Sains Malaysiana 48(7)(2019): 1357–1366
http://dx.doi.org/10.17576/jsm-2019-4807-05

L-Moment-Based Frequency Analysis of High-Flow at Sungai Langat, Kajang, Selangor, Malaysia

(Analisis Kekerapan berdasarkan L-momen Aliran Tinggi di Sungai Langat, Kajang, Selangor, Malaysia)

FIRDAUS MOHAMAD HAMZAH*, SITI HAWA MOHD YUSOFF & OTHMAN JAAFAR

ABSTRACT

Annual maximum daily streamflow data were used to examine flood frequency for Sungai Langat in Kajang, Selangor, Malaysia. The objectives of this study were to identify the best fit probability distribution to the streamflow data and estimate the return period of the extreme flood events. The L-moment method was implemented to estimate the parameter of probability, by using distributions namely Gamma, LN3, GEV, PE3, GLO and Kappa. It was found that Kappa distribution was the best fitting distribution to the data after being tested using the goodness-of-fit tests. The Kappa distribution gave the most appropriate to the annual maximum series data of Sungai Langat, Kajang, Selangor. The return values were calculated using Kappa distribution model. The return period of 2 years gave the return value of 49.09 m^3/s, while return period of 100 years gave the return value of 390.54 m^3/s.

Keywords: Goodness-of-fit test; L-moment; probability distribution; return period; streamflow

ABSTRAK

Kajian terhadap frekuensi banjir dilakukan bagi kawasan Sungai Langat, Kajang, Selangor, Malaysia dengan menggunakan data maksimum tahunan yang diperoleh daripada data aliran sungai harian. Objektif kajian ini adalah untuk mengenal pasti taburan kebarangkalian terbaik bagi aliran sungai dan seterusnya menganggarkan tempoh pulangan banjir. Kajian ini mengaplikasikan kaedah L-momen untuk menganggarkan parameter kebarangkalian iaitu Gamma, LN3, GEV, PE3, GLO dan Kappa. Didapati taburan Kappa merupakan taburan terbaik setelah diuji menggunakan ujian kebagusan penyuaian. Oleh itu, taburan Kappa dipilih sebagai taburan yang paling sesuai bagi data siri maksimum tahunan untuk Sungai Langat, Kajang, Selangor. Seterusnya, nilai tempoh pulangan dapat dihitung dengan menggunakan parameter Kappa. Tempoh pulangan 2 tahun memberi nilai 49.09 m^3/s, manakala tempoh pulangan 100 tahun memberi nilai 390.54 m^3/s.

Kata kunci: Aliran sungai; L-momen; taburan kebarangkalian; tempoh pulangan; ujian kebagusan penyuaian

INTRODUCTION

Most extreme events in hydrology can cause danger to human beings, material damages, and infectious diseases. In predicting extreme event, extreme value theory can be used to characterise the tail of the probability distribution. By definition, an extreme event is unusual and rarely happens. Extreme value theory is frequently used in environmental studies (Katz 2002). The particular interest in this study is on extreme flood events.

Flood event is one of the natural disasters that occur everywhere around the world. Extreme flood is one of the natural disasters that have become a normal occurrence due to the condition of our environment today. Previous town planning mostly never included any extreme flood prevention in the calculation. Rapid house development is inevitable these days but bad planning further disrupts the water flow. These phenomena among others are the main reason flood has become a common occurrence and remains among the most destructive natural hazards in many countries around the world. Protecting population from the rare flood by forecasting and signalling early warning is one of the mechanisms that government should employ (Petersen 2009). Borujeni et al. (2009) explained that one of the prevention systems created by human to control flood is the hydrology system. One of the functions of the system is to determine the extreme value and estimation of the flood event using the data gathered by the hydrology system that later on can be calculated using statistical methods depending on the structural design and size of the watershed.

Malaysia is located near the equator that often experience flood events, especially during monsoon season and due to convectional rainfall. Several flood events in Malaysia were reported by Gasim et al. (2010). In December 2006, Segamat was hit by very big flood due to abnormal rainfall while Rahman (2007) reported that Kuala Lumpur was hit by the worst flash flood in June 2007.

Fréchet (1927) was the first to obtain an asymptotic distribution of the largest value. Fisher and Tippett (1928) also came out with extreme value theory. Gumbel (1958) applied extreme value distributions to analyse flood flow data. Thereafter, several researchers have provided useful applications of extreme value distributions to flood data from the different regions of the world. The same method was applied by Bhunya et al. (2013) in Scotland, Zaman et al. (2012) in Maryland, U.S.A.

Extreme events can be analysed by using flood frequency analysis. Flood frequency analysis is used to analyse past hydrological record to make prediction of future occurrences. Usually, the observed frequency of occurrence of an event with an associated risk of failure is represented by its return period. Thus, the return period at the site of interest or across homogeneous region can be computed by using suitable probability distribution. The frequency analysis methods do not predict the future with certainty but they offer good models for explaining and making efficient use of the extreme events that had occurred in the past (Khaliq et al. 2006). The method used to predict extreme event and estimate their return period is annual maxima series (AMS) (Adamowski 2000; Hosking & Wallis 1987; Madsen et al. 1997). It contains only the largest magnitude value that occurs every single year.

The L-moment method was used to estimate the parameter of distributions and this method was developed by Hosking and Wallis (1997). This method has been applied successfully in flood frequency analysis in many countries such as in Malaysia (Lim & Lye 2003), China (Hassan & Ping 2012), India (Sai Krishna & Veerendra 2015), Norway (Hailegeorgis & Alfredsen 2017), Iran (Mosaffaie 2015), Poland (Rutkowska et al. 2017), Turkey (Seckin et al. 2011), and Iran (Malekinezhad et al. 2011).

This study discussed the L-moment approach in flood frequency analysis of annual maximum streamflow of Sungai Langat, Kajang. The objectives of this study were to identify the best fit probability distribution to the streamflow data and estimate the return period of the extreme flood events. The distributions considered in this study were Gamma, LN3, GEV, PE3, GLO and Kappa.

STUDY AREA AND DATASETS

Sungai Langat which is situated in Selangor, Peninsular Malaysia has a total catchment area of approximately 1815 km^2. It is one of the most important basins that supply water to two-third of the state of Selangor. However, Sungai Langat has several tributaries with the principle ones being Sungai Semenyih and Sungai Lui. There are two reservoirs, the Langat Reservoir and the Semenyih Reservoir, respectively. Sungai Langat generally flows from the Titiwangsa Range at the Northeast of Hulu Langat District and drains into the Straits of Malacca. Along Sungai Langat, there are four gauging stations, namely Lui (station no. 3118445), Kajang (station no. 2917401), Semenyih (station no. 2918401), and Dengkil (station no. 2816441). The hydrological characteristics of Sungai Langat are greatly influenced by two heavy rainy seasons during South-West (May-September) and North-East (November-March) monsoons. Meanwhile, convectional rain is common during inter-monsoon period. Sungai Langat Basin receives between 1900 mm and 3000 mm of rainfall per year.

The study area in Figure 1 focuses on Sungai Langat in Kajang (02° 59' 40" N, 101° 47' 10" E). The size of the catchment above the gauging station is 389.4 square kilometre. It is subjected to tremendous urban development pressure with high rate of population growth and intensity pressure for urban land development. One of the places located near to Sungai Langat is Kajang. The total population in Kajang itself is 311785. Kajang is one of the towns which are good for investment properties and location for new home that is located near to Sungai Langat. Unfortunately, it is also a flood prone area. In 2014, Kajang faced flood three times on 29th September (Bernama 2014), 12th November (Bernama 2014) and 20th December (Brown et al. 2014). Streamflow data used for

FIGURE 1. The Google maps (2015) showed the area of study which is Sungai Langat Basin in Kajang

this study were extracted from Kajang Gauging Station. This station provides daily discharge (m^3/s) and has been maintained by the Department of Irrigation and Drainage (DID), Ministry of Natural Resources and Environment, Malaysia since 1978. Annual maximum discharge data observed from 1978 to 2016 were used in this analysis.

METHODS

PARAMETER ESTIMATION TECHNIQUES

In order to perform flood frequency analysis, the L-moment method was used by estimating the annual maximum streamflow data. It was obtained by taking the largest value in each year of interest (Adamowski 2000; Esteves 2013). Beside the L-moment, there are several other methods of performing parameter estimation such as method of moment and maximum likelihood estimation. As a general framework for extreme value modelling, maximum likelihood estimation method has many advantages because it can be constructed for complex modelling situation, enabling for non-stationarity, covariate effects and regression modelling (Coles & Dixon 1999). However, previous study showed that the method of maximum likelihood estimation is unstable and can give unrealistic estimates for the shape parameter for the small sample size (Hosking & Wallis 1997; Martins & Stedinger 2000).

The shape of a probability distribution has been described by using the moment method. In this study, the L-moment method was used to estimate the distribution. L-moment is a modification from probability weighted moments. Traditionally, the probability weighted moment used weight of cumulative distribution function. However, they are difficult to interpret directly as measures of the scale and shape of probability distribution. To overcome this problem, the L-moment technique introduced by Hosking and Wallis (1997) was used. The L-moment carried in certain linear combinations of data arranged in ascending order. The L-moment technique is more accurate for small sample size and more reliable as it is less sensitive to outliers. In term of probability weighted moment, the L-moments are given by (Hosking & Wallis 1997; Millington et al. 2011). The first four L-moment formulas are as follows:

$$\lambda_1 = \beta_0 \tag{1(a)}$$

$$\lambda_2 = 2\beta_1 = \beta_0 \tag{1(b)}$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \tag{1(c)}$$

$$\lambda_4 = 20\beta_3 - 12\beta_1 - \beta_0 \tag{1(d)}$$

where λ_1 is the mean of the distribution; λ_2 is a measure of dispersion; λ_3 is a measure of skewness; and λ_4 is a measure of kurtosis. The four L-moments are derived from the following probability weighted moments:

$$\beta_0 = \frac{1}{n}\sum_{i=1}^{n} Q_i \tag{2(a)}$$

$$\beta_1 = \frac{1}{n}\sum_{i=2}^{n} \frac{(i-1)}{(n-1)} Q_i \tag{2(b)}$$

$$\beta_2 = \frac{1}{n}\sum_{i=3}^{n} \frac{(i-1)(i-2)}{(n-1)(n-2)} Q_i \tag{2(c)}$$

$$\beta_3 = \frac{1}{n}\sum_{i=4}^{n} \frac{(i-1)(i-2)(i-3)}{(n-1)(n-2)(n-3)} Q_i \tag{2(d)}$$

where n is the sample size; Q is the data value; and i is the rank of the value in ascending order. Other useful ratios are:

$$\text{Coefficient of L- variation, } \tau_2 = \frac{\lambda_2}{\lambda_1} \tag{3(a)}$$

$$\text{L- skewness, } \tau_3 = \frac{\lambda_3}{\lambda_2} \tag{3(b)}$$

$$\text{L-kurtosis, } \tau_4 = \frac{\lambda_4}{\lambda_2} \tag{3(c)}$$

where L-moment ratios satisfy $|\tau_r| < 1$ for all $r \geq 3$.

STATISTICAL DISTRIBUTION

The selection of probability distributions is based on probability distributions with the shape parameter. This is because the shape parameter can be represented as the skewness parameter. There are six types of probability distributions used in this analysis, which are Gamma distribution (Yue 2001), Three-parameter Lognormal distribution (LN3) by Cohen and Whitten (1980), Generalized Extreme Value distribution (GEV) by Provost et al. (2018), and Pearson type 3 distribution (PE3) by Wu et al. (2012), Generalized Logistic Distribution (GLO) by (Kjeldsen & Jones 2004) and Four-parameter Kappa distribution (Kjeldsen et al. 2017; Shabri & Jemain 2010). Table 1 shows the probability density functions for each distribution.

GOODNESS-OF-FIT TESTS

Goodness-of-fit tests can be used to decide whether two samples belong to the same population or the probability distribution of data belongs to a specific theoretical distribution. The goodness-of-fit tests based on empirical density function (EDF) measure the different distance between the empirical cumulative density function and theoretical cumulative density function (D'Agostino 1986; Moralles et al. 2013).

In goodness-of-fit tests, the test statistics are calculated and then the p-value can be obtained. If the p-value is

TABLE 1. Probability density function for Gamma, LN3, GEV, PE3, GLO and Kappa distributions

No.	Distribution	Probability density function
1	Gamma	$f(x)=\frac{\beta^{-\alpha}x^{\alpha-1}}{\Gamma(\alpha)}\exp\left(\frac{-x}{\beta}\right), \quad \alpha>0,\ \beta>0,\ x>0$ where α is the location parameter and β is the scale parameter
2	Three-parameter Lognormal (LN3)	$f(x)=\frac{1}{(x-\gamma)\sigma\sqrt{2\pi}}\exp\left[-\frac{\left[\ln(x-\gamma)-\mu\right]^2}{2\sigma^2}\right]$ where $\sigma^2>0$, $\gamma<x<\infty$ where γ, μ and σ are respectively, the location, scale and shape parameters
3	Generalized extreme value (GEV)	$f(x)=\alpha^{-1}\exp\left[-(1-\kappa)y-\exp(-y)\right]$ where $y=-\kappa^{-1}\log\left[1-\frac{\kappa(x-\xi)}{\alpha}\right]$, when $\kappa\neq 0$ where ξ, α and κ are respectively, the location, scale and shape parameters
4	Pearson Type 3 (PE3)	$f(x)=\frac{1}{\alpha\Gamma(\gamma+1)}\left(\frac{x-\mu}{\alpha}\right)^{\gamma}\exp\left[-\left(\frac{x-\mu}{\alpha}\right)\right]$ where μ, σ and γ are respectively, the location, scale and shape parameters
5	Generalized Logistic (GLO)	$f(x)=\alpha^{-1}\frac{\exp(-(1-\kappa)y)}{(1+\exp(-y))^2}$ where $y=\begin{cases}-\kappa^{-1}\ln\left(1-\frac{\kappa(x-\xi)}{\alpha}\right) & k\neq 0\\ \frac{x-\xi}{\alpha} & k=0\end{cases}$ where ξ, α and κ are respectively, the location, scale and shape parameters
6	Four-parameter Kappa (Kappa)	$f(x)=\frac{1}{\alpha}\left[1-\frac{k}{\alpha}(x-\xi)\right]^{1/k-1}\left\{1-h\left[1-\frac{k}{\alpha}(x-\xi)\right]^{1/k}\right\}^{1/h-1}$ where ξ is the location parameter, α is the scale parameter, and κ and h are the shape parameters

greater than the significance level alpha, α = 0.05, the null hypothesis, H_0 will be accepted which means that the data follow a specified distribution. Meanwhile, if the p-value is less than the significance level alpha, the alternative hypothesis, H_1, will be accepted which means that the data do not follow a specified distribution. The Anderson-Darling test (AD) in (4) and Kolmogorov-Smirnov test (KS) in (5) were used for testing the goodness-of-fit.

ANDERSON-DARLING TEST

The Anderson-Darling test belongs to the class of quadratic EDF statistics to give more weight to the tails of the distribution. This characteristic is interesting, since the maximum streamflow reflects the distribution tails well instead of the central values of distribution. This test is the most powerful EDF test according to Razali and Wah (2011). Anderson and Darling (1954) defined the statistics for this test as,

$$A_n^2=-n-\sum_{i=1}^{n}\frac{2i-1}{n}\left[\log\left(\hat{F}(x_{(i)})\right)+\log\left(1-\hat{F}(x_{(n+1-i)})\right)\right] \quad (4)$$

where $\hat{F}(x_{(i)})$ is the cumulative distribution function of the theoretical distribution and $\hat{F}(x_{(n+1-i)})$ is the empirical distribution function.

KOLMOGOROV-SMIRNOV TEST

The Kolmogorov-Smirnov test is based on the largest difference between the empirical CDF and theoretical distribution. It belongs to the supremum class of EDF statistics. Thus, under the null hypothesis, it is expected that the difference between the empirical distribution function of the theoretical distribution, $F_n(x)$ and the cumulative distribution function, $F(x)$ is small. Razali and Wah (2011) defined the test statistic as the maximum deviation of the largest difference in absolute value,

$$D=max_x\left|F(x)-F_n(x)\right| \quad (5)$$

If $D\leq D_{1-\alpha}$, it means that the sample has the same distribution with the tested theoretical distribution. The $D_{1-\alpha}$ are critical values for the Kolmogorov-Smirnov goodness-of-fit at a level of significant α.

RETURN PERIOD

The return period can be defined as the average number of trials usually in a year to the first occurrence of an event of magnitude greater than a predefined critical event (Benjamin & Cornell 1970). Many current flood management policies and designs are based on an estimate of the 100-year flood, an event that has a 1% chance of occurring in a given year. However, the existing methods to estimate the 100-year flood assumed flood records are stationary even though there are multiple non-stationary factors such as climate change and urbanisation that can influence measured hydrological data (Gilroy & McCuen 2012). The return period was expressed as (Mélice & Reason 2007):

$$T_x = \frac{1}{P_x} \tag{6(a)}$$

$$F_x = 1 - \frac{1}{T_x} \tag{6(b)}$$

where T_x corresponds to years of return period of such a design high flow and P_x is an exceedance probability, where $P_x = P(X \geq x_T)$ of occurrence of the event $\geq x_T$. While F_x is the cumulative probability distribution function.

RESULTS AND DISCUSSION

DESCRIPTIVE STATISTICS

The maximum streamflow of time series plots of Sungai Langat, Kajang is shown in Figure 2. The y-axis represents the streamflow data in cubic meter per second (m^3/s) and x-axis represent years from 1978 to 2016. In Table 2, the average of maximum streamflow was 74.29 m^3/s, the median of maximum streamflow was 48.50 m^3/s and the standard deviation was 67.77 m^3/s. Since the value of mean was far from the median, it seems that the extreme values gave an effect to the mean. While for standard deviation, it gave a large value, giving an assumption that the data points are far to the mean value. It can be observed that the highest value of extreme flow was 360.8 m^3/s recorded in 2009 and the second highest streamflow was recorded in 2004 with value 212.0 m^3/s.

A flow duration curve in Figure 3 represents the relationship between the magnitude of flow in m^3/s and frequency of daily streamflow for Sungai Langat basin. The flow duration curve graphically displays the percentage of time that daily streamflow is equal or exceeded for a particular river basin (Li et al. 2010). The daily average exceeded 57.4 m^3/s at 1 % of the days are considered extreme flood events. From the total of 14245

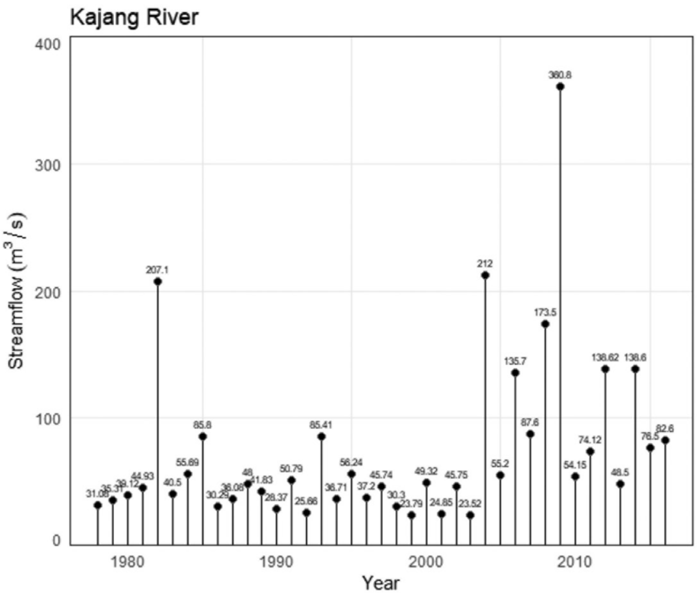

FIGURE 2. Time series plot of maximum streamflow of Sungai Langat, Kajang from 1978 to 2016

TABLE 2. The descriptive statistics of annual maximum streamflow of Sungai Langat, Kajang

Descriptive Statistics	Average	Median	Standard deviation	Maximum	Minimum
Annual Maximum Streamflow (m^3/s)	74.29	48.50	67.77	360.8	23.52

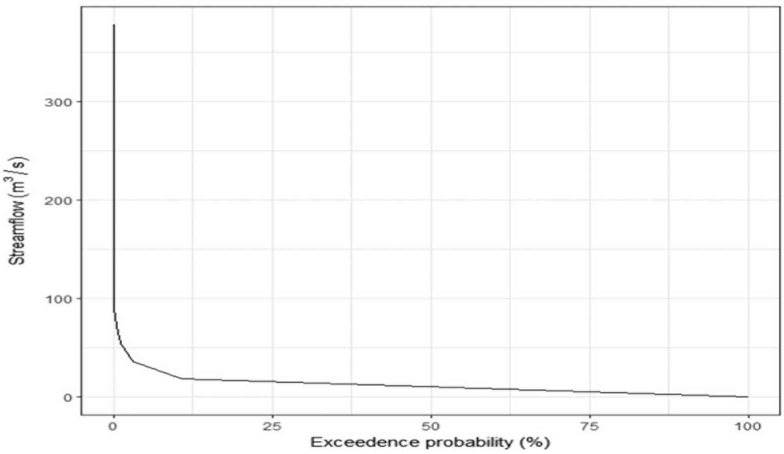

FIGURE 3. Flow duration curve for Sungai Langat, Kajang

observations, 142 days were considered as extreme flows which exceeded 57.4 m^3/s. The graph has a steep slope. According to Searcy (1959) and Shao et al. (2009), a curve with a steep slope denotes a highly variable stream which flow is largely from a direct runoff. The more urbanised the catchment, the higher flood peak and the total runoff (Hashim 2003).

L-MOMENT PARAMETER

Sample L-moment was computed for Sungai Kajang maximum streamflow using (1), (2), and (3). The result is presented at Table 3.

From Table 3, the value of L-location describes the central value of the maximum streamflow data with the mean is 74.289. The L-scale, 30.689 describes the spread of the distribution. The larger the scale parameter the more spread out the distribution. The coefficient of variation shows that the ratio between the mean and standard deviation is 0.413. As a rule of thumb, if the coefficient of variation is more or equal to 1 indicates relatively high variation, while a coefficient of variation less than 1 can be considered a low variation. Since the value is 0.413 below than 1, it is can be considered to be low variance. The parameter L-skewness tells that the distribution is skewed to the right based on the positive value of 0.499. It indicates that the tail on the right side is longer than the left side. Meanwhile, the L-kurtosis 0.303 indicates measures of the peak of the distribution.

Table 4 shows the parameter of distribution that can be estimated. The following distributions were considered as one for 2-parameter distribution (Gamma), four for

TABLE 3. Parameters estimate using L-moment method

Descriptive statistics	Parameter
λ_1, L-location	74.289
λ_2, L-scale	30.689
τ_2, coefficient of L-variation	0.413
τ_3, L-skewness	0.499
τ_4, L-kurtosis	0.303

TABLE 4. Parameter estimation of six probability distributions for Sungai Kajang station by using L-moment method

No.	Distribution	Parameters
1	Gamma	$\alpha = 1.599, \beta = 46.463$
2	GEV	$\xi = 42.273, \alpha = 22.991, \kappa = -0.457$
3	PE3	$\mu = 74.289, \sigma = 70.571, \gamma = 3.078$
4	LN3	$\gamma = 19.489, \mu = 3.407, \sigma = 1.092$
5	GLO	$\xi = 51.992, \alpha = 19.545, \kappa = -0.499$
6	Kappa	$\xi = 14.809, \alpha = 39.415, \kappa = -0.285, h = 1.278$

3-parameter distributions (GEV, PE3, LN3 and GLO) and one for 4-parameter distribution (Kappa). The parameters of the six probability distributions were computer by L-moment method. The results are presented in Table 4. Table 4 shows the parameters of L-moment for each distribution. It is represented to the location, scale and shape parameters that have been shown in Table 2.

The extreme value theory was applied to Annual Maximum Series (AMS) to find the probability of the maximum flow from each year of record. The positively skewed distribution was selected based on the positive value of L-skewness. Hosking (1992) has shown that L-moments have good properties as measures of distribution shape and are useful for fitting distributions to data. In Figure 4, the data were fitted using Gamma distribution, Three-parameter Lognormal distribution (LN3), Generalized Extreme Value (GEV) distribution, and Pearson type 3 distributions (PE3), Generalized Logistic (GLO) distribution and Four-parameter Kappa distribution. From Figure 4, it can be seen that the shape of Kappa probability density function was much closer to the shape of histogram. The goodness-of-fit test will be used for better decision of the best fit model to the Sungai Langat, Kajang data.

GOODNESS OF FIT TESTS

Goodness-of-fit tests were used for the comparison among the probability distribution for finding the best distribution to use to fit the given data. The concept of the test was to compare the empirical distribution function (EDF) which was estimated based on the data with the cumulative distribution function (CDF) to see if there was a good agreement between them (Razali & Wah 2011). The Anderson-Darling (AD) test in (4) and Kolmogorov-Smirnov (KS) in (5) test were used for the goodness-of-fit tests in this study. The comparison of the results was given in Table 5.

If the *p*-value is greater than the significance level alpha, $\alpha = 0.05$, the null hypothesis, H_0 will be accepted which means that the data follow a specified distribution. Meanwhile, if the *p*-value is less than the significance level alpha, the alternative hypothesis, H_1, will be accepted which means that the data do not follow a specified distribution. Table 5 shows that the data appropriate for the Kappa, LN3, GEV, and PE3 distributions after being tested using the Anderson-Darling. Also, the data appropriate for the GEV, PE3, LN3, GLO and Kappa after being tested using the Kolmogorov-Smirnov tests. Meanwhile, there was enough evidence to conclude that the data did not

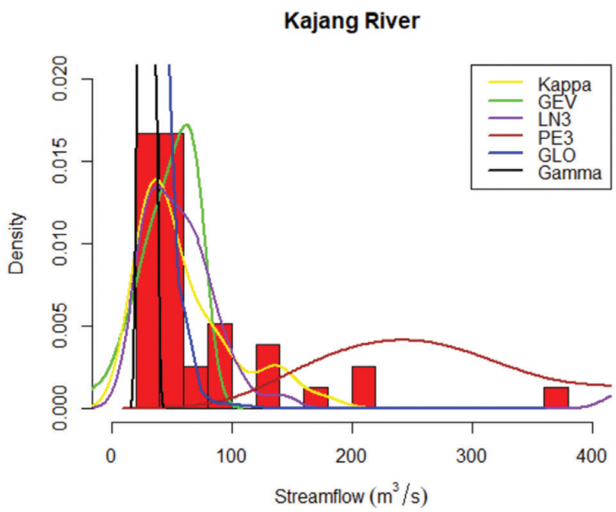

FIGURE 4. Comparison between probability density function empirical and probability density function theoretical

TABLE 5. Goodness-of-fit tests according to Anderson-Darling and Kolmogorov-Smirnov test

No.	Distribution	Anderson Darling	*p*-value	Kolmogorov Smirnov	*p*-value
1	Gamma	1.416	0.001	0.185	0.002
2	GEV	0.685	0.068	0.099	0.440
3	PE3	0.497	0.201	0.091	0.566
4	LN3	0.455	0.255	0.090	0.584
5	GLO	0.759	0.044	0.099	0.424
6	**Kappa**	**0.366**	**0.417**	**0.085**	**0.679**

follow the gamma distribution for both goodness-of-fit tests. The most appropriate distribution was chosen based on the highest p-values in each Anderson-Darling and Kolmogorov-Smirnov tests. As a rule of thumb, the p-value close to 1 indicates the most appropriate distribution to be compared to the actual data. The Kappa distribution showed the most appropriate distribution according to AD (0.366) since the value is smaller and close to the actual data. In addition, the p-value (0.417) shows the highest p-value in AD. Same goes to KS, the Kappa distribution also become the most appropriate distribution with KS (0.085) and the p-value of Kappa is 0.679.

RETURN PERIOD

After the suitable probability distribution was selected, it was important to estimate the return values for certain return periods. The return period of flood occurrence in 6(a) and 6(b) is crucial for determining the magnitude and frequency of floods and such information is valuable in accessing and mitigating the flood hazard in future (Hashim 2003). The return values were calculated using Kappa parameters. In order to calculate the return period of the Sungai Langat data, the Kappa inverse cumulative distribution function need to be derived from the probability density function of Kappa in Table 2. The following is the Kappa inverse cumulative distribution function, (7);

$$x(F) = \xi + \frac{\alpha}{k}\left\{1 - \left(\frac{1-F^h}{h}\right)^k\right\} \tag{7}$$

Table 6 shows the return values of Sungai Langat by using Kappa distribution. The return value calculated by using formula in (7). Meanwhile, Figure 5 shows the return period plot. The horizontal axis corresponds to return period in years, while the vertical axis corresponds to the flow in m³/s. From Figure 5, it can be concluded that the flow was getting an increase as the return periods increase.

TABLE 6. The estimated return values of streamflow for several return periods of Sungai Langat by using Kappa distribution

Return period (Year)	Return value (m^3/s)
2	49.09
5	97.16
10	144.18
20	201.96
30	241.58
40	272.64
50	298.57
60	321.02
70	340.94
80	358.92
90	375.36
100	390.54

From the maximum streamflow of Sungai Langat data in Figure 2, the maximum flow was 360.8 m^3/s on year 2009. By using Kappa, it can calculate the associated return period for the particular flow.

$$F(x) = \{1 - h[1 - k(x-\xi)/\alpha]^{1/k}\}^{1/h} \tag{8}$$

By substituting the Kappa parameters taken from Table 4, which are $\xi = 14.809$, $\alpha = 39.415$, $k = -0.285$ and $h = 1.278$ in the Kappa cumulative distribution function in (8), the probability was 0.0123.

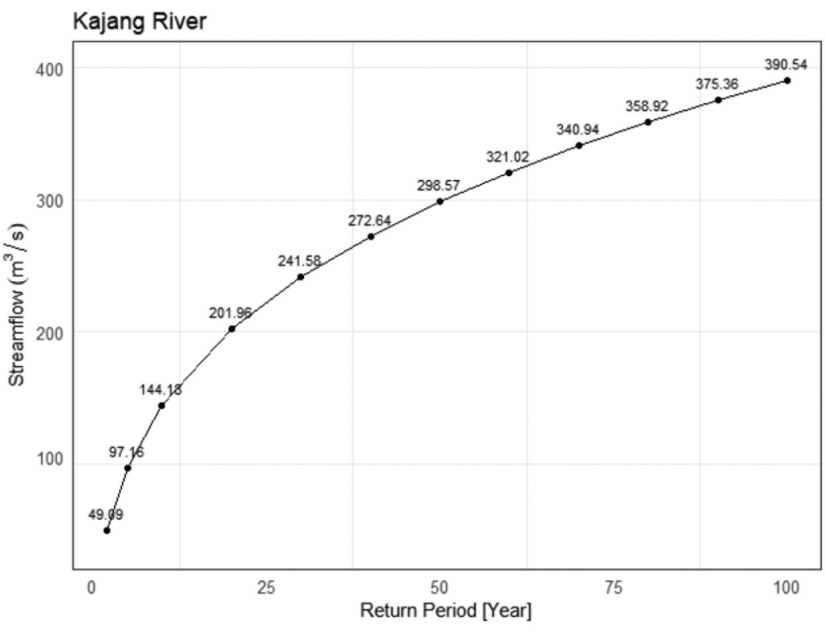

FIGURE 5. Graph of return period versus flow according to Kappa distribution

$$P(annual\ maximun > 360.8) = F(360.8) = 0.0123$$
$$\text{Estimated return period} = (0.0123)^{-1} \approx 81\ years.$$

Then, the estimated return period for the flow of 360.8 m^3/s was equal or exceeds 81 years.

Conclusion

Historical streamflow data from gauging station of Sungai Langat, Kajang from 1978 to 2016 were used in the study to analyse the extreme event. The data were taken from daily streamflow record and only the extreme value in that particular year was taken.

The maximum streamflow data were estimated by using L-moment method. It has been proven that L-moment has good properties to measure the distribution shape and is useful for fitting distributions to data. Six probability distributions were used to identify the best fit distribution such as Gamma, three-parameter lognormal (LN3), generalised extreme value (GEV) distribution, Pearson type 3 distributions (PE3), Generalized Logistic distribution (GLO) and Kappa.

It was found that the Four-parameter Kappa distribution became the most appropriate distribution to the data after being tested using the Anderson-Darling and Kolmogorov-Smirnov tests. The return values were calculated using Kappa parameters and the 100-year return period can be estimated. This is in line with the design storm water system that being used in Malaysia, where the design should be able to manage minor and major storm events. Especially during the major storm events, the average recurrence interval (ARI) could be excess of 5-year and up to 100-year. This system hope can protect community from severe flood damage, injury and loss of life.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the Earth Observation Center, Universiti Kebangsaan Malaysia and the Department of Irrigation and Drainage (DID), Ministry of Natural Resources and Environment, Malaysia for providing the data for this research. The authors would also like to thank the Ministry of Higher Education (MOHE), Malaysia and Yayasan Sime Darby (YSD) for supporting this research via research grants FRGS/1/2016/STG06/UKM/02/1 and ZF-2017-008, respectively.

REFERENCES

Adamowski, K. 2000. Regional analysis of annual maximum and partial duration flood data by nonparametric and L-moment methods. *Journal of Hydrology* 229(3-4): 219-231.

Anderson, T.W. & Darling, D.A. 1954. A test of goodness-of-fit. *Journal of the American Statistical Association* 49(268): 765-769.

Benjamin, J. & Cornell, C. 1970. *Probability, Statistics, and Decision for Civil Engineers*. New York: McGraw Hill.

Bernama. 2014. Kajang town flooded after downpour. *https://www.thestar.com.my/news/nation/2014/09/29/kajang-town-flooded/*.

Bernama. 2014. Flash floods hit Kajang. *https://www.thestar.com.my/news/nation/2014/11/12/kjang-flash-floods/*.

Bhunya, P.K., Berndtsson, R., Jain, S.K. & Kumar, R. 2013. Flood analysis using negative binomial and generalized pareto models in partial duration series (PDS). *Journal of Hydrology* 497: 121-132.

Borujeni, S.C., Nor, W. & Sulaiman, A. 2009. Development of L-moment based models for extreme flood events. *Malaysian Journal of Mathematical Sciences* 3(2): 281-296.

Cohen, A.C. & Whitten, B.J. 1980. Estimation - Parameter in the three lognormal distribution. *Journal of the American Statistical Association* 75(370): 399-404.

Coles, S.G. & Dixon, M.J. 1999. Likelihood-based inference for extreme value models. *Extremes* 2(1): 5-23.

D'Agostino, R. 1986. *Goodness-of-Fit-Techniques*. New York: Marcel Dekker, Inc.

Esteves, L.S. 2013. Consequences to flood management of using different probability distributions to estimate extreme rainfall. *Journal of Environmental Management* 115: 98-105.

Fisher, R.A. & Tippett, L.H.C. 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society* 24(02): 180-190.

Fréchet, M. 1927. Sur la loi de probabilité de l'écart maximum. *Annales de La Societe Polonaise de Mathematique Cracovie* 6: 93-116.

Gasim, M.B., Surif, S., Mokhtar, M., Toriman, M.E.H., Rahim, S.A. & Bee, C.H. 2010. Analisis Banjir Disember 2006: Tumpuan di Kawasan Bandar Segamat, Johor. *Sains Malaysiana* 39(3): 353-361.

Gilroy, K.L. & McCuen, R.H. 2012. A nonstationary flood frequency analysis method to adjust for future climate change and urbanization. *Journal of Hydrology* 414-415: 40-48.

Google Maps. 2015. *Langat River at Kajang*. https://www.google.com/maps/place/2%C2%B059'40.0%22N+101%C2%B047'10.0%22E/@3.0072332,101.6447652,93654m/data=!3m1!1e3!4m5!3m4!1s0x0:0x0!8m2!3d2.9944444!4d101.7861111.

Gumbel, E. 1958. *Statistics of Extremes*. Mineola, New York: Dover Publications Inc.

Hailegeorgis, T.T. & Alfredsen, K. 2017. Regional flood frequency analysis and prediction in ungauged basins including estimation of major uncertainties for mid-Norway. *Journal of Hydrology: Regional Studies* 9: 104-126.

Hashim, N. 2003. Flash flood, imperviousness and urban drainage in Malaysia: The geographical perspective. In *Urban Ecosystem Studies in Malaysia: A Study of Change,* edited by Hashim, N. & Rainis, R. California: Universal Publisher.

Hassan, B.G.H. & Ping, F. 2012. Regional rainfall frequency analysis for the Luanhe Basin - by using L-moments and cluster techniques. *APCBEE Procedia* 1(January): 126-135.

Hosking, J.R.M. 1992. Moments or L moments? An example comparing two measures of distributional shape. *The American Statistician* 46(3): 186-189.

Hosking, J. & Wallis, J. 1987. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* 29(3): 339-349.

Hosking, J. & Wallis, J. 1997. *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge: Cambridge University Press.

Katz, R.W. 2002. Techniques for estimating uncertainty in climate change scenarios and impact studies. *Climate Research* 20: 167-185.

Khaliq, M.N., Ouarda, T.B.M.J., Ondo, J.C., Gachon, P. & Bobee, B. 2006. Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *Journal of Hydrology* 329(3-4): 534-552.

Kjeldsen, T.R., Ahn, H. & Prosdocimi, I. 2017. On the use of a four-parameter kappa distribution in regional frequency analysis. *Hydrological Sciences Journal* 62(9): 1354-1363.

Kjeldsen, T.R. & Jones, D.A. 2004. Sampling variance of flood quantiles from the generalised logistic distribution estimated using the method of L-moments. *Hydrology and Earth System Sciences* 8(2): 183-190.

Li, M., Shao, Q., Zhang, L. & Chiew, F.H.S. 2010. A new regionalization approach and its application to predict flow duration curve in ungauged basins. *Journal of Hydrology* 389(1-2): 137-145.

Lim, Y.E.O.H. & Lye, L.M. 2003. Regional flood estimation for ungauged basins in Sarawak, Malaysia. *Hydrological Sciences Journal* 48(1): 79-94.

Madsen, H., Pearson, C.P. & Rosbjerg, D. 1997. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 2. Regional modeling. *Water Resources Research* 33(4): 759-769.

Malekinezhad, H., Nachtnebel, H.P. & Klik, A. 2011. Comparing the index-flood and multiple-regression methods using L-moments. *Physics and Chemistry of the Earth, Parts A/B/C* 36(1-4): 54-60.

Martins, S. & Stedinger, J.R. 2000. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research* 36(3): 737-744.

Mélice, J.L. & Reason, C.J.C. 2007. Return period of extreme rainfall at George, South Africa. *South African Journal of Science* 103(11-12): 499-501.

Millington, N., Das, S. & Simanovis, S.P. 2011. The comparison of GEV, Log-Pearson type 3 and Gumbel distributions in the upper thames river watershed under global climate models. *Water Resources Research Report*. p. 40.

Moralles, H.F., Rebelatto, D.A.D.N. & Sartoris, A. 2013. Parametric VaR with goodness-of-fit tests based on EDF statistics for extreme returns. *Mathematical and Computer Modelling* 58(9-10): 1648-1658.

Mosaffaie, J. 2015. Comparison of two methods of regional flood frequency analysis by using L-moments. *Water Resources* 42(3): 313-321. https://doi.org/10.1134/S0097807815030112.

Petersen, G. 2009. Managing extreme flood events: Analysing, forecasting, warning, protecting and informing-case studies from the RIMAX projects. *Managing Extreme Flood Events*. pp. 1-106.

Provost, S.B., Saboor, A., Cordeiro, G.M. & Mansoor, M. 2018. On the *q*-generalized extreme value distribution. *REVSTAT - Statistical Journal* 16(1): 45-70.

Rahman, H.A. 2007. Suatu tinjauan terhadap permasalahan banjir kilat di Lembah Klang. *Persidanagan Geografi UPSI* September: 8-9.

Razali, N. & Wah, Y. 2011. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Llilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics* 2(1): 21-33.

Sai Krishna, G. & Veerendra, G.T.N. 2015. Flood frequency analysis of Prakasam Barrage Reservoir Krishna District, Andhra Pradesh using Weibull, Gringorten and L-moments formula. *International Journal of Civil, Structural, Environmental and Infrastructure Engineering Research and Development* 5(2): 57-62.

Searcy, J.K. 1959. Flow-duration curves. *Manual of Hydrology*. Part 2. Low-Flow Technique. Geological Survey Water-Supply Paper 1542-A.

Seckin, N., Haktanir, T. & Yurtal, R. 2011. Flood frequency analysis of Turkey using L-moments method. *Hydrological Processes* 25(22): 3499-3505.

Shabri, A. & Jemain, A.A. 2010. LQ-moments: Parameter estimation for Kappa distribution. *Sains Malaysiana* 39(5): 845-850.

Shao, Q., Zhang, L., Chen, Y.D. & Singh, V.P. 2009. A new method for modelling flow duration curves and predicting streamflow regimes under altered land-use conditions/ Une nouvelle méthode de modélisation des courbes de débits classés et de prévision des régimes d'écoulement sous conditions modifiées d'o. *Hydrological Sciences Journal* 54(3): 606-622.

Victoria Brown, Beh Yuen Hui & Nicholas Cheng. 2014. Kajang struck by flash floods after heavy rain. https://www.thestar.com.my/news/nation/2014/12/20/kajang-flash-floods/.

Vogel, R.M., Thomas, W.O. & Mcmahon, T.A. 1993. Flood-flow frequency model selection in Southwestern United States. *Water Resources* 119(3): 353-366.

Wu, Y.C., Liou, J.J., Su, Y.F. & Cheng, K.S. 2012. Establishing acceptance regions for L-moments based goodness-of-fit tests for the Pearson type III distribution. *Stochastic Environmental Research and Risk Assessment* 26(6): 873-885.

Yue, S. 2001. A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrological Processes* 15(6): 1033-1045.

Zaman, M.A., Rahman, A. & Haddad, K. 2012. Regional flood frequency analysis in arid regions: A case study for Australia. *Journal of Hydrology* 475: 74-83.

Firdaus Mohamad Hamzah*, Siti Hawa Mohd Yusoff & Othman Jaafar
Department of Civil and Structural Engineering
Faculty of Engineering and Built Environment
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor Darul Ehsan
Malaysia

Siti Hawa Mohd Yusoff
Department of Science and Biotechnology
Faculty of Engineering and Life Sciences
Universiti Selangor
45600 Bestari Jaya, Selangor Darul Ehsan
Malaysia

*Corresponding author; email: fir@ukm.edu.my

Received: 12 November 2018
Accepted: 5 April 2019