Journal of Applied Sciences Research, 4(12): 1846-1857, 2008

Fitting the Statistical Distribution for Daily Rainfall in Peninsular Malaysia Based on AIC Criterion

[1]Jamaludin Suhaila and [2]Abdul Aziz Jemain

[1]Department of Mathematics, Faculty of Science, Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia.
[2]School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600, Bangi, Selangor, Malaysia.

Abstract: This paper presents several types of exponential distributions to describe rainfall distribution in Peninsular Malaysia over a multi-year period. The exponential, gamma, mixed exponential and mixed gamma distributions are compared to identify the optimal model for daily rainfall amount based on data recorded at rain gauges stations in Peninsular Malaysia. The models are evaluated based on the Akaike Information criterion (AIC). The log likelihood ratio test has been employed to determine whether the differences in AIC between tested models are statistically significant. However, this test is restricted by the need of the models to be nested. Since the gamma is not nested in the mixed exponential model so comparison has been done indirectly using the mixed gamma as the nested model. Overall, this study has shown that the mixture of two distributions is better than single distributions for describing the daily rainfall amount in Peninsular Malaysia based on the AIC criterion and their differences in AIC are statistically significant.

Key words: Mixture of two distribution; Akaike information criterion, nested model; mixed gamma; mixed exponential.

INTRODUCTION

Haze is no longer a new phenomenon to the Southeast Asian countries. It has become a regular problem that has to be faced by the country such as Brunei, Indonesia, Thailand, Philippines and Malaysia. This problem is caused by land and forest fires in the zones with high temperature levels in Indonesia. Haze occurs when dust and smoke particles accumulate in relatively dry air. Severe haze can cause huge damages to the forests and agricultural products. These have serious consequences to transport, tourism and other economic endeavours. Besides, haze also poses threats to peoples' health.

In this respect, rainfall is important in that it can reduce, or eliminate the effect of haze. Unfortunately, heavy rainfalls could bring disaster such as floods and landslides. Of course, the shortage of rainfall could also affect the water management system in such a way it could bring problems to the economic activities. Therefore, there are needs to investigate the characteristic of rainfall of a country intensively and comprehensively.

Modeling of daily rainfall using various mathematical models has been done throughout the world to give a better understanding about the rainfall pattern and its characteristics which involve the study on the sequence of dry and wet days and also the rainfall amount on the wet days. Markov chain models have been widely used in modeling the sequence of dry and wet days (Gabriel and Neumann[1]; Roldan and Woolhiser[2]; Stern and Coe[3]; Jimoh and Webster[4].) On the other hand, the Gamma with two parameters distribution is often used in fitting rainfall amount because it represents the large sizes of drop size distributions better than simple Exponential (Ison *et al.*[5]; Katz[6]; Buishand[7]; Aksoy[8], May[9]). Meanwhile, some other theoretical distributions that have been employed in the analysis of rainfall are the Exponential the Kappa[11], the S_B[12], the mixed Exponential[13,14,15,16] the Weibull[17] and the skew Normal[14,15].

These mathematical models of rainfall have been employed in various applications. Mostly they are used in the study of agriculture and crop planning. Sharda and Das[17] compared the two and three parameter probability distributions in order to identify the most suitable distribution that best describe the weekly rainfall data. The results from the model have

Corresponding Author: Jamaludin Suhaila, Department of Mathematics, Faculty of Science, Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia.
E-mail: suhaila@mel.fs.utm.my Tel: (607)-5534317, Fax: (607)-5566162

been used to study the effect of rainfall variability during the cropping season in India. The same kind of study has also been conducted in Uganda[18]. On the other hand, the first-order Markov chain, the Gamma and Weibull distributions have been selected as the models to generate the daily rainfall data such as the studies done in Argentina[19] and in Western Australia[20]. In addition, these mathematical models of rainfall are also important to serve other purpose. For example, Yoo *et al*.[21] have used the results from the parameter estimations of the mixed gamma distribution to study the effect of the global warming in Korea. Therefore, the importance of these mathematical models in rainfall studies should not be neglected.

In the Malaysian context, these kinds of studies received less attention. The studies that have been conducted in Malaysia are more on the general aspects such as pattern, trend and variability of rainfall[22,23] Most of the data are outdated and not analyzed comprehensively especially in the area of statistics. Shaharuddin[24] who studied the trends and variability of rainfall in Malaysia only considered the simple descriptive statistics such as the mean, standard deviation and coefficient of variation in his analysis. Nevertheless, Zalina *et al*.[25,26] who studied the distribution of extreme rainfall series over 17 rain gauge stations in Peninsular Malaysia has considered the method of L-moment in the parameters estimation while the probability plot correlation coefficient test (PPCC), the root mean squared error (RMSE), the relative root mean squared error (RRMSE) and the maximum absolute error (MAE) has been employed as the goodness-of-fit (GOF) tests to determine the best fit model. In her study, Generalized Extreme Value was found to be the best distribution that fit the annual maximum rainfall for hourly data.

The purpose of our study is to find the most appropriate distribution for describing the daily rainfall amount that involves all types of rainfall events in Malaysia. We note that the rainfall distribution in Malaysia is not only characterized by heavy rains but also in terms of light and moderate rains where the frequency of such amounts is greater than the heavy rains. Therefore, we intend to employ the single distributions of Exponential and Gamma together with their mixed distributions. However, the mixtures of distributions are restricted only to two components because if more than two components are considered, it will involve many parameters and things will get complicated especially in the estimation process. Furthermore, modeling with extra components does not bring much difference to the fitted model.

The Akaike information criterion (AIC) will be used as the indicator throughout the studies in selecting the best model since it is much easier to compare the models that came from the same type of Exponential distributions. Based on this criterion, the model which has the minimum AIC is considered to be the best model. In addition, the difference in AICs between the tested models will be investigated to determine whether the differences are statistically significant.

Topography and Climate: Malaysia is situated in the tropics between 1° and 7° north of the equator. It occupies a total area approximately 330,400 square km and is separated by the South China Sea into West Malaysia, which is known as Peninsular Malaysia, and East Malaysia (Sabah and Sarawak). Malaysia in general experiences a wet and humid tropical climate throughout the year that is characterized by high annual rainfall, humidity and temperature. Malaysia has uniform temperatures throughout the year of 25.5° to 32° C. Normally, the annual rainfall amount is between 2000 mm and 4000 mm while the annual number of wet days ranges from 150 to 200 days. There are major differences of climate observed within the country especially between the west and east coasts of Malaysia and slightly less so between north and south. These differences arise from the discrepancy of altitude and the exposure of the coastal lowlands to the southwest and northeast monsoon winds. The southwest monsoon is usually occurred in mid of May and ends in August. The wind is generally light, below 15 knots. On the other hand, the northeast monsoon usually begins in early November and ends in February with speed between 10 to 20 knots. During this season, the more severely affected areas are the east coast states of Peninsular Malaysia where the wind may reach more than 30 knots. The coasts that are exposed to the northeast monsoon in Malaysia tend to be wetter than those exposed to the southwest monsoon. The period of the south-west monsoon is a drier period for the whole country, particularly for the other states of the west coast of the Peninsula. The period of change between the two monsoon is the inter monsoon which occur in Mac/April and Sept/October. These two inter monsoons are usually associated with heavy rainfall. Thus, in general the rainfall distribution in Malaysia is governed by those monsoons.

Rainfall Data: Daily rainfall series data for this study have been obtained from the Malaysian Meteorology Department for the periods ranging from 21 to 35 years. For this study, eighteen rain gauge stations were chosen based on the completeness of the data. The stations are selected to represent rainfall pattern for the whole Peninsular Malaysia. The details about these stations are provided in Table 1 and the specific locations of the stations are shown in Fig. 1.

Modeling Rainfall Amount: In this study, four models were tested. These four models are described below with their probability density function. Note that X is the random variable representing the daily rainfall amount.

- The exponential distribution, with one parameter β which represents the scale parameter determines the variation of rainfall amount series that is given in the same unit as the random variable X.

$$f(x) = \frac{1}{\beta}\exp\left(\frac{-x}{\beta}\right), \qquad x \geq 0 \tag{1}$$

The maximum likelihood estimation (MLE) for β is given as $\hat{\beta} = \overline{X}$.

- The gamma distribution with two parameters, α and β denote the shape and scale parameters respectively.

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

The shape parameter governs the shape of the rainfall distribution and the scale parameter determines the variation of rainfall amount series which is given in the same unit as the random variable X. If $\alpha < 1$, the rainfall distribution is tend to be positively skewed and the maximum of gamma density is located at 0mm/day. For $\alpha = 1$, the rainfall distribution exhibit an exponential shape and the probability density function approaches 0 mm/day asymptotically. In the case of $\alpha > 1$, the gamma density exhibits a single mode at $x = \beta(\alpha - 1)$ and will result in the distribution to be less skewed and the probability density function will be shifted to the right. The MLE for the gamma distribution is easily calculated using the two approximations methods as described by Thom[27] and Greenwood and Duran[28]. However, in this paper the maximum likelihood equations for this distribution were solved by using the quasi-Newton algorithm in the nonlinear programming. The formula shown in Eq. (3) from the method of moment was used to determine the initial values of the parameters for the iterative procedure in the calculation of the gamma density function.

$$\hat{\alpha} = \frac{\bar{x}^2}{s^2} \; and \; \hat{\beta} = \frac{s^2}{\bar{x}} \tag{3}$$

- The mixed exponential distribution with three parameters is the mixture of two one-parameter exponential distributions where p denotes the mixing probability that give the weights to the two exponential distributions with scale parameters β_1 and β_2. The mixed exponential distribution has the same characteristic as the single parameter exponential where the scale parameters for both components represent the variation of rainfall amount series which have the same values as random variable X. Large values of scale parameters give large variation of rainfall amount series.

$$f(x) = \left(\frac{p}{\beta_1}\right)\exp\left[\frac{-x}{\beta_1}\right] + \left(\frac{1-p}{\beta_2}\right)\exp\left[\frac{-x}{\beta_2}\right] \tag{4}$$

$$0 \leq p \leq 1, \quad \beta_1 > 0, \quad \beta_2 > 0, \quad x > 0 \tag{5}$$

The maximum likelihood equation for this distribution is in implicit form. Hence, the nonlinear programming has been used to determine the parameters through the quasi-Newton algorithm with the restrictions as described in Eq. (5). The iterative procedure requires initial values where in this paper; they were estimated by the method of moments as shown by Rider[29] and some other initial values as suggested below.

$$\beta_2 \geq \beta_1 \; and \; p = 0.2, 0.6, 0.8 \tag{6}$$

The mixed gamma with five parameters is the mixture of two two-parameter gamma distributions where p denotes the mixing probability that determines the weights given to the two gamma distributions in which α_1, α_2, β_1 and β_2 denote the shape and scale parameters respectively.

$$f(x) = \frac{p x^{\alpha_1 - 1}}{\Gamma(\alpha_1)\beta_1^{\alpha_1}}\exp\left(\frac{-x}{\beta_1}\right) + \frac{(1-p)\,x^{\alpha_2 - 1}}{\Gamma(\alpha_2)\beta_2^{\alpha_2}}\exp\left(\frac{-x}{\beta_2}\right) \tag{7}$$

Table 1: The latitude, longitude and period of records data obtained for each of the eighteen rain gauge stations.

Code	Stations	Latitude	Longitude	Period of records
1	Senai	1°38 N	103°40 E	1974-2005
2	Kluang	2°01 N	103°19 E	1974-2005
3	Malacca	2°16 N	102°15 E	1971-2005
4	Mersing	2°27 N	103°50 E	1971-2005
5	Petaling Jaya	3°06 N	101°39 E	1971-2005
6	Subang	3°07 N	101°33 E	1971-2005
7	Temerloh	3°28 N	102°23 E	1978-2005
8	Kuantan	3°47 N	103°13 E	1971-2005
9	Batu Embun	3°58 N	102°21 E	1982-2005
10	Sitiawan	4°13 N	100°42 E	1971-2005
11	Cameron Highlands	4°28 N	101°22 E	1983-2005
12	Ipoh	4°34 N	101°06 E	1971-2005
13	Kuala Trengganu	5°23 N	103°06 E	1985-2005
14	Kuala Krai	5°32 N	102°12 E	1984-2005
15	Bayan Lepas	5°18 N	100°16 E	1971-2005
16	Kota Bharu	6°10 N	102°17 E	1971-2005
17	Alor Star	6°12 N	100°24 E	1971-2005
18	Chuping	6°29 N	100°16 E	1979-2005

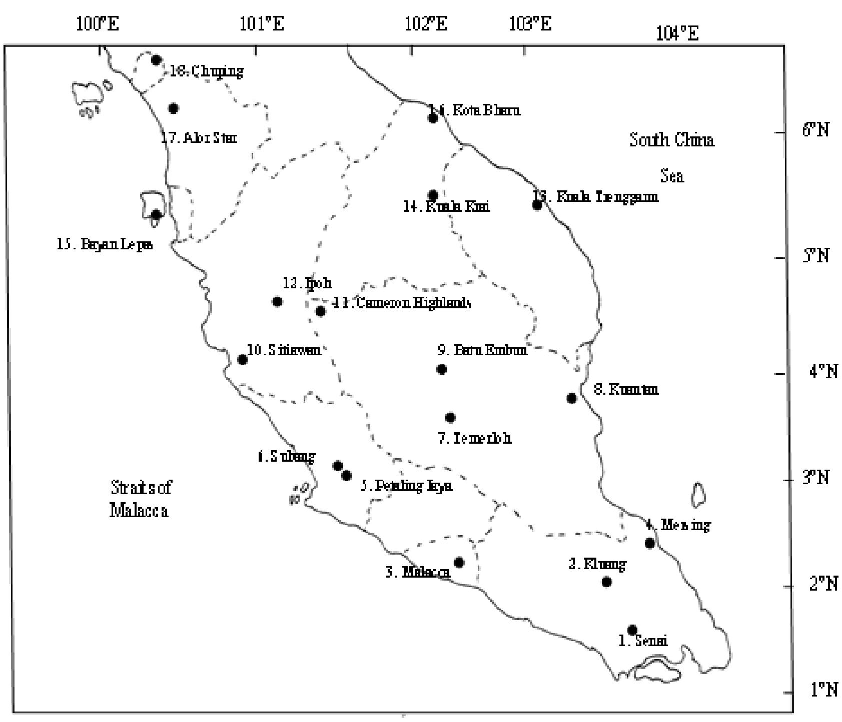

Fig. 1: Map of Peninsular Malaysia showing the location of eighteen rain gauge stations.

$$0 \le p \le 1, \quad \alpha_1 > 0, \; \alpha_2 > 0, \; \beta_1 > 0, \; \beta_2 > 0 \text{ and } x > 0 \qquad (8)$$

Once again, the maximum likelihood equation for this distribution is in implicit form and complicated and we will not go into its details in this paper. The quasi-Newton algorithm has been employed in the iteration procedure to solve the likelihood equation. The choice of initial values is depended on the constraints as described in Eq. (8). Different initial values have been tested for this procedure as suggested in Eq. (9) by trial and error.

$$\alpha_1 \le \alpha_2, \; \beta_2 \ge \beta_1 \; and \; p = 0.2, 0.6, 0.8 \qquad (9)$$

If the initial values converge to the same values and have the largest likelihood, it is considered to be the chosen estimated parameter.

Goodness-Of-Fit Tests (GOF): Several criteria of GOF exist in determining which distribution is the best model to describe the rainfall process. Among them were the Empirical Distribution Function Statistics which include Kolmogorov-Smirnov, Anderson Darling and Cramer-von-Mises tests. Since all the tested models are derived from the same exponential distributions, the AIC has been used to search for the best model. The AIC is based on the value of the log likelihood function. The formula is given as follows;

$$\text{AIC} = -2 \log L + 2k \qquad (10)$$

where log L is the log likelihood of the proposed model and k is the number of model parameters. The best model is the one that gives the smallest AIC. However, the question may arise whether the differences in AIC are statistically meaningful and significant. Therefore, the likelihood ratio test has been used to evaluate the statistical differences in the values of AIC for the tested models. Unfortunately, the use of likelihood ratio test is restricted by the needs of the models to be nested[30].

A nested model is defined as a model which can be derived from the higher order model with more parameters by restricting one or more parameters of that higher order model. Several examples given here will illustrate the idea of the nested model. The exponential distribution given in Eq. (1) is nested in the gamma distribution since it can be derived from the gamma distribution by restricting the shape parameter $\alpha = 1$ in Eq. (2). The exponential distribution is also nested in the mixed exponential distribution by setting $p = 0$ in Eq. (4). If we use the same procedure for mixed gamma, and set $p = 0$ and either α_1 orα_2 equals to 1, then the exponential distribution results. This shows that the gamma distribution and exponential distribution are nested in the mixed gamma.

Suppose we have two models with the AICs given as

$$\text{AIC}_i = -2 \log (L_i) + 2i \qquad (11)$$
$$\text{AIC}_{i+j} = -2 \log (L_{i+j}) + 2(i+j)$$

Then the likelihood ratio test can be written as

$$-2 (\log (L_i) - \log (L_{i+j})) = \text{AIC}_i - \text{AIC}_{i+j} + 2j \sim \chi_j^2 \qquad (12)$$

where the subscript indicates the number of parameters with i is the number of common parameters in the two models. The difference in AIC for those two models are statistically significant at the α level if the value of Eq. (12) is greater than χ_j^2 (α).

In this paper, four models have been tested and the AIC for each model has been calculated. The comparisons of the AIC values of the gamma with mixed gamma, mixed exponential with mixed gamma, and the exponential with those three models are shown below.

$$\text{AIC}_{\text{exponential}} - \text{AIC}_{\text{gamma}} + 2 \sim \chi_1^2 \; (0.05) \qquad (13)$$

$$\text{AIC}_{\text{exponential}} - \text{AIC}_{\text{mixed exponential}} + 4 \sim \chi_2^2 \; (0.05) \qquad (14)$$

$$\text{AIC}_{\text{exponential}} - \text{AIC}_{\text{mixed gamma}} + 8 \sim \chi_4^2 \; (0.05) \qquad (15)$$

$$\text{AIC}_{\text{gamma}} - \text{AIC}_{\text{mixed gamma}} + 6 \sim \chi_3^2 \; (0.05) \qquad (16)$$

$$\text{AIC}_{\text{mixed exponential}} - \text{AIC}_{\text{mixed gamma}} + 4 \sim \chi_2^2 \; (0.05) \qquad (17)$$

The gamma is not nested in the mixed exponential distributions so the Eq. (12) is not valid for these two models. However, as the gamma and mixed exponential distributions are nested in the mixed gamma distributions, we still can compare these two models indirectly based on their nested model.

RESULTS AND DISCUSSION

First, we will have a brief discussion on the descriptive statistics for each of the eighteen rain gauge stations and then proceed to comment on the results of fitting distributions that are based on AIC criterion.

Finally the remarks on the estimated parameters for the best model will be made.

Descriptive Statistics: Descriptive statistics of the daily rainfall amount for each of the eighteen rain gauge stations are summarized in Table 2 where the mean, standard deviation, coefficient of variations, skewness, kurtosis, number of wet days and maximum amount of daily rainfall of each station are given. Kota Bharu station received the highest mean rainfall amount followed by stations Kuantan, Kuala Trengganu and Mersing. All of these stations are located along the east coast of Peninsular Malaysia, and we have already mentioned before that the east coast is exposed to the northeast monsoon which is known to bring heavy rainfall during the monsoon. Hence, those stations are very much influenced by the monsoon. According to Zalina[26], normally 45 to 55 percent of the annual maximum rainfall events for those stations were experienced during this period and the rainy events during this period are very long with heavy and moderate rains occurring intermittently. Meanwhile, Sitiawan station which is located in the west coast of Peninsular Malaysia indicates the lowest rainfall amount. This probably occurred because of the northeasterly winds are blocked by the Main Range (Banjaran Titiwangsa) that affect most of the stations along the west coast of Peninsular. That explains why the west coast is drier than the east coast.

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. Those four stations have high coefficient of variations which ranged between 180% and 200% compared to other stations that ranged between 130% and 160%. In terms of the skewness, the shape of the rainfall distribution for these four stations is strongly skewed. This may be due to the effect of extreme values in the rainfall amount time series or the maximum amount of rainfall that were recorded at those stations. We noticed that Cameron Highlands station which is located on the Main Range showing the lowest variability as well as the smallest value of skewness and the lowest maximum amount of daily rainfall. This shows that the rainfall distribution at this station is more evenly distributed than other stations.

The amount of rainfall is found to be uncorrelated to the number of wet days of the stations. For example, Kota Bharu station has a smaller number of wet days compared to other stations, but has the highest mean rainfall amount. Meanwhile, Petaling Jaya station in the west coast has the highest number of wet days but a smaller mean daily rainfall amount compared to the other stations. This indicates that the higher mean amount of rainfall is not due to the large number of wet days, but possibly contributed by heavy rainfalls. Finally, we could summarize that the differences between the east and the west coast of Peninsular Malaysia are mostly affected by their geographical sites, topographical and climate change in both sites.

Fitting Distributions Based on AIC Criterion: The values for AIC criterion have been calculated and the results are shown in Fig. 2. Based on the results, it shows the complete dominance of the mixed gamma distribution. Of the four models tested, the mixed gamma is found to be the best fitting distribution for all stations studied. These were followed closely by the mixed exponential which ranked second. Since the difference between AIC for both models is quite small and could not be seen clearly from Fig. 2, so the values of AIC are given in Table 3. The pattern of the ranking of the AIC values can be shown as

$$AIC_{\text{exponential}} > AIC_{\text{gamma}} > AIC_{\text{mixed exponential}} > AIC_{\text{mixed gamma}} \tag{18}$$

In general, the outcomes shown that the mixture of two distributions is better than the single distributions in describing daily rainfall amount in Malaysia.

It is important to view the relative difference between the best AIC and other AICs for each of the eighteen rain gauge stations. Using Cahill's approach, the relative difference is defined as

$$r = \frac{AIC_{\text{lesser model}} - AIC_{\text{best model}}}{AIC_{\text{best model}}} \tag{19}$$

Since there are four models, so there are three values of r calculated for each of the eighteen rain gauge stations. The results of this relative difference are plotted on Fig. 3.

The relative difference between the AIC values for the best and worst models range 3% to 8% which we consider high. For all stations the relative difference between the best model and the second best model is less than 0.5% with most of the stations having less than a 0.25% difference. Based on this relative difference, we can say that there is not much difference between mixed exponential with three parameters and the mixed gamma distribution that has five parameters. To ascertain this, the comparison of AIC values between models were carried out to determine whether the differences in AIC are statistically significant.

Table 2: Statistics of daily rainfall amount on wet days for each of the eighteen rain gauge stations.

Stations	Mean	Stdev	CV(%)	Skewness	Kurtosis	Number of Wet days	Maximum amount of rainfall (mm)
Senai	11.94	17.53	147	4.06	40.25	6415	364.4
Kluang	11.22	17.79	159	4.88	63.89	5935	433.4
Malacca	11.50	16.96	148	3.11	18.17	5982	275.2
Mersing	14.38	26.28	183	5.50	48.32	6468	430
Petaling Jaya	13.83	18.40	133	2.39	8.01	7001	177.2
Subang	12.43	17.07	137	2.58	9.43	6926	171.5
Temerloh	11.39	17.15	151	3.07	14.40	4661	200.1
Kuantan	15.84	29.01	183	5.42	47.89	6493	527.5
Batu Embun	11.41	16.82	147	2.73	9.84	4375	160.8
Sitiawan	10.40	15.76	152	2.93	12.45	5964	178.7
Cameron Highlands	11.82	14.18	120	2.08	5.58	5329	107.6
Ipoh	12.59	16.88	134	2.24	6.04	6829	135.4
Kuala Trengganu	15.66	30.31	194	5.44	44.87	3494	432.9
Kuala Krai	13.15	22.19	169	4.59	35.02	3928	356
Bayan Lepas	13.42	19.79	148	3.06	15.61	6192	288.2
Kota Bharu	15.87	31.02	195	6.13	63.90	5660	591.5
Alor Star	12.06	16.75	139	2.85	12.52	5878	178.8
Chuping	10.72	15.81	148	3.65	26.68	4461	267.2

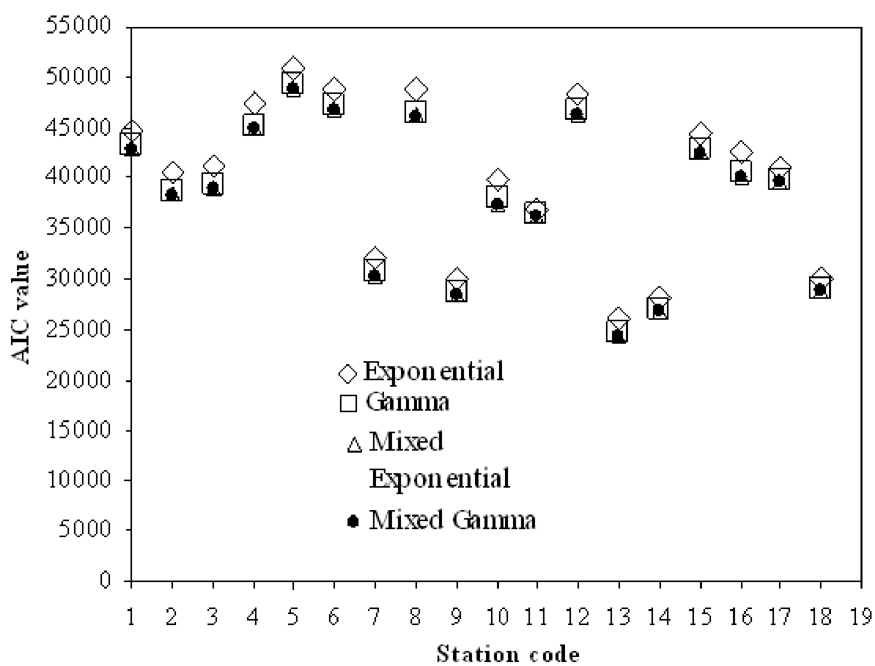

Fig. 2: The AIC values of each of the eighteen rain gauge stations

Table 3: The AIC values of each of the eighteen rain gauge stations

Station Code	Exponential	Gamma	Mixed Exponential	Mixed Gamma
1	44647.14	43205.12	42828.72	42806.42
2	40566.74	38822.36	38434.62	38381.86
3	41181.6	39461.44	38945.62	38907.44
4	47421.3	45228.96	44939.86	44834.92
5	50785.1	49337.46	48917.4	48896.98
6	48758.34	47279	46858.18	46819
7	32003.9	30636.38	30284.32	30269.8
8	48865.48	46549.68	46187.16	46082.82
9	30051.52	28667.7	28378.84	28331.14
10	39861.58	38043.8	37420.84	37384.04
11	36982.38	36467.12	36304.56	36290.46
12	48255.1	46769.94	46382.04	46364.2
13	26215.68	24737.32	24392.92	24347.4
14	28098.48	26953.84	26740.9	26708.16
15	44543.4	42875	42512.78	42484.14
16	42614.74	40557.48	40149.44	40058.7
17	41030.2	39885.52	39601.32	39578.7
18	30089.12	29038.94	28789.28	28772.32

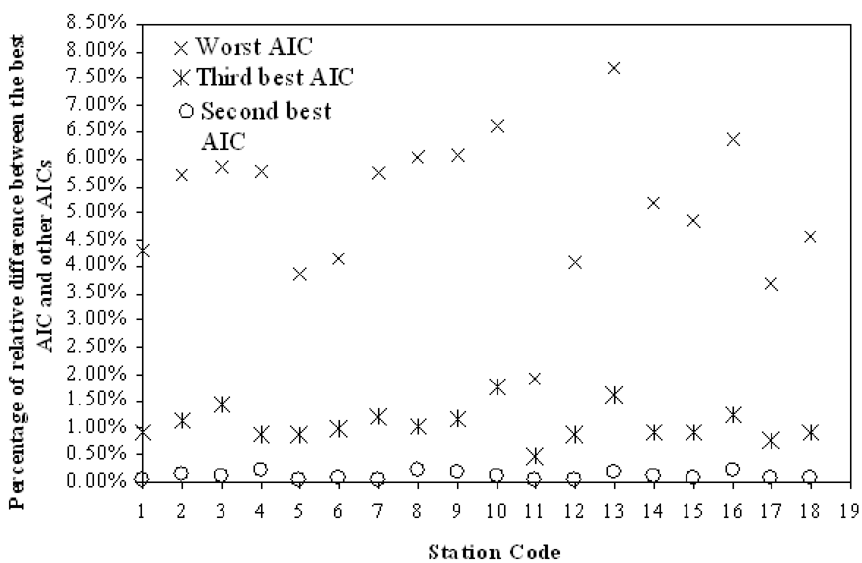

Fig. 3: The relative difference in AIC between the best model and the other three models

Table 4: Estimated parameters for the mixed gamma as the best fitting distribution

Code Stations	mixing probability (p_1)	shape1 (α_1)	scale1 (β_1)	Estimated Mean (1)	shape1 (α_2)	scale1 (β_2)	Estimated Mean (2)
1	0.26	1.19	1.03	0.32	0.83	19.00	11.62
2	0.24	1.30	0.67	0.21	0.76	19.15	11.01
3	0.29	1.27	0.79	0.29	0.82	19.11	11.21
4	0.25	1.11	1.21	0.33	0.69	26.96	14.05
5	0.25	1.25	0.94	0.29	0.88	20.32	13.54
6	0.23	1.36	0.73	0.23	0.84	18.81	12.20
7	0.31	1.11	1.09	0.37	0.81	19.68	11.02
8	0.27	1.12	1.35	0.40	0.70	30.28	15.44
9	0.19	1.61	0.33	0.10	0.73	19.10	11.31
10	0.32	1.29	0.78	0.32	0.82	18.06	10.08
11	0.17	1.39	0.66	0.15	0.94	14.87	11.67
12	0.24	1.23	0.85	0.25	0.85	19.06	12.34
13	0.37	1.04	1.86	0.71	0.70	34.00	14.95
14	0.34	0.95	2.22	0.71	0.75	25.15	12.44
15	0.25	1.17	1.04	0.30	0.78	22.40	13.12
16	0.44	0.87	3.80	1.44	0.70	36.81	14.43
17	0.22	1.27	0.84	0.24	0.84	18.17	11.83
18	0.25	1.19	0.85	0.25	0.81	17.16	10.47

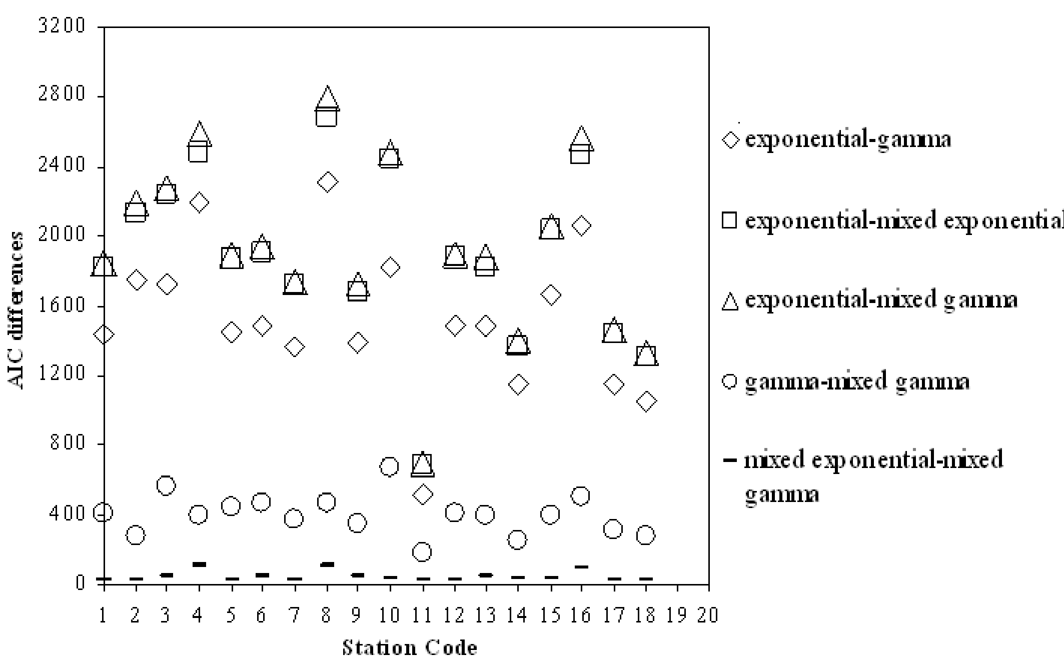

Fig. 4: The comparison between the differences in AIC for each of the tested models with their nested models.

Comparisons based on AIC differences between each tested model with their nested models are shown in Fig. 4. From this figure it is clearly shown that there are large differences in AIC between the exponential distribution with the other three models. The differences for these models are statistically significant at the 5% level. That means the assumption of the AIC values are the same fails at the 95% level for each rain gauge station.

We next consider the gamma and mixed exponential distributions. As noted before the two distributions are not related by nesting but both are nested in a mixed gamma distribution. The results of Eq. (16) and Eq. (17) show that the differences for these two models with their nested models are statistically significant at the 5% level as shown in Fig. 4, means that the assumption of the AIC values are the same fails at the 95% level for each rain gauge station. Combining these results with Eq. (18), the outcomes can be described as follow. The AIC value of the gamma is greater than the mixed gamma, and the difference is statistically significant. Also, the AIC value of the mixed exponential is also greater than the mixed gamma and the difference is statistically significant although it is not much different. Indirectly, we can say that the AIC value of the gamma distribution is significantly different from the mixed exponential since the difference between these two distributions is generally greater than the AIC values between the mixed exponential distribution with mixed gamma distribution. However, this conclusion is based on a heuristic approach, not a formal test. The formal test can be carried out using Monte Carlo analysis without the need of the nested models[30]. However, it is not in covered in this paper. The important findings are that the mixture of two gamma distributions is found to be the best fitting distribution for daily rainfall amount in Malaysia and the differences in AIC with other tested models are statistically significant.

Estimated Parameters: The estimated parameters for the mixed gamma distribution as the best fitting distribution are given in Table 4. The mixing probability indicates the weights given to the first and second component while the shape parameter control the shape of the rainfall amount time series, and the scale parameter represents the variation of rainfall amount series where it has the same unit as rainfall amount (X), in this case millimeters (mm).

We have noticed that the shape parameter of the second component for all stations is less than 1 ($\alpha_2 < 1$), indicates that the distribution is very strongly skewed to the right. Their scale parameters also show a large variation of daily rainfall with extremely low and high rainfalls. In the case of the first component, almost all stations have a shape parameter greater than 1($\alpha_1 > 1$) exceptstations[14,16]. Large values of the shape parameter make the distribution to be less skewed with a shifting of probability density function to the right. In addition, the scale parameters of the first component are much smaller than the second component. Hence, it is likely to have smaller estimated mean for the first component than the second component.

As we have mentioned earlier, the rainfall distribution for each station is different because of its geographical, topographical and climatic changes. For that reason, it is rather difficult to define the threshold of light, moderate and heavy rainfall for each station. Therefore, we could only conclude that the distribution of rainfall amount in Peninsular Malaysia is very well described by two components.

We have seen that Kota Bharu station received the highest mean amount of rainfall and at the same time highest maximum amount of rainfall. Based on its first component, we noticed that its shape parameter is less than one; therefore its distribution tends to be positively skewed. At the same time its scale parameter which determines the variation in rainfall amount series is quite large compared to the values of the first component for other stations. Here we could say the first component for Kota Bharu station is more likely to represent moderate rainfall. For Batu Embun station, the shape parameter of the first component is greater than one automatically indicates that the shape is less skewness along with the small value of scale parameter. This phenomenon can give a very small estimated mean for the first component of Batu Embun station. Therefore, we are unable to make real comment on whether the first component represents either light or moderate rains. In this paper, the terms “light” and “moderate” rains are quite difficult to distinguish since the threshold for both terms are still being debated. Nevertheless, we are convinced that the second component represent heavy rains since most places in Malaysia experience heavy rains.

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

In this study, a comparison of the exponential, gamma, mixed exponential and mixed gamma shows that for eighteen Malaysia rain gauge stations, the mixed gamma best describes the distribution of daily rainfall amount based on Akaike information criterion. The mixed exponential ranked second followed by the gamma and finally the exponential distribution. We have compared AIC values of four models to see if the differences in AIC are statistically significant. By using a likelihood ratio test to compare the AIC values, we have found that the exponential is significantly different from the gamma, mixed exponential and mixed gamma. The same results also hold for the comparison between the gamma with mixed exponential, where their difference in AIC is statistically significant. This comparison was done indirectly, since the two models are not nested, but they are both nested in a mixed gamma distribution model.

The results of this study have shown that mixture of two distributions is better than single distributions for describing the daily rainfall amount in Malaysia. Based on these findings, we can conclude that the pattern of rainfall distribution in Malaysia could be categorized into two types of components that most probably represent heavy and light rains or also could be between heavy and moderate rains. We stress that further studies must be carried out to give a reasonable definition for "light", "moderate" or "heavy" rains. However, we are confident that the total amount of rainfall in Malaysia is mainly contributed by heavy rains.

REFERENCES

1. Gabriel, K.R. and J. Neumann, 1962. A Markov chain model for daily rainfall occurrence at Tel Aviv. Quarterly Journal of Royal Meteorology Society, 88: 90-95.
2. Roldan., J. and D.A. Woolhiser, 1982. Stochastic daily precipitation models 1. A comparison of occurence process. Water Resources Research, 18(5): 1451-1459.
3. Stern, R.D. and R. Coe, 1984. A model fitting analysis of daily rainfall data. Journal of Royal Statistical Society Series A, 147: 1-34.
4. Jimoh, O.D. and P. Webster, 1996. Optimum order of Markov chain for daily rainfall in Nigeria. Journal of Hydrology, 185: 45-69.
5. 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.
6. Katz, R.W., 1977. Precipitation as chain-dependent process. Journal of Applied Meteorology, 16: 671-676.
7. Buishand, T.A., 1978. Some remarks on the use of daily rainfall models. Journal of Hydrology, 36: 295-308.
8. Aksoy, H., 2000. Use of gamma distribution in hydrological analysis. Turkey Journal of Engineering Environmental Sciences, 24: 419-428.
9. 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.
10. Todorovic, P. and D.A. Woolhiser, 1975. A Stochastic model of n-day precipitation. Journal of Applied Meteorology, 14(1): 17-24.
11. Mielke, P.W., 1973. Another family of distributions for describing and analyzing precipitation data. Journal of Applied Meteorology, 12: 275-280.
12. 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.
13. 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.
14. Chapman, T.G., 1997. Stochastic Models for Daily Rainfall in the Western Pacific. Mathematics and Computers in Simulation, 43: 351-358.
15. 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: 317-324.16.
16. Wilks, D.S., 1998. Multisite Generalization of a Daily Stochastic Precipitation Generation Model. Journal of Hydrology, 210: 178-191.
17. 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.
18. Rugumayo, A.I., N. Kiiza, J. Shima. 2003. Rainfall reliability for crop production. A case study in Uganda. Diffuse Pollution Conference Dublin.
19. Castellvi, F., I., Mormeneo and P.J., Perez, 2004. Generation of daily amounts of precipitation from standard climatic data: a case study for Argentina. Journal of Hydrology, 289: 286-302.
20. Mummery, D. and M. Battaglia, 2004. Significance of rainfall distribution in predicting eucalypt plantation growth, management options, and risk assessment using the process-based model CABALA. Forest Ecology and Management., 193: 283-296.
21. Yoo, C., K.S. Jung and W.K. Tae, 2005. Rainfall

frequency analysis using a mixed Gamma distribution: evaluation of the global warming effect on daily rainfall. Hydrological Processes, 19: 3851-3861.
22. Dale, W.L., 1959. The rainfall of Malaya, Part 1. Journal of Tropical Geography, 13: 23-27.
23. Chia, L.S., 1968. An analysis of rainfall patterns in Selangor. Journal of Tropical Geography, 27: 1-18.
24. 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.
25. 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(2): 63-68.
26. Zalina, M.D., 2001. Statistical modeling of extreme rainfall process in Malaysia. Ph.D. Thesis, Universiti Teknologi Malaysia, Malaysia.
27. Thom, H.C.S., 1958. A note on the gamma distribution. Monthly Weather Review, 86: 117-122.
28. Greenwood, J.A. and D. Durand, 1960. Aids for fitting the gamma distribution by maximum likelihood. Technometrics, 2: 55-65.
29. Rider, P.R., 1961. The method of moments applied to a mixture of two exponential distributions. Annals of Mathematical Statistics., 32: 143-147.
30. Cahill, A.T., 2003. Significance of AIC differences for precipitation intensity distributions. Advances in Water Resources. 26: 457-464.