The Gumbel logistic model for representing a multivariate storm event

Sheng Yue *

Meteorological Services of Canada, Ontario Region, Environment Canada, 75 Farquhar Street, Guelph, Ontario, Canada N1H 3N4

Received 25 April 1999; received in revised form 5 May 2000; accepted 24 May 2000

Abstract

This study proposes the use of the Gumbel logistic model, the bivariate extreme value distribution with Gumbel marginals, to analyze the joint distribution of annual maximum storm peaks (maximum rainfall intensities) and the corresponding storm amounts which are mutually correlated. Parameters of the distribution are estimated using the method of moments (MM). On the basis of the marginal distributions of these random variables, the joint distribution, the conditional probability distribution, and the associated return periods can be readily deduced. The model is verified using observed daily rainfall data from the Tokushima meteorological station of Tokushima Prefecture, Japan. Results show that the model is suitable for representing the joint distribution of correlated storm peaks and amounts that are Gumbel distributed. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Storm frequency analysis; Storm peak; Storm amount; Gumbel distribution; Bivariate extreme value distribution; Joint probability distribution; Marginal distribution; Conditional distribution

1. Introduction

In recent years, it has been recognized that a hydrological extreme event such as a storm event appears to be a multiple episodic event that might be characterized by a few correlated variables such as storm maximum intensity (peak) and storm total amount. The severity of the storm event is a function of both storm peak and storm total amount. However, storm frequency analysis has often centered on storm peak analysis only. This single-variable storm frequency analysis provides a limited assessment of storm events. Many hydrological engineering planning, design, and management problems require a more complete description of a storm event, i.e., the joint distribution as well as the conditional distribution of these two correlated variables. Some meaningful works on the joint distribution with regard to storm events have been done by Hashino [14], Singh and Singh [18], Bacchi et al. [1], Yue [23], etc. Hashino [14] utilized the Freund bivariate exponential distribution [6] to represent the joint probability distribution of rainfall intensities and the corresponding maximum storm surges at the Osaka Bay, Japan. Singh and Singh [18] derived a bivariate probability density function with exponential marginals to describe the joint distribution of rainfall intensities and the corresponding depths. Bacchi et al. [1] proposed another bivariate exponential model with exponential marginals to analyze the joint distribution of rainfall intensities and durations. Yue [23] explored the use of the bivariate normal model for representing the joint statistical properties of non-normally distributed storm events that can be normalized by a transformation technique such as the Box–Cox approach [2].

In practice, extreme events such as maximum storm peak and the corresponding storm amount might be represented by the Gumbel distribution [3, 8]. Thus, it is useful to develop a bivariate extreme value distribution for representing two correlated Gumbel distributed random variables. Theoretically, two bivariate extreme value distributions with Gumbel marginals, namely the Gumbel mixed model and the Gumbel logistic model, have been studied by Gumbel [9–11], Gumbel and Mustafi [12], and Oliveira [16]. However, these models have been seldom applied to analyze joint statistical properties of correlated hydrological extreme events. The recent work by Yue et al. [22] investigated the suitability of the Gumbel mixed model [9] for describing a multivariate flood event. But the limitation of this
model is that the correlation coefficient between two random variables must be in the range $0 \leq \rho \leq 2/3$.

This study makes another attempt to use the Gumbel logistic model, the bivariate extreme value distribution model with Gumbel marginals to represent the joint probability distribution of positively correlated storm peaks and amounts. On the basis of the marginal distributions of these random variables, the joint distribution, the conditional probability distribution, and the associated return periods can be readily derived. Section 3 presents a practical application, the use of the model for representing the joint distribution of annual storm peak and the corresponding storm amount at the Tokushima meteorological station in Japan. The results indicate the usefulness of the model for describing joint statistical properties of two positively correlated random variables with Gumbel marginals.

### 2. The Gumbel logistic model

The Gumbel logistic model with standard Gumbel marginal distributions was originally proposed by Gumbel [10,11] as follows:

$$F(x, y) = \exp \left\{ - \left[ \frac{1}{m} \left( - \ln F(x) \right)^m + \left( - \ln F(y) \right)^m \right]^{1/m} \right\}$$

$$F(x) = \exp \left\{ - \exp \left( - \frac{x - u_x}{\alpha_x} \right) \right\}, \quad (1 \leq m), \quad (6a)$$

$$F(y) = \exp \left\{ - \exp \left( - \frac{y - u_y}{\alpha_y} \right) \right\}, \quad (6b)$$

where $(u_x, \alpha_x)$ and $(u_y, \alpha_y)$ are the location and scale parameters of the Gumbel distributions of $X$ and $Y$, respectively.

The joint probability density function (p.d.f.) is derived using Eq. (1) and is expressed as follows:

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} = \frac{F(x, y)}{x \alpha_x} \frac{F(x, y)}{y \alpha_y} \cdot \left( e^{\frac{-u_x}{\alpha_x}} + e^{\frac{-u_y}{\alpha_y}} \right)^{(2m)/m} \cdot \left[ \left( e^{\frac{-u_x}{\alpha_x}} + e^{\frac{-u_y}{\alpha_y}} \right)^{1/m} + m - 1 \right]$$

$$\cdot e^{\frac{-(x - u_x)^2}{2\alpha_x^2}} e^{\frac{-(y - u_y)^2}{2\alpha_y^2}}.$$ \( \tag{7} \)

The joint cumulative distribution function (cdf) of the random variables $X$ and $Y$ takes the same form as Eq. (1) in which the marginal distributions are given by Eqs. (6a) and (6b).

The conditional cdf $F_{X|Y}(x|y) = \Pr[X \leq x|Y = y]$ of $X$ given $Y = y$ is presented by

$$F_{X|Y}(x|y) = \frac{F(x, y)}{F_Y(y)} = \exp \left\{ \exp \left( - \left( y - u_y \right) / \alpha_y \right) \right.$$}

$$+ \exp \left( - m(x - u_x)/\alpha_x \right) \right\}.$$ \( \tag{8} \)

Similarly, the conditional probability distribution function of $Y$ given $X \leq x$ can be expressed by an equivalent formula.

The return periods exceeding certain values of the variables $X$ and $Y$ are represented as follows:

$$T_x = \frac{1}{1 - F(x)} \quad (F(x) = \Pr[X \leq x]), \quad (9a)$$

$$T_y = \frac{1}{1 - F(y)} \quad (F(y) = \Pr[Y \leq y]). \quad (9b)$$

On the basis of the same principle, the joint return period $T(x, y)$ of $X$ and $Y$ associated with the event $(X > x, Y > y, or X > x and Y > y$, i.e., at least one value of $X$ and $Y$ exceeds) can be represented by

$$T(x, y) = \frac{1}{1 - F(x, y)} \quad (F(x, y) = \Pr[X \leq x, Y \leq y]). \quad (9c)$$
The conditional return period of $X$ given $Y \leq y$ and the conditional return period of $Y$ given $X \leq x$ are, respectively, presented as follows:

$$T_{X|Y} = \frac{1}{1 - F(x|y)} \quad (F(x|y) = \Pr[X \leq x|Y \leq y]), \quad (9d)$$

$$T_{Y|X} = \frac{1}{1 - F(y|x)} \quad (F(y|x) = \Pr[Y \leq y|X \leq x]). \quad (9e)$$

3. Application

In order to demonstrate the applicability of the Gumbel logistic model, 102-year daily rainfall data from 1892 to 1996 (except the years 1896, 1965, and 1976) were made use of, which were observed at the Tokushima meteorological observation station (TMOA) in Japan. The data from 1892 to 1990 were recorded in the Monthly Meteorological Report by TMOA [19]. The data after 1990 were obtained in the Gumbel logistic model, 102-year daily rainfall data from 1892 to 1996 (except the years 1896, 1965, and 1976) were made use of, which were observed at the Tokushima meteorological observation station (TMOA) in Japan. The data from 1892 to 1990 were recorded in the book by TMOA [19]. The data after 1990 were obtained in the Monthly Meteorological Report by TMOA. The corrections concerning the changes in the station location, the measurement equipment, and the measurement procedure have been made by TMOA [19].

3.1. Definition of a storm sequence

A storm time series can be constructed using the annual maximum series (AMS) approach, or the partial duration series (PDS) (or peak over threshold (POT)) approach. An AMS is constructed by selecting the annual maximum value of each year, i.e., only one event per year is retained. This naturally leads to events that are generally independently and identically distributed. The PDS consists of all values that exceed a certain threshold. The main advantage of the PDS approach is that it is not confined to only one event per year and it allows for more extreme events to be considered. The key unresolved problem of the PDS is how to select appropriate thresholds. This issue has been addressed in the work of Cunnane [4], Valadares Tavares and Evaristo Da Silva [20], Wang [21], and Rasmussen et al. [17]. Wang [21] has documented that the AMS and the PDS are similar for a long-term time series. In the present study, the AMS approach is employed to construct a storm peak time series and the corresponding storm amount time series using the observed daily rainfall data.

During the monsoon season in Japan, standing rainy fronts or hurricanes or typhoons causes the heaviest storm both in storm peak (maximum rainfall intensity) and storm amount. As this type of rain continues to fall over a few days, it is considered to be a storm event. Because recorded historical rainfall data is often in the form of averages over a period of one day, one storm is defined as continuous daily rainfalls, as shown in Fig. 1.

Let the storm peak $I$ (mm/day), be the maximum daily rainfall in a year selected by Eq. (10a) and the corresponding total rainfall amount $A$ (mm) be given by Eq. (10b).

$$I_n = \max\{u_{ni}, i = 1, 2, \ldots, 365(366)\} \quad (n = 1, 2, \ldots, 102), \quad (10a)$$

$$A_n = \sum_{i=1}^{D_n} u_{ni} \quad (n = 1, 2, \ldots, 102), \quad (10b)$$

where $u_{ni}$ is the $i$th daily rainfall amount (mm/day) for the $n$th year, $u_{n}$ the $n$th daily rainfall amount of the storm $n$ corresponding to the storm peak $I_n$, and $D_n$ is the storm duration (day) of the storm $n$.

The maximum daily rainfall (storm peak $I_n$) for each year is first selected using Eq. (10a) from the observed 102-year daily rainfall data at the Tokushima meteorological observation station; then the corresponding storm duration $D_n$ is counted and the storm amount $A_n$ is computed using Eq. (10b).

3.2. Marginal distributions of storm peaks and amounts

3.2.1. Empirical probabilities

The non-exceedance probability is estimated using the Gringorton formula [5,7,13]

$$P_k = \frac{k - 0.44}{N + 0.12}, \quad (11)$$

where $P_k$ is the cumulative frequency, the probability that a given value is less than the $k$th smallest observation in the data set of $N$ observations.

3.2.2. Parameter estimation

In a single-variable frequency analysis domain, the location and scale parameters of the Gumbel distribution can be estimated using a few approaches such as the maximum likelihood method (ML) and the method of moments (MM). While for the analysis of the joint distribution of two correlated Gumbel distributed
random variables, if the ML can provide a reliable estimate of the association parameter between the variables remains unknown. This paper presents the simplest way to derive the model’s parameters, i.e., to completely identify the model’s parameters through the marginal distributions of the variables. Gumbel and Mustafi [12] proposed to estimate the association parameter via the MM. This study follows their idea that the association parameter is derived using Eqs. (3) and (4). The correlation coefficient (\( \rho \)) between storm peaks and amounts is estimated by Eq. (4), and takes 0.731. The association parameter \( m \) between storm peaks and amounts is computed using Eq. (3), and is equal to 1.928. In order to keep the consistency of the methodology, the location and scale parameters are also estimated using the MM as follows:

\[
\alpha = \sqrt{6} S, \tag{12a}
\]

\[
u = M - 0.577 \alpha, \tag{12b}
\]

where \( M \) and \( S \) are the mean and standard deviation of the sample data, respectively. The estimated mean and standard deviation of storm peaks and amounts from the sample data are listed in Table 1. The estimated parameters of the Gumbel distribution are also presented in Table 1.

The \( \chi^2 \) test is executed to test the goodness of fit of the Gumbel distribution. The \( \chi^2 \) test statistics are 7.65 for the storm peak and 8.82 for the storm amount. The critical value of the \( \chi^2 \) test \( \chi^2_{0.05}(9) \) is 16.92. Thus, the null hypothesis \( H_0 \) that the underlying distributions of both the storm peaks and amounts are Gumbel distributed is accepted at the significant level 0.05. The empirical probabilities computed by Eq. (11) and the theoretical probabilities calculated using Eqs. (6a) and (6b) for the observed storm peaks and amounts are respectively illustrated in Figs. 2(a) and (b).

### 3.3. Statistics of the joint distribution of storm peaks \( I \) and amounts \( A \)

#### 3.3.1. Validity of the proposed model

The empirical non-exceedance joint probabilities of storm peaks \( I \) and amounts \( A \) are estimated using an equivalent form of Eq. (11) in order to keep consistent with the marginal case and is given by Yue et al. [22]

\[
P_{ij} = \text{Pr}[I \leq i_k, A \leq a_j] = \frac{\sum_{m=1}^{d} \sum_{l=1}^{f} n_{ml} - 0.44}{N + 0.12}, \tag{13}
\]

where \( N \) is the total number of observations \( (N = 102) \), and \( n_{ml} \) is the number of occurrences of the combinations of \( i_m \) and \( a_l \).

Theoretical joint probabilities of the real occurrence combinations of \( i_k \) and \( a_j \) are estimated using Eq. (1). The empirical and theoretical joint probabilities of storm peaks and amounts are plotted in Fig. 3 in which
the solid line represents the theoretical probabilities (arranged in ascending order), and the corresponding empirical probabilities are indicated by the plus sign. The $x$-axis is the corresponding order number of a combination of $i_k$ and $a_j$. It is evident that no significant difference can be detected. Thus, it is concluded that the model is suitable for representing the joint distribution of the correlated Gumbel distributed storm peaks and amounts.

3.3.2. Contours of the joint cdf $F(i,a)$ and joint return period $T(i,a)$ of $I$ and $A$

The joint cdf and joint return period corresponding to given $I = [0, 10, 20, \ldots, 500]$ and $A = [0, 20, 40, \ldots, 800]$ are computed using Eqs. (1) and (9c) and are displayed in Figs. 4(a) and (b), respectively. The corresponding contours of the joint cdf and the joint return period of $I$ and $A$ are plotted in Figs. 4(c) and (d), respectively. These contours indicate that the proposed method can contribute meaningfully in solving several problems of hydrological engineering design and management, for which single variable storm frequency analysis cannot provide answers. For example, given an occurrence probability or a return period of a storm event, we can obtain various occurrence combinations of storm peaks and amounts, and vice versa. These results are especially useful in the case that both storm peak and amount must be considered for hydrological engineering design and management.
3.3.3. Conditional return periods

The conditional return period $T_{I/A}$ of storm peak $I$ given storm amount $A$ is estimated based on Eq. (8) and (9d), and is shown in Fig. 5(a). Similarly, the conditional return period $T_{A/I}$ of $A$ given $I$ is plotted in Fig. 5(b). It can be seen that the proposed method also allows one to obtain information concerning the occurrence return periods of the storm peak under the condition that a given storm amount occurs, and vice versa.

For the purpose of comparison, the return period $T_I$ of the storm peak $I$ is computed using Eq. (9a) and is displayed by the dashed line in Fig. 5(a). In fact, it is the conditional return period $T_{I/A}$ of the storm peak $I$ given the storm amount $A$ when $A \to \infty$. It can be seen that given a return period (for example, 100 years), the corresponding storm peak value obtained by single-variable frequency analysis is greater than those obtained by the joint distribution. This implies that if we ignore the close correlation between the storm peak and amount and execute single-variable frequency analysis on storm peak only, then the severity of the storm event will be overestimated in the study case. If hydrological engineering planning, design, and management are based on this storm peak, then this overestimation will lead to increased cost.

Similarly, the return period $T_A$ of storm amount $A$ is computed using Eq. (9b) and is presented by the dashed line in Fig. 5(b). We can obtain the same inference as that of the storm peak case.

4. Conclusions

This study presents a methodology for using the Gumbel logistic model to analyze the joint distributions of two positively correlated extreme random variables that are Gumbel distributed. The model is used to develop the joint distribution of combinations of storm peaks and amounts. Based on this model, if the marginal distributions of two random variables can be represented by the Gumbel distribution, one can readily obtain the joint probability distribution, the conditional distributions, and the associated return periods of these variables. The parameters of the model are easily estimated from the sample data based on the marginal distributions of the random variables.

This approach is verified using observed daily rainfall data from the Tokushima meteorological station of Tokushima prefecture, Japan. A good agreement is observed between the theoretical and empirical distributions. The results point out that the proposed method provides additional information which cannot be obtained by single variable storm frequency analysis, such as the joint return periods, and the conditional return periods of these variables. These results also indicate that the proposed model can contribute meaningfully in solving several problems of hydrological engineering design and management. For example, given a storm event return period, it is possible to obtain various occurrence combinations of storm peaks and amounts, and vice versa. These various scenarios can be of great usefulness in the analysis and assessment of the risk associated with several hydrological problems.

Acknowledgements

The work had been accomplished when the author was in INRS-EAU, University of Quebec, Canada. The author gratefully acknowledges the Natural Sciences and Engineering Research Council (NSERC) of Canada, Hydro-Quebec, and INRS-EAU for partially supporting this study. The author also thanks Prof. M. Hashino of University of Tokushima in Japan for his kind assistance concerning this research. The construc-
tive comments of anonymous reviewers significantly improved the quality of this paper.

References