MATEMATIKA: MJIAM, 2021, Volume 37, Number 2, 47–62

Flood Frequency Analysis using L-Moment For Segamat River

[1]Basri Badyalina[*], [1]Nurkhairany Amyra Mokhtar, [2]Nur Amalina Mat Jan, [1]Nor Hidayah Hassim and [1]Haslenda Yusop

[1]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Johor, Kampus Segamat, 85000 Segamat, Johor, Malaysia

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

[*]Corresponding author: basribdy@uitm.edu.my

Article history
Received: 25 May 2021
Received in revised form: 29 July 2021
Accepted: 6 August 2021
Published online: 25 August 2021

Abstract Flood frequency analysis is critical in water system design and estimating flood recurrence. This study aims to conduct the flood frequency analysis on Segamat River stream flow site to find the optimum distribution that fits the flood frequency data. In terms of estimating parameters, the L-moment method is more robust and more efficient than the maximum likelihood method. Besides, the L-moment method is not affected by sampling variability. Therefore, in this study, we applied the technique of L-moment for parameter estimation on five candidate distributions, the generalized Pareto (GPA) distribution, generalized extreme value (GEV) distribution, generalized logistic (GLO) distribution, log-Pearson 3 (LP3) distribution, and log-normal (LN3) distribution. The rank score approach is implemented to determine the optimum distribution for the annual Segamat River peak flow. Probability distribution identification is essential and it is a fundamental step in statistical analysis. The goodness of fit test and efficiency assessment are employed to evaluate the distributions' performance. The results show that the LN3 distribution is selected as the optimum function for the yearly peak flow for the Segamat River stream flow site. The outcome of this study can be used to understand the flood frequency analysis for the Segamat River.

Keywords L-Moment, flood, probability distribution, parameter estimation, performance measures

Mathematics Subject Classification 46N30

1 Introduction

Floods are natural disasters that inflict widespread devastation around the world. Natural disasters result in infrastructure degradation, environmental and agricultural land disruption, mortality, and economic losses. Flood disasters in Malaysia occur annually, affecting approximately 29,720 km^2, impacting over 4.915 million people and triggering up to RM 915 million loss in infrastructure [1]. Rahman *et al* [2] stated that the most direct approach for assessing peak discharge estimation is the flood frequency analysis (FFA). Flood frequency analysis aims to determine how long it would take for a given flood severity to return. It depicts the relationship between an event's magnitude and the frequency by which it is surpassed [3,4,5].

Various FFA models were employed in previous studies to determine hydrologic frequency. Numerous probability models were constructed to define the frequency distribution of extreme hydrologic phenomena; however, there is no common understanding on which distribution must be employed to analyse the frequency of extreme hydrologic phenomena [6]. The most commonly used distributions in FFA for annual peak flow estimation of river sites in Malaysia are generalized pareto distribution, generalized extreme value distribution (GEV), generalized logistic distribution (GLO), log-Pearson distribution (P3), and log-normal distribution (LN3) [4, 7, 8].

The identification of an effective probability distribution to characterize the FFA at a given location is often crucial. The methods that are typically used for parameter estimation of the distribution are the maximum likelihood estimation and the L-moment method [9, 10, 11]. Many studies have explored and applied the L-moment approach to FFA since 1990, including hydrological studies [12]. The implementation of L-moments by Hosking and Wallis marked a revolution in flood frequency analysis [13]. L-moments are recognized for their robustness [14]. Hassan *et al.* [15] mentioned that the best-fitted distributions used the L-moment estimation approach at most sites. This approach is also frequently utilised to determine the scale and shape of probability distributions [16]. Furthermore, there are no sample size restrictions in using this approach. [10].

Daud *et al.* [17] found that the GEV distribution is the most effective for yearly maximum rainfall distribution in Peninsular Malaysia. Lim and Lye used the L-moment for the index-flood estimation for river sites in Sarawak and found that GEV and GLO distributions are adequate to represent the annual flood data [18]. In a study that applied the L-moment estimation for regional FFA in Northern Uganda, Kizza *et al.* [19] discovered that the log-normal distribution estimates the flood return period in Northern Uganda. A study by Badyalina & Shabri estimated the flood quantile for 70 stream flow sites located in Peninsular Malaysia and found that no single distribution appears to be the best fit for all 70 stream flow sites [20]. For the frequency study, Liang *et al.* [21] proposed using L-moments of GEV distribution to determine the Taihu site's annual daily peak rainfall return period in China.

Hamzah *et al.* [22] studied the L-moment to estimate the yearly peak daily stream flow to evaluate FFA for Langat stream flow site in Kajang, Selangor, using distributions LN3 GEV, PE3, GLO. In addition, Drissia *et al.* [15] applied the L-moments approach in conducting at-site FFA for west-flowing rivers in Kerala. GPA, GLO and GEV distributions were found to be the best fit. Hassan *et al.* [15] investigated the most suitable probability distribution for at-site FFA of the Torne River and its parameter estimation. They found that the LN3 distribution with the L-moment approach performs better than other distributions for Pajala Pumphus and Abisko stream flow sites located at Torne River. Lescesen *et al.* [23] applied the L-moment method in FFA for the Tisza stream flow site and identified the P3 distribution as the optimum distribution. Das [24] studied the L-moment method in finding the optimum distribution for at-site FFA in China. In other environmental applications, the L-moment method was used to analyse the regional frequency of peak daily rainfall sites in Selangor and Kuala Lumpur [25].

The previous studies on flood frequency of the Segamat River only consider the return period of flood discharge using Generalized Pareto distribution [4]. Also, Romali *et al.* [26] investigated the return period of flood level using the HEC-RAS model. Hence this paper seeks to add to the body of knowledge of flood frequency by applying the L-moment estimation method on five potential distributions, namely GEV, GLO, GPA, PE3 and LN3 for the Segamat stream flow site. The data used in this study is dated from the year 1960 until 2020. Section 2 describes the methodology used in the study, which includes the study area, the probability distributions, and the performance measures of the distributions. Section 3 summaries the findings, and Section 4 concludes the paper.

2 Methodology

2.1 Study Area

Segamat is a district in Johor, Malaysia, known for its frequent floods during the northeast monsoon season, which usually occurs from November to March [27]. The Segamat River is located at latitudes which are 102°49'E and longitudes of 2°30.5'N. It is 14 meters above sea level, 23 kilometers in length and has an average width of 40 meters. Figure 1 shows the location of the Segamat River. Table 1 presents the summarized descriptive statistics of Segamat gauged river site's annual peak flow or annual maximum flow data from 1960 until 2020. For the purpose of this study, the data analysed is limited to the annual peak flow. Therefore, the weekly and monthly peak flow data are not considered in this study. The mean peak flow for Segamat gauged river site is 234.59 with a median of 119, which is quite far from the mean. It shows that this data is highly spread data, and the extreme event has occurred in the data. The standard deviation is 271.84 with minimum and maximum values of 35.7 and 1559, respectively.

Table 1: Descriptive Statistics for Annual Peak Flow of Segamat River

Station	n	Mean	SD	Min	Max	Skewness	Kurtosis
Segamat	61	234.58	271.84	35.7	1559	2.73	9.19

The skewness of the streamflow data of the Segamat river is 2.73, indicating that the stream flow data is skewed to the right. The kurtosis of the Segamat River peak flow data is highly positive (9.19), demonstrating that it is leptokurtic and prone to extreme values. These results support our suggestion that the peak flow data series at the Segamat River is fat-tailed and non-normal. Therefore, the outcomes indicate that the peak flow data fit the non-normal distributions. Based on the data, this study implements five different distributions for annual maximum flow in Peninsular Malaysia, namely GEV, GLO, GPA, PE3 and LN3.

2.2 L-Moment

L-moments are a linear combination of probability-weighted moments (PWMs) proposed by Hosking (1990) [28]. L-moments are considered more efficient than PWMs since they are easily comprehended as a scale and shape measurement of probability distributions. L-moments are also regarded as being better than traditional moments. In terms of application, L-moments are frequently calculated from a finite sample. Assume $x_{1:n} \leq x_{2:n} \leq \cdots \leq x_{n:n}$ are the data in a specific order with a sample size of n. Landwehr *et al.* [29] described the L-moment method's unbiased sample estimator as follows:

$$b_r = \frac{1}{n}\binom{n-1}{r}^{-1}\sum_{i=r+1}^{n}\binom{i-1}{r}x_{i:n} \tag{1}$$

The first four components of an unbiased sample estimator are as follows

$$b_0 = \frac{1}{n}\sum_{i=1}^{n}x_{i:n} \tag{2}$$

$$b_1 = \frac{1}{n}\sum_{i=2}^{n}\frac{(i-1)}{(n-1)}x_{i:n} \tag{3}$$

$$b_2 = \frac{1}{n}\sum_{i=3}^{n}\frac{(i-1)(i-2)}{(n-1)(n-2)}x_{i:n} \tag{4}$$

$$b_3 = \frac{1}{n}\sum_{i=4}^{n}\frac{(i-1)(i-2)(i-3)}{(n-1)(n-2)(n-3)}x_{i:n} \tag{5}$$

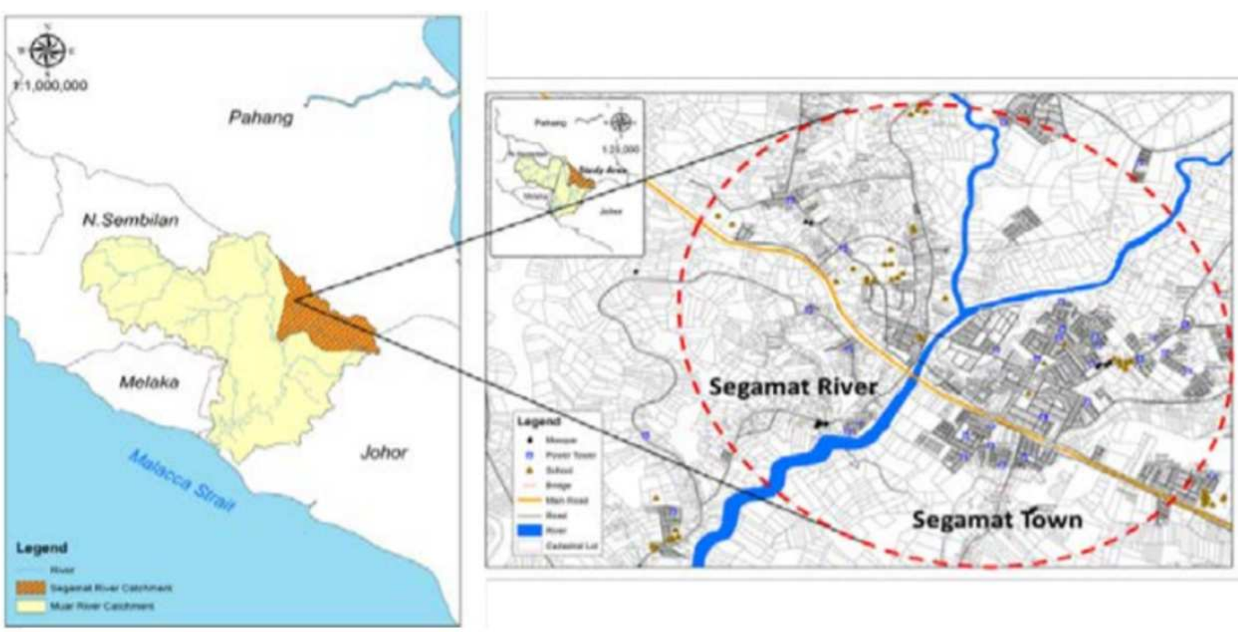

Figure 1: Location of Segamat River by Romali *et al.* (2018) [24]

The first four sample estimates for L-moments referred to as:

$$\boldsymbol{l}_1 = b_0 \tag{6}$$

$$l_2 = 2b_1 - b_0 \tag{7}$$

$$l_3 = 6b_2 - 6b_1 + b_0 \tag{8}$$

$$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0 \tag{9}$$

The samples of the L-moments ratio are addressed as follows:

$$t_2 = \frac{l_2}{l_1} \tag{10}$$

$$t_3 = \frac{l_3}{l_2} \tag{11}$$

$$t_4 = \frac{l_4}{l_2} \tag{12}$$

From the sample, L-moments, l_1 or known as L-location, describes the central value of the maximum stream flow data while L-scale (l_2) means the spread of the distribution. The larger the scale parameter, the more spread out the distribution. The coefficient variation (t_2) is more or equal to 1, indicating higher or lower variation. Meanwhile, L-skewness and L-kurtosis indicate the position of the distribution's tail and peak value of the distribution [30]. Table 2 describes the candidates of the probability distribution to fit the flood frequency data, which are GEV, GLO, GPA, PE3, and LN3. The cumulative density function and the parameter estimations of the distributions are explained in Table 2. From Table 2, $x(F)$ represents the quantile estimate and F is a corresponding probability of non-exceedance for a given return period T. F can be defined as $F = 1 - 1/T$. Meanwhile, $\hat{\xi}$, $\hat{\alpha}$ and $\hat{k}$ represent the location, scale and shape parameter of the candidate distributions.

2.3 Performance Measurement

This section is divided into three sub-sections: accuracy measure performance, L-moment Diagram (LMR), and goodness-of-fit test (GOF) to identify the optimum distribution at the Segamat River site.

2.3.1 Accuracy Measure Performance

There are five accuracy measure performance used: the mean absolute percentage error (MAPE), mean absolute error (MAE), coefficient of determination (R^2), root mean square error (RMSE) and root mean square percentage error (RMSPE), and the definition of each accuracy measurement is shown in equation (13) to equation (17).

Table 2: Estimated Distribution Parameters using the L-Moment Technique

Dist	Cumulative Density function	Parameter Estimation
GEV	$x(F)= \hat{\xi}+\frac{\hat{a}}{\hat{k}}\left\{1-(-\ln(F))^{\widehat{k}}\right\}$	$\widehat{k}=7.85890c+2.9554c^2$ where $c=\frac{2}{3+t_3}-\frac{\ln 2}{\ln 3}$ $\widehat{\alpha}=\frac{l_2}{\Gamma(\widehat{k})(1-2^{\widehat{k}})}$; $\hat{\xi}=l_1-\frac{\hat{\alpha}}{\widehat{k}}+\hat{\alpha}\Gamma(\widehat{k})$
GLO	$x(F)=\hat{\xi}+\frac{\hat{\alpha}}{\hat{k}}\left[1-\left\{\frac{(1-F)}{F}\right\}^{\hat{k}}\right]$	$\hat{k}=-t_3$; $\hat{\alpha}=\frac{l_2}{\Gamma(\hat{k})\left[\Gamma(1-\hat{k})-\Gamma(2-\hat{k})\right]}$ $\hat{\varepsilon}=l_1-\frac{\hat{\alpha}}{\hat{k}}+\hat{\alpha}\Gamma(\hat{k})\Gamma(1-\hat{k})$
GPA	$x(F)=\hat{\xi}+\frac{\hat{a}}{\hat{k}}\left\{1-[1-F]^{\hat{k}}\right\}$	$\hat{k}=\frac{1-3t_3}{1+t_3}$; $\hat{a}=l_2(\hat{k}+1)(\hat{k}+2)$ $\hat{\xi}=l_1-\frac{\hat{a}}{\hat{k}}+\frac{\hat{a}}{\hat{k}(\hat{k}+1)}$
LN3	$x(F)=\alpha+e^{\xi+uk}; u=\Phi^{-1}[1-\Phi]$	$\hat{k}=-0.001005+0.997386z+0.001027z^2$ $-0.005853z^3-0.000154z^4+0.000141z^5$ where $z=\sqrt{\frac{8}{3}}\Phi^{-1}\left[\frac{1+t_3}{2}\right]$, Φ^{-1} is an inverse CDF of Normal Distribution. $\alpha=\ln(l_2)-\ln\left(2S_1(k)-e^{\frac{k^2}{2}}\right)$; $\xi=l_1-e^{a+\frac{k^2}{2}}$
P3	$x(F)=\hat{\xi}\hat{k}+K_T\sqrt{\hat{\xi}^2\hat{k}}$ where $K_T=\frac{2}{C_s}\left[\left\{\frac{C_s}{6}\left(\Phi^{-1}[1-F]-\frac{C_s}{6}\right)+1\right\}^3-1\right]$ $C_s=\frac{2}{\sqrt{\hat{k}}}$	$\hat{k}=0.0127331632+\frac{1.0246130369}{3p(t_3)^2}-\frac{0.0024863669}{(3p(t_3)^2)^2}+\frac{0.0001169073}{(3p(t_3)^2)^3}-\frac{0.0000027751}{(3p(t_3)^2)^4}+\frac{0.0000000323}{(3p(t_3)^2)^5}-\frac{0.0000000001}{(3p(t_3)^2)^6}$ $\widehat{\alpha}=\frac{l_2}{2S_1(\hat{k})-\hat{k}}$; $\hat{\xi}=l_1-\hat{k}\hat{a}$ $S_r(k)=\int_0^\infty\left[\int_0^x\frac{1}{\Gamma(\hat{k})}t^{\hat{k}-1}e^{-t}dt\right]^r\frac{1}{\Gamma(\hat{k})}x^{\hat{k}}e^{-x}dx$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |F(y_i) - F(\widehat{y}_i)| \tag{13}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n} \left|\frac{F(y_i) - F(\hat{y}_i)}{F(y_i)}\right| \tag{14}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (F(y_i) - F(\hat{y}_i))^2}{n}} \tag{15}$$

$$RMSPE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(\frac{F(y_i) - F(\hat{y}_i)}{F(y_i)}\right)^2} \times 100 \tag{16}$$

$$R^2 = \frac{\sum_{i=1}^{n} (F(\hat{y}_i) - \bar{F}(y_i))^2}{\sum_{i=1}^{n} (F(\hat{y}_i) - \bar{F}(y_i))^2 + \sum_{i=1}^{n} (F(y_i) - F(\hat{y}_i))^2} \tag{17}$$

where $F(y_i)$ represents the actual stream flow data (observed data), n represent the total number of data, $\hat{F}(y_i)$ represents the average of the actual stream flow, and $F(\bar{y}_i)$ represents the estimated return period obtained from the selected distribution. The numerical accuracy performances such as MAE, MAPE, RMSE, and RMSPE, measure the difference between the actual observation of annual peak flow and theoretical estimates of peak flow from the distribution. The R^2 measures how well the theoretical estimated value obtained from distribution is able to fit the actual data. The larger value of R^2 indicates a better fit for the model.

2.3.2 L-Moment Ratio Diagram (LMR)

Hosking and Wallis [30] proposed L-Moment Ratio Diagram (LMR) in identifying an ideal distribution. For a three-parameter distribution, the LMR illustrates the theoretical relationship between t_3 (skewness) and t_4 (kurtosis), defined in Eq. 11 and Eq. 12, respectively. The value of t_3 is drawn from the Segamat river peak flow data and plot to LMR to recognise which distribution line it lies closely.

2.3.3 Goodness of Fit Test

To evaluate whether the proposed probability distributions are appropriate, suitable statistical techniques, such as the GOF tests may be used. The GOF tests can be utilised as the justification to select the most optimum distribution in FFA [31]. In this study, two GOF-tests, the Kolmogorov-Smirnov (KS) and Anderson Darling (AD) tests are used to examine how well the observed data represents the distributions. In the hydrological field, the KS test is commonly employed to investigate the suitability of the distribution [32]. Hypothesis testings were carried out to assess if the data follows the specific distributions, using KS and AD tests.

H_0: The data follow a specified distribution.

H_1: The data do not follow the specified distribution.

The p-value must be greater than $a = 0.05$ to accept the null hypothesis.

The equation of KS test is as follows;

$$\Delta = \max_i |F(x_i) - F_a(x_i)| \tag{18}$$

where $F(x_i)$ is the corresponding to the selected x_i distribution function, and $F_a(x_i)$ is the additive frequency distribution ordinal calculated from the observed sample [27]. The AD test is a statistical test that can detect deviations from normality in sample distributions. The AD test, in particular, converges quickly to the asymptote [33]. The formula for AD test is given by

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln(\ F(X_i)) + \ln(\ 1 - F(X_{n-i+1})\right] \tag{19}$$

3 Results and Discussion

Maximum flows are mainly linked to the engineering of flooding, the facilities of drainage systems, and the development of flood management measures. Table 1 presents the summarized descriptive statistics of the annual peak flow of Segamat gauged river site. The peak flow data of the Segamat river is measured in cubic meters per second. Table 1 shows that the observed flow data for Segamat River site is highly skewed, indicating that the data fits the non-normal distribution. The annual maximum series of daily stream flow in Segamat, Johor, Malaysia, is analysed and fitted to five non-normal distributions. This research implements five different distributions that are suitable for annual maximum flow in Peninsular Malaysia, namely GEV, GLO, GPA, PE3 and LN3. The parameters of the five probability distributions are obtained using the L-moment method. The estimated parameters for GEV, GLO, GPA, PE3 and LN3 distributions are reported in Table 3. The Gringorton unbiased plotting position formula plots the observed annual maximum flood data shown in Figure 2.

Figure 2 shows the cumulative distribution function for the candidate distribution, the GEV, GLO, GPA, PE3 and LN3 distributions. The CDF plot is used to compare the estimated peak flow from the candidate distribution with the actual data graphically. From the observations of the CDF plot, all the candidate distributions precisely represent the left tail and the central part. Usually, the right tail of the data is the key for any flood risk management project. From Figure 2, the observations show that the LN3 distribution is the closest in representing the right tail of the actual data.

Table 3 shows the estimated parameter for GEV, GLO, GPA, PE3, and LN3 distributions using the L-moment method. From this table, GEV, GLO, and GPA distributions exhibit negative values of the shape parameter through the indication of short-tailed or bounded distribution for GEV, GLO, and GPA distributions [34]. However, PE3 distribution produces the greatest absolute value of shape parameter and the highest value of scale parameter. These parameters correspond to greater maximum streamflow and the gradient of the maximum streamflow [34].

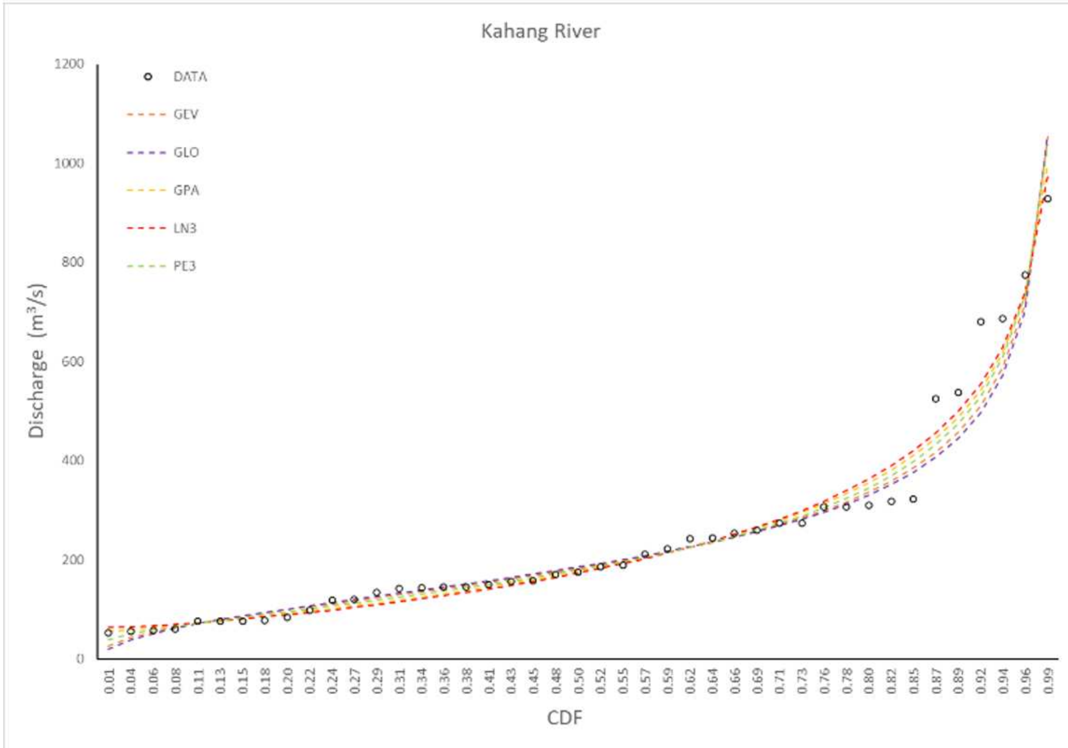

Figure 2: CDF plot for the Segamat Annual Peak Flow and the Candidate Distribution

Figure 3 shows a probability plot or QQ-plot between annual peak flow and the candidate distributions. The QQ-plot is a graphical method to compare the candidate distributions. Based on observation, the LN3 distribution lies close to the black line compared to other candidate distributions. In order to support our observation based on the graphical method, accuracy performance measure, goodness-of-fit test, and L-moment Diagram are added in the analysis. Evaluating the significance of the variations between the models solely based on the graphical display is tough to comprehend or identify. Therefore, in this study, the optimum model is identified using three types of measurement: the GOF test, numerical accuracy measure, and L-moment ratio diagram. Table 4 shows the performance measurement for candidate distributions. The performance measurements used are the KS test, AD test, RMSE,

Table 3: Estimated Distribution Parameters using the L-Moment Technique

Parameters			
Dist.	$\hat{\alpha}$	$\widehat{\xi}$	$\hat{k}$
GEV	108.64	86.49	-0.47
GLO	145.25	74.24	-0.51
GPA	37.33	126.30	-0.35
PE3	234.58	280.83	3.18
LN3	25.78	4.70	1.13

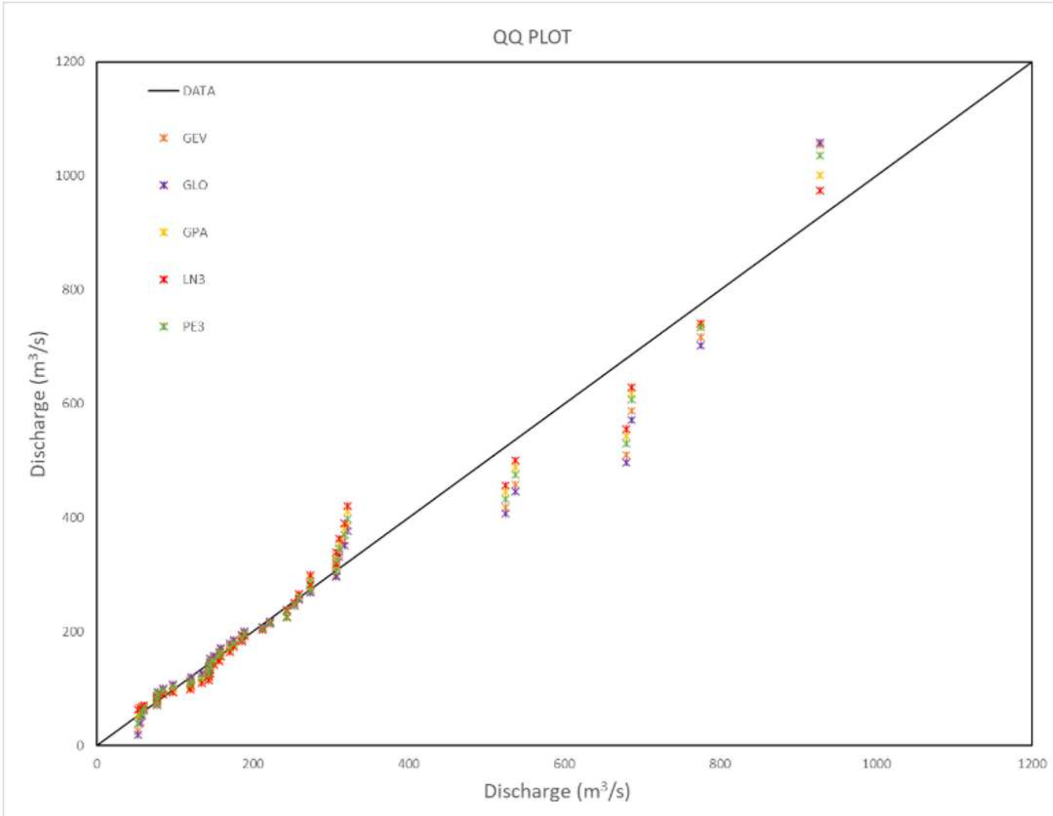

Figure 3: QQ-Plot for Segamat Annual Peak Flow and Candidate Distribution

MAPE, RMSPE, R^2, and LMR. The KS and AD test is measured as p-value. From Table 4, we can see that the p-values of the KS and AD tests are greater than 0.05. Thus null hypothesis in Section 2.3.3 is accepted for all distributions. We chose the distribution with the highest p-value to fit the data as the higher significant p-value indicates that the distribution fits the actual data better. RMSE, MAPE, RMSEPE, and MAE are relative accuracy measures, whereby the lowest value for each measurement better fits the estimation from the candidate distribution to the actual data. From Table 4, LN3 distribution is the best-fitted distribution when using KS test, AD test, MAPE, RMSEPE and LMR. PE3 is the best-fitted distribution when using RMSE, MAE and R^2.

For the purpose of fair evaluation, the rank score approach is implemented in this study to identify the best-fitted distribution at the target site. This approach required giving each distribution a score based on how well the distribution fits the actual observation. The best-fitted distribution is given a score of 5, and the least fit distribution is given a score of 1. The GOF test of distributions was measured with p-value. The higher p-value indicates that the distribution fit the actual data well. Sample estimates of the dimensionless ratios are compared in the LMD. The LMD is shown in Figure 4. The distribution with the smallest RMSE, smallest MAE, smallest RMSPE, smallest MAEP and the highest R^2 is given the highest score value of 5. The distribution with the highest RMSE, highest MAE, highest RMSPE, highest MAEP and the lowest R^2 is given the lowest score value of 1. Table 5 shows the performance measurement for candidate distributions. Table 5 presents the score result for each distribution selected in this study.

Table 4: P-value of GOF Test and Performance Measurement for Candidate Distributions

Candidate Distribution	*GEV*	*GLO*	*GPA*	*PE3*	***LN3***
KS	0.8167	0.8165	0.9865	0.6705	**0.9866**
AD	0.8188	0.7301	0.9544	0.6362	**0.9844**
RMSE	51.4190	54.8169	40.0422	**31.9708**	39.2965
MAPE	0.1222	0.1341	0.0879	0.1091	**0.0798**
MAE	27.7876	29.6819	21.0709	**19.9644**	20.4559
RMSPE	0.2592	0.2734	0.1156	0.1372	**0.1075**
R^2	0.9840	0.9824	0.9901	**0.9934**	0.9899
LMR	0.0928	0.0847	0.0389	0.0402	**0.0350**

Table 5: Rank Score for Candidate Distribution

Candidate Distribution	*GEV*	*GLO*	*GPA*	*PE3*	***LN3***
KS	3	2	4	1	**5**
AD	3	2	4	1	**5**
RMSE	2	1	3	5	**4**
MAPE	2	1	4	3	**5**
MAE	2	1	3	5	**4**
RMSPE	2	1	4	3	**5**
R^2	2	1	4	5	**3**
LMR	1	2	4	3	**5**
Total Score	17	11	30	26	**36**

The outcomes in Tables 4 and 5 reveal that the LN3 distribution is the optimum distribution to represent the Segamat River peak flow data. After assessing each of the distributions individually, the data Table 5 suggests that the GLO distribution is not suitable, as it has the lowest rank score. The LN3 distribution has a better total rank score compared to the other distributions used in this study. Figure 2 shows the CDF plot curve for the Segamat River site. The CDF plots show that most distributions precisely represent the left tail and central part. The right tail is the key because it is the most crucial in water resource design and planning. GEV, GLO, GLPA, and P3 frequency models did a mediocre job describing the right-tail parts. The frequency mechanism of the data influences the selection of the optimum model for FFA at a location with similar physical and hydrological characteristics. These results indicate that it is difficult to propose a specific probability distribution for the annual peak flow of a river in Johor because each river has a unique data series. The peak flow data rapidly changes from year to year due to weather uncertainty caused by climate change. The objective of choosing the optimum model for each site is to estimate extreme stream flow with different return periods. As a result, various stream flow return periods are estimated using an adequate frequency model.

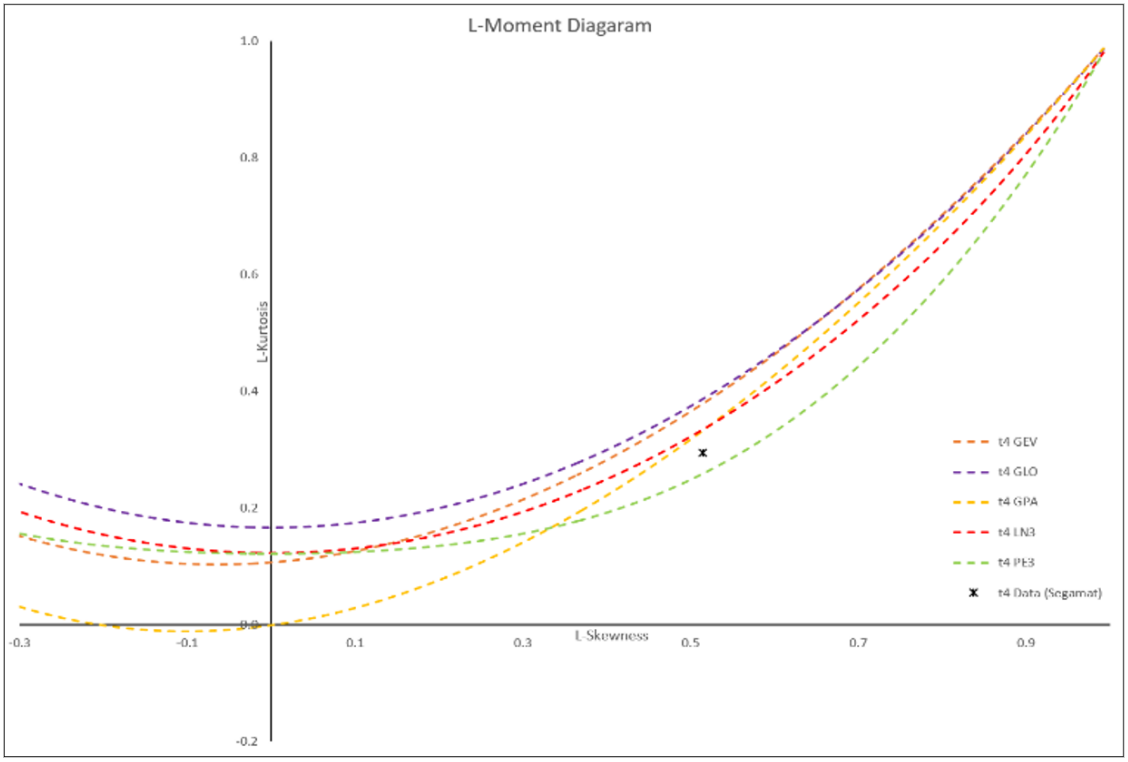

Figure 4: L-Moment Diagram

Table 6 shows the estimated return period for all distributions. From the table, the LN3 distribution produces a large value of estimated flood discharge. These values indicate the probability of having a flood magnitude for T = 2, 10, 25, 50, 100, and 200 years. The estimated flows for 50, 100, and 200 years for the Segamat River site on LN3 distribution are 1149.1211 m^3/s, 1554.535 m^3/s, and 2052.6002 m^3/s, respectively.

Table 6: Estimated Flood Discharge for the Segamat River Site

Return Period (Years)	**Estimated Flood Discharge (m^3/s)**				
	GEV	**GLO**	**GPA**	**PE3**	**LN3**
2	143.2785	145.2569	136.7669	121.1037	136.0053
10	457.2904	448.0712	490.0393	557.4518	495.0469
25	759.7659	741.8052	803.8349	874.0756	823.2788
50	1090.9282	1070.8430	1120.2796	1128.5833	1149.1211
100	1549.7497	1537.7372	1526.3191	1391.7534	1554.5351
200	2186.6812	2202.5991	2047.3204	1661.3290	2052.6002

4 Conclusion

FFA is a well-known hydrologic engineering method that has attracted attention from researchers. This study aims to find the best candidate distribution for FFA to portray the annual peak flow variable of a river site of interest. The GOF test is used to assess the accuracy of five three-parameter probability distributions in this study which are P3, GLO, GPA, LN3, and GEV. The most widely used parameter estimation method, the L-moment approach, is used to obtain the parameter for each distribution used in this paper. The annual maximum series data, extracted from the historical daily streamflow record of the Segamat River site in Johor, Malaysia, is used for testing. The models' assessment was conducted using a goodness of fit test, LMR, and numerical performance criteria. The data's fitness assessment using various measurement tools ensures that the selected distribution is robust and accurate to estimate the return period. The best-fit function for annual peak flows for the Segamat River site is the LN3 distribution. The distribution may be implemented for regional FFA or in-situ FFA of several other rivers located in Johor for further studies.

References

[1] Government of Malaysia Department Of Irrigation And Drainage. *Flood Management (1)*. 2009.

[2] Rahman, A. S., A. Rahman, M. A. Zaman, K. Haddad, A. Ahsan and M. Imteaz. A study on selection of probability distributions for at-site flood frequency analysis in Australia. *Natural Hazards*. 2013. 69(3):1803–1813.
DOI: https://doi.org/10.1007/s11069-013-0775-y

[3] Selaman, O. S., S. Said and F. Putuhena. Flood frequency analysis for Sarawak using Weibull, Gringorten and L-moments formula. *Journal-The Institution of Engineers*. 2007. 68(1): 43–52.

[4] Romali, N. S. and Z. Yusop. Frequency analysis of annual maximum flood for Segamat River. *MATEC Web of Conferences*. 2017. 103 (04003): 1–9.
DOI : https://doi.org/10.1051/matecconf/201710304003

[5] Jan, N. A. M., A. Shabri, J. Hounkpè and B. Badyalina. Modelling non-stationary extreme streamflow in Peninsular Malaysia. *International Journal of Water*. 2018. 12(2): 116-140.
DOI : https://doi.org/10.1504/IJW.2018.091380

[6] Singo, L., P. Kundu, J. Odiyo, F. Mathivha, and T. Nkuna. Flood frequency analysis of annual maximum stream flows for Luvuvhu River Catchment, Limpopo Province, South Africa. *University of Venda, Department of Hydrology and Water Resources Thohoyandou, South Africa*. 2012. 1–9.

[7] Jan, N. A. M., A. Shabri, S. Ismail, B. Badyalina, S. S. Abadan and N. Yusof. Three-Parameter Log-normal Distribution: Parametric Estimation Using L-Moment And Tl-Moment Approach. 2016. *Jurnal Teknologi*. 78(6–11).
DOI: https://doi.org/10.11113/jt.v78.9202

[8] Badyalina, B., and A. Shabri. Streamflow Forecasting at Ungauged Sites using Multiple Linear Regression. *MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics*. 2013. 29 (1b): 67–75.
DOI: https://doi.org/10.11113/matematika.v29.n.360

[9] Mokhtar, N. A., B. Badyalina, K. L. Chang, F. F. Yaa'cob, A. F. Ghazali, and P. Shamala. Error-in-Variables Model of Malacca Wind Direction Data with the von Mises Distribution in Southwest Monsoon. *Applied Mathematical Sciences*. 2021. 15(9): 471–479.

DOI: https://doi.org/10.12988/ams.2021.914521

[10] Mokhtar, N. A., Y. Z. Zubairi, A. G. Hussin, and R. M. Yunus. On parameter estimation of a replicated linear functional relationship model for circular variables. *MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics*. 2017. 33(2): 159–163.
DOI: https://doi.org/10.11113/matematika.v33.n2.1010

[11] Jan, N. A. M., A. Shabri, B. Badyalina. Selecting Probability Distribution for Regions of Peninsular Malaysia Streamflow. *AIP Conference Proceedings*. 2016. 1750 (060014): 1–6.
DOI: https://doi.org/10.1063/1.4954619

[12] Peel, M. C., Q. Wang, R. M. Vogel, and T. A. McMahon. The utility of L-moment ratio diagrams for selecting a regional probability distribution. *Hydrological Sciences Journal*. 2001. 46(1): 147–155.
DOI : https://doi.org/10.1080/02626660109492806

[13] Hosking, J., and J. Wallis. Some statistics useful in regional frequency analysis. *Water resources research*. 1993. 29(2): 271–281.
DOI : https://doi.org/10.1029/92WR01980

[14] Drissia, T., V. Jothiprakash, and A. Anitha. Flood frequency analysis using L moments: a comparison between at-site and regional approach. *Water Resources Management*. 2019. 33(3): 1013–1037.
DOI : https://doi.org/10.1007/s11269-018-2162-7

[15] Hassan, W. H. W., A. Z. Jidin, S. A. C. Aziz, and N. Rahim. Flood disaster indicator of water level monitoring system. *International Journal of Electrical and Computer Engineering*. 2019. 9(2): 1694–1699.
DOI : 10.11591/ijece.v9i3.pp1694-1699

[16] Hailegeorgis, T. T., and K. Alfredsen. Regional flood frequency analysis and prediction in ungauged basins including estimation of major uncertainties for mid-Norway. *Journal of Hydrology: Regional Studies*. 2017. 9: 104–126.
DOI : https://doi.org/10.1016/j.ejrh.2016.11.004

[17] Daud, Z. M., A. M. Kassim, M. M. Desa, and V. T. V. Nguyen. Statistical analysis of at-site extreme rainfall processes in peninsular Malaysia. *International Association of Hydrological Sciences, Publication*. 2002. 274: 61–68.

[18] Lim, Y. H., and L. M. Lye. Regional flood estimation for ungauged basins in Sarawak, Malaysia. *Hydrological Sciences Journal*. 2003. 48(1): 79–94.
DOI : https://doi.org/10.1623/hysj.48.1.79.43477

[19] Kizza, M., H. Ntale, A. Rugumayo and M. Kigobe. Regional flood frequency analysis for Northern Uganda Using the L-Moment approach. *International Conference on Advances in Engineering and Technology*. 2006. 44(3): 387–398.
DOI : https://doi.org/10.1016/B978-008045312-5/50027-3

[20] Badyalina, B., and A. Shabri. Flood estimation at ungauged sites using group method of data handling in Peninsular Malaysia. *Jurnal Teknologi*. 2015. 76(1): 373–380.
DOI : https://doi.org/10.11113/jt.v76.2640

[21] Liang, Y., S. Liu, Y. Guo and Hua, H. L-moment-based regional frequency analysis of annual extreme precipitation and its uncertainty analysis. *Water Resources Management*. 2017. 31(7): 3899–3919.
DOI : https://doi.org/10.1007/s11269-017-1715-5

[22] Hamzah, F. M., S. H. M. Yusoff and O. Jaafar. L-moment-based frequency analysis of high-flow at Sungai Langat, Kajang, Selangor, Malaysia. *Sains Malaysiana*. 2019. 48(7): 1357–1366.
DOI : http://dx.doi.org/10.17576/jsm-2019-4807-05

[23] Leščešen, I., M. Urošev, D. Dolinaj, M. Pantelić, T. Telbisz, G. Varga, and D. Milošević. Regional flood frequency analysis based on L-moment approach (case study tisza river basin). *Water Resources*. 2019. 46(6): 853–860.
DOI : https://doi.org/10.1134/S009780781906006X

[24] Das, S. Assessing the regional concept with sub-sampling approach to identify probability distribution for at-site hydrological frequency analysis. *Water Resources Management*. 2010. 34(2): 803–817.
DOI : https://doi.org/10.1007/s11269-019-02475-6

[25] Shabri, A., and N. A. M. Ariff. Frequency analysis of maximum daily rainfalls via l-moment approach. *Sains Malaysiana*. 2009. 38(2): 149–158.

[26] Romali, N. S., Z. Yusop, A. Z. Ismail. Application of HEC-RAS and Arc GIS for floodplain mapping in Segamat town, Malaysia. *International Journal*. 2018. 14(43): 125–131.
DOI : https://doi.org/10.21660/2018.47.3656

[27] Jaafar, J. B., A. N. B. Ishak, S. B. Hassan, K. F. Bin, and M. I. Q. Adrutdin. A study of customer satisfaction with planning movement of goods during disaster aid programs: A case study of flood hit in Segamat, Johor. 2010 *Journal of Environmental Treatment Techniques*. 8(1): 419–428.

[28] Hosking, J. R. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. 1990. *Journal of the Royal Statistical Society: Series B (Methodological)*. 52(1): 105–124.
DOI : https://doi.org/10.1111/j.2517-6161.1990.tb01775.x

[29] Landwehr, J. M., N. Matalas, and J. R. Wallis. Some comparisons of flood statistics in real and log space. *Water Resources Research*. 1978. 14(5): 902–920.
DOI: https://doi.org/10.1029/WR014i005p00902

[30] Hosking, J. R. M., and Wallis, J. R. *Regional frequency analysis: an approach based on L-moments*. 2005. Cambridge University Press. UK.

[31] Millington, N., S. Das, and S. P. Simonovic. *The comparison of GEV, log-Pearson type 3 and Gumbel distributions in the Upper Thames River watershed under global climate models.* 2011. Water Resources Research Report. 077. Department of Civil and Environmental Engineering, University of Western Ontario London, Ontario, Canada.
DOI: http://ir.lib.uwo.ca/wrrr/40.

[32] Sandalci, M. Flood Frequency Analysis of Akçay Stream. *Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi*. 2018. 22(5): 1309–1313.
DOI: https://doi.org/10.16984/saufenbilder.402190

[33] Dey, S., E. Raheem, S. Mukherjee, and H. K. T. Ng. Two parameter exponentiated Gumbel distribution: properties and estimation with flood data example. 2017. *Journal of Statistics and Management Systems*. 20(2): 197–233.
DOI: https://doi.org/10.1080/09720510.2016.1228261

[34] Hasan, H., N. Salam, and M. B. Adam. Modelling extreme temperature in Malaysia using generalised extreme value distribution. *International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering*. 2013. 7(6): 618–624.
DOI: https://doi.org/10.1063/1.4801267