European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 7, Issue 8, 2020

RAINFALL FORECAST THROUGH ROOT MEAN SQUARED ERROR USING DOUBLE EXPONENTIAL & LOG-PEARSON III PROBABILITY DISTRIBUTIONS.

B. R. Sreedhar[1] and K. Muthyalappa[2]
[1]Department of Mathematics C.B.I.T., Gandipet, Hyderabad
[2]Department of Statistics, Sri Krishnadevaraya University, Ananthapuramu

Abstract - The study of analysis of rainfall is vital to find the relevant distribution model to anticipate the natural phenomena (earthquake, floods, rainfall, etc.;). The main theme of this study to determine the best fit of probability distribution in the case of frequency of daily rainfall in past 37 years (1982-2019) from all districts of the state of Andhra Pradesh, India, by using different statistical analysis and continuous probability distributions. The daily rainfall data are analysed using two different probability models, those are Double Exponential Distribution and Log-Pearson III Probability Distribution. Efficiency of the all probability models are compared using Root Mean Squared Error (RMSE) value of Chi-square Goodness of Fit. It is precisely witnessed that the Double Exponential Distribution was identified to be the best fit for forecasting daily rainfall (mm).

Keywords: Rainfalls, Probability distributions, Root mean squared error, Double Exponential Distribution and Log-Pearson-III Probability Distribution.

INTRODUCTION:

The study of rainfall data is one of the important events in hydrological cycle. It is the major component of the water cycle for accumulating the large amount of water on the universe. It provides many types of ecosystems, for crop irrigation and hydroelectric power stations. This plays a important role in many non agricultural and agricultural applications. The average rainfall in our country is 1185mm per year and it ranges from 339 mm to 2250 mm annually. Normally 80 to 85% of the total annual rainfall in India accounting from the months of June to September. Rainfall is a unique phenomenon that is highly diversified with respect to space and time. Analysis of Rainfall and computation of daily rainfall should improve the management of water resources application and the effective utilization of water. Probability and frequency study of rainfall data enables us to determined the expected rainfall at various cases, this information is also used to prevent floods and droughts and apply to development and designing of water resources associated to technology such as reservoir design, flood control work and soil and water conservation setting up like dames.

MATERIAL/METHODS

The rainfall data (1982-2019) collected from Indian meteorological department. The present study determined on versatile of rainfall using Double Exponential and Log-Pearson III Probability Distribution for stochastically analysis.

STOCHASTIC ANALYSIS

The following formulae are used for the basic statistical analysis such as Arithmetic Mean, Standard error, Coefficient of variation and coefficient of Skew ness.

Arithmetic Mean $(\bar{y}) = \frac{\sum y_i}{n}$

Standard Error $(S_n) = \sqrt{\dfrac{\sum(x-\bar{x})}{n-1}}$

Coefficient of Variation $(CV) = \dfrac{\sigma}{\bar{x}} \times 100$

Coefficient of Skewness $(C_{sk}) = \dfrac{n\sum(w_i - \bar{w})^2}{(n-1)(n-2)S_n}$

W = log value of rainfall data

$\overline{W} =$ mean value of Rainfall data

(n) = Sample size

Rainfall data fitted using various probability distributions Those are Double Exponential distribution, Lognormal distribution and Chi square goodness of fit.

DOUBLE EXPONENTIAL DISTRIBUTION

$E_T = P + KS_n$

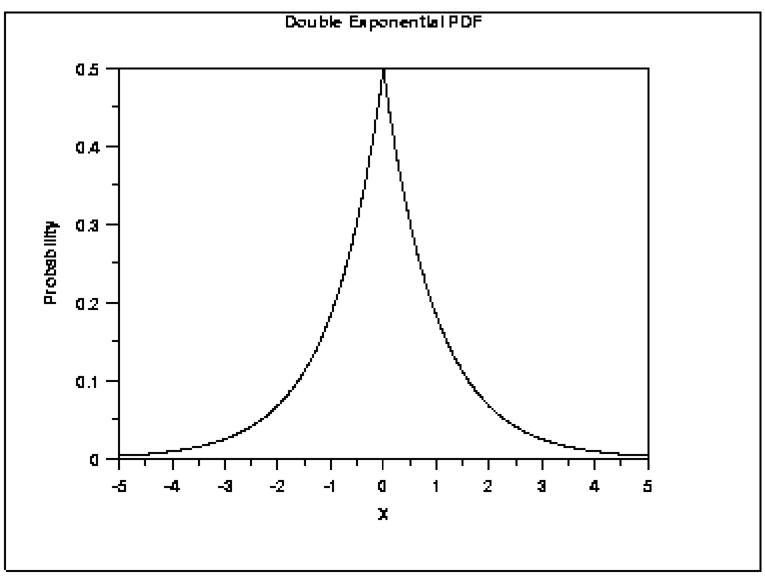
Where E_T is Perdition of Rainfall amount for a return period of 't'"

The general formula for the probability density function of the double exponential distribution is

$$f(x) = \frac{e^{-\left|\frac{x-\mu}{\beta}\right|}}{2\beta}$$

Where μ is the location parameter and β is the scale parameter. The scale where μ=0 and β=1 is called the Double exponential distribution. The equation for the standard double exponential distribution is

$$f(x) = \frac{e^{-|x|}}{2}$$



The formula for the cumulative distribution function of the double exponential distribution is

$$F(x) = \begin{cases} \frac{e^x}{2} & \text{for } x < 0 \\ 1 - \frac{e^{-x}}{2} & \text{for } x \geq 0 \end{cases}$$

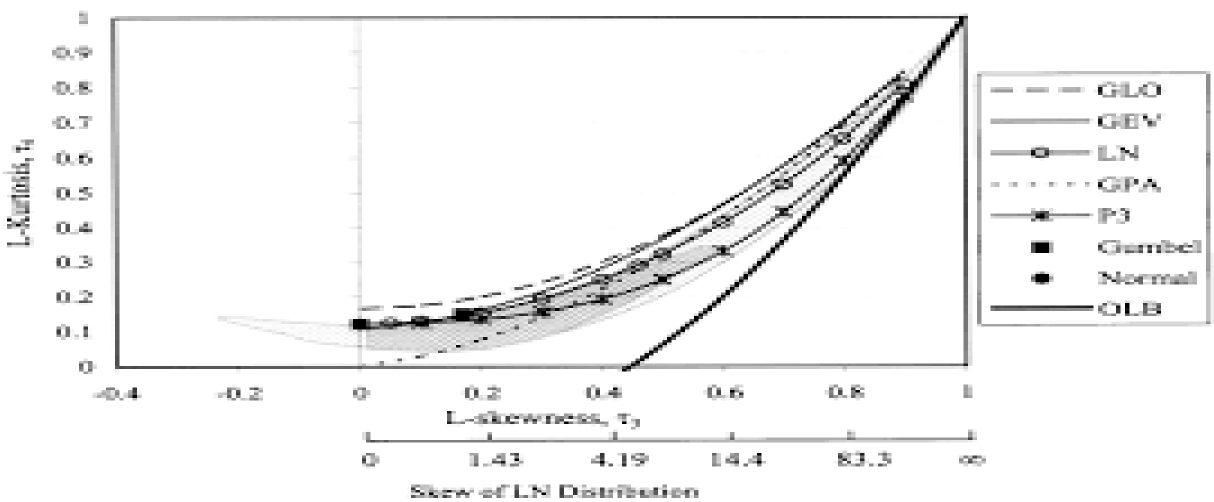
THE LOG-PEARSON TYPE (LPT) III DISTRIBUTION

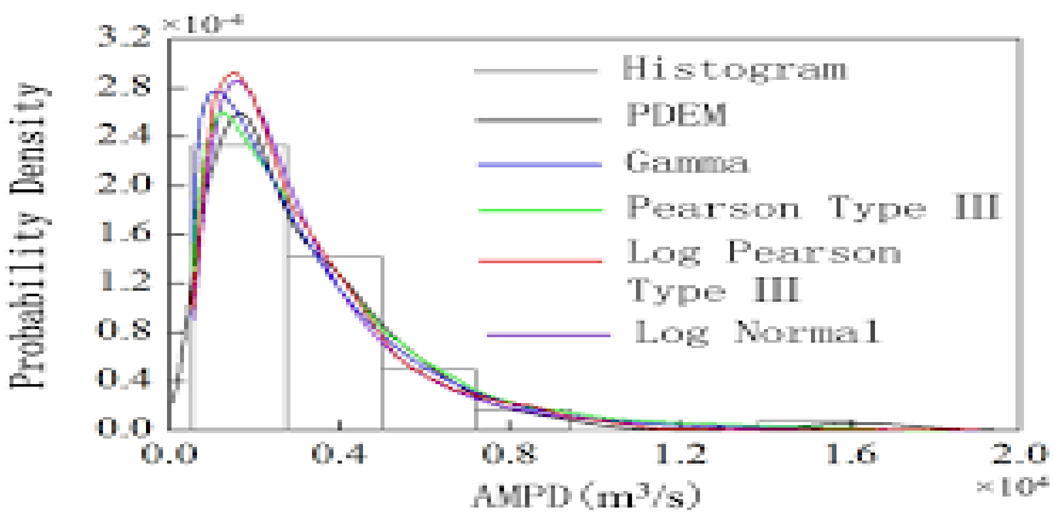
The log-Pearson type (LPT) III distribution is extensively used in hydrologic frequency analysis (2, -4, 12, 15, 16, 18, 24). Its use was recommended by the working group of the Water Resources Council on flow frequency methods, as reported by Benson (1), which concluded that "The log-Pearson III distribution has been selected as the base method, with provisions for departures from the base method where justified the probability density function is defined by

Let y = In x where x is a positive random variable. If y has a LPT III distribution, then x will have an LPT III distribution with the probability density function given by

$$f(x) = \frac{1}{ax\Gamma b}\left(\frac{\ln x - c}{a}\right)^{b-a} e^{-\left(\frac{\ln x - c}{a}\right)}$$

Where $a > 0, b > 0$ and $0 < c < \ln x$ are parameters





TESTING THE CHI-SQUARE GOODNESS OF FIT:

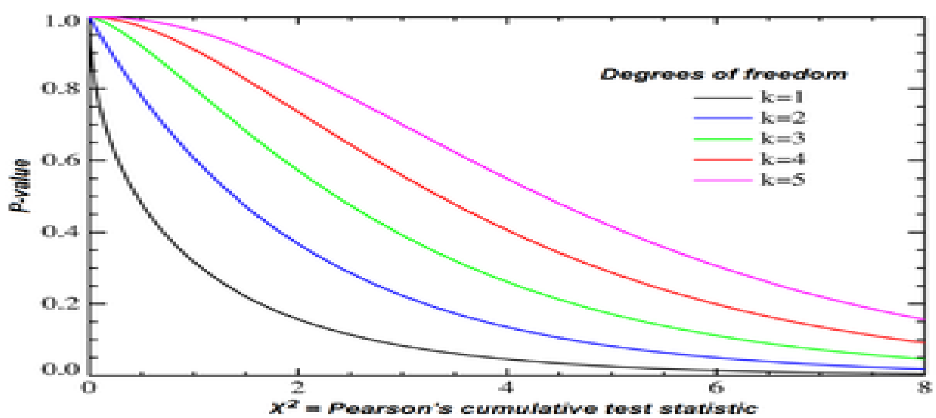
The Root Mean Square Deviation (RMSD) or Root Mean Square Error (RMSE) is a frequently used measure of the differences between values predicted by a model and the values observed. The RMSD represents the sample standard deviation of the differences between predicted values and observed values. The RMSE is the square root of the variance of the differences. It shows the absolute fit of the model to the data, how close the observed data are to the model's predicted values. While R-squared is a relative measure of fit, RMSE is an absolute measure of fit. Being the square root of a variance, RMSE can be interpreted as the standard deviation of the unexplained the variance, and has the helpful property of being in the same units as the response variable. Lower values of RMSE indicate better fit of distribution models. RMSE is a fine measure of how precisely the model predicts the response, and is the best decisive factor for fit if the major objective of the model is prediction.

It is calculated as $RMSE = \sqrt{\sum \frac{(o-e)^2}{e}}$

It is calculated as RMSE = $\sqrt{\sum \frac{(o-e)^2}{e}}$

O: Observed values

E: Expected values



Graphical Representation

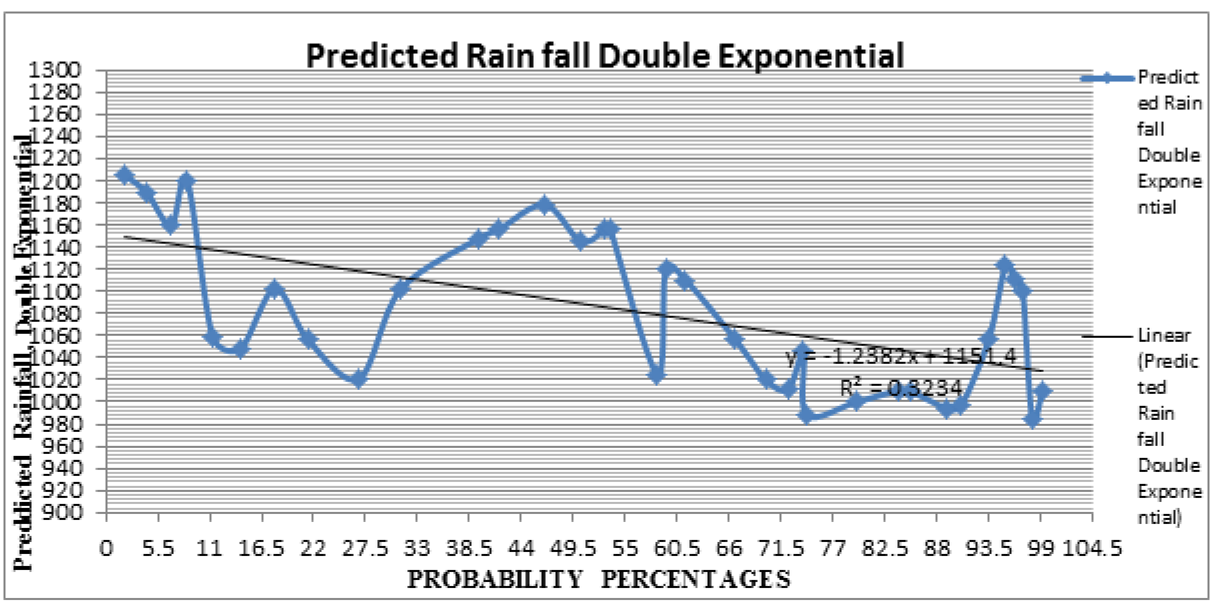


Fig-1 Comparison of observed and predicted rainfall Double Exponential Distribution at Various Levels

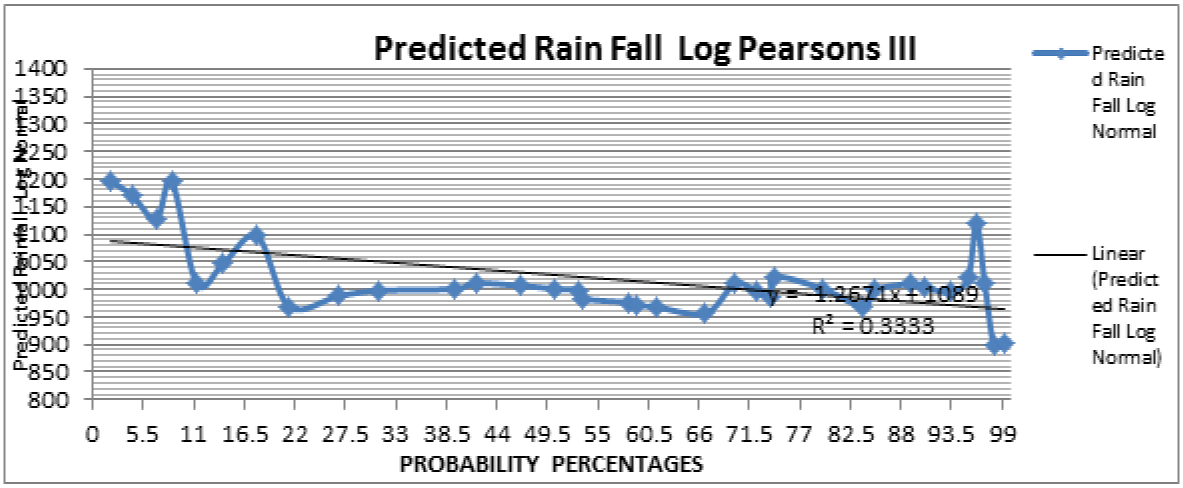


Fig-2 Comparison of observed and predicted rainfall Log Normal at various distribution levels

Table (I): Annual Maximum rainfalls at different return periods in Years (1982-2017)

Probability (%)	*Recurrence Interval*	*Observed Rainfall(O)*	*Predicted Rain fall Double Exponential*	*Predicted Rain Fall* LOG-PEARSON TYPE (LPT) III DISTRIBUTION
1.9	36	1047.2	1205.6	1196.6
4.3	21.6	954.3	1189.3	1170.3
6.9	16.3	1100	1159.3	1126.5
8.6	14.3	1096	1200.3	1196.2
11.3	11.3	963	1059	1011.1
14.2	10.6	1023.2	1047.2	1046.7
17.9	10	1120	1102.1	1096.9
21.4	9.3	1089.3	1056.2	968.7
26.8	8.6	998.3	1021.3	988.2
31.2	5.6	1000.2	1101.5	997.6
39.4	4.6	1123.6	1147.2	1000.7
41.7	4.5	1150.3	1156.3	1011.1
46.5	4.1	1189.2	1178.2	1006.8
50.3	3.9	1089.4	1145.1	1001.1
52.9	3.6	1099.2	1156.1	998.1
53.4	3.5	1023.9	1156.2	983.1
58.4	3.4	1102.5	1025.1	976.1
59.3	3.2	1000.6	1120.3	971.2
61.4	2.9	1158.3	1110.1	968.1
66.7	2.7	1097.2	1056.2	956.1
69.9	2.6	1110.3	1020.1	1010.2
72.3	2.5	1112.8	1011.9	996.1
73.9	2.3	1102.3	1045.7	986.1
74.2	2.1	1000.9	988.2	1023.1
79.5	1.9	985.6	1000.1	1001.1
83.9	1.8	1056.7	1010.1	968.2
85.2	1.8	1100.6	1009.8	999.1
89.1	1.5	1105.9	994.2	1010.4

90.6	1.4	1189.1	997.1	1003.6
93.5	1.4	1147.2	1056.1	995.2
95.3	1.3	1056.2	1123.8	1020.4
96.2	1.3	998.6	1111.1	1120.4
97.1	1.2	987.2	1100.8	1010.1
98.1	1.1	1056.1	985.2	899.1
99.2	1.0	1152.3	1009.1	902.6

Table-2 RMSE value for goodness of fit

Distribution Models	***RMSE Value***
Double Exponential	108.9
LOG-PEARSON TYPE (LPT) III	111.6

RESULTS& DISCUSSION

The predictions of maximum rainfall of 37 years were estimated by two most widely used probability distribution method by Double Exponential and Log-Pearson Type (LPT) III model.

For analyzing of results, the maximum rainfall recorded from last 37 years data were taken down in descending order scale. These we remark has input in log normal distribution model. Other to this recurrence interval, CS and CV values, the frequency factor values were achieved from table. In Double Exponential 35 years of data used were given as input and the annual maximum rainfall were arranged in descending order of magnitude. Recurrence intervals were computed for the Double Exponential as shown in [Table-1].

The predicted annual maximum rainfalls at the different probability levels are tabulated in [Table-1] for Double Exponential Log-Pearson type (LPT) III Distribution. On the basis of basis of calculation the graphs plotted [Fig-1 and Fig-2] showed similarity between observed and predicted values points next to precise, except variations at highest rainfall during both the probability distribution models.

The evaluation of Double Exponential and Log-Pearson type (LPT) III Distribution were conducted using statistical factor RMSD and RMSE for goodness of fit. The minimum value of the RMSE value is taken as the best. The result of annual maximum rainfall is tabulated in [Table-1]. It shows that the value of RMSE for Double Exponential and Log normal distributions comes out to be is 108.9mm and 111.6mm respectively [Table-2]. Since the Double Exponential has the smallest value of RMSE as compared to the Log-Pearson Type (LPT) III. So the Double Exponential gave the best fit for yearly rainfall data. Therefore, it may be concluded that the Double Exponential was found to be the best model for predicting the annual maximum rainfall of India, which reveals the overall accuracy of the model for predicting rainfall. The graphical representation showed that the Double Exponential distribution is predicting the rainfall very near to the observed rainfall [Fig-1].

The observed data were much successfully described using the predicted values, which were taken on basis of recorded data from natural process. Every predicted value are not precisely standard values but proved approximate to principal phenomenon.

Agricultural production can be significantly expanded with proficient application of rainfall. Though the nature of rainfall is erratic and varies with time and space, however it is possible to predict design rainfall quite accurately for certain return periods using various probability distributions functions. Frequency analysis rainfall data has been attempted for different places in India. Frequency analysis of rainfall is an important tool for solving various water management problems and is used to assess the extent of crop failure due to deficiency or excess of rainfall. Probability analysis of annual maximum daily rainfall for different returns periods has been suggested for the design of small and medium hydraulic structure. The rainfall distribution pattern of any are a strongly effects analysis of rainfall data. Establishing probability distribution for knowing daily rainfall activities has always been the matter of research in meteorological field. Various successful rainfall analysis and probability distribution models such as Normal, Log-Pearson Type (LPT) III Distribution, Double Exponential, Weibulls and Pearson type distribution were identified after wide and efficient studies conducted within India and international levels. Influence of rainfall on the yield of wheat and distribution of rainfall during a season rather than the total amount of rainfall which influence the crop yield. Rainfall distribution transformed the skew frequency of rainfall to approximate directly to the theoretical normal distribution.

CONCLUSION

The present study concluded that data of Thirty-Seven (1982-2019) is sufficient to obtain annual maximum rainfall (mm) distribution of India. The selection of probability distribution function to be used for representing the observed data influentially depends on rainfall pattern of the place. As rainfall pattern varies from place to place. The annual maximum rainfall was 1463.9mm in the year 1982 and minimum of 930.1 mm in the year 2019 respectively was observed for analysis. The Root Mean Square Deviation (RMSD) or Root Mean Square Error (RMSE) for goodness of fit was conducted Double Exponential and Log-Pearson Type (LPT) III Distribution. The minimum value of the RMSE value is taken as the best for goodness of fit. The predicted rainfalls are fairly close to the observed rainfall according to the analysis. It shows that the Double Exponential distribution has the least value as compared to the Log-Pearson Type (LPT) III Distribution method according to RMSE. Therefore, prediction by Double Exponential method was found to be the best model for India.

Having precise and standard information of rainfall pattern proves useful for preparing crop calendar, designing of different storage structures and also managing and executing up of irrigation strategies during drought spells. Well knowledge of consecutive days of return periods proves a fundamental parameter of safe, sound and effective economic planning and in designing of different structural and non-structural measures small and medium hydraulic structure such as culverts, bridges, check dams and ponds.

REFERENCES

1). Asati.S.R.(2012)International Journal of Life Sciences Biotechnology and Pharma Research,1

2). Sbudi R.(2007)Indian Journal of Soil Conservation,35,84-85

3). BhaskarS.R.Iqbal M., MukeshDevanda,NeerajChhajed and Banasai A.K.(2008)Journal Of Agricultural Research,42,201-206.

4). Bhati V K ,Tiwari A K and Sharma A.K.(1996) Indian Journal of Soil Conservation,1,25-27.

5). Sharma M.A. and Singh J.B.(2010) New York science journal,3(9),1-10.

6). Manning H.L.(1950)Journal of Agriculture Science,40,169.