Research Journal of Mathematical and Statistical Sciences ISSN 2320–6047
Vol. **2(9),** 1-8, September **(2014)** *Res. J. Mathematical and Statistical Sci.*

An Investigation on the Stochastic Modeling of Daily Rainfall Amount in the Mahanadi Delta Region, India

M.K. Sukla[1], A.K. Mangaraj[2] and L.N. Sahoo[3]
[1]Department of Statistics, S.V. College, University of Delhi, New Delhi 110021, INDIA
[2]Department of Statistics, Rajendra Junior College, Bolangir 767002, INDIA
[3]Department of Statistics, Utkal University, Bhubaneswar 751004, INDIA

Available online at: www.isca.in, www.isca.me
Received 29th August 2014, revised 8th September 2014, accepted 11th September 2014

Abstract

Our primary objective in the present paper is to search for an appropriate probability distribution for modeling of the distribution of daily rainfall amount in the Mahanadi Delta of Odisha. To achieve this objective, 17 types of probability distributions were fitted to the data on daily rainfall volume for a period of 28 years during rainy season. Kolmogorov-Sminov, Anderson-Darling and Chi-Squared goodness-of-fit tests were employed in order to test the strength of the fitted distributions and to identify the most appropriate one.

Keywords: Goodness-of-fit test, Mahanadi Delta, probability distribution, rainfall amount. **MSC 2010 Subject Classification:** 62P12.

Introduction

Stochastic modeling of daily precipitation amount utilizing a probability distribution is very essential for hydrological and meteorological studies. Because, such a modeling of daily rainfall volume is related to precipitation frequency analysis, the key steps of which involve selection of a suitable distribution for representing precipitation depth, to investigate various rainfall characteristics, for the trend detection, and to forecast rainfall amounts at different probability levels. Proper understanding of the shape of the underlying distribution of rainfall amount is therefore of vital importance in efficient planning and execution of water resource management program for agricultural and industrial developments, and environmental strategies, as well as in conducting researches in the fields of hydrology, meteorology, fisheries, health, ecology, environment and others. However, the choice of a suitable distribution is still one of the major problems since there is no general agreement as to which distribution, or distributions, that should be used for the frequency analysis of rainfall amount. Hence, it is necessary to evaluate many available distributions in order to find a suitable one that could provide accurate rainfall estimates. If a distribution provides a good fit to the data set, then the statistical properties of the rainfall amounts are approximated by those properties of the distribution.

Probability distributions connected with daily rainfall amount have been studied since quite sometimes by many researchers[1-8]. But, there is no guarantee that a specific distribution can be considered to have a good fit for all situations.

The Mahanadi Delta, as is situated on the eastern coast of India parallel to the Bay of Bengal, gets rainfall from the south-west monsoon with an average annual rainfall 1572 mm and the rainy day in a year ranging from 55 to 80 days. The most pre-dominant crop in this region is paddy covering about 95% of the total area under cultivation. As sufficient supplementary irrigation facilities are not available in the most parts, people mainly depend on autumn and winter paddy which are grown during monsoon season (June-September) and harvested during post-monsoon season (October and November). During monsoon season a large variety of vegetables are also grown here. The quantum of annual rainfall received by this river basin is fairly good. But, its irregular distribution and variation in time and space is a cause of great stress to the farming activities, crop production and crop yield as the agriculture is mostly rain fed. An appropriate modeling of the daily rainfall volume is therefore of crucial importance in planning agricultural activities and managing the associated water supply systems at various locations of the study domain.

Our investigation is aimed at searching for a probability distribution that is most suitable for modeling the daily rainfall amount of the study area, Mahanadi Delta as well as for 4 representative stations spread across the area during the rainy season. Based on our prior knowledge and experience gathered from the various past studies available in the literature, we include 17 probability distributions as candidate distributions for our investigation. Three goodness-of-fit (GOF) tests, namely, Kolmogorov-Sminov (KS), Anderson-Darling (AD) and Chi-Squared (CS) tests were used to evaluate fitness of the competing distributions under considerations.

Methodology

Source and Nature of Data: The present study utilizes data on daily rainfall amount of the four meteorological stations –

Bhubaneswar, Cuttack, Paradip and Puri of the Mahanadi Delta region for 28 years (1982-2009). The relevant data were collected from the Meteorological Centre, Bhubaneswar, Odisha. We are confined to the rainy season only *i.e*., the period from 1st June to 31st October, because during this season our study site receives more than 85% of its total annual rainfall and this period also coincides with the growth season of the paddy crop, the major cash crop in the tract. For proper representation of the whole study area, figures on the daily rainfall amount are obtained by taking average of such figures for the four recording stations.

Modeling of Rainfall Amount: Here our main concern is on the daily rainfall amount that can be assumed as a continuous random variable denoted by X. In order to model the distributional pattern of the said random variable in the four meteorological stations as well as in the whole study domain, the probability density functions $f(x)$ of the considered 17 distribution are given below. We consider one 1 – parameter, twelve 2 – parameter and four 3 – parameter distributions.

Beta Distribution

$f(x) = \frac{1}{B(\alpha_1,\alpha_2)} \frac{(x-a)^{\alpha_1-1}(b-x)^{\alpha_2-1}}{(b-a)^{\alpha_1+\alpha_2-1}}$; $\alpha_1, \alpha_2 > 0, a \le x \le b$;

α_1 and α_2 are the shape parameters, and a and b are the boundary parameters $(a < b)$.

Dagum Distribution

$f(x) = \frac{\alpha k \left(\frac{x}{\beta}\right)^{\alpha k-1}}{\beta\left[1+\left(\frac{x}{\beta}\right)^{\alpha}\right]^{k+1}}$; $k, \alpha, \beta > 0, 0 \le x < \infty$;

k and α are the shape parameters, and β is the scale parameter.

Exponential Distribution

$f(x) = \lambda \exp[-\lambda(x-\gamma)]$; $\lambda > 0, \gamma \le x < +\infty$;

λ is the inverse scale parameter, and γ is the location parameter.

Frechet Distribution

$f(x) = \frac{\alpha}{\beta}\left(\frac{\beta}{x}\right)^{\alpha+1} \exp\left[-\left(\frac{\beta}{x}\right)^{\alpha}\right]$; $\alpha, \beta > 0, 0 \le x < +\infty$;

α is the shape parameter, and β is the scale parameter.

Gamma Distribution

$f(x) = \frac{x^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} \exp\left[-\left(\frac{x}{\beta}\right)\right]$; $\alpha, \beta > 0, 0 \le x < +\infty$;

α is the shape parameter, and β is the scale parameter.

Generalized Extreme Value Distribution

$f(x) = \begin{cases} \frac{1}{\sigma}\exp\left[-(1+kx)^{-\frac{1}{k}}\right](1+kx)^{-1-\frac{1}{k}}, k \ne 0 \\ \frac{1}{\sigma}\exp[-x - \exp(-x)], k = 0 \end{cases}$; $\sigma > 0, 1 + k\frac{(x-\mu)}{\sigma} > 0$ for $k \ne 0$ and $-\infty < x < +\infty$ for $k = 0$; k is the shape parameter, σ is the scale parameter, and μ is the location parameter.

Generalized Pareto Distribution

$f(x) = \begin{cases} \frac{1}{\sigma}\left(1 + k\frac{(x-\mu)}{\sigma}\right)^{-1-\frac{1}{k}}, k \ne 0 \\ \frac{1}{\sigma}\exp\left(-\frac{(x-\mu)}{\sigma}\right), k = 0 \end{cases}$; $\sigma > 0, \mu \le x < +\infty$ for $k \ge 0$ and $\mu \le x \le \mu - \frac{\sigma}{k}$ for $k < 0$; k is the shape parameter, σ is the scale parameter, and μ is the location parameter.

Inverse Gaussian Distribution

$f(x) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left[-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right]$; $\lambda, \mu > 0, 0 \le x < +\infty$;

λ and μ are the parameters.

Laplace Distribution

$f(x) = \frac{\lambda}{2}\exp(-\lambda|x-\mu|)$; $\lambda > 0, -\infty < x < +\infty$;

λ is the inverse scale parameter, and μ is the location parameter.

Log-Logistic Distribution

$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\left[1+\left(\frac{x}{\beta}\right)^{\alpha}\right]^{-2}$; $\alpha, \beta > 0, 0 \le x < +\infty$;

α is the shape parameter, and β is the scale parameter.

Logistic Distribution

$f(x) = \frac{\exp(-x)}{\sigma[1+\exp(-x)]^2}$; $\sigma > 0, -\infty < x < +\infty$;

σ is the scale parameter, and μ is the location parameter.

Lognormal Distribution

$f(x) = \frac{1}{x\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2\right]$; $\sigma > 0, 0 \le x < +\infty$;

σ and μ are the parameters.

Normal Distribution

$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$; $\sigma > 0, -\infty < x < +\infty$;

σ is the scale parameter, and μ is the location parameter.

Pearson 5 Distribution

$f(x) = \frac{\exp\left(-\frac{\beta}{x}\right)}{\beta\,\Gamma(\alpha)\left(\frac{x}{\beta}\right)^{\alpha+1}}$; $\alpha, \beta > 0, 0 \le x < +\infty$;

α is the shape parameter, and β is the scale parameter.

Pearson 6 Distribution

$f(x) = \frac{\left(\frac{x}{\beta}\right)^{\alpha_1-1}}{\beta\, B(\alpha_1,\alpha_2)\left(1+\frac{x}{\beta}\right)^{\alpha_1+\alpha_2}}$; $\alpha_1, \alpha_2, \beta > 0, 0 \le x < +\infty$;

α_1 and α_2 are the shape parameters, and β is the scale parameter.

Power Function Distribution

$f(x) = \frac{\alpha(x-a)^{\alpha-1}}{(b-a)^{\alpha}}$; $\alpha > 0, a \le x \le b$;

α is the shape parameter, and a and b are the boundary parameters $(a < b)$.

Weibull Distribution

$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right]$; $\alpha, \beta > 0, 0 \leq x < +\infty$;

α is the shape parameter, and β is the scale parameter.

The Goodness-of-fit Tests: Denoting cumulative distribution function (CDF) of the random variable X by $F(\cdot)$, and $x_1, x_2, \ldots, x_n$ as n sample observations, the three GOF tests used at 5% level of significance, are described as follows:

The Kolmogorov-Smirnov Test: The KS test is based on the largest vertical difference between the theoretical CDF $F_0(x)$ and the observed (empirical) CDF $F_n(x)$ of the random sample of n observations. The hypothesis for the GOF under this test procedure is $H_0: F_0(x) = F_n(x)$ *vs.* $H_1: F_0(x) \neq F_n(x)$, and the KS test statistics is defined by

$d_{max} = \max_x |F_n(x) - F_0(x)|$.

The null hypothesis H_0 is rejected at 5% level of significance if the calculated value of d_{max} exceeds the tabulated value $D_{0.05} = 1.36/\sqrt{n}$.

The Anderson-Darling Test: The Anderson-Darling (AD) test compares the fit of an observed CDF to an expected (theoretical) CDF. This test gives more weight to the tail of the distribution than the KS test and the test statistic (A^2) is defined by

$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\ln F(x_i) + \ln(1 - F(x_{n-i+1}))]$.

The Chi-Squared Test: The Chi-Squared (CS) test simply compares how well the theoretical distribution fits the empirical distribution. The CS statistic is defined by

$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$,

where $k =$ number of classes, O_i and E_i are respectively the observed and expected frequencies for bin i, and $E_i = F(x_2) - F(x_1)$ such that x_1 and x_2 are the lower and upper limits for bin i.

Analysis of Data: For the purpose of data analysis that involves in fitting of the selected probability distributions to the data sets on daily rainfall amounts, in selecting the best model, and in applying the analysis results to make better decisions, we relied on the Software Easy Fit 5.5 Professional.

Some Descriptive Statistics: Before proceeding further, first we will have a brief a discussion on some important descriptive statistics such as range, mean, standard deviation (SD), coefficient of variation (CV), skewness and kurtosis of the daily rainfall amount of different locations. These statistics are computed from the raw data for the period under consideration and summarized in Table-1. Since the minimum amount of daily rainfall is 0 mm, the range of daily rainfall amount is equal to the maximum amount of daily rainfall for each location. The maximum amount of daily rainfall of Paradip is comparatively higher than that of other stations. Because Paradip, due to its location, is more influenced by the depressions and storms originated from the Bay of Bengal than any other part of the deltaic region. Bhubaneswar receives the highest mean rainfall amount *i.e.*, 9.3 mm. This occurs because according to the study conducted by Sukla[9], the expected number of wet days in a month for Bhubaneswar is the maximum with longer expected wet spell lengths. This means that during rainy season, Bhubaneswar experiences more rainfall events than the other three stations and these rainy events during this period are also longer with heavy and moderate rains occurring intermittently.

Puri and Cuttack possess respectively the maximum and minimum value of CV. As the irregularity of the daily rainfall is represented by the CV, our findings indicate that the daily rainfall at Cuttack is more consistent and at Puri is more irregular than other stations. In terms of the shape characteristics *i.e.*, skewness and kurtosis, the daily rainfall distributions of four stations as well as for the Mahanadi Delta region are strongly positively skewed and highly leptokurtic type.

As demonstrated by table-1, Cuttack shows the lowest variability, smallest values of skewness and kurtosis as well as the lowest value of maximum amount of daily rainfall. This means that the rainfall distribution at this station is more evenly distributed than other stations.

Table-1
Descriptive statistics for different locations

Statistic	Metrological Station				
	Bhubaneswar	Cuttack	Paradip	Puri	Mahanadi Delta
No. of Days	4284	4284	4284	4284	4284
Range	4.3E+2	3.3E+2	5.2E+2	3.6E+2	3.5E+2
Mean	9.3	9.0	9.1	8.4	9.0
SD	21.0	20.0	23.0	22.0	17.0
CV	2.3	2.2	2.6	2.7	1.9
Skewness	5.5	4.9	6.4	6.3	5.5
Kurtosis	58.0	40.0	77.0	61.0	57.0

Table-2
Estimated parameters of fitted distributions

Distribution	Bhubaneswar	Cuttack	Paradip	Puri	Mahanadi Delta
Beta	$\alpha_1 = 0.06$ $\alpha_2 = 8.7$ $a = 3.4E-15$ $b = 1.2E+3$	$\alpha_1 = 0.08$ $\alpha_2 = 1.4$ $a = 2.8E-15$ $b = 1.7E+3$	$\alpha_1 = 0.06$ $\alpha_2 = 2.0$ $a = -6.9E-16$ $b = 1.1E+3$	$\alpha_1 = 0.03$ $\alpha_2 = 2.1$ $a = -9.8E-16$ $b = 8.2E+2$	$\alpha_1 = 0.03$ $\alpha_2 = 0.14$ $a = 2.4E-15$ $b = 1.1E+3$
Dagum	$k = 0.41$ $\alpha = 1.6$ $\beta = 19.0$	$k = 0.5$ $\alpha = 1.6$ $\beta = 17.0$	$k = 0.6$ $\alpha = 1.2$ $\beta = 12.0$	$k = 0.48$ $\alpha = 1.4$ $\beta = 16.0$	$k = 0.4$ $\alpha = 1.7$ $\beta = 14.0$
Exponential	$\lambda = 0.11$ $\gamma = -1.0E-14$	$\lambda = 0.11$ $\gamma = -1.0E-14$	$\lambda = 0.11$ $\gamma = -1.0E-14$	$\lambda = 0.12$ $\gamma = -1.0E-14$	$\lambda = 0.11$ $\gamma = -1.0E-14$
Frechet	$\alpha = 0.58$ $\beta = 2.3$	$\alpha = 0.68$ $\beta = 3.5$	$\alpha = 0.58$ $\beta = 2.1$	$\alpha = 0.57$ $\beta = 2.2$	$\alpha = 0.59$ $\beta = 1.9$
Gamma	$\alpha = 0.2$ $\beta = 47.0$	$\alpha = 0.21$ $\beta = 43.0$	$\alpha = 0.15$ $\beta = 60.0$	$\alpha = 0.14$ $\beta = 60.0$	$\alpha = 0.27$ $\beta = 33.0$
Generalized Extreme Value	$k = 0.64$ $\sigma = 3.5$ $\mu = 1.4$	$k = 0.63$ $\sigma = 3.5$ $\mu = 1.4$	$k = 0.69$ $\sigma = 2.9$ $\mu = 1.1$	$k = 0.71$ $\sigma = 2.6$ $\mu = 0.89$	$k = 0.53$ $\sigma = 4.1$ $\mu = 2.1$
Generalized Pareto	$k = 0.58$ $\sigma = 4.5$ $\mu = -1.3$	$k = 0.56$ $\sigma = 4.5$ $\mu = -1.3$	$k = 0.65$ $\sigma = 3.6$ $\mu = -1.1$	$k = 0.67$ $\sigma = 3.1$ $\mu = -0.99$	$k = 0.44$ $\sigma = 5.7$ $\mu = -1.1$
Inverse Gaussian	$\lambda = 1.4$ $\mu = 16.0$	$\lambda = 2.9$ $\mu = 17.0$	$\lambda = 1.3$ $\mu = 17.0$	$\lambda = 1.3$ $\mu = 17.0$	$\lambda = 1.2$ $\mu = 12.0$
Laplace	$\lambda = 0.07$ $\mu = 9.3$	$\lambda = 0.07$ $\mu = 9.0$	$\lambda = 0.06$ $\mu = 9.1$	$\lambda = 0.06$ $\mu = 8.4$	$\lambda = 0.08$ $\mu = 9.0$
Log-Logistic	$\alpha = 1.0$ $\beta = 6.1$	$\alpha = 1.2$ $\beta = 7.8$	$\alpha = 0.98$ $\beta = 5.5$	$\alpha = 0.98$ $\beta = 5.7$	$\alpha = 1.1$ $\beta = 5.0$
Logistic	$\sigma = 12.0$ $\mu = 9.3$	$\sigma = 11.0$ $\mu = 9.0$	$\sigma = 13.0$ $\mu = 9.1$	$\sigma = 12.0$ $\mu = 8.4$	$\sigma = 9.5$ $\mu = 9.0$
Lognormal	$\sigma = 1.7$ $\mu = 1.7$	$\sigma = 1.4$ $\mu = 2.0$	$\sigma = 1.7$ $\mu = 1.6$	$\sigma = 1.7$ $\mu = 1.7$	$\sigma = 1.6$ $\mu = 1.5$
Normal	$\sigma = 21.0$ $\mu = 9.3$	$\sigma = 20.0$ $\mu = 9.0$	$\sigma = 23.0$ $\mu = 9.1$	$\sigma = 22.0$ $\mu = 8.4$	$\sigma = 17.0$ $\mu = 9.0$
Pearson 5	$\alpha = 0.45$ $\beta = 0.58$	$\alpha = 0.58$ $\beta = 1.4$	$\alpha = 0.45$ $\beta = 0.55$	$\alpha = 0.44$ $\beta = 0.55$	$\alpha = 0.46$ $\beta = 0.51$
Pearson 6	$\alpha_1 = 0.69$ $\alpha_2 = 3.1$ $\beta = 49.0$	$\alpha_1 = 0.9$ $\alpha_2 = 3.1$ $\beta = 40.0$	$\alpha_1 = 0.7$ $\alpha_2 = 2.0$ $\beta = 27.0$	$\alpha_1 = 0.69$ $\alpha_2 = 2.3$ $\beta = 34.0$	$\alpha_1 = 0.76$ $\alpha_2 = 3.2$ $\beta = 35.0$
Power Function	$\alpha = 0.04$ $a = 2.0E-15$ $b = 6.0E+2$	$\alpha = 0.05$ $a = 2.0E-15$ $b = 4.6E+2$	$\alpha = 0.03$ $a = 2.0E-15$ $b = 7.3E+2$	$\alpha = 0.04$ $a = 2.0E-15$ $b = 5.1E+2$	$\alpha = 0.05$ $a = 2.7E-15$ $b = 5.4E+2$
Weibull	$\alpha = 0.69$ $\beta = 12.0$	$\alpha = 0.78$ $\beta = 15.0$	$\alpha = 0.65$ $\beta = 12.0$	$\alpha = 0.66$ $\beta = 12.0$	$\alpha = 0.73$ $\beta = 9.6$

Results and Model Selection

Parameters of the distributions are estimated by the method of maximum likelihood and their values are compiled in table-2. Taking into account the estimated parametric values, the aforesaid seventeen distributions were fitted to the data.

The three GOF tests *viz.*, KS, AD and CS tests mentioned earlier were conducted to test the discrepancies between observed and theoretical frequency distributions. The test statistic in each case were computed and tested at 5% level of significance. Accordingly, the ranking of different probability distributions were marked from 1 to 17 based on the minimum test statistic value. No rank was given to a distribution when the concerned test failed to fit the distribution. Results on the GOF tests for all locations are summarized in tables 3 – 7.

Table-3
Goodness-of-fit summary for Bhubaneswar

Distribution	KS Test		AD Test		CS Test	
	Statistic	Rank	Statistic	Rank	Statistic	Rank
Beta	0.25	3	6.3E+3	15	NA	-
Dagum	0.42	11	2.6E+3	13	9.1E+3	9
Exponential	0.44	15	2.6E+4	17	1.2E+4	15
Frechet	0.44	14	2.5E+3	9	1.1E+4	14
Gamma	0.42	10	1.8E+2	1	7.9E+3	6
Gen. Extreme Value	0.23	1	3.4E+2	3	1.4E+3	4
Gen. Pareto	0.23	2	3.0E+2	2	1.1E+3	1
Inv. Gaussian	0.45	17	1.9E+3	7	1.0E+4	11
Laplace	0.30	4	6.9E+2	5	1.8E+3	5
Log-Logistic	0.42	12	2.7E+3	14	1.1E+4	13
Logistic	0.31	6	6.7E+2	4	1.4E+3	3
Lognormal	0.43	13	2.5E+3	11	9.8E+3	10
Normal	0.33	7	7.1E+2	6	1.3E+3	2
Pearson 5	0.44	16	2.2E+3	8	1.0E+4	12
Pearson 6	0.42	9	2.6E+3	12	9.1E+3	8
Power Function	0.31	5	6.8E+3	16	NA	-
Weibull	0.42	8	2.5E+3	10	9.1E+3	7

Table-4
Goodness-of-fit summary for Cuttack

Distribution	KS Test		AD Test		CS Test	
	Statistic	Rank	Statistic	Rank	Statistic	Rank
Beta	0.18	1	5.8E+3	15	NA	-
Dagum	0.42	13	2.6E+3	13	9.1E+3	9
Exponential	0.44	8	2.6E+4	17	1.2E+4	10
Frechet	0.44	15	2.5E+3	10	1.1E+4	14
Gamma	0.42	12	1.8E+2	1	7.9E+3	6
Gen. Extreme Value	0.23	4	3.4E+2	3	1.4E+3	4
Gen. Pareto	0.23	3	3.0E+2	2	1.1E+3	2
Inv. Gaussian	0.45	17	1.9E+3	7	1.0E+4	12
Laplace	0.3	5	6.9E+2	5	1.8E+3	5
Log-Logistic	0.42	11	2.7E+3	14	1.1E+4	13
Logistic	0.31	6	6.7E+2	4	1.4E+3	3
Lognormal	0.43	14	2.5E+3	11	9.8E+3	11
Normal	0.33	7	7.1E+2	6	1.3E+3	1
Pearson 5	0.44	16	2.2E+3	8	1.0E+4	15
Pearson 6	0.42	10	2.6E+3	12	9.1E+3	8
Power Function	0.31	2	6.8E+3	16	NA	-
Weibull	0.42	9	2.5E+3	9	9.1E+3	7

Table-5
Goodness-of-fit summary for Paradip

Distribution	KS Test		AD Test		CS Test	
	Statistic	Rank	Statistic	Rank	Statistic	Rank
Beta	0.37	7	6.7E+2	4	4.5E+3	6
Dagum	0.46	12	2.9E+3	13	1.1E+4	10
Exponential	0.48	16	3.1E+4	17	1.4E+4	16
Frechet	0.47	14	2.8E+3	10	1.2E+4	15
Gamma	0.46	11	-1.1E+2	1	9.7E+3	7
Gen. Extreme Value	0.24	2	4.0E+2	3	1.7E+3	4
Gen. Pareto	0.24	1	3.5E+2	2	1.2E+3	1
Inv. Gaussian	0.48	17	2.2E+3	8	1.2E+4	13
Laplace	0.33	4	7.8E+2	6	2.1E+3	5
Log-Logistic	0.46	10	2.9E+3	15	1.2E+4	12
Logistic	0.33	5	7.7E+2	5	1.6E+3	3
Lognormal	0.46	13	2.8E+3	12	1.2E+4	11
Normal	0.35	6	8.2E+2	7	1.5E+3	2
Pearson 5	0.48	15	2.5E+3	9	1.2E+4	14
Pearson 6	0.46	9	2.9E+3	14	1.1E+4	9
Power Function	0.31	3	7.4E+3	16	NA	-
Weibull	0.46	8	2.8E+3	11	1.0E+4	8

Table-6
Goodness-of-fit summary for Puri

Distribution	KS Test		AD Test		CS Test	
	Statistic	Rank	Statistic	Rank	Statistic	Rank
Beta	0.27	2	6.0E+2	4	3.6E+3	6
Dagum	0.51	12	3.3E+3	13	1.3E+4	10
Exponential	0.51	13	3.7E+4	16	1.6E+4	16
Frechet	0.51	14	3.2E+3	11	1.5E+4	15
Gamma	0.51	11	-4.2E+2	1	1.2E+4	7
Gen. Extreme Value	0.28	3	4.6E+2	3	2.0E+3	4
Gen. Pareto	0.26	1	4.1E+2	2	1.6E+3	1
Inv. Gaussian	0.52	16	2.6E+3	8	1.4E+4	13
Laplace	0.33	4	8.1E+2	6	2.1E+3	5
Log-Logistic	0.51	10	3.4E+3	15	1.4E+4	12
Logistic	0.34	5	8.0E+2	5	1.7E+3	3
Lognormal	0.51	9	3.3E+3	12	1.4E+4	11
Normal	0.35	6	8.5E+2	7	1.6E+3	2
Pearson 5	0.52	15	2.9E+3	9	1.5E+4	14
Pearson 6	0.51	8	3.3E+3	14	1.3E+4	9
Power Function	NA	-	NA	-	NA	-
Weibull	0.51	7	3.2E+3	10	1.3E+4	8

Table-7
Goodness-of-fit summary for Mahanadi Delta

Distribution	KS Test		AD Test		CS Test	
	Statistic	Rank	Statistic	Rank	Statistic	Rank
Beta	0.4	17	4.5E+3	15	NA	-
Dagum	0.25	7	1.6E+3	8	3.4E+3	9
Exponential	0.29	14	9.2E+3	17	5.2E+3	15
Frechet	0.27	10	1.7E+3	14	4.2E+3	13
Gamma	0.25	6	8.6E+2	6	2.3E+3	6
Gen. Extreme Value	0.16	1	1.5E+2	2	5.9E+2	2
Gen. Pareto	0.18	2	1.2E+2	1	3.5E+2	1
Inv. Gaussian	0.28	13	1.6E+3	10	4.2E+3	12
Laplace	0.25	3	5.1E+2	3	1.2E+3	5
Log-Logistic	0.27	8	1.7E+3	13	4.0E+3	11
Logistic	0.28	12	5.1E+2	4	9.9E+2	3
Lognormal	0.27	9	1.6E+3	11	3.9E+3	10
Normal	0.3	15	5.7E+2	5	1.0E+3	4
Pearson 5	0.28	11	1.7E+3	12	4.7E+3	14
Pearson 6	0.25	4	1.6E+3	9	3.4E+3	8
Power Function	0.39	16	4.7E+3	16	NA	-
Weibull	0.25	5	1.5E+3	7	2.9E+3	7

Referring to the results provided in tables 3 – 7, we see that it is difficult to identify the best statistical distribution for a particular location as three different GOF criteria have been used. The selected distribution for the same data set based on one test is different for other test. For example, in case of Bhubaneswar, General Extreme Value distribution is ranked in the first position under KS test but the distribution is ranked in third and fourth places under the AD and CS tests respectively. For this situation, as pointed out by Suhaila and Jemain[2,3], we may choose the best fitting distribution based on the majority of the tests as we have no scope for investigating which is the most powerful test. With this objective, we have selected five distributions holding the first five ranks based on all three tests independently for all locations. Hence, 8 distributions – Beta, Gamma, General Extreme Value, General Pareto, Laplace, Logistic, Normal and Power Function for Bhubaneswar, Cuttack and Paradip; 7 distributions – Beta, Gamma, General Extreme Value, General Pareto, Laplace, Logistic and Normal for Puri; and 7 distributions – General Extreme Value, General Pareto, Laplace, Logistic, Normal, Pearson 6 and Weibull for the Mahanadi Delta have got selected as the candidate distributions for competition.

Identification of the Best Probability Model: Considering the criterion of best fit in respect of the majority of tests, we may identify General Pareto as the first candidate distribution for modeling our daily rainfall volume data for Bhubaneswar, Paradip, Puri and the study domain as a whole, and the General Extreme Value distribution as the second candidate for Bhubaneswar and the study domain. However, on this ground, selections of the best and the second best fitted distributions in respect of Cuttack, and second best in respect of Paradip and Puri are not straight forward.

In order to have better determination of the best fit models at each location, the assessments of all probability models were made on the basis of the total test score obtained by combining the three GOF tests[1-5]. Maximum score 17 was awarded to the probability distribution securing rank 1 based on the test statistic and further less score was awarded to the distribution having rank more than 1, that means 2 was awarded to the distribution securing rank 16. A distribution was awarded a zero score for a test if the distribution has not a significant fit to the data set. The total score of the entire three tests are summarized in table-8 with a view to identify the best fit distribution for different locations on the basis of the highest scores obtained. The scores in respect of the best fit and second best fit and third best fit distributions are boldly printed whereas the scores in respect of the distributions coming out in fourth and fifth positions are underlined.

Based on the results available in table-8, General Pareto and General Extreme Value distributions were found to be the best fit and second best fit models respectively among the seventeen probability distributions tested for the all metrological stations and for the entire study domain. The Logistic distribution comes third for Bhubaneswar, Cuttack and Paradip while Beta and Laplace distributions come third for Puri and Mahanadi Dela respectively.

Table-8
Summary of the statistical score results at each location

Distribution	Bhubaneswar	Cuttack	Paradip	Puri	Mahanadi Delta
Beta	18	20	<u>37</u>	**42**	4
Dagum	21	19	19	19	30
Exponential	7	19	5	9	8
Frechet	17	15	15	14	17
Gamma	37	35	35	35	<u>36</u>
Gen. Extreme Value	**46**	**43**	**45**	**44**	**49**
Gen. Pareto	**49**	**47**	**50**	**50**	**50**
Inv. Gaussian	19	18	16	17	19
Laplace	<u>40</u>	<u>39</u>	<u>39</u>	<u>39</u>	**43**
Log-Logistic	15	16	17	17	22
Logistic	**41**	**41**	**41**	<u>41</u>	<u>35</u>
Lognormal	20	18	18	22	24
Normal	<u>39</u>	<u>40</u>	<u>39</u>	<u>39</u>	30
Pearson 5	18	15	16	16	17
Pearson 6	25	24	22	23	33
Power Function	15	18	17	0	4
Weibull	29	29	27	29	<u>35</u>

Conclusion

Our data analysis and scientific assessment procedure in this paper has been successfully employed and General Pareto and General Extreme Value were respectively emerged out as the best and second best fitted distributions for all locations. However, our study leads to an overall conclusion that the General Pareto distribution is one of the best probability models for describing the distribution of daily rainfall volume of the study site. Appropriate planning and hydrological design of soil conservation and drainage structures at the Mahanadi Delta region can therefore be effectively carried out on the basis of predicted amount of daily rainfall using General Pareto distribution.

Another interesting result of this study is that the daily rainfall distribution for each rainfall recording station is more or less similar and is not much influenced as is expected due to their topographical, geographical and climatic changes.

References

1. Akhter S.S. and Hossain M.F., A study on rainfall occurrences using probability approach, *Journal of Statistical Studies*, **18**, 37-50 **(1998)**
2. Suhaila J. and Jemain A.A., Fitting daily rainfall amount in peninsular Malaysia using several types of exponential distributions, *Journal of Applied Sciences Research*, **3**, 1027-1036 **(2007)**
3. Suhaila, J. and Jemain A.A., Fitting the statistical distributions to the daily rainfall amount in peninsular Malaysia, *Jurnal Teknologi*, **46**, 33-48 **(2007)**
4. Suhaila J. and Jemain A.A., Fitting the statistical distributions for daily rainfall in peninsular Malaysia based on AIC criterion, *Journal of Applied Science Research*, **4**, 1846-1857 **(2008)**
5. Olofintoye O.O., Sule B.F. and Salami A.W., Best-fit probability distribution model for peak daily rainfall of selected cities in Nigeria, *New York Science Journal*, **2**, 1-12 **(2009)**
6. Mangaraj A.K. and Sahoo L.N., A study on the probability distribution of daily rainfall amounts in Western Orissa, *International Journal of Agricultural and Statistical Sciences*, **6**, 53-60 **(2010)**
7. Dan’azumi S., Shamsudin S. and Rahman A.A., Probability distribution of rainfall depth at hourly time-scale, *World Academy of Science, Engineering and Technology*, **72**, 760-764 **(2010)**
8. Dan’azumi S., Shamsudin, S. and Aris A., Modeling the distribution of rainfall intensity using hourly data, *American Journal of Environmental Sciences*, **6**, 238-243 **(2010)**
9. Sukla M.K., Stochastic Modeling of Rainfall Occurrence for the Mahanadi Delta Region, Unpublished Ph. D. Thesis submitted to the Utkal University, Bhubaneswar, India, **(2012)**
10. Murray R.S. and Larry J.S., Theory and Problems of Statistics (3rd Edition), Tata Mc Graw-Hill Publishing Company Ltd., New Delhi, **(2000)**
11. Sharma M.A. and Singh J.B., Use of probability distribution in rainfall analysis, *New York Science Journal*, **3**, 40-49 **(2010)**