Journal of Environmental Science and Technology

ISSN 1994-7887

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Environmental Science and Technology 5 (6): 419-429, 2012
ISSN 1994-7887 / DOI: 10.3923/jest.2012.419.429

Modeling the Distributions of Flood Characteristics for a Tropical River Basin

[1]Mohsen Salarpour, [1]Zulkifli Yusop and [2]Fadhilah Yusof

[1]Department of Hydraulics and Hydrology, Faculty of Civil Engineering, Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia

[2]Department of Mathematics, Faculty of Science, Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia

Corresponding Author: Mohsen Salarpour, Department of Hydraulics and Hydrology, Faculty of Civil Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

ABSTRACT

Understanding the distribution of flood variable is crucial for the purposes of flood prediction and hydraulic design. Three major parameters that are useful to describe a flood event are peak flow or flood peak, flood volume and flood duration. This study aimed at exploring the statistical distribution of these parameters for Johor River in south of Malaysia. Hourly data were recorded for 45 years from Rantau Panjang gauging station. The annual flood peak was selected from the maximum flow in each water year (July-June). Five probability distributions namely Gamma, Weibull, Exponential, Gumbel and Generalize Extreme Value (GEV) were used to model the distribution of flood parameters. Kolmogorov-Smirnov and chi-squared goodness-of-fit tests were used to evaluate the best fit. Goodness-of-fit tests at 5% level of significance indicated that all the models can be used to model the distribution of peak flow, duration and volume. However, the Exponential distribution is the most suitable model when tested with the Kolmogorov-Smirnov test whereas GEV is the best when chi-squared test was used. The result can be used as a basis to improve flood frequency analysis for Malaysian rivers.

Key words: Flood frequency characteristic, goodness-of-fit test, Johor River, probability distribution

INTRODUCTION

Malaysia is a tropical country with high rainfall intensity. Over the last few years, a number of extreme floods had occurred, bringing great economic losses. Areas which are prone to flooding are mostly highly populated and industrialized due to rapid development. Therefore, the estimation of flood variable distribution has become a popular and practical mean to provide flood frequency analysis at different sites to manage flood control and for economical purpose (Salarpour *et al*., 2011).

Knowledge of the flood distribution plays a vital role in water resources planning of a country. The result of such studies have been found to be useful in decision making related to the planning and management of the water resources of the countries involved. Statistical distributions are being used to model the long term characteristics of flood. This paper was aimed at analyzing the statistical distribution of flood variables such as peak flow, duration and volume for long term data.

A complex flood event generally has a few parameters which are mutually correlated, namely the flood peak, volume and duration. However, past research mainly focused on flood peak and/or flood volume only Zhang and Singh (2006) and Samaniego *et al.* (2010), among others, had many reviews done on single-value flood frequency analysis. Many hydrological solutions require more complete information regarding the flood event, which should include flood volume, flood duration and time to peak. Ironically, the conventional flood frequency analysis can only provide limited assessment on the flood events and this topic has only been highlighted by a few researchers like Durrans (1998), Yue *et al.* (1999) and Yue (1999, 2000, 2001).

Thom (1951) suggested two rather important parameters for the Gamma (G2) distribution function for wet-day amounts (Buishand, 1978). The ratio of the empirical Coefficient of Variation to Coefficient of Skewness was quite close to the theoretical value for a Gamma distribution. A simple regression was also used by Geng *et al.* (1986) to show that the beta parameter of the Gamma distribution for a particular month can be reasonably predicted by the average rainfall per wet-day of the same month.

On the other hand, a three-parameter Mixed Exponential distribution was suggested by Renard and Lang (2007) instead of the Gamma distribution. The Mixed Exponential distribution became a more favorable distribution instead of the Gamma distribution through a variety of goodness-of-fit tests and log-likelihood analyses. Besides that, the Weibull and sometimes, the Exponential distribution have also been recommended for daily rainfall amount modeling (Burgueno *et al.*, 2005).

For many years, the work of Hershfield (1962), which is TP-40, serves as the most common approach to summarize flood frequency analyses in the United States. More recently, the L-moments and other powerful methods have also been employed to complete these types of analysis, rather than the more traditional goodness-of-fit measures (Bonnin *et al.*, 2006). Bonnin *et al.* (2006) also fitted a Generalized Extreme Value (GEV) distribution to the AMS of rainfall.

In this analysis, five probability distributions were considered as potential candidates. These are the two Gumbel parameters and three Gamma parameters, Generalized Extreme Value (GEV), Weibull and Exponential. The reason for selecting these distributions for analysis is due to the fact that they are commonly used in flood frequency studies (Zalina *et al.*, 2002; Danazumi and Shamsudin, 2011). The objective of the study was to explain the statistical distribution of parameters that useful to describe a flood event, volume and duration.

MATERIALS AND METHODS

Data collection and study area: Precipitation data, consisting of rainfall depth, recorded at hourly intervals were obtained from the Department of Irrigation and Drainage, Malaysia. Discharge data at the Rantau Panjang gauging station (Latitude 01 46 50 and Longitude 103 44 45) was used for flood analysis. The data, covered 45 years. Figure 1 shows the map of south Malaysia and the location of the streamflow station.

The most significant characteristics of a flood event are the peak flow (Q_p), volume (Q_v) and duration (Q_D) (Yue *et al.*, 1999; Yue and Rasmussen, 2002). In this study, base flow defines the starting and ending times of the flood event and hence, is considered as a criterion to delineate flood hydrographs. Base flow is determined largely through graphical observation of hydrograph using the methods as described by (Yusop *et al.*, 2006). The starting point of the surface runoff is marked by the abrupt rise of the rising limb of the hydrograph and ends at a point where the recession limb crosses with a separation line. The separation line in this study had a slope of 0.0005 $m^3 sec^{-1} km^{-2} h$ extended from the point of initial rise.

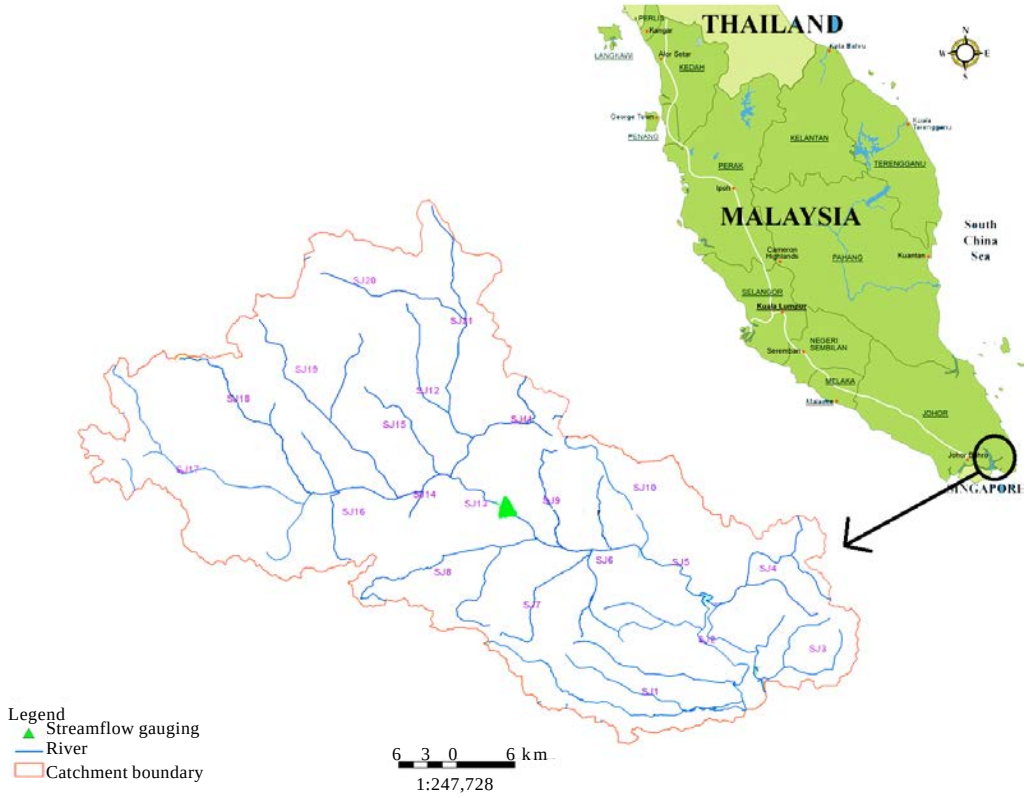

Fig. 1: Map of Johor River and the Rantau Panjang Gauging station

Modeling the peak flow, duration and volume: Gamma, Weibull, Exponential, Gumbel and GEV1 were used to model the distribution of the flood variables. Empirical cumulative distribution function was determined using the equation:

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

So the theoretical Cumulative Distribution Function (CDF) is displayed as a continuous curve. The empirical CDF is denoted by:

$$F_n(x) = \frac{1}{n}[\text{No. of observations} \le x] \tag{2}$$

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:

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

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

$$f(x) = p(X = x) \quad (4)$$

The empirical PDF is depicted in a histogram consisting of equal-width vertical bars (bins) where each represents the number of sample data values (falling into the corresponding interval), divided by the total number of data points. Depending on the number of intervals, the theoretical PDF, displayed as a continuous curve, should be scaled accordingly.

The PDF and CDF for the five models are given as follows:

Gamma distribution: The Gamma distribution with continuous shape parameter (α), continuous scale parameter (β) and continuous location parameter (γ) have PDF and CDF given by:

$$f(x) = \frac{(x-\gamma)^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} \exp(-(x-\gamma)/\beta) \quad (5)$$

$$F(x) = \frac{\Gamma_{(x-\gamma)/\beta}(\alpha)}{\Gamma(\alpha)} \quad (6)$$

Where:

$$\gamma \leq x < +\infty$$

Weibull distribution: The Weibull distribution with continuous shape parameter (α), continuous scale parameter (β) and continuous location parameter (γ) have PDF and CDF given by:

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x-\gamma}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right) \quad (7)$$

$$F(x) = 1 - \exp\left(-\left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right) \quad (8)$$

Where:

$$\gamma \leq x < +\infty$$

Exponential distribution: The exponential distribution with continuous inverse scale parameter (λ) and continuous location parameter (γ) have PDF and CDF given by:

$$f(x) = \lambda\exp(-\lambda(x-\gamma)) \quad (9)$$

$$F(x) = 1-\exp(-\lambda(x-\gamma)) \quad (10)$$

Where:

$$\gamma \leq x < +\infty$$

Gumbel distribution: The Gumbel distribution with continuous scale parameter (σ) and continuous location parameter (μ) have PDF and CDF given by:

$$f(x)=\frac{1}{\sigma}\exp(-z-\exp(-z)) \tag{11}$$

$$F(x) = \exp(-\exp(-z)) \tag{12}$$

Where:

$$z \equiv \frac{x-\mu}{\sigma}$$

Generalized extreme value (GEV): The general extreme value with continuous shape parameter (κ), continuous scale parameter (σ) and continuous location parameter (μ) have PDF and CDF given by:

$$f(x)=\begin{cases}\frac{1}{\sigma}\exp(-(1+kz)^{-1/k})(1+kz)^{-1-1/k} & k\neq 0\\ \frac{1}{\sigma}\exp(-z-\exp(-z)) & k=0\end{cases} \tag{13}$$

$$F(x)=\begin{cases}\exp(-(1+kz)^{-1/k} & k\neq 0\\ \exp(-\exp(-z)) & k=0\end{cases} \tag{14}$$

Where:

$$z \equiv \frac{x-\mu}{\sigma}$$

$$1+k\frac{(x-\mu)}{\sigma}\rangle 0 \quad \text{for} \quad k\neq 0$$
$$-\infty\langle x\langle+\infty \quad \text{for} \quad k=0$$

Goodness-of-fit tests: Goodness-of-fit (GOF) test measures the compatibility of a random sample with a theoretical probability distribution function. In other words, these tests show how well the selected distribution fits to the data. Two goodness-of-fit tests, namely Kolmogorov-Smirnov (K-S) and chi-squared (C-S) were conducted at 5% level of significance. Note that X denotes the random variable; and n, the sample size. The mathematical explanation of the two goodness-of-fit tests is as follows:

Kolmogorov-Smirnov (K-S) test: This nonparametric test is used in statistics to test the equality of continuous, which is to compare a sample with a reference probability distribution. It quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution or between the empirical distribution functions of two samples. For a random variable X and sample (x_1, x_2, x_3,..., x_n) the empirical CDF of X ($F_x(x)$) is given by:

$$F(x)=\frac{1}{n}\sum_{i=1}^{n}I(x_i \le x) \quad (15)$$

where, I (condition) = 1 if true and 0 otherwise.

Given two cumulative probability functions F_x and F_y, the Kolmogorov-Smirnov statistic test (D_+ and D_-) are given by:

$$D_+ = \max(F_x(x)-F_y(x))$$

$$D_- = \max(F_y(x)-F_x(x))$$

Chi-squared (χ^2) test: This test is a statistical hypothesis test to simply compare how well the theoretical distribution fits the empirical distribution PDF. The chi-squared test statistic is given by:

$$\chi^2 = \sum\frac{(O_i - E_j)^2}{E_i} \quad (16)$$

Where:
O_i = The observed frequency for bin i
E_i = The expected frequency for bin i
K = The number of classes

E_i is given by:

$$E_i = f(X_2)-F(X_1) \quad (17)$$

and X_1 and X_2 are the lower and upper limits for bin i, respectively.

The probability difference method: This method is used to identify how well the theoretical distribution fits into the observed data and to compare several fitted distribution for goodness-of-fit test. In other words, the probability difference graph plots the difference between the empirical CDF and the theoretical CDF. The theoretical equation is as follow:

$$\text{Difference } (x) = F_n(x)-F(x) \quad (18)$$

RESULTS AND DISCUSSION

Table 1 presents the descriptive statistic of the flood variables. The statistic amount of peak flow, duration and the volume for 34 years are calculated as the basic information for fitting the distribution. The fitting result parameters for various distributions of flood variables are as shown in Table 2. The result of Table 2 shows continuous shape parameter (α), continuous scale parameter (β) and continuous location parameter (γ) are shown in different distributions for each variables.

The search for the best fitting distribution for flood variables amount has been the main interest in several studies. Thus, various forms of distributions were tested in order to find the best

Table 1: The statistics of the flood variables

Parameters	Min.	Max.	Mean	SD
Peak flow, Q_p (m^3 sec^{-1})	76.89	724.734	248.22	163.86
Duration, Q_d (h)	144	600	349.41	125.50
Volume, Q_v (mm)	19.95	231.16	104.80	49.49

The number of storm is 34

Table 2: Fitting result parameters for various distributions of flood variables

	Flood variables		
Distributions	Peak flow (P)	Duration (D)	Volume (v)
Gamma			
α	0.88255	6.0071	1.6746
β	177.13	1.9005E+5	874.65
γ	76.899	1.1621E+5	209.92
Weibull			
α	0.96654	1.934	1.3846
β	157.85	9.3755E+5	1566.1
γ	76.899	4.2546E+5	237.27
Exponential			
λ	0.00584	1.3523E-6	7.1962E-4
γ	76.899	5.1840E+5	285.03
Gumbel			
μ	127.77	3.5227E+5	777.81
σ	174.48	1.0545E+6	1225.7
Gen. extreme value (GEV)			
κ	0.21558	-0.20041	-0.04092
σ	98.515	4.4081E+5	857.47
μ	164.97	1.0777E+6	1213.2

Generalized extreme value distribution is the best fitted to peak flow and Weibull distribution is the best fitted to Volume and duration

Table 3: Goodness-of-fit test ranking for various distributions of flood variables

	Goodness-of fit tests					
	Kolmogorov-Smirnov			Chi-squared		
Distributions	Peak flow	Duration	Volume	Peak flow	Duration	Volume
Gamma	4	3	2	2	3	2
Weibull	3	2	1	4	1	1
Exponential	1	5	5	3	5	4
Gumbel	5	4	4	5	4	3
Gen. extreme value	2	1	3	1	2	5

Ranking is in the order of 1, 2, 3, 4 and 5. 1 is the best ranking and 5 the worst ranking

fitting distribution. Earlier finding based on shorter observation found that GEV ranked the best to describe peak flow distribution (Suhaila and Jemain, 2008; Danazumi and Shamsudin, 2011). The present study is based on a longer data record but the results also confirmed that GEV is still the best for describing peak flow distribution (Table 3). The peak flow was best fitted by GEV distribution whereas flood volume and flood duration fitted better by Weibull distribution in compare between Kolmogorov-Smirnov and chi-squared.

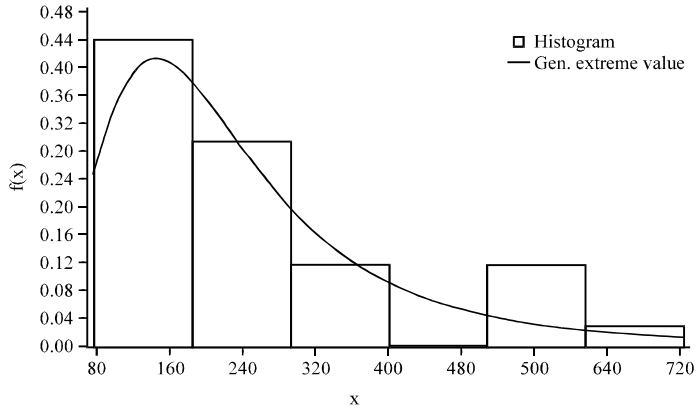

Fig. 2: Probability density function for generalized extreme value distribution fitted to the peak flow

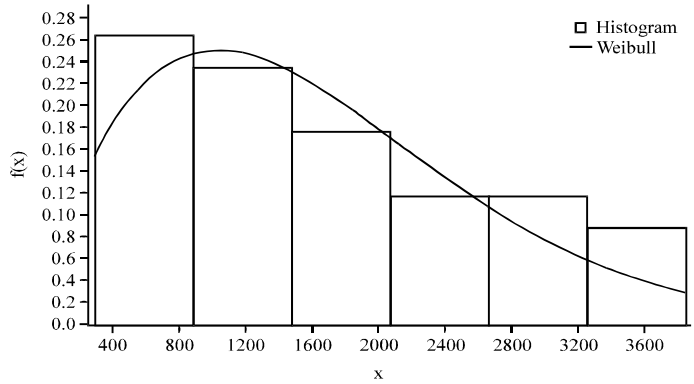

Fig. 3: Probability density function for Weibull distribution fitted to the volume of peak flow

The occurrence of the wet and dry days sequence researched using daily rainfall data has also been the main focus of past researches in Malaysia. Probability models that fitted with the dry and wet spell distribution were the mixture of log series distribution; the mixture of log series Poisson distribution and the mixture of log series geometric distribution. The result generated through chi-squared goodness-of-fit test showed that the mixture of log series geometric distribution and mixture of log series Poisson distribution had better fit (Deni and Jemain, 2009). On the other hand, seven theoretical distributions were fitted to mixture of wet and dry days in another research. The results showed that a compound geometric distribution and the truncated negative binomial distribution would be sufficient for most stations (Deni *et al*., 2008). Similarly, Fig. 2 presents the PDF for the GEV distribution fitted to the peak flow and Fig. 3 presents the PDF of Weibull distribution fitted to the volume of stormflow.

Meanwhile, tests were done on log normal, skew normal and mixed log-normal distributions and the results indicated that the mixed log-normal distribution performed better than the rest of the models (Suhaila and Jemain, 2007). However, researches on rainfall characteristics modeling for Peninsular Malaysia using hourly rainfall data are very rare. A particular research was done for 12 stations in Wilayah Persekutuan, Malaysia for four candidate distributions, namely Exponential, Gamma, Weibull and Mixed Exponential. The results concluded that the Mixed

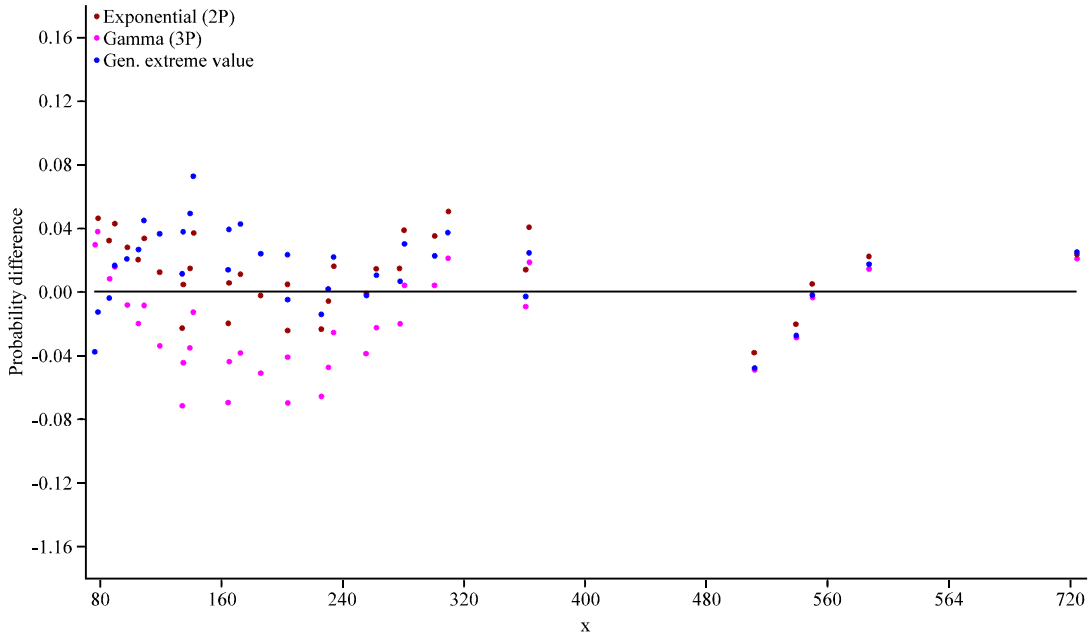

Fig. 4: Probability difference for Gamma, generalized extreme value and Weibull distributions of peak flow

Exponential distribution fitted better for hourly rainfall volume in that area (Fadhilah *et al*., 2007). Figure 4 shows the probability difference plot between the empirical CDF and the theoretical CDF.

CONCLUSION

Hourly rainfall and flow data were used to analyze the statistical distribution for the peak flow, duration and volume annual flood for Johor River at Rantau Panjang gauging station. Five probability distributions, namely Gamma, Weibull, Exponential, Gumbel and GEV were tested. Based on the chi-squared test, the GEV1 distribution was found to be the most suitable for modeling the peak flow. On the other hand, Weibull is the most suitable distribution for modeling the flood duration and volume. Goodness-of-fit tests at 5% level of significance indicated that all the models can be used to model the distribution of peak flow, duration and volume. For further study it is recommended to use other distributions such as Generalized Pareto, Pearson, Exponential and Beta. In addition, different Goodness-of-fit test such as Anderson-Darling (A-D) can be attempted.

ACKNOWLEDGMENTS

The authors are grateful to the Department of Irrigation and Drainage (DID), Malaysia for the supply of rainfall and hydrograph data. The support from Institute of Environmental and Water Resources Management (IPASA) and Research Management Centre of Universiti Teknologi Malaysia is also highly appreciated.

REFERENCES

Bonnin, G.M., D. Martin, B. Lin, T. Parzyok, M. Yekta and D. Riley, 2006. Precipitation-frequency atlas of the United States, Volume 1, Version 4.0: Semiarid Southwest (Arizona, Southeast California, Nevada, New Mexico, Utah). NOAA Atlas 14. http://www.nws.noaa.gov/oh/hdsc/PF_documents/Atlas14_Volume1.pdf

Buishand, T.A., 1978. Some remarks on the use of daily rainfall models. J. Hydrol., 36: 295-308.

Burgueno, A., M.D. Martinez, X. Lana and C. Serra, 2005. Statistical distributions of the daily rainfall regime in Catalonia (Northeastern Spain) for the years 1950-2000. Int. J. Climatol., 25: 1381-1403.

Danazumi, S. and S. Shamsudin, 2011. Modeling the distribution of inter-event dry spell for Peninsular Malaysia. J. Applied Sci. Res., 7: 333-339.

Deni, S.M., A.A. Jemain and K. Ibrahim, 2008. The spatial distribution of wet and dry spells over Peninsular Malaysia. Theor. Applied Climatol., 94: 163-174.

Deni, S.M. and A.A. Jemain, 2009. Fitting the distribution of dry and wet spells with alternative probability models. Meteorol. Atmosphere Phys., 104: 13-27.

Durrans, S.R., 1998. Total Probability Methods for Problems in Flood Frequency Estimation. In: Statistical and Bayesian Methods in Hydrological Science, Parent, E., P. Hubert, B. Bobee and J. Miquel (Eds.). UNESCO, Paris, France, pp: 299-326.

Fadhilah, Y., M.D. Zalina, V.T.V. Nguyen, S. Suhaila and Y. Zulkifli, 2007. Fitting the best-fit distribution for the hourly rainfall amount in the Wilayah Persekutuan. J. Teknol., 46: 49-58.

Geng, S., F.W.T.P. de Vries and I. Supit, 1986. A simple method for generating daily rainfall data. Agric. For. Meteorol., 36: 363-376.

Hershfield, D.M., 1962. Rainfall frequency atlas of the United States for Durations from 30 minutes to 24 hours and return periods from 1 to 100 years. U.S. Weather Bureau Technical Paper No. 40, Washington, DC., USA. http://www.nws.noaa.gov/oh/hdsc/PF_documents/TechnicalPaper_No40.pdf

Renard, B. and M. Lang, 2007. Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. Adv. Water Resour., 30: 897-912.

Salarpour, M., N.A. Rahman and Z. Yusop, 2011. Simulation of flood extent mapping by InfoWorks RS-case study for tropical catchment. J. Software Eng., 5: 127-135.

Samaniego, L., A. Bardossy and R. Kumar, 2010. Streamflow prediction in ungauged catchments using copula-based dissimilarity measures. Water Resour. Res., Vol. 46. 10.1029/2008WR007695.

Suhaila, J. and A.A. Jemain, 2007. Fitting daily rainfall amount in Peninsular Malaysia using several types of exponential distributions. J. Applied Sci., 3: 1027-1036.

Suhaila, J. and A.A. Jemain, 2008. Fitting the statistical distribution for daily rainfall in Peninsular Malaysia based on AIC criterion. J. Applied Sci. Res., 4: 1846-1857.

Thom, H.C.S., 1951. A frequency distribution for precipitation. Bull. Am. Meteorol. Soc., 32: 397-397.

Yue, S., 1999. Applying bivariate normal distribution to flood frequency analysis. Water Int., 24: 248-252.

Yue, S., T.B.M.J. Ouarda, B. Bobee, P. Legendre and P. Bruneau, 1999. The Gumbel mixed model for flood frequency analysis. J. Hydrol., 226: 88-100.

Yue, S., 2000. The bivariate lognormal distribution to model a multivariate flood episode. Hydrol. Processes, 14: 2575-2588.

Yue, S., 2001. A bivariate γ distribution for use in multivariate flood frequency analysis. Hydrol. Processes, 15: 1033-1045.

Yue, S. and P. Rasmussen, 2002. Bivariate frequency analysis: discussion of some useful concepts in hydrological application. Hydrol. Process, 16: 2881-2898.

Yusop, Z., I. Douglas and A.R. Nik, 2006. Export of dissolved and undissolved nutrients from forested catchments in Peninsular Malaysia. For. Ecol. Manage., 224: 26-44.

Zalina, M.D., M.N.M. Desa, V.V. Nguyen and A.H.M. Kassim, 2002. Selecting a probability distribution for extreme rainfall series in Malaysia. Water Sci. Technol., 45: 63-68.

Zhang, L. and V.P. Singh, 2006. Bivariate flood frequency analysis using the copula method. J. Hydrol. Eng., 11: 150-164.