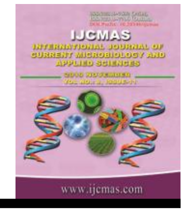
Int.J.Curr.Microbiol.App.Sci (2019) *8*(11): 1437-1444

EXCELLENT PUBLISHERS

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Volume 8 Number 11 (2019)**
Journal homepage: http://www.ijcmas.com



Case Study https://doi.org/10.20546/ijcmas.2019.811.169

Probability Distribution of Daily Maximum Rainfall Data for Six Different Geographical Locations in West Bengal-A Case Study

Jit Sankar Basak*, Debaditya Gupta, Mriganka Saha and Sayan Pan

Bidhan Chandra Krishi Viswavidyalaya, India

**Corresponding author*

Keywords

Rainfall, Probability Distributions, Goodness of fit Test, EasyFit

Article Info

Accepted:
12 October 2019
Available Online:
10 November 2019

A B S T R A C T

The study of probability analysis of rainfall is necessary to find the most suitable distribution model that could visualize rainfall extremities. In this present study probability distribution viz. Normal, Log-Normal, Exponential, Gumble, Generalized Extreme Value, Weibull, Generalized gamma, Log Pearson type 3, logistic distribution with two parameter and two-parameter Log-Logistic distribution for daily maximum rainfall data of 37 years for six different locations of West Bengal to find the best fit distribution model among each location by using Goodness of Fit test (χ^2 test), Kolmogorov-Smirnov test (K-S test) and Anderson Darling test. The result of the sum of rank it is found that Log Pearson type 3 distribution fitted best for three geographical locations viz. Kharagpur, Bolpur and Balurghat followed by Gumble in Kolkata, Log logistic in Darjeeling and generalized extreme value in Berhampur.

Introduction

Rainfall is a crucial factor in the hydrological cycle. It plays an essential role in global water balance, irrigation scheduling, water resources planning and management. Extreme rainfall causes flood and erosion hazard whereas, scarcity of rainfall causes drought conditions. The analysis of rainfall data deals with interpreting the past record of rainfall events in terms of future probabilities of occurrence (Kumar and Bhardwaj, 2015). Rainfall patterns vary from country to country as well as from weather station to station. However, homogeneity can also be found bytaking annual extreme precipitation as an extreme event (Khudri and Sadia, 2013). Frequency analysis of rainfall datahad been done for different places in India (Sharda and Bhushan, 1985; Prakash and Rao, 1986; Rizvi *et al.*, 2001). Sen *et al.*, (1999) found that the gamma probability distribution provided the best fit for monthly maxima rainfall in arid regions in Libya. Lee (2005) showed that LP3 distributions fitted best to 50% of the stations for the rainfall distribution of the Chia-Nan plain area of Taiwan. Bhakar *et al.*, (2006) did a frequency analysis of consecutive days peak rainfall at Banswara, Rajasthan, India, was found gamma distribution as the best fit as compared to other methods after due testing with Chi-square value.

Mandal and Choudhury (2015) studied the annual, seasonal and monthly maximum daily rainfall data for Sagar Island, located on the continental shelf of the Bay of Bengal and found that normal (N) distributions were fitted best to annual, post-monsoon and summer seasons, while Lognormal (LN2), Weibull (W2) and Pearson type 5 were fitted best to pre-monsoon, monsoon and winter seasons respectively.

West Bengal has an agriculture-based economy where precipitation plays a vital role important. The extreme events of rainfall may affect agriculture, ecosystems, biodiversities and livelihood patterns (Alam *et al.*, 2018).

It is very important to extract the patterns of extreme rainfall events to determine risk factors that can be used to develop long-term measures to save human lives and property as well as soil and water conservation and irrigation planning.

Therefore in the present study probability distribution of maximum daily rainfall data for six different locations in West Bengal and to check the best fit of the distributions in these locations.

Materials and Methods

Daily rainfall data (0.5° x 0.5° resolution) of 37 years (1st January 1982 to 31st December 2018) for six different locations of West Bengal viz. Kharagpur (Lat 22.346° N, Lon 87.232° E), Kolkata (Lat 22.572° N, Lon 88.363°E), Darjeeling (Lat 27.041° N, Lon 88.266° E), Bolpur (Lat 23.669° N, Lon 87.688°E), Balurghat (Lat 25.237° N, Lon 88.783° E), Berhampur (Lat 24.098° N, Lon 88.268° E) were taken from NASA Power Data Access Viewer Website (https://power.larc.nasa.gov/data-access-viewer/). From the daily rainfall data, the maximum daily rainfall data from each year among all 37 years were selected for the probability distribution for each location. The geographical locations are shown in figure 1.

Nine different distributions are used to fit the data by using the EasyFit software package. The probability density functions for each distribution are given below. Abbreviations are given in parenthesis.

Normal distribution (N)

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

$$-\infty < x < \infty, -\infty \leq \mu \leq \infty \text{ and } \sigma > 0$$

Log normal distribution (LN)

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2\right\}$$

$$0 < x < \infty, \ 0 < \mu < \infty \ \ and \ \ \sigma > 0$$

Exponential distribution (EXP)

$$f(x) = \lambda \exp[-\lambda(x-\gamma)]$$

$$\gamma \leq x < \infty \ \ and \ \lambda > 0$$

Gumble distribution (GUM)

$$f(x) = \frac{1}{\sigma} \exp[-(z) - \exp(-z)]$$

where,

$$z = \frac{x - \mu}{\sigma}$$

$$-\infty < x < \infty \text{ and } \sigma > 0$$

Generalized Extreme Value distribution (GEV)

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

where,

$$z = \frac{x - \mu}{\sigma}$$

$$0 < x < \infty, \quad 0 < \mu < \infty \text{ and } \sigma > 0$$

Weibull distribution (W)

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left[-\left(\frac{x}{\beta}\right)\right]^{\alpha}$$

$$0 \leq x < \infty \text{ and } \alpha, \beta > 0$$

Generalized gamma distribution with three parameter (GG)

$$f(x) = \frac{kx^{k\alpha-1}}{\beta^{k\alpha}\Gamma(\alpha)}\exp\left(-\left(\frac{x}{\beta}\right)^{k}\right)$$

$$0 \leq x < \infty \text{ and } k, \alpha, \beta > 0$$

Log-Pearson type 3 distribution (LP)

$$f(x) = \frac{1}{x|\beta|\Gamma(\alpha)}\left(\frac{\ln(x)-\gamma}{\beta}\right)^{\alpha-1} \exp\left(-\frac{\ln(x)-\gamma}{\beta}\right)$$

$$\alpha > 0, \beta \neq 0$$

$$0 < x \leq e^{\gamma} \quad, \beta < 0$$

$$e^{\gamma} \leq x < \infty \quad, \beta > 0$$

Logistic distribution with two parameters (LG)

$$f(x) = \frac{\exp(-z)}{\sigma(1+\exp(z))^2}$$

$$z = \frac{x - \mu}{\sigma}$$

$$-\infty < x < \infty, -\infty < \mu < \infty \text{ and } \sigma > 0$$

Log-logistic distribution with two parameters (LLG)

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\left(1+\left(\frac{x}{\beta}\right)^{\alpha}\right)^{-2}$$

$$0 \leq x < \infty \text{ and } \alpha, \beta > 0$$

The best fit probability distribution criteria are tested by using Goodness of fit test (Chi-squared test), Kolmogorov-Smirnov test and Anderson Darling test. The description of the tests are expressed below

Chi-Squared (χ^2) Test

The Goodness of fit Chi-Squared (χ^2) test is a non-parametric test. It is used to determine whether a sample comes from a population with a specific distribution or not. It is used for continuous sample data only. The Chi-Squared statistic is defined as,

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

where,

O_i is the observed frequency for class i and E_i is the expected frequency for class i

H_0: the data follows a specified distribution, against

H_1: the data do not follow the specified distribution.

H_0 is rejected at α % level of significance if calculated χ^2_{k-1} > tabulated $\chi^2_{\alpha,k-1}$. Otherwise it is accepted.

Kolmogorov-Smirnov (K-S) Test

It is also a popular non-parametric test. This test is used to decide if a sample comes from a hypothetical continuous distribution or not. It test is based on the empirical cumulative distribution function (ECDF). Assume that, we have a random sample $x_1,..., x_n$ from some distribution with CDF F(x). The empirical CDF is denoted by,

$$F_n(x) = \frac{1}{n}.[\text{No. of observations} \leq x]$$

The Kolmogorov-Smirnov statistic is denoted by D, is based on the maximum positive difference between the theoretical and the empirical cumulative distribution function:

$$D_n = \max_{1 \leq i \leq n} |F_n(x_i) - F_o(x_i)|$$

where,

F_0 is the hypothetical cumulative distribution function

H_0: the data follows a specified distribution, against

H_1: the data do not follow the specified distribution.

H_0 is rejected at α % level of significance if calculated D_n>tabulated$D_{\alpha,n}$. Otherwise it is accepted. Otherwise it is accepted.

Anderson-Darling Test

The Anderson-Darling is also a distribution-free method or non-parametric method of the testing procedure to compare the fit of an observed cumulative distribution function to an expected cumulative distribution function. This test gives more weight to the tails than the Kolmogorov-Smirnov test. The Anderson-Darling statistic (A^2) is defined as

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

where,

F denotes the cumulative distribution function.

H_0: the data follows a specified distribution, against

H_1: the data do not follow the specified distribution.

The critical values for the Anderson-Darling test depend on the specific distribution that is being tested. The hypothesis regarding the specific distributional is rejected at the significance level of α % if the test statistic, A^2, is greater than the critical value obtained from a table.

Results and Discussion

The data in this study shows the maximum daily rainfall in the year from 1982 to 2018. The maximum daily rainfall varies in a different station based on geographical location i.e. latitude, longitude, altitude and environmental factors such as proximity to the sea. The descriptive statistics of the maximum daily rainfall data along with best-fit results of K-S test, A-D test, Goodness of fit and the best sum of rank result of six different geographical locations of West Bengal are shown in the following table (Table 1 and Table 2).

In table 2, figure in the parenthesis for K-S, A-D and Goodness of fit test given the test statistics value and for the sum of the rank of three tests are given in sum of the rank column.

Fig.1 Geographical location of the six locations in West Bengal map

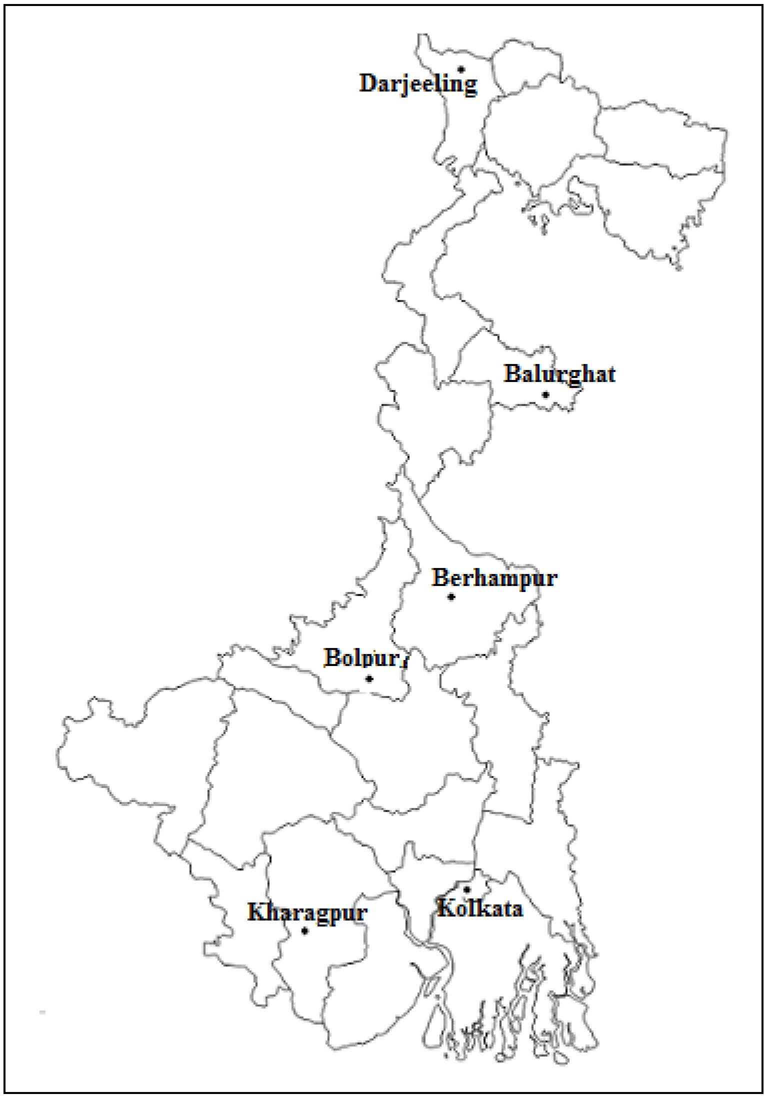


Table.1 Descriptive statistics of daily maximum rainfall data

Location	Minimum	Maximum	Mean	Standard Deviation	Coefficient of skewness
Kharagpur	35.800	144.450	61.107	21.024	1.833
Kolkata	29.660	161.020	64.595	26.651	1.404
Darjeeling	26.350	100.210	46.427	15.673	1.447
Bolpur	16.920	80.220	44.082	16.268	0.577
Balurghat	18.080	118.630	55.067	25.666	0.576
Berhampur	12.520	81.760	43.440	15.774	0.532

Table.2 Results of K-S test, A-D test, Goodness of fit and sum of ranks

Location	Kolmogorov Smirnov Test		Anderson- Darling Test		Chi-Squared Test		Sum of ranks	
	Best fit	Worse fit	Best fit	Worse fit	Best fit	Worse fit	Best fit	Worse fit
Kharagpur	LP(0.076)	LG(0.092)	LP(0.223)	EXP(1.839)	GG(0.536)	LLG(2.314)	LP(4)	LG(26)
Kolkata	W(0.079)	EXP(0.160)	GEV(0.672)	EXP(1.844)	GUM(0.402)	EXP(2.182)	GUM(5)	EXP(30)
Darjeeling	LL(0.079)	LG(0.161)	LL(0.079)	LG(0.858)	GEV(0.655)	N(3.065)	LL(4)	N(28)
Bolpur	GEV(0.096)	EXP(0.233)	LP(0.439)	EXP(3.442)	EXP(2.788)	W(6.683)	LP(6)	W(24)
Balurghat	GEV(0.090)	EXP(0.140)	LP(0.354)	EXP(1.972)	LL(0.342)	LL(5.941)	LP(5)	EXP and LG(28)
Berhampur	GEV(0.071)	EXP(0.266)	GEV(0.162)	EXP(4.908)	LL(0.629)	EXP(4.816)	GEV(7)	EXP(30)

Table.3 Parameters of ten different distributions

SL No	Distribution	Kharagpur	Kolkata	Darjeeling	Bolpur	Balurghat	Berhumpur
1	Normal	μ = 61.107 σ = 21.024	μ = 64.595 σ = 26.651	μ = 64.427 σ = 15.673	μ = 44.082 σ = 16.268	μ = 55.066 σ = 25.666	μ = 43.44 σ = 15.774
2	Log Normal	μ = 4.064 σ = 0.301	μ = 4.094 σ = 0.381	μ = 3.789 σ = 0.304	μ = 3.719 σ = 0.367	μ = 3.898 σ = 0.480	μ = 3.703 σ = 0.383
3	Gumble	μ = 51.645 σ = 16.392	μ = 52.601 σ = 20.78	μ = 39.374 σ = 12.22	μ = 36.76 σ = 12.684	μ = 43.515 σ = 20.012	μ = 36.341 σ = 12.299
4	Generalized gamma	k = 1.047 α = 9.379 β = 7.234	k = 1.037 α = 6.301 β = 10.996	k = 1.038 α = 9.562 β = 5.291	k = 1.010 α = 7.507 β = 6.004	k = 1.004 α = 4.634 β = 11.963	k = 0.997 α = 7.548 β = 5.728
5	Gen Extreme value	k = 0.109 μ = 51.254 σ = 14.109	k = 0.055 μ = 52.135 σ = 19.625	k = 0.106 μ = 38.928 σ = 10.821	k = -0.006 μ = 36.419 σ = 13.43	k = -0.041 μ = 43.262 σ = 21.917	k = -0.069 μ = 36.428 σ = 13.682
6	Logistic	μ = 61.107 σ = 11.591	μ = 64.595 σ = 14.694	μ = 46.427 σ = 8.641	μ = 44.082 σ = 8.969	μ = 55.066 σ = 14.151	μ = 43.44 σ = 8.697
7	Log logistic	α = 5.956 β = 56.782	α = 4.506 β = 58.347	α = 5.749 β = 43.242	α = 4.423 β = 40.497	α = 3.373 β = 48.114	α = 4.22 β = 39.789
8	Log Pearson type 3	α = 9.663 β = 0.098 γ = 3.116	α = 73.389 β = 0.045 γ = 0.785	α = 15.697 β = 0.078 γ = 2.569	α = 870.85 β = -0.013 γ = 14.704	α = 156.56 β = -0.039 γ = 9.989	α = 10578 β = -0.119 γ = 4.965
9	Exponential	λ = 0.039 γ = 35.8	λ = 0.029 γ = 29.66	λ = 0.032 γ = 12.52	λ = 0.037 γ = 16.92	λ = 0.027 γ = 18.08	λ = 0.032 γ = 12.52
10	Weibull	α = 4.139 β = 64.709	α = 3.166 β = 69.222	α = 3.969 β = 49.556	α = 3.103 β = 48.21	α = 2.393 β = 60.32	α = 3.028 β = 47.573

Berhampur shows the lowest minimum daily rainfall and Kolkata received maximum daily rainfall and also maximum daily rainfall data variation exists in Kolkata which is revealed by the standard deviation of 26.651 mm. Daily rainfall data for each location showing positively skewed means all the maximum rainfalls are on the left-hand side of the distribution. Best fitted (rank 1) criteria based on the three above mentioned test shows different results and rank in different cases. According to K-S test Log Pearson type 3 distribution, Weibull, Log logistic distribution are best fitted for the station Kharagpur, Kolkata and Darjeeling respectively and Generalized extreme value distribution is fitted for Bolpur, Balurghat, and Berhampur. For Kharagpur and Darjeeling Logistic distribution shows the worse result (rank 10) and for the other locations, exponential distribution (2P) shows the worse result. Darling test conferred that Log Pearson type 3 distribution fitted (rank 1) for the stations Kharagpur, Bolpur and Balurghat, generalized Extreme Value distribution for station Kolkata and Berhampur and Log logistic distribution for Darjeeling. The worse fit (rank 10) result is obtained by exponential distribution in all the stations except Darjeeling. For Darjeeling Logistic distribution shows a worse fit. According to Chi-Squared test for goodness of fit Generalized gamma, Gumble, Generalized extreme value, the exponential distribution shows best fit for the stations Kharagpur, Kolkata, Darjeeling, Bolpur respectively and Log logistic distribution for Balurghat and Berhampur. The worse result is observed in Log logistic, Normal, Weibull, Logistic in the station Kharagpur, Darjeeling, Bolpur, Balurghat respectively and Exponential distribution for Kolkata and Berhampur. Since the result of the present study shows the fitness of different distributions in a different location, therefore, the rank of three tests in each location are summed to get the best distribution for that station. The best and worse fitted distribution is considered as the lowest and highest value of the rank-sum respectively. According to this Gumble, Log logistic and Generalized Extreme Value shows the best fit for the station Kolkata, Darjeeling and Berhumpur respectively and Log Pearson type 3 distribution shows the best fit for the remaining three-station viz. Kharagpur, Bolpur and Balurghat. Worse fit is observed in case of Logistic, Exponential, Normal distribution in the case of Kharagpur, Kolkata, Darjeeling respectively, Normal and Weibull jointly (equal rank-sum) for Bolpur and exponential and Logistic together in Balurghat, Exponential in Berhampur. The parameters for the different distributions are expressed in the following table (Table 3).

In the present study, ten different probability distributions are fitted for the 37-year dataset (1982-2018) of six geographical locations in West Bengal and from the results of the sum of rank, it is found that Log Pearson type 3 distribution fitted based for three geographical locations viz. Kharagpur, Bolpur and Balurghat followed by Gumble in Kolkata, Log logistic in Darjeeling and generalized extreme value in Berhampur.

References

Alam, M., Emura, K., Farnham, C., and Yuan, J. 2018. Best-fit probability distributions and return periods for maximum monthly rainfall in Bangladesh. Climate, 6(1), 9.

Bhakar, S. R., Bansal, A. K., Chhajed, N., and Purohit, R. C. 2006. Frequency analysis of consecutive days maximum rainfall at Banswara, Rajasthan, India. ARPN Journal of Engineering and Applied Sciences, 1(3), 64-67.

Gupta, S. C. 1982. Fundamentals of mathematical statistics: A modern approach. Sultan Chand and Sons.

Khudri, M. M., and Sadia, F. 2013.

Determination of the best-fit probability distribution for annual extreme precipitation in Bangladesh. European Journal of Scientific Research, 103(3), 391-404.

Kumar, R., and Bhardwaj, A. 2015. Probability analysis of return period of daily maximum rainfall in annual data set of Ludhiana, Punjab. Indian Journal of Agricultural Research, 49(2).

Lee, C. 2005. Application of rainfall frequency analysis on studying rainfall distribution characteristics of Chia-Nan plain area in Southern Taiwan. Crop Environ. Bioinf, 2, 31-38.

Mandal, S., and Choudhury, B. U. 2015. Estimation and prediction of maximum daily rainfall at Sagar Island using best-fit probability models. Theoretical and applied climatology, 121(1-2), 87-97.

Prakash, C., and Rao, D. H. 1986. Frequency analysis of rainfall data for crop planning-Kota. Indian J. Soil Cons, 14(2), 23-26.

Rizvi, R. H., Singh, R., Yadav, R. S., Tewari, R. K., Dadhwal, K. S., and Solanki, K. R. 2001. Probability analysis of annual maximum daily rainfall for Bundelkhand region of Uttar Pradesh. Indian J. Soil Conser, 29(3), 259-262.

Şen, Z., and Eljadid, A. G. 1999. The rainfall distribution function for Libya and rainfall prediction. Hydrological sciences journal, 44(5), 665-680.

Sharda, V. N., and Bhushan, L. S. 1985. Probability analysis of annual maximum daily rainfall for Agra. Indian J. Soil Cons, 13(1), 16-20.

How to cite this article:

Jit Sankar Basak, Debaditya Gupta, Mriganka Saha and Sayan Pan. 2019. Probability Distribution of Daily Maximum Rainfall Data for Six Different Geographical Locations in West Bengal-A Case Study. *Int.J.Curr.Microbiol.App.Sci*. 8(11): 1437-1444.
doi: https://doi.org/10.20546/ijcmas.2019.811.169