International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:4, No:12, 2010

# Probability Distribution of Rainfall Depth at Hourly Time-Scale

S. Dan'azumi, S. Shamsudin and A. A. Rahman

***Abstract*—**Rainfall data at fine resolution and knowledge of its characteristics plays a major role in the efficient design and operation of agricultural, telecommunication, runoff and erosion control as well as water quality control systems. The paper is aimed to study the statistical distribution of hourly rainfall depth for 12 representative stations spread across Peninsular Malaysia. Hourly rainfall data of 10 to 22 years period were collected and its statistical characteristics were estimated. Three probability distributions namely, Generalized Pareto, Exponential and Gamma distributions were proposed to model the hourly rainfall depth, and three goodness-of-fit tests, namely, Kolmogorov-Sminov, Anderson-Darling and Chi-Squared tests were used to evaluate their fitness. Result indicates that the east cost of the Peninsular receives higher depth of rainfall as compared to west coast. However, the rainfall frequency is found to be irregular. Also result from the goodness-of-fit tests show that all the three models fit the rainfall data at 1% level of significance. However, Generalized Pareto fits better than Exponential and Gamma distributions and is therefore recommended as the best fit.

***Keywords*—**Goodness-of-fit test, Hourly rainfall, Malaysia, Probability distribution.

## I. INTRODUCTION

A good understanding of the pattern and distribution of rainfall is important for water resource management of a country. Knowledge of rainfall characteristics, its temporal and spatial distribution plays a major role in the design and operation of agricultural systems, telecommunications, run-off control, erosion control as well as water quality systems. Urban areas have short time of concentration, and therefore, the sensitivity of stormwater management systems to rainfall input is very quick and as such, data is mostly needed at a high resolution and hourly rainfall data is most useful. Similarly newer watershed models for urban stormwater management utilize the rainfall data at hourly time scale to predict runoff and its effects on stormwater storage and conveyance systems.

Historical rainfall data exists, in many parts of the world, at daily and monthly time scales and therefore, existing researches characterizing the sequence of rainfall and its probability distributions are mostly conducted at course resolutions. Comparatively, there exist little data on the distribution of rainfall at hourly time-scale and as such few studies are available on that, particularly in the developing world.

Researches investigating the modeling of rainfall can be classified into three main areas, namely, (1) Stochastic models of rainfall relating to global climate change, (2) Stochastic rainfall models describing the generation of sequence of dry and wet spells, and (3) Models of frequency analysis of rainfall.

The effect of global climate change necessitated researches in to the first class of models. They are predictive models that use the existing data to model the trend of future climate as a result of global warming [1]-[4]. The second category includes stochastic models of rainfall processes describing the rainfall occurrence and its characteristics. In this category, generation of the sequence of wet and dry spells is mostly conducted with Markov chain and other models [5]-[8].

This paper falls in the third category. To effectively carry out the stochastic techniques, knowledge of probability distribution of rainfall characteristics is very imperative. Data on the frequency distribution of rainfall events at micro level is scarce in developing countries and therefore, little is known on the characteristics of rainfall at resolutions less than 24hrs. The paper is aimed at exploring the characteristics of hourly rainfall depth in Malaysia. Rao and Biazi [9] applied truncated negative probability model and Markov chain probability model to fit daily rainfall data. A set of 169 longest available records were tested for a period of 100-154 years and the study concluded that the Three Parameter Extreme value Type 2 distribution is the most appropriate in modeling annual maximum rainfall series [10]. Several types of Exponential distributions were tested to model the daily rainfall data for Malaysia and the results indicated that mixture of distributions are better in modeling the daily rainfall data [11]-[12]. Four candidate distributions, namely, Exponential, Gamma, Weibul and Mixed exponential were tested from twelve stations in Wilayah Persekutuan, Malaysia, and it was concluded that the Mixed exponential distribution fits better for hourly rainfall in the area [13]. Also, with IETD of 1hr, 2hr, and 3hr; Gamma distribution is found to fit well to the statistics of rainfall depth in Barcelona, Spain [14].

S. Dan'azumi is with the Faculty of Civil Engineering, Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia (Phone: +60177809907; fax: +6075566157; e-mail: sdanazumi@gmail.com).

S. Shamsudin is with the Faculty of Civil Engineering, Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia (e-mail: supiah@utm.my).

A. A. Rahman is with the Faculty of Management and Human Resource Development, Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia (e mail: m-azmi@utm.my).

## II. Methodology

### *A. Data Collection and Analysis*

Hourly rainfall data, for 12 locations spread across the Peninsular Malaysia, were collected for time period ranging from 10 to 22 years. The mean record length is 137,970 hours. The long term hourly data was examined and missing records were removed. The stations were chosen to represent different geographical areas of the Peninsular (see Fig. 1.).

Six hour inter-event time definition (IETD) is used to separate the individual rainfall data. This is because 6hr is appropriate for urban stormwater management systems design [15]-[17]. Precipitations that are separated by less than six hours are considered as single event. The average annual number of rainfall for the 12 stations based on the 6hr IETD was 163 events.

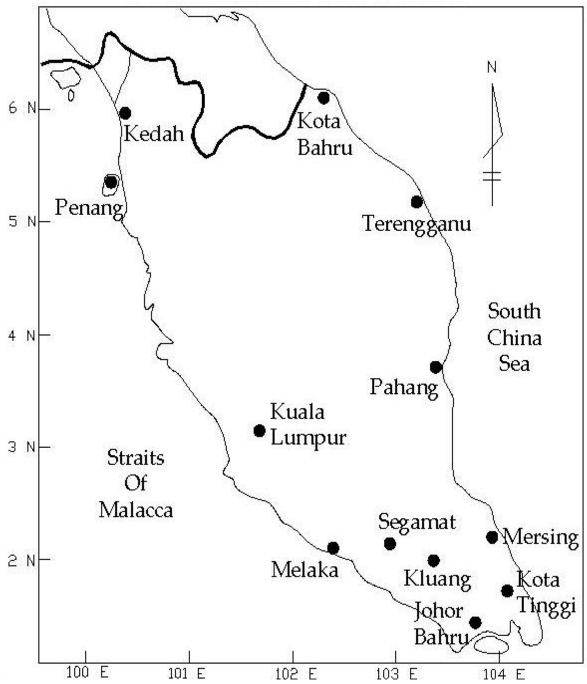


Fig. 1 Map of Peninsular Malaysia showing the rainfall stations

### *B. Modeling the Hourly Rainfall*

Three continuous probability distributions were selected to model the hourly rainfall depth. The three models are presented below along with their probability density functions (PDF). Note that $x$ is the random variable representing the hourly rainfall volume.

1) Generalized Pareto (GP) Distribution

The Generalized Pareto distribution with continuous shape parameter ($k$), continuous scale parameter ($\sigma$), and continuous location parameter ($\mu$) has a PDF given by:

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

$$where \quad \mu \leq x \leq +\infty \qquad for\ k \geq 0$$

$$\mu \leq x \leq \mu - \sigma/_{k} \qquad for\ k < 0$$

2) Exponential (EXP) Distribution

The One parameter exponential distribution with scale parameter ($\lambda$) is represented as the PDF given by:

$$f(x) = \lambda e^{-\lambda x} \qquad for\ \lambda > 0 \quad (2)$$

3) Gamma (GM) Distribution

The Two parameter gamma distribution with continuous shape parameter ($\alpha$), continuous scale parameter ($\beta$) is represented by the PDF given by:

$$f(x) = \frac{x^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)}\exp\left(-x/_{\beta}\right), \quad for\ \gamma \leq x < +\infty \quad (3)$$

where г is the gamma function

### *C. Goodness-of- Fit Tests*

Three goodness-of-fit tests were performed at 1% significant levels. Note also that $X$ denotes the random variable, and; $n$, the sample size. The tests are as follows:

1) 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 ($x_1, x_2, \ldots\ldots\ldots x_n$) the empirical CDF of $X$ ($F_x(x)$) is given by

$$F_x(x) = \frac{1}{n}\sum_{i=1}^{n} I(x_i \leq x) \quad (4)$$

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

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

$$D_+ = \max_x\left(F_x(x) - F_y(x)\right) \quad (5)$$

$$D_- = \max_x\left(F_y(x) - F_x(x)\right) \quad (6)$$

2) 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:

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln F(X_i) + \ln\left(1 - F(X_{n-i+1})\right)\right] \quad (7)$$

3) 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:

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - E_i)^2}{E_i} \quad (8)$$

where $o_i$ and $E_i$ are the observed and expected frequency for bin i respectively and $k$ is the number of classes. $E_i$ is given by

$$E_i = F(x_2) - F(x_1) \quad (9)$$

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

## III. Result And Discussions

Table 1 presents the basic statistics of the hourly rainfall with 6hr IETD. The mean, standard deviation, coefficient of variation (CV), skewness and kurtosis for each of the twelve stations are presented. The rainfall frequency ranges from 115 to 198 per annum. Khota-Bahru; having the lowest, and Kuala-Lumpur; the highest frequency. Generally there is no specific trend on the frequency of rainfall events received over the peninsular. Statistics of highest depth of rainfall over the period of data shows that Pahang received the highest amount of 542.5mm while Kuala-Lumpur, received the lowest amount of 142.4mm. It is also observed that stations located on the east coast of the peninsular receive the highest depth of rainfall as compared to those on the west coast.

Statistics of mean depth indicates that Terangganu, Khota-Bahru and Pahang (located on the east coast) received the highest mean rainfall of more than 16mm while Segamat, Johor-Bahru, Kedah and Melaka (located on the west coast) received the lowest mean rainfall of less than 12mm. Based on the above finding, Khota-Bahru; having being on the top, in terms of the mean rainfall depth and the lowest in terms of average annual number of events will obviously have received rainfall at higher intensity over the period of record. The higher amount of rainfall on the east coast of the peninsular is due to the occurrence of north east monsoon which brings about heavy rainfall. Stations on the west coast are blocked by the main range of Banjaran Titiwangsa, thus affecting the quantity of rainfall they receive (Suhaila and Jemain, 2007).[11].

The variability of rainfall depth indicates that Pahang, Khota-Bahru, Kluang and Mersin have the highest value of CV while Kuala-Lumpur, Melaka and Kedah have the lowest. This indicated that rainfall is more evenly distributed in the west coast as compared to the east coast. All stations are positively skewed and skewness values ranges from 2.37 to 9.67.

Result for goodness-of-fit using K-S, A-D and C-S tests is presented in Table 2 and Figs. 2 and 4 show the three probability distributions fitted to the histogram of the hourly rainfall data at stations Kedah and Johor-Bahru. The goodness-of-fit tests at 1% significance level, indicates that all the three models can be used to fit the hourly rainfall data. However, it is observed that, among the three probability distributions; GP distribution is the best fit in most of the tests. In both K-S and A-D tests; GP distribution was the first candidate in 92% of the total data and then 83% of the total data with C-S test. The overall result shows that GP distribution scores 88% as the first candidate for modeling hourly rainfall data. Similarly, EXP distribution comes second with 67% of the total data and GM distribution comes third with 67% of the total data.

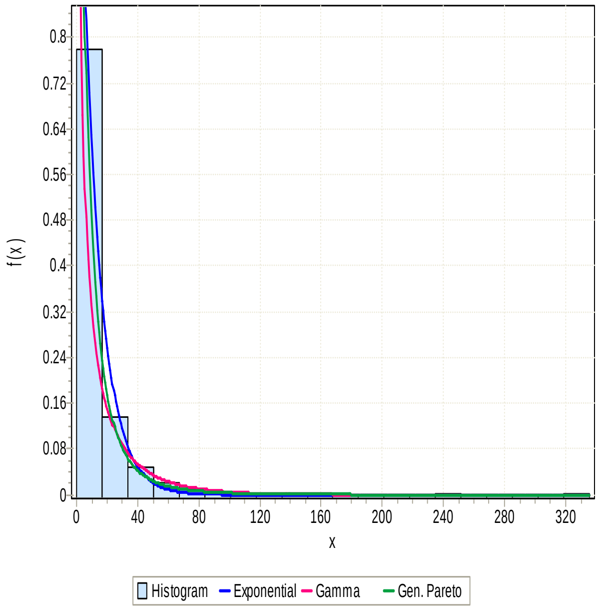


Fig. 2 GP, EXP and GM PDFs fitted to histogram of rainfall depth for Kedah

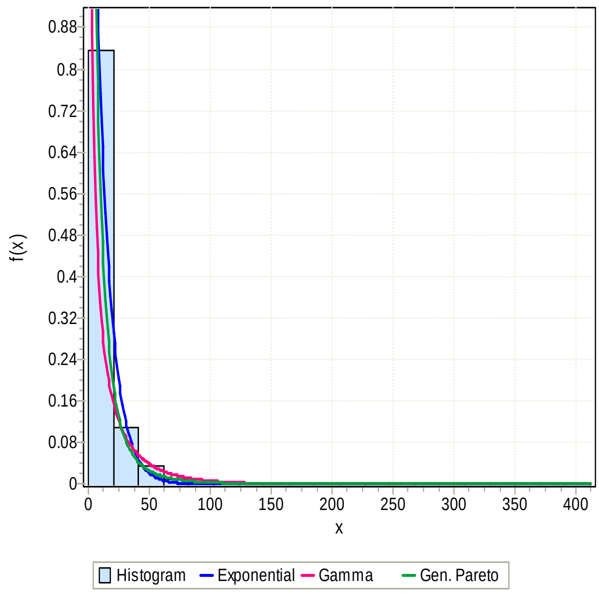


Fig. 3 GP, EXP and GM PDFs fitted to histogram of rainfall depth for Johor

TABLE I
STATISTICAL ANALYSIS OF HOURLY RAINFALL DEPTH AT 6HR IETD

| Station Location | Station ID | Period of Data Collection (years) | Rainfall Frequency per Annum | Highest Depth Observed (mm) | Mean Depth (mm) | Standard Deviation (mm) | CV | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|---|
| Kedah | 6108001 | 1996-2009 | 194 | 335.4 | 11.98 | 20.17 | 1.68 | 6.38 | 75.75 |
| Penang | 5302003 | 1999-2009 | 184 | 363.5 | 13.08 | 13.08 | 1.74 | 5.46 | 50.26 |
| Kota- Tinggi | 1737001 | 1996-2008 | 164 | 251.4 | 12.07 | 19.53 | 1.62 | 4.52 | 33.74 |
| Johor- Bahru | 1834001 | 1996-2009 | 175 | 411.7 | 11.31 | 21.26 | 1.88 | 8.73 | 128.52 |
| Melaka | 2224038 | 1988-2009 | 150 | 283.8 | 11.38 | 16.63 | 1.46 | 3.71 | 30.01 |
| Pahang | 3833002 | 1988-2009 | 152 | 542.5 | 16.69 | 40.54 | 2.43 | 7.34 | 70.10 |
| Kuala Lumpur | 3116006 | 1996-2009 | 198 | 142.4 | 14.29 | 19.07 | 1.33 | 2.37 | 6.99 |
| Terangganu | 4131001 | 1996-2009 | 173 | 515.6 | 16.97 | 33.74 | 1.99 | 6.80 | 71.14 |
| Khota- Bahru | 6122064 | 1988-2009 | 115 | 463.1 | 16.87 | 35.66 | 2.11 | 6.66 | 60.26 |
| Mersin | 2237164 | 1988-2009 | 183 | 519.7 | 14.14 | 28.37 | 2.01 | 6.71 | 71.03 |
| Kluang | 2033002 | 2000-2009 | 137 | 482.3 | 12.71 | 26.59 | 2.09 | 9.67 | 143.74 |
| Segamat | 2528012 | 1999-2009 | 134 | 243.4 | 11.28 | 19.81 | 1.76 | 5.53 | 48.43 |

TABLE II
RESULT OF GOODNESS-OF-FIT TEST RANKING FOR THREE PROBABILITY DISTRIBUTIONS AT 1% LEVEL OF SIGNIFICANCE

| Station Location/Goodness of Fit Test | Kolmogorov-Smirnov (K-S) | | | Anderson-Darling (A-D) | | | Chi- Squared (C-S) | | | Best Fitted Model |
|---|---|---|---|---|---|---|---|---|---|---|
| G.O.F. Test Ranking | 1st | 2nd | 3rd | 1st | 2nd | 3rd | 1st | 2nd | 3rd | |
| Kedah | GP | EXP | GM | GP | GM | EXP | GP | GM | EXP | GP |
| Penang | GP | EXP | GM | GP | GM | EXP | GP | EXP | GM | GP |
| Kota Tinggi | GP | EXP | GM | GP | GM | EXP | GP | EXP | GM | GP |
| Johor-Bahru | GP | EXP | GM | GP | EXP | GM | GP | EXP | GM | GP |
| Melaka | GP | GM | EXP | GP | GM | EXP | GM | GP | EXP | GP |
| Pahang | GP | EXP | GM | GP | EXP | GM | GP | GM | EXP | GP |
| Kuala Lumpur | GM | GP | EXP | GM | GP | EXP | GM | GP | EXP | GM |
| Terangganu | GP | EXP | GM | GP | GM | EXP | GP | EXP | GM | GP |
| Khota-Bahru | GP | EXP | GM | GP | EXP | GM | GP | EXP | GM | GP |
| Mersin | GP | EXP | GM | GP | EXP | GM | GP | EXP | GM | GP |
| Kluang | GP | EXP | GM | GP | EXP | GM | GP | EXP | GM | GP |
| Segamat | GP | EXP | GM | GP | EXP | GM | GP | EXP | GM | GP |

Note: GP: Generalized Pareto, EXP: Exponential and GM: Gamma Distributions

## IV. CONCLUSIONS

In this study, Statistical analysis of hourly rainfall was conducted and three models namely GP, EXP and GM distributions were employed as candidates for modeling hourly rainfall depth in Peninsular Malaysia. Three goodness-of-fit tests namely, C-S, A-D and C-S were used at 1% level of significance.

Result shows that the rainfall frequency across the Peninsular is irregular, but generally towns located on the eastern cost of the peninsular receive higher mean and maximum rainfall depth as compared to those on the western cost. The coefficient of variation is also higher at the eastern cost, indicating high variability. Result of goodness-of-fit test shows that all the three models can be accepted. However, GP distribution fits better than EXP and GM and is therefore recommended as the best model to fit hourly rainfall depth in Peninsular Malaysia. The paper highlights the need to also model other rainfall characteristics of duration, intensity and length of dry spell preceding the rainfall event at hourly time scale in Malaysia.

ACKNOWLEDGMENT

The authors wish to acknowledge the support from Universiti Teknologi Malaysia and MOSTI for sponsoring this research. We sincerely appreciate the help from the staff of Department of Irrigation and Drainage, Malaysia in supplying rainfall data.

REFERENCES

[1] Burlando, P. and R. Rosso, *Effects of transient climate change on basin hydrology. 1. Precipitation scenarios for the Arno River, central Italy.* Hydrol. Process. 16, Wiley Interscience, 2002. **16**: p. 1151-1175.
[2] Hutchinson, M.F., *Stochastic space-time weather models from ground-based data* Agricultural and Forest Meteorology, 1995. **73**(3-4): p. 237-264.
[3] Semenov, M.A. and E.M. Barrow, *Use of a Stochastic Weather Generator in the Development of Climate Change Scenarios.* Climatic Change, 1997. **35**: p. 397-414.
[4] Toews, M.W. and D.M. Allen, *Evaluating different GCMs for predicting spatial recharge in an irrigated arid region.* Journal of Hydrology, Elsevier 2009. **374**: p. 265-284.
[5] Bonta, J.V., *Stochastic Simulation of Storm Occurrence, Depth, Duration, And Within−Storm Intensities.* Transactions of the American Society of Agricultural Engineers, ASAE 2004. **47**(5): p. 1573-1584.
[6] Khalili, M., R. Leconte, and F. Brissette, *Stochastic Multisite Generation of Daily Precipitation Data Using Spatial Autocorrelation.* Journal of Hydrometeorology, 2006. **8**: p. 396-412.
[7] Srikanthan, R., T.A. McMahon, and A. Sharma, *Stochastic Generation of Monthly Rainfall Data*, in *Technical Report 02/8*. 2002, Cooperative Research Centre for Catchment Hydrology.
[8] Unal, N.E., H. Aksoy, and T. Akar, *Annual and monthly rainfall data generation schemes.* Stoch Envir Res and Risk Ass., 2004. **18**: p. 245-257.
[9] Rao, N.J.M. and E. Biazi, *Probability Distribution Models for Daily Rainfall Data for an Interior Station of Brazil.* Arch. Met. Geoph. Biocl., 1983. **Set. B**(33): p. 261-265.
[10] Koutsoyiannis, D., *Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records.* Hydrological Sciences–Journal–des Sciences Hydrologiques, 49(4) August 2004, 2004. **49**(4): p. 591-610.
[11] Suhaila, J. and A.A. Jemain, *Fitting Daily Rainfall Amount in Peninsular Malaysia Using Several Types of Exponential Distributions.* Journal of Applied Scinces Research, 2007. **3**(10): p. 1027-1036.
[12] Suhaila, J. and A.A. Jemain, *Fitting the Statistical Distribution for Daily Rainfall in Peninsular Malaysia Based on AIC Criterion.* Journal of Applied Sciences Research, 2008. **4**(12): p. 1846-1857.
[13] Fadhilah, Y.; Zalina, M.; Nguyen, V. T. V.; Suhaila, S.; and Zulkifli, Y., *Fitting the Best-Fit Distribution for the Hourly Rainfall Amount in the Wilayah Persekutuan.* Jurnal Teknologi, UTM, 2007. **46**(C): p. 49-58.
[14] Burguefio, A.; Codina, B.; Redafio, A.; and Lorente J., *Basic Statistical Characteristics of Hourly Rainfall Amounts in Barcelona (Spain).* Theor. Appl. Climatol. Springer Verlag, 1994. **49**: p. 175-181.
[15] Adams, B.J.; Fraser, H.G, Charles, D.D., and Hanafy, M.S.; *Meteorological Data Analysis For Drainage System Design.* Journal of Environmental Engineering, ASCE, 1986. **112**(5): p. 827-848.
[16] Guo, J.C.Y., *Overflow Risk Analysis for Stormwater Quality Control Basins.* Journal of Hydrologic Engineering, 2002. **7**(6): p. 428-434.
[17] Adams, B.J. and F. Papa, *Urban Stormwater Management Planning with Analytical Probabilistic Models*. 2000, New York: John Wiley & Sons. p. 53-79.