# BIVARIATE FLOOD FREQUENCY ANALYSIS OF SUNGAI LUI CATCHMENT IN MALAYSIA USING ARCHIMEDEAN COPULAS

**[1]ATAN I., [2]MUHAMAD NAJMI N.S., [3]SAMSUDIN M.B., [4]JAAFAR J., [5]ASHAARI Y., [6]MOHAMMAD S.N., [7]BAKI A.**

[1]Department of Civil Engnineering, Kuala Lumpur Infrastructure University, Selangor, Malaysia,
[2]Faculty of Civil Engnineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[3]Envirab Services, P.O. Box 7866, GPO Shah Alam, 40730, Selangor, Malaysia
Faculty of Civil Engnineering, Universiti Teknologi MARA, Cawangan Johor Kampus Pasir Gudang, Johor, Malaysia
City University College, Petaling Jaya, Selangor, Malaysia
E-mail: ismail585@gmail.com, jucyber@yahoo.com, yasminashaari@yahoo.com, envirab@gmail.com, shahrul9688@johor.uitm.edu.my, aminbaki@msn.com

**Abstract-** In order to formulate mitigating measures against flood damages, flood frequency analyses are required. There are several methods of flood frequency analyses applied throughout the world. This study will focus on bivariate flood frequency analyses. Archimedean copulas have been adopted, using Gumbel, Frank and Clayton copulas. Sungai Lui catchment in Malaysia was adopted as the study area. Akaike Information Criterion (AIC) was used to assess the copulas. L-moment was used to assess the distributions. Graphical methods were used to find relationships between Flood Peak, Flood Volume and Flood Duration. Results show that Generalised Extreme Value (GEV) was the best fitted distribution for Flood Peak, and Weibul were the best fitted distribution for Flood Volume and Flood Duration. Gumbel was found to be the best copulas with minimum AIC values. Relationships were established between Flood Peak-Flood Volume, Flood Peak-Flood Duration and Flood Volume-Flood Duration.

**Index Terms-** Akaike Information Criterion, Clayton Copula, Flood frequency analysis, Frank Copula, Gumbel Copula, Joint Cumulative Distribution, Marginal Distribution Parameter.

## I. INTRODUCTION

Floods are disasters that have taken many lives, damaged properties and the environment. Floods are common events occurring throughout the world. In Malaysia, there are 189 river systems, of which 89 are in Peninsular Malaysia, 78 are in Sabah and 22 are in Sarawak. All river systems discharge immediately to the sea [1]. There are several types of flood in Malaysia, including: rainstorm flood, city-based flash flood, debris flow and mud flow, landslide, tidal flood, dam release, and bund failure [2]. However, Smith [3] classified floods into two main categories: river floods and coastal floods, to differentiate the effects of tidal for coastal and no tidal effects for river.

In order to reduce fatalities and damages, mitigation measures are required. Flood frequency analyses are often required to estimate the peak flows and frequencies, which will enable mitigation measures to be formulated. Flood frequency analyses are commonly applied in many countries including: Malaysia [2,4], United Kingdom [5], United States [6], Australia [7], New Zealand [8], Canada [9] and Italy [10].

Several approaches have been used in flood frequency analyses. The univariate flood frequency analysis is widely used in hydrological studies. Often only flood peak or flood volume is statistically analyzed. There were also studies where more variables or multivariate flood frequency analyses have been used [11,12,13,14]. In this study however, the focus will be on bivariate flood frequency analyses.

Several methods can be used to perform multivariate flood frequency analyses [15]. One of the methods was to use copulas [16,17,18]. Copula can be defined a multivariate probability distribution for which the not important probability of each variable is uniform [19]. It was used to describe the dependence between random variables and popular variables in high-dimensional statistical applications. Copulas enable one to easily model and calculate the distribution of random vectors by estimating not important and copula separately [20].

Copulas are used to determine the dependence between random variables [21]. The Sklar's theorem [19] states that, the joint behaviour of a random variable (X, Y) with continuous marginal $u = Fxy(X) = P(X \leq x)$ and $v = FY(Y) = P(Y \leq y)$ can be expressed by its associated dependence function or copula C. For two (2) dimensional case, for all $(u, v) \in [0,1]2$the relationship can be formulated as below:

$$Fx,y(X,Y) = C[Fx(X), Fy(Y)] = C(u,v) \quad (1)$$

Fx,y(x,y) is joint cumulative distribution function (CDF) of random variables X and Y.

Archimedean copulas admit a clear formula and known as a class of copulas [18]. Archimedean copulas are popular because they allow modelling dependence in randomly high dimensions with only one limit/guideline, governing the strength of addiction [22]. Examples of copulas are Clayton, Gumbel and Frank Copula, which will be used in this

study. Table 1 shows summary of the Bivariate Archimedean copulas with their corresponding generator.

**Table 1: The Bivariate Archimedean Copulas with their corresponding generator**

| Name | Bivariate copula $C_\theta(u,v)$ | Generator $\varphi_\theta(t)$ | Parameter $\theta$ |
|---|---|---|---|
| Clayton | $(\max\{u+v-1;0\})^{-\frac{1}{\theta}}$ | $\frac{1}{\theta}(t^{-\theta}-1)$ | $\theta \in [-1,\infty)\setminus\{0\}$ |
| Gumbel | $\exp(-((-\log(u))^{\theta}+(-\log(v))^{\theta}))$ | $(-\log(t))^{\theta}$ | $\theta\epsilon[1,\infty)$ |
| Frank | $-\frac{1}{\theta}\log(1+\frac{(\exp(-\theta u)-1)(\exp(-\theta v)-}{\exp(-\theta)-1}$ | $-\log\left(\frac{\exp(-\theta t)-}{\exp(-\theta t)-}\right.$ | $\theta\epsilon R\setminus\{0\}$ |

The not important distribution is probability explanation and static is the probability distribution of the variables contained in the subset. It presented the probabilities of different values of the variables in the subset without reference to the values to other variables. Basically term not important distribution is applied to those variables in the subset of variables being held. It also refers to the process of removing the influence of one or more event form probability [23]. For example if X and Y are separate random variables and f (x,y) is the value of their joint probability distribution at (x,y), the function will be given as:

$$g(x) = f(x) = \sum_y (f(x,y) and\ h(y) \sum_x f(x,y) \quad (2)$$

In this study, the objective is to perform bivariate flood frequency analyses using Archimedean copula. Data from Sungai Lui catchment in Malaysia will be used to assess the approach.

## II. MATERIALS AND METHODS

Sungai Lui catchment is subjected to anthropogenic manipulations in hydrological status. Some landforms resulting from urban growth and agricultural activities were not caught in the land use and topography maps. This omission included most of the pool, which can affect the sedimentation process via an increase in deposition rate [24]. Sungai Lui catchment comprises steep slopes, and is exposed to high intensity rain.

Sungai Lui catchment is located 300 07' – 300 12' N, and 10100 52' – 10100 58' E on the upstream of the Langat River with a drainage area of 68.25 km2 and a basin length of 11.5 km as shown Figure 1. The lower limit and maximum heights of the basin are 61 m and 1207 m, respectively, while the average altitude is approximately 354 m above the sea level. Figure 2 shows Sungai Lui catchment is steep with an average slope 35%. Approximately 60% of the catchment areas have a gradient that is higher than 30%. Sungai Lui (river) hydrometer station is situated at the outlet of Sungai Lui catchment with an average annual water discharge of 55.05 x 106 m3 and an average annual sediment load of 5.88 x 103 tonnes. The average annual precipitation at Kg. Sungai Lui rain gauge station is approximately 2188.3 mm. In terms of land utilization, Sungai Lui catchment comprises 80.35% forest, 9.85% rubber and 2.6% orchards. The remaining part of this catchment area consists of mixed horticulture and crops, urban areas, and mining land. Three (3) characteristic of floods: flood peak, flood volume and flood duration will be used in this study. Bivariate joint distributions of the flood characteristics will be derived using the concept of copula considering a lot of parameters and nonparametric marginal distribution for these three characteristics. A set of parametric distribution functions, and nonparametric methods based on density estimation and orthonormal series will be used to determine these three characteristics. In conventional methods of flood frequency analysis, the marginal distribution functions of peak flow, mass and duration are assumed to conform to some specific parametric distribution function. Each parameter was evaluated by using L-moment method [25,10]. The concept of copula relaxes the restriction of the traditional flood frequency analysis by selecting marginal from different families of probability distribution functions. Akaike Information Criterion (AIC) [26] and graphical method will be applied in determination of copula model and joint distribution between flood variables. In this study, the Archimedean copula has been adopted.

AIC is a measure of the relative quality of statistical model for a given set of data [26]. AIC deals with two (2) models: a trade-off between the goodness of fit of the model and complexity of the model. Basically AIC does not provide testing a null hypothesis as it can say nothing about the quality of the model in an absolute sense. For any statistical model, the AIC value is:

$$AIC = 2k - 1\ln(L) \quad (3)$$

where, k is the number of parameters in the model

L is the maximized value of the likelihood/parameters of statistical model. L-moment standardised are called L-moment ratios and it is a sequence of statistics used to summarise the shape of a probability distribution [25]. L-moments are analogous to standardized moments which compute sample statistics of individual data at the site, defining distribution parameters and the suitability of the probability distribution.

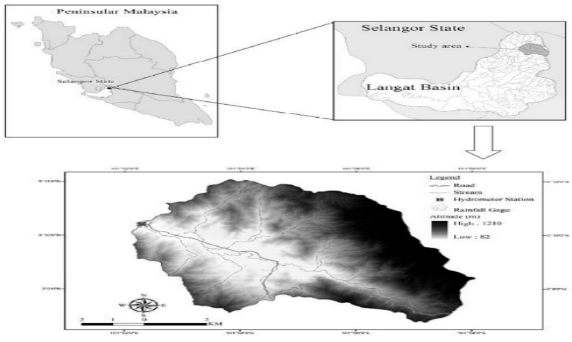



**Figure 1: Geographic location of Sungai Lui catchment**

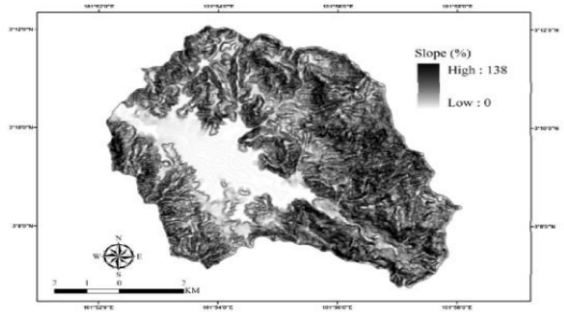



**Figure 2: Slope map of Sungai Lui catchment**

## III. RESULTS AND DISCUSSIONS

The joint cumulative distribution function of peak flow and duration, peak flow and volume and duration and volume are illustrated in Figures 3, 5 and 7, respectively. The contour lines of joint distribution of each pairs of flood characteristics are illustrated in separate Figures 4, 6 and 8, respectively.

The contour lines of joint cumulative distribution of peak flow and duration as indicated in Figure 3 described for each cumulative distribution (CUV) contour. There is an opposite relationship between peak flow and duration. If flood peak is high, the flood duration will be low and vice versa. With higher cumulative distribution centres, it allows higher peak flow and flood duration. In Figure 4, the joint probability graph presented the scenes of two conditions with peak flow and flood duration occurring at the same time with the combination distribution contours interval at 0.2, 0.4, 0.6, 0.8 and 1.0 in X and Y axis.

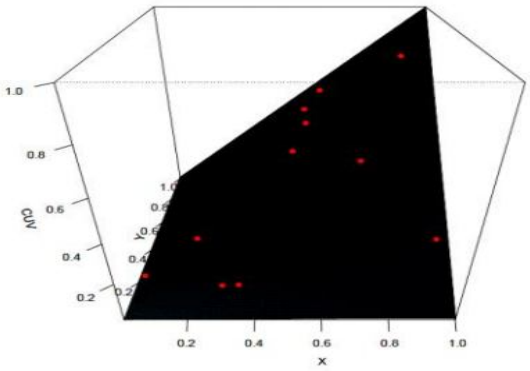



**Figure 3: Joint cumulative distribution function of peak flow and duration**

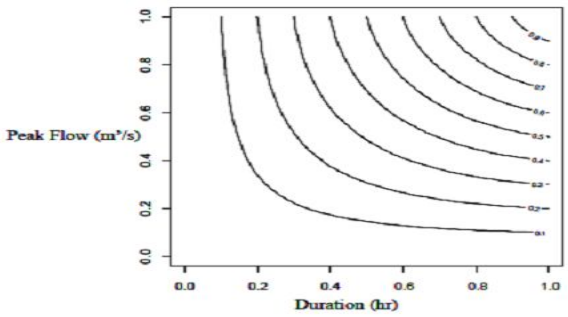



**Figure 4: Contours showing two-dimensional view of joint cumulative distribution function of peak flow and duration**

Contours lines of joint distribution of peak flow and volume are presented in Figure 5. The flood characteristics are often important for flood volume based on cumulative distribution (CUV). Figure 6 shows the flood volume observation and prediction of daily flows for all years are close to each other based on the flood volume. Madiero et al. [27] found that linear relationship exists between standardised peaks and volumes in homogeneous regions.

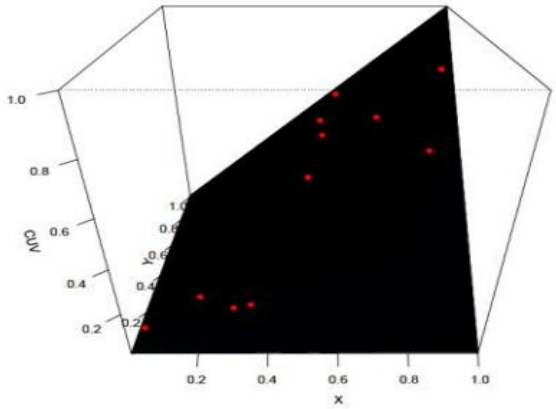



**Figure 5: The joint cumulative distribution function of peak flow and flood volume**

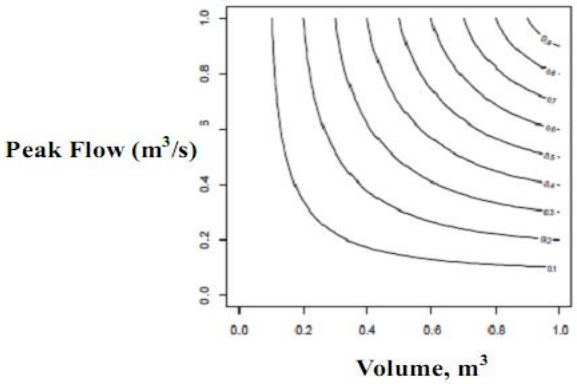



**Figure 6: Contours showing the two-dimensional view of joint cumulative distribution function of peak flow and volume.**

Figure 7 shows positive correlations between flood volume and duration. Figure 8 illustrates the chance of two flood characteristic with peak duration and flood volume occurring at the same time which also indicates positive correlation between volume and duration.

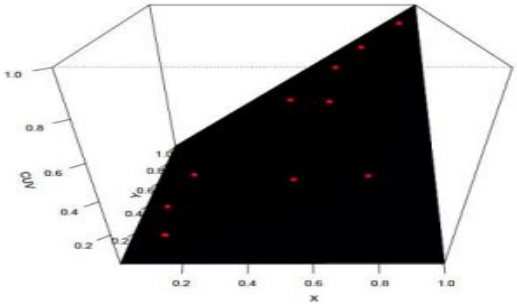



**Figure 7: Joint cumulative distribution function of flood volume and flood duration**

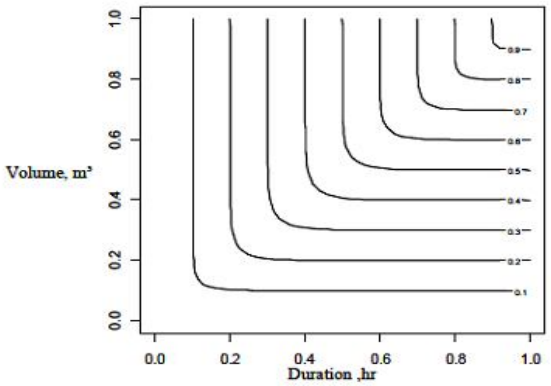



**Figure 8: Contours showing the two-dimensional view of joint cumulative**

Table 2 presents the parameter for α, β and γ based on L-Moment method and Table 3 is the summary of value τ, θ and AIC. The best value of α, β and γ from Table 2 is the best fitted distribution was Generalised Extreme Value (GEV) from peak flow flood variable. This was consistent with the findings of Salarpour et al. [28] where GEV is the best marginal distribution fit for peak flow. Noto and La Loggia [10] also found GEV to be the most robust distributions. However, for flood duration and flood volume, Weibull was found to be the best fitted distribution.

**Table 2: Parameter of marginal distribution by using method of L-Moment**

| Flood Variable | Best Fitted Distribution | Parameter |
|---|---|---|
| **Peak Flow** | Gev. Extreme Value (GEV) | α = 0.4421<br>β = 66.243<br>γ = 141.462 |
| **Duration** | Weibull | α = 0.2762<br>β = 76.88<br>γ = 135.65 |
| **Volume** | Weibull | α = 0.8842<br>β = 86.542<br>γ = 168.99 |

Flood Variable Best Fitted Distribution Parameter

Peak Flow
Gev. Extreme Value (GEV) α = 0.4421
β = 66.243
γ = 141.462

Duration
Weibull α = 0.2762
β = 76.88
γ = 135.65

Volume
Weibull α = 0.8842
β = 86.542
γ = 168.99

In Table 3, the best copula for flood frequency was Gumbel compared to Clayton and Frank copula because of the AIC value. Zhang and Singh [29] stated that the best model of copula is when the AIC value was at minimum value. All the parameters for τ and θ were based on dependence structure into a single number and the τ was measured using Kendall's tau approach. Similar methods were used by Reddy and Ganguli [18] but they found Frank copula to be more suitable for their sets of data.

**Table 3: Estimation of τ, θ, and AIC value for Clayton, Gumbel and Frank Copula**

| | Flow-Duration | | | Flow-Volume | | | Volume-Duration | | |
|---|---|---|---|---|---|---|---|---|---|
| | Clayton | Gumbel | Frank | Clayton | Gumbel | Frank | Clayton | Gumbel | Frank |
| τ | 0.0909091 | 0.0909091 | 0.0909091 | 0.5636364 | 0.5636364 | 0.5636364 | 0.3454545 | 0.3454545 | 0.3454545 |
| θ | 1.10 | 1.204721 | 1.829922 | 2.291667 | 1.507343 | 1.347723 | 1.527778 | 5.982673 | 3.336723 |
| AIC | -2.84201 | *-2.19223* | -3.55621 | -58.72132 | *-42.66982* | -86.16722 | -345.16782 | *-189.44721* | -567.23567 |

Copulas provide a convenient way to model and test related variables. Several copulas with different builds are

Flow-Duration Flow-Volume Volume-Duration
Clayton Gumbel Frank Clayton Gumbel Frank Clayton Gumbel Frank
τ 0.0909091 0.0909091 0.0909091 0.5636364 0.5636364 0.5636364 0.3454545 0.3454545 0.3454545
θ 1.10 1.204721 1.829922 2.291667 1.507343 1.347723 1.527778 5.982673 3.336723
AIC -2.84201 -2.19223 -3.55621 -58.72132 -42.66982 -86.16722 -345.16782 -189.44721 -567.23567

available for modelling these relationships. Shape differences among copulas can be seen with the descriptive functions. These can be used both in fitting copulas to date and in applying informed judgment to select a copula for a given application. This study focused on Bivariate Copulas but many of the estimates can be extrapolated to the multivariate cases. The result accomplished from this study was by using Archimedean copula, which was useful to estimate peak flow, flood volume, flood duration and the joint distribution functions of flood characteristics.

## CONCLUSION

Bivariate flood frequency analyses of Sungai Lui catchment has been satisfactory using Archimedean copulas. Assessment using AIC has found Gumbel copula to be the most satisfactory. Assessment using L-moment has found that GEV was the best fitted distribution for Flood Peak, whilst Weibull was the best fitted distribution for Flood Volume and Flood Duration. Relationships were also established between

flood peak-flood volume, flood peak-flood duration and flood volume-flood duration. In conclusion, copulas can be used in flood frequency analyses to assist water managers in flood prediction and formulating appropriate mitigating measures.

## REFERENCES


[1] DID (2009), DID Manuals: Volume 5 – Irrigation and Agricultural Drainage, Department of Irrigation and Drainage Malaysia.

[2] DID (2000), Urban Stormwater Management Manual for Malaysia, MSMA, Department of Irrigation and Drainage Malaysia.

[3] D.I. Smith (1998). Water in Australia: resources and management. Oxford University Press, USA.

[4] DID (2012), Urban Stormwater Management Manual for Malaysia, MSMA 2nd Edition, Department of Irrigation and Drainage Malaysia.

[5] Natural Environmental Research Council (NERC) (1975), Flood studies report, II, Meteorological studies, Swindon, England.

[6] H.C. Riggs (1973). Regional analyses of streamflow characteristics. US Government Printing Office.

[7] D. Pilgrim (2001), Australian Rainfall and Runoff: Volume 1 – A Guide to Flood Estimation, The Institution of Engineers Australia.

[8] M.E. Beable (1982). Regional flood estimation in New Zealand.

[9] GREHY, G. D. R. E. S. (1996). Presentation and review of some methods for regional flood frequency analysis. Journal of hydrology (Amsterdam), 186(1-4), 63-84.

[10] L.V. Noto and G. La Loggia (2009), Use of L-Moments Approach for regional flood frequency analysis in Sicily, Italy. Water Resources Management, 23 (11): 2207-2229

[11] B. Sackl and H. Bergmann (1987). A bivariate flood model and its application. In: Hydrologic Frequency Modelling. Ed: V.P. Singh, Reidel, Dordrecht, The Netherlands 571-582

[12] N.K. Goel, S.M. Seth and S. Chandra (1998). Multivariate modelling of flood flows. Journal of Hydraulic Engineering, 124(2):146-155.

[13] A.W. Marshall and I. Olkin (1988). Families of multivariate distributions. Journal of the American Statistical Association, 83(403), 834-841.

[14] S. Grimaldi and F. Serinaldi (2006). Asymmetric copula in multivariate flood frequency analysis. Advances in Water Resources, 29: 1155-1167.

[15] H. Joe (1997). Multivariate Models and Dependence Concepts. Chapman & Hall. London.

[16] G. Salvadori and C. De Michele (2004). Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. Water Resources Research, Vol. 40. 10.1029/2004WR003133

[17] T. Li, S. Guo, L. Chen and J. Guo (2012). Bivariate flood frequency analysis with historical information based on copula. Journal of Hydrologic Engineering, 10.1061/ (ASCE) HE.1943-5584.0000684

[18] J.M. Reddy and P. Ganguli (2012), Bivariate Flood Frequency Analysis of Upper Godavari River Flows Using Archimedean Copulas, Water Resources Management, November 2012, Volume 26, Issue 14, pp 3995-4018.

[19] Sklar (1959), Fonctions de répartition à n dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris, 8, pp. 229–231.

[20] E.W. Frees and E.A. Valdez (1998), Understanding Relationships Using Copulas, North American Actuarial Journal, Vol. 2, Issue 1, pp. 1-25,

[21] R.B. Nelsen (1999). An introduction to copulas. Springer.

[22] F. Wu, E. Valdez and M. Sherris (2007). Simulating from exchangeable Archimedean copulas. Communications in Statistics—Simulation and Computation®, 36(5), 1019-1034.

[23] C. Genest, K. Ghoudi and L.P. Rivest (1998). "Understanding Relationships Using Copulas," by Edward Frees and Emiliano Valdez, January 1998. North American Actuarial Journal, 2(3), 143-149.

[24] H. Memarian, S.K. Balasundram, J.B. Talib, A.M. Sood and K.C. Abbaspour (2012). Trend analysis of water discharge and sediment load during the past three decades of development in the Langat basin, Malaysia. Hydrological Sciences Journal, 57(6), 1207-1222.

[25] J.R.M. Hosking and J.R. Wallis (2005). Regional frequency analysis: an approach based on L-moments. Cambridge University Press.

[26] H. Akaike (1974). A new look at the statistical model identification, IEEE Transactions on Automatic Control, 19(6), 716-723.

[27] L. Madiero, A. Jimenez-Alvarez and L. Garrote (2010), Design Flood Hydrographs from the Relationship between Flood Peaks and Volume, Hydrology and Earth Sciences Discussions, 7, 4817–4849, 2010, www.hydrol-earth-syst-sci-discuss.net/7/4817/2010/ doi:10.5194/hessd-7-4817-2010

[28] M. Salarpour, Z. Yusop, F. Yusof, S. Shahid and M. Jajarmizad (2013). Flood Frequency Analysis Based on t-copula for Johor River, Malaysia. Journal of Applied Sciences, 13, 1021-1028.

[29] L. Zhang and V.P. Singh (2006). Bivariate flood frequency analysis using the Copula method. Journal of Hydrologic Engineering 11 (2): 150 – 164.


★★★