BIVARIATE FLOOD FREQUENCY ANALYSIS USING GUMBEL COPULA

Tarmizi Ismail[1]*, Kamal Ahmed, Mahiuddin Alamgir, Mohammad Noor Kakar & Abu Bakar Fadzil

Department of Hydraulics and Hydrology, Faculty of Civil Engineering, Universiti Teknologi Malaysia

Corresponding Author: tarmiziismail@utm.my

Abstract: A copula based methodology is presented in this study for bivariate flood frequency analysis over a station over a Kelantan river basin located in Northeast Malaysia. The joint dependence structures of three flood characteristics, namely, peak flow, flood volume and flood duration were modelled using Gumble Copula. Various univariate distribution functions of flood variables were fitted with observed flood variables to find the best distributions (eg. generalized pareto, log-normal, exponential, gamma distribution, weibull, gumbel, cauchy). The results of study revealed that different variable fits with different distributions and the correlation analysis among variables showed a strong association. Cumulative joint distribution functions (CDF) of peakflow and volume, peakflow and duration and volume and duration revealed that return period of joint return periods are much higher.

1.0   Introduction

Reliable assessment of flood frequency and magnitude is essential for effective mitigation planning and designing hydraulic structures. Univariate analysis based on flood peak is generally used for predicting flood occurrence. However, devastating flood does not always depend on peak flow. Destructive floods occur when high flood peak sustain for a longer period or huge volume of flood water inundates an area for a longer time. The duration of flood is often very important, particularly for structural designing. Flood volume is also required for designing flood protecting measures. Therefore, flood hazard can be considered as a multivariate phenomenon depends on different variables such as, flood peak, flood duration and flood volume. Several studies has been conducted to analysis the flood variables using multivariate method in order to evaluate flood characteristics (Zhang, 2012; Chen, 2013). Most of the studies pointed that marginal distribution that best describe the flood parameters are often not from the same probability distribution function (Requena, 2013; Salvadori, 2013) and therefore, make

the bivariate analysis complicated. To overcome this problem the concept of Copula was introduced into flood frequency analysis (Chowdhary *et al*., 2011; Kao and Change, 2011). Copula allows user to model the correlations among flood variables without considering the type of marginal distributions of flood variables (Chen, 2012; Salvadori and Michele, 2013). The evolution of joint distribution of flood frequency analyses using Copula initiated number of studies on joint distribution of flood variables across the world. Chen (2012) used Copula function for multivariate analysis of flood coincidence analysis. Chowdhary *et al*. (2011) compared different copulas for identification of best fitted Copula for bivariate frequency analysis of flood peak and flood volume. Kao and Change (2011) employed Copula for flood frequency analysis in ungauged river basin of Nashville, USA. Li *et al*. (2012) adopted Copula for bivariate flood frequency analysis using historical information. Reddy and Ganguli (2012) used Archimedean Copulas for bivariate flood frequency analysis of flood upper Godavari River. Salvadori and Michele (2013) adopted multivariate extreme value methods for analysis for flood. Xie and Wang (2013) used joint probability methods for precipitation and flood frequencies analysis. All the above studies indicated Copula as a better option for joint parametric distribution. However, application of joint analysis of flood variables is still very rare in structural designing and flood management. Therefore, the objectives of this study are (a) to determine the flood variables namely, annual peak, corresponding duration and volume from historical river discharge data, (b) to estimate the univariate distribution of flood variables, and (c) to perform Copula analysis to modelled joint distribution of flood. The proposed method is applied over a station located in Kelantan river basin Malaysia.

2.0   Study Area and Data

The station of Sg. Nenggiri di Jam. Bertam located over Kelantan River, Malaysia (Figure 1) is selected as a study area. Forty-three years (1972-2014) hourly river discharge data records was used for the study.

Hourly stream flow data was collected from the Department of Irrigation and Drainage (DID), Malaysia. The years having complete streamflow record were used for analysis. The rainfall over the area varies between 0 mm in the dry season (March–May) to 1750 mm in the monsoon season (November–January). The average runoff from the area is about 500 m³/s. Flood is a common phenomenon in the river basin; during the past twenty years, the basin has been severely flooded for many times. This further magnifies the need to generate joint flood characteristics probability for the welfare of mankind.
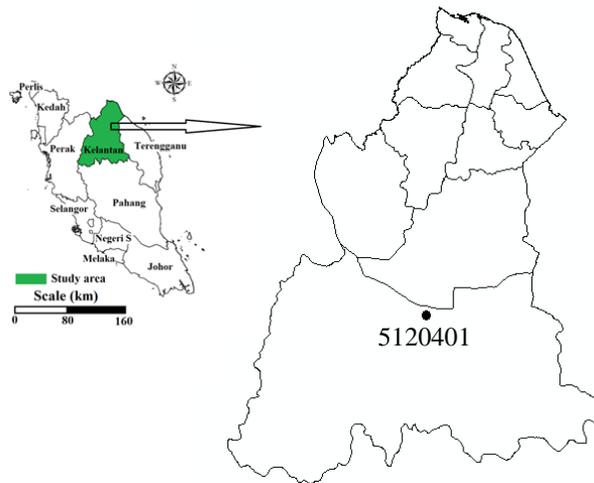
Figure 1: Location of the study area (Kelantan) in the map of peninsular Malaysia

3.0   Methodology

This study was mainly conducted in four major steps. In the first step hourly stream flow data was collected from the Department of Irrigation and Drainage (DID), Malaysia. The quality of data was evaluated before determining the annual flood and its corresponding volume and duration. In the second step, various distribution of flood parameters namely, flood duration, flood volume and peak flow was conducted using Generalized Pareto, Normal, Log-normal, Exponential, Gamma, Weibull, Gumbel, Cauchy distributions. The hourly river discharge data was used to determine the annual flood and its corresponding volume and duration. The initiation and ending of a flood event were marked using the method of (Salarpour *et al*. 2013). A flood event begins from a point when the hydrograph start to rise to a point on the recession limb where the separation line with a constant slope. The time lapse between the initiation time (t_s) and end time (t_e) represents the flood duration (D). The flood volume (V) is given by the area and peak flow is the maximum flow during the flood event. The annual flood series was determined based on water year similar to that done by (Salarpour *et al*. 2013).

The cumulative distribution function (CDF) is defined as:

$$F(x) = \int_{-\infty}^{x} fx(t)dt \qquad\qquad (1)$$

The theoretical CDF is displayed as a continuous curve. The empirical CDF is denoted by:

$$F_n(x) = \frac{1}{n} \ (Number \ of \ Observation \ \le x) \tag{2}$$

Where, x is the random variable representing the hourly rainfall intensity.

The Probability Density Function (PDF) is the probability that the variate has the value x

$$\int_b^a f(x)dx \ = P(a \le X \le b) \tag{3}$$

For discrete distributions, the empirical (sample) PDF is displayed as vertical lines representing the probability mass at each integer X:

$$f(x) = P(X = x) \tag{4}$$

The empirical PDF is represented as a histogram with equal-width vertical bars (bins). The bins represent the number of sample data that belong to a certain interval divided by the total number of data points. Ideally, a continuous curve can be properly scaled to the number of intervals to form a continuous curve.

Kolmogorov-Smirnov (KS) test was conducted to estimate the goodness of fit to a particular distribution. If the estimated value of KS test statistics is found less than critical table value, it is assumed that the sample data is drawn from the same distribution. In the third step, Copula analyses were performed to model the joint distribution of flood variables
.
A copula captures the dependence of two or more random variables. The Sklar's theorem (Sklar, 1959) states that the joint behaviour of random variables (X, Y) with continuous marginal distributions $u = F_x(x) = P(X \le x) \ and \ v = F_y(y) = P(Y \le y)$ can be characterized uniquely by its associated dependence function or copula, C. For 2-dimensional cases, all (u, v) relationships c a n be written as:

$$F_{x,y}(X,Y) = C\big[F_x \ (x), F_y \ (y)\big] = C(u,v) \tag{5}$$

where, $F_{x,y}(X,Y)$ is the joint CDF of random variables X and Y and also $\forall_x, y \in R$ When I = [0,1], the bivariate copula has a distribution function of C = $I^2 \to I$ which normally satisfies the following basic properties:

The boundary conditions:

$$C(t,0) = C(0,t) = 0 \ and \quad C(t,1) = C(1,t) = t, \forall t = I$$

Increasing property:

$$C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0 \ \forall u_1, u_2, v_1, \in I \quad such \ that \quad u_1 \leq u_{2 \ and} \quad v_1 \leq v_2$$

The bivariate copula density, *C (u, v)*, is the double derivative of *C* with respect to its marginal and can be written as:

$$C(u,v) = \frac{\partial^2 C(u,v)}{\partial u \partial v} \tag{6}$$

3.1   Gumbel Copula

The Gumbel Copola has the generator $\varphi(\mu) = (-\ln(\mu))^\alpha$, with $\alpha \epsilon [1,\infty]$. For all $\alpha > 1$, the copula read as:

$$C(\mu_1, \dots \mu_n) = exp \left\{ -\left[\sum_{i=1}^{n} (-\ln \mu_i)^\alpha \right]^{1/\alpha} \right\} \tag{7}$$

The Kendall's tau can be computed as $\tau = 1 - \alpha^{-1}$. It can be shown that Gumbel Copulas have upper tail dependences $\lambda_U = 2 - 2^{1/\alpha}$ and lower tail dependence vanishing as α diverges to infinity.

4.0   Results and Discussion

4.1   Determination of Flood Variables

The mean, maximum, minimum, standard deviation and skewness of station are determined and listed in table 1. It can be seen that the mean duration of 68.35 hr can generate a volume 1.23 km^3 with a peak flow of 7139.71m^3/sec. Similar types of results were also obtained for other parameters such as maximum minimum, standard deviation and skewness.

Table 1: Summary statistics of flood parameter

Index	*Duration (hr)*	*Volume (km³)*	*Flow (m³/sec)*
Mean	68.35	1.23	7139.71
Maximum	142	7.07	19237.16
Minimum	3.00	0.04	1388.19
Standard Deviation	41.53	1.71	5259.44
Skewness	-0.08	2.33	0.92

4.2 Determination of Flood Variable Distributions

Various univariate distribution functions including pareto, log-normal, exponential, gamma distribution, Weibull, gumbel, Cauchy were used to determine the best fit for flood variables. The goodness of fit for different distributions was carried out based on Kolmogorov-Smirnov (KS). The obtained results are provided in Table 2. It was observed that there is no consistency of distributions for duration, volume and peak flow. For example; normal was found best fit for duration, exponential for volume and gamma for flow.

Table 2: Fitting distribution of flood variable

Variable	*Fitting Distribution*	*Parameter*	
Duration	Normal	σ	68.353
		μ	40.291
Volume	Exponential	λ	0.816
Flow	Gamma	α	0.738
		β	0.603

4.3 Copula Analysis for Flood Variables

After univariate analysis, the pairs of variables were modelled by Gumbel copula. The Copula linear correlation parameters among flood variables are given in Table 3. Correlation analysis measures the degree of association between two variables and, therefore, it can be used as a measure to show how flood variables are allied with each other. The table shows that correlation is highest between duration and volume (0.96) indicating volume of flood will be high with longer duration. The correlation between duration and peak flow was found 0.1 which lowest relationship among variable. On the

other side relationship between flow and volume was 0.20.

Table 3: Correlation parameter of copula

Static	*Duration*	*Volume*	*Flow*
Duration	1.00	0.96	0.10
Volume	0.96	1.00	0.20
Flow	0.10	0.20	1.00

Figure 2 demonstrates the bivariate return period of flood events over station (5120401). The contour lines in the figure represents the return period of flood event while x and y-axis are different combinations of flood peak, volume and durations. Red dots represent the observed data while grey dots represent the simulations. It can be seen that flood are more frequent at low return periods and duration of peak flood is low. Figures provide quick information over volume, duration and peak. For example, a flood volume of 2000 mcm can generate a peak flow of around 8000 cms at 2 years return period. Similarly, a flood volume of 20,000 mcm can happen in 2 days. On the other hand a flood peak of 10,000 cms can occur approximately in 1.8 days. Similar types of information can be extracted from the figures of other contour lines.
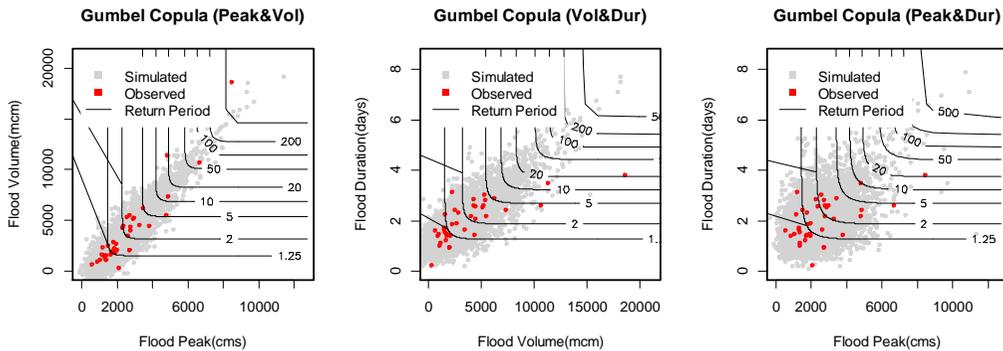

Figure 2: Contour plots for return periods of peak, volume and duration at station (5120401)

5.0   Conclusions

Destructive floods occur when high flood peak sustain for a longer period or huge volume of flood water inundates an area for a longer time. This emphasizes the need to study the joint distribution of flood variables together. In this study, three flood variables namely; peak flow, duration and volume are separately modelled by a

probabilistic distribution function. A flood event was defined from the point when the hydrograph start to rise while the time lapse between the initial time and end times was used to represent the flood duration. The peak flow is defined as the maximum flow during the flood event. The flood parameters were further fitted using the concept of Copula to construct joint bivariate distribution function of peak flow-duration, peak flow-volume and volume-duration. Based on Kolmogorov-Smirnov (KS), the different distribution fits were found for different flood variables. The correlation among flood variables including duration, volume and flow indicates that there is a good agreement between duration and volume, volume and flow. However relationship is relatively weak between duration and flow. The bivariate copula provides all information, which cannot be obtained by single-variable flood frequency analysis. These results can be useful for hydrological structure design studies and also expected that same methodology can be applied to any station over Malaysia.

6.0   Acknowledgements

References

Chen L, Singh, VP, Shenglian G, Hao Z, & Li T (2012). Flood coincidence risk analysis using multivariate copula functions. Journal of Hydrologic Engineering, 17(6): 742-755.

Chen LS, Tzeng IS, Lin CT (2013). Bivariate generalized gamma distributions of Kibble's type. Statistics, (ahead-of-print), 1-17.

Chowdhary H, Escobar LA, Singh VP (2011). Identification of suitable Copulas for bivariate frequency analysis of flood peak and flood volume data. Hydrology Research; 42: 193-215.

De Michele, C., & Salvadori, G. (2003). A generalized Pareto intensity‐duration model of storm rainfall exploiting 2‐copulas. Journal of Geophysical Research: Atmospheres, 108(D2).

Fu, G., & Kapelan, Z. (2013). Flood analysis of urban drainage systems: Probabilistic dependence structure of rainfall characteristics and fuzzy model parameters. Journal of Hydroinformatics, 15(3), 687-699

Kao, S. C., & Chang, N. B. (2011). Copula-based flood frequency analysis at ungauged basin confluences: Nashville, Tennessee. Journal of Hydrologic Engineering, 17(7), 790-799.

Keef, C., Tawn, J. A., & Lamb, R. (2013). Estimating the probability of widespread flood events. Environmetrics, 24(1), 13-21.

Reddy MJ, Ganguli P (2012). Bivariate flood frequency analysis of upper Godavari River flows using Archimedean Copulas. Water Resources Management, 26(14): 3995-4018.

Requena A I, Mediero L, Garrote L (2013). Bivariate return period based on copulas for hydrologic dam design: comparison of theoretical and empirical approach. Hydrology and Earth System Sciences Discussions, 10: 557-596.

Salarpour M, Yusop Z, Yusof F, Shahid S, Jajarmizad M (2013) Flood frequency analysis based on t-copula for Johor River, Malaysia Journal of Applied Sciences 13:1021-1028

Salvador G, Michele CD (2004). Frequency analysis via Copulas: Theoretical aspects and applications to hydrological events. Water Resources Research; 12:W12511.

Salvadori, G., & De Michele, C. (2013). Multivariate extreme value methods. In Extremes in a changing climate (pp. 115-162). Springer Netherlands.

Sklar, A. Fonctions de répartition à n dimensions et leurs marges. Paris Institute of Statistics. 1959, 8, 229-231.

Xie H, Wang K (2013). Joint-probability Methods for Precipitation and Flood Frequencies Analysis. Third International Conference on Intelligent System Design and Engineering Applications.

Zhang, L.; Singh, V.P. Bivariate Rainfall and Runoff Analysis Using Entropy and Copula Theories. Entropy. 2012. 14, 1784-1812. DOI: 10.3390/e14091784.