Analysis of annual maximum and partial duration rainfall series

Jing Lin Ng1, Soon Kim Tiang1, Yuk Feng Huang2, Nur Ilya Farhana Md Noh1 and Ramez A Al-Mansob1

1Department of Civil Engineering, Faculty of Engineering, Technology and Built Environment, UCSI University, 56000 Cheras, Kuala Lumpur, Malaysia.
2Department of Civil Engineering, Lee Kong Chian, Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Selangor, Malaysia

Email: ngjl@ucsiuniversity.edu.my

Abstract. Kelantan River Basin is affected by two significant monsoon seasons, namely the Northeast and Southwest monsoons that lead to flood and heavy downpour events. Consequently, analysis of rainfall series is gaining more attention from researchers. The aim of this study is to analyse the annual maximum series (AMS) and partial duration series (PDS) by fitting different probability distributions. Generalized Extreme Value (GEV), Generalized Pareto (GP), Log Pearson Type 3 (LP3), Log Normal (LN) and Log Normal 3 (LN3) were used in this study. The performances of these probability distributions were evaluated using different goodness-of-fit tests, namely the chi-square ($\chi^2$), Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests. Subsequently, the performances of probability distributions were compared and the best fit probability distribution was selected. The GEV and GP distributions were selected as the best fit probability distributions for AMS and PDS, respectively. The findings can provide useful information for flood mitigation and water resources management.

1. Introduction
Flooding is natural disaster that brings many bad impacts to the population and society. In order to address this issue, hydrological frequency analysis that is associated with deducing the probability of occurrence of flood events is carried out. Frequency analysis is very important for the design of hydraulic structures and hydrological planning [1]. The fundamental types of time series are annual maximum series (AMS) and partial duration series (PDS) [2]. AMS fits the probability distribution to one of the highest rainfall data for each year so that the number of selected data is equivalent to the number of years in a given period. In comparison, PDS handles the selected data which exceed a predefined base value in a given period. PDS is often known as the Peaks Over Threshold (POT) series due to its sampling method. A threshold is predetermined during the selection of the rainfall data and all the peaks exceeding the threshold are selected.

Both AMS and PDS have been fitted by different types of probability distributions. Many countries, including Malaysia have developed guidelines suggesting the most suitable distribution to be applied. Department of Irrigation and Drainage (DID) Malaysia revealed that the most appropriate frequency distribution for AMS is Gumbel distribution or Generalized Extreme Value (GEV) whereas Generalized Pareto (GP) or Exponential distribution is the most suitable for PDS [3]. This is similar to the finding of [4] who conducted an extreme rainfall frequency analysis using AMS. GEV distribution was the only
distribution model which fits most of the extreme value series. Apart from that, a comprehensive study on flood frequency analysis based on a large Australian annual maximum flood data set was investigated. The results revealed that the LP3, GEV and GP distributions were rated as the top three best-fit distributions [5]. Moreover, [6] carried out a low magnitude flood frequency analysis using AMS and PDS in the Great Barrier Reef (GBR) lagoon catchments, North-eastern Australia. The result depicted that GEV, GP and LPS demonstrated relatively similar results between AMS and PDS during large magnitude of flood.

From the historical overview, Kelantan was one of the states in Malaysia that encountered numerous disastrous flood events caused by the unexpected rainfall pattern. Kota Bahru is the capital of Kelantan which is prone to flood. This is attributed to the occurrence of Northeast (NE) monsoon that delivers significant large amount of rainfall [7,8]. Consequently, the safety of Kelantan residents’ is at risk and greater damages on properties are noticed [9,10]. This issue has highlighted the urge to analyse the AMS and PDS to manage the annual flood events. Accordingly, this study aims to evaluate the performance of different probability distributions in fitting the AMS and PDS and select the best fit probability distribution in Kelantan River Basin. The outcomes from this study are expected to benefit the researchers or design engineers in providing them an in-depth analysis on different types of probability distributions for AMS and PDS.

2. Study area

Kelantan state is located at the east coast of Peninsular Malaysia. Kelantan River Basin covers 80% of the Kelantan state and it is located between 4° 40’ to 6° 12’ North, and 101° 20’ to 102° 20’ East. Agricultural activities such as oil palm plantations, rubber estates and paddy farms play an important part in the economy of Kelantan state [11,12]. Unfortunately, Kelantan River Basin is severely affected by flood events annually attributed by the Northeast (NE) Monsoon which takes place between November and February. The hydrological data for the period of 1973-2013 were retrieved from the Malaysian Meteorological Department (MMD). It can be highlighted that only rainfall data with less than 10% missing data were considered for subsequent analysis. Figure 1 shows the rainfall station distribution map of Kelantan River Basin and the study area details are shown in table 1.

Figure 1. Rainfall station distribution map.
Table 1. Study area details.

| Station code | Station name          | Record Period | Duration | Latitude      | Longitude     |
|--------------|-----------------------|---------------|----------|---------------|---------------|
| 40431        | Pos Blau              | 1978-2013     | 36       | 04° 39' N    | 101° 41' E   |
| 40502        | Pos Bihai             | 1978-2013     | 36       | 05° 00' N    | 101° 34' E   |
| 40510        | Pos Tehoi             | 1978-2013     | 36       | 05° 03' N    | 101° 45' E   |
| 40512        | Pos Wias              | 1978-2013     | 36       | 05° 07' N    | 101° 49' E   |
| 40516        | Pos Gob               | 1978-2013     | 36       | 05° 17' N    | 101° 38' E   |
| 40547        | Mardi Jeram Pasu      | 1984-2013     | 30       | 05° 48' 46" N| 102° 20' 40" E|
| 40546        | Pusat Ternakan Haiwan | 1980-2013     | 34       | 5° 48' 40" N| 102° 00' 33" E|
| 40664        | Pusat Pertanian       | 1974-2013     | 40       | 06° 06' N    | 02° 14' E    |
| 48615        | Kota Bharu            | 1954-2013     | 60       | 06° 10' N    | 102° 18' E   |
| 48616        | Kuala Krai            | 1986-2013     | 28       | 05° 32' N    | 102° 12' E   |

3. Methods

3.1. Homogeneity test

A total of four homogeneity tests, namely the standard normal homogeneity test (SNHT), Buishand Range (BR) test, Pettitt test and Von-Neumann Ratio test (VNR) were used to detect the inhomogeneity of rainfall. The results of all methods were evaluated at 5% significant level. Therefore, any tests of p-value fall under 0.05 will be considered as inhomogeneous.

3.1.1. Standard Normal Homogeneity Test (SNHT). The standard normal homogeneity test is employed to detect the inhomogeneity by pinpointing the break near the starting and ending of the time series. The statistic, $T_k$, is expressed as:

$$T_k = k \bar{z}_1^2 + (n - k) \bar{z}_2^2, \quad k = 1, 2, 3, \ldots, n$$  \hspace{1cm} (1)

where

$$\bar{z}_1 = \frac{1}{k} \sum_{i=1}^{k} (Y_i - \bar{Y}) / s$$  \hspace{1cm} (2)

and

$$\bar{z}_2 = \frac{1}{n - k} \sum_{i=k+1}^{n} (Y_i - \bar{Y}) / s$$  \hspace{1cm} (3)

$Y_i$ represents the time series to be tested, $\bar{Y}$ is the mean value of the time series and s represents standard deviation. $T_k$ will reach a maximum value near the year $k = K$ if a break takes place in the year $k$.

3.1.2. Buishand Range test (BR). The Buishand range test discovers the break in the midpoint of the time series. The adjusted partial sum is defined as:

$$S_0^* = 0$$  \hspace{1cm} (4)
\[ S_0^* = \sum_{i=1}^{k} (Y_i - \bar{Y}), k = 1,2,3,...,n \]  

\( Y_i \) represents the time series to be tested and \( \bar{Y} \) is the mean value of the time series. The value of \( S_0^* \) will vary around zero in the case where time series is classified as homogenous.

3.1.3. Pettitt test. The Pettitt test is a non-parametric homogeneity test which detects the inhomogeneity on the basis of data ranking. The statistic is defined as:

\[ X_k = 2 \sum_{i=1}^{k} r_i - k(n + 1), k = 1,2,3,...,n \]

where \( r_1, r_2, r_3, ..., r_n \) represent the ranks of the time series in ascending order. The statistic value will reach its maximum or minimum value near to the year \( k = E \).

3.1.4. Von-Neumann Ratio test. The Von-Neumann Ratio test evaluates whether the time series is randomly distributed. The statistic is expressed as follows:

\[ N = \frac{\sum_{i=1}^{n-1}(Y_i - Y_{i+1})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \]

The \( N \) value is equal to 2 when the time series is homogeneous. If the \( N \) value is less than 2, there is a break within the time series.

3.2. Extraction of data using AMS and PDS

AMS (Annual Maximum Series) and PDS were used in this study. Annual maximum series (AMS) only selected the highest peaks for each year, therefore, the number of selected data was equivalent to the number of years. For partial duration series (PDS), all the peaks exceeding the threshold were selected and the threshold values were predetermined during the selection of the rainfall data.

3.3. Parameter estimation

3.3.1. Maximum Likelihood Method. In this study, the parameter estimation method used for fitting distributions to data is Maximum Likelihood Method (MLM). MLM provides an upstanding representation of the relationship between parameter values and sample data. The likelihood function can be expressed as:

\[ ML(\varepsilon) = \Pi_{t=1}^{n} f(\varepsilon_t, \alpha_1, ..., \alpha_m) \]

Maximizing the function \( ML(\varepsilon) \), maximum likelihood is obtained,

\[ \log (ML(\varepsilon)) = \log \Pi_{t=1}^{n} f(\varepsilon_t, \alpha_1, ..., \alpha_m) = \sum_{t=1}^{n} \log[f(\varepsilon_t, \alpha_1, ..., \alpha_m)] \]

Partial Derivatives should be equated to 0,

\[ \frac{\partial \log (ML(\varepsilon))}{\partial \alpha_1} = 0; \frac{\partial \log (ML(\varepsilon))}{\partial \alpha_m} = 0 \]

where \( f(\varepsilon_t, \alpha_1, ..., \alpha_m) \) is the probability distribution function of the sample.
3.4 Probability distribution fitting

3.4.1. Generalized Extreme Value distribution. The GEV distribution is a combination from all three types of EV distribution which consist of three parameters, which are the location, scale and shape. The probability density function (PDF) of GEV distribution is defined as:

\[ Q(p; \mu, \sigma, \xi) = \mu + \frac{\sigma((- \log(p))^{-\xi} - 1)}{\xi}; \]

if \( \xi > 0 \) and \( p \in [0,1) \); \( \xi < 0 \) and \( p \in (0,1) \)

If \( \xi = 0 \), \( p \in (0,1) \)

where \( \mu \) is the location parameter, \( \sigma \) is the scale parameter and \( \xi \) is the shape parameter while the quantile density function of \( Q = \frac{dq}{dp} \) is:

\[ q(p; \sigma, \xi) = \frac{\sigma}{p(- \log(p))^{\xi + 1}} \]

3.4.2. Generalized Pareto distribution. The Generalized Pareto (GP) distribution fits best to the sample data with high frequencies due to its characteristics of long and thick upper tail. The PDF of GP is illustrated as:

\[ f(x) = \frac{1}{\sigma} (1 + \xi \frac{(x - \mu)}{\sigma})^{-1 - \frac{1}{\xi}} \]

where \( \mu \) is the location parameter, \( \sigma \) is the scale parameter and \( \xi \) is the shape parameter.

3.4.3. Log-Pearson Type 3 distribution. Log-Pearson Type 3 (LP3) distribution is a three-parameter Gamma distribution. The purpose of applying transformed variable log x to P3 is to generate another reduced skewness distribution. The general equation of LP3 is as follows:

\[ \log Q_T = \log Q^- + K(C_s, T)S_{\log Q} \]

where \( Q_T \) is estimated rainfall time series, \( K \) is frequency factor, \( C_s \) is skew coefficient and \( S_{\log Q} \) represents the standard deviation of logarithm of estimated rainfall time series.

3.4.4. Log-Normal distribution. Log-Normal (LN) Distribution is a spinoff from normal distribution. LN derived its probability density function from normal distribution and it is more commonly used when the random variables cannot be less than 0. The PDF of LN is as follows:

\[ f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp \left[ -\frac{(\log x - \bar{x})^2}{2\sigma^2} \right] \]

where \( \sigma \) and \( \bar{x} \) are the standard deviation and mean of the logarithmic transformed rainfall amount, respectively.

3.4.5. Log-Normal 3 distribution. The application of Log-Normal 3 (LP3) distribution is similar with LN but the only difference was the number of parameters used when fitting to data. The additional parameter added to LN3 is location parameter. The PDF of LN3 is as follows:
\[
f(x) = \frac{1}{(x - y)\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln(x - y) - \mu}{\sigma}\right)^2\right) \quad \text{for } x > 0
\]

where \( \mu \) is the location parameter and \( \sigma \) is the scale parameter.

3.5. Evaluation of performances

Goodness-of-fit (GoF) tests were used to evaluate the performances of different distributions. The implementation of GoF test was to compute the gap between the fitted parametric distributions along with empirical distributions. In this study, three types of GOF tests were selected, namely the chi-square \((\chi^2)\), Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests. The equation of \( \chi^2 \), KS and AD test are defined as follows:

\[
\chi^2 = \sum_{i=1}^{N} \frac{(O(i) - E(i))^2}{E(i)}
\]

\[
KS = \max_{1 \leq i \leq n} \left( F(x_i) - \frac{i - 1}{n}, i \right)
\]

\[
AD = n \int_{-\infty}^{\infty} \frac{[F_N(x) - F(x)]^2}{F(x)\{1 - F(x)\}} dF(x)
\]

where \( N \) is the number of class intervals and \( O(i) \) is the observed frequency of the \( i^{th} \) histogram class while \( E(i) \) is the expected frequency for \( \chi^2 \) test and \( F_N(x) \) and \( F(x) \) are the sample and hypothesis distribution function for AD test.

3.6. Identification of the best fit probability distribution

The scores of each goodness of fit tests were summed up and the best fit probability distribution was selected based on the highest total scores. The results were ranked from 5 (best fit) to 0 (least fit) for the probability distributions. Apart from that, the comparison results of all probability distributions will be illustrated using Q-Q plot. The best distribution will be determined by computing the gap between fitted parametric distributions along with empirical distribution.

4. Results and discussion

4.1. Homogeneity test

For homogeneity tests, the rainfall data obtained from the MMD were aggregated into monthly time scale and yearly time scale. Then, the rainfall data in the respective time scales are validated using BRT, Pettitt test, SNHT and VNR. The rainfall data are said to be homogeneous if the null hypothesis is accepted with a p-value larger than 0.05 at the significance level of 5%. The results of the homogeneity for monthly time scale, yearly time scale and a summary on the classification of the rainfall stations are shown in table 2, table 3 and table 4, respectively. For the monthly time scale, there were 8 out of 10 rainfall stations were identified as useful and 2 out of 10 rainfall stations were identified as suspect. Figure 2 shows the test statistics for the homogeneity tests at Pos Blau rainfall station. It is noticeable that there were breaks in the monthly rainfall data during February 1999, March 1999 and July 1999 for Pettitt test, SNHT and BRT, respectively. It was found that Pos Gob and Pos Blau rainfall stations
rejected three and all four null hypotheses, respectively as shown in table 2. Therefore, these two rainfall stations contain heterogeneity in the rainfall data and shall not be included in any further analysis. However, there were no doubtful rainfall stations in the monthly time scale. Besides that, it is to note that all the rainfall stations rejected the VNR for the monthly time scale. Since VNR evaluates the time series as randomly distributed, the results had shown the exact opposite where the monthly rainfall data were normally distributed.

Table 2. Summary of homogeneity tests for monthly time series.

| Station Name                  | Station ID | Pettit | SNHT | BRT  | VNR test | Results  |
|------------------------------|------------|--------|------|------|----------|----------|
| MARDI JERAM PASU             | 40547      | 0.880  | 0.391| 0.636| < 0.0001 | Useful   |
| POS GOB                      | 40516      | 0.066  | 0.036| 0.038| < 0.0001 | Suspect  |
| POS BIHAI                    | 40502      | 0.762  | 0.530| 0.397| 0.000    | Useful   |
| POS BLAU                     | 40431      | 0.004  | 0.031| 0.001| < 0.0001 | Suspect  |
| PUSAT PERTANIAN LUNDANG      | 40664      | 0.464  | 0.297| 0.769| < 0.0001 | Useful   |
| P. TER. HAIWAN TANAH MERAH  | 40546      | 0.199  | 0.335| 0.090| 0.000    | Useful   |
| POS TEHOI                    | 40510      | 0.553  | 0.486| 0.198| < 0.0001 | Useful   |
| POS WIAS                     | 40512      | 0.390  | 0.489| 0.350| < 0.0001 | Useful   |
| KUALA KRAI                   | 48616      | 0.193  | 0.552| 0.195| 0.003    | Useful   |
| KOTA BHARU                   | 48615      | 0.231  | 0.180| 0.346| < 0.0001 | Useful   |

*Bolded value represents the heterogeneous time series

Table 3. Summary of homogeneity tests for yearly time series.

| Station Name                  | Station ID | Pettit | SNHT | BRT  | VNR test | Results  |
|------------------------------|------------|--------|------|------|----------|----------|
| MARDI JERAM PASU             | 40547      | 0.355  | 0.758| 0.538| 0.721    | Useful   |
| POS GOB                      | 40516      | 0.127  | 0.113| 0.158| 0.409    | Useful   |
| POS BIHAI                    | 40502      | 0.878  | 0.790| 0.616| 0.166    | Useful   |
| POS BLAU                     | 40431      | 0.042  | 0.020| 0.012| 0.035    | Suspect  |
| PUSAT PERTANIAN LUNDANG      | 40664      | 0.545  | 0.654| 0.722| 0.339    | Useful   |
| P. TER. HAIWAN TANAH MERAH  | 40546      | 0.145  | 0.104| 0.054| 0.484    | Useful   |
| POS TEHOI                    | 40510      | 0.280  | 0.718| 0.944| 0.868    | Useful   |
| POS WIAS                     | 40512      | 0.984  | 0.346| 0.325| 0.581    | Useful   |
| KUALA KRAI                   | 48616      | 0.113  | 0.126| 0.056| 0.144    | Useful   |
| KOTA BHARU                   | 48615      | 0.271  | 0.036| 0.144| 0.035    | Doubtful |

*Bolded value represents the heterogeneous time series

Table 4. Summary of homogeneity tests result for monthly and yearly time series.

| Class           | Monthly | Yearly |
|-----------------|---------|--------|
| Useful          | 8(80%)  | 8(80%) |
| Doubtful        | 0(0%)   | 1(10%) |
| Suspect         | 2(20%)  | 1(10%) |
Figure 2. Test statistic of Pettitt test, SNHT and BR test of monthly rainfall at Pos Blau station.

For the yearly time scale, there were 8 out of 10 rainfall stations identified as useful, 1 out of 10 rainfall stations identified as doubtful and 1 out of 10 rainfall stations identified as suspect. It was found that Kota Bahru rainfall station rejected two null hypotheses. Although the Kota Bahru rainfall station may contain inhomogeneity, it can still be applied for further analysis if handled with care. In addition to that, Pos Blau has again rejected all four null hypotheses. Based on figure 3, a break was detected in the yearly rainfall data during year 1999 for Pettitt test, SNHT and BRT. The break observed in the yearly time scale could be due to the significant increase in the amount of rainfall at Kelantan River Basin [13]. On the side note, the practice of using different techniques in measurement could also lead to the presence of inhomogeneity in the rainfall data [14].

Figure 3. Test statistic of Pettitt test, SNHT and BR test of yearly rainfall at Pos Blau station.
4.2 Selection of best fit probability distribution for annual maximum series (AMS) and partial duration series (PDS)

Five different probability distributions namely the Generalized Extreme Value (GEV) Distribution, Generalized Pareto (GP) Distribution, Log-Pearson Type 3 (LP3) Distribution, Log-Normal (LN) Distribution and Log-Normal 3 (LN3) Distribution were fitted to the annual maximum rainfall series and partial duration rainfall series. The LN and LN3 are simpler distributions consisting of two parameters while the GEV, GP and LP3 are complex distributions consisting of three parameters. These wide range of probability distributions can be used to explore and examine the best fit distribution for AMS and PDS. Three GoF tests, namely the chi-square test, Kolmogorov-Smirnov test and Anderson-Darling test were applied for the AMS and PDS to evaluate the performance for each of the probability distribution as shown in table 5. The results were ranked from 5 (best fit) to 0 (least fit) for all probability distributions. A summary on the Best Fit Distribution at each rainfall stations for the AMS and PDS were shown in table 6.

Table 5. Results of GoF tests for annual maximum series (AMS) and partial duration series (PDS).

| Station         | Probability Distribution | Annual Maximum Series | Partial Duration Series |
|-----------------|--------------------------|-----------------------|------------------------|
|                 |                          | KS | AD | χ² | Total | KS | AD | χ² | Total |
| Mardi Jeram     | GEV                      | 4  | 5  | 1  | 10    | 5  | 4  | 4  | 13    |
|                 | GP                       | 5  | 3  | 2  | 10    | 2  | 0  | 0  | 2     |
|                 | LP3                      | 3  | 4  | 4  | 11    | 4  | 3  | 5  | 12    |
|                 | LN                       | 1  | 1  | 5  | 7     | 0  | 0  | 2  | 2     |
|                 | LN3                      | 2  | 2  | 3  | 7     | 3  | 5  | 3  | 11    |
|                 | GEV                      | 5  | 4  | 3  | 12    | 0  | 0  | 0  | 0     |
|                 | GP                       | 2  | 0  | 0  | 2     | 5  | 0  | 0  | 5     |
|                 | LN                       | 1  | 2  | 4  | 7     | 0  | 0  | 0  | 0     |
|                 | LN3                      | 4  | 5  | 2  | 11    | 0  | 0  | 0  | 0     |
|                 | GEV                      | 5  | 5  | 4  | 14    | 3  | 2  | 2  | 7     |
|                 | GP                       | 2  | 0  | 0  | 2     | 5  | 5  | 5  | 15    |
|                 | LN                       | 1  | 0  | 0  | 1     | 5  | 5  | 4  | 14    |
|                 | LN3                      | 4  | 4  | 3  | 11    | 3  | 4  | 2  | 9     |
|                 | GEV                      | 5  | 5  | 5  | 15    | 4  | 2  | 3  | 9     |
|                 | GP                       | 1  | 0  | 0  | 1     | 5  | 5  | 4  | 14    |
|                 | LP3                      | 3  | 2  | 2  | 7     | 0  | 0  | 1  | 1     |
|                 | LN                       | 2  | 3  | 4  | 9     | 2  | 3  | 5  | 10    |
|                 | LN3                      | 4  | 4  | 5  | 13    | 3  | 2  | 2  | 7     |
|                 | GEV                      | 5  | 5  | 3  | 13    | 5  | 5  | 5  | 15    |
|                 | GP                       | 1  | 0  | 0  | 1     | 5  | 5  | 4  | 14    |
|                 | LP3                      | 3  | 3  | 2  | 8     | 2  | 3  | 3  | 8     |
|                 | LN                       | 2  | 2  | 4  | 8     | 4  | 4  | 4  | 12    |
|                 | LN3                      | 4  | 4  | 5  | 13    | 3  | 2  | 2  | 7     |
|                 | GEV                      | 5  | 5  | 3  | 13    | 5  | 5  | 5  | 15    |
| P.Ter.Haiwan     | GEV                      | 5  | 4  | 4  | 14    | 4  | 4  | 4  | 12    |
|                 | GP                       | 1  | 0  | 0  | 1     | 5  | 5  | 4  | 14    |
|                 | LP3                      | 3  | 3  | 3  | 9     | 2  | 2  | 3  | 7     |
|                 | LN                       | 2  | 1  | 1  | 4     | 0  | 0  | 0  | 0     |
|                 | LN3                      | 1  | 4  | 5  | 10    | 5  | 5  | 5  | 15    |
|                 | GEV                      | 5  | 5  | 5  | 15    | 0  | 0  | 0  | 0     |
| Pos Tehoi       | GEV                      | 5  | 4  | 4  | 14    | 4  | 4  | 4  | 12    |
|                 | GP                       | 2  | 3  | 3  | 8     | 0  | 0  | 0  | 0     |
For the AMS, the GEV can be identified as the best fit distribution for 8 out of 10 rainfall stations in the Kelantan River Basin based on the results shown in Table 6. The results can be seen at the Pos Wias rainfall station where most of the distributions (GP, LP3, LN and LN3) were close to the low theoretical quantile ranges but the GEV distribution was the best fit to the medium and high theoretical quantiles as the gaps were smaller as compared to other probability distributions. However, the other two rainfall stations namely, Mardi Jeram Pasu and Pusat Ternakan Haiwan Tanah Merah preferred the LP3 and GP distribution, respectively, in the identification of the AMS. Besides that, it is notable to mention that Pusat Pertanian Lundang station has two best fit distributions which are the GEV and GP distributions. These two distributions performed equally well with a total score of 13 although there were some differences observed in each of the GoF tests applied as seen in Table 5. These findings are consistent with the findings by [15] in the Peninsular Malaysia. A logical explanation to this finding is that the three-parameter GEV distribution is more flexible than other probability distribution in analyzing annual maximum data and it is able to capture the behavior of extreme rainfall characteristics.

As for the distribution of LN and LN3, none of these two were selected for any of the rainfall stations in the Kelantan River Basin. The ability of the three parameter distributions such as GEV, GP and LP3 were able to capture more climatic characteristics of the Kelantan River Basin as compared to the LN and LN3. Additionally, the two parameter distributions of the LN and LN3 are not as flexible when compared to the three parameter distributions [14].

For the PDS, the GP can be identified as the best fit distribution based on Table 6 for 7 out of 10 rainfall stations in the Kelantan River Basin. The possible reason where GP was selected as the best fit distribution is that it is sensitive towards the high frequency time series due to its long and thick upper tail characteristics [16]. This can be seen from the Q-Q plot for the Kuala Krai rainfall station where most of the probability distributions were close to the low theoretical quantile ranges but the GP distribution was the best fit to the medium theoretical quantiles and second best fit to high theoretical quantiles as the gaps were smaller as compared to other probability distributions. However, the other three rainfall stations namely, Mardi Jeram Pasu, Pos Wias and Kota Bahru preferred the GEV, LP3 and LN3 distribution in the identification of the PDS. These results are consistent with previous findings conducted by [17] and [18]. In short, the three parameters GP distribution is more suitable than the other two parameters distribution in capturing the behavior of extreme rainfall characteristics of PDS.

|       | LN   | LN3  | GEV  | GP   | LP3  | LN3  | GEV  | GP   | LP3  | LN3  | Pos Wias |
|-------|------|------|------|------|------|------|------|------|------|------|----------|
|       | 4    | 2    | 8    | 0    | 0    | 0    | 0    | 0    | 0    | 0    |          |
|       | 3    | 1    | 4    | 8    | 0    | 0    | 0    | 0    | 0    | 0    | Kuala Krai |
|       | 5    | 5    | 2    | 12   | 4    | 3    | 4    | 11   |      |      |          |
|       | 3    | 0    | 0    | 3    | 2    | 0    | 0    | 0    | 2    |      | Kota Bahru |
|       | 4    | 3    | 3    | 10   | 5    | 5    | 5    | 15   |      |      |          |
|       | 1    | 2    | 5    | 8    | 1    | 2    | 2    | 5    |      |      |          |
|       | 2    | 4    | 4    | 10   | 3    | 4    | 3    | 10   |      |      |          |
|       | 5    | 5    | 5    | 15   | 0    | 2    | 0    | 2    |      |      |          |
|       | 2    | 0    | 0    | 2    | 5    | 5    | 5    | 15   |      |      |          |
|       | 3    | 3    | 3    | 9    | 0    | 3    | 0    | 3    |      |      |          |
|       | 1    | 2    | 4    | 7    | 0    | 0    | 0    | 0    |      |      |          |
| Kuala Krai | 4    | 4    | 2    | 10   | 4    | 4    | 4    | 12   |      |      |          |
|       | 4    | 5    | 5    | 14   | 3    | 4    | 0    | 7    |      |      |          |
|       | 5    | 0    | 0    | 5    | 5    | 0    | 0    | 5    |      |      |          |
|       | 3    | 4    | 4    | 11   | 2    | 3    | 5    | 10   |      |      |          |
|       | 2    | 2    | 2    | 6    | 0    | 0    | 0    | 0    |      |      |          |
| Kota Bahru | 1    | 3    | 3    | 7    | 4    | 5    | 4    | 13   |      |      |          |
|       | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    |      |      |          |

* Bolded value represents the best fit probability distribution
Figure 4. Q-Q plot for annual maximum series (AMS) at Pos Wias station.

Figure 5. Q-Q plot for partial duration series (PDS) at Kuala Krai station.
Table 6. Best fit probability distribution for annual maximum series (AMS) and partial duration series (PDS).

| Station               | Annual Maximum Series            | Partial Duration Series         |
|-----------------------|----------------------------------|---------------------------------|
| Mardi Jeram Pasu      | Log-Pearson Type 3               | Generalized Extreme Value       |
| Pos Gob               | Generalized Extreme Value        | Generalized Pareto              |
| Pos Bihai             | Generalized Extreme Value        | Generalized Pareto              |
| Pos Blau              | Generalized Extreme Value        | Generalized Pareto              |
| Pusat Pert.Lundang    | Generalized Extreme Value,       | Generalized Pareto              |
|                       |                                  |                                 |
| P.Ter.Haiwan Tanah    | Generalized Pareto               | Generalized Pareto              |
| Merah                 |                                  |                                 |
| Pos Tehoi             | Generalized Extreme Value        | Generalized Pareto              |
| Pos Wias              | Generalized Extreme Value        | Log-Pearson Type 3              |
| Kuala Krai            | Generalized Extreme Value        | Generalized Pareto              |
| Kota Bahru            | Generalized Extreme Value        | Log Normal 3                    |

5. Conclusion
In a nutshell, this study aims to determine the best fit distribution for the AMS and PDS and evaluate their performances in Kelantan River Basin. A few distributions were identified based on the findings from the previous studies such as the Maximum Likelihood Distribution, the Generalized Extreme Value Distribution, Generalized Pareto Distribution, Log-Pearson Type 3 Distribution, Log-Normal Distribution and Log-Normal 3 Distribution. Then, these distributions were applied to the rainfall data of the Kelantan River Basin and their performances were evaluated using the goodness-of-fit tests such as chi-square, Kolmogorov-Smirnov and Anderson-Darling tests.

Based on the findings from this study, we can conclude that the complexity and durability of three parameters distributions were able to capture more climatic characteristics in both the annual maximum series and partial duration series. This can be seen from the results of the three parameters distribution namely, GEV and GP which were frequently selected as the best fit probability distribution for most of the rainfall stations in Kelantan River Basin. These two distributions have smaller gaps as compared to the other distributions for medium and high theoretical quantiles. Although the three parameters distributions can describe the climatic characteristics better, but there was no single probability distribution that can best fit all the stations in for the different the time scales analyzed in this study. Nonetheless, the result of this finding may provide useful insights in selecting the probability distribution for hydrological modelling.

In addition to that, this study adapted the rainfall data from ten rainfall stations, but the findings of this study cannot represent the overall rainfall characteristics of Kelantan River Basin due to the sparse network of rainfall stations in the basin. However, this study can be useful for developing hydrological model in analysing risk and damage from extreme rainfall events or floods. Subsequently, these findings can be applied in different practical field such as hydrological, construction and agriculture.

For further studies, recommendation on the inclusion of Artificial intelligence (AI) to evaluate the time series in determining the best fit distribution may be considered for each rainfall stations. Besides that, the use of a longer length of historical data may directly affect the accuracy of the statistically derived quantiles [13]. An application of more than five distribution are also recommended for different time series as there was no single best fit probability distribution for all rainfall stations.

Acknowledgments
The authors are grateful to the Malaysian Meteorological Department for providing the daily rainfall data in Kelantan River Basin. The authors also would like to acknowledge the sincere appreciation
towards the financial support from UCSI University through Pioneer Scientist Incentive Fund (PSIF) with project code Proj-2019-In-FETBE-065.

References

[1] Kabiri R, Ramani B and Chan, A 2012 Estimation of Climate Change Impacts on Frequency of Precipitation Extremes Case Study: Klang Watershed, Malaysia IPCEBEE 49 144-149

[2] Karamouz M, Nazif S and Falahi, M 2012 Hydrology and hydroclimatology: principles and applications CRC Press New York 740

[3] Department of Irrigation and Drainage (DID) Malaysia 2015 Hydrological Procedure No. 1 (Estimation of Design Rainstorm in Peninsular Malaysia)

[4] Alias N E, Mohamad H, Chin W Y and Yusop Z 2016 Rainfall analysis of the Kelantan big yellow flood 2014 J. Teknol. 78 83-90

[5] Rahman A S, Rahman A, Zaman M A, Haddad K, Ahsan A and Imteaz M 2013 A study on selection of probability distributions for at-site flood frequency analysis in Australia Nat hazards 69 1803-1813

[6] Karim F, Hasan M and Marvanek S 2017 Evaluating annual maximum and partial duration series for estimating frequency of small magnitude floods Water 9 481

[7] Ng C K, Ng J L, Huang Y F, Xun T Y and Mirzaei M 2020 Tropical rainfall trend and stationarity analysis Water Supply

[8] Lian C Y, Huang Y F, Ng J L, Mirzaei M, Koo C H and Tan K W 2019 A proposed hybrid rainfall simulation model: bootstrap aggregated classification tree–artificial neural network (BACT-ANN) for the Langat River Basin, Malaysia J. Water Clim. Change

[9] Ng J L, Yap S Y, Huang Y F, Noh N M, Al-Mansob R A and Razman R 2020 Investigation of the best fit probability distribution for annual maximum rainfall in Kelantan River Basin. In IOP Conf. Ser. Earth Environ. Sci. 47 012118

[10] Amanda Sean Peik L 2017 Assessment of climate change impacts on rainfall series in Peninsular Malaysia using statistical methods 308

[11] Teodorescu S 2010 On the truncated composite lognormal-Pareto model Math. Rep. 12 71-84

[12] Ben-Zvi A 2010 Rainfall intensity–duration–frequency relationships derived from large partial duration series J. Hydrol. 307 104-114

[13] Pham H X, Shamseldin A Y and Melville B 2014 Statistical properties of partial duration series: Case study of North Island, New Zealand J. Hydrol. Eng. 19 807- 815