Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

ISSN: 0972-6268	Vol. 14	No. 4	pp. 951-956	2015

Original Research Paper

Stochastic Modelling of Monthly Rainfall Volume During Monsoon Season over Gangetic West Bengal, India

Soumen Pal*† and Debasis Mazumdar**
*Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute, New Delhi-110 012, India
**Department of Agricultural Statistics, Bidhan Chandra Krishi Viswavidyalaya, Nadia, West Bengal-741 252, India
†Corresponding author: Soumen Pal

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com
Received: 21-02-2015
Accepted: 15-03-2015

Key Words:
Stochastic modelling
Monthly rainfall
Probability distributions
Return periods

ABSTRACT

The aim of the present study is to find suitable probability distributions which describe best the monthly rainfall volume during monsoon (MRM) for each of the thirteen districts which constitute Gangetic West Bengal (GWB) meteorological sub-division of India. For this purpose, 18 continuous probability distributions are employed to examine the fitness of time series data of 113 years (1901-2013). The parameters of the distributions are estimated either by maximum likelihood, moments or L-moments method. Based on goodness of fit (GOF) tests viz., Kolmogorov-Smirnov, Anderson-Darling and Chi-Squared, three separate rankings are given to each of the distributions for all the districts. Further, a method of scoring, based on ranking, has been adopted to pick up the best fitted distribution for each district. Log-Logistic (LL) distribution comes out as the most suitable for five districts whereas, for a particular district, Generalized Extreme Value (GEV) and Pearson 5 (P5) distributions jointly rank first among all the distributions. Individually, P5 explains best the monthly rainfall of two districts while GEV is found to be most suitable for another district. Log-Pearson 3 (LP3), 3-parameter Dagum (D3), 4-parameter Generalized Gamma (GG4) and 3-parameter Generalized Gamma (GG3) are the others which are found to be most candidate distributions for one district each. Finally, MRM, for all the districts, is estimated for different return periods. Probability of occurrences of MRM, at various points of exceedance, has also been calculated.

INTRODUCTION

West Bengal is an agro-based State of India. The major part of this State falls under the GWB meteorological subdivision (66, 228 sq. km). The region consists of 13 districts viz., Bankura, Birbhum, Burdwan, East Midnapore, Hooghly, Howrah, Kolkata, Murshidabad, Nadia, North 24 Parganas, Purulia, South 24 Parganas and West Midnapore (Fig. 1). On an average, more than 75% of rainfall, in this area, is received during the monsoon season, which not only nourishes the crops grown in the rainy (kharif) season, but enriches all sources of irrigation to enable cultivation of a wide variety of crops during post rainy (rabi) and summer (pre-kharif) season. Surplus monsoon water can be used to recharge underground aquifers. Almost 100 percent of the cultivable area of the region is brought under cultivation, particularly utilizing monsoon rainfall. Thus, the agriculture in GWB is basically dependent on the monsoon and a delay of a few days in arrival of monsoon, can badly affect the economy. However, due to ecological and climatic differences from one place to another, the region has always experienced variation in the spatial distribution of precipitation. Good monsoons correlate with a meaningful economy, whereas weak or failed monsoons (droughts) result in widespread agricultural losses and sustainability, hinder overall economic growth (URL1). Hence, a proper study of rainfall is critical to crop and hydrological planning of this region. Although some work has already been done on precipitation of West Bengal or GWB (Basu et al. 2004, Dastidar et al. 2010, Samanta et al. 2011), studies on the rainfall distribution pattern for this region is restricted locally (Mandal & Choudhury 2014).

The fitting of data on rainfall amount to suitable probability distribution has been a topic for long, as the agriculture in any area does not depend only on the timely occurrence and the amount of rainfall, but also on the maximum and minimum rainfall, range and distribution pattern of rainfall of that particular region. In fact, spatial and temporal distribution of rainfall is more important than total rainfall. This kind of study is not only necessary from an agricultural point of view, as identification of rainfall pattern and probability of occurrence of the rainfall events play a vital role in hydrologic and economic evaluation of water resources projects.

The distribution of summer monsoon rainfall has been studied since long by several researchers (Parida 1999, May 2004). However, there exists no unique probability distribution which can be considered as the one to be fitted best for the rainfall time series of any place. Instead, it depends

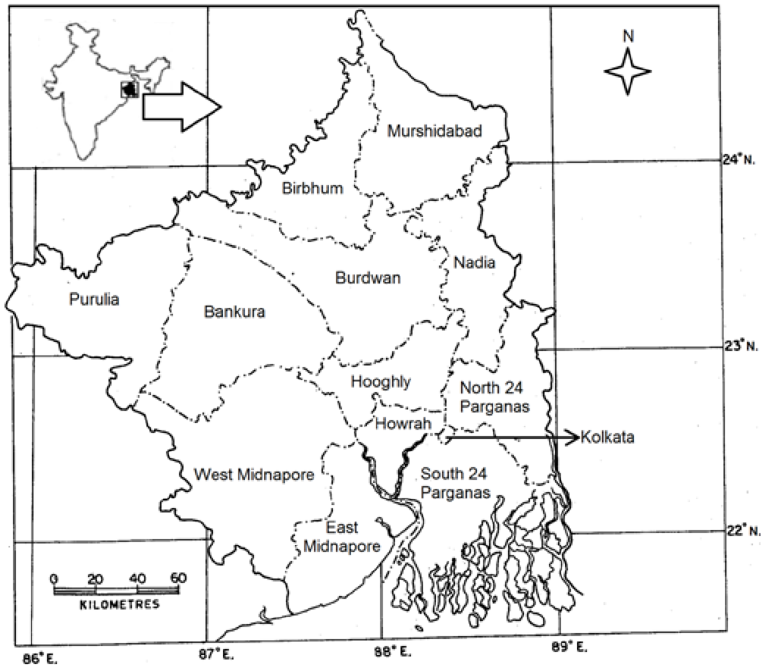

Fig. 1: GWB meteorological sub-division.

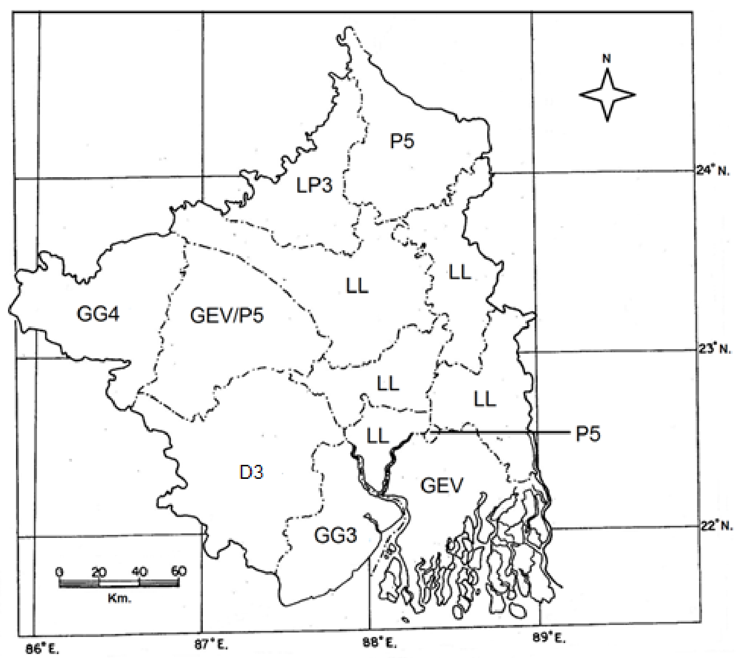

Fig.2: Best fitted distributions for MRM in the districts of GWB.

entirely upon the time scale and place under consideration. There are several previous works on stochastic modelling of rainfall volume to a wide variety of continuous probability distributions. Gamma distribution has been found as the most suitable by many researchers (Husak et al. 2006, Atroosh & Moustafa 2012) for explaining the amount of precipitation in different parts of the world. Another well known distribution in this regard is the Weibull (Muralidharan & Lathika

2005, Yusof & Hui-Mean 2012). The other distributions include exponential (Al-Suhili & Khanbilvardi 2014), generalized extreme value (Zalina et al. 2002, Chu et al. 2013, Win & Win 2014), generalized logistic (Deka et al. 2009, Zin et al. 2009) etc.

In the present study, the standard procedure to find out the suitable distributions for GWB and its districts involves consideration of a number of distributions of similar types and then, based on GOF criteria, selecting a few which are most eligible. If a distribution fits well to the dataset, then the characteristics of rainfall amount have been approximated by the properties of that particular distribution.

MATERIALS AND METHODS

Study area and data collection: District-wise monthly rainfall data (MRD) for 11 districts from 1901 to 2003 have been taken from India Meteorological Department (IMD) for conducting the present study. From 2004 to 2013, the MRD which is maintained by IMD has been collected online from http://www.indiawaterportal.org (2004-2008) and http://www.imd.gov.in (2009-2013). For other two districts, viz., Kolkata and West Midnapur, MRD from 1901 to 2000 has been taken from IMD. From 2001 to 2008, data for these two districts are not available at IMD. Alternatively, those are collected from the website of West Bengal State Marketing Board (http://www.wbagrimarketingboard. gov.in) and again from 2009 to 2013, MRD are obtained from the IMD site (http://www.imd.gov.in). Only 1.9% of the total observations, which are missing in the time series, have been replaced by 100 year's (1901-2000) average monthly rainfall values. All the observations are measured in millimeter (mm).

Fitting probability distributions: MRM data are fitted to a total number of 18 continuous distributions. These include Burr, Dagum (3 and 4 parameters), Erlang, Fatigue Life, Gamma (2 and 3 parameters), Generalized Extreme Value, Generalized Gamma (3 and 4 parameters), Generalized Logistic, Inverse Gaussian, Log-Logistic, Log-Pearson 3, Log normal, Pearson 5 and Pearson 6 (3 and 4 parameters). The parameters of the distributions are estimated either by maximum likelihood, method of moments or method of L-moments.

GOF tests: Once the parameters of the distributions are estimated from sample data, three well-known statistical tests viz., Kolmogorov-Smirnov, Anderson-Darling and Chi-Squared are employed at α (0.01) level of significance to examine the fitness of distributions. The hypothesis under the GOF test is as below:

H_0: MRM data follow the specified distribution.
H_1: MRM data does not follow the specified distribution.

In each case, the hypothesis regarding the distributional form is rejected at the chosen significance level, if the test statistic is greater than the critical value obtained from standard statistical table value. Since the GOF test statistics indicate the distance between the data and the fitted distributions, the distribution with the lowest statistic value is considered as the best fitted model. Based on the test statistic value, each of the 18 distributions used in the present study has been assigned with a rank from 1 to 18. The best model is assigned with rank 1, 2 is given for the next best model likewise. No rank is given to a distribution when the concerned test fails to fit the data. This allows easily comparing the fitted models and selecting the most valid one.

Identification of best fit probability distribution: As the ranks are given for each of GOF separately, it is difficult to identify the best fitted distribution based on three GOF tests for a district. Hence, an approach of scoring (Olofintoye et al. 2009, Sharma & Singh 2010) has been adopted to find out the best fitted model for each district. In this method, for the 18 candidate distributions, a highest score of 18 is given to the one which ranks first and fewer score is awarded to the distribution having rank more than 1 i.e., 17 is given to the distribution securing rank 2 likewise. A lower score of 1 is provided to the distribution which ranks 18 and 0 is given when a distribution fails to fit the data. The score is given to all the distributions for each of the GOF tests ranking separately and the final score is obtained by adding these three scores.

RESULTS AND DISCUSSION

Descriptive statistics: Descriptive statistics viz., minimum, maximum, range, mean, standard deviation (SD), coefficient of variation (CV), skewness and kurtosis of MRM for the 13 districts are summarized in Table 1.

Over the period of 113 years in GWB meteorological sub-division, it has been observed that the maximum MRM was highest (1106.6 mm) and lowest (720.2 mm) in Murshidabad and Nadia districts respectively. Within the same period, the minimum MRM was as low as 22.80 mm in Nadia and highest (73.3 mm) in North 24 Parganas district. The highest and lowest ranges of MRM are observed in Murshidabad (1071.20 mm) and Burdwan (707.90 mm) district respectively. South 24 Parganas district has received the highest mean MRM of 331.48 mm over the time span of 113 years while in Nadia district, the lowest mean MRM of 239.03 mm has been observed during the same period. The SD of MRM in GWB region varies from 107.68 mm to 147.95 mm. South 24 Parganas district obtains the highest SD implying the large variation in MRM while Burdwan district has the lowest SD which indicates comparatively

Table 1: Descriptive statistics of MRM in districts of GWB.

Districts	Statistic							
	Min(mm)	Max(mm)	Range(mm)	Mean(mm)	SD(mm)	CV	Skewness	Kurtosis
Bankura	56.80	814.80	758.00	266.64	113.31	0.42	1.02	1.88
Birbhum	35.20	1021.60	986.40	268.35	117.07	0.44	1.24	3.92
Burdwan	48.60	756.50	707.90	257.47	107.68	0.42	1.01	1.57
East Midnapore	30.30	792.00	761.70	297.89	124.42	0.42	1.01	1.72
Hooghly	37.60	774.60	737.00	267.99	119.30	0.45	1.16	1.78
Howrah	40.70	883.80	843.10	299.49	135.01	0.45	1.13	1.95
Kolkata	40.70	946.20	905.50	315.34	141.55	0.45	1.12	2.27
Murshidabad	35.40	1106.60	1071.20	265.36	121.17	0.46	1.95	8.94
North 24 Parganas	73.30	1041.50	968.20	294.36	130.04	0.44	1.36	3.15
Nadia	22.80	720.20	697.40	241.51	110.95	0.46	1.25	2.24
Purulia	27.10	842.00	814.90	268.79	117.57	0.44	0.77	1.20
South 24 Parganas	33.60	1029.80	996.20	331.48	147.95	0.45	1.18	2.18
West Midnapore	58.80	920.00	861.20	288.27	117.17	0.41	1.15	3.27

Table 2: Score-wise best fit probability distribution with parameter values for districts of GWB.

Districts	Distributions with highest score				
	Distribution	Score	Parameters		
Bankura	GEV	49	κ = -0.02983	σ = 91.393	μ = 216.51
	P5	49	α = 18.211	β = 7863.4	γ = -190.22
Birbhum	LP3	51	α = 37.517	β = -0.07197	γ = 8.1996
Burdwan	LL	54	α = 5.1596	β = 290.71	γ = -50.196
East Midnapur	D3	54	κ = 0.61609	α = 5.1252	β = 325.39
Hooghly	LL	53	α = 4.8824	β = 293.48	γ = -44.688
Howrah	LL	53	α = 5.4399	β = 375.72	γ = -96.833
Kolkata	P5	54	α = 16.697	β = 8505.0	γ = -226.49
Murshidabad	P5	49	α = 12.376	β = 4295.0	γ = -112.46
North 24 Parganas	LL	50	α = 3.8462	β = 250.02	γ = 18.693
Nadia	LL	52	α = 4.5351	β = 252.27	γ = -30.798
Purulia	GG4	48	κ = 1.2078	α = 4.374	
			β = 88.624	γ = -27.211	
South 24 Parganas	GEV	49	κ = 0.0333	σ = 111.37	μ = 263.41
West Midnapur	GG3	53	κ = 1.0101	α = 6.1709	β = 47.626

Table 3: Estimated MRM in various return periods for districts of GWB.

District	Distribution	Return periods (years)						
		2	5	10	15	25	50	100
		Estimated MRM (mm)						
Bankura	GEV	249.82	350.57	415.43	447.11	495.32	553.15	609.36
	P5	249.60	349.95	414.63	446.44	495.31	555.03	614.66
Birbhum	LP3	250.50	356.29	422.76	454.65	502.48	558.89	612.93
Burdwan	LL	240.51	330.12	394.85	429.74	488.03	567.88	658.14
East Midnapur	D3	282.05	382.52	451.55	488.21	549.00	631.84	725.22
Hooghly	LL	248.79	345.16	415.59	453.18	518.00	606.58	707.49
Howrah	LL	278.89	387.94	465.87	507.63	577.05	671.53	777.58
Kolkata	P5	293.22	417.95	499.14	539.28	601.20	677.26	753.63
Murshidabad	P5	244.14	346.36	415.47	450.34	505.00	573.55	643.85
Nadia	LL	221.47	311.67	378.72	415.45	477.60	564.26	664.08
N. 24 Parganas	LL	268.71	377.21	461.36	508.53	589.94	706.43	844.41
Purulia	GG4	254.69	361.80	425.88	455.95	500.36	551.84	600.38
S. 24 Parganas	GEV	304.48	434.70	523.66	568.69	639.29	727.46	817.05
West Midnapur	GG3	273.32	378.10	441.96	472.22	517.24	569.91	620.05

Table 4: Probabilities of MRM at various point of exceedance for districts of GWB.

District	Best fitted distribution	Probability (%) of getting MRM more than 100 mm	200 mm	300 mm	400 mm	500 mm	600 mm	700 mm	800 mm
Bankura	GEV	97.0	69.8	32.7	11.9	3.8	1.1	0.3	0.08
	P5	97.6	68.0	31.1	11.7	4.2	1.5	0.4	0.12
Birbhum	LP3	97.1	69.1	33.7	12.8	4.1	1.2	0.3	0.08
Burdwan	LL	96.8	68.4	27.7	9.5	3.6	1.5	0.8	0.39
East Midnapur	D3	97.6	79.5	43.4	16.8	6.3	2.6	1.2	0.61
Hooghly	LL	96.9	70.1	31.3	11.6	4.7	2.1	1.1	0.57
Howrah	LL	97.1	78.0	42.6	17.9	7.5	3.4	1.6	0.87
Kolkata	P5	97.9	79.6	47.9	23.1	9.9	4.0	1.6	0.66
Murshidabad	P5	97.6	68.1	31.1	11.7	4.2	1.5	0.6	0.23
N. 24 Parganas	LL	98.7	77.5	38.9	16.5	7.5	3.8	2.1	1.23
Nadia	LL	95.2	60.0	22.6	8.1	3.3	1.5	0.8	0.45
Purulia	GG4	95.4	69.4	35.4	13.4	4.0	1.0	0.2	0.04
S. 24 Parganas	GEV	98.8	83.1	51.4	26.0	12.0	5.5	2.5	1.14
West Midnapur	GG3	98.3	76.7	40.8	15.9	5.0	1.3	0.3	0.07

smaller variation in MRM.

The CV values indicate the irregularities in distribution of MRM in the region. Nadia and West Midnapur district posses the highest (0.48) and lowest (0.41) values of CV respectively. Therefore, it can be concluded that MRM in West Midnapur is more consistent than that of any other district, whereas in Nadia district, MRM is most irregular. Skewness measures the asymmetry of a distribution around the mean. For all the districts, the values of skewness are positive indicating that MRM are positively skewed. The maximum skewness (1.95) is obtained in Murshidabad district, while in Purulia district, the MRM distribution is least positively skewed (0.77). Kurtosis provides an idea about the flatness or peaks of the frequency curve. The values of kurtosis, in all the districts, are positive within a range of 1.20 to 9.94. Murshidabad district has the highest value of kurtosis implying the possibility of a distribution having a distinct peak near to the mean with a heavy tail. Purulia district has the smallest value of kurtosis which indicates that the distribution is probably characterized by a relatively flat peak near to the mean.

Distribution fitting: The MRM data for each of the 13 districts are fitted to the 18 continuous probability distributions. Based on Kolmogorov-Smirnov, Anderson-Darling and Chi-squared GOF test statistic values, 3 different rankings have been given to each of the distributions for all the districts. After giving the scores to individual distribution based on GOF test rankings, the total scores of all the distributions have been obtained for each district. The distribution which receives highest score is the best fitted distribution for the corresponding district. Fig. 2 shows the most candidate distributions for all the districts in the GWB region. It is observed that in Bankura district, both the distributions, GEV as well as P5 jointly rank first as most fitted distribution. Otherwise, a single distribution for each of the districts has been found as most suited.

The total score of the best fitted probability distribution along with respective estimated parameters is given for all the districts in Table 2. The estimated MRM in different return periods for the districts is depicted in Table 3. For Bankura district, as two probability distributions (GEV, P5) are fitted best, MRM using both of them are estimated and as expected, the values against different return periods for GEV and P5 are almost same.

In order to have an idea regarding the chances of getting expected MRM, probability of occurrences at various points of exceedance have been calculated (Table 4). For all the districts in the GWB region, there is almost 100% probability of getting MRM of >100 mm while there exits high chances (60 to 83.1%) for receiving MRM of >200 mm. Further analysis reveals a moderate chance (22.6 to 51.4%) in getting MRM of >300 mm, however, for the amount of > 400mm, the probability decreased considerably (26 to 8.1%). Chances are comparatively less (3.3 to 12%) of getting a rainfall of >500 mm in any of the monsoon months while the probability of MRM of >600 mm is considerably low (1 to 5.5%). Receiving MRM of >700 mm and >800 mm can be considered rare events as the probability (0.2 to 2.5% and 0.04 to 1.23% respectively) of occurrence is minute in both the cases.

CONCLUSION

A number of continuous probability distributions have been explored to capture the pattern of monthly rainfall volume during monsoon in the districts of GWB regions of India. No single distribution has uniquely been identified to be the most plausible for the entire region. On the contrary, different distributions prove their goodness of fit for different districts. The predicted rainfall amounts, at various return periods, however, show homogeneity among the districts to some extent. The probabilities, particularly at lower exceedance points, are also indicative to support the similarity. However, as the value of the exceedance points is increased, the dissimilarity in probabilities for precipitation volume shows the differences among the districts. This is also evident from the descriptive statistics for the respective districts. This study, specifically in the summer monsoon season, will be helpful for the crop and hydrological planning of the region.

REFERENCES

Al-Suhili, R.H. and Khanbilvardi, R. 2014. Frequency analysis of the monthly rainfall data at Sulaimania region, Iraq. American Journal of Engineering Research, 3(5): 212-222.

Atroosh, K.B. and Moustafa, A.T. 2012. An estimation of the probability distribution of Wadi Bana flow in the Abyan Delta of Yemen. J. Agric. Sci., 4(6): 80-89.

Basu, G.C., Bhattacharjee, U. and Ghosh, R. 2004. Statistical analysis of rainfall distribution and trend of rainfall anomalies districtwise during monsoon period over West Bengal. MAUSAM, 55(3): 409-418.

Chu, L., McAleer, M. and Wang, S. 2013. Statistical modelling of recent changes in extreme rainfall in Taiwan. International Journal of Environmental Science and Development, 4(1): 52-55.

Dastidar, A.G., Ghosh S., De, U.K. and Ghosh, S.K. 2010. Statistical analysis of monsoon rainfall distribution over West Bengal, India. Mausam, 61(4): 487-498.

Deka, S., Borah, M. and Kakaty, S.C. 2009. Distributions of annual maximum rainfall series of North-East India. European Water, 27/28: 3-14.

Husak, G.J., Michaelsen, J. and Funk, C. 2006. Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications. Int. J. Climatol., 27(7): 935-944.

Mandal, S. and Choudhury, B.U. 2014. Estimation and prediction of maximum daily rainfall at Sagar Island using best fit probability models. Theor. Appl. Climatol., 1-11.

May, W. 2004. Variability and extremes of daily rainfall during the Indian summer monsoon in the period 1901-1989. Global Planet. Change, 44(1): 83-105.

Muralidharan, K. and Lathika, P. 2005. Statistical modelling of rainfall data using modified Weibull distribution. Mausam, 56(4): 765-770.

Olofintoye, O.O., Sule, B.F. and Salami, A.W. 2009. Best-fit probability distribution model for peak daily rainfall of selected cities in Nigeria. New York Science Journal, 2(3): 1-12.

Parida, B.P. 1999. Modelling of Indian summer monsoon rainfall using a four-parameter kappa distribution. Int. J. Climatol., 19(12): 1389-1398.

Samanta, S., Pal, D.K., Lohar, D. and Pal, B. 2011. Modeling of temperature and rainfall of West Bengal through remote sensing and GIS techniques. International Journal of Geoinformatics, 7(2): 2011.

Sharma, M.A. and Singh, J.B. 2010. Use of probability distribution in rainfall analysis. New York Science Journal, 3(9): 40-49.

URL1 http://yojana.gov.in/CMS/(S(et5vxcfhzguhdqf2i03tyw55))/DocumentViewer.aspx?#page1

Win, N.L. and Win, K.M. 2014. The probability distributions of daily rainfall for Kuantan river basin in Malaysia. International Journal of Science and Research, 3(8): 977-983.

Yusof, F. and Hui-Mean, F. 2012. Use of statistical distribution for drought analysis. Applied Mathematical Sciences, 6(21): 1031-1051.

Zalina, M.D., Desa, M.N.M., Nguyen, V-T-A and Kassim, A.H.M. 2002. Selecting a probability distribution for extreme rainfall series in Malaysia. Water Sci. Technol., 45(2): 63-68.

Zin, W.Z.W., Jemain, A.A. and Ibrahim, K. 2009. The best fitting distribution of annual maximum rainfall in Peninsular Malaysia based on methods of L-moment and LQ-moment. Theor. Appl. Climatol., 96(3): 337-344.