Jurnal Teknologi, 46(C) Jun 2007: 33–48

FITTING THE STATISTICAL DISTRIBUTIONS TO THE DAILY RAINFALL AMOUNT IN PENINSULAR MALAYSIA

SUHAILA JAMALUDIN[1] & ABDUL AZIZ JEMAIN[2]

Abstract. Daily rainfall data have been classified according to four rain types of sequence of wet days (Type 1, 2, 3 and 4). The Gamma, Weibull, Kappa and Mixed Exponential are the four distributions that have been tested to fit daily rainfall amount in Peninsular Malaysia. Parameters for each distribution were estimated using the maximum likelihood method. The selected model is chosen based on the minimum error produced by seven goodness of-fit (GOF) tests namely the median of absolute difference (MAD) between the empirical and hypothesized distributions, the traditional Empirical Distribution Function (EDF) Statistics which include Kolmogorov-Smirnov statistic D, Anderson Darling statistic A^2 and Cramer-von-Mises statistic W^2 and the new method of EDF Statistics based on likelihood ratio statistics. Based on these goodness-of-fit tests, the Mixed Exponential is found to be the most appropriate distribution for describing the daily rainfall amount in Peninsular Malaysia.

Keywords: Dairy rainfall amount, goodness-of-fit test, mixed exponential

Abstrak. Data hujan harian dibahagikan kepada empat jenis rentetan hujan (jenis 1, 2, 3 dan 4). Taburan Gamma, Weibull, Kappa dan Gabungan Eksponen ialah empat taburan statistik yang diuji dalam memadankan data jumlah hujan harian di Semenanjung Malaysia. Parameter bagi setiap taburan dianggar dengan menggunakan kaedah kebolehjadian maksimum. Model dipilih berdasarkan nilai ralat yang minimum terhasil dari tujuh ujian kesesuaian model iaitu median bagi perbezaan nilai mutlak antara taburan empirik dengan taburan yang diuji, statistik fungsi empirik iaitu Kolmogorov-Smirnov D, Anderson Darling A^2 dan Cramer-von-Mises W^2 serta kaedah baru statistik fungsi empirik yang berasaskan kepada ujian nisbah kebolehjadian. Berdasarkan nilai ujian kesesuaian model, didapati taburan Gabungan Eksponen adalah yang paling sesuai dalam memadankan data jumlah hujan harian di Semenanjung Malaysia.

Kata kunci: Jumlah hujan harian, ujian kesesuaian model, gabungan eksponen

1.0 INTRODUCTION

Water is one of the vital natural resources. It is the first requirement for survival. It plays an important role for agriculture, industry and daily domestics. Shortage of water supply will cause tremendous negative impact to the country. However, excessive water

[1] Mathematics Department, Faculty of Science, Universiti Teknologi Malaysia, 81310, Skudai, Johor. Tel: 07-5534317, Fax: 07-5566162, Email: suhaila@mel.fs.utm.my
[2] School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600, Bangi, Selangor. Tel: 03-89215712, Fax: 03-89256086, Email: azizj@pkrisc.cc.ukm.my

supply could also contribute to natural disasters such as landslides and floods. Some diseases are also related to improper water supply and drainage management. Therefore, it is crucial to manage water resources in an optimum manner since it has major implications on the country's economic prosperity. In order to manage the water resources, we need to understand and identify the characteristics of rainfall since it is a major source of input in water resource system.

Modeling of daily rainfall using various mathematical models has been carried out throughout the world to develop a better understanding about the rainfall pattern and its characteristics. Several methods have been presented in the literature for modeling rainfall amount on wet days. There are two common approaches in analyzing the rainfall amount. In the first approach, it is assumed that the daily rainfalls are independent and a theoretical distribution is fitted. In the second approach, it is assumed that the amount is independent but its distribution depends on whether the previous day was wet or dry. This approach is known as a process of chain dependent. A comprehensive literature on chain dependent process was written by Katz [1]. The most widely used theoretical distribution is a two-parameter Gamma distribution as described by Buishand [2], Ison *et al.* [3] and May [4]. Some other distributions are the Exponential (Todorovic and Woolhiser [5]), the Kappa (Mielke [6]), the S_B (Llyod and Schreuder [7]), the Mixed Exponential (Woolhiser and Roldan [8]; Chapman [9,10]; Wilks [11]), the Weibull (Sharda and Das [12]) and the Skew Normal (Chapman [9,10]).

In studies carried out by Cole and Sheriff [13], Buishand [2] and Chapman [9, 10], daily rainfall data are classified according to the number of adjacent wet days (0, 1 and 2). There are three different types of wet days, namely solitary or single wet days, first day or last day of a wet spell and the interior of wet spell. Buishand [2] found that the mean for these three types of wet days are significantly different and it increased with the number of adjacent wet days. Meanwhile, Chapman [9, 10] has proven that the distributions of rainfall amount depend strongly on the number of adjoining wet days, and these models performed consistently better than the models which do not consider these differences. Chapman [9, 10] insists that serious errors might result from these models.

However, in Malaysia these kinds of studies received less attention. The studies that have been conducted in Malaysia focused more on the general aspects such as pattern, trend and variability of rainfall (Dale [14] and Chia [15]). Most of the data are outdated and not analyzed comprehensively. Shaharuddin [16] who studied the trend and variability of rainfall in Malaysia only considered the simple descriptive statistics such as mean, standard deviation and coefficient of variation in his study.

Compared to the study done by Zalina *et al.* [17] who investigated the distribution of extreme rainfall series over 17 rain gauge stations in Peninsular Malaysia, more advanced statistical techniques such as the L-moment method to estimate the parameters while the Probability Plot Correlation Coefficient (PPCC) test, the Root Mean Squared Error (RMSE), the Relative Root Mean Squared Error (RRMSE) and the Maximum

Absolute Error (MAE) were used as the goodness-of-fit (GOF) tests to determine the best fit model. In the study, Generalised Extreme Value (GEV) was found to be the best distribution that fit the annual maximum rainfall for hourly data.

The selection of the best fitting distribution has always been a key interest in the studies of rainfall amount. Thus, in this study we would like to find the best fitting distribution for daily rainfall amount based on several criteria of GOF tests and to determine whether they came from the same probability distributions. In this study, daily rainfall data are separated into four different types of wet days. Type 1 rain is referred as a single wet day, Type 2 rain is for the first day of wet spell, Type 3 rain represents the second day of wet spell and finally, Type 4 rain is for the third day of wet spell. Assumptions of independence and serial correlation are taken into account since the daily rainfalls have been divided into those four types of wet days. Besides, the results from autocorrelation already proved that the daily rainfall series are independent and serially uncorrelated.

The new method of GOF tests based on likelihood ratio statistics developed by Zhang [18, 19] will be employed together with the traditional GOF tests in finding the best fitting distribution. Referring to Zhang's approach, these new GOF tests are more powerful than the traditional GOF. Other additional criterion included in the analysis is the median of absolute difference between the hypothesized and the empirical distribution. The best distribution will be chosen based on the minimum error specified by all those criteria.

2.0 CASE STUDY

Malaysia is situated in the tropics between 1° and 6° degrees north of the equator. It consists of two major regions; West Malaysia, which is known as Peninsular Malaysia, and East Malaysia (Sabah and Sarawak). In general, Malaysia experiences wet and humid tropical climate throughout the year that is characterized by high annual rainfall, humidity and temperature. Malaysia has uniform temperature throughout the year with temperature between 25.5° to 32° Celsius. Approximately, the annual rainfall amount is between 2000 mm and 4000 mm with annual number of wet days ranging from 150 to 200 days. The distribution of rainfall in Malaysia is very much affected by two types of monsoon, the Northeast monsoon and Southwest monsoon. During Northeast monsoon (Nov-Mac), the northeastern coasts of Malaysia received heavy rainfalls meanwhile Southwest monsoon (May-Sept) bring heavy rainfalls to the west coast region of Malaysia.

Daily rainfall series data for this study have been obtained from the Malaysian Meteorology Department for the duration of 35 years. For this study, five rain gauge stations were chosen based on the completeness of the data. The information about the stations is given in the Table 1. The rain gauge stations are selected to represent rainfall pattern for the whole Peninsular Malaysia. The rain gauge stations for Bayan

Table 1 Latitude and longitude of the rain gauge stations

Station no.	Stations	Latitude	Longitude
48601	Bayan Lepas	5°18'N	100°16'E
48615	Kota Bharu	6°10'N	102°17'E
48657	Kuantan	3°47'N	103°13'E
48674	Mersing	2°27'N	103°50'E
48647	Subang	3°07'N	101°33'E

Lepas and Subang are in the West Cost of Peninsular Malaysia while stations for Kota Bharu, Mersing and Kuantan are in the East Cost of Peninsular Malaysia.

A wet day is defined as a day with a rainfall amount above some threshold δ (Buishand [2]). The threshold was fixed at 0.1 mm. Thus, the data that was considered in this analysis are ≥ 0.1 mm.

3.0 MODELING RAINFALL AMOUNT

Four models for daily rainfall amount are tested with their probability density functions are given as follows. Note that the random variable X represents the daily rainfall amount.

(i) The Gamma distribution with two parameters α and β denote the shape and scale parameters respectively.

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

(ii) The Weibull distribution with two parameters α and β denote the shape and scale parameters respectively.

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right], \qquad \alpha > 0, \quad \beta > 0, \quad x > 0 \tag{2}$$

(iii) The Kappa distribution with two parameters α and β denote the shape parameter and scale parameter respectively.

$$f(x) = \left(\frac{\alpha}{\beta}\right)\left[\alpha + \left(\frac{x}{\beta}\right)^{\alpha}\right]^{-(\alpha+1)/\alpha}, \qquad \alpha > 0, \quad \beta > 0, \quad x > 0 \tag{3}$$

(iv) The Mixed Exponential distribution with three parameters is the mixture of two one-parameter exponential distributions where p denotes the mixing probability which determines the weights given to the two exponential distributions with scale parameters β_1 and β_2.

$$f(x) = \left(\frac{p}{\beta_1}\right)\exp\left[\frac{-x}{\beta_1}\right] + \left(\frac{1-p}{\beta_2}\right)\exp\left[\frac{-x}{\beta_2}\right], \quad 0 \le p \le 1,\ \beta_1 > 0,\ \beta_2 > 0,\ x > 0 \quad (4)$$

These distributions are selected for the analysis due to the fact that they are commonly used in modeling rainfall amount such as studies done by Woolhiser and Roldan [8], Coe and Stern [20], Chapman [9, 10], Wilks [11], May [4], Sharda and Das [12].

4.0 GOODNESS-OF-FIT TESTS (GOF)

Seven criteria of GOF tests have been used in this study to identify the best fit models. The tests are based on the degree of similarity between the empirical distribution $F_n(x_{(i)})$ and the hypothesized distribution $F(x_{(i)}, \hat{\theta})$. The chosen distribution that fit the daily rainfall amount is based on the minimum error indicated by all these seven tests.

(i) First criteria involve the median absolute difference (MAD) between the hypothesized distribution $F(x_{(i)}, \hat{\theta})$ and the empirical distribution function $F_n(x_{(i)})$. The formula is given as below:

$$\text{MAD} = \text{Med}\,|\,F_n(x_{(i)}) - F(x_{(i)}, \hat{\theta})\,| \quad (5)$$

$x_{(i)}$ represents the ordered data meanwhile $F(x_{(i)}, \hat{\theta})$ with $\hat{\theta}$ represents the vector of estimated parameters. In addition, the empirical distribution function is given as $F_n(x_{(i)}) = \dfrac{i}{n+1}$ with n is the number of sample size.

(ii) Next, we will discuss the traditional GOF tests. The Kolmogorov-Smirnov (K-S), Cramer-von-Mises (CvM) and Anderson Darling (AD) test are among the tests that have been selected in our analysis. After using the Probability Integral Transformation as discussed by Stephens [21], the computing formulas for those three tests are as given below.

(a) K-S test calculates the maximum difference between the hypothesized distribution $Z_{(i)} = F(x_{(i)}, \hat{\theta})$ and the empirical distribution function, $F_n(x_{(i)})$ with $x_{(i)}$ representing the ordered data.

$$D^{+} = \max_i \{i/n - Z_{(i)}\}, \qquad D^{-} = \max_i \{Z_{(i)} - (i-1)/n\}$$
$$D = \max(D^{+}, D^{-}) \tag{6}$$

(b) CvM test calculates the squared difference between the hypothesized distribution $Z_{(i)} = F\left(x_{(i)}, \hat{\theta}\right)$ and the empirical distribution function, $F_n\left(x_{(i)}\right)$ with $x_{(i)}$ representing the ordered data.

$$W^2 = \sum_{i=1}^{n} \{Z_{(i)} - (2i-1)/2n\}^2 + \frac{1}{12n} \tag{7}$$

(c) AD test calculates the weighted squared difference between the hypothesized distribution $Z_{(i)} = F\left(x_{(i)}, \hat{\theta}\right)$ and the empirical distribution $F_n\left(x_{(i)}\right)$ where the weight function is $\left[F\left(x_{(i)}, \hat{\theta}\right)\left(1 - F\left(x_{(i)}, \hat{\theta}\right)\right)\right]^{-1}$ with $x_{(i)}$ representing the ordered data.

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}\left[(2i-1)\log Z_{(i)} + (2n+1-2i)\log\left(1 - Z_{(i)}\right)\right] \tag{8}$$

(iii) The new powerful GOF tests are based on the likelihood ratio statistics between the hypothesized distribution $Z_{(i)} = F\left(x_{(i)}, \hat{\theta}\right)$ and the empirical distribution function $F_n\left(x_{(i)}\right)$. Using Zhang's approach, the computed formulas are given as.

(a) New Kolmogorov-Smirnov test

$$Z_K = \max_{1 \le i \le n}\left[(i-0.5)\log\frac{i-0.5}{nZ_{(i)}} + (n-i+0.5)\log\frac{n-i+0.5}{n\left(1 - Z_{(i)}\right)}\right] \tag{9}$$

(b) New Cramer-von-Mises test

$$Z_C = \sum_{i=1}^{n}\left[\log\frac{Z_{(i)}^{-1} - 1}{(n-0.5)/(i-0.75) - 1}\right]^2 \tag{10}$$

(c) New Anderson-Darling test

$$Z_A = -\sum_{i=1}^{n}\left[\frac{\log Z_{(i)}}{n-i+0.5} + \frac{\log\left(1 - Z_{(i)}\right)}{i-0.5}\right] \tag{11}$$

5.0 RESULTS AND DISCUSSION

First, we would like to discuss the result of descriptive statistics for each rain type for five rain gauge stations. Next, we will proceed with the discussion on fitting distribution.

5.1 Descriptive Statistics

A summary of daily rainfall amount of four types of wet days is provided in Table 2. The differences among the stations can be compared through the mean, standard deviation and coefficient of variations of daily rainfall. The results clearly showed that the standard deviation for each rain gauge stations is always larger than the mean. This is due to the fact that there are some large values in the dataset which could possibly be affected by the extreme values. Therefore, the shape of the rainfall distribution for each rain gauge station is skewed to the right as shown in Figure 1(a), 1(b), 1(c) and 1(d).

In addition, the mean rainfall amount increases as the number of days of wet spell increases. Among the five rain gauge stations, Kuantan station showed the highest mean rainfall amount for Type 1 and Type 2 rain. On the other hand, Kota Bharu station has the highest mean rainfall amount for Type 3 and Type 4 rain. For Type 4 rain, we could see that the stations located in the east coast of Peninsular Malaysia received high mean rainfall amount compared to those stations in the west coast. In addition, the standard deviations for these stations also indicate a large variation in their daily rainfall amount series. These results however could be explained by the

Table 2 Statistics of daily rainfall amount for five rain gauge stations

Type 1 stations	Mean	St.dev	CV (100%)	Type 2 stations	Mean	St.dev	CV (100%)
Bayan Lepas	10.255	15.754	154	Bayan Lepas	12.3181	17.7235	144
Subang	9.233	13.156	143	Subang	12.583	16.700	133
Kota Bharu	8.313	11.577	140	Kota Bharu	11.162	16.8571	151
Kuantan	10.766	16.924	157	Kuantan	12.86	21.664	169
Mersing	9.513	13.950	147	Mersing	11.293	16.993	151
Type 3 stations	**Mean**	**St.dev**	**CV (100%)**	**Type 4 stations**	**Mean**	**St.dev**	**CV (100%)**
Bayan Lepas	13.307	19.124	144	Bayan Lepas	15.224	22.180	146
Subang	11.770	16.646	141	Subang	13.736	18.409	134
Kota Bharu	14.448	24.507	170	Kota Bharu	22.317	42.267	190
Kuantan	14.384	21.704	151	Kuantan	19.845	37.030	187
Mersing	12.702	21.861	172	Mersing	18.688	34.098	183

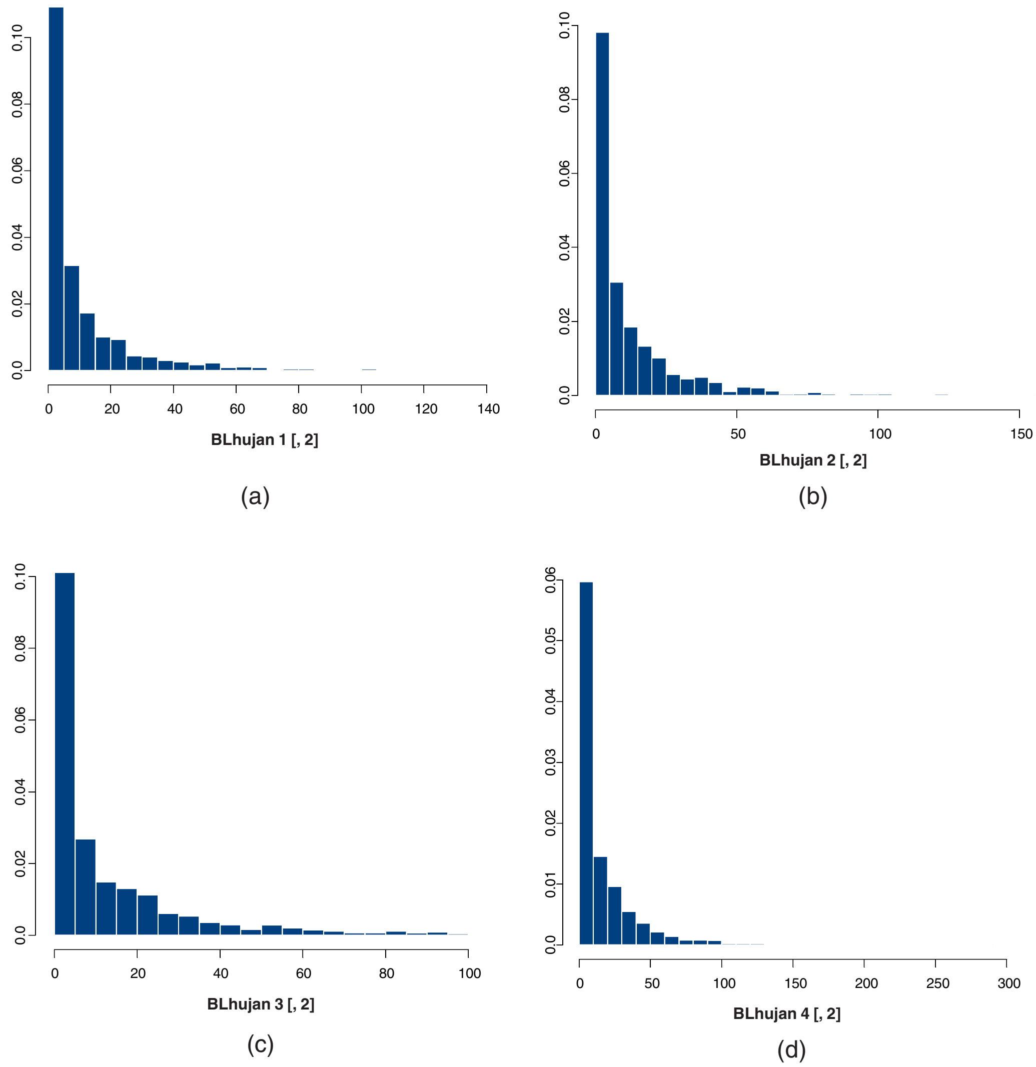

Figure 1 Histograms of daily rainfall amount in Bayan Lepas station for (a) Type 1 rain, (b) Type 2 rain, (c) Type 3 rain and (d) Type 4 rain

influences of the monsoons. The coasts that are exposed to the northeast monsoon in Malaysia tend to be wetter than those exposed to the southwest monsoon since they received heavy rainfall during the time. These possibilities could explain the pattern of the outcomes.

The irregularity of the daily rainfall between stations is represented by the coefficient of variation (CV) which is evident in all cases that the 100% is clearly exceeded. Again for Type 4 rain, the results show a major difference between stations located in the east coast and west coast of Peninsular Malaysia. Stations in the east coast shown high variability of rainfall amount which ranged between 180% and 190% compared to the other two stations in the west coast that ranged between 130% and 140%. These results give strong indication that those stations might not have the same rainfall pattern or distribution.

5.2 Fitting Distributions

The values of seven goodness-of-fit criteria have been calculated for each rain type and are shown in Tables 3 to 6. The best distribution Bold Type in Tables 3-6 was chosen based on the minimum error of GOF tests. The distributions were then ranked in ascending order based on those values. Unfortunately, when many criteria are used to identify the best distribution, it is more difficult to make the selection decision. The

Table 3 Values of seven goodness-of-fit tests for Type 1 rain

Stations	**Distribution**	**MAD** $\times 10^{-4}$	D $\times 10^{-4}$	W^2 $\times 10^{-2}$	A^2 $\times 10^{-1}$	Z_K	Z_C	Z_A $\times 10^{-2}$
Bayan Lepas	Gamma	320.46	824.76	145.43	87.11	55.96	181.15	356.85
	Weibull	125.63	618.36	54.47	40.96	48.27	132.59	351.08
	Kappa	326.43	530.33	96.76	69.43	22.50	179.04	349.92
	Mixed Exp	99.33	328.52	18.88	17.57	29.18	91.53	341.50
Subang	Gamma	303.53	723.94	150.74	88.61	53.59	175.78	353.47
	Weibull	163.83	496.48	60.41	43.71	46.90	139.80	349.24
	Kappa	259.73	536.21	85.32	63.11	26.84	181.36	350.33
	Mixed Exp	67.76	300.28	9.96	11.53	28.82	79.50	339.81
Kota Bharu	Gamma	341.36	628.29	111.84	74.27	53.55	170.27	354.80
	Weibull	150.04	540.70	47.69	41.58	46.92	143.88	350.95
	Kappa	311.13	809.50	104.56	82.62	30.08	201.90	353.69
	Mixed Exp	113.55	497.45	25.04	22.87	32.33	98.38	342.75
Kuantan	Gamma	391.14	893.29	226.48	127.36	63.82	219.36	359.93
	Weibull	186.01	599.81	88.47	61.04	55.26	161.03	353.67
	Kappa	259.88	585.69	84.53	62.29	26.15	185.81	351.10
	Mixed Exp	81.97	347.88	14.87	14.43	30.14	86.04	340.71
Mersing	Gamma	318.42	758.53	151.04	92.56	65.69	189.46	358.93
	Weibull	200.23	560.02	70.90	53.74	56.57	156.35	353.59
	Kappa	362.52	608.20	123.14	94.00	29.40	233.31	355.25
	Mixed Exp	82.25	386.98	20.11	18.67	38.02	102.40	344.12

Table 4 Values of goodness-of-fit tests for Type 2 rain

Stations	**Distribution**	**MAD** $\times 10^{-4}$	D $\times 10^{-4}$	W^2 $\times 10^{-2}$	A^2 $\times 10^{-1}$	Z_K	Z_C	Z_A $\times 10^{-2}$
Bayan Lepas	Gamma	257.14	521.61	119.41	74.92	60.14	187.07	350.32
	Weibull	115.22	423.16	40.57	34.03	50.46	138.78	345.55
	Kappa	285.33	651.70	134.38	103.83	29.97	236.79	348.03
	Mixed Exp	124.98	318.51	25.39	20.00	29.21	92.93	337.94
Subang	Gamma	151.56	505.05	80.75	57.26	59.08	167.99	347.96
	Weibull	126.11	477.86	49.70	39.28	48.50	147.86	344.33
	Kappa	338.57	734.47	197.85	154.14	42.48	307.21	351.85
	Mixed Exp	95.35	258.08	16.77	16.39	34.73	95.26	338.84
Kota Bharu	Gamma	322.37	750.27	168.79	97.27	54.31	208.10	351.33
	Weibull	140.34	523.24	60.99	44.47	46.96	142.27	346.60
	Kappa	246.56	573.50	96.75	69.12	23.47	182.00	344.53
	Mixed Exp	64.04	289.12	13.46	12.47	26.56	104.18	338.03
Kuantan	Gamma	265.50	649.00	140.50	87.58	64.81	301.65	353.41
	Weibull	76.32	466.88	52.48	41.97	54.30	176.55	347.68
	Kappa	340.20	614.46	168.26	117.73	29.26	234.24	345.18
	Mixed Exp	76.64	273.36	16.66	16.39	33.22	352.26	341.31
Mersing	Gamma	235.29	596.87	127.00	86.44	77.11	232.74	356.34
	Weibull	141.87	522.36	63.75	52.28	64.83	172.75	350.48
	Kappa	368.82	689.23	205.59	153.68	35.28	303.54	352.40
	Mixed Exp	104.74	342.23	31.92	28.40	45.37	165.39	344.03

Table 5 Values of goodness-of-fit tests for Type 3 rain

Stations	**Distribution**	**MAD** $\times 10^{-4}$	D $\times 10^{-4}$	W^2 $\times 10^{-2}$	A^2 $\times 10^{-1}$	Z_K	Z_C	Z_A $\times 10^{-2}$
Bayan Lepas	Gamma	251.01	734.82	169.51	97.19	66.55	210.79	353.13
	Weibull	147.44	446.58	63.39	44.41	55.29	166.96	348.55
	Kappa	373.39	520.35	113.59	92.75	46.55	253.61	353.61
	Mixed Exp	366.57	310.13	15.34	17.10	28.69	94.47	337.89
Subang	Gamma	378.76	913.79	330.83	176.76	59.04	224.14	349.80
	Weibull	163.21	516.90	75.44	55.89	60.25	174.51	348.54

Table 5 (Continues)

Stations	Distribution	MAD $\times 10^{-4}$	D $\times 10^{-4}$	W^2 $\times 10^{-2}$	A^2 $\times 10^{-1}$	Z_K	Z_C	Z_A $\times 10^{-2}$
	Kappa	364.86	658.89	193.61	143.18	41.34	301.41	353.70
	Mixed Exp	75.34	297.15	16.22	17.19	40.37	113.64	340.87
Kota Bharu	Gamma	299.73	693.12	171.42	101.70	57.25	281.67	354.15
	Weibull	111.76	460.73	39.33	34.54	48.00	150.54	347.25
	Kappa	257.64	521.91	90.83	69.28	18.23	170.52	342.74
	Mixed Exp	103.71	362.68	31.77	28.39	21.74	152.41	337.96
Kuantan	Gamma	177.77	519.99	99.41	69.82	69.04	204.89	353.42
	Weibull	166.65	440.82	58.39	46.22	56.03	152.02	347.58
	Kappa	369.61	707.57	218.77	167.77	36.69	312.02	351.83
	Mixed Exp	131.62	328.20	32.90	27.59	36.79	155.96	341.36
Mersing	Gamma	322.12	604.29	149.13	91.27	75.59	306.32	356.70
	Weibull	80.23	468.42	38.53	34.32	63.39	168.25	349.79
	Kappa	321.06	572.29	146.84	110.98	26.04	244.42	346.09
	Mixed Exp	118.46	346.38	37.09	30.54	34.36	281.13	341.59

Table 6 Values of goodness-of-fit tests for Type 4 rain

Stations	Distribution	MAD $\times 10^{-4}$	D $\times 10^{-4}$	W^2 $\times 10^{-2}$	A^2 $\times 10^{-1}$	Z_K	Z_C	Z_A $\times 10^{-2}$
Bayan Lepas	Gamma	247.69	475.18	186.54	120.89	120.99	336.92	353.67
	Weibull	135.98	386.98	78.26	65.29	97.99	255.70	347.92
	Kappa	354.05	619.55	314.93	258.77	75.36	540.25	352.33
	Mixed Exp	97.56	342.38	52.58	43.77	57.56	204.33	339.70
Subang	Gamma	175.55	537.51	201.69	128.86	122.58	334.22	350.07
	Weibull	154.00	462.02	116.51	84.96	100.31	282.02	345.75
	Kappa	352.61	650.79	389.02	300.69	81.63	607.92	352.88
	Mixed Exp	56.91	228.16	21.88	20.63	67.07	166.10	338.99
Kota Bharu	Gamma	428.87	665.35	424.30	246.74	119.85	582.05	360.52
	Weibull	161.49	458.32	73.15	67.75	95.63	285.56	350.64
	Kappa	256.94	497.77	178.45	132.38	36.38	316.95	344.23
	Mixed Exp	187.19	482.73	136.20	105.76	32.17	375.45	338.22

Table 6 (Continues)

Stations	Distribution	MAD $\times 10^{-4}$	D $\times 10^{-4}$	W^2 $\times 10^{-2}$	A^2 $\times 10^{-1}$	Z_K	Z_C	Z_A $\times 10^{-2}$
Kuantan	Gamma	393.45	647.86	462.79	269.76	146.82	711.53	361.19
	Weibull	132.14	430.08	69.81	68.25	118.56	333.91	351.38
	Kappa	232.68	501.34	202.02	160.21	41.20	387.29	344.63
	Mixed Exp	164.22	482.98	145.48	125.47	41.73	452.50	339.06
Mersing	Gamma	341.63	639.08	396.91	231.24	128.67	671.56	358.03
	Weibull	123.80	396.37	56.75	57.13	104.71	307.91	349.21
	Kappa	212.05	501.17	175.75	144.36	35.79	347.08	342.85
	Mixed Exp	146.91	482.00	125.60	112.32	36.40	380.85	337.94

selected statistical distribution for the same data may be different for different analysis. In this study, we chose the best fitting distribution based on the majority of the tests, since we did not investigate which is the most powerful test.

Referring to the results shown in Tables 3-6, the two most frequently selected distributions are the Mixed Exponential and Weibull. The Mixed Exponential provided a very good fit for Type 1, Type 2 and Type 3 rains. However, Type 4 rain showed slightly different pattern. Bayan Lepas station and Subang station in the west coast of Peninsular Malaysia could be fitted best by Mixed Exponential model, meanwhile the Weibull was selected as the best statistical distribution for stations Kuantan, Mersing and Subang in the east coast. These showed that different rain types have different statistical distributions for different regions. Overall results for each rain type are shown on the map of Peninsular Malaysia as given in Figure 2(a) to 2(d).

The mixture of two distributions is certainly better than single distributions in describing the daily rainfall amount in Peninsular Malaysia for Type 1, Type 2 and Type 3 rains. On the other hand, Type 4 rain which refers to the amount of the third day of wet spell is usually associated with heavy amount of rainfall. As we mentioned earlier, the stations in the east coast of Peninsular Malaysia tend to be wetter than those stations in the west coast since the influence of Northeast monsoon that are known to bring heavy rainfall to the east coast. That does explain the reason stations in the east coast are better described with single component meanwhile stations in the west coast of Peninsular Malaysia are better represented with two components for this type rain. In addition, the topographical, geographical and climatic changes in both regions also could influence the results on fitting distribution.

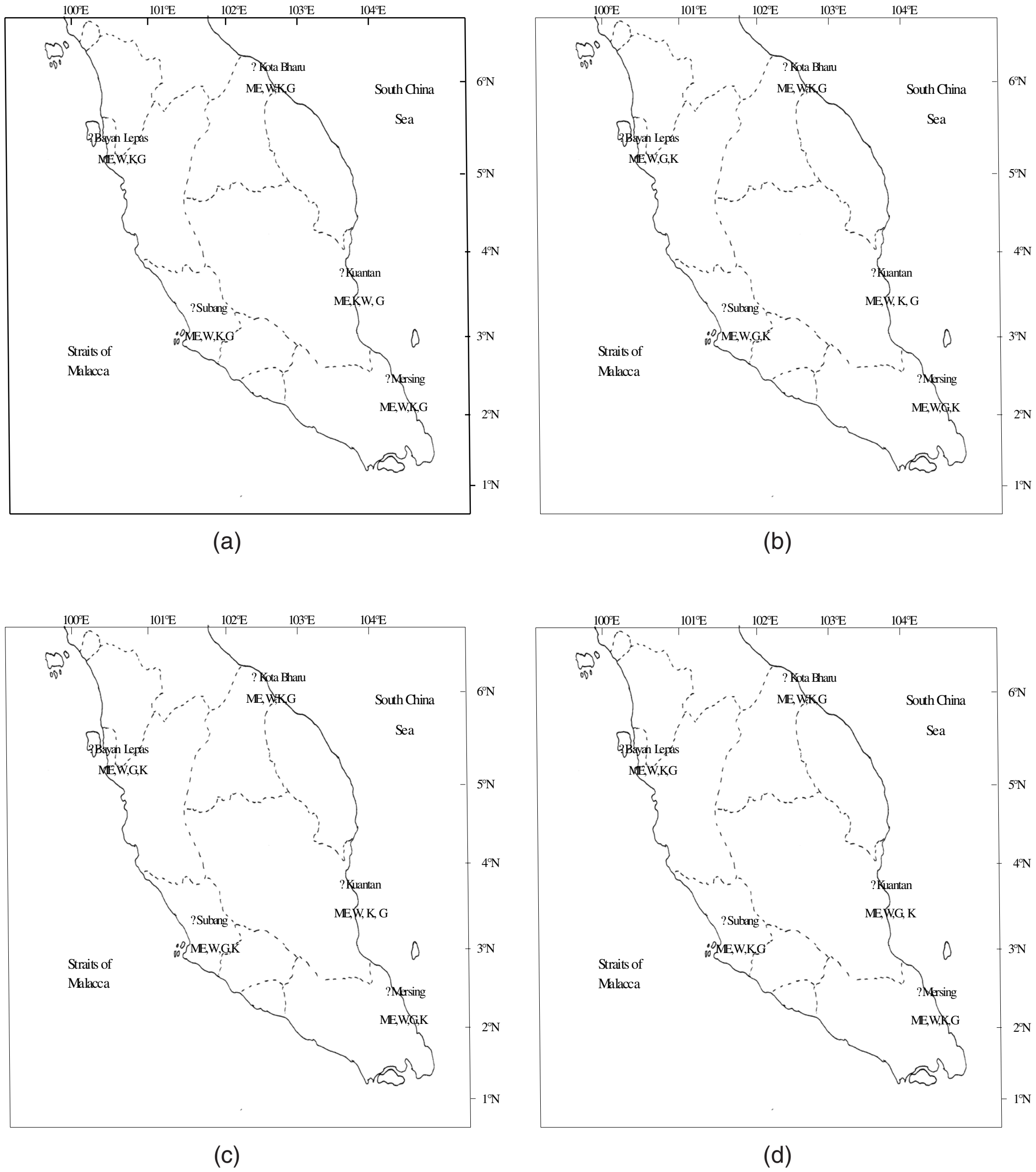

Figure 2 Ranking of distributions for (a) Type 1 rain (b) Type 2 rain (c) Type 3 rain (d) Type 4 rain for all five rain gauge stations

6.0 CONCLUSIONS

The search for the best distribution in fitting daily rainfall amount has been the main interest in several studies. Various forms of distributions have been tested in order to find the best fitting distribution. Different tests of goodness-of-fit have been attempted in the studies.

Daily rainfall amount in this study has been classified according to the four types of wet days. It has already shown that the mean daily rainfall amount increases as the number of days of wet spell increases. The Gamma, Weibull, Mixed Exponential and Kappa distributions have been tested on these dataset of four types of wet days. Results shown that, for five Malaysia rain gauge stations, the Mixed Exponential best describes the distribution of daily rainfall amount on the basis of the seven criteria of GOF tests. The Weibull and Kappa ranked second and third respectively.

The results of this study also shown that the mixture of two distributions is better than single distributions for describing the daily rainfall amount in Peninsular Malaysia particularly for Type 1 rain, Type 2 rain and Type 3 rain. However, there are differences between stations located in the west coast and east coast of Peninsular Malaysia for Type 4 rain. The single Weibull distribution has been chosen as the best model for daily rainfall amount in east coast stations while the mixture of two exponentials are still the best distribution for stations in the west coast. Despite of this, overall findings are still indicated that the daily rainfall amount in Peninsular Malaysia is very well described by using the mixture of two components. This indicated that the rainfall distribution in Peninsular Malaysia is consists of two components, most probably light and heavy rainfalls. We stress that further analysis must be carried out for all rain types by including more rain gauge stations and tested with other mixture of two distributions since daily rainfall amount for the majority sites in Malaysia is very well represented using two components.

ACKNOWLEDGEMENTS

The authors are thankful to the Malaysian Meteorology Department for providing daily rainfall data for this paper. The study would not have been possible without the sponsorship from Universiti Teknologi Malaysia.

REFERENCES

[1] Katz, R. W. 1977. Precipitation as a Chain-dependent Process. *Journal of Applied Meteorology*. 16: 671-676.

[2] Buishand, T. A. 1978. Some Remarks on the Use of Daily Rainfall Models. *Journal of Hydrology*. 36: 295-308.

[3] Ison, N. T., A. M. Feyerherm, and L. Dean Bark. 1971. Wet Period Precipitation and the Gamma Distribution. *Journal of Applied Meteorology*. 10: 658-665.

[4] May, W. 2004. Variability and Extremes of Daily Rainfall during the Indian Summer Monsoon in the Period 1901-1989. *Global and Planetary Change*. 44: 83-105.

[5] Todorovic, P. and D. A. Woolhiser. 1975. A Stochastic Model of n-day Precipitation. *Journal of Applied Meteorology*. 14(1): 17-24.

[6] Mielke, P. W. 1973. Another Family of Distributions for Describing and Analyzing Precipitation Data. *Journal of Applied Meteorology*. 12: 275-280.

[7] Llyod, S. W., Jr. and H. T. Schreuder. 1981. Fitting Daily Precipitation Amounts Using the S_B Distribution. *Monthly Weather Review*. 109(12): 2535-2540.

[8] Woolhiser, D. A. and J. Roldan. 1982. Stochastic Daily Precipitation Models 2. A Comparison of Distribution of Amounts. *Water Resources Research*. 18(5): 1461-1468.

[9] Chapman, T. G. 1997. Stochastic Models for Daily Rainfall in the Western Pacific. *Mathematics and Computers in Simulation*. 43: 351-358.

[10] Chapman, T. G. 1998. Stochastic Modelling of Daily Rainfall: The Impact of Adjoining Wet Days on the Distribution of Rainfall Amounts. *Environmental Modelling and Software*. 13: 317-324.

[11] Wilks, D. S. 1998. Multisite Generalization of a Daily Stochastic Precipitation Generation Model. *Journal of Hydrology*. 210: 178-191.

[12] Sharda, V. N. and P. K. Das. 2005. Modelling Weekly Rainfall Data for Crop Planning in a Sub-humid Climate of India. *Agricultural Water Management*. 76: 120-138.

[13] Cole, J. A. and J. D. F. Sheriff. 1972. Some Single and Multisite Models of Rainfall within Discrete Time Increments. *Journal of Hydrology*. 17: 97-113.

[14] Dale, W. L. 1960. The Rainfall of Malaya, Part II. *Journal of Tropical Geography*. 14: 11-28.

[15] Chia, L. S. 1968. An Analysis of Rainfall Patterns in Selangor. *Journal of Tropical Geography*. 27: 1-18.

[16] Shaharuddin, A. 2004. Trends and Variability of Rainfall in Malaysia: A Case of Kuala Lumpur and Kuantan, Pahang. Proceedings of Hawaii International Conference on Social Sciences. Wakiki, Honolulu, USA. 16-19.

[17] Zalina, M. D., M. N. M. Desa, V-T-V. Nguyen, and A. H. M. Kassim. 2002. Selecting a Probability Distribution for Extreme Rainfall Series in Malaysia. *Water Science and Technology*. 45: 63-68.

[18] Zhang, J. 2002. Powerful Goodness-of-fit Tests based on the Likelihood Ratio. *Journal of Royal Statistical Society Series B*. 64(2): 281-294.

[19] Zhang, J. and Y. Wu. 2005. Likelihood Ratio Tests for Normality. *Computational Statistics & Data Analysis*. 49(3): 709-721.

[20] Coe, R. and R. D. Stern. 1982. Fitting Models to Daily Rainfall Data. *Journal of Applied Meteorology*. 21(7): 1024-1031.

[21] Stephens, M. A. and R. B. D'Agostino. 1986. *Goodness-of-fit Techniques*. Marcel Dekker, INC: New York and Basel.