WEST INDIAN JOURNAL OF ENGINEERING TECHNICAL PAPER

COMPARATIVE EVALUATION OF FOUR PROBABILITY DISTRIBUTIONS FOR TRINIDAD MONTHLY RAINFALL

R.J. Stone*

ABSTRACT

A comparative study of four probability distributions for representing monthly rainfall amounts was conducted using data recorded at six locations in Trinidad. The parameters of the distributions were estimated by the method of maximum likelihood while their performances were assessed by the χ^2 goodness-of-fit test. The results indicate that, generally, the gamma distribution was the most acceptable ($\alpha = 0.05$). However, the performances of the Gaussian (Normal) and log-Gaussian (Lognormal) distributions were superior to the gamma distribution at some locations for various months. The log-Pearson type 3 distribution performed poorly at all the locations. This study, therefore, clearly demonstrates that the present practice of using the log-Gaussian distribution throughout Trinidad for all months is not appropriate. Instead, the engineer should also consider the gamma and Gaussian distributions as possible candidates for the representation of monthly rainfall data.

1.0 INTRODUCTION

The analysis of hydrologic data, such as monthly rainfall, is of critical importance for the efficient planning and design of water resources engineering projects. However, since it is not possible to predict, with certainty, monthly rainfall in future years, estimates are routinely computed using a probabilistic approach. In this approach, the non-exceedance probabilities of various rainfall amounts are determined by fitting an appropriate probability distribution to the historic data recorded at the location of interest. In many countries, including Trinidad and Tobago, the distribution most widely used for this purpose is the log-Gaussian [1].

However, several studies have indicated that better fits may be obtained with other probability distributions. The alternative distributions reported in the literature are the Gaussian [2], the gamma [2,3,4] and the log-Pearson type 3 [2].

Choosing an appropriate distribution from the suite of available distributions is essential for optimal planning and design. The choice of an inappropriate distribution, for example, could result in under- or over-design of an irrigation or drainage system. Under-design will result in frequent failures of the system leading to serious losses while over-design constitutes a waste of scarce and expensive capital resources.

The objective of this study, therefore, was to carry out a comparative evaluation of the Gaussian (GAU), log-Gaussian (LGAU), gamma (GAM) and log-Pearson type 3 (LP3) distributions to assess their suitability for representing monthly rainfall data in Trinidad. These four distributions were chosen because they have been reported in the technical literature to have performed well and/or are routinely used in various countries.

2.0 DESCRIPTION OF THE PROBABILITY DISTRIBUTIONS

Any random variable X can be represented by a probability distribution function. The distribution function specifies the chance that an observation χ of the variable will fall within a particular range of X. For example, if X is the January monthly rainfall at a particular location, then the probability distribution function expresses the chance that the January rainfall will be less than or equal to a given amount e.g. r mm (say). This then will be the probability of non-exceedance of r mm rainfall for the month of January at the particular location.

If $p(\chi)$ is the probability density function of X, then for a given value of χ, $P(\chi)$ is the probability distribution function and is expressed as the integral of the probability density function over the range $X \leq \chi$, i.e.

$$P(x) = \int_{-\infty}^{x} p(u)\, du \tag{1}$$

where u is a dummy variable of integration.

Thus, the probability of non-exceedance of any value r, $P(X \leq r)$, can be determined using equation 1 which gives

$$P(X \leq r) = \int_{-\infty}^{r} p(x)\, dx \tag{2}$$

* *Department of Crop Science, The University of the West Indies*

Pertinent discussion will be published in July 1995 West Indian Journal of Engineering if received by May, 1995

2.1 The Gaussian Distribution

The probability density function of the Gaussian distribution is given by

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \qquad (3)$$

$$-\infty < x < \infty, \; -\infty < \mu < \infty, \; \sigma > 0,$$

where μ and σ are the location and scale parameters respectively.

2.2 The Log-Gaussian Distribution

The probability density function of the log-Gaussian distribution is given by

$$p(x) = \frac{1}{x\sigma_y\sqrt{2\pi}} \exp\left(-\frac{(y-\mu_y)^2}{2\sigma_y^2}\right), \qquad (4)$$

$$x > 0, \; -\infty < \mu_y < \infty, \; \sigma_y > 0,$$

where $y = \ln\chi$ and, μ_y and σ_y are the scale and shape parameters respectively.

2.3 The Gamma Distribution

The probability density function of the gamma distribution is given by

$$p(x) = \frac{\lambda^\beta x^{(\beta-1)} e^{-\lambda x}}{\Gamma(\beta)} \qquad (5)$$

$$x \geq 0, \; \lambda > 0, \; \beta > 0,$$

where λ and β are the scale and shape parameters respectively. Γ is the well known gamma function.

2.4 The Log-Pearson Type 3 Distribution

The probability density function of the log-Pearson type 3 distribution is given by

$$p(x) = \frac{\lambda^\beta (y-\epsilon)^{(\beta-1)} e^{-\lambda(y-\epsilon)}}{x\,\Gamma(\beta)} \qquad (6)$$

$$\ln x \geq \epsilon, \; \lambda > 0, \; \beta > 1, \; \epsilon > 0,$$

where $y = \ln\chi$, and λ, β and ε are the scale, shape and location parameters respectively.

3.0 ESTIMATION OF PARAMETERS

Although there are several methods for estimating the parameters of probability distributions, the preferred method is the method of maximum likelihood [5,6]. The likelihood function, L, is given by

$$L = \prod_{i=1}^{n} p(x_i; \mu, \sigma, ...) \qquad (7)$$

where $p(\chi_i; \mu\sigma, ...)$ is the probability density function of the selected distribution, $\mu, \sigma, ...,$ are the parameters to be estimated, and $\chi_1, \chi_2, ..., \chi_n$ are the n rainfall data points. In the method of maximum likelihood, the parameters $\mu, \sigma, ...,$ are estimated so that L attains its maximum value. In practice, ln L, which is a monotonically increasing function of L, referred to as the log likelihood function, is employed in preference to L for convenience and ease of computation [7]. The likelihood equation may therefore be written as

$$\ln L = \ln \prod_{i=1}^{n} p(\chi_1; \mu, \sigma, ...) = \sum_{i=1}^{n} \ln p(\chi_1; \mu, \sigma, ...) \quad (8)$$

Estimates of the parameters are obtained by taking the partial derivatives of ln L with respect to each parameter and equating the results to zero i.e.

$$\frac{\partial}{\partial\mu}(\ln L) = 0; \qquad \frac{\partial}{\partial\sigma}(\ln L) = 0; \; ... \qquad (9)$$

4.0 ASSESSMENT OF PERFORMANCE

The chi-square (χ^2) test is the oldest, most commonly used, and perhaps the most versatile procedure for assessing the goodness-of-fit of probability distributions to sample data [8]. In this test, the data is divided into a discrete number of intervals and a comparison is made between the actual number of observations and the expected number of observations (expected according to the distribution under test) that fall in the various class intervals. The expected number of observations for a particular interval is calculated by multiplying the probability that an observation falls within the interval by the total number of actual observations. The chi-square test statistic, χ_c^2, is calculated using the relationship

$$\chi_c^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad (10)$$

where k is the number of class intervals, O_i is the actual and E_i the expected (according to the distribution under test) number of observations in the ith class interval. The distribution of χ_c^2 is a chi-square distribution with k - m - 1 degrees of freedom where m is the number of parameters estimated from the data. The null hypothesis, H_0, that the data are from the distribution tested is rejected if the following occurs

$$\chi_c^2 > \chi_{\alpha,k-m-1}^2 \quad (11)$$

where $\chi_{\alpha,k-m-1}^2$ is the critical chi-square value obtained from statistical tables of the chi-square distribution and α the level of significance [9]. The most commonly used values of α are 0.05, 0.01 and 0.001. The larger the value of α, the less likely is the acceptance of a false null hypothesis. A value of α = 0.05 was used in this study.

5.0 DATA USED IN THE STUDY

The data used in the study were obtained from the Piarco Meteorological Office and the Water Resources Agency. **Figure 1** shows the six locations which were chosen to be representative of Trinidad covering the northern, central and southern regions of the island. The locations and the corresponding period of the rainfall data used were: Piarco (Meteorological Office), 1946 - 1992; Port of Spain (Botannical Gardens), 1862 - 1991; Sangre Grande (Grosvenor Estate), 1927 - 1957, 1965 - 1992; Couva (Exchange), 1941 - 1992; Rio Claro (Poole Syndicate), 1932 - 1991; and Cedros (Perseverance Estate), 1925 - 1992.

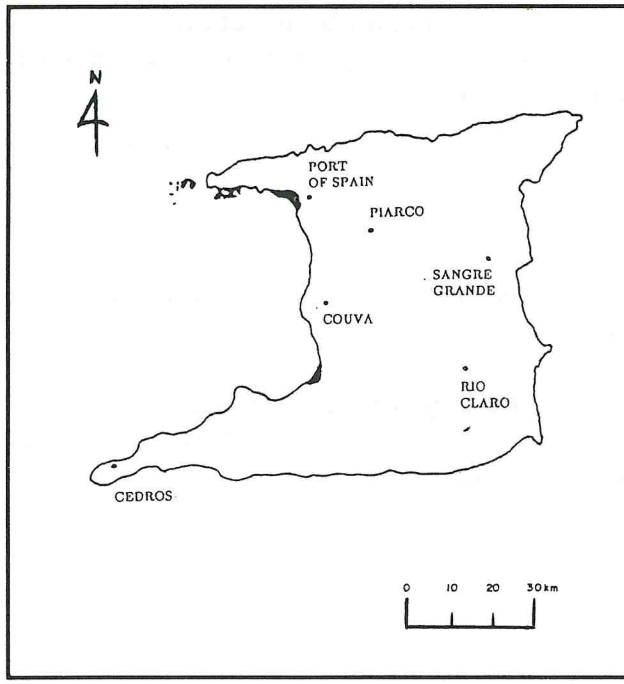

Figure 1. Geographical positions of the six locations in Trinidad.

6.0 RESULTS AND DISCUSSION

Tables 1-6 show the calculated chi-square values for the four distributions at the six locations in Trinidad, namely, Piarco, Port of Spain, Sangre Grande, Couva, Rio Claro and Cedros respectively. At Piarco, only GAM was acceptable for all months. However, LGAU also performed acceptably for all months except January. On the other hand, although GAU only performed acceptably for the rainy season months (June to December) and January, its performance during these months were generally superior to the other distributions. LP3 performed poorly for all the months.

At Port of Spain, GAM performed acceptably for all the months except February for which its performance was only marginally unacceptable. GAU was acceptable for the rainy season months (June to December) and LGAU for eight months. LP3 was only acceptable for four months. However, for two of those months, January and February, LP3 performed better than the other distributions.

At Sangre Grande, only GAM was acceptable for all the months. GAU and LGAU were acceptable for 10 and 11 months respectively while LP3 was unacceptable for all the months. On the other hand, LGAU perfomed best for more months than the other distributions.

At Couva, GAM, LGAU, GAU and LP3 were acceptable for ten, seven, six and two months respectively. However, only GAU was acceptable for September and November. Generally, GAM performed the best.

At Rio Claro, both LGAU and GAM were acceptable for all months. GAU was acceptable for nine months while LP3 was only acceptable for one month. GAM generally performed better than the other distributions.

At Cedros, only GAM was acceptable for all months. LGAU and GAU were acceptable for eleven and six months repectively while LP3 was only acceptable for one month. GAM generally performed better than the other distributions.

From the above, it is clear that overall, GAM is the most appropriate distribution for describing Trinidad monthly rainfall data. However, at some locations LGAU and GAU performed better than GAM for some months. LP3 gave the worst overall performance. In practice then, the engineer should place at his disposal all three distributions, GAM, LGAU and GAU, since no one distribution might be applicable for all months

MONTH	GAU	LGAU	GAM	LP3
January	9.1	16.4	11.8	32.1
February	17.2	6.8	2.6	33.2
March	29.0	9.5	12.2	22.5
April	45.9	2.6	6.0	17.9
May	32.1	3.0	11.0	22.5
June	4.5	6.0	3.7	84.2
July	4.9	6.0	5.7	55.5
August	5.7	6.0	10.3	22.5
September	4.5	9.5	7.6	77.7
October	6.0	10.6	10.3	91.5
November	1.5	10.3	7.2	51.2
December	4.9	7.2	3.0	44.7

The critical chi-square value, $\chi^2_{0.05,8} = 15.5$.

Table 1: The χ^2_c values for the Four Distributions at Piarco.

MONTH	GAU	LGAU	GAM	LP3
January	46.3	54.0	25.7	20.2
February	96.2	79.9	30.9	15.9
March	48.2	81.1	23.2	28.5
April	82.6	30.0	8.5	11.5
May	46.6	26.3	15.2	44.2
June	19.9	19.2	13.4	51.5
July	22.6	23.5	28.2	122.3
August	20.8	24.2	13.4	33.7
September	19.9	16.5	18.3	106.3
October	29.1	22.0	21.7	117.1
November	27.9	47.2	28.5	98.3
December	18.3	28.2	15.9	131.9

The critical chi-square value, $\chi^2_{0.05,19} = 30.1$.

Table 2: The χ^2_c values for the Four Distributions at Port of Spain.

MONTH	GAU	LGAU	GAM	LP3
January	8.3	4.2	4.6	50.1
February	15.4	12.0	12.0	104.1
March	6.1	14.6	3.8	23.2
April	19.9	25.8	16.5	29.2
May	27.3	10.2	11.3	20.6
June	3.5	8.3	6.1	118.7
July	14.3	5.3	10.2	35.5
August	8.7	10.9	12.8	108.6
September	11.7	3.1	3.8	113.8
October	14.3	3.5	3.8	36.6
November	6.1	18.0	7.6	48.2
December	9.8	4.6	10.2	35.9

The critical chi-square value, $\chi^2_{0.05,10} = 18.3$.

Table 3: The χ^2_c values for the Four Distributions at Sangre Grande.

MONTH	GAU	LGAU	GAM	LP3
January	24.2	28.0	8.0	38.0
February	43.0	17.2	6.9	76.5
March	23.8	29.9	13.0	14.2
April	28.8	13.4	8.0	23.0
May	21.1	3.4	6.5	30.3
June	14.9	4.9	13.4	44.9
July	7.6	3.0	4.2	51.1
August	6.9	3.0	1.5	82.6
September	6.1	20.3	18.4	33.8
October	17.2	5.3	6.1	15.7
November	8.0	68.4	23.8	24.9
December	15.7	5.7	3.0	28.8

The critical chi-square value, $\chi^2_{0.05,9} = 16.9$.

Table 4: The χ^2_c values for the Four Distributions at Couva.

MONTH	GAU	LGAU	GAM	LP3
January	16.8	10.4	9.2	55.2
February	19.2	8.0	8.0	61.2
March	24.0	10.4	7.2	46.4
April	34.8	9.2	9.2	35.6
May	34.4	18.0	16.0	87.2
June	12.0	6.4	3.6	14.8
July	6.0	9.6	5.2	65.2
August	10.8	10.4	6.8	60.4
September	6.4	10.0	7.6	37.6
October	7.2	9.2	8.0	120.8
November	10.0	14.4	7.6	20.0
December	10.4	5.2	6.8	49.2

The critical chi-square value, $\chi^2_{0.05,11} = 19.7$.

Table 5: The χ^2_c values for the Four Distributions at Rio Claro.

MONTH	GAU	LGAU	GAM	LP3
January	23.8	13.1	18.8	57.4
February	21.5	10.8	18.0	53.6
March	57.4	10.0	10.8	9.6
April	59.3	21.9	11.2	63.5
May	24.5	5.8	4.7	40.2
June	14.2	7.7	6.9	33.7
July	10.4	9.6	11.2	36.0
August	10.0	12.7	13.8	133.9
September	18.4	15.4	11.5	103.7
October	9.2	18.0	6.9	97.2
November	21.5	13.1	18.8	38.3
December	5.8	5 0	3.9	31.0

The critical chi-square value, $\chi^2_{0.05,12} = 21.0$.

Table 6: The χ^2_c values for the Four Distributions at Cedros.

Return Period (yr)	Monthly Rainfall (mm)			
	GAU	LGAU	GAM	LP3
2	41	24	29	30
5	69	89	66	63
10	84	175	93	92
25	99	360	130	138
50	109	571	156	180
100	119	879	184	216

Table 7: Monthly Rainfall Estimates of Various Return Periods for the Four Distributions at Port of Spain for March.

throughout Trinidad.

Table 7 shows the monthly rainfall estimates of various return periods for the four distributions at Port-of-Spain for March. In this particular case, GAM provides the best fit of the data (from **Table 2**). It can be seen that if LGAU were to be used instead of GAM, gross errors would result. For example, the 2-yr return period estimate of LGAU is 17% less than the corresponding estimate for GAM. In addition, the 5, 10, 25, 50 and 100-yr return period estimates of LGAU are greater than the corresponding GAM estimates by 35, 88, 177, 266 and 378% respectively. This example therefore clearly demonstrates to the practising engineer the importance of selecting an appropriate distribution for a given situation.

7.0 CONCLUSIONS

A comparative study was conducted to assess the suitability of four probability distributions for representing monthly rainfall data in Trinidad. The results indicate that the order of decreasing overall performance was gamma, log-Gaussian, Gaussian, log-Pearson type 3. The study, therefore, clearly demonstrates that the present practice of using the log-Gaussian distribution to describe monthly rainfall data throughout Trinidad for all months is inappropriate. Instead, the gamma and Gaussian distributions must be added to the analytical toolbox of the engineer and used whenever they more accurately represent the data.

ACKNOWLEDGEMENTS

The author wishes to thank the Piarco Meteorological Office and the Water Resources Agency for providing the data used in the study.

REFERENCES

1. Shaw, E.M. (1983). *"Hydrology in Practice"*, Van Nostrand Reinhold (UK) Co. Ltd, Berkshire.

2. Linsley, R. K., Kohler, M. A. and Paulhus, J.L.H. (1982). *"Hydrology for Engineers"*, 3rd edition, McGraw-Hill Inc., New York.

3. Haan, C.T. (1977). *"Statistical Methods in Hydrology"*, Iowa State University Press, Ames, Iowa.

4. Chow, V.T., Maidment, D.R. and Mays, L.W. (1988). *"Applied Hydrology"*, McGraw-Hill Inc., Singapore.

5. Law, A.M. and Kelton, W.D. (1991). *"Simulation Modeling and Analysis"*, McGraw-Hill Inc., New York.

6. Kite, G.W. (1991). *"Hydrologic Applications: Computer Programs for Water Resources Engineering"*, Water Resources Publications, Littleton, Colorado.

7. Kottegoda, N. T. (1980). *"Stochastic Water Resources Technology"*, The Macmillan Press Ltd., London.

8. Hahn, G.J. and Shapiro, S.S. (1967). *"Statistical Models in Engineering"*, John Wiley & Sons Inc., New York.

9. Walpole, R.E. and Myers, R.H. (1989). *"Probability and Statistics for Engineers and Scientists"*, 4th edition, Macmillan Inc., New York.