American Journal of Environmental Sciences 6 (3): 238-243, 2010
ISSN 1553-345X
© 2010 Science Publications

Modeling the Distribution of Rainfall Intensity using Hourly Data

[1]Salisu Dan'azumi, [1]Supiah Shamsudin and [2]Azmi Aris
[1]Department of Hydraulics and Hydrology, Faculty of Civil Engineering,
[2]Faculty of Civil Engineering, Institute of Environmental and Water Resource Management,
University Technology Malaysia, 81310, Skudai, Johor Malaysia

Abstract: Problem statement: Design of storm water best management practices to control runoff and water pollution can be achieved if a prior knowledge of the distribution of rainfall characteristics is known. Rainfall intensity, particularly in tropical climate, plays a major role in the design of runoff conveyance and erosion control systems. This study is aimed to explore the statistical distribution of rainfall intensity for Peninsular Malaysia using hourly rainfall data. **Approach:** Hourly rainfall data were collected from twelve stations spread across the Peninsular. Six hour separation time was used to divide the data into individual rainfall events and four probability distributions namely, Generalized Pareto (GP), Exponential (EXP), Beta (BT) and Gamma (GM) distributions were used to model the distribution of the hourly rainfall intensity. Kolmogorov-Sminov anderson-Darling and Chi-squared goodness-of-fit tests were used to evaluate the best fit. **Results:** The rainfall frequency, based on 6 h minimum inter-event time, ranges from 115-198 events. The distribution of the rainfall frequency and that of the highest intensity observed, over the recorded period, across the peninsular, is however irregular. The mean rainfall intensity ranges from 2.32-3.88 mm h^{-1}. Kuala-Lumpur and Penang received the highest, while Segamat and Kedah received the lowest. Conversely, over the period of record, Segamat recorded the highest CV, skewness and kurtosis while Pahang has the least value for these parameters. Goodness-of-fit tests at 5% level of significance indicate that all the models can be used to model the distribution of rainfall intensity in Peninsular Malaysia. However, GP is found to be the most suitable model among the four probability distributions tested. **Conclusion:** Basic statistics of hourly rain intensity were obtained and probability distributions compared. It was found that GP is the most suitable model. Results can be useful, particularly, to agricultural and storm water management planning.

Key words: Hourly rainfall intensity, probability distribution, goodness-of-fit test, peninsular Malaysia

INTRODUCTION

Precipitation is caused, in the atmosphere, as a result of rising and cooling air masses. In Malaysia, rainfall is formed by two major wind systems; the North-east monsoon, which occurs from November to March and the South-west monsoon which occurs from May to September. Rainfall occurs throughout the year but the North-east monsoon brings in more rainfall. The country has an average annual rainfall of more than 2000 mm, while humidity ranges from 70-90%. The highly frequent precipitation experienced affects the quantity and quality of surface runoff and ground water, which are essential for human existence. Rainfall pattern affects mans activities in so many ways and as such, designs of agricultural, storm water management, telecommunication, erosion and sediment control systems are highly dependent on rainfall characteristics. High intensity rainfall, particularly if sustained over a long duration, is mostly responsible for altering the geomorphology of a watershed and therefore, knowledge of the distribution of rainfall intensity is important in the design of erosion control and runoff conveyance systems. The aim of this research is to explore the statistical distribution of rainfall intensity for Peninsular Malaysia using hourly rainfall data.

The spatial characteristic of rain rate was developed from a rainfall data in a tropical rainfall measuring mission. Gamma and Log-normal distributions were used to model the rain rate and

Corresponding Author: Salisu Dan'azumi, Department of Hydraulics and Hydrology, Faculty of Civil Engineering, University Technology Malaysia, 81310, Skudai, Johor, Malaysia Tel: +60177809907

results indicate that both the models fit well with the PDF of rainfall data (Cho *et al*., 2004). Similarly, parametric family of PDFs were derived alongside Gamma and Log-normal PDFs for comparing the best fit in modeling rain rate in Darwin and Florida. Result indicates that the Log-normal distribution out-performs both the Gamma and the other family of PDFs (Kedem *et al*., 1994). Two parameter gamma distribution, alongside Markov model were used to describe the distribution of daily rainfall at two sites in Ghana and to predict the occurrence of the rainfall in order to develop a rainfall simulation model; using twenty years of rainfall data; the research confirms the applicability of the gamma distribution (Adiku *et al*., 1997). Continuous hourly rainfall data were derived from rainfall rate recordings at Jardi gauge, Barcelona, Spain. The result shows that the duration of rainfall is Exponentially distributed, that of rainless interval is Generalized pareto and the cumulative rainfall in the cumulative rain duration is Beta distributed (Burgueno *et al*., 1994).

Long term rainfall data at hourly resolution is uncommon in developing countries and modeling of rainfall characteristics for Peninsular Malaysia is mostly conducted using the available daily rainfall data. Deni and Jemain (2008) compared Mixture of geometric distribution with truncated poisons distribution alongside five other distributions namely, Geometric distribution, Compound geometric distribution, Geometric log series distribution, Log series distribution, Modified log series distribution and Truncated negative binomial distribution for modeling the sequence of wet days in Peninsular Malaysia. Result shows that all the data fits the Mixture of geometric distribution with truncated poisons distribution successfully. Similarly, the distribution of dry and wet spell was fitted with probability models namely, Mixture of Log-series distribution, Mixture of log series poisons distribution and Mixture of log series geometric distribution. Chi-square goodness of fit test was used and results indicate that the Mixture of log series geometric distribution and Mixture of log series poisons distribution showed a better fit (Deni and Jemain, 2009). Several types of exponential distributions were used to model daily rainfall data for Peninsular Malaysia and results indicated that mixture of distributions are better in modeling daily rainfall data (Suhaila and Jemain, 2007). Eight probability distributions namely Two parameter gumbel and Gamma, Three parameter generalized normal, Generalized pareto, Generalized extreme value, Pearson type 3, Log-Pearson type 3 and the 5 parameter wakeby were tested using Probability plot correlation coefficient test combined with root mean squared error, relative root mean squared error and maximum absolute deviation and it was concluded that the Generalized extreme value distribution is the most appropriate distribution for describing the annual maximum rainfall series in Malaysia (Zalina *et al*., 2002).

MATERIALS AND METHODS

Data collection and analysis: Precipitation data, consisting of rainfall depth, recorded at hourly intervals were collected from the Department of Irrigation and Drainage, Malaysia. Twelve locations spread across different geographical regions of Peninsular Malaysia were selected for the data collection. The data, which covered period of 10-22 years, was examined and missing records were removed. Figure 1 shows the map of Peninsular Malaysia and locations of the data collection.

Six hour storm separation time was used as Minimum Inter-Event Time (MIT) to separate the data into individual rainfall events. This follows discussions by Guo (2002); Adams *et al*. (1986) and Adams and Papa (2000) for storm water management applications.

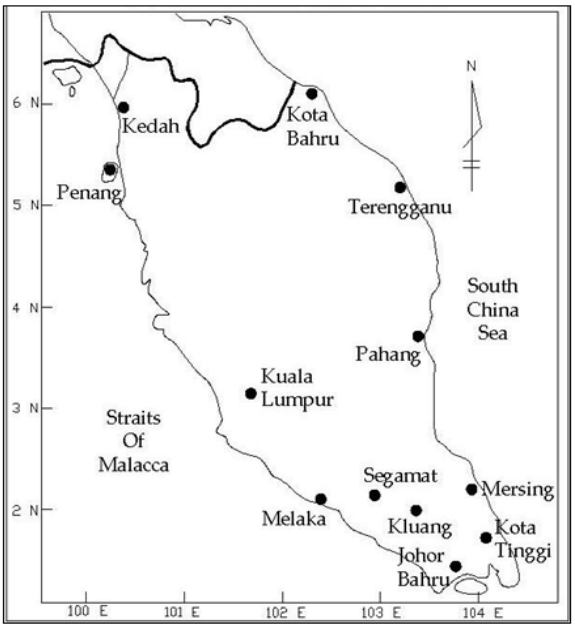


Fig. 1: Map of Peninsular Malaysia showing the rainfall stations

Rainfall events separated by less than 6 h are merged and considered as single event. This gives an average annual number of 163 events for the Peninsular based on the 6hr MIT. The rainfall intensity was obtained by dividing the rainfall depth with the duration from the beginning of the rainfall.

Distribution of rainfall intensity: Generalized pareto, Exponential, Beta and Gamma distributions were used to model the distribution of the rainfall intensity. Empirical cumulative distribution function was determined using the equation:

$$f_n(x) = \frac{1}{n}[\text{Number of observation} \le x] \quad (1)$$

The Probability Density Functions (PDF) and Cumulative Distribution Functions (CDF) for the four models are given as follows: Note that x is the random variable representing the hourly rainfall intensity.

Generalized pareto distribution: The Generalized pare to distribution with continuous shape parameter (k), continuous scale parameter (σ) and continuous location parameter (μ) has a PDF and CDF given by:

$$f(x) = \begin{cases} \frac{1}{\sigma}\left(1+k\left(\frac{x-\mu}{\sigma}\right)^{-1-1/k}\right), k \neq 0 \\ \frac{1}{\sigma}\exp\left(-\frac{(x-\mu)}{\sigma}\right), \quad k = 0 \end{cases} \quad (2)$$

$$F(x) = \begin{cases} 1-\left(1+k\left(\frac{x-\mu}{\sigma}\right)^{-1/k}\right), k \neq 0 \\ 1-\exp\left(-\frac{(x-\mu)}{\sigma}\right) k = 0 \end{cases} \quad (3)$$

Where:

$$\mu \le x \le +\infty \quad \text{for } k \ge 0$$
$$\mu \le x \le \mu - \sigma/k \quad \text{for } k < 0$$

Exponential distribution: The One parameter exponential distribution with scale parameter (λ) is represented by the PDF and CDF given by:

$$f(x) = \lambda e^{-\lambda x} \quad (4)$$

$$F(x) = 1 - e^{-\lambda x} \quad (5)$$

where: λ>0.

Beta distribution: The PDF and CDF of Beta distribution with continuous shape parameters α1 and α2 is given by:

$$f(x) = \frac{1(x-a)^{\alpha 1-1}(b-x)^{\alpha 2-1}}{B(\alpha 1,\alpha 2)\ (b-a)^{\alpha 1+\alpha 2-1}} \quad (6)$$

$$F(x) = 1_z(\alpha 1,\alpha 2) \quad (7)$$

Where:
B = A Beta function
I_Z = The Regularized incomplete beta function
a and b = Boundary parameters (a<b)

Gamma distribution: The Two parameter gamma distribution with continuous shape parameter (α), continuous scale parameter (β) is represented by the PDF and CDF given by:

$$f(x =) \frac{x^{\alpha-1}}{\beta\alpha\Gamma(a)}\exp(-x/\beta) \quad (8)$$

$$F(x) = \frac{\Gamma x/\beta(\alpha)}{\Gamma(\alpha)} \quad (9)$$

Where:
Γ = The gamma function
Γ_x = The Incomplete gamma function

Goodness-of-fit tests: Three goodness-of-fit tests were conducted at 5% level of significance. Note that X denotes the random variable and; n, the sample size. The tests are as follows:

Kolmogorov-Smirnov (K-S) test: This test is used to decide if a sample comes from a hypothesized continuous PDF. It is based on the largest vertical difference between the theoretical and empirical CDF. For a random variable X and sample (x_1, x_2,.....x_n) the empirical CDF of X ($F_x(x)$) is given by:

$$F(x) = \frac{1}{n}\sum_{i=1}^{n} I(x_i \le x) \quad (10)$$

where, I (condition) = 1 if true and 0 otherwise,

Given two cumulative probability functions F_x and F_y, the Kolmogorov-Smirnov test statistics (D_+ and D_-) are given by:

$$D_+ = \max_x\left(F_x(x) - F_y(x)\right) \quad (11)$$

$$D_- = \max_x\left(F_y(x) - F_y(x)\right) \quad (12)$$

Anderson-Darling (A-D) test: The A-D test compares the fit of an observed CDF to an expected CDF. It gives more weight to the tail of the distribution and the test statistic (A^2) is given by:

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln F(X_i) + \ln\left(1 - F(X_{n-i+1})\right)\right] \quad (13)$$

Chi-squared (C-S) test: This test simply compares how well the theoretical distribution fits the empirical distribution PDF. The C-S test statistic is given by:

$$X^2 = \sum_{i=1}^{k}\frac{(o_i - E_i)^2}{E_i} \quad (14)$$

Where:

o_i = The observed frequency for bin i
Ei = The expected frequency for bin i
K = The number of classes

E_i is given by:

$$E_1 = F(x_2) - F(x_1) \quad (15)$$

and x_1 and x_2 are the lower and upper limits for bin i.

RESULTS

Table 1 presents the result of basic rainfall intensity statistics for the twelve stations. The goodness-of -fit test ranking result is shown in Table 2. Similarly, Fig. 2 and 3 present probability difference plot between the empirical CDF and the theoretical CDF. Figure 4 and 5 show the GP distribution fitted on the histogram of rainfall intensity.

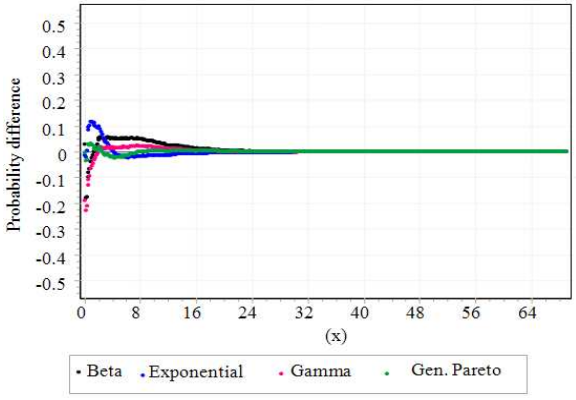


Fig. 2: Probability difference plot showing generalized Pareto, exponential, beta and gamma distributions for Segamat

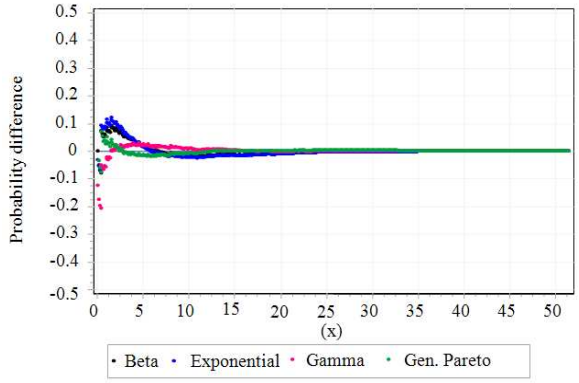


Fig. 3: Probability difference plot showing generalized Pareto, exponential, beta and gamma distributions for Penang

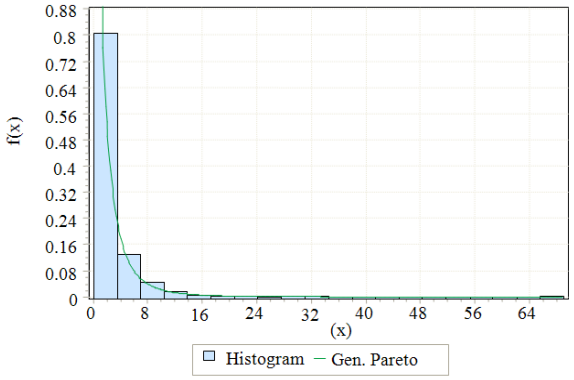


Fig. 4: GP model fitted to the histogram of rainfall intensity for Segamat

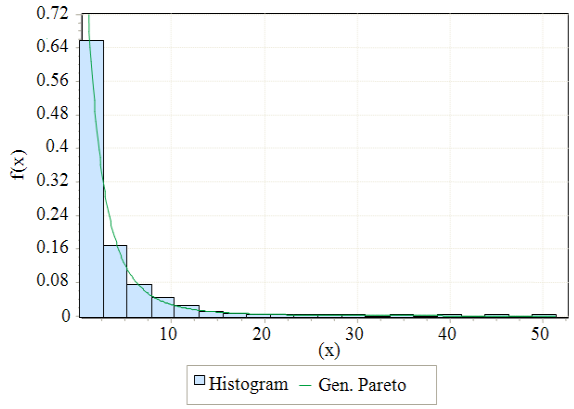


Fig. 5: GP model fitted to the histogram of rainfall intensity for Penang

Table 1: Basic Statistics of hourly rainfall Intensity using 6hr MIT

Station location	Station ID	Period of data collection (years)	Rainfall frequency per Annum	Highest intensity observed (mm h^{-1})	Mean intensity (mm h^{-1})	Standard deviation (mm h^{-1})	CV	Skewness	Kurtosis
Kedah	6108001	1996-2009	194	38.25	2.39	3.21	1.34	3.74	23.14
Penang	5302003	1999-2009	184	51.50	3.13	4.27	1.36	3.71	22.32
Kuala-Lumpur	3116006	1996-2009	198	52.50	3.88	5.24	1.35	3.20	14.98
Melaka	2224038	1988-2009	150	35.00	2.82	3.94	1.40	3.21	13.86
Segamat	2528012	1999-2009	134	69.00	2.32	3.63	1.56	7.45	107.21
Kluang	2033002	2000-2009	137	33.30	2.68	3.83	1.43	3.41	15.98
Johor-Bahru	1834001	1996-2009	175	36.20	2.66	3.63	1.36	3.31	15.97
Khota- Bahru	6122064	1988-2009	115	38.45	2.46	3.71	1.51	3.94	23.16
Terangganu	4131001	1996-2009	173	52.20	3.08	4.63	1.50	3.54	17.20
Pahang	3833002	1988-2009	152	32.00	2.60	3.43	1.32	3.04	13.01
Mersin	2237164	1988-2009	183	44.00	3.04	4.21	1.38	3.35	16.15
Kota-Tinggi	1737001	1996-2008	164	39.00	2.82	3.94	1.40	3.39	16.41

Table 2: Goodness-of-fit test ranking with GP, EXP, BT and GM Distributions

Station Locations	Goodness-of fit tests: Kolmogorov-Smirnov				Anderson-darling				Chi-squared			
	GP	EXP	BT	GM	GP	EXP	BT	GM	GP	EXP	BT	GM
Kedah	1	2	3	4	1	2	4	3	1	2	NA	3
Penang	2	3	1	4	1	2	3	4	NA	2	1	3
Kuala-Lumpur	2	3	1	4	1	3	2	4	1	2	NA	3
Melaka	1	3	2	4	1	3	2	4	1	2	NA	4
Segamat	1	2	3	4	1	2	3	4	1	2	3	4
Kluang	1	2	3	4	1	2	3	4	1	2	3	NA
Johor- Bahru	1	2	4	3	1	2	4	3	1	2	NA	3
Khota- Bahru	1	2	3	4	1	2	4	3	1	2	NA	3
Terangganu	1	2	3	4	1	2	3	4	1	2	3	4
Pahang	1	2	3	4	1	2	3	4	1	2	4	3
Mersin	1	3	2	4	1	3	2	4	NA	1	NA	2
Kota- Tinggi	1	2	3	4	1	2	3	4	1	2	4	3

Note: Ranking is in the order of 1,2,3,4. 1 is the best ranking and 4 the worst ranking; NA: implies not applicable; GP: Generalized Pareto, EXP: Exponential; BT: Beta and GM: Gamma Distribution

DISCUSSION

The rainfall frequency, based on the 6 h MIT, ranges from 115-198 events. Result indicates that Khota-Bahru has the lowest value of 115 while Kuala-Lumpur, the highest value of 198 events per annum. The distribution of rainfall frequency as well as that of the observed highest intensity, over the recorded period, across the peninsular is however irregular. Over the same period, rainfall of as high as 69 mm h^{-1} was observed at a station in Segamat while a closer station in Kluang recorded less than half of this amount. The mean rainfall intensity ranges from 2.32-3.88 mm h^{-1}. Kuala-Lumpur and Penang received the highest, while Segamat and Kedah received the lowest. Conversely, over the period of record, Segamat recorded the highest CV, skewness and kurtosis indicating the spread and peakedness of rainfall intensity in the station while Pahang has the least value for these parameters.

K-S, A-D and C-S goodness-of-fit tests at 5% level of significance indicate that all the models can be used to describe the distribution of rainfall intensity in Peninsular Malaysia. However, GP is found to be the most suitable model among the four probability distributions tested. EXP distribution comes second while BT and GM distributions come third and fourth respectively. Therefore, GP is recommended as the best model.

The probability difference plot shown in Fig. 2 and 3 is calculated as the difference between the empirical CDF and the theoretical CDF given by the equation:

$$\text{Diff}(x) = F_n(x) - F(x) \quad (16)$$

The low intensity events occur at higher frequency and therefore have higher probability difference as compared to the higher intensity events. Also Fig. 4 and 5 displays the GP model (i.e., the best model) matched on the histograms of rainfall intensity for Segamat and Penang stations.

CONCLUSION

Basic statistical characteristics of hourly rainfall intensity for Peninsular Malaysia have been obtained

using hourly data recorded at twelve stations spread across the Peninsular. Four probability distributions, namely, Generalized pareto, Exponential, Beta and Gamma distributions were tested to model the distribution of the hourly rainfall intensity and Kolmogorov-Sminov anderson-Darling and Chi-squared goodness-of-fit tests were used to evaluate the best fit at 5% level of significance. GP is found to be the most suitable distribution for modeling the hourly rainfall intensity. Based on these findings, it is recommended as the best model for describing the hourly rainfall intensity.

ACKNOWLEDGEMENT

The researchers expresse their appreciation to the Department of Irrigation and Drainage, Malaysia for the supply of rainfall data. We sincerely acknowledge the support from University Technology Malaysia and MOSTI for sponsoring this research.

REFERENCES

Adams, B.J. and F. Papa, 2000. Urban Storm water Management Planning with Analytical Probabilistic Models. 1st Edn., John Wiley and Sons, New York, ISBN: 978-0-471-33217-6, pp: 53-79.

Adams, B.J., H.G. Fraser, C.D.D. Howard and M.S. Hanafy, 1986. Meteorological data analysis for drainage system design. J. Environ. Eng., ASCE, 112: 827-848. DOI: 10.1061(ASCE)0733-9372(1986)112:5(827)

Adiku, S.G.K., P.W.A. Dayananda, C.W. Rose and G.N.N. Dowuona, 1997. Analysis of within-season rainfall characteristics and simulation of the daily rainfall in two savannah zones in Ghana. Agric. Forest Meteorol., 86: 51-62. DOI: 10.1016/s0168-1923(96)02414-8

Burgueno, A., B. Codina, A. Redafio and J. Lorente, 1994. Basic statistical characteristics of hourly rainfall amounts in Barcelona (Spain). Theor. Applied. Climatol., 49: 175-181. DOI: 10.1007/bf00865532

Cho, H.K., K.P. Bowman and G.R. North, 2004. A comparison of gamma and lognormal distributions for characterizing satellite rain rates from the tropical rainfall measuring mission. J. Applied Meteorol., 43: 1586-1597. DOI: 10.1175/jam2165.1

Deni, S.M. and A.A. Jemain, 2008. Mixed Geometic truncated Poisson model for sequence of wet days. J. Applied Sci., 8: 3975-3980. DOI: 10.3923/jas.2008.3975.3980

Deni, S.M. and A.A. Jemain, 2009. Fitting the distribution of dry and wet spells with alternative probability models. Meteorol. Atmosphere Phys. 104: 13-27. DOI: 10.1007/s00703-008-0010-7

Guo, J.C.Y., 2002, Overflow risk analysis for Stormwater quality control basins. J. Hydrol. Eng., 7: 428-434. DOI: 10.1061(ASCE)1084-0699(2002)7:6(428)

Kedem, B., Pavlopoulos, H., Guan, X. and D.A. Short, 1994. Probability distribution model for rain rate J. Applied Meteorol., 33: 1486-1493. DOI: 10.1175/1520-0450

Suhaila, J. and A.A. Jemain, 2007. Fitting daily rainfall amount in Peninsular Malaysia using several types of exponential distributions. J. Applied Sci., 3: 1027-1036. DOI: 10.3923/jas.2007.1880.1886

Zalina, M.D., M.N.M. Desa, V.V. Nguyen and A.H.M. Kassim, 2002. Selecting a probability distribution for extreme rainfall series in Malaysia. Water Sci. Technol., 45: 63-68. PMID: 11890166