Malaysian Journal of Civil Engineering 30(3):395-404 (2018)

BIVARIATE FREQUENCY ANALYSIS OF FLOOD VARIABLES USING COPULA IN KELANTAN RIVER BASIN

Mahiuddin Alamgir, Tarmizi Ismail & Muhammad Noor*

Department of Hydraulics & Hydrology, Faculty of Civil Engineering, Universiti Teknologi Malaysia (UTM), 81310 Johor Bahru, Malaysia

*Corresponding Author: *mnkakar@gmail.com*

Abstract: A copula-based methodology is presented in this study for bivariate flood frequency analysis of Kelantan river basin located in Northeast Malaysia. The joint dependence structures of three flood characteristics, namely, peak flow (Q), flood volume (V) and flood duration (D) were modelled using t-Copula. Various univariate distribution functions of flood variables were fitted with observed flood variables to find the best distributions. Cumulative joint distribution functions (CDF) of peak flow and volume (Q-F), peak flow and duration (Q-D) and volume and duration (V-D) revealed that return period of joint return periods are much higher compared to univariate return period. The joint probabilities of occurrence of 0.8, 0.6, 0.4, 0.2 and 0 can be expected when flood duration greater than 65 h, 54 h, 46 h, and 32 h, and the flood volume higher than 0.62 km^3, 0.33 km3, 0.25 km^3, and 0.22 km^3 respectively.

Keywords: *Flood frequency analysis, t-copula, bivariate probability distribution.*

Article history:
Received 25 Sept 2017
Received in revised form 15 Sept 2018
Accepted 17 October 2018
Published online 1 Dec 2018

1.0 Introduction

Increase of heavy rainfall events may cause more frequent floods (Shahid, 2011). Estimation of return period of certain magnitude flood is very important for effective flood mitigation planning and reliable designing of hydraulic structure. Usually, this is done using a flood frequency analysis on a long-term flood record. Univariate frequency analysis of flood magnitude is generally carried out for estimation of flood return period. However, 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 all flood variables namely, flood peak, flood volume and flood duration together.

Flood duration is a major factor in analyzing the costs and impacts of a flooding event or any future flooding event (FEMA, 2006). Building materials that are in prolonged contact (more than 72 hours) with floodwaters may be affected very differently than materials in short-term contact (FEMA, 2006). Long-duration flooding can also damage building materials that are relatively flood-resistant and would not be damaged in short-term contact with floodwaters. The accessibility and mobility in urban area are affected more severely by long inundation compared to short duration flood. Similarly, flood volume is required to consider for protecting the newly developed regions of Johor river basin. Structural measure like building dams is the major policy of Malaysian government to adapt with growing frequency and severity of floods. In order to design of retention basins and spillways of reservoirs, as well as other hydraulic structures where storage is involved, the entire hydrograph, or at least volume estimates are required along the peak discharges, in order to calculate the effect of the inflow on the storage, and therefore failure probabilities (Mediero *et al.*, 2010, and Gaál *et al.*, 2015).

It is very clear from above discussion that all flood parameters, such as peaks, volumes and duration are needed for practical applications in hydrology. The cause of often failure of structural in Malaysia from moderately severe of floods may be due to structural development based on the flood peaks alone, which have been historically conducted in Malaysia. Multivariate frequency analysis has been employed by several researchers after recognizing the weaknesses of univariate frequency analysis. An emerging statistical tool called Copula was employed to construct the dependency structure and joint probabilistic distributions (Zhang *et al.*, 2006, Zhang *et al.*, 2012, and Chen *et al.*, 2014), which allows to model the correlations among flood variables (Salvadori *et al.*, 2004, and Li *et al.*, 2012). Number of studies on joint distribution of flood frequency analyses using Copula has been conducted across the world. Li *et al.*, (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 *et al.*, (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. Xie and Wang, (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.

The objective of this study determine probability of occurrence of three flood variables were peak flow, duration and volume using t-Copula. For this purpose, the flood variables are separately modelled by a probabilistic distribution function. The fitted models are linked using the concept of Copula to construct joint bivariate distribution function of peak flow-duration, peak flow-volume and volume-duration. The joint flood variable’s curve of various return intervals for Kelantan River in Malaysia was developed by the proposed approach.

2.0 Materials and Methods

2.1 Materials

Forty-three years (1972-2014) hourly river discharge data records from twelve stations distributed over the Kelantan catchment was used for the study. Hourly stream flow data of Kelantan river basin was collected from the Department of Irrigation and Drainage (DID), Malaysia. Kelantan is a rural state located in the northeast of Peninsular Malaysia, having an area of about 5,099 km² (Figure 1).

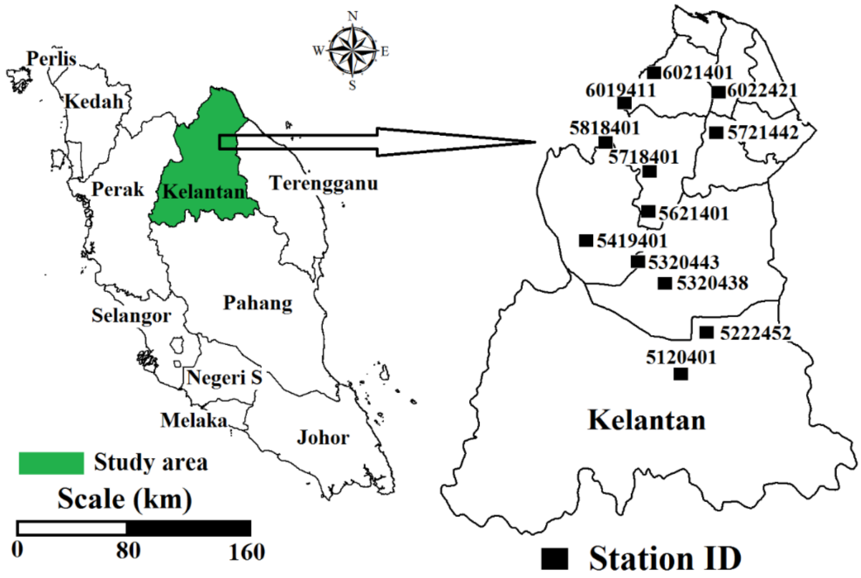

Figure 1: Location of the study area (Kelantan) in the map of peninsular Malaysia

2.2 Methodology

Univariate 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 cumulative distribution function (CDF) is defined as:

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

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

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

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

$$\int_b^a f(x)dx \ = P(a \leq X \leq b) \qquad (3)$$

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

$$f(x) = P(X = x) \qquad (4)$$

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.

The joint probability of occurrence of flood variables was conducted using t-Copula. A copula captures the dependence of two or more random variables. The Sklar's theorem (Sklar, 1959) states that the joint behavior of random variables (X, Y) with continuous marginal distributions u= F_x (x)=P(X≤x) and v=F_y (y)=P (Y≤y) can be characterized uniquely by its associated dependence function or copula, C. For 2-dimensional cases, all (u, v) relationships can be written as:

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

where, $F_{x,y}$ (X,Y) is the joint CDF of random variables X and Y and also V x,y ER When I = [0,1], the bivariate copula has a distribution function of C= I2→I which normally satisfies the following basic properties:

3.0 Result and Discussion

The summary statistics of the flood parameters at twelve station are given in Table 1. The averages of flood duration, flood volume and peak flow of the study area varies between 23 to 171 hours, 0.1 to 1.38 km3, and 81 to 7140 m3/sec, respectively.

Table 1: Summary statistics of flood parameter

Station ID	*Statistics*	*Duration (hour)*	*Volume (km^3)*	*Flow (m^3/s)*	*Station ID*	*Statistics*	*Duration (hour)*	*Volume (km^3)*	*Flow (m^3/s)*
5120401	Mean	68	1.2	7140	5718401	Mean	171	0.0	94
	Max	142	7.1	19237		Max	4216	0.1	219
	Min	3	0.0	1388		Min	10	0.0	13
	SD	42	1.7	5259		SD	729	0.0	61
5222452	Mean	58	0.3	1669	5721442	Mean	82	1.4	6223
	Max	110	1.1	4785		Max	159	7.1	19237
	Min	5	0.0	96		Min	3	0.0	835
	SD	25	0.3	1164		SD	41	1.5	4401
5320438	Mean	23	0.2	1593	5818401	Mean	54	0.1	346
	Max	67	0.3	2310		Max	392	0.3	1153
	Min	3	0.0	303		Min	2	0.0	12
	SD	25	0.1	699		SD	75	0.1	325
5320443	Mean	48	0.4	2563	6019411	Mean	136	0.1	435
	Max	91	1.9	8481		Max	354	0.5	1218
	Min	5	0.0	585		Min	3	0.0	81
	SD	19	0.4	1596		SD	72	0.1	207
5419401	Mean	31	0.0	496	6021401	Mean	58	0.1	206
	Max	75	0.3	1275		Max	150	0.2	467
	Min	12	0.0	102		Min	11	0.0	34
	SD	14	0.1	302		SD	46	0.1	147
5621401	Mean	53	0.0	192	6022421	Mean	77	0.0	81
	Max	273	0.1	360		Max	243	0.0	187
	Min	11	0.0	21		Min	11	0.0	25
	SD	65	0.0	130		SD	51	0.0	49

Fitting result for various distributions of flood variables are given in Table 2. The table shows that different distribution best fit different flood variables at different station. Overall, normal or log-normal is found to fit flood duration in most of the station, exponential or log-normal for flood volume, and gamma or Gumble for peak flood flow. The parameters of fitted distribution are also given in Table 3.

Table 2: Fitted distribution and distribution parameters of flood variables

Station ID	*Variable*	*Distribution*	*Parameter*	*Station ID*	*Variable*	*Distribution*	*Parameter*
5120401	D	Normal	68.3, 40.3	5718401	D	Cauchy	28.96,10.91
	V	Exponential	0.82		V	Log-normal	-4.99,1.13
	F	Gamma	0.74, 0.60		F	Gumbel	0.01,0.01
5222452	D	Exponential	58.3	5721442	D	Normal	81.9, 40.7
	V	Normal	25.1, 3.70		V	Exponential	0.73
	F	Gamma	0.95, 3.51		F	Gamma	0.97, 0.70
5320438	D	Normal	0.04	5818401	D	Log-normal	3.29, 1.26
	V	Exponential	6.70		V	Weibull	0.69, 0.05
	F	Gamma	0.08, 0.11		F	Gumbel	0.03, 0.04
5320443	D	Normal	47.6, 18.4	6019411	D	Gumbel	103.3, 60.41
	V	Log-normal	-1.31, 0.82		V	Gumbel	0.10, 0.07
	F	Gamma	1.67,4.45		F	Gumbel	0.10,0.07
5419401	D	Log-normal	3.34, 0.43	6021401	D	Log-normal	3.77, 0.79
	V	Log-normal	-3.71, 1.07		V	Log-normal	-4.05,1.50
	F	Gamma	0.02, 0.03		F	Gumbel	0.02,0.04
5621401	D	Log-normal	3.62, 0.73	6022421	D	Gumbel	1.39,0.02
	V	Exponential	50.0		V	Weibull	0.01
	F	Gamma	0.01, 0.01		F	Gumbel	0.01

NB: D for Duration (hour), V for Volume (km3), and F for Flow (m3/sec)

After univariate analysis, the pairs of variables were modelled by t-copula. The t-Copula linear correlation parameters among flood variables are given in Table 3. The table shows high dependency among flood variables, particularly dependency between flood volume and duration was found very high at all stations. Association between peak flow and flood volume was also found very high is many stations. Correlations among flood variables at a station are depicted in Figure 2 as an example.

Table 3: Correlation among parameters, duration (D), volume (V), and flow (F)

Station ID		*D*	*V*	*F*	*Station ID*		*D*	*V*	*F*
5120401	D	1.00	0.96	0.10	5718401	D	1.00	0.71	0.31
	V	0.96	1.00	0.20		V	0.71	1.00	0.67
	F	0.10	0.20	1.00		F	0.31	0.67	1.00
5222452	D	1.00	0.70	0.49	5721442	D	1.00	0.90	0.41
	V	0.70	1.00	0.88		V	0.90	1.00	0.61
	F	0.49	0.88	1.00		F	0.41	0.61	1.00
5320438	D	1.00	0.67	0.25	5818401	D	1.00	0.37	-0.08
	V	0.67	1.00	-0.12		V	0.37	1.00	0.68
	F	0.25	-0.12	1.00		F	-0.08	0.68	1.00
5320443	D	1.00	0.77	0.43	6019411	D	1.00	0.69	0.08
	V	0.77	1.00	0.87		V	0.69	1.00	0.55
	F	0.43	0.87	1.00		F	0.08	0.55	1.00
5419401	D	1.00	0.52	0.31	6021401	D	1.00	0.79	0.72
	V	0.52	1.00	0.77		V	0.79	1.00	0.99
	F	0.31	0.77	1.00		F	0.72	0.99	1.00
5621401	D	1.00	0.74	0.21	6022421	D	1.00	0.44	0.18
	V	0.74	1.00	0.78		V	0.44	1.00	0.86
	F	0.21	0.78	1.00		F	0.18	0.86	1.00

NB: D for Duration (hour), V for Volume (km3), and F for Flow (m3/sec)

Each pair of flood variables are fitted with t-Copula and Bivariate cumulative distribution (CDF) were prepared. Bivariate CDF of flood duration and volume is given in Figure 3. The joint probability curve for flood volume and duration is presented in Figures 3 as an example. In the figure the probabilities are shown in intervals of 0.1, from 0.1 to 0.9. From Figure 3 the joint probabilities of occurrence of 0.8, 0.6, 0.4, 0.2 and 0 can be expected when flood duration greater than 65 h, 54 h, 46 h, and 32 h, and the flood volume higher than 0.62 km3, 0.33 km3, 0.25 km3, and 0.22 km3 respectively.

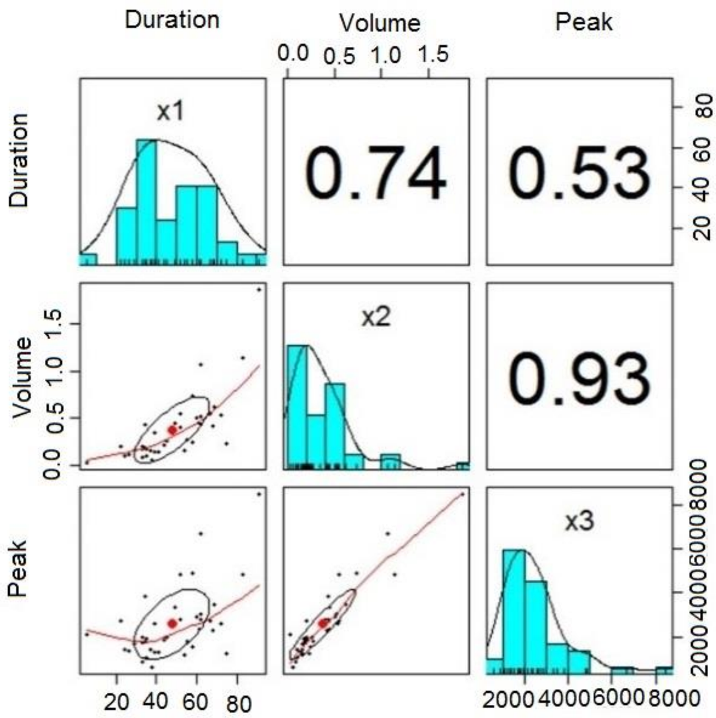

Figure 2: Correlation among flood variable duration, volume and peak

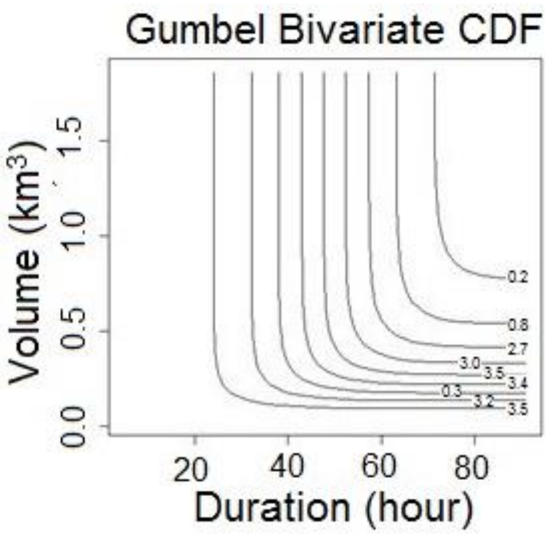

Figure 3: Bivariate cumulative distribution function of volume and duration

4.0 Conclusions

Information related to the peak flow, flood duration and flood volume are necessary to design hydraulic structures for water resources development and management as well as flood assessment and mitigation. The t-copula has been proposed in this study for the modelling of joint dependence of these flood characteristics i.e peak flow-volume, volume-duration and peak flow-duration for the Kelantan river basin. Obtained results reveal that join return period of flood variables are much higher compared to univariate return period of flood variables. It is expected that the Bivariate cumulative distribution of flood variables prepared in this article will help in effective flood mitigation planning and reliable designing of hydraulic structure in the study area.

5.0 Acknowledgements

We are grateful to the Ministry of Higher Education (MOHE) Malaysia and Universiti Teknologi Malaysia (UTM) for financial support of this research through the FRGS research project (Vote No. R. J130000.2522.13H07).

References

Chen, Li-Shya, I-Shiang Tzeng, and Chien-Tai Lin (2014). *Bivariate generalized gamma distributions of Kibble's type*. Statistics 48(4) 933-949.

Chowdhary, H., Escobar, L. A., and Singh, V. P. (2011). *Identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data*. Hydrology Research. 42(2-3), 193-216.

FEMA (2006). *Hurricane Katrina in the Gulf Coast (FEMA 549).* Federal Emergency Management Agency, Washington DC.

Gaál, L., Szolgay, J., Kohnová, S., Hlavčová, K., Parajka, J., Viglione, A., and Blöschl, G. (2015). *Dependence between flood peaks and volumes: a case study on climate and hydrological controls*. Hydrological Sciences Journal, 60(6) 968-984.

Janga Reddy, M., and Ganguli, P. (2012). *Application of copulas for derivation of drought severity–duration–frequency curves*. Hydrological Processes. *26*(11) 1672-1685.

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

Li, T., Guo, S., Chen, L., and Guo, J. (2012). *Bivariate flood frequency analysis with historical information based on copula*. Journal of Hydrologic Engineering. 18(8) 1018-1030.

Mediero, L., Jiménez-Álvarez, A., and Garrote, L (2010). *Design flood hydrographs from the relationship between flood peak and volume*. Hydrology and Earth System Sciences. 14(12), 2495-2505.

Salvadori, G., and De Michele, C. (2004). *Frequency analysis via copulas: Theoretical aspects and applications to hydrological events.* Water Resources Research. 40(12).

Shahid, S (2011*). Trends in extreme rainfall events of Bangladesh*, Theoretical and Applied Climatology. 104(3-4): 489-499.

Sklar, M. (1959). *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8.

Xie, H., and Wang, K. (2013). *Joint-probability methods for precipitation and flood frequencies analysis.* In Intelligent System Design and Engineering Applications (ISDEA), 2013 Third International Conference on (pp. 913-916). IEEE.

Zhang, L., and V. P. Singh. Bivariate flood frequency analysis using the copula method (2006). Journal of Hydrologic Engineering. 11 (2) 150-164.

Zhang, Qiang, Mingzhong Xiao, Vijay P. Singh, and Jianfeng Li. (2012). *Regionalization and spatial changing properties of droughts across the Pearl River basin, China*. Journal of hydrology 472 355-366.