Journal of Applied Sciences Research, 3(10): 1027-1036, 2007

Fitting Daily Rainfall Amount in Peninsular Malaysia Using Several Types of Exponential Distributions

[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 a statistical study on fitting distribution of rainfall amount in Peninsular Malaysia using several types of exponential distributions for periods from 21 to 35 years. The mixed gamma, mixed weibull and mixed exponential are proposed and tested together with their single distributions to identify the optimal model for daily rainfall amount in several rain gauge stations in Peninsular Malaysia. The selected model will be determined based on the minimum error produced by some criteria of goodness of-fit (GOF) tests. The results indicated that the mixture distributions are better than the single distributions in modeling rainfall amount where the mixed weibull is identified as the most appropriate model for the majority of sites in Peninsular Malaysia. These results however are varied between the rain gauge stations which are strongly influenced by their geographical, topographical and climatic changes.

Keywords: Mixture of two distributions; goodness-of-fit test; mixed exponential, mixed weibull; mixed gamma.

INTRODUCTION

Modeling daily rainfall data using various mathematical models has been an important research in hydrology for the last 30 years. The use of mathematical models of rainfall has been applied worldwide in order to give a better understanding about the rainfall pattern and its characteristics. This process involves the study on the sequence of dry and wet days as well as the study of rainfall amount on the wet days. Markov chain models are normally used to model the sequence of dry and wet days[1,2,3,4] while rainfall amount is often modeled using a two-parameter gamma distribution[5,6,7,8,9]. Other theoretical distributions that have also been employed in the analysis of rainfall are exponential distribution[10], kappa distribution[11], S_B distribution[12], mixed exponential distribution[13,14,15,16], weibull distribution[17] and skew normal distribution [14,15].

The selection of the best fitting distribution has always been a key interest in the study of rainfall amount. Suhaila and Jemain[18] compared several types of normal transform distributions to fit daily rainfall amount in Peninsular Malaysia. They found that the mixture of two lognormal distributions gave the best fitting model. In this paper, we will focus on mixture distributions which come from several types of exponential distributions that will be employed together with their single distributions in order to find the best model in fitting rainfall data. In order to verify the suitable distribution that best describes the rainfall amount, the new method of goodness-of-fit tests (GOF) based on the likelihood ratio statistics which has been developed by Zhang[19] and Zhang and Wu[20] will be engaged together with the traditional GOF tests. An additional criterion that is included in the analysis is the median of absolute difference between the hypothesized and empirical distribution function. The final result on the best fitting distribution will be chosen based on the minimum error specified by these GOF criteria.

MATERIALS AND METHODS

Case Study: Daily rainfall series data for this study have been obtained from the Malaysian Meteorological 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 details are shown in Table 1.The stations are selected to represent rainfall pattern for the whole Peninsular Malaysia.

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)-553 4317, Fax: (607)-556 6162

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

Code	Stations	Alt	Latitude	Longitude	Period of records
1	Senai	37.8 m	1°38 N	103°40 E	1974-2005
2	Kluang	88.1 m	2°01 N	103°19 E	1974-2005
3	Malacca	8.5 m	2°16 N	102°15 E	1971-2005
4	Mersing	43.6 m	2°27 N	103°50 E	1971-2005
5	Petaling Jaya	60.8 m	3°06 N	101°39 E	1971-2005
6	Subang	16.5 m	3°07 N	101°33 E	1971-2005
7	Temerloh	39.1 m	3°28 N	102°23 E	1982-2005
8	Kuantan	15.3 m	3°47 N	103°13 E	1971-2005
9	Batu Embun	59.5 m	3°58 N	102°21 E	1978-2005
10	Sitiawan	7.0 m	4°13 N	100°42 E	1971-2005
11	Cameron Highlands	1545 m	4°28 N	101°22 E	1983-2005
12	Ipoh	40.1 m	4°34 N	101°06 E	1971-2005
13	Bayan Lepas	2.8 m	5°18 N	100°16 E	1971-2005
14	Kuala Trengganu	5.2 m	5°23 N	103°06 E	1985-2005
15	Kuala Krai	68.3 m	5°32 N	102°12 E	1984-2005
16	Kota Bharu	4.6 m	6°10 N	102°17 E	1971-2005
17	Alor Star	3.9 m	6°12 N	100°24 E	1971-2005
18	Chuping	21.7 m	6°29 N	100°16 E	1979-2005

There are major differences of climate observed within the country especially between the west and east coasts of Peninsular Malaysia and slightly less so between the 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 usually occurs in mid of May and ends in September. The wind is generally light going below 15 knots. On the other hand, the northeast monsoon usually begins in early November and ends in March with a 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 Peninsular Malaysia tend to be wetter than those exposed to the southwest monsoon. The period of the southwest 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 monsoons is a transitional period which occurs in April and October. These two transitional periods often result in heavy rainfall which usually occurs in the form of convective rains. During these periods, the west coast is generally wetter than the east coast. Thus, in general the rainfall distribution in Peninsular Malaysia is mostly affected by those two monsoons and the two transitional periods.

Modeling Daily Rainfall Amount: Five models of rainfall amount are described as follows with their probability density functions. Note that X is the random variable representing the daily rainfall amount.

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

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

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.

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

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

Slightly similar to the gamma distribution, the weibull distribution shares the ame properties of nonnegativity and positive skewness. The weibull and gamma distribution are identical for $\alpha = 1$, where both of these distributions reduce to the exponential distribution.

The mixed exponential distribution with three parameters is a 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.

$$f(x)=\left(\frac{p}{\beta_1}\right)\exp\left[\frac{-x}{\beta_1}\right]+\left(\frac{1-p}{\beta_2}\right)\exp\left[\frac{-x}{\beta_2}\right] \qquad (3)$$

$$0\le p\le 1,\quad \beta_1>0,\quad \beta_2>0,\quad x>0 \qquad (4)$$

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.

The mixed weibull with five parameters is a mixture of two two-parameter weibull distributions where p denotes the mixing probability which determines the weights given to the two weibull distributions. Here parameters α_1 and α_2 represent the shape parameters whereas β_1 and β_2 represent the scale parameters in the distribution.

$$f(x)=p\left(\frac{\alpha_1}{\beta_1}\right)\left[\frac{x}{\beta_1}\right]^{\alpha_1-1}\exp\left[-\left(\frac{x}{\beta_1}\right)^{\alpha_1}\right]$$
$$+(1-p)\left(\frac{\alpha_2}{\beta_2}\right)\left(\frac{x}{\beta_2}\right)^{\alpha_2-1}\exp\left[-\left(\frac{x}{\beta_2}\right)^{\alpha_2}\right] \qquad (5)$$
$$0\le p\le 1,\quad \alpha_1>0,\quad \alpha_2>0,\quad \beta_1>0\quad \beta_2>0 \text{ and } x>0. \qquad (6)$$

The mixture of two weibull distribution can be reduced to a single parameter of exponential distribution if we set $p = 1$ with either $\alpha_1 = 1$ or $\alpha_2 = 1$. The mixture of two weibull distributions share the same properties as the single weibull distribution where their shape parameters determine the shape of the rainfall distribution for both components while the scale parameters again represent the variation of rainfall amount series which has the same units as the random variable X. The differences between the first and second components are determined by their mixing probability.

The mixed gamma with five parameters is a mixture of two two-parameter gamma distributions where p denotes the mixing probability which determines the weights given to the two gamma distributions where $\alpha^1,\alpha^2,\beta^1$ and β^2 denote the shape and scale parameters respectively.

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

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

Once again, the mixture of two gamma distributions shares the same properties as the single gamma distribution. The shape and scale parameters represent the same properties as we discussed before. So we did not elaborate further.

Estimation Procedure: Many methods are available for parameter estimations, which include the method of moments (MM), maximum likelihood estimation (MLE), the least squares method (LS), L-moments and generalized probability weighted moments (GPWM). The MLE method is considered in this study because it provides the smallest variance as compared to other methods. The idea of this method is to find a set of parameters that will maximize the likelihood function. The parameters are obtained by differentiating the log likelihood function with respect to the parameters of the distribution.

Gamma: The logarithm of the likelihood function of the considered distribution is given as follow.

$$\ln L = -N\ln\Gamma(\alpha)-N\alpha\ln\beta+(\alpha-1)\sum_{i=1}^{N}\ln x_i-\frac{\sum_{i=1}^{N}x_i}{\beta} \qquad (9)$$

The MLE for the gamma distribution is easily calculated using the two approximation methods as described by Thom[21] and Greenwood and Duran[22]. 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. (10) from the MM method was used to determine the initial values of the

$$\hat{\alpha} = \frac{\bar{x}}{s^2} \quad \text{and} \quad \hat{\beta} = \frac{s^2}{\bar{x}} \qquad (10)$$

$$\ln L = -N\ln\Gamma(\alpha) - N\ln(\beta) + (\alpha - 1)\sum_{i=1}^{N}\ln\left(\frac{x_i}{\alpha}\right) - \frac{1}{\beta^{\alpha}}\sum_{i=1}^{N}x_i^{\alpha} \qquad (11)$$

parameters for the iterative procedure in the calculation of the gamma density function.

The quasi-Newton algorithm once again was applied to solve the nonlinear equation given by Eq. (11). The method of moments given by Wong[23] was used to determine the initial values for the iteration.

The MLE estimations of the single gamma and weibull distributions are easier to compute since many relevant softwares are available to compute the gamma and weibull MLE directly such as SPLUS and MATLAB.

Mixture of two exponential distribution

$$\ln L = \sum_{i=1}^{N}\ln\left\{ \begin{array}{l} p\left[\frac{1}{\beta_1}\exp\left[-\left(\frac{x_i}{\beta_1}\right)\right]\right] \\ +(1-p)\left[\frac{1}{\beta_2}\exp\left[-\left(\frac{x_i}{\beta_2}\right)\right]\right] \end{array} \right\} \qquad (12)$$

The maximum likelihood equation for this distribution is in implicit form. Once again the quasi-Newton algorithm has been used to solve Eq. (12) with the restrictions as described in Eq. (4). The iterative procedure requires initial values where they were estimated by the method of moments as shown by Rider[24] and some other initial values as suggested below.

$$\beta_2 \ge \beta_1 \text{ and } p = 0.2, 0.6, 0.8 \qquad (13)$$

Mixture of two weibull distribution

In this paper, the MLE method has been used to solve the log likelihood equation as expressed in Eq. (14). We used the quasi-Newton algorithm to resolve this nonlinear equation, with the restrictions as described in Eq. (6).

$$\ln L = \sum_{i=1}^{N}\ln\left\{ \begin{array}{l} p\left[\frac{\alpha_1}{\beta_1}\left(\frac{x_i}{\beta_1}\right)^{\alpha_1-1}\exp\left[-\left(\frac{x_i}{\beta_1}\right)^{\alpha_1}\right]\right] \\ +(1-p)\left[\frac{\alpha_2}{\beta_2}\left(\frac{x_i}{\beta_2}\right)^{\alpha_2-1}\exp\left[-\left(\frac{x_i}{\beta_2}\right)^{\alpha_2}\right]\right] \end{array} \right\} \qquad (14)$$

We next consider the initial values for the iterative procedure. Regarding to the restrictions in Eq. (6), we tried several values given in Eq. (15) by trial and error.

$$\alpha 1 \le \alpha 2,\ \beta 2 \le \beta 1 \text{ and } p = 0.2, 0.6, 0.8 \qquad (15)$$

Another alternative is to use the initial values as shown in Eq. (16) that was given in the study done by Carta and Ramirez[25] where m_1 and s^2 refer to the mean and variance of rainfall data.

$$\alpha_1 = \alpha_2 = \left(\frac{\sqrt{s^2}}{m_1}\right)^{-1.086};\ \beta_1 = \beta_2 = m_1\left[\Gamma\left(1+\frac{1}{\alpha_1}\right)\right]^{-1}\ p = 0.25, 0.5, 0.75 \qquad (16)$$

Using both approaches, we found that they give the same results for the estimated parameters.

Mixture of two gamma distribution

Again the MLE method was used to solve the following nonlinear equation;

$$\sum_{i=1}^{N}\ln\left\{ \begin{array}{l} \frac{p x_i^{\alpha_1-1}}{\Gamma(\alpha_1)\beta_1^{\alpha_1}}\exp\left(\frac{-x_i}{\beta_1}\right) \\ +\frac{(1-p)x_i^{\alpha_2-1}}{\Gamma(\alpha_2)\beta_2^{\alpha_2}}\exp\left(\frac{-x_i}{\beta_2}\right) \end{array} \right\} \qquad (17)$$

The same estimation procedures for the mixture of two gamma distribution were carried out to determine their estimate parameters following the restriction specified in Eq. (8) by trying some initial values as suggested in Eq. (15).

Goodness-of-fit Tests (GOF): Seven different 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 function $Fn(x_{(1)})$ and the hypothesized distribution function $Fn(x_{(1)}\ \theta)$. The chosen distribution that best fits the daily rainfall amount is based on the minimum error indicate by all these seven tests.

The first criterion invc lute difference (MAD) between $F(x_{(i)}, \hat{\theta})$ and $F_n(x_{(i)})$ Fhe formula is given as below.

$$\text{MAD} = \text{Med}\left|F_n(x_{(i)}) - F(x_{(i)}, \hat{\theta})\right| \qquad (18)$$

In this equation, ☻represents the ordered data whereas the vector of estimated parameters is represented by θ

The second criterion involves the classical EDF statistics which are also called as classical GOF tests. They are usually divided into two classes; the supremum and the quadratic. This paper concentrates on the well known Kolmogorov-Smirnov statistic D which belongs to the supremum class of EDF statistics.

This stati [illegible] est vertical differences between $F(x_{(i)},\hat{\theta})$ and $F_n(x_{(i)})$ For the quadratic class, this paper focuses on the Cramer-von-Mises statistic W^2 and the Anderson Darling statistic A^2 test. This class of statistics is based on the squared difference $\left[F_n(x_{(i)})-F\left(x_{(i)},\hat{\theta}\right)\right]^2$

Let $\left[F_n(x_{(i)})-F\left(x_{(i)},\hat{\theta}\right)\right]^2$ with $x_{(i)}$ represents the ordered data whereas the vector of estimated parameters is represented by $\hat{\theta}$ Then the above EDF statistics have the following formula:

$$D^+ = \max_i \left\{ i/n - F_{(i)} \right\},\ D^- = \max_i \left\{ F_{(i)} - (i-1)/n \right\}$$

$$D = \max(D^+, D^-) \qquad (19)$$

$$W^2 = \sum_{i=1}^{n} \left\{ F_{(i)} - (2i-1)/2n \right\}^2 + \frac{1}{12n} \qquad (20)$$

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n} \left[\begin{array}{l} (2i-1)\log F_{(i)} + \\ (2n+1-2i)\log\left(1-F_{(i)}\right) \end{array} \right] \qquad (21)$$

The third criterion involves new powerful GOF tests which are based on the likelihood [illegible] tween the hypothesized distribution $F_{(i)} = F(x_{(i)},\hat{\theta})$ and the empirical distribution function Fn $(x_{(i)})$. The computed formulas used in this paper are shown in Eq. (22)-Eq. (24) as in Zhang[19] and Zhang and Wu[20].

a) New Kolmogorov-Smirnov test

$$Z_K = \max_{1\le i\le n} \left[\begin{array}{l} (i-0.5)\log\dfrac{i-0.5}{nF_{(i)}} \\ +(n-i+0.5)\log\dfrac{n-i+0.5}{n(1-F_{(i)})} \end{array} \right] \qquad (22)$$

b) New Cramer-von-Mises test

$$Z_C = \sum_{i=1}^{n} \left[\log \frac{F_{(i)}^{-1} - 1}{(n-0.5)/(i-0.75)-1} \right]^2 \qquad (23)$$

c) New Anderson-Darling test

$$Z_A = -\sum_{i=1}^{n} \left[\frac{\log F_{(i)}}{n-i+0.5} + \frac{\log(1-F_{(i)})}{i-0.5} \right] \qquad (24)$$

RESULTS AND DISCUSSIONS

First, we will have a brief discussion on the descriptive statistics for each of the eighteen rain gauge stations and then proceed to give comments on the results of fitting distributions that are based on GOF criteria. Finally the remarks on the estimated parameters for the best model will be made.

Descriptive Statistics: A summary of the basic statistics of daily rainfall amount for each of the eighteen rain gauge stations is 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 daily rainfall amount followed by three other stations which are Kuantan, Kuala Trengganu and Mersing. In terms of the standard deviation, these four stations once again showed high values which indicate a large variation in their daily rainfall amount series. We noticed that all of these four stations are located along the east coast of Peninsular Malaysia. As we mentioned before, the entire east coast is exposed to the northeast monsoon which is known to bring heavy rainfall during 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. Therefore, those areas are subjected to flooding during the monsoon. Meanwhile, Sitiawan station 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 the Peninsular. That explains why the west coast is drier than the east coast in terms of rainfall amount.

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. Stations which are located in the east coast have shown high variability of rainfall amount which ranged between 180% and 200% compared to other stations that ranged between 130% and 160%. The lowest variability is indicated by Cameron Highlands station which is located on the Main Range. Based on the results given in Table 2, this station indicates 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 also found to be uncorrelated to the number of wet days of the stations. For example, Kota Bharu station has a smaller number

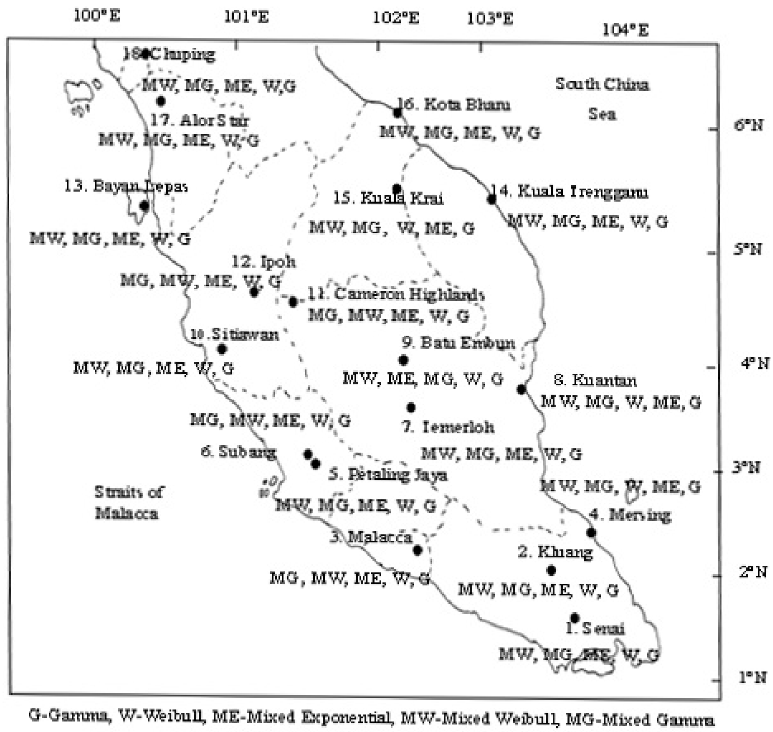

Fig. 1: Map of Peninsular Malaysia showing the eighteen rain gauge stations and the ranking distributions for each station.

of wet days compared to other stations, but has the highest mean rainfall amount. Meanwhile, Petaling Jaya station 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 rainfall amount is not due to the large number of wet days, but it is possibly contributed by heavy rainfall.

Fitting Distribution Based on GOF Criteria: The values of seven goodness-of-fit criteria have been calculated and the best distribution 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. The 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 effective test.

The overall results of the analysis are plotted onto the map of Peninsular Malaysia as shown in Fig. 1. Based on the results, 78 percent of the studied stations chose the mixed weibull as the best distribution in describing daily rainfall amount while the mixed gamma and mixed exponential are placed in the second and third ranked distributions respectively. But for

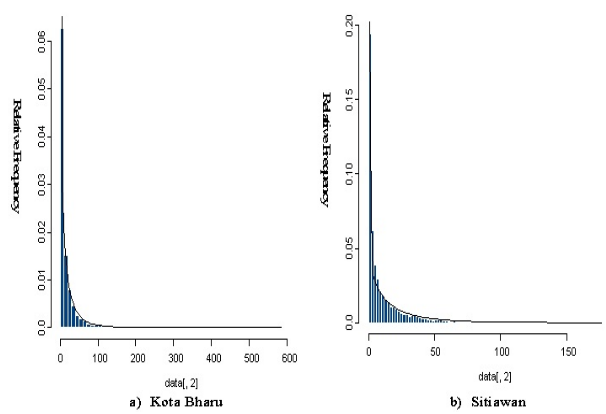

Fig. 2: Histogram of daily rainfall series for stations (a) Kota Bharu and (b) Sitiawan with their best fitting distribution (mixture of two weibull distributions).

stations Malacca, Subang, Ipoh and Cameron Highlands which are located in the western and central part of Peninsular Malaysia, the mixed gamma has been chosen as the best model. The obvious difference in the rank of the distributions could be seen for certain stations in the east coast of Peninsular Malaysia where the single weibull distribution placed as the third rank distribution for stations Kuala Krai, Kuantan and

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
Bayan Lepas	13.42	19.79	148	3.06	15.61	6192	288.2
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
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

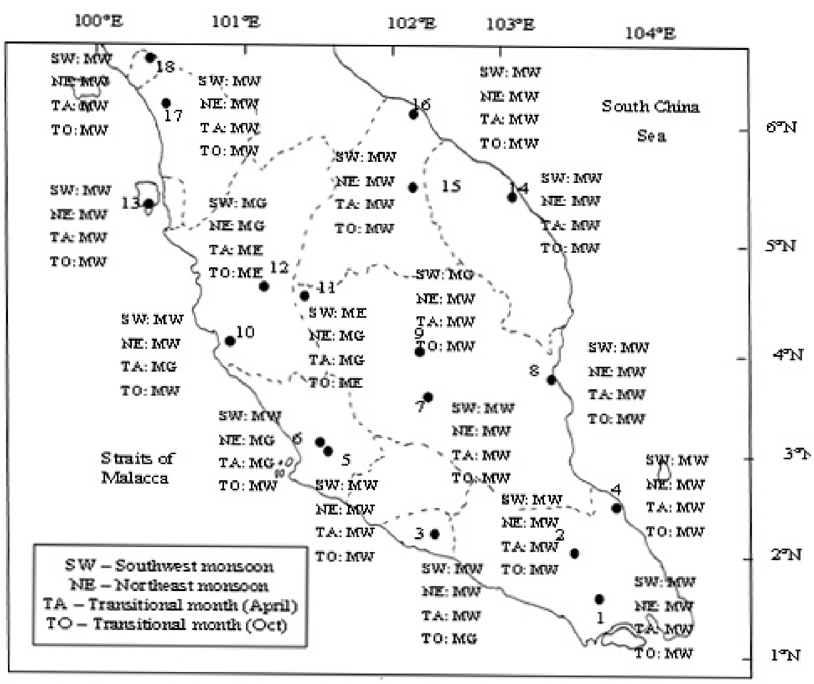

Fig. 3: The best model of rainfall amount for each type of monsoons

Table 3: Estimated parameters for mixed weibull 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	0.219	1.135	1.099	0.23	0.847	13.717	11.677
2	0.183	1.262	0.709	0.121	0.788	11.763	11.013
3	0.263	1.166	0.97	0.242	0.854	14.050	11.233
4	0.136	1.340	0.743	0.093	0.730	13.297	13.998
5	0.236	1.127	1.175	0.265	0.912	16.982	13.560
6	0.201	1.237	0.924	0.174	0.865	14.222	12.233
7	0.266	1.062	1.084	0.282	0.836	13.744	11.083
8	0.169	1.218	1.005	0.159	0.735	15.314	15.403
9	0.150	1.471	0.449	0.061	0.777	11.459	11.280
10	0.278	1.184	0.933	0.245	0.839	12.783	10.126
11	0.153	1.210	0.937	0.135	0.944	13.442	11.682
12	0.212	1.136	0.983	0.199	0.877	14.72	12.382
13	0.184	1.204	0.911	0.157	0.805	14.339	13.204
14	0.261	1.107	1.438	0.362	0.732	16.740	15.036
15	0.142	1.306	0.817	0.107	0.750	12.579	12.856
16	0.198	1.148	1.408	0.266	0.723	15.473	15.244
17	0.178	1.208	0.924	0.154	0.851	13.291	11.877
18	0.199	1.149	0.862	0.164	0.832	11.913	10.521

Mersing. For these three stations, the weibull is better than the mixed exponential in describing the daily rainfall amount. However, the overall results still indicate that the mixture of two distributions is generally better than the single distributions. Taking stations Kota Bharu and Sitiawan as the representative stations, we give the histograms of daily rainfall series for those stations together with their best fitting distribution as shown in Fig. 2.

We have already seen before that there are evidences showing the differences in variations of rainfall amount between the east and west coast of Peninsular Malaysia during the monsoons. These monsoonal flows and transitional seasons as we mentioned earlier contribute to heavy rainfall on the east and west coast at different times of the year. So, once again the analysis was done according to the monsoons and the results are shown in Fig. 3.

The tests results have shown that all rain stations which are located in the east coast of Peninsular Malaysia chose the mixture of two weibull distributions as the most appropriate model to describe the daily rainfall amount during the monsoons. Those stations that are located in the north with approximately 5°N latitude and to the west also hold the same results as the stations in the east coast. The major difference in the rainfall distribution during the monsoons could be seen especially for the stations that are located in the west coast particularly for two stations; Cameron Highlands and Ipoh. The mixed weibull distribution has never been selected by any of the monsoons for both stations. On the other hand, the mixed gamma and mixed exponential have been considered as the suitable model for these stations but varied according to the type of monsoons. We have already noticed that Cameron Highlands station is located on the Main Range with the highest altitude compared to other stations while Ipoh station is considered to be the nearest neighbor to Cameron Highland. As we have mentioned earlier the discrepancy of altitude and exposure of the coastal lowlands to the southwest and northeast monsoons could influence the results on fitting distribution. Besides, these differences in the rainfall distributions between the studied stations are strongly influenced by their geographical sites, topographical and climatic changes.

Estimated Parameters: The estimated parameters for the best model are displayed in Table 3. The mixture of two weibull distributions is considered as the best model among other types of exponential distributions in describing the daily rainfall amount in Malaysia.

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 or 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.

Based on the results from Table 3, the mixing probability indicates that less than 30 percent of daily rainfall amount series in Peninsular Malaysia represents the first component while the remainders describes second component. For a detail explanation examine stations Kota Bharu in the east coast of Peninsular Malaysia, which has the highest mean daily rainfall amount, and Sitiawan in the west coast of Peninsular Malaysia that has the lowest mean daily rainfall amount. The rainfall amount series in Kota Bharu station consists of 20 percent of first component with estimated mean approximately 0.27 mm/day while the remainder is from the second component with estimated mean approximately 15.24 mm/day. Comparatively at the Sitiawan station, the rainfall amount series comprised of 28 percent of first component with estimated mean 0.245 mm/day while 72 percent of the second component produce the estimated mean approximately 10.126 mm/day. We are unable to make a real comment on what types of rainfall that each component represent since the threshold for terms "light", "moderate" and "heavy" rainfall are still being questioned. However, we are convinced that rainfall distribution in Peninsular Malaysia is very well described using two components.

Conclusion: The search for the best fitting distribution for 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.

In this study, the mixture of two weibull distributions has been identified as the best fitting distribution for the majority of sites in Peninsular Malaysia. However, the mixed gamma distribution was also considered as the best models for certain stations particularly for those stations which are located in the west coast of the Peninsular Malaysia such as stations Cameron Highland and Ipoh. The same results also hold for the rainfall analysis during the monsoons. In general, we could say that the major differences in rainfall distribution between the rain stations in Penisular Malaysia are mostly influenced by their geographical sites, topographical and climatic changes. In conclusion, we have shown that mixtures of two distributions are better than single distributions for describing the daily rainfall amount in Peninsular Malaysia. Based on these findings, we can conclude that the pattern of rainfall distribution in Peninsular Malaysia could be categorized into two types of components. We stress that further studies must be carried out to give a reasonable definition for "light", "moderate" or "heavy" rains to support our results.

ACKNOWLEDGEMENTS

Authors are thankful to the staff of Malaysian Meteorology Department for providing daily rainfall data for this paper. This paper would not have been possible without the sponsorship from Universiti Teknologi Malaysia and Universiti Kebangsaan Malaysia.

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.D. 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. Swift Llyold 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. Wilks, D.S., 1999. Interannual variability and extreme-value characteristics ofseveral stochastic daily precipitation models. Agricultural and Forest Meteorology, 93: 153-169.
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. Suhaila, J. and A.A. Jemain, 2007.Fitting daily rainfall amount in Malaysia using the normal transform distribution. Journal of Applied Sciences 2007. ISSN, 1812-5654.
19. Zhang, J., 2002. Powerful goodness-of-fit tests based on the likelihood ratio. Journal of Royal Statistical Society Series B, 64(2): 281-294.
20. Zhang, J. and Y. Wu, 2005. Likelihood ratio tests for normality. Computational Statistics & Data Analysis, 49(3): 709-721.
21. Thom, H.C.S., 1958. A note on the gamma distribution. Monthly Weather Review, 86: 117-122.
22. Greenwood, J.A. and D. Durand, 1960. Aids for fitting the gamma distribution by maximum likelihood. Technometrics, 2: 55-65.
23. Wong, R.K.W., 1977. Weibull distribution, iterative likelihood techniques and hydrometeorological data. Journal of Applied Meteorology, 16: 1360-1364.
24. Rider, P.R., 1961. The method of moments applied to a mixture of two exponential distributions. Annals of Mathematical Statistics, 32: 143-147.
25. Carta, J.A. and P. Ramirez, 2007. Analysis of two component mixture weibull statistics for estimation of wind speed distributions. Renewable Energy, 32: 518-531.
26. Zalina, M.D., 2001. Statistical modeling of extreme rainfall process in Malaysia. Ph.D Thesis, Universiti Teknologi Malaysia, Malaysia.