Use of Probability Distribution in Rainfall Analysis

Mohita Anand Sharma and Jai Bhagwan Singh

School of Basic and Applied Science, Shobhit University, Meerut, Uttar Pradesh, 250110, India
Email: mohita_anand@rediffmail.com

Abstract: The daily rainfall data of 37 years were collected from the IMD approved Meteorological Observatory situated at GB Pant University of Agriculture and Technology, Pantnagar, India. The data were then processed to identify the maximum rainfall received on any one day (24hrs duration), in any week (7 days), in a month (4 weeks), in a monsoon season (4 months) and in a year (365 days period). The data were also analyzed to find out the standard deviation and coefficient of variation during all the four periods of study. The data showed that the annual daily maximum rainfall received at any time ranged between 49.32mm (minimum) to 229.40mm (maximum) indicating a very large range of fluctuation during the period of study. The rainfall data were analysed to identify the best fit probability distribution for each period of study and the trend has been presented in this study. Three statistical goodness of fit test were carried out in order to select the best fit probability distribution on the basis of highest rank with minimum value of test statistic. Fourth probability distribution was identified using maximum overall score based on sum of individual point score obtained from three selected goodness of fit test. Random numbers were generated for actual and estimated maximum daily rainfall for each period of study using the parameters of selected distributions. The best fit probability distribution was identified based on the minimum deviation between actual and estimated values. The lognormal and gamma distribution were found as the best fit probability distribution for the annual and monsoon season period of study, respectively. Generalized extreme value distribution was observed in most of the weekly period as best fit probability distribution. The best fit probability distribution of monthly data was found to be different for each month. The scientific results clearly established that the analytical procedure devised and tested in this study may be suitably applied for the identification of the best fit probability distribution of weather parameters. [New York Science Journal 2010;3(9):40-49]. (ISSN: 1554-0200).

Keywords: Peak rainfall, probability distributions, goodness-of-fit tests.

Introduction

Analysis of rainfall data strongly depends on its distribution pattern. It has long been a topic of interest in the fields of meteorology in establishing a probability distribution that provides a good fit to daily rainfall. Several studies have been conducted in India and abroad on rainfall analysis and best fit probability distribution function such as normal, log-normal, gumbel, weibull and Pearson type distribution were identified.

Fisher (1924) studied the influence of rainfall on the yield of wheat in Rothamasted. He showed that it is the distribution of rainfall during a season rather than its total amount which influence the crop yield. Tippet (1929) subsequently applied the technique on sunshine distribution and found that sunshine has beneficial effect through out the year on wheat crop. Another useful line of work relating to the study of rainfall distribution was introduced by Manning (1950). He transformed the skew frequency distribution of rainfall to approximate closely to the theoretical normal distribution.

Moaley et al., (1970) studied statistical distribution of rainfall during south west and north east monsoon season at representative stations in India and Gamma distribution has been fitted to rainfall data. Bhargava et al., (1971) also showed that for a number of crops the distribution of rainfall over the season has a great influence on the yield. RamanRao et al., (1975) analyzed the daily rainfall data collected at Bijapur for the year from 1921 to 1970.

Kulandaivelu (1984) analysed the daily precipitation data of Coimbatore for a period of 70 years for weekly totals by fitting incomplete Gamma distribution model. The data indicate the likely commencement of rains, period of drought length of growing season and end of growing season. Based on the assured rainfall at (50%) probability level, suitable cropping system was suggested for Coimbatore. Phien and Ajirajah (1984) showed that for the annual flood, annual maximum rainfall, annual stream flow and annual rainfall, the log-Pearson type III distribution was highly suitable after evaluating by Chi-square and Kolmogorov- Smirnov tests.

Biswas and Khambete (1989) computed the lowest amount of rainfall at different probability level by fitting gamma distribution probability model to week by week total rainfall of 82 stations in dry farming tract of Maharashtra. Lin et al., (1993) stated that in accordance with the probability distribution all stations in same area can be classified in different clusters and special characteristic among a clusters can have spatial relationship to a certain extent in that cluster. Chapman (1994) evaluated five daily rainfall generating models with several methods and analysed that Srikanthan-McMahon model performed well when calibrated with long rainfall records. Duan et al., (1995) suggested that for modeling daily rainfall amounts, the weibull and to a lesser extent the exponential distribution is suitable. Upadhaya and Singh (1998) stated that it is possible to predict rainfall fairly accurate using various probability distributions for certain returns periods although the rainfall varies with space, time and have erratic nature. Sen and Eljadid (1999) reported that for monthly rainfall in arid regions, gamma probability distribution is best fit.

Ogunlela (2001) evaluated that log-person type III distribution best describe the stochastic analysis of peak daily rainfall. Tao et al., (2002) recommended generalized extreme value model as the most suitable distribution after a systematic assessment procedure for representing extreme-value process and its relatively simple parameter estimation. Topaloglu (2002) reported that gumbel probability model estimated by the method of moments and evaluated by chi-square tests was found to be the best model in the Seyhan river basin. Salami (2004) studied the meteorological data for Texas and found that Gumbel distribution fits adequately for both evaporation and temperature data, while for precipitation data log-Pearson type III distribution conforms more accurate. Lee (2005) indicated that log-Pearson type III distribution fits for 50% of total station number for the rainfall distribution characteristics of Chia-Nan plain area.

Baskar et al., (2006) observed the frequency analysis of consecutive days peaked rainfall at Banswara, Rajasthan, India, and found gamma distribution as the best fit as compared by other distribution and tested by Chi-square value. Deidda and Puliga (2006) found for left-censored records of Sardinia that some weak are evident for the generalized Pareto distribution. Kwaku et al., (2007) revealed that the log-normal distribution was the best fit probability distribution for one to five consecutive days' maximum rainfall for Accra, Ghana. Hanson et al., (2008) analysis indicated that Pearson type III distribution fits the full record of daily precipitation data and Kappa distribution best describes the observed distribution of wet-day daily rainfall. Olofintoye et al., (2009) examined that 50% of the total station number in Nigeria follows log-Pearson type III distribution for peak daily rainfall, while 40% and 10% of the total station follows Pearson type III and log-Gumbel distribution respectively.

On the basis of above it can be said that generally Log-Pearson/ Pearson type III distribution is fitted for the data and tested by Chi-square test. The present study is planned for establishing the methodology for identifying the pattern of probability distribution of weather parameter using least square method and the best fit probability distribution was evaluated on the basis of three goodness of fit test. The maximum rainfall data of a single site was used to select a best fit probability distribution for the value of weather parameters.

Material and Methods

The present study is based on time series data related to maximum daily rainfall annually, seasonally, monthly and weekly. The daily rainfall data of 37 years were collected from the IMD approved Meteorological Observatory situated at GB Pant University of Agriculture and Technology, Pantnagar, India. It is located at 29 N latitude, 79.3 E longitude and altitude 243.84m. Obove mean sea level and lies in Tarai belt of Uttaranchal. The dry season from October to May and wet season from June to September are found in this area. The soils of this region have good moisture holding capacities and have about 7.3 pH. The annual rainfall of this region is about 1400 mm, which is subjected to large variation. The data were then processed to identify the maximum rainfall received on any one day (24hrs duration), in any week (7 days), in a month (4 weeks), in a monsoon season (4 months) and in a year (365 days period). The annual maximum daily rainfall is ranging from 49.32mm to 229.40mm during the study period as presented in Figure 1.

On an average the region has a humid subtropical climate having hot summers (40-42 ^{0}C) and cold winters (2-4 ^{0}C) with monsoon rains occurring from June to September. More than 80% of the rain is received from south-west monsoon during four month period from June to September, and the rainfall of rainy season is significantly different from that of dry season. The best fit probability distribution was evaluated by using the following systematic steps.

Step I: Fitting the probability distribution

The probability distributions viz. normal, lognormal, gamma, weibull, pearson, generalized extreme value were identified to evaluate the best fit probability distribution for rainfall. In addition the different forms of these distributions were also tried and thus total 16 probability distributions viz. normal, lognormal (2P, 3P), gamma (2P, 3P), generalized gamma (3P, 4P), log-gamma, weibull (2P, 3P), pearson 5 (2P, 3P), pearson 6 (3P, 4P), log-pearson 3, generalized extreme value were applied to find out the best fit probability distribution The description of various probability distribution functions viz. density function, range and the parameter involved are presented in table 1.

Step II: Testing the goodness of fit

The goodness of fit test measures the compatibility of random sample with the theoretical probability distribution. The goodness of fit tests is applied for testing the following null hypothesis:

H_0 : the maximum daily rainfall data follow the specified distribution

H_A : the maximum daily rainfall data does not follow the specified distribution.

The following goodness-of-fit tests viz. Kolmogorov-Smirnov test and Anderson-Darling test were used along with the chi-square test at α (0.01) level of significance for the selection of the best fit Probability distribution.

(i) Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov statistic (D) is defined as the largest vertical difference between the theoretical and the empirical cumulative distribution function (ECDF):

$$D = \max_{1\le i\le n}\left(F(x_i) - \frac{i-1}{n}, \frac{i}{n} - F(x_i) \right) \quad (1)$$

Where,

X_i = random sample, i =1, 2,….., n.

$$\text{CDF}= F_n(x) = \frac{1}{n}.\left[\text{Number of observations} \le x\right] \quad (2)$$

This test is used to decide if a sample comes from a hypothesized continuous distribution.

(ii) Anderson-Darling Test

The Anderson-Darling statistic (A^2) is defined as

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1).[\text{In } F(X_i) + \text{In}(1-F(X_{n-i+1}))] \quad (3)$$

It is a test to compare the fit of an observed cumulative distribution function to an expected cumulative distribution function. This test gives more weight to the tails then the Kolmogorov-Smirnov test.

(iii) Chi-Squared Test

The Chi-Squared statistic is defined as

$$\chi^2 = \sum_{i=1}^{k}\frac{\left(O_i - E_i\right)^2}{E_i} \quad (4)$$

where

O_i = observed frequency

E_i = expected frequency

'i'= number of observations (1, 2, …….k)

calculated by

$$E_i = F(x_2) - F(x_1) \quad (5)$$

F = the CDF of the probability distribution being tested

The observed number of observation (k) in interval 'i' is computed from equation given below

$$k = 1 + \log_2 n \quad (6)$$

n = sample size

This test is for continuous sample data only and is used to determine if a sample comes from a population with a specific distribution.

Step III: Identification of best fit probability distribution

The three goodness of fit test mentioned above were fitted to the maximum rainfall data treating different data set. The test statistic of each test were computed and tested at (α=0.01) level of significance. Accordingly the ranking of different probability distributions were marked from 1 to 16 based on minimum test statistic value. The distribution holding the first rank was selected for all the three tests independently. The assessments of all the probability distribution were made on the bases of total test score obtained by combining the entire three tests. Maximum score 16 was awarded to rank first probability distribution based on the test statistic and further less score were awarded to the distribution having rank more than 1, that is 2 to 16. Thus the total score of the entire three tests were summarized to identify the best fit distribution on the bases of highest score obtained.

The probability distribution having the maximum score was included as a fourth probability distribution in addition to three probability distributions which were previously identified. Thus on the bases of the four identified probability distribution the procedure for obtaining the best fitted probability distribution is explained below:

(i) Generating random numbers

The four probability distributions identified for each data set were used to select the best probability distribution. The parameters of these four probability distributions were used to generate the random numbers.

(ii) Least square method

The least square method was used to identify the best fit probability. The random numbers were generated for the distributions and residuals (R) were computed for each observation of the data set.

$$R = \left| \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right) \right| \quad (7)$$

Where, Y_i = is the actual observation

$\hat{Y}_i$ = is the estimated observation (i=1, 2,…..,n)

The distribution having minimum sum of residuals was considered to be the best fit probability distribution for that particular data set.

Finally the best fit probability distributions for maximum rainfall on different sets of data were obtained and the best fit distribution for each set of data was identified. The above methodology can also be used for studying the probability distribution pattern for other weather variables.

Result and Discussion

The methodology presented above was applied to the 37 years weather data in which maximum rainfall in mm. were taken from Meteorological observatory, Pantnagar. Accordingly, the data was classified into 23 data sets as mentioned in table 2. These 23 data sets were classified as 1 annual, 1 seasonal, 4 months and 17 weekly to study the distribution pattern at different levels.

The summary of statistics mean, standard deviation, skewness coefficient, coefficient of variation, maximum and minimum values of daily maximum rainfall is presented in table 2. Where, the mean of maximum daily rainfall of all years annually is 123.35mm, seasonally is 123.31mm and monthly it is ranging from 50.64mm to 101.84mm and weekly is varying from 13.74mm to 61.97mm. The maximum daily rainfall in a year/ monsoon season is 229.40mm and monthly maximum daily rainfall in monsoon season is ranging from 110.50mm to 229.40mm and weekly maximum daily rainfall is in between 68.60mm to 229.40mm.

It was also observed that the minimum among the maximum daily rainfall was 0.00mm in the month of June and also in most of the weeks except in the second and third weeks of the month of July and first and fourth weeks of the month of August. The maximum value of coefficient of variation was observed in the last week which indicates a large fluctuation in the rainfall data set.

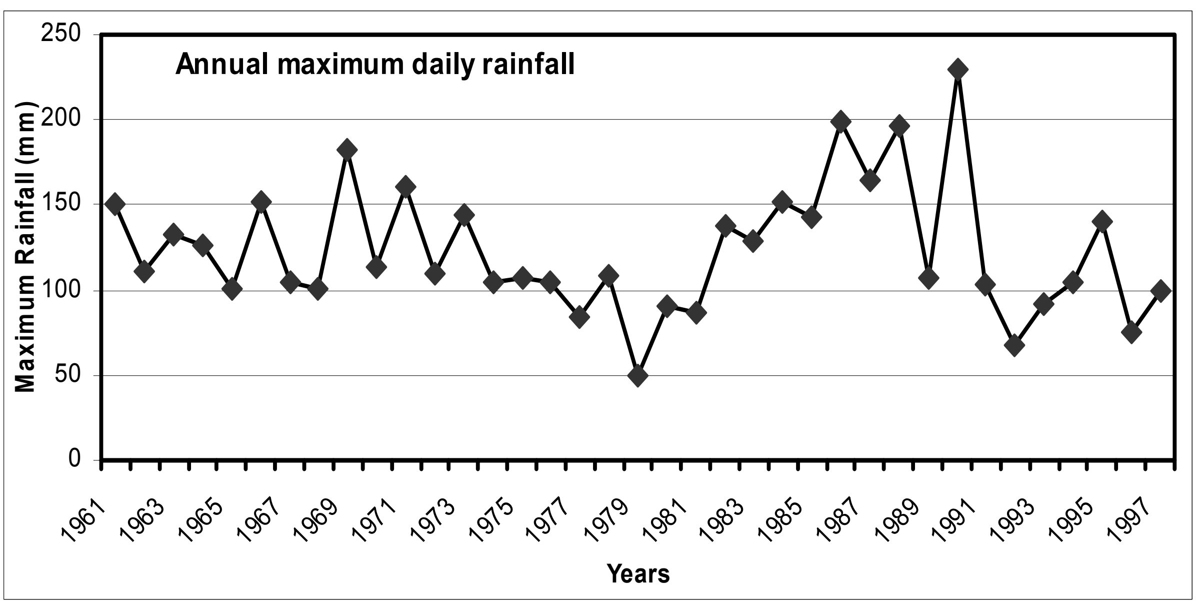

Figure-1. Year wise annual maximum daily rainfall (in mm).

Table 1. Description of various probability distribution functions.

Distribution	Probability density function	Range	Parameters
Gamma (3P)	$f(x)=\frac{(x-\gamma)^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)}\exp\left(-(x-\gamma)/\beta\right)$	$\gamma \le x < +\infty$	$\sigma = shape\ parameter$ $(\sigma>0)$ $\beta = scale\ parameter$ $(\beta>0)$ $\gamma = location\ parameter$ ($\gamma \equiv 0$ yields the two parameter gamma distribution) $\Gamma = Gamma\ function$
Gamma (2P)	$f(x)=\frac{(x)^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)}\exp\left(-(x)/\beta\right)$		
Generalized Extreme Value	$f(x)=\begin{cases}\frac{1}{\sigma}\exp\left(-(1+kz)^{-1/k}\right)(1+kz)^{-1-1/k} & k\neq 0\\ \frac{1}{\sigma}\exp\left(-z-\exp(-z)\right) & k=0\end{cases}$	$1+k\frac{(x-\mu)}{\sigma}>0$ for $k\neq 0$ $-\infty<x<+\infty$ for $k=0$	$\sigma = scale\ parameter$ $(\sigma>0)$ $k = shape\ parameter$ $\mu = location\ parameter$ where $z \equiv \frac{x-\mu}{\sigma}$
Generalized Gamma (4P)	$f(x)=\frac{k(x-\gamma)^{k\alpha-1}}{\beta^{k\alpha}\Gamma(\alpha)}\exp\left(-\left((x-\gamma)/\beta\right)^{k}\right)$	$\gamma \le x < +\infty$	$k = shape\ parameter$ $(k>0)$ $\alpha = shape\ parameter$ $(\alpha>0)$ $\beta = scale\ parameter$ $(\beta>0)$ $\gamma = location\ parameter$ ($\gamma \equiv 0$ yields the three parameter Generalized gamma distribution)
Generalized Gamma (3P)	$f(x)=\frac{kx^{k\alpha-1}}{\beta^{k\alpha}\Gamma(\alpha)}\exp\left(-\left(x/\beta\right)^{k}\right)$		
Log-Gamma	$f(x)=\frac{(In(x))^{\alpha-1}}{x\beta^{\alpha}\Gamma(\alpha)}\exp\left(-In(x)/\beta\right)$	$0<x<+\infty$	$\alpha = shape\ parameter$ $(\alpha>0)$ $\beta = scale\ parameter$ $(\beta>0)$
Lognormal (3P)	$f(x)=\frac{\exp\left[\frac{-1}{2}\left(\frac{In(x-\gamma)-\mu}{\sigma}\right)^{2}\right]}{(x-\gamma)\sigma\sqrt{2\pi}}$	$\gamma<x<+\infty$	$\sigma = scale\ parameter$ $(\sigma>0)$ $\mu = shapeparameter$ $(\mu>0)$ $\gamma = location\ parameter$ ($\gamma \equiv 0$ yields the two parameter lognormal distribution)
Lognormal (2P)	$f(x)=\frac{\exp\left[\frac{-1}{2}\left(\frac{In(x)-\mu}{\sigma}\right)^{2}\right]}{(x)\sigma\sqrt{2\pi}}$		
Log-Pearson 3	$f(x)=\frac{1}{x\lvert\beta\rvert\Gamma(\alpha)}\left(\frac{In(x)-\gamma}{\beta}\right)^{\alpha-1}\exp\left(-\frac{In(x)-\gamma}{\beta}\right)$	$0<x\le e^{\gamma}$ $\beta<0$ $e^{\gamma}\le x<+\infty$ $\beta>0$	$\alpha = shape\ parameter$ $(\alpha>0)$ $\beta = scale\ parameter$ $(\beta\neq 0)$ $\gamma = location\ parameter$
Normal	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^{2}\right]$	$-\infty<x<+\infty$	$\mu = mean$ $\sigma = standard\ Deviation$ $(\sigma>0)$

Table 1. continue

Distribution	Probability density function	Range	Parameters
Pearson 5 (3P)	$f(x) = \dfrac{\exp\left(-\beta/\left(x-\gamma\right)\right)}{\beta\,\Gamma(\alpha)\left(\left(x-\gamma\right)/\beta\right)^{\alpha+1}}$	$\gamma < x < +\infty$	$\alpha = shape\ parameter \quad (\alpha > 0)$ $\beta = scale\ parameter \quad (\beta > 0)$ $\gamma = location\ parameter$ ($\gamma \equiv 0$ yields the two parameter pearson 5 distribution)
Pearson 5 (2P)	$f(x) = \dfrac{\exp\left(-\beta/x\right)}{\beta\,\Gamma(\alpha)\left(x/\beta\right)^{\alpha+1}}$		
Pearson 6 (4P)	$f(x) = \dfrac{\left(\left(x-\gamma\right)/\beta\right)^{\alpha_1-1}}{\beta B(\alpha_1,\alpha_2)\left(1+(x-\gamma)/\beta\right)^{\alpha_1+\alpha_2}}$	$\gamma \le x < +\infty$	$\alpha_1 = shape\ parameter \quad (\alpha_1 > 0)$ $\alpha_2 = shape\ parameter \quad (\alpha_2 > 0)$ $\beta = scale\ parameter \quad (\beta > 0)$ $\gamma = location\ parameter$ ($\gamma \equiv 0$ yields the three parameter pearson 6 distribution)
Pearson 6 (3P)	$f(x) = \dfrac{\left(x/\beta\right)^{\alpha_1-1}}{\beta B(\alpha_1,\alpha_2)\left(1+x/\beta\right)^{\alpha_1+\alpha_2}}$		
Weibull (3P)	$P(x) = \dfrac{\alpha}{\beta}\left(\dfrac{x-\gamma}{\beta}\right)^{\alpha-1}\exp\left[-\left(\dfrac{x-\gamma}{\beta}\right)^{\alpha}\right]$	$\gamma \le x < +\infty$	$\alpha = shape\ parameter \quad (\alpha > 0)$ $\beta = scale\ parameter \quad (\beta > 0)$ $\gamma = location\ parameter$ ($\gamma \equiv 0$ yields the two parameter weibull distribution)
Weibull (2P)	$P(x) = \dfrac{\alpha}{\beta}\left(\dfrac{x}{\beta}\right)^{\alpha-1}\exp\left[-\left(\dfrac{x}{\beta}\right)^{\alpha}\right]$		

Table 2. Summary of statistics for maximum daily rainfall.

Study Period		Parameters					
		Mean	**Standard Deviation**	**Skewness coefficient**	**Coefficient of variation**	**Maximum**	**Minimum**
Annual	1Jan-31Dec	123.35	38.36	0.77	31.09	229.40	49.32
Seasonal	1June-30Sep	123.31	38.38	0.77	31.12	229.40	49.32
June	1June-30June	50.64	30.04	0.59	59.32	110.50	0.00
July	1July-31July	101.84	44.25	1.19	43.45	229.40	43.60
August	1Aug-31Aug	84.64	34.19	0.06	40.39	144.00	27.83
September	1Sep-30Sep	79.64	47.94	0.56	60.20	203.60	8.80
1 week	4 june-10 June	13.74	19.06	1.53	138.71	68.60	0.00
2 week	11 june-17 june	22.51	27.76	1.68	123.35	104.00	0.00
3 week	18 june-24 june	27.69	27.95	1.83	100.96	110.50	0.00
4 week	25 june-1 july	28.85	26.82	1.20	92.95	109.20	0.00
5 week	2 july-8 july	41.88	46.60	1.72	111.26	196.40	0.00
6 week	9 july-15 july	61.97	51.01	1.49	82.32	229.40	1.20
7 week	16 july-22 july	47.23	27.92	0.58	59.11	111.30	6.30
8 week	23 july-29 july	55.50	39.55	0.84	71.26	151.60	0.00
9 week	30 july-5 aug	43.31	28.44	0.54	65.66	104.00	3.60
10 week	6 aun-12 aug	45.02	35.39	1.14	78.62	144.00	0.00
11 week	13 aug-19 aug	44.03	36.71	1.13	83.38	140.40	0.00
12 week	20 aug-26 aug	52.70	38.23	0.72	72.55	137.80	1.50
13 week	27 aug-2 sep	43.18	37.83	0.85	87.61	138.40	0.00
14 week	3 sep-9 sep	46.48	41.92	0.93	90.21	164.20	0.00
15 week	10 sep-16 sep	40.80	45.71	1.76	112.02	203.60	0.00
16 week	17sep-23 sep	26.20	37.57	1.97	143.38	164.60	0.00
17 week	24 sep-30 sep	22.71	39.85	2.64	175.48	182.00	0.00

The test statistic D, A^2 and χ^2 for each data set were computed for 16 probability distribution. The probability distribution having the first rank along with their test statistic is presented in table 3. It has been observed that Pearson 6 (3P) using Kolmogorov Smirnov test, Generalized Extreme value using Anderson Darling test and Gamma (3P) using Chi-square test obtained the first rank for maximum daily annual rainfall. Thus the three probability distributions were identified as the best fit based on these three tests independently.

Table 3. Study period wise first ranked probability distribution using goodness of fit tests.

Study period	**Test ranking first position**					
	Kolmogorov Smirnov		**Anderson Darling**		**Chi-square**	
	Distribution	**Statistic**	**Distribution**	**Statistic**	**Distribution**	**Statistic**
Annual	Pearson 6	0.1103	Gen. Extreme	0.3992	Gamma (3P)	1.8771
Seasonal	Log gamma	0.1126	Gen. Extreme	0.3915	Gamma (3P)	1.8775
June	Gamma (3P)	0.0714	Gen. Extreme	0.3114	Gamma (2P)	0.3963
July	Pearson 6 (4P)	0.0811	Gen. Extreme	0.2784	Weibull (2P)	1.103
August	Gen. Extreme	0.0668	Gen. Extreme	0.2387	Gen. Extreme	0.2507
Sept	Normal	0.0856	Gen. Extreme	0.377	Weibull (2P)	1.3491
1 Week	Gen. Extreme	0.1719	Gen. Extreme	1.7003	Gen. Extreme	2.7383
2 Week	Gen. Extreme	0.1419	Gen. Extreme	1.5830	Gen. Extreme	3.8349
3 Week	Gen. Extreme	0.0901	Gen. Extreme	0.4485	Gamma (3P)	0.4839
4 Week	Gen. Extreme	0.0817	Gen. Extreme	0.3125	Gen. Extreme	0.3615
5 Week	Gamma (3P)	0.0609	Gen. Extreme	0.4178	Gen. Extreme	0.2052
6 Week	Gen. Extreme	0.1117	Gen. Extreme	0.3299	Weibull (3P)	0.7568
7 Week	Log Pearson 3	0.0843	Log Pearson 3	0.2646	Weibull (3P)	0.4682
8 Week	Gen. Extreme	0.0651	Gen. Extreme	0.2212	Pearson 6 (3P)	0.5732
9 Week	Log Pearson 3	0.0686	Log Pearson 3	0.2070	Gen. Extreme	0.3695
10 Week	Gen. Extreme	0.0744	Gen. Extreme	0.3268	Lognormal (3P)	0.4152
11 Week	Gen. Extreme	0.0749	Gen. Extreme	0.2823	Gen. Extreme	0.6289
12 Week	Gen. Extreme	0.0784	Gen. Extreme	0.3008	Weibull (3P)	0.8678
13 Week	Gen. Extreme	0.1004	Gen. Extreme	0.3869	Gen.gamma (3P)	0.5226
14 Week	Gen. Extreme	0.0933	Gen. Extreme	0.4223	Gen. Extreme	0.8971
15 Week	Gamma (3P)	0.0952	Gen. Extreme	0.4418	Lognormal (3P)	0.4831
16 Week	Gen. Extreme	0.1662	Gen. Extreme	1.514	Gen. Extreme	4.4507
17 Week	Gen. Extreme	0.1726	Gen. Extreme	1.3549	Gen. Extreme	3.0257

The combination of total test score were obtained for each data set for all 16 probability distribution. This was done to identify the fourth probability distribution in addition to three identified earlier for obtaining the best fit probability distribution. This distribution was identified using maximum overall score based on sum of individual point score obtained from three selected goodness of fit test. The distributions identified which were having highest score are presented in Table 4.

Those distributions which were having same highest score were also included in the selected probability distribution, for annual data set Lognormal (3P) and Pearson 5 (3P) were having 36 as highest score so both these distributions were selected. It was also observed that some of the probability distribution already having the first rank in table 3 was also having the highest scores and hence three or less distributions were identified. The distributions so identified are listed in Table 5 where the parameter of these identified distribution for each data set are mentioned.

Table 4. Score wise best fit probability distribution.

Study Period	Distributions with highest Score	
	Distribution	Score
Annual	Lognormal (2P) & Pearson 5 (3P)	36
Seasonal	Gen. Extreme	36
June	Gamma (2P)	35
July	Pearson 5 (3P)	38
August	Gen. Extreme	47
Sept	Normal	44
1 Week	Gen. Extreme	42
2 Week	Gen. Extreme	42
3 Week	Gen. Extreme	38
4 Week	Gen. Extreme	42
5 Week	Gen. Extreme & Gamma (3P)	39
6 Week	Gen. Extreme & Lognormal (2P)	38
7 Week	Log Pearson 3	45
8 Week	Gen. Extreme	38
9 Week	Log Pearson 3	42
10 Week	Gamma (3P)	37
11 Week	Gen. Extreme	42
12 Week	Gen. Extreme	42
13 Week	Gen. Extreme	33
14 Week	Gen. Extreme	42
15 Week	Gen. Extreme	38
16 Week	Gen. Extreme	42
17 Week	Gen. Extreme	42

These values of the parameter were used to generate random numbers for each data set and the least square method was used for the rainfall analysis. The random numbers were generated for actual and estimated observations for all the 37 years. The residuals were computed for each data set. Sum of these deviation were obtained for all identified distribution. The probability distribution having minimum deviation was treated as the best selected probability distribution for the individual data set. The best selected probability distribution for each data set is presented in Table 6.

Conclusion

The result of rainfall analysis for identifying the best fit probability distribution revealed that the distribution pattern for different data set can be identified out of a large number of commonly used probability distributions by using different goodness of fit tests.

The data showed that the annual daily maximum rainfall received at any time ranged between 49.32mm (minimum) to 229.40mm (maximum) indicating a very large range of fluctuation during the period of study. It was observed that the best probability distributions obtained for the maximum daily rainfall for different data set are different. The lognormal and gamma distribution were found as the best fit probability distribution for the annual and monsoon season period of study, respectively. Generalized extreme value distribution was observed in most of the weekly period as best fit probability distribution. The best fit probability distribution of monthly data was found to be different for each month. The scientific results clearly established that the analytical procedure devised and tested in this study may be suitably applied for the identification of the best fit probability distribution of weather parameters.

Table 5. Parameters of the best fitted distributions.

Study Period	Distributions	Parameters
Annual	Gamma (3P)	α=9.2806 β=12.313 γ=9.0742
	Gen. Extreme Value	k=0.00215 σ=30.534 μ=105.66
	Lognormal (2P)	σ=0.30778 μ=4.7685
	Pearson 5 (3P)	α=29.458 β=5637.3 γ=-74.753
	Pearson 6 (3P)	α_1=26.983 α_2=18.149 β=77.942
Seasonal	Gamma (3P)	α=9.2096 β=12.367 γ=9.4188
	Gen. Extreme Value	k=0.00223 σ=30.555 μ=105.61
	Log-Gamma	α=233.31 β=0.02044
June	Gamma (2P)	α=2.5411 β=20.484
	Gamma (3P)	α=2.8415 β=17.823
	Gen. Extreme Value	k=-0.03019 σ=25.124 μ=36.872
July	Gen. Extreme Value	k=0.12763 σ=30.153 μ=80.117
	Pearson 5 (2P)	α=6.6399 β=576.36
	Pearson 6 (4P)	α_1=69.525 α_2=6.6044 β=7.9088 γ=4.0848
	Weibull (2P)	α=2.8985 β=110.24
August	Gen. Extreme Value	k=-0.25796 σ=34.667 μ=71.85
September	Gen. Extreme Value	k=-0.12586 σ=43.458 μ=59.413
	Normal	σ=47.942 μ=79.637
	Weibull (2P)	α=1.4832 β=87.627
1 week	Gen. Extreme Value	k=0.41694 σ=7.748 μ=3.9059
2 week	Gen. Extreme Value	k=0.29523 σ=14.142 μ=8.5866
3 week	Gamma (3P)	α=0.98108 β=28.223
	Gen. Extreme Value	k=0.28511 σ=14.234 μ=13.956
4 week	Gen. Extreme Value	k=0.14693 σ=17.979 μ=15.445
5 week	Gamma (3P)	α=0.80779 β=51.847
	Gen. Extreme Value	k=0.31815 σ=22.959 μ=18.23
6 week	Gen. Extreme Value	k=0.08574 σ=35.618 μ=38.129
	Lognormal (3P)	σ=0.57786 μ=4.2884 γ=-23.657
	Weibull (3P)	α=1.0694 β=62.311 γ=1.03
7 week	Log-Pearson 3	α=6.3474 β=-0.28657 γ=5.4628
	Weibull (3P)	α=1.5535 β=48.053 γ=3.8657
8 week	Gen. Extreme Value	k=-0.03104 σ=32.641 μ=37.631
	Pearson 6 (3P)	α_1=2.3608 α_2=29617.0 β=7.4421E+5
9 week	Gen. Extreme Value	k=-0.02735 σ=24.12 μ=30.025
	Log-Pearson 3	α=6.6347 β=-0.32385 γ=5.6431
10 week	Gamma (3P)	α=1.6177 β=27.827
	Gen. Extreme Value	k=0.17471 σ=22.899 μ=27.062
	Lognormal (3P)	σ=0.93308 μ=3.5125
11 week	Gen. Extreme Value	k=0.15915 σ=24.328 μ=25.484
12 week	Gen. Extreme Value	k=-0.01611 σ=31.581 μ=34.967
	Weibull (3P)	α=1.2691 β=55.905 γ=0.46572
13 week	Gen. Extreme Value	k=0.07246 σ=28.517 μ=24.529
	Gen. Gamma (3P)	k=3.5131 α=0.21383 β=113.53
14 week	Gen. Extreme Value	k=0.08526 σ=30.955 μ=25.773
15 week	Gamma (3P)	α=0.79685 β=51.203
	Gen. Extreme Value	k=0.31084 σ=22.826 μ=17.629
	Lognormal (3P)	σ=1.3585 μ=3.1717
16 week	Gen. Extreme Value	k=0.43334 σ=14.285 μ=7.3782
17 week	Gen. Extreme Value	k=0.57577 σ=9.5565 μ=4.6363

Table 6. Best fit probability distribution.

STUDY PERIOD	BEST-FIT
Annual	Lognormal (2P)
Seasonal	Gamma (3P)
June	Gamma (2P)
July	Pearson 5 (2P)
August	Gen. Extreme value
September	Normal
1 Week	Gen. Extreme value
2 Week	Gen. Extreme value
3 Week	Gamma (3P)
4 Week	Gen. Extreme value
5 Week	Gamma (3P)
6 Week	Lognormal (3P)
7 Week	Log Pearson 3
8 Week	Pearson 6 (3P)
9 Week	Log Pearson 3
10 Week	Lognormal (3P)
11 Week	Gen. Extreme value
12 Week	Weibull (3P)
13 Week	Gen. Gamma (3P)
14 Week	Gen. Extreme value
15 Week	Lognormal (3P)
16 Week	Gen. Extreme value
17 Week	Gen. Extreme value

References

1. Bhakar, S.R., Bansal, A.N., Chhajed, N., and Purohit, R.C. Frequency analysis of consecutive day's maximum rainfall at Banswara, Rajasthan, India. ARPN Journal of Engineering and Applied Sciences. 2006;1(3): 64-67.
2. Bhargava, P.N. The influence of rainfall on crop production. Research Journal of Jawahar Lal Nehru Krishi Vishwavidyalaya. 1974;32, No. 1 and 2.
3. Biswas, B.C. and Khambeta, N.K. Distribution of short period rainfall over dry farming tract of Maharashtra. Journal of Maharashtra Agricultural University. 1989.
4. Chapam, T. Stochastic models for daily rainfall. The Institution of Engineers, Australia, National Conference Publication. 1994;94 (15): 7-12.
5. Deidda, R. and Puliga, M. Sensitivity of goodness-of-fit statistics of rainfall data rounding off. Physics and Chemistry of the Earth. 2006; 31: 1240-1251.
6. Duan, j., Sikka, A.K., and Grant, G.E. A comparison of stochastic models for generating daily precipation at the H.J. Andrews Experiment Forest. Northwest Science. 1995; 69(4): 318-329.
7. Fisher R.A. The influence of the rainfall on the yield of wheat at Rothamsted. Philosophical transaction of the Royal Society of London. 1924; Series B, Vol. 213.
8. Hanson, L.S., and Vogel, R. The probability distribution of daily rainfall in the United States. Proc. In World Environment and Water Resources Congress Conference. 2008.
9. Kulandaivelu, R. Probability analysis of rainfall and evolving cropping system for Coimbatore. Mausam. 1984;5, 3: 257-258.
10. Kwaku, X.S., and Duke, O. Characterization and frequency analysis of one day annual maximum and two to five consecutive days maximum rainfall of Accra, Ghana. ARPN Journal of Engineering and Applied Sciences. 2007; vol. 2, no. 5: 27-31.
11. Lee, C. Application of rainfall frequency analysis on studying rainfall distribution characteristics of Chia-Nan plain area in Southern Taiwan. Journal of Crop, Environment & Bioinformatics. 2005;(2): 31-38.
12. Manning, H.L. Confidence limits of monthly rainfall. Jour, Agril. Sci. 1950;40: 169.
13. Ogunlela, A.O. Stochastic Analysis of Rainfall Events in Ilorin, Nigeria. Journal of Agricultural Research and Development. 2001;Volume 1: 39-50.
14. 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. 2009;2(3).
15. Phien, H.N., and Ajirajah, T.J. Applications of the log-Pearson Type-3 distributions in hydrology. Journal of hydrology. 1984;73: 359-372.
16. RamaRao, B.V., Kavi, P.S. and Sridharan, P.C. Study of rainy days and wet spells at Bijapur. Annual Arid Zone. 1975;14, 4: 371-372.
17. Salami, A.W. Prediction of the annual flow regime along Asa River using probability distribution models. AMSE periodical, Lyon, France. Modelling C. 2004;65(2): 41-56.
18. Sen, Z., and Eljadid, A.G. Rainfall distribution functions for Libya and Rainfall Prediction. Hydrol. Sci. J. 1999;4(5): 665-680.
19. Tao, D.Q., Nguyen, V.T., and Bourque, A. On selection of probability distributions for representing extreme precipitations in Southern Quebec. Annual Conference of the Canadian Society for Civil Engineering. 2002;5th- 8th June: 1-8.
20. Tippet, L.H.C. On the effect of sunshine on wheat yield at Rothamsted. Jour. Agril. Sci. 1929;60, 2.
21. Topalogu, F. Determining Suitable Probability Distribution Models for Flow and Precipitation Series of the Seyhan River Basin. Turk. Journal of Agric. 2002;26: 189-194.
22. Upadhaya, A., and Singh, S.R. Estimation of consecutive day's maximum rainfall by various methods and their comparison. Indian Journal of S. Cons. 1998;26 (2): 193-2001.
23. Wilks, D.S. Multi-site generalization of a daily stochastic precipitation model. J. Hydrol. 1998;210: 178-191.

Dated: 02/07/2010