ISSN: 0974-5823 Vol. 7 (Special Issue 4, 2022)

International Journal of Mechanical Engineering

Multiparameter probability distributions of at-site L-moment-based frequency analysis in Malaysia

Siti Hawa Mohd Yusoff [1,2], Firdaus Mohamad Hamzah [1*], Othman Jaafar [1]

[1]*Department of Civil and Structural Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia.*

[1,2]*Department of Science and Biotechnology, Faculty of Engineering and Life Sciences, Universiti Selangor, 45600 Bestari Jaya, Selangor, Malaysia.*

Abstract

In the present study, flood frequency analysis of annual maximum streamflow was investigated for Langat River Basin which is located in Selangor, Malaysia. There are four streamflow gauging stations along the Langat River Basin. The study aimed to identify the best fit probability distribution to the streamflow data and estimate the return period of the extreme flood events. In this study, the L-moment method was implemented to estimate the parameter of probability, namely Exponential, Gamma, Kappa, Three-Parameter Lognormal (LN3), Generalized Extreme Value (GEV), and Pearson type 3 (PE3). The Exponential distribution became the best fit distribution for the Langat-Lui River. The Kappa distribution gave the best fit to the annual maximum series data of the Langat-Kajang and Semenyih Rivers. Meanwhile, the Gamma distribution was the best distribution for the Langat-Dengkil River. The return period was plotted by using selected probability distribution for each streamflow gauging station.

Keywords: Extreme events, Goodness-of-fit test, L-moment, Probability distribution, Return period, Streamflow.

1.0 Introduction

Extreme meteorological and hydrological events all over the world may lead to major disasters and result in heavy social and economic losses (Tian et al., 2011). One of the extreme meteorological and hydrological events is extreme streamflow which can lead to flood disaster. The extreme streamflow studies have been reported in some regions such as in China (Gao et al., 2017), USA (Duan et al., 2017), West Africa (Andersson et al., 2016), and New Zealand (Nagy, Mohssen, & Hughey, 2017). The characteristic of daily streamflow has been known as heavy-tailed probability distributed variable (Basso, Schirmer, & Botter, 2015). In the annual cycle streamflow data, streamflow time series often exhibits spikes that rise far above the typical values in the series. To capture this behaviour, the heavy-tailed distributions were used to represent the extreme streamflow data (Bowers, Tung, & Gao, 2012).

In the flood frequency analysis, the L-moment (Hosking & Wallis, 1997) is the common method used. Furthermore, this method was applied to identify an appropriate distribution type for representing a hydrological variable of interest of a site or a region. There have been a few studies to identify the application of L-moment and probability type of maximum streamflow. Wu, Zhang, & She (2012) carried out the frequency analysis on 32 hydrological stations in the Huai River Basin of China. The generalized logistic (GLO) and Three-Parameter Lognormal (LN3) were proven to be the optimal marginal distributions in modelling the spring, summer, and autumn in most sub-regions. A similar study was also conducted by Mosaffaie (2015) who reported that the generalized Pareto (GP) and generalized logistic (GLO) were the best fit distributions in 15 gauging stations at Qazvin, Iran. Anilan, et.al (2016) investigated the probability distribution of annual maximum flood of 38 gauging stations in the Eastern Black Sea Basin, Turkey using L-moment method. They found out that the log normal (LN) distribution was identified as the most appropriate distribution for the selected regions. In the Upper Vistula River Basin, Poland, Rutkowska, Zelazny, Kohnova, Lyp, and Banasik (2016) investigated the distributions of annual maximum streamflow of 52 mid-sized catchments. The catchment areas were clustered according to similar morphometric, land use, and rainfall variables.

In Malaysia, the regional frequency analysis was also conducted using an index flood estimation procedure based on L-moments where Lim & Lye (2003) found that the generalized extreme value (GEV) and generalized logistic (GLO) distributions were fit for the distribution of extreme flood events in the Sarawak region in Malaysia. Selaman, Said, & Putuhena (2007) carried out the study on magnitude and frequency of floods for Sarawak using plotting position formulas, for example, Weibull, Gringorten and L-Moments in the application of Gumbel distribution. Zalina, et.al (2002) claimed that generalized extreme value (GEV) distribution was the most appropriate distribution for describing the annual maximum rainfall series in Malaysia by using the L-moment method. Shabri & Ariff (2009) performed regional frequency analysis using the L-moment method and found that generalized logistic (GLO) distribution was the most suitable distribution to fit the data of maximum daily rainfalls for stations in Selangor and Kuala Lumpur. A similar study conducted by Zawiah, et.al (2009) also reported that generalized extreme value (GEV) and generalized Pareto (GP) distributions were the best distributions on 50 rain-gauge stations in Peninsular Malaysia by using L-moment. Zakaria, Shabri, & Ahmad (2012) found that generalized extreme value (GEV) and generalized logistic (GLO) distributions were identified as the best distributions to represent the extreme rainfalls in Selangor by using the Partial L-moment method.

In this study, the objective is to identify the best fit probability distribution to the streamflow data in four flow stations along the Langat River Basin, Selangor by using the L-moment method. There were six distributions has been selected, namely Exponential, Gamma, Kappa, Three-Parameter Lognormal (LN3), Generalized Extreme Value (GEV), and Pearson Type 3 (PE3). Finally, the value for the return period will be determined based on the best-fit distribution.

2.0 Area study and data set

The Langat River shown in Fig 1 is situated in the state of Selangor, Peninsular Malaysia and has a total catchment area of approximately 1815 km^2. It lies within the longitude of 101°17′E to 101°55′E and latitudes of 2°40′N to 3°17′N (Yang, et.al, 2011). It is one of the most important basins which supply water to two third of the state of Selangor. However, the Langat River has several tributaries with the principle ones being the Semenyih and Lui Rivers. There are two reservoirs, the Langat and Semenyih Reservoir. The Langat River generally flows from the Titiwangsa Range at the Northeast of Hulu Langat District and drains into the Straits of Malacca. From Table 1, along with the Langat River basin, there are four flow gauging stations, namely Sg. Lui at Kg Sg. Lui (Langat-Lui River) with station no. 3118445, Sg. Langat at Kajang (Langat-Kajang River) with station no. 2917401, Sg. Semenyih at Kg. Rinching (Semenyih River) with station no. 2918401 and Sg. Langat at Dengkil (Langat-Dengkil River) with station no. 2816441. The Langat River hydrological characteristics are greatly influenced by two heavy rainy seasons during the South-West (May – September) and North-East (November-March) monsoons. Meanwhile, convectional rain is common during the inter-monsoon period. Malaysia receives heavy rainfall between 2000 and 3000 mm per year (Hamzah, Saimi, & Jaafar, 2017). Meanwhile, the Langat River Basin receives between 1900 mm to 3000 mm of rainfall per year. It is shown that the Langat River Basin receives a high value of precipitation and can cause extreme streamflow in that area. All the streamflow data were taken from the Department of Drainage and Irrigation (DID), Malaysia.

Table 1: Flow gauging stations in the Langat River Basin

No.	Site name	Station number	Catchment areas (km^2)	Latitude	Longitude	Mean elevation (m)	Sample size (years)	Period of data
1.	Sg. Lui at Kg. Sg Lui	3118445	68.1	03°10′25″N	101°52′20″E	76.8	52	1965 – 2016
2	Sg. Langat at Kajang	2917401	389.4	02°59′40″N	101°47′10″E	22.9	39	1978 – 2016
3	Sg. Semenyih at Kg. Rinching	2918401	226.6	02°54′55″N	101°49′25″E	22.0	41	1976 – 2016
4.	Sg. Langat at Dengkil	2816441	1251.4	02°51′20″N	101°40′55″E	3.9	57	1960 – 2016

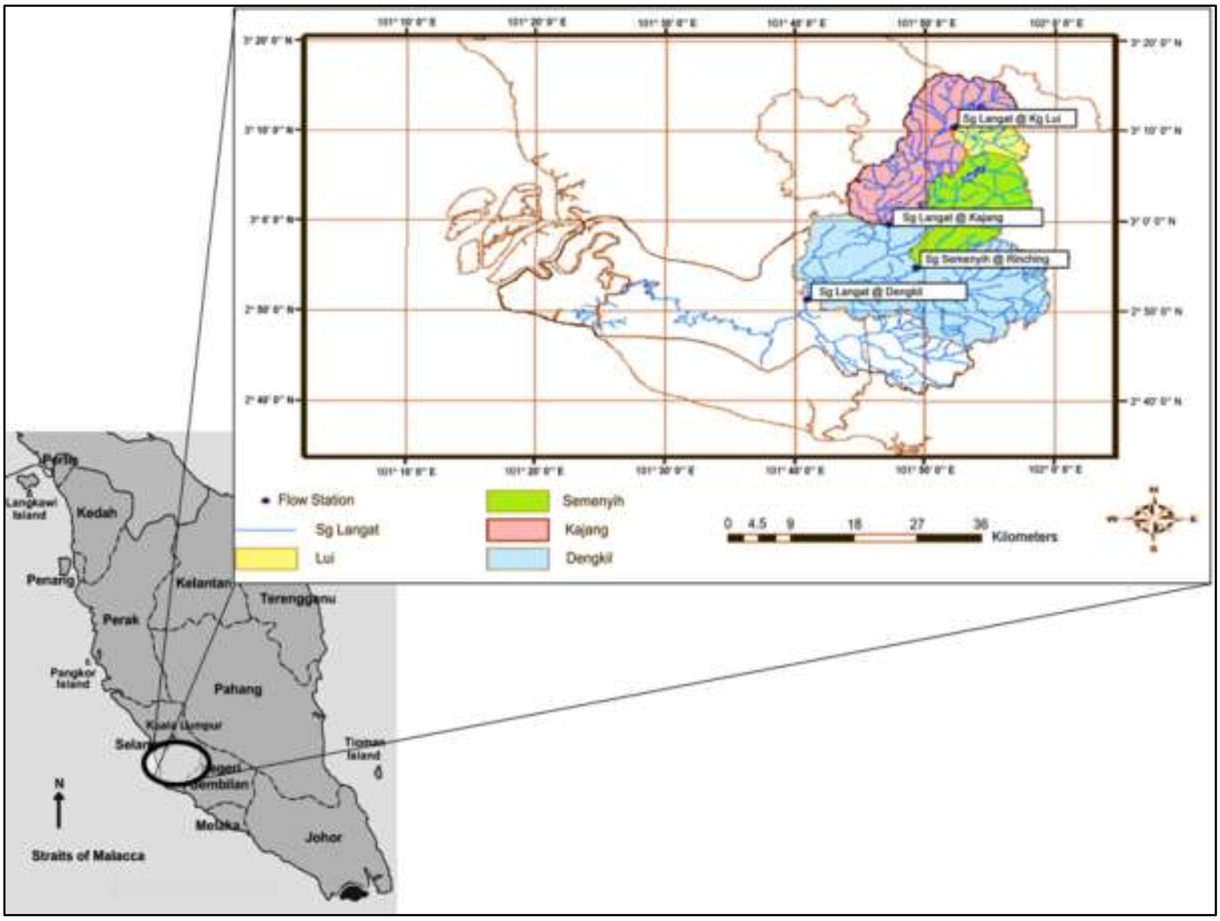

Fig 1. Map of the Langat River basin located in Selangor, Peninsular Malaysia.

3.0 Methodology

3.1 L-Moment Method

The L-moment method was used to estimate the flood frequency analysis of annual maximum streamflow data. It was obtained by taking the largest value in each year of interest (Esteves, 2013). 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. This method offers more stable and realistic estimate for the shape parameter of the small sample size compared to other estimation methods (Hosking & Wallis, 1997; Martins & Stedinger, 2000; Zhou, et.al, 2017). The L-moment summary statistics can be used to select which distributions are consistent with a given data sample (Noto & La Loggia, 2009). Also it can be used to estimate parameters when fitting a distribution to a sample, by equating sample and distribution L-moments (Hosking, 1990). The shape of a probability distribution has been described by using the moment method. In addition, the equations of the parameter estimation of the L-moment method in this study was mentioned by (Hosking & Wallis ,1997; Millington, Das, & Simanovis, 2011). The first four L-moment formulas are followed:

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

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

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

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \tag{1(d)}$$

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 above are derived from the following probability weighted moments:

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

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

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

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

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:

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

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

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

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

3.2. Probability Distribution

In this study, six types of probability distributions were used, namely exponential (Salarpour et al., 2012) in Eq. (4); gamma, (Yue 2001) in Eq. (5); four parameter kappa (Shabri & Jemain 2010) in Eq. (6); LN3 by Cohen & Whitten (1980) in Eq. (7); GEV by Millington et al. (2011) in Eq. (8); and PE3 by Wu, et.al, (2012) in Eq. (9). The following is the probability density function for each distribution.

a) Exponential distribution

$$f(x) = \alpha^{-1}\exp\left[\frac{-(x-\beta)}{\alpha}\right], \quad \beta \leq x < \infty \quad (4)$$

Where β is the location parameter and α is the scale parameter.

b) Gamma distribution

$$f(x) = \frac{\beta^{-\alpha}x^{\alpha-1}}{\Gamma(\alpha)}\exp\left(\frac{-x}{\beta}\right), \quad \alpha > 0,\ \beta > 0,\ x > 0 \quad (5)$$

Where α is the shape parameter and β is the scale parameter.

c) Four Parameter Kappa distribution (Kappa)

$$f(x) = \frac{1}{\alpha}\left[1-\frac{k}{\alpha}(x-\xi)\right]^{1/k-1}\left\{1-h\left[1-\frac{k}{\alpha}(x-\xi)\right]^{1/k}\right\}^{1/h-1} \quad (6)$$

Where ξ is the location parameter, α is the scale parameter, and κ and h are the shape parameters

d) Three-Parameter Log Normal distribution (LN3)

$$f(x) = \frac{1}{(x-\gamma)\sigma\sqrt{2\pi}}\exp\left[-\frac{\left[\ln(x-\gamma)-\mu\right]^2}{2\sigma^2}\right] \quad (7)$$

Where $\sigma^2 > 0,\ \gamma < x < \infty$

Where γ, μ and σ are respectively, the location, scale and shape parameters

e) Generalized Extreme Value distribution (GEV)

$$f(x) = \alpha^{-1}\exp\left[-(1-\kappa)y-\exp(-y)\right] \quad (8)$$

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.

f) Pearson Type 3 distribution (PE3)

$$f(x)=\frac{1}{\alpha\Gamma(\gamma+1)}\left(\frac{x-\mu}{\alpha}\right)^{\gamma}\exp\left[-\left(\frac{x-\mu}{\alpha}\right)\right] \tag{9}$$

Where μ, σ and γ are respectively, the location, scale and shape parameters.

3.3 Goodness-Of-Fit Test

The 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 were used to measure the different distance between the empirical and theoretical cumulative density functions (Moralles, Rebelatto, & Sartoris, 2013; Shin, et.al, 2012). In this study, the Anderson-Darling (AD) test in Eq. (10) and Kolmogorov-Smirnov (KS) test in Eq. (11) were used for testing the goodness-of-fit tests. The AD test statistics is defined by:

$$A_n^2=-n-\sum_{i=1}^{n}\frac{2i-1}{n}\left[\log\left(\hat{F}(x_{(i)})\right)+\log\left(1-\hat{F}(x_{(n+1-i)})\right)\right] \tag{10}$$

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. While KS test statistics is given by:

$$D=\max_x\left|F(x)-F_n(x)\right| \tag{11}$$

If $D\le D_{1-\alpha}$, where $D_{1-\alpha}$ is the critical values at a significant α, it means that the sample has the same distribution with the tested theoretical distribution. $F(x)$ is the cumulative distribution function of the theoretical distribution and $F(x_n)$ is the empirical distribution function.

3.4 Return Period

The return period or recurrence interval can be defined as the average number of trials usually in the year to the first occurrence of an event of magnitude greater than a predefined critical event. It is a measure of how often flood event of certain magnitude is likely to happen (Mélice & Reason, 2007). 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 (Benjamin & Cornell, 1970). However, the existing methods to estimate the 100-year flood assume 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 can be expressed as

$$T(x)=\frac{1}{P(x)} \tag{12}$$

Where $T(x)$ corresponds to years of return period of such a design flood and $P(x)$ is an exceedance probability (Gumbel, 1941; Salas & Jayantha Obeysekera, 2014).

4.0 Results and discussion

An attempt has been made to estimate the annual maximum streamflow by six distributions at Langat-Lui River at Kg. Lui station, Semenyih River at Kg. Rinching station, Langat-Kajang River station and Langat-Dengkil River station. These gauging stations are located at west coast of Peninsular Malaysia, mainly at Selangor state. The data can be shown by using graph in Fig 2. Fig 2 shows the time series plot of high streamflow for the four rivers along the Langat River, Malaysia. Each plot represents the highest streamflow taken each year of the study period. The Langat-Dengkil River station had the longest historical record of streamflow readings followed by the Langat-Lui River at Kg. Lui station, Semenyih River at Kg. Rinching station and lastly the Langat-Kajang River station.

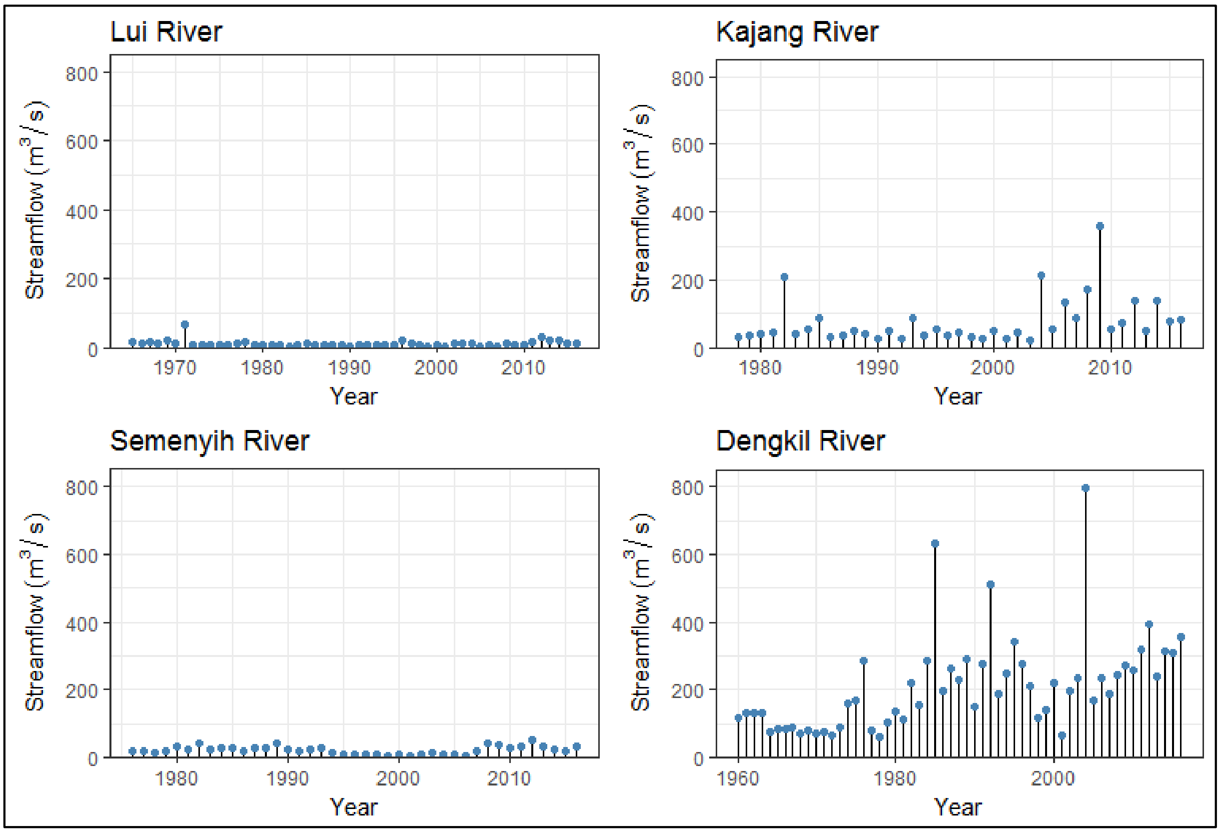

Fig 2. Time series plot of annual maximum streamflow for the Langat-Lui, Langat-Kajang, Semenyih, and Langat-Dengkil Rivers

The Langat-Lui gauging station located at latitude 03°10′25″N and longitude 101°52′20″E represent the upstream of the Langat River. The annual maximum streamflow values for Langat-Lui flow gauging station is fluctuating in a smaller range and quite stable based on the 52 years data ranges between 2.75m^3/s to 65.02 m^3/s. The same goes for the Semenyih River located at latitude 02°54′55″N and longitude 101°49′25″E which represent the lower part of the Langat River. The annual maximum streamflow for Semenyih River fluctuated in a smaller range based on 41 years data ranges between 6.22m^3/s to 52.27 m^3/s. The Langat-Kajang River station located at latitude 02°59′40″N and longitude 101°47′10″E represent the middle part of the Langat River showed the fluctuation of annual maximum streamflow data beginning the year 2003. It is because of tremendous development in that area, resulting in changes of river morphologies. Toriman (2008) stated that the environment pattern in the Langat Catchment area has changed because of the development pressure in that region. It is also the primary factor resulting in altered catchment hydrology and sediment production. Langat-Dengkil flow gauging station located at latitude 02°51′20″N and longitude 101°40′55″E has higher minimum value for every month compared to the other stations due to the much bigger catchment area located in the downstream part of the Langat River (Yang et al., 2011). In contrast, the Langat-Dengkil flow gauging station showed higher inconsistency of flow rate with more rigorous fluctuation. Langat-Dengkil River station has 57 years of data with a range of annual maximum streamflow between 59.08 m^3/s to 796.6 m^3/s.

Table 2. L-Moments statistics for the Langat-Lui, Langat-Kajang, Semenyih, and Langat-Dengkil Rivers

No.	Site name	L-location	L-scale	L-CV	L-skewness	L-kurtosis
1.	Sg. Lui at Kg. Sg Lui (Langat-Lui River)	10.86	3.906	0.359	0.429	0.316
2	Sg. Langat at Kajang (Langat-Kajang River)	74.29	30.689	0.413	0.499	0.303
3	Sg. Semenyih at Kg. Rinching (Semenyih River)	22.22	6.499	0.293	0.117	0.069
4.	Sg. Langat at Dengkil (Langat-Dengkil River)	209.62	70.082	0.334	0.264	0.178

Table 2 gives a description of the L-moment parameters that can be used as a guideline for selecting suitable probability distributions. The value of L-location describes the mean of the annual maximum streamflow data. The L-location of Langat-Lui

River is 10.86 m^3/s smaller compared to Langat-Dengkil River 209.62 m^3/s. This is because of the location of Langat-Lui River is at the upstream while Langat-Dengkil River is at the downstream of the Langat River basin. The L-scale describes the spread of the distribution. The larger the scale parameter the more spread out the distribution. From Table 2, Langat-Dengkil River (70.082) showed the high variation of maximum streamflow compared to the other three site stations. The L-coefficient variation showed the ratio between mean and standard deviation. The parameter L-skewness indicated that the distribution was skewed to the right based on the positive value of the L-skewness. It indicated that the tail on the right side was longer than the left side. Langat-Lui River (0.429), Langat-Kajang River (0.499) and Langat-Dengkil River (0.264) showed the right tail extremes because there were a few extreme streamflow data in the data set. Meanwhile, the L-kurtosis measures the peak of the distribution. Langat-Lui River (0.316) and Langat-Kajang River (0.303) showed that the central peak is high compared to the other stations. The candidate distributions based on the L-moment parameters are exponential, gamma, GEV, kappa, LN3 and PE3 distributions.

The data from four stations were fitted using the exponential distribution Eq. (4), Gamma distribution Eq. (5), four-parameter Kappa distribution Eq. (6), LN3 Eq. (7), GEV distribution Eq. (8), and PE3 distributions Eq. (9). It can be seen in Fig 3. In Fig 3, from the Langat-Lui River, the data fit closely to the exponential distribution compared to the other five distributions. From the Langat-Kajang River and Semenyih River, it can be seen that the shape of probability density function kappa was much closer to the shape of the histogram. Meanwhile, for Langat-Dengkil River, the probability density function of Gamma showed the closest shape to the histogram.

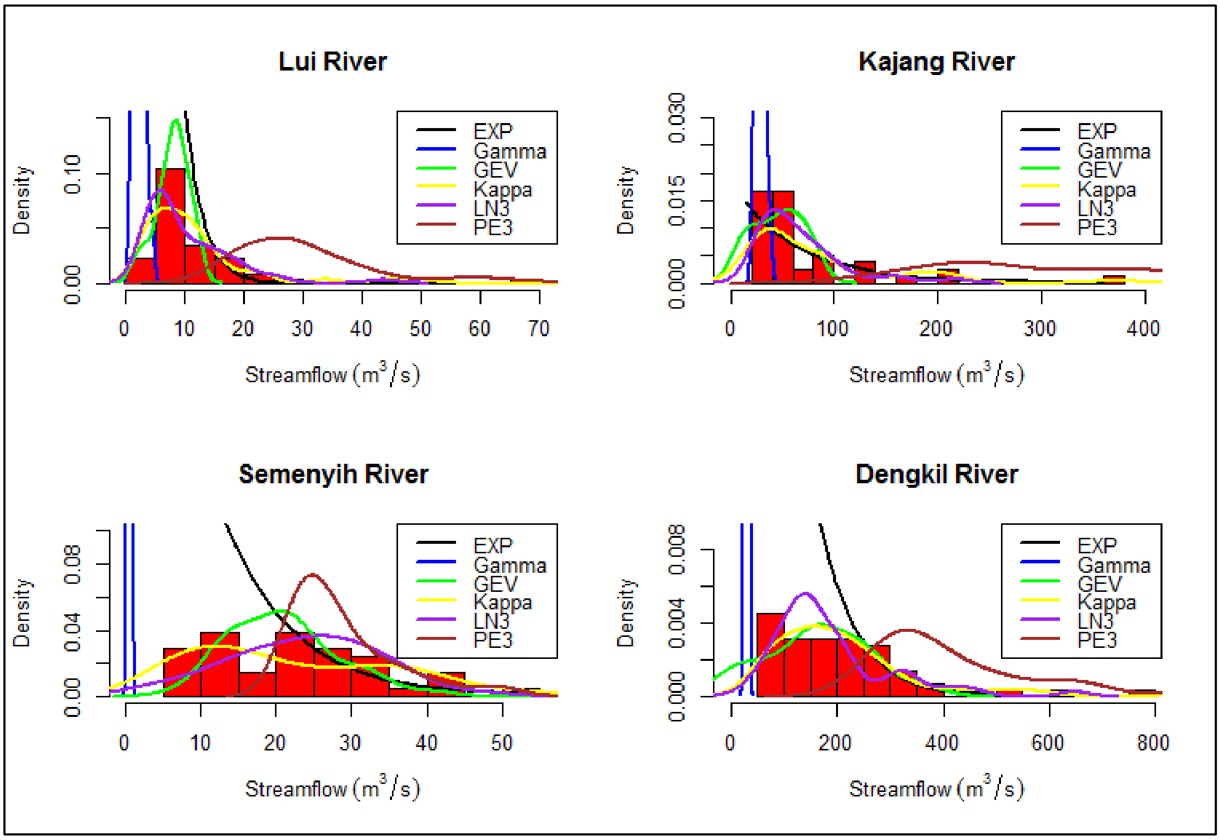

Fig 3. Probability density function of six types of distributions plot for the Langat-Lui, Langat-Kajang, Semenyih, and Langat-Dengkil Rivers

The goodness-of-fit test was used for better decision of the best fit model to the four stations at the Langat River. If the p-value is greater than 0.05 at a significance level it indicates that the data follow a specified distribution. Table 3 represents the goodness-of-fit test results for four flow stations with six selected probability distributions. From the Langat-Lui River, it showed that the data follows all the six distributions after being tested using the AD and KS tests. The best distribution for both goodness-of-fit tests is highlighted in bold. The best distribution was chosen based on the highest p-values in each AD and KS tests. The exponential distribution showed the best distribution according to AD and KS with p-value is 0.603 and 0.829 respectively. While the Langat-Kajang River showed that the data follows GEV, Kappa, PE3, and LN3 distributions. The highest p-value according to AD and KS belong to Kappa distribution with the value of 0.417 and 0.679, respectively. For the Semenyih River, based on AD, only Kappa became the best fit distribution with p-value 0.069. Meanwhile, based on KS, all distributions fit the data except exponential. But only Kappa distribution was selected from KS since the p-value was the highest among other distributions. Lastly, for the Langat-Dengkil River, only Gamma distribution was selected according to KS since the p-value (0.099) was more than 0.05 of the level of significance. Meanwhile, according to AD, all the distributions were not fit to the distribution of the Dengkil River data. In Table 2, the value of L-skewness for the Langat-Dengkil River data was less than the Langat-Lui and Langat-Kajang Rivers with the reading of 0.264. It means that the Langat-Lui and Langat-Kajang River distributions were more skewed compared to the Langat-Dengkil River. The reasons why AD test did not give any significant result is because the Langat-

Dengkil River distribution was less skewed and AD test is much more sensitive to the tails of the distribution (Engmann & Cousineau, 2011).

The location of the gauging station can affect the result of the distribution selection since all the gauging stations have a different mean of elevation (Table 1). The best fit distribution for the Langat-Lui River is exponential distribution, since the location of the Langat-Lui River is at the upstream of the Langat River Basin with the reading of normal stage height from the datum is 76.8 m. The Langat-Kajang River and Semenyih River share the same distribution, which is Kappa distribution, since the stage height of these two gauging stations are approximately same. The normal stage height of Langat-Kajang River from the datum is 22.9 m while the Semenyih River the normal stage height is 22.0 m. Meanwhile, the best fit distribution for the Langat-Dengkil River was Gamma distribution. The location of the Langat-Dengkil River is at the downstream of the Langat River Basin with the reading of normal stage height is 3.9 m.

Table 3: Goodness of fit tests according to AD and KS test for the Langat-Lui, Langat-Kajang, Semenyih, and Langat-Dengkil Rivers

Station	Distribution	Anderson Darling	p-value	Kolmogorov Smirnov	p-value
Lui River	**Exponential**	**0.289**	**0.603**	**0.066**	**0.829**
	Gamma	0.737	0.052	0.105	0.158
	GEV	0.658	0.081	0.078	0.591
	Kappa	0.721	0.056	0.083	0.499
	PE3	0.383	0.384	0.067	0.818
	LN3	0.481	0.224	0.071	0.755
Kajang River	Exponential	0.979	0.012	0.157	0.016
	Gamma	1.416	0.001	0.185	0.002
	GEV	0.685	0.068	0.099	0.440
	Kappa	**0.366**	**0.417**	**0.085**	**0.679**
	PE3	0.497	0.201	0.091	0.566
	LN3	0.455	0.255	0.090	0.584
Semenyih River	Exponential	1.589	0.000	0.149	0.023
	Gamma	0.970	0.013	0.125	0.112
	GEV	0.935	0.016	0.128	0.088
	Kappa	**0.684**	**0.069**	**0.114**	**0.199**
	PE3	0.907	0.019	0.127	0.095
	LN3	0.932	0.016	0.128	0.089
Dengkil River	Exponential	1.371	0.001	0.134	0.012
	Gamma	1.114	0.006	**0.107**	**0.099**
	GEV	1.424	0.001	0.123	0.031
	Kappa	1.333	0.002	0.122	0.035
	PE3	1.151	0.005	0.118	0.047
	LN3	1.332	0.002	0.122	0.035

After the best fit distribution has been selected by using goodness-of-fit method, the return period for each station can be generated. In Fig 4, the return period plot is represented by each gauging station in the Langat River basin. From the plots, the Langat-Lui River was using an exponential distribution from Eq. (4), both the Langat-Kajang and Semenyih Rivers were using Kappa distribution from Eq. (6), while the Langat-Dengkil River was using Gamma distribution from Eq. (5). All the selected distributions were chosen by the goodness-of-fit test in Table 3. The return period plot is a convenient way to see the tail of the distribution compared to the histogram, as it draws attention to the middle of the distribution. The return period indicated the likelihood of an event occurring. For example, from the Langat-Dengkil River, the 100 years return period had a 1% chance of occurring in any given year with a flow rate of 800 m^3/s.

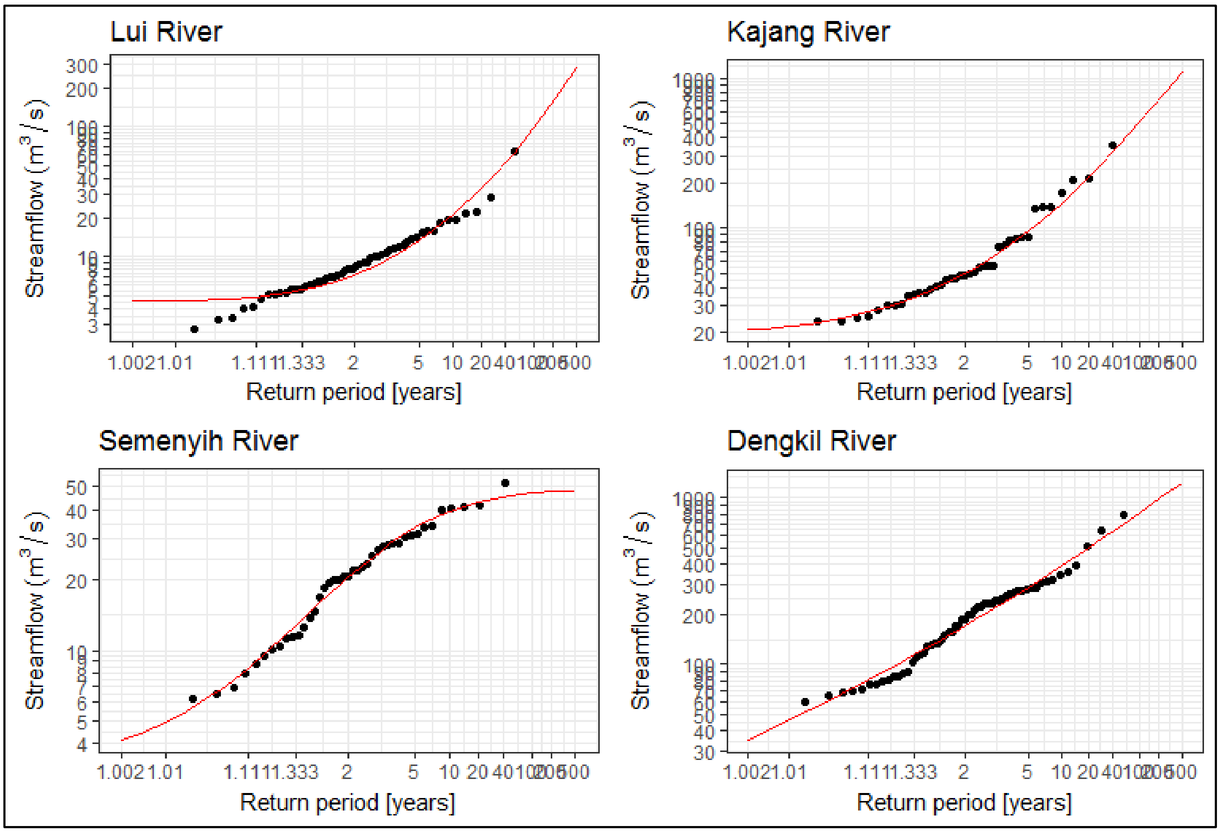

Fig 4. Return period plot for the Langat-Lui, Langat-Kajang, Semenyih and Langat-Dengkil Rivers. The red line represents the return period

5.0 Conclusion

Historical streamflow data from 4 gauging stations along the Langat River Basin 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.

There were four flow stations involved in this study, namely the Sg. Lui at Kg. Sg.Lui (Langat-Lui River), Sg. Langat at Kajang (Langat-Kajang River), Sg. Semenyih at Kg. Rinching (Semenyih River) and Sg. Langat at Dengkil (Langat-Dengkil River) which observation records spanned more than 30 years. The study used the application of the L-moment method on the annual maximum streamflow data. It has been proven that L-moment has good properties to measure the distribution shape and is useful for fitting distributions to data. This study also showed the relationship between heavy-tailed probability distribution with extreme streamflow data.

There were six candidates of probability distributions distribution such as exponential, Gamma, Kappa, three-parameter lognormal (LN3), generalised extreme value (GEV) distribution and Pearson type 3 distributions (PE3) to represent the heavy-tailed distributions. In order to select the best fit distribution to match with the actual annual maximum streamflow data, the goodness-of-fit tests were used. In this study, the Anderson-Darling test (AD) and Kolmogorov-Smirnov test (KS) were used as the goodness-of-fit tests. The best distribution chosen for the Langat-Lui River was exponential distribution. Kappa distribution was the best distribution for the Langat-Kajang River and Semenyih River. Meanwhile, for the Langat-Dengkil River, the Gamma distribution was the best distribution according to the Kolmogorov-Smirnov (KS) test. The height of the gauging station from the datum can affect the selection of the distribution.

After that, each of the flow stations was presented by the return period plot. The return period plot was constructed by using selected distribution. From the result, it can be concluded that the flow was getting an increase as the return periods increase.

6.0 Acknowledgments

The authors would like to acknowledge the Earth Observation Centre, Universiti Kebangsaan Malaysia and 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

1. Andersson, J. C. M., Ali, A., Arheimer, B., Gustafsson, D., & Minoungou, B. (2016). Providing peak river flow statistics and forecasting in the Niger River basin. *Physics and Chemistry of the Earth*, 1–10. https://doi.org/10.1016/j.pce.2017.02.010
2. Anilan, T., Satilmis, U., Kankal, M., & Yuksek, O. (2016). Application of Artificial Neural Networks and Regression Analysis to L-Moments Based Regional Frequency Analysis in the Eastern Black Sea Basin , Turkey, *KSCE Journal of Civil Engineering*, 20(5), 2082-2092. https://doi.org/10.1007/s12205-015-0143-4
3. Basso, S., Schirmer, M., & Botter, G. (2015). On the emergence of heavy-tailed streamflow distributions. *Advances in Water Resources*, *82*, 98–105. https://doi.org/10.1016/j.advwatres.2015.04.013
4. Benjamin, J., & Cornell, C. (1970). *Probability, statistics, and decision for civil engineers*. McGraw Hill.
5. Bowers, M. C., Tung, W. W., & Gao, J. B. (2012). On the distributions of seasonal river flows: Lognormal or power law? *Water Resources Research*, *48*(5), 1–12. https://doi.org/10.1029/2011WR011308
6. Cohen, A. C., & Whitten, B. J. (1980). Estimation - Parameter in the Three Lognormal Distribution. *Journal of the American Statistical Association*, *75*(370), 399–404.
7. Duan, J. G., Bai, Y., Dominguez, F., Rivera, E., & Meixner, T. (2017). Framework for incorporating climate change on flood magnitude and frequency analysis in the upper Santa Cruz River. *Journal of Hydrology*, *549*, 194–207. https://doi.org/10.1016/j.jhydrol.2017.03.042
8. Engmann, S., & Cousineau, D. (2011). Comparing distributions : the two-sample anderson-darling test as an alternative to the kolmogorov-smirnoff test. *Journal of Applied Quantitative Methods*, *6*(3), 1–17.
9. Esteves, L. S. (2013). Consequences to flood management of using different probability distributions to estimate extreme rainfall. *Journal of Environmental Management*, *115*, 98–105. https://doi.org/10.1016/j.jenvman.2012.11.013
10. Gao, Z., Long, D., Tang, G., Zeng, C., Huang, J., & Hong, Y. (2017). Assessing the potential of satellite-based precipitation estimates for flood frequency analysis in ungauged or poorly gauged tributaries of China's Yangtze River basin. *Journal of Hydrology*, *550*, 478–496. https://doi.org/10.1016/j.jhydrol.2017.05.025
11. 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. https://doi.org/10.1016/j.jhydrol.2011.10.009
12. Gumbel, E. J. (1941). The return period of flood flows. *JSTOR*, *12*(2), 163–190. Retrieved from http://www.jstor.org/stable/2235766
13. Hamzah, F. M., Saimi, F. M., & Jaafar, O. (2017). Identifying the Monotonic Trend in Climate Change Parameter in Kluang and Senai, *Sains Malaysiana*, *46*(10), 1735–1741. https://doi.org/10.1016/j.jmr.2004.05.021
14. Hosking, J. R. M. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*. https://doi.org/10.2307/2345653
15. Hosking, J., & Wallis, J. (1997). *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press.
16. 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. https://doi.org/https://doi.org/10.1623/hysj.48.1.79.43477
17. Martins, S., & Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, *36*(3), 737–744.
18. 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.
19. 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*, *77*(September).
20. 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. https://doi.org/10.1016/j.mcm.2013.07.002
21. 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
22. Nagy, B. K., Mohssen, M., & Hughey, K. F. D. (2017). Flood frequency analysis for a braided river catchment in New Zealand: Comparing annual maximum and partial duration series with varying record lengths. *Journal of Hydrology*, *547*, 365–374. https://doi.org/10.1016/j.jhydrol.2017.02.001
23. Noto, L. V., & La Loggia, G. (2009). Use of L-moments approach for regional flood frequency analysis in Sicily, Italy. *Water Resources Management*, *23*(11), 2207–2229. https://doi.org/10.1007/s11269-008-9378-x
24. Puchner, H., & Selberherr, S. (1995). Simulation of ion implantation using the four- parameter kappa distribution function. *Institute for Microelectronics, TU Vienna Gusshausstrasse*, 295–297.
25. Rutkowska, A., Zelazny, M., Kohnova, S., Lyp, M., & Banasik, K. (2016). Regional L-Moment-Based Flood Frequency Analysis in the Upper Vistula River Basin , Poland. *Pure and Applied Geophysics*, *174*(2), 701–721. https://doi.org/10.1007/s00024-016-1298-8
26. Salarpour, M., Yusop, Z., & Yusof, F. (2012). Modeling the distributions of flood characteristics for a tropical river basin. *Journal of Environmental Science and Technology, 5(6), 419-429.*
27. Salas, J. D., & Jayantha Obeysekera. (2014). Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. *Journal of Hydrologic Engineering, ASCE*, *19*, 554–568. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000820.
28. Selaman, O. S., Said, S., & Putuhena, F. J. (2007). Flood Frequency Analysis for Sarawak using Weibull, Gringorten and L-Moments Formula. *Journal - the Institution of Engineers, Malaysia*, *68*(1), 43–52.

29. Shabri, A., & Ariff, N. M. (2009). Frequency analysis of maximum daily rainfalls via L-moment approach. *Sains Malaysiana*, *38*(2), 149–158.
30. Shabri, A., & Jemain, A. A. (2010). LQ-moments: Parameter estimation for kappa distribution. *Sains Malaysiana*, *39*(5), 845–850.
31. Shin, H., Jung, Y., Jeong, C., & Heo, J. H. (2012). Assessment of modified Anderson-Darling test statistics for the generalized extreme value and generalized logistic distributions. *Stochastic Environmental Research and Risk Assessment*, *26*(1), 105–114. https://doi.org/10.1007/s00477-011-0463-y
32. Tian, P., Zhao, G., Li, J., & Tian, K. (2011). Extreme value analysis of streamflow time series in Poyang Lake Basin, China. *Water Science and Engineering*, *4*(2), 121–132. https://doi.org/10.3882/j.issn.1674-2370.2011.02.001
33. Toriman, M. E. (2008). Developing A Conceptual Model Of Geomorphic Change For Channel Management Proposes : A Case Of Langat River Basin An integrated approach to river basin development devolves upon recognition of the hydrological and geomorphological integrity of the fluvial. *Jurnal E-Bangi*, *3*(12), 1–12.
34. Wu, S., Zhang, X., & She, D. (2012). Joint pattern of seasonal hydrological droughts and floods alternation in China ' s Huai River Basin using the multivariate L-moments. *Journal of Earth System Science*. https://doi.org/10.1007/s12040-017-0824-0
35. 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. https://doi.org/10.1007/s00477-011-0519-z
36. Yang, H. H., Jaafar, O., El-Shafie, A., & Syed Abdullah, S. M. (2011). Analysis of hydrological processes of Langat river sub basins at Lui and Dengkil. *International Journal of Physical Sciences*, *6*(32), 7390-7409. https://doi.org/10.5897/IJPS11.1036
37. Yue, S. (2001). A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrological Processes*, *15*(6), 1033–1045. https://doi.org/10.1002/hyp.259
38. Zakaria, Z. A., Shabri, A., & Ahmad, U. N. (2012). Regional Frequency Analysis of Extreme Rainfalls in the West Coast of Peninsular Malaysia using Partial L-Moments. *Water Resources Management*, *26*(15), 4417–4433. https://doi.org/10.1007/s11269-012-0152-8
39. Zalina, M. D., Desa, M. N. M., Nguyen, V. T. V, & Kassim, A. H. M. (2002). Selecting a probability distribution for extreme rainfall series in Malaysia. *Water Science and Technology*, *45*(2), 63–68.
40. Zawiah, W. Z. W., Jemain, A. A., Ibrahim, K., Suhaila, J., & Sayang, M. D. (2009). A comparative study of extreme rainfall in peninsular malaysia: with reference to partial duration and annual extreme series. *Sains Malaysiana*, *38*(5), 751–760.
41. Zhou, Z., Liu, S., Hu, Y., Liang, Y., & Lin, H. (2017). Analysis of precipitation extremes in the Taihu Basin of China based on the regional L-moment method. *Hydrology Research*, *48*(2), 468–479. https://doi.org/10.2166/nh.2016.071