Jurnal Teknologi

Full Paper

CHOOSING THE BEST FIT DISTRIBUTION FOR RAINFALL EVENT CHARACTERISTICS BASED ON 6H-IETD WITHIN PENINSULAR MALAYSIA

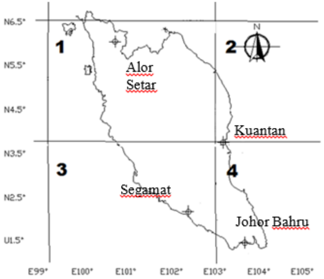
Article history
Received
5 June 2014
Received in revised form
1 February 2015
Accepted
15 June 2015

Zulkarnain Hassan[a,b*], Supiah Shamsudin[c], Sobri Harun[a]

[a]Faculty of Civil Engineering, Universiti Teknologi Malaysia, 81310, UTM Johor Bahru, Johor, Malaysia
[b]The School of Environmental Engineering, Universiti Malaysia Perlis, Arau, Perlis, Malaysia
[c]Razak School of Engineering and Advanced Technology, Universiti Teknologi Malaysia-Kuala Lumpur, Kuala Lumpur, Malaysia

*Corresponding author
Zulkarnain23@live.utm.my

Graphical abstract



Abstract

In selecting the best-fit distribution model for the rainfall event characteristics based on the inter-event time definition (IETD) of 6 hours for the selected rainfall in the Peninsular of Malaysia, seven distributions were utilized namely the beta (B4), exponential (EX1), gamma (G2), generalized extreme value (GEV), generalized Pareto (GP), Log-Pearson 3 (LP3), and Wakeby (WKB). Maximum likelihood estimation (MLE) was applied to estimate the parameters of each distribution. Based on the results, GP, WKB and GEV were found to be the most suitable distribution for describing the rainfall event characteristics in the studied regions.

Keywords: Rainfall event, distribution model, inter-event time definition, Peninsular of Malaysia

Abstrak

Dalam permilihan model taburan yang bersesuaian dengan sifat musim hujan berdasarkan definasi masa antara kejadian untuk 6 jam bagi kawasan Semenanjung Malaysia, 7 jenis taburan telah digunakan iaitu the beta (B4), exponential (EX1), gamma (G2), generalized extreme value (GEV), generalized Pareto (GP), Log-Pearson 3 (LP3), dan Wakeby (WKB). Penganggar kebolehjadian maksimum (MLE) telah digunakan untuk menganggarkan parameter untuk setiap taburan. Bedasarkan keputusan analisa, WKB, GP dan GEV menjadi taburan yang bersesuaian untuk memperihal sifat musim hujan bagi kawasan kajian.

Kata kunci: Musim hujan, model taburan, definasi antara kejadian, Semenanjung Malaysia

1.0 INTRODUCTION

Managing the water resources in an optimum mode has a huge benefit towards economic, social and environmental water demands [1]. Understanding the behavior of rainfall is indeed crucial since it is highly related with the water resources. Hence, it is important to understand the characteristics and pattern of rainfall [2] for that purpose. Many assessments (e.g. Wan Zin *et al*. [3]; Wilks [4]; Blain and Camargo [5]) had been widely done in many regions to identify the best theoretical distribution models to modeling the rainfall series. These models describe the random (stochastic) behavior of a process, especially in hydrologic studies such as rainfall and flood process. Hence, it is crucial to select the best-fit distribution of those processes [5]. By fitting a distribution to the rainfall series, the rainfall amount (quantile) can be projected correlatedly with the high return period although using a shorter period of rainfall records [6].

Nowadays, intensive studies related to those assessments have been done in Malaysia. Fadhilah *et al*. [7] applied the exponential (EX1), gamma (G2), Weibul and mixed-exponential to fit the hourly rainfall amount in the Federal Territory, Malaysia. Their research found mixed-exponential could describe most of hourly rainfall amount. Jamaludin and Jemain [2] determined that mixed-exponential is able to fit the daily amount of rainfall in the Peninsular of Malaysia. Zalina *et al*. [8] further an assessment by examining the best distribution that fit the annual maximum rainfall for hourly rainfall series. Their research found generalized extreme value (GEV) is able to fit well in the Peninsular of Malaysia.

Although the studies in Malaysia, as discussed above, focused more on the investigation of the selection of best fitting distribution on the rainfall process, there was only one study that focused on rainfall event characteristics based on inter-event time definition (IETD), which was conducted by Dan'azumi *et al*. [9]. There is also no intensive recent studies on fitted distribution model to the IETD rainfall characteristics at worldwide studies, except for the studies by Eagleson [10.11], Howard [12], Adams and Bontje [13] and Adams *et al*. [14], which found the EX2 distribution always fits the histograms of rainfall series satisfactory.

Dan'azumi *et al*. [9] only focused on the intensity of rainfall event and examined four distributions named as beta (B4), EX1, G2, and generalized Pareto (GP) and the use of complex parametric distributions was not examined very well. Morgan *et al*. [15] stated those complex distributions were able to describe the probabilistic structure of the natural process (e.g. rainfall) and able to display shapes that other distribution cannot do. Therefore, an assessment of the multi parameters and other potential distributions to the rainfall event characteristics in the Malaysian region becomes the interest of this study.

This paper aims to explore the best fitted distributions for the rainfall event characteristics based on IETD of 6 hours within the Peninsular of Malaysia. Goodness-of-fit tests and seven distributions include B4, EX1, G2, GEV, GP, Log-Pearson 3 (LP3), and Wakeby (WKB) were utilized. In the following sections, the data and materials are introduced, followed by the results. Then, the discussions and conclusion are presented.

2.0 DATA AND METHODOLOGY

2.1 Case Study

Hourly rainfall series from the four stations at different climate regions in the Peninsular of Malaysia were the input in this study. Each station can be described as different climate regions in the peninsular including the north (region 1), east (region 2), west (region 3) and south (region 4) regions. The rainfall data were obtained from the Department of Irrigation and Drainage (DID) Malaysia. The details and location of the used data can be referred to Table 1 and Figure 1, respectively. Those data are homogenized and contain a smaller missing data (<5%).

Table 1 Details of hourly rainfall station

ID	Name of Station	Coordinate		Year
		Lat (°N)	Long (°E)	
6108001	Alor Setar	6.11	100.85	2001-2012
3833002	Kuantan	3.81	103.33	2001-2012
2528012	Segamat	2.52	102.81	2001-2012
1437116	Johor Baharu	1.47	103.75	2002-2012

2.2 Defining Rainfall Event Characteristics

A long rainfall series consists of a series of rainfall pulse and it can be separated into an individual rainfall event. Each two rainfall events is divided by a dry period. In order to define rainfall event, the start and the end of the event should be identified. The definition that can be applied to separate the events is inter-event time definition (IETD) (as shown in Figure 2). In general, the IETD value functions as the minimum inter-event period between two pulses of rainfall and is located between single rainfall events in the rainfall series. Two pulses of rainfall are categorized as the same event if the time between pulses (*h*) is less than the IETD, and vice versa (Figure 2).

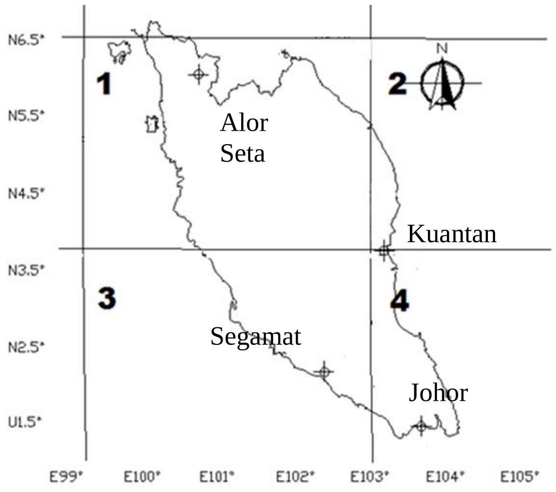


Figure 1 Location of hourly rainfall station

From the individual rainfall event (j), which is defined by IETD, its characteristics (Figure 3) can be analyzed statistically such as the rainfall duration (t_j), intensity (i_j), inter-event (b_j) and amount (v_j). As shown in Figure 3, the t_j of the rainfall event is the duration of the rainfall event within the single event. The i is the average intensity, which is determined by dividing the amount (v_j) with t_j. The b_j can be obtained by summing all dry period within the single event. The v_j of the rainfall event can be determined by summing all the rainfall amounts of each sampling interval within the single event. Therefore, each individual event will provide their own characteristics from the long rainfall series and will be averaged to get the average characteristic, which can be written in the Equations 1-4.

$$Average\ duration, t = \frac{\sum_{j=1}^{n} t_j}{n} \tag{1}$$

$$Average\ intensity, i = \frac{\sum_{j=1}^{n} t_j}{n} \tag{2}$$

$$Average\ interevent, b = \frac{\sum_{j=1}^{n} b_j}{n} \tag{3}$$

$$Average\ amount, v = \frac{\sum_{j=1}^{n} v_j}{n} \tag{4}$$

where, n is the total of rainfall events within studied period. The details and selection of each characteristic are discussed in detail by Adam and Papa [16]. In this study, the IETD of 6 hour (named as 6H IETD) was applied. The 6H IETD is found adequate to be applied in many forms of application [17].

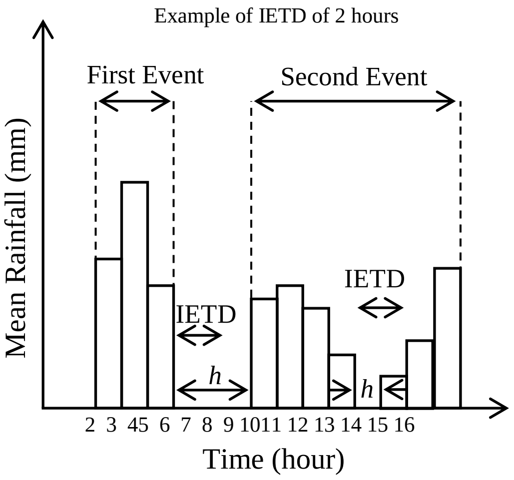


Figure 2 Concept of IETD

2.3 Model Distribution Fittings

In this study, we consider several potential distributions, which are commonly used in describing rainfall series (including extreme rainfall) and flood series in previous studies. The studied distributions include beta (B4), exponential (EX1), gamma (G2), generalized extreme value (GEV), generalized Pareto (GP), Log-Pearson 3 (LP3), and Wakeby (WKB). The parameters were estimated using maximum likehood method (MLE). The details of MLE in estimating the probability parameters are discussed by Myung [18]. For some cases, method of moments, least squares moments, or L-moments moments are applied when the estimation of MLE is difficult or unavailable. The following section describe search of the distributions. x is denoted as random variable, pdf or $f(x)$ denotes the probability distribution function, and cdf or $F(x)$ denotes the cumulative distribution function.

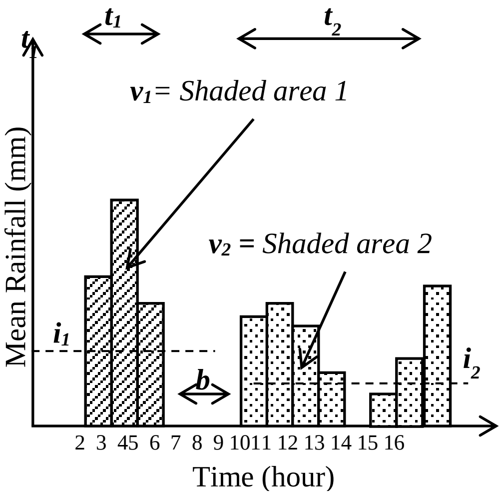


Figure 3 Defining characteristics of storm event

2.3.1 Beta Distribution

The pdf and cdf of Beta distribution can be written as:

$$f(x) = \left[\frac{1}{B(\alpha_1,\alpha_2)}\right]\left[\frac{(x-a)^{(\alpha_1-1)}(b-x)^{(\alpha_2-1)}}{(b-a)^{(\alpha_1+\alpha_2-1)}}\right] \tag{5}$$

and,

$$F(x) = I_z(\alpha_1, \alpha_2) \tag{6}$$

where,

$$z \equiv \frac{x-a}{b-a} \tag{7}$$

In those equations, α_1 and α_2 are the continuous shape parameters, and *a* and *b* are the continuous boundary parameters. $\alpha_1 > 0$, $\alpha_2 > 0$ and $a < b$. B is the beta function (Equation 8) and I_z is the regularized incomplete beta function (Equation 9).

$$B = \beta(\alpha_1, \alpha_2) = \int_0^1 (t^{\alpha_1-1})\,(1-t)^{\alpha_2-1} dt\,, \quad (\alpha_1, \alpha_2 > 0) \tag{8}$$

$$I_x(\alpha_1, \alpha_2) = \frac{B_x(\alpha_1,\alpha_2)}{B(\alpha_1,\alpha_2)} \tag{9}$$

2.3.2 Exponential Distribution

In this study, one parameter exponential distribution was applied. The pdf and cdf of this distribution can be written as;

$$f(x) = \lambda exp(-\lambda x) \tag{10}$$

and,

$$F(x) = 1 - exp(-\lambda x) \tag{11}$$

Where, λ is the continuous inverse scale parameter (λ>0) and x lies [λ, ∞).

2.3.3 Gamma distribution

Two-parameter Gamma distribution was–applied. The pdf and cdf of this distribution can be summarized as;

$$f(x) = \frac{x^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} exp\left(\frac{-x}{\beta}\right) \tag{12}$$

and,

$$F(x) = \frac{\alpha\Gamma_{x/\beta}}{\alpha\Gamma} \tag{13}$$

where, α and β are the continuous shape and scale parameters, respectively (α>0 and β>0) and x lies [λ, ∞). Γ is the Gamma (Equation 14) and Γ_z Incomplete Gamma functions (Equation 15).

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt \qquad , \alpha > 0 \tag{14}$$

$$\Gamma_x(\alpha) = \int_0^{x} t^{\alpha-1} e^{-t} dt \qquad , \alpha > 0 \tag{15}$$

2.3.4 Generalized Extreme value distribution

Pdf and cdf Generalized Extreme value (GEV) distribution can be written as:

$$f(x) = \begin{cases} \frac{1}{\sigma} exp\left[-(1+kz)^{-\frac{1}{k}}\right](1+kz)^{-1-\frac{1}{k}} & ; k \neq 0 \\ \frac{1}{\sigma} exp(-z - exp(-z)) & ; k = 0 \end{cases} \tag{16}$$

and,

$$F(x) = \begin{cases} exp\left[-(1+kz)^{-\frac{1}{k}}\right] & ; k \neq 0 \\ exp[-exp(-z)] & ; k = 0 \end{cases} \tag{17}$$

where,

$$z \equiv \frac{x-\mu}{\sigma} \tag{18}$$

In these equations, k is the continuous shape parameter, σ is the continuous scale parameter (σ>0), and μ is the continuous location parameter. For k ≠ 0, 1+ (k/σ)(x-μ) > 0 and k=0, x is lying between (-∞,∞)

2.3.5 Generalized Pareto Distribution

Pdf and cdf Generalized Pareto (GP) distribution can be written as:

$$f(x) = \begin{cases} \frac{1}{\sigma}\left[1 + k\left(\frac{x-\mu}{\sigma}\right)\right]^{-1-\frac{1}{k}} & ; k \neq 0 \\ \frac{1}{\sigma} exp\left(-\frac{x-\mu}{\sigma}\right) & ; k = 0 \end{cases} \tag{19}$$

and,

$$F(x) = \begin{cases} 1 - \left[1 + k\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{k}} & ; k \neq 0 \\ 1 - exp\left(-\frac{x-\mu}{\sigma}\right) & ; k = 0 \end{cases} \tag{20}$$

where, k is the continuous shape parameter, σ is the continuous scale parameter (σ > 0), and μ is the continuous location parameter. x lies in the following condition:

$$\mu \leq x < \infty \qquad \text{for} \qquad k \geq 0 \tag{21}$$

$$\mu \leq x < \mu - (\sigma / k) \qquad \text{for} \qquad k < 0 \tag{22}$$

2.3.6 Log-Pearson 3 Distribution

The pdf and cdf of Log-Person 3 distribution can be written as:

$$f(x) = \frac{1}{x|\beta|\Gamma(\alpha)}\left[\frac{ln(x)-\gamma}{\beta}\right]^{\alpha-1} exp\left[-\frac{ln(x)-\gamma}{\beta}\right] \tag{23}$$

and,

$$F(x) = \frac{\Gamma_{\frac{ln(x)-\gamma}{\beta}}(\alpha)}{\Gamma(\alpha)} \tag{24}$$

In these equations, a is the continuous shape parameter (a > 0), β is the continuous scale parameter (β ≠ 0), and γ is the continuous location parameter. If β < 0 and β > 0, then x are in the range (0, e^{γ}] and [e^{γ},∞) respectively.

2.3.7 Wakeby Distribution

Pdf and cdf of Wakeby distribution are not explicitly defined and can be defined by its quartile function, which can be written as:

$$x(F) = \mu + \frac{\alpha}{\beta}\left[1 - (1-F)^{\beta}\right] - \frac{\gamma}{\sigma}\left[1 - (1-F)^{-\sigma}\right] \tag{25}$$

where,

$$\mu \leq x < \infty \text{ if } \delta \geq 0 \text{ and } \gamma > 0, \tag{26}$$

$$\mu \leq x \leq \mu + (\alpha/\beta) - (\gamma/\delta) \text{ if } \delta < 0 \text{ or } \gamma = 0. \tag{27}$$

In those equations, β,σ, α, γ, and μ are the continuous parameters.

2.4 Exploratory Data Analysis

To explore the performance of each distribution on how well the distributions fit to the data, goodness-of-fit assessments were applied, which are the Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D) and Q-Q plot. Those techniques are revealed as the most valid distribution describing the storm event characteristics in the studied region.

Table 2 Description of 6H-IETD storm event for historical period

	Mean	SD	Skew	Max	Events
d (hour)					
Alor Setar	6.43	7.08	3.17	63	2051
Kuantan	7.11	11.15	5.64	167	1795
Segamat	4.35	5.56	3.74	55	1349
Johor Bahru	8.34	7.11	4.32	87	1767
v (mm)					
Alor Setar	12.19	17.42	3.12	168.9	2051
Kuantan	16.42	39.87	7.99	618.2	1795
Segamat	10.83	19.57	6.34	323.2	1349
Johor Bahru	17.11	27.16	6.56	526.4	1767
i (mm/hour)					
Alor Setar	2.47	3.91	5.81	74.25	2051
Kuantan	2.55	3.56	3.09	32	1795
Segamat	2.71	3.77	3.16	31.5	1349
Johor Bahru	2.35	3.3	3.81	38	1767
b (hour)					
Alor Setar	44.85	94.71	10.78	2208	2051
Kuantan	51.49	87.27	6.65	1413	1795
Segamat	60.63	111.72	6.39	1580	1349
Johor Bahru	46.23	78.82	7.01	1170	1767

The K-S and A-D tests measure the performance of a random sample with a theoretical probabilistic distribution function by comparing the fit of a sample cdf to the theoretical cdf. The statistics for K-S and A-D are labeled with D and A^2 respectively. Those statistics can be written as Equations 28 and 29.

$$D = \max_{1 \le i \le n} \left(F(x_i) - \frac{i-1}{n}, \frac{i}{n} - F(x_i) \right) \qquad (28)$$

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i-1). \left[lnF(x_i) + ln\left(1 - F(x_{n-i+1})\right) \right] \qquad (29)$$

The lower values of D and A2, the performance of models are better.

Generally, Q-Q plot visually presents the performance by comparing the sample data with fitted distribution graph. It technically plots the quantile (amount) of selected distribution and compares it with the corresponding quantile of the sample (observed) data. The different between quantiles of selected distribution and observed data are checked using root mean square (rmse) and coefficient of correlation (*r*) as follow:

$$R = \frac{\sum(obs - \overline{obs})(pred - \overline{pred})}{\sqrt{\sum(obs - \overline{obs})^2 \sum(pred - \overline{pred})^2}} \qquad (30)$$

$$RMSE = \sqrt{\frac{(obs - pred)^2}{n}} \qquad (31)$$

In which, *obs* = quantile of observed data; *pred* = quantile of distribution data; $\overline{obs}$ = mean quantile of observed data, and $\overline{pred}$ = mean quantile of distribution data. The closer *R* and *NSE* values are to 1 and *RMSE* value is to 0, the better the prediction.

3.0 RESULT

3.1 Identification of Rainfall Event from Hourly Rainfall Series

Table 2 shows the statistical description (mean, standard deviation (SD), skewness (Skew), maximum (max) and number of rainfall events) of the IETD of 6 hour. From the table, it is depicted that most regions did not face the same average duration (*t*) of rainfall events, in which the mean ranges between 4.35hour and 8.34hour. It is also shown that the maximum *t* is recorded at Kuantan (167 hours). In terms of average amount (*v*) of 6H-IETD, Kuantan received a higher average *v* of rainfall event, compared to other regions with mean and max value are 16.42 mm and 618.20 mm, respectively. However, Segamat, in general, obtained a high intensity (*i*) of storm event with the value of 2.71 mm/hour followed by Kuantan (2.55 mm/hour), Alor Setar (2.47 mm/hour) and Johor Bahru (2.35 mm/hour).

In terms of *i*, Segamat area faced a higher average *i*; with mean value is 2.71 mm/hour, followed by Kuantan, Alor Setar and Johor Bahru. For inter-event (*b*), the table shows that Segamat obtained a higher mean value compared to other regions. However, the skewness of *b* for Alor Setar is higher and the maximum record of *b* is recorded in this region with the value of 2208 hour. Based on the number of events, Alor Setar showed many events which are 2051 events, compared to other regions within the studied periods.

3.2 Ranking the fitted distributions

Tables 3-4 describe the best fitted probability distributions for rainfall event characteristics such as duration (*t*), amount (*v*), intensity (*i*) and inter-event (*b*) based on Equations 28-29. The results reveal that WKB and GEV are consistent to fit the *t* storm event (Table 3a) which becomes the top four in the mean rank. Table 3a also shows that the GP distribution is able to become a first rank to fit the *t* at Alor Setar, Kuantan and Segamat. However, this distribution is not consistent in which it becomes the fourth rank at Johor Bahru.

Table 3 Goodness-of-fit test ranking for various distributions of duration and amount series

Distribution	**a.** *t*					**b.** *v*				
	K-S		A-D		Mean Rank	K-S		A-D		Mean Rank
	D	Rank	A2	Rank		D	Rank	A2	Rank	
Alor Setar										
B4	0.231	7	1617.9	7	7	0.127	5	38.739	5	5
EX1	0.144	4	31.95	3	3.5	0.168	7	136.97	7	7
G2	0.185	6	47.05	6	6	0.145	6	31.345	3	4.5
GEV	0.125	3	37.74	4	3.5	0.12	3	42.055	6	4.5
GP	0.12	1	29.38	1	1	0.125	4	27.959	2	3
LP3	0.149	5	43.06	5	5	0.08	2	19.404	1	1.5
WKB	0.12	2	29.38	2	2	0.064	1	37.498	4	2.5
Kuantan										
B4	0.422	7	542.91	7	7	0.111	3	41.068	5	4
EX1	0.144	4	55.66	5	4.5	0.25	6	288.45	7	6.5
G2	0.347	6	174.9	6	6	0.426	7	348.15	9	8
GEV	0.133	3	35.02	3	3	0.132	4	38.71	4	4
GP	0.13	1	29.11	1	1	0.135	5	28.252	3	4
LP3	0.155	5	36.83	4	4.5	0.089	1	20.128	2	1.5
WKB	0.13	2	29.11	2	2	0.097	2	14.705	1	1.5
Segamat										
B4	0.634	7	4414.7	7	7	0.228	6	58.551	5	5.5
EX1	0.205	5	53.83	5	5	0.224	5	145.42	7	6
G2	0.319	6	95.9	6	6	0.286	7	92.774	6	6.5
GEV	0.195	3	46.45	4	3.5	0.133	2	34.057	4	3
GP	0.183	1	40.46	1	1	0.141	3	25.024	2	2.5
LP3	0.205	4	46.04	3	3.5	0.112	1	19.433	1	1
WKB	0.183	2	40.46	2	2	0.141	4	25.024	3	3.5
Johor Bahru										
B4	0.175	6	460.17	7	6.5	0.124	5	36.481	4	4.5
EX1	0.19	7	103.57	5	6	0.158	6	96.64	7	6.5
G2	0.14	5	50.22	4	4.5	0.182	7	65.315	5	6
GEV	0.105	2	18.81	2	2	0.108	4	30.195	3	3.5
GP	0.118	4	434.43	6	5	0.105	3	17.877	2	2.5
LP3	0.112	3	24.38	3	3	0.055	2	8.9117	1	1.5
WKB	0.061	1	6.68	1	1	0.041	1	72.921	6	3.5

Table 3b shows the goodness-of-fit result for the *v* series of rainfall event. All studied areas are distributedly fitted with the LP3 distribution for the *v*. For the second and third rank, the WKB, GP and GEV distributions are constantly changing position with each other, with a quite similar mean rank to fit the amount. The results also reveal that LP3 is consistent to become the top three in the fitted distribution for *i* series (Table 4a). However, GP and WKB are more likely to be at a higher rank, compared to LP3 for this series. This pattern of fitting distributions is similar for the *b* (Table 4b) in which the LP3, GP and GEV distributions are listed in the highest ranking by the K-S and A-D tests, compared to other distributions.

3.3 Quantile Estimation

Figures 4-7 and Table 5 illustrate the Q-Q plots of the 6H IETD storm event and its performance indicators respectively. This is the underlying procedure in the probability distribution selection.

As shown in Figure 4, the quantiles for *t* data and its estimation at Alor Setar, Kuantan and Segamat appear to come from populations with the GP and WKB distributions, in which they are able to plot near to the 45-degree reference line. This result also agrees with the performance indicator (Table 5), where it can be seen that the GP and WKB distributions give the best estimation (Table 5) in which rmse<1.600 hour and r>0.950. At Johor Bahru, the study found that GEV is able to plot near to the reference line, compared to other distributions for the *t* case in which rmse=1.997 hour and r=0.965.

In terms of probability distribution of the average *v* as shown in Figure 5, the results reveal that the points of WKB distributions lie nearly along the reference line and followed by GP at most of studied regions. This result also agrees with the result of Table 5 in which the lowest value of rmse and highest value of *r* are shown for almost studied region. The same plot trend of quantiles of the probability distribution for the *v* data is illustrated for the Q-Q plot for *i* as shown in Figure 6. However, G2 and B4 show that the plotted points with the reference lines are straight line except a few outliers, and the differences are smaller with the GP, WKB, and GEV distributions.

In terms of the quantiles for *b*, Figure 7 shows that similar result is obtained like other storm event characteristics. The GP, WKB and GEV give a promising result in fitted distribution by giving a fairly curve close to the 45° reference line.

Table 4 Goodness-of-fit test ranking for various distributions of intensity and inter-event series

Distribution	**a.** *i*					**b.** *b*				
	K-S		A-D		Mean Rank	K-S		A-D		Mean Rank
	D	Rank	A2	Rank		D	Rank	A2	Rank	
Alor Setar										
B4	0.233	7	424.94	7	7	0.250	6	483.17	6	6
EX1	0.154	5	69.49	5	5	0.185	5	112.16	3	4
G2	0.229	6	151.67	6	6	0.497	7	566	7	7
GEV	0.089	4	21.55	4	4	0.068	4	9.0473	2	3
GP	0.07	2	8.35	2	2	0.051	2	278.89	4	3
LP3	0.068	1	5.76	1	1	0.051	1	3.2928	1	1
WKB	0.07	3	8.35	3	3	0.051	3	278.89	5	4
Kuantan										
B4	0.179	7	78.19	7	7	0.201	6	406.42	7	6.5
EX1	0.161	5	68.06	5	5	0.127	5	55.221	5	5
G2	0.173	6	69.79	6	6	0.364	7	260.19	6	6.5
GEV	0.121	4	33.07	4	4	0.066	4	9.269	4	4
GP	0.09	1	18.25	2	1.5	0.029	1	1.6747	1	1
LP3	0.106	3	11.21	1	2	0.033	3	2.7538	3	3
WKB	0.09	2	18.25	3	2.5	0.029	2	1.6747	2	2
Segamat										
B4	0.116	3	42	5	4	0.23	6	314.18	7	6.5
EX1	0.146	6	53.28	6	6	0.148	5	54.573	5	5
G2	0.222	7	70.45	7	7	0.39	7	223.94	6	6.5
GEV	0.127	4	29.13	4	4	0.072	4	8.2719	4	4
GP	0.102	1	22.66	2	1.5	0.037	2	1.6174	1	1.5
LP3	0.134	5	15.96	1	3	0.035	1	2.1693	3	2
WKB	0.102	2	22.66	3	2.5	0.037	3	1.6174	2	2.5
Johor Bahru										
B4	0.041	2	8.39	3	2.5	0.156	6	155.77	4	5
EX1	0.12	6	54.5	7	6.5	0.14	5	78.74	3	4
G2	0.16	7	42.71	6	6.5	0.397	7	322.3	7	7
GEV	0.102	5	22.96	5	5	0.096	4	18.433	2	3
GP	0.093	4	11.5	4	4	0.073	1	263.36	6	3.5
LP3	0.043	3	6.98	2	2.5	0.077	3	11.02	1	2
WKB	0.034	1	1.54	1	1	0.075	2	262.24	5	3.5

3.4 Fitted Parameter Distributions

Tables 6-7 (all rainfall event characteristics) summarized the variations of the continuous shape (α, k), scale (σ, β), location (μ, γ) and inverse scale (λ) parameters of the studied distributions by using Equations 5-25. In general, the result of each storm event characteristic and the studied distribution do not show a correlation and an identical result. The variation of each parameter is unique and contributes to the different size of curve as shown in Figure 8. As depicted in Figure 8, the study compares the PDF curve of the GEV distribution for all studied rainfall event characteristics. It can be seen that the PDF of the figures is long tail to the right towards the characteristics of the rainfall event. In terms of *t*, the result reveals that Segamat has a higher peak of the PDF and smaller long tail to the right compared to other regions, followed by Kuantan and Alor Setar. However, the PDF of *t* for Johor Bahru is not similar with other regions, and it seems that the curve is shifted to the right. For *v*, Segamat is giving the highest occurrence of *v*, compared to other regions. In terms of the *i* and *b*, the graphs consistently show that Alor Setar gives a highest peak of those characteristics, compared to other regions.

4.0 DISCUSSION

This study aims to evaluate the fitted distribution model from the rainfall event characteristics in the Peninsular of Malaysia for the current period. In order to separate the rain event, inter-event time definition (IETD) of 6 hours is introduced. The selection of hours of IETD is relative and depends on the purpose on the analysis [19, 20]. In the Malaysian region, an application of IETD of 6 hours is widely applied in the urban storm water management such as Dan'azumi *et al*. [21, 22] and Shamsudin *et al*. [23].

In this study, the application of IETD reveals that the rainfall event with their characteristics (*t*, *v*, *i* and *b*) can be described well in terms of statistical distribution in the selected region in the Peninsular of Malaysia. The comparison between each characteristic for Johor Bahru with other regions may not fit well because the time period of the selected study is not similar with the other three regions. However, for the sake of this study, it is assumed that the length of the period (as shown in Table 1) for Johor Bahru is similar with the other regions and hence, enables the characteristics' comparison in terms of its statistical and fitted distribution.

Table 5 The performance of the distribution based on the Q-Q plot (bold text refers to the best performance)

	t		***v***		***i***		***b***	
	rmse (hour)	***r***	**rmse (mm)**	***r***	**rmse (mm/hour)**	***r***	**rmse (hour)**	*r*
Alor Setar								
B4	2.047	0.971	15.289	0.746	1.195	0.964	90.452	0.865
EX1	1.537	0.979	15.920	0.740	1.855	0.930	65.120	0.794
G2	1.309	0.983	14.820	0.757	0.917	0.972	39.645	0.912
GEV	1.751	0.974	17.360	0.699	1.666	0.953	29.318	0.972
GP	**0.995**	**0.990**	15.489	0.740	**0.825**	**0.985**	**15.781**	**0.989**
LP3	2.534	0.972	22.799	0.692	1.292	0.980	30.806	0.981
WKB	**0.995**	**0.990**	**14.818**	**0.757**	**0.825**	**0.985**	**15.781**	**0.989**
Kuantan								
B4	3.714	0.974	23.219	0.876	0.813	0.988	61.415	0.921
EX1	5.488	0.914	27.864	0.827	1.202	0.977	48.585	0.880
G2	3.070	0.962	12.270	0.952	**0.284**	**0.997**	28.252	0.947
GEV	3.763	0.961	19.723	0.919	2.216	0.890	33.043	0.952
GP	**1.653**	**0.990**	14.159	0.949	1.332	0.951	**18.103**	**0.980**
LP3	4.411	0.972	33.715	0.940	2.699	0.905	29.386	0.970
WKB	**1.653**	**0.990**	**7.421**	**0.984**	1.332	0.951	**18.103**	**0.980**
Segamat								
B4	5.340	0.776	7.272	0.944	0.813	0.988	77.042	0.922
EX1	1.864	0.959	10.668	0.912	1.202	0.977	65.885	0.873
G2	1.140	0.979	5.106	0.966	**0.284**	**0.997**	33.657	0.954
GEV	2.747	0.921	6.905	0.967	2.216	0.890	48.069	0.936
GP	**1.572**	**0.968**	**3.486**	**0.989**	1.332	0.951	**31.472**	**0.966**
LP3	3.442	0.922	20.217	0.946	2.699	0.905	48.922	0.954
WKB	**1.572**	**0.968**	**3.486**	**0.989**	1.332	0.951	**31.472**	**0.966**
Johor Bahru								
B4	3.754	0.920	10.613	0.939	0.920	0.983	40.758	0.869
EX1	2.905	0.941	13.120	0.924	1.292	0.973	46.136	0.854
G2	2.589	0.934	8.028	0.956	**0.401**	**0.994**	29.713	0.929
GEV	**1.997**	**0.965**	6.794	0.981	2.350	0.886	34.746	0.929
GP	5.037	0.713	**3.922**	**0.990**	2.350	0.886	23.042	0.961
LP3	2.779	0.925	13.928	0.975	3.479	0.871	33.107	0.946
WKB	2.778	0.938	6.607	0.971	1.501	0.944	**19.020**	**0.971**

Table 6 Fitted parameters for Alor Setar and Kuantan

	a. Alor Setar				**b. Kuantan**			
	t	***v***	***i***	***b***	***t***	***v***	***i***	***b***
B4								
$\alpha 1$	0.446	0.448	0.324	0.247	0.135	0.498	0.378	0.287
$\alpha 2$	4.662	4.894	9.793	7.383	7.564	53.665	4.330	6.012
EX1								
λ	0.155	0.082	0.405	0.022	0.141	0.061	0.392	0.019
G2								
α	0.825	0.489	0.399	0.224	0.406	0.170	0.513	0.348
β	7.797	24.912	6.196	199.980	17.499	96.794	4.972	147.920
GEV								
κ	0.310	0.431	0.495	0.606	0.482	0.604	0.466	0.529
σ	3.167	6.069	1.029	13.459	2.895	6.112	1.124	18.638
μ	3.225	4.225	0.900	16.988	2.828	3.883	0.952	20.485
GP								
κ	0.113	0.296	0.386	0.538	0.368	0.534	0.346	0.434
σ	5.390	9.216	1.478	17.676	4.204	8.046	1.656	26.027
μ	0.361	-0.900	0.061	6.607	0.450	-0.838	0.021	5.541
LP3								
α	1730.4	468.4	335.2	4.526	28.699	225.39	142.72	12.184
β	0.024	-0.069	0.066	0.438	0.199	0.107	0.101	0.282
γ	-40.181	33.779	-21.796	1.232	-4.365	-22.575	-14.126	-0.072
WKB								
α	0	-18.926	0	0	0	-18.832	0	0
β	0	1.373	0	0	0	0.215	0	0
γ	5.390	19.229	1.478	17.676	4.204	22.365	1.656	26.027
σ	0.113	0.029	0.386	0.538	0.368	0.299	0.346	0.434
μ	0.361	0.365	0.061	6.607	0.450	0.000	0.021	5.541

Table 7 Fitted parameters for Segamat and Johor Bahru

	a. Segamat				**b. Johor Bahru**			
	t	**v**	**i**	**b**	**t**	**v**	**i**	**b**
B4								
$\alpha1$	0.012	0.303	0.754	0.244	0.891	0.411	0.636	0.541
$\alpha2$	4.456	7.417	123.160	4.751	9.540	10.329	68.723	11.704
EX1								
λ	0.230	0.092	0.368	0.016	0.120	0.058	0.426	0.022
G2								
α	0.613	0.306	0.520	0.295	1.378	0.397	0.506	0.344
β	7.099	35.366	5.223	205.850	6.057	43.125	4.643	134.390
GEV								
κ	0.465	0.512	0.480	0.562	0.150	0.438	0.422	0.555
σ	1.676	4.798	1.136	21.300	3.792	8.353	1.135	14.842
μ	1.972	3.189	1.043	21.920	5.502	5.975	0.891	19.759
GP								
κ	0.344	0.410	0.366	0.478	-0.150	0.306	0.283	0.469
σ	2.470	6.796	1.653	28.969	7.619	12.605	1.738	20.296
μ	0.582	-0.692	0.109	5.122	1.721	-1.051	-0.073	8.015
LP3								
α	11.266	272.680	23.202	11.534	13.071	181.090	702.510	5.137
β	0.274	0.091	0.229	0.308	-0.215	-0.109	-0.050	0.386
γ	-2.088	-23.506	-4.957	-0.106	4.659	21.649	35.013	1.358
WKB								
α	0	0	0	0	16.315	-18.388	-2.069	-7.248
β	0	0	0	0	2.014	2.098	1.354	0.805
γ	2.470	6.796	1.653	28.969	1.107	19.279	2.802	24.067
σ	0.344	0.410	0.366	0.478	0.557	0.145	0.112	0.422
μ	0.582	-0.692	0.109	5.122	0.432	0.481	0.072	8.603

In general, the study found that Kuantan faced a higher average *i* and *v* of 6H-IETD event compared to other regions. It happened because the area, which is located on the east part of the Peninsular of Malaysia, faced a heavy rainfall distribution that is influenced by the north-east monsoon. This result is also consistent with the study by Suhaila *et al*. [24]. Although Kuantan faced a long average of *i* and a high *v*, the higher rate of *i* based on the 6H IETD is recorded at Segamat with the rate 2.71 mm/hour.

In order to select the best fitted distribution for the selected studies, numerous closely-related distributions were used, named as beta (B4), exponential (EX1), gamma (G2), generalized extreme value (GEV), generalized Pareto (GP), Log-Pearson 3 (LP3), and Wakeby (WKB). Previous study by Dan'Azumi *et al*. [9] applied G2, B4, EX1 and GP in the Peninsular of Malaysia, but the other potential distribution, which was commonly used in the rainfall studies in Malaysia (e.g: Zalina *et al*., [8]; Wan Zin *et al*. [3]) were not evaluated for the rainfall event. In order to estimate the parameters of each distribution, the MLE method is applied for almost distributions with 100 iterations and accuracy of 1×10^{-5}. In addition, the Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D) tests, and the Q-Q plot were applied for the goodness-of-fit. The results reveal that GP, GEV, LP3 and WKB become a favorite distribution using K-S and A-D test. Since this study deals with the event characteristics, special consideration must be given for the probabilistic structure especially in the upper tail of the distributions by using the Q-Q plot. Wilks [4] also supported that this plot is able to verify the pattern of distribution compared to other approaches (e.g: cumulative distribution function or pdf). Although LP3 is listed in the highest rank in the K-S and AD tests, this distribution is not fitted very well and tends to show a larger bias of quantile estimation. In summary, WKB, GP and GEV seemed fitted using this plotting, which consistently give very high or low estimation of the quantiles. The result of the Q-Q plot is also verified using the performance of indictors, namely rmse and *r*.

From the good-ness-fit, the study is able to conclude that the WKB, GP and GEV distributions are consistently fitted for every assessment. This result is also consistent with the result by Dan'Azumi *et al*. [9], in which it is revealed that GP is suitable to be applied in the Peninsular of Malaysia. The GEV distribution is also found to have the potential to be applied in the storm event characteristics, similar as the other rainfall analysis in the Peninsular of Malaysia such as annual rainfall (e.g: Wan Zin *et al*. [3]) and extreme rainfall (e.g. Zalina *et al*. [8]).

5.0 CONCLUSION

This study presents the identification of the "best" distribution model to represent the rainfall event characteristics based on IETD of 6 hours in the selected regions in the Peninsular Malaysia. The long-term hourly rainfall data series from the current period (2001-2012, except for Johor Bahru, which is 2002-2012) were applied and converted to the series of rainfall event characteristics particularly *t*, *v*, *i* and *b* series of events. Those characteristics were fitted using B4, EX1, G2, GEV, GP, LP3, and WKB.

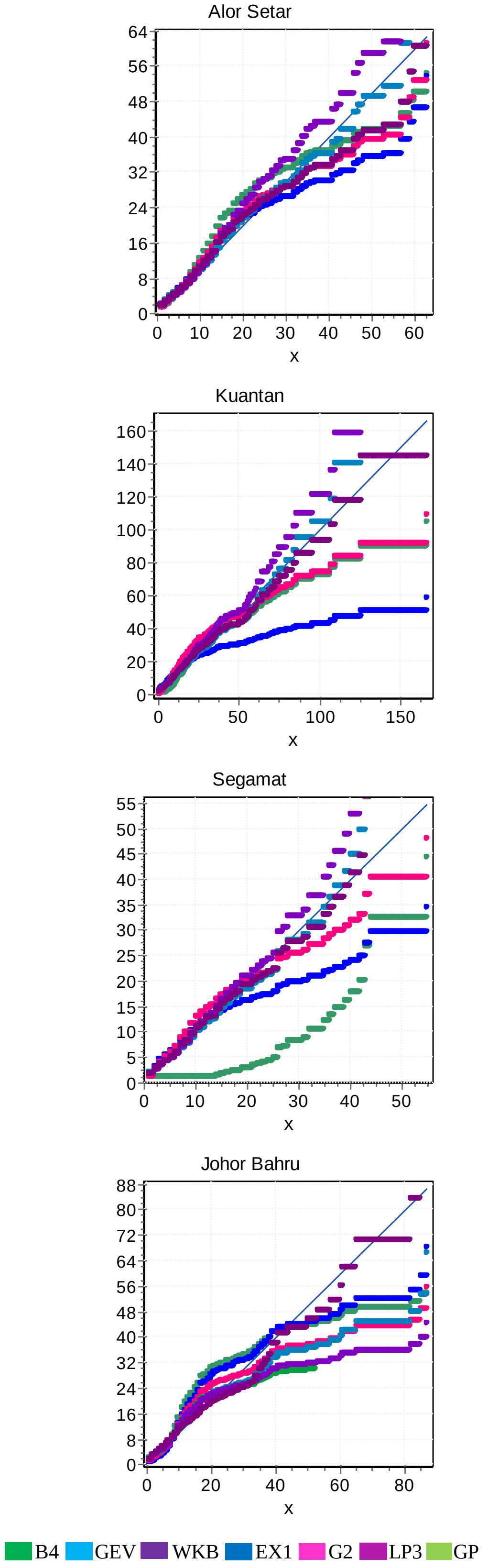


Figure 4 Q-Q plot for *t*

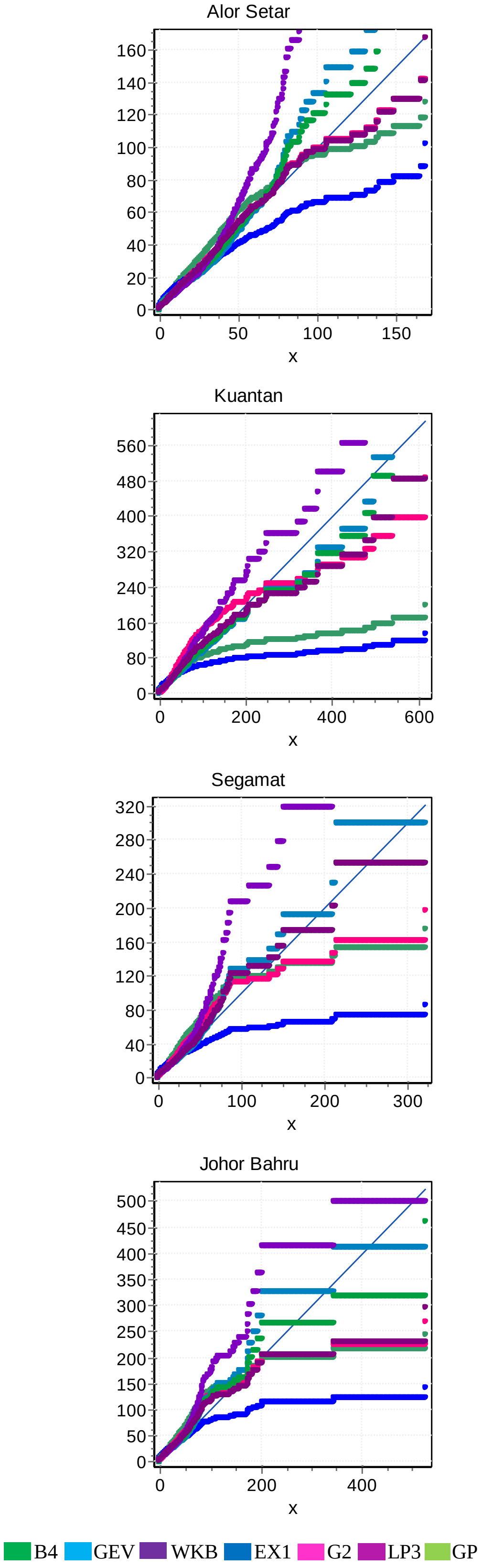


Figure 5 Q-Q plot for *v*

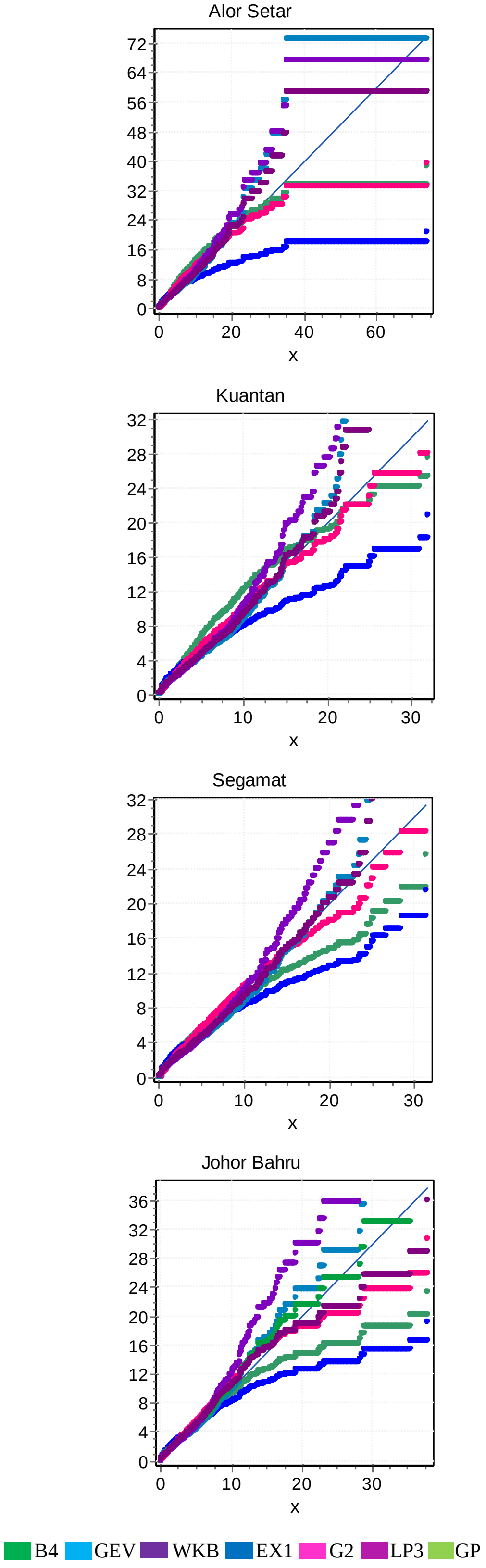


Figure 6 Q-Q plot for *i*

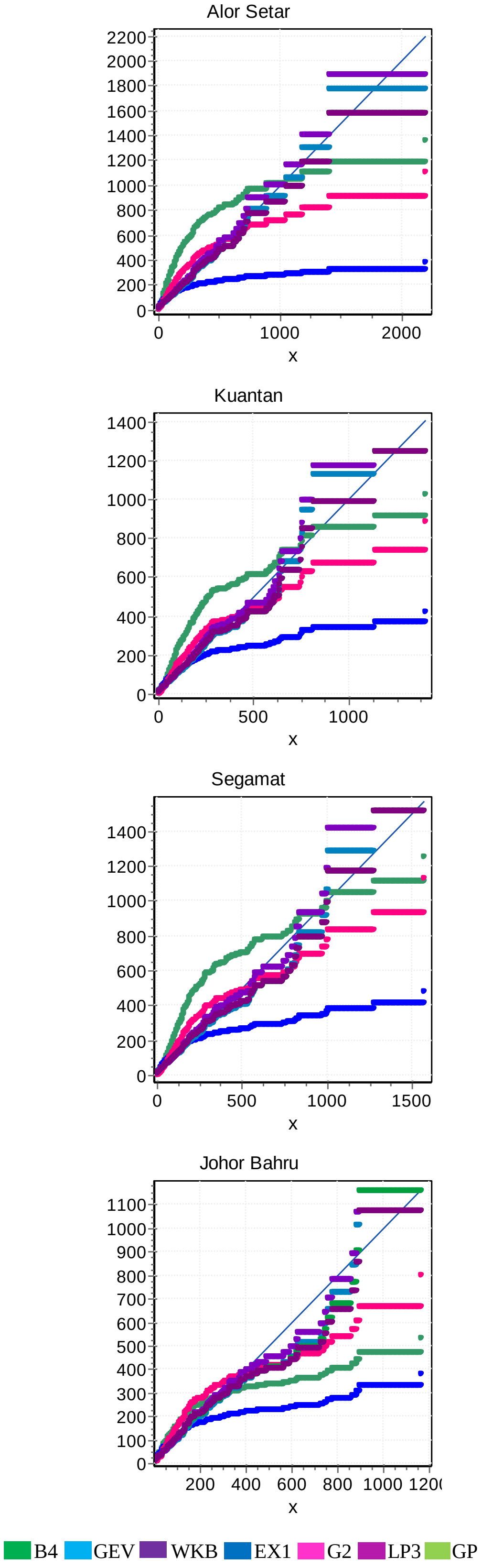


Figure 7 Q-Q plot for *b*

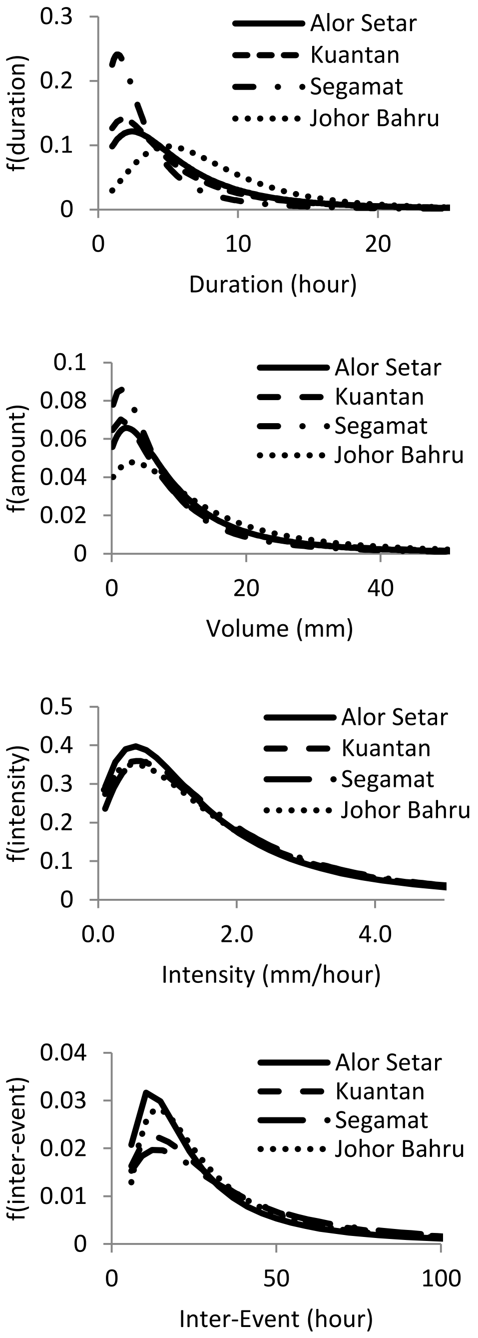


Figure 8 The PDF curve of the GEV distribution models for the storm event characteristics

Based on the results (K-S, A-D and Q-Q plot), WKB is the favorite distribution to describe all rainfall event characteristics. This is followed by the GP and GEV distributions.

The result of parameters of PDF in this study will be extremely useful for the urban storm water modeling, especially for analytical probabilistic approaches. The result will become preliminary results of an ongoing study.

Acknowledgements

The first author is grateful to Universiti Malaysia Perlis and the Ministry of Education Malaysia for a study leave in Universiti Teknologi Malaysia. Thanks to the Department of Irrigation and Drainage Malaysia for providing the data and technical support.

References

[1] Draper, A. J., M. W. Jenkins, K. W. Kirby, J. R. Lund, R. E. Howitt. 2003. Economic-engineering Optimization for California Water Management. *J. Water Resour. Plan. Manage.-ASCE*. 129(3): 155-164.

[2] Jamaludin, S., A. A. Jemain. 2007. Fitting the Statistical Distributions to the Daily Rainfall Amount in Peninsular Malaysia. *Jurnal Teknologi*. 46(C): 33-48.

[3] Wan Zin, W. Z., A. A. Jemain, K. Ibrahim. 2008. The Best Fitting Distribution of Annual Maximum Rainfall in Peninsular Malaysia Based on Methods of L-moment and LQ-moment. *Theoretical and Applied Climatology*. 96(3-4): 337-344. doi:10.1007/s00704-008-0044-2.

[4] Wilks, D. S. 2011. *Statistical Methods in the Atmospheric Sciences*. 3. Ed. San Diego: Academic Press.

[5] Blain, G. C., M. B. P. de Camargo. 2012. Probabilistic Structure of an Annual Extreme Rainfall Series of a Coastal Area of the State of São Paulo. *Brazil. Engenharia Agrícola*. 32(3): 552-559.

[6] Rosenberg, E. A., P. W. Keys, D. B. Booth, D. Hartley, J. Burkey, A. C. Steinemann, D. P. Lettenmaier. 2010. Precipitation Extremes and the Impacts of Climate Change on Stormwater Infrastructure in Washington State. *Clim Change*. 102: 319-349.doi:10.1007/s10584-010-9847-0.

[7] Fadhilah, Y., M. D. Zalina, V. T. V. Nguyen, S. Suhaila, Y. Zulkifli. 2007. Fitting the Best-Fit Distribution for the Hourly Rainfall Amount in the Wilayah Persekutuan. *Jurnal Teknologi*. 46(C): 49-58.

[8] Zalina, M. D., V. T. V. Nguyen, M. K. Amir Hashim M. D. Mohd Nor. 2002. Selecting a Probability Distribution for Extreme Rainfall Series in Malaysia. *Water Science Technology Journal*. 45(2): 63-68.

[9] Dan'azumi, S., S. Shamsudin, A. Aris. 2010. Modeling the Distribution of Rainfall Intensity Using Hourly Data. *American Journal of Environmental Sciences*. 6(3): 238-243.

[10] Eagleson, P. S. 1978. Climate, Soil, And Vegetation: 2. The Distribution of Annual Precipitation Derived from Observed Storm Sequences. *Water Resour. Res.* 14(5): 713-721.

[11] Eagleson, P. S. 1972. Dynamics of Flood Frequency. *Water Resour. Res.* 8(4): 878-898.

[12] Howard, C. D. D. 1976. Theory of Storage and Treatment Plant Overflows. *J. Environ. Eng.* 102(EE4): 709-722.

[13] Adams, B. J., J. B. Bontje. 1984. Microcomputer Applications of Analytical Models For Urbanstormwater Management, in Emerging Computer Techniques In Stormwater and Flood Management. *Am. Soc. Civ. Eng.* New York.

[14] Adams, B. J., H. G. Fraser, C. D. D. Howard, M. S. Hanafy. 1986. Meteorologic Data Analysis for Drainage System Design. *J. Environ. Eng*. 112(5): 827-848.

[15] Morgan, E. C., M. Lackner, R. M. Vogel, L. G. Baise. 2011. Probability Distributions for Offshore Wind Speeds. *Energy Conversion and Management*. 52: 15-26.

[16] Adams, B. J., F. Papa. 2000. *Urban Storm Water Management Planning with Analytical Probabilistic Models*. 1st Edn. John Wiley and Sons, New York. 53-79.

[17] Guo, Y., B. J. Adams. 1998. Hydrologic Analysis of Urban Catchments with Event-based Probabilistic Models 1 .Runoff Volume. *Water Resources Research*. 34(12): 3421-3431.

[18] Myung, I. J. 2003. Tutorial on Maximum Likelihood Estimation. *Journal of Mathematical Psychology*. 47: 90-100.

[19] Burgueno, A., B. Codina, A. Redafio, J. Lorente. 1994. Basic Statistical Characteristics of Hourly Rainfall Amounts in Barcelona (Spain). *Theor. Appl. Climatol.* 49: 175-181.

[20] Dunkerley. D. 2008. Identifying Individual Rain Events from Pluviograph Records: A Review with Analysis of Data from an Australian Drylandsite. *Hydrol. Process.* 22: 5024-5036.

[21] Dan'azumi, S., S. Shamsudin, A. Aris. 2013. Development of Analytical Probabilistic Model Parameters for Urbanstormwater Management. *Sains Malaysiana.* 42(3): 325-332.

[22] Dan'azumi, S., S. Shamsudin, A. Aris. 2013. Optimization of Pollution Control Performance of Wet Detention Ponds in Tropical Urban Catchments Using Particle Swarm Optimization. *Journal of Hydroinformatics*. 15(2): 529-539. doi:10.2166/hydro.2012.206.

[23] Shamsudin, S., S. Dan'azumi, A. Aris, Z. Yusop. 2014. Optimum Combination of Pond Volume and Outlet Capacity of a Stormwater Detention Pond Using Particle Swarm Optimization. *Urban Water Journal*. 11(2): 127-136. doi: 10.1080/1573062X.2013.768680.

[24] Suhaila, J., S. Mohd Deni, W. Z. Wan Zin, A. A. Jemain. 2010. Trends in Peninsular Malaysia Rainfall Data During the Southwest Monsoon and Northeast Monsoon Seasons, 1975–2004. *Sains Malaysiana*. 39(4): 533-542.