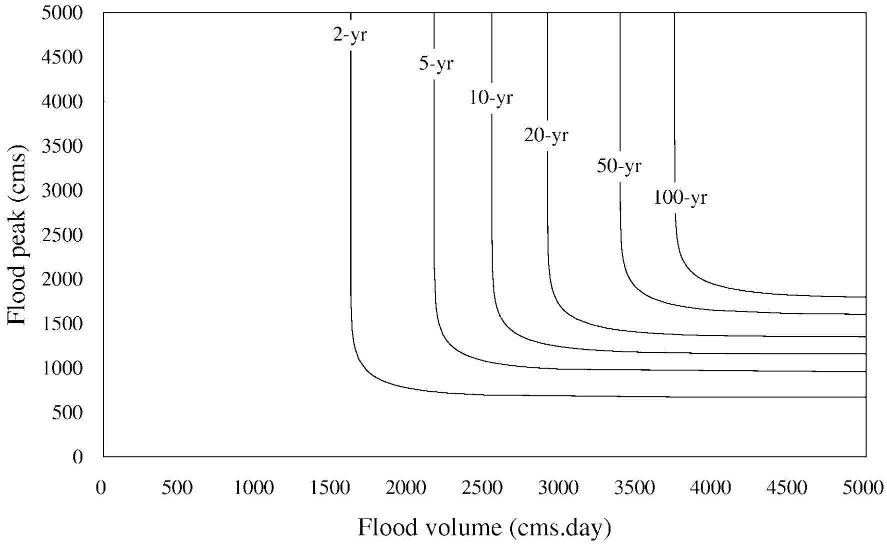

Fig. 10. Joint return period of flood volume and flood peak $E(T_{VQ})$

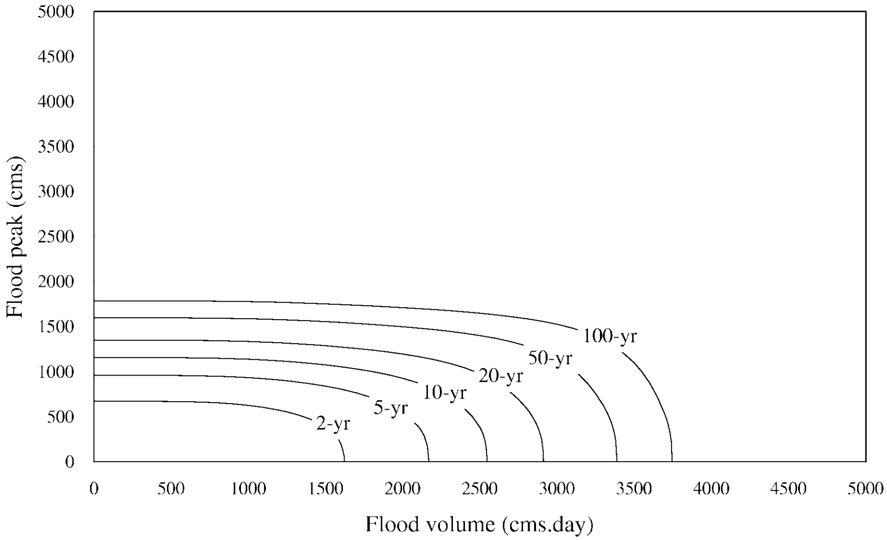

Fig. 11. Joint return period of flood volume and flood peak $E(T'_{VQ})$

Not like the return period defined by a single random variable, the specific joint return periods can be achieved using various combinations of the two random variables. Hence, the joint return period for flood volume and flood peak must be illustrated using the contour lines. However, the joint return periods $E(T_{VQ})$ and $E(T'_{VQ})$ exhibit different characteristics. The contour lines for various specific joint return period years defined by $E(T'_{VQ})$ are bounded by the horizontal and vertical axes, while the contour lines for various specific joint return period years defined by $E(T_{VQ})$ have no bounds. In addition, for the same value of v and q, $E(T'_{VQ})$ is greater than $E(T_{VQ})$ according to Eqs. (10) and (11). For example, $v = 1000$ cms $\cdot$ day and $q = 1000$ cms, $E(T_{VQ}) = 0.95$ years, and $E(T'_{VQ}) = 6.23$ years.

4.4
Discussion

The bivariate distribution of extreme hydrological events can offer important and useful information for hydrological structure design. However, comprehensively interpreting and comparing the univariate frequency analysis is the key issue. If the capacity of hydrological structures can be determined using flood peak or flood volume separately and the correlation between them is not significant, then Eq. (32a) or (32b) will be sufficient for the design criterion. If both the flood volume and flood peak are important factors for design, then Eq. (32a) or (32b) is insufficient because the correlation between them is not considered. For example, the flood peak is an essential design factor for a reservoir spillway. If the reservoir storage capacity is small compared to the flood inflow, then the flood volume is also an important safety factor in the reservoir design. The joint return periods for flood volume and flood peak can offer more useful information for design criterion. The use of Eq. (35) or (36) as the design criterion depends on what situations will destroy the structure. Under the condition that either flood peak or flood volume exceeding a certain magnitude will cause damage, then Eq. (35) can be used to evaluate the average recurrence interval. On the other hand, when the flood volume and flood peak must exceed a certain magnitude that will cause damage, then Eq. (36) is used.

It is worthwhile to observe the relationship between the joint return period for the flood volume and flood peak and the return periods defined solely by the flood peak or flood volume. Comparing Figs. 10 and 11, the values of horizontal part of specific return period defined by $E(T_{VQ})$ and $E(T'_{VQ})$ is the same and equal to the return period defined by the flood peak solely. The values of vertical part of specific return period defined by $E(T_{VQ})$ and $E(T'_{VQ})$ is equal to the return period defined by the flood volume solely. For example, a 10-year flood event defined by the flood volume solely is 2546 cms · day and a 10-year flood event defined by the flood peak solely is 1157 cms. These two magnitudes are also the lower limits for a 10-year flood defined by $E(T_{VQ})$ and the upper limits for a 10-year flood defined by $E(T'_{VQ})$.

5
Summary and conclusions

This study presented a methodology to define the return period for bivariate distributed extreme hydrological events. If the extreme hydrological events must be described using two random variables, the return periods for the extreme hydrological events can be defined using a single random variable separately or jointly using two random variables. If the correlation between random variables is insignificant in the design criterion, the return periods for the extreme events can be defined using one random variable. The joint consideration of two random variables leads to a more complex return period form. The return periods for a bivariate distribution can be defined in two ways. The first method defines the return periods using one random variable equaling or exceeding a certain magnitude and/or another random variable equaling or exceeding another certain magnitude. The second method defines the conditional return periods for one random variable given that another random variable equals or exceeds a specific magnitude.

In this study, a bivariate extreme value distribution with the Gumbel marginal distributions (also known as the logistic model) was used to fit extreme flood events defined by a daily streamflow exceeding a specific threshold. Both the flood volume and flood peak were then used to characterize the extreme flood events. If the flood volume and flood peak are considered separately, the return periods for extreme events can be defined using the flood volume or flood peak, as in the traditional univariate flood frequency analysis. Joint consideration of the flood volume and flood peak results in a return period definition in two different ways. The first method defines the joint return periods for flood volume and/or flood peak equaling or exceeding a certain magnitude. For the same magnitudes of flood volume and peak, $E(T'_{VQ})$ is greater than $E(T_{VQ})$. The second method defines the conditional return period for flood volume given flood peak, or vice versa. In addition to the conditional return periods, the conditional flood volume distribution given flood peak and the conditional flood peak distribution given flood volume can also be derived. It is worthwhile to note that the contour lines for various specific joint return period years defined by $E(T'_{VQ})$ are bounded by the horizontal and vertical axes, while the contour lines for various specific joint return period years defined by $E(T_{VQ})$ have no bounds.

The proposed methodology was applied to the recorded daily streamflow from the Ichu gauge station of the Pachang River located in Southern Taiwan. The results exhibited a good agreement between the theoretical models and observed data and provided more useful information than the univariate random variable frequency analysis.

References

Anderson CW, Dwyer IJ, Nadarajah S, Reed DW, Tawn JA (1994) Maximum reservoir water level. In Reservoir Safety and Environment, pp. 200–213, Thomas Telford, London

Anderson CW, Nadarajah S (1993) Environmental factors affecting reservoir safety. In Barnett V, Turkman KF (eds), Statistics for the Environment pp. 163–182. John Wiley & Sons, Chichester

Bacchi B, Becciu G, Kottegoda NT (1994) Bivariate exponential model applied to intensities and duration of extreme rainfall. J. Hydrol., 155: 225–236

Chow VT, Maidment DR, Mays LR (1988) Applied Hydrology, McGraw-Hill Inc., New York

Goel NK, Seth SM, Chandra S (1998) Multivariate modeling of flood flows. ASCE J. Hydraul. Eng., 124(2): 146–155

Gumbel EJ (1958) Statistics of Extremes. Columbia University Press, New York

Gumbel EJ, Mustafi CK (1967) Some analytical properties of bivariate extremal distributions. J. Am. Statist. Assoc., 62: 569–588

Gupta VK, Duckstein L, Peebles RW (1976) On the joint distribution of the largest flood and its time of occurrence. Water Resour. Res., 12(2): 295–304

Nadarajah S, Withers CS (2001) Modeling dependency between climate extremes for New Zealand. World Resour. Rev., 13(4): 526–539

Kotz S, Nadarajah S (2000) Extreme Value Distributions: Theory and Applications, Imperial College Press, London

Kotz S, Balakrishnan N, Johnson NL (2000) Continuous Multivariate Distribution, Vol 1, Models and Applications, John Wiley & Sons, Inc., New York

Shiau JT, Shen HW (2001) Recurrence analysis of hydrologic droughts of differing severity. ASCE J. Water Resour. Plng. and Mgmt., 127(1): 30–40

Singh K, Singh VP (1991) Derivation of bivariate probability density functions with exponential marginals. Stochast. Hydrol. Hydraul., 5(1): 55–68

Stone CJ (1996) A Course of Probability and Statistics, Wadsworth Publishing Company, New York

Yue S, Ouarda TBMJ, Bobee B, Legendre P, Bruneau P (1999) The Gumbel mixed model for flood frequency analysis. J. Hydrol., 226: 88–100

Yue S (2001) A statistical measure of severity of El Niño events. Stochast. Envir. Res. and Risk Assess., 15(2): 153–172