International Journal of Advanced Research and Development
ISSN: 2455-4030, Impact Factor: RJIF 5.24
www.advancedjournal.com
Volume 2; Issue 4; July 2017; Page No. 239-243

Determining best fit probability distributions and estimation of annual rainfall of Marathwada region of Maharashtra

Yashwant SS, Sananse SL

Department of Statistics, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India

Abstract
The objective of this paper is to identify best fit probability distribution of eight station of Marathwada region. Based on this fitting probability distribution annual rainfall value of different return period estimate. Marathwada region consists of eight districts. The total area of the region is 64,798 sq. km. The annual rainfall data 32 yea (1978 to 2010) have been collected IMD. Three statistical goodness of fit test such as Kolmogorov–Smirnov test (K-S), Anderson Darling test (A^2) and Chi-Square test (χ^2) was applied to annual rainfall data to select best fit probability distribution among 17 probability distribution. The select 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. On the basis of fitting probability distribution estimate parameter and return period of annual rainfall. The analysis of data indicated that different distribution fitted. Aurangabad district it was Weibull, For Jalna it was Weibull (3P), Latur it was Lognormal (3P), for Nanded it was Gamma Distribution, for Osmanabad it was Weibull and for Parbhani district fitted distribution was Gamma (3P).

Keywords: estimation, probability distributions, goodness of fit, MLE, return period

1. Introduction
Marathwada region consists of eight districts viz. Aurangabad, Beed, Hingoli, Jalna, Latur, Nanded, Osmanabad and Parbhani. The total area of the region is 64,798 sq. km. The entire region is situated at the height of about 300–650 meter. The highest peak at Aurangabad district is Surpal Nath (960 m). Dry deciduous forests, open scrub jungles and vast tracts of grasslands form the components of vegetation. Estimation of rainfall is one of the most important natural input resources to crop and operation of various hydraulic structures such dam,bridges, crop planning. Probability distribution is most important statistical tool for rainfall estimation of magnitude with different return periods. The study of consequences of global climate change on these systems requires scenarios of future precipitation change as input to hydrologic cycle. Hydrological and meteorological data show no random behavior. Then they can be analyzing by some statistical methods based on frequency analyses of precipitation and flood data. Therefore, statistical distributions can be employed for the studies such as the design of water structure, the management of water resource and watershed, and the determination of effective factors about hydrologic cycle. However, it is necessary to determine the best-fitted distribution to studied data. As result. selection of selection of location of best fit model &estimation of desired amount of rainfall for different return period of various probability levels has been attempted by the several researchers over the year. Analysis of rainfall and determination of annual maximum daily rainfall would enhance the management of water resources applications as well as the effective utilization of water resources. (Fisher (1924) [9]. 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) [21] subsequently applied the technique on sunshine distribution and found that sunshine has beneficial effect throughout 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. Upadhaya and Singh (1998) [22]. reported that it is likely to predict precipitation fairly accurate employing different probability distributions for certain return periods although the precipitation varies with respect to space, time and have erratic nature. 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. Phien and Ajirajah (1984) [17] 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) [4].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.

2. Data and Methodology
2.1 Study area and data collection
The present day Marathwada forms a revenue division of Maharashtra Geographically the region is situated between 170 -35 N and 200- 40 N latitude and 740 - 40 E and 780 – 15 E longitudes. The total geographically area of the region is 64525 square Kms. The land of Marathwada of region is flat

with an evaluation ranging between 300 & 900 meters. Marathwada has extensive hilly ranges and sprats but these range neither provide water cultivation nor attract rains and hence are used form economic point of views. The present study is based annual rainfall data of Marathwada Region of eight district of Maharashtra. The annual rainfall data of 32 years years (1978 to 2010) were collected from the IMD and Department of Agriculture, Govt. of Maharashtra. The rainfall was measured in millimeter (mm).

2.2 Selection of the probability distribution.

In this study different type of probability distribution such as Weibull, gamma (3P), generalized extreme value and Pearson 3 were used probability models for evaluating the best fitted probability distribution for rainfall. In addition the different forms of these distributions were also tried and thus total 17 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 for eight station of Marathwada region. The model parameters μ and σ represents mean and standard deviation for normal distribution, scale and shape parameter for lognormal distribution, while α (α1 α2), β and σ indicated the shape, scale and location. location parameters for rest of the distributions (Table 3). In generalized gamma (3P, 4P) and in generalized extreme value, the shape and location parameters were demonstrated by μ and k, respectively. Shape parameter (α) is a measure of skewness of the distribution.

2.3 Fitting of the probability distribution.

The goodness of fit tests measures the compatibility of a random sample with a theoretical probability distribution function. The results are presented in the form of interactive tables that help you decide describe your data in the best way. The goodness of fit test is designed to compare the sample obtained with the type of sample one would expect from the hypothesized distribution and to check whether the hypothesized distribution function fits the data in the sample The goodness of fit test is performed in order to test the following hypotheses.

H0: the precipitation data follow the specified distribution

H1: the precipitation data do not follow the specified distribution.

The goodness-of-fit tests have been conducted were several pervious researchers such as Kolmogorov-Smirnov test, Anderson-Darling test (Husak *et al.* 2007; Olofintoye *et al.* 2009 [16]; Sharma and Singh 2010 [19]; Roman *et al.* 2012; Mandal & Choudhury 2014) [15].

The following goodness of test was used α (0.05) level of significance for the selection of the best fit probability distribution.

i) Kolmogorov-Smirnov Test

This test is used to decide if sample come from a hypothesized continuous distribution. It is based on the empirical cumulative distribution. Assume that we have a random sample x_1, ---x_i from some distribution with cumulative distribution function F(x).

The empirical cumulative distribution is denoted by.

Fn(x) = 1/n [Number of observation ≤ x].

Definition:

The Kolmogorov-Smirnov test statistic (D) is a function of the greatest vertical distance between

Distribution functions, either hypothesized or empirical distribution functions. K-S test calculates the maximum difference between the hypothesized distributions $Z_{(i)} = F(x_{(i)}, \hat{\theta})$ and empirical cumulative distribution function $F_n(x_i)$ with X_i representing the ordered data

$$D^+ = \max_i(i/n - z_i) \qquad D^- = \max_i[-(i-1)/n] \qquad (1)$$

$$D^+ = \max_i(D^+, D^-)$$

Where x_1,,---x_i are daily rainfall values In case of small sample, the Kolmogorov test is preferable over the chi-square test for goodness of fit. This test is used to decide if a sample comes from a hypothesized continuous distribution (Chakravarti *et al.* 1967) [5].

II. Anderson-Darling Test

The Anderson-Darling test is a statistical test of whether a given sample of data is drawn from form a given probability distribution.

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

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\{\ln F(\chi_i) + \ln[1 - F(\chi_{n-i+1})\} \qquad (2)$$

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 (Stephens, 1974, 1976, 1977)

III. Chi-square Test

Pearson's chi-squared test is uses a measure of goodness of fit which is the sum of differences between observed and expected outcome frequencies (that is, counts of observations), each squared and divided by the expectation.

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

Where

O_i = observed frequency

E_i = expected (theoretical) frequency for asserted by the null hypothesis

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

Calculated by $E_i = F(\chi_2) - F(\chi_1)$

F = the CDF of the probability distribution being tested The observed number of observation (k) in interval 'i' is computed from equation given below:

$$k = 1 + \log_2 n$$

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.

2.4 Identification of the best fit probability distribution

Identification of best fit probability distribution was depends on total score obtain from applying three statistical goodness of fit test for 32 year of rainfall data of Marathwada region. The ranking was assigned to different probability distribution based on minimum test statistics value. The identification of all probability distribution was based on the total score obtain by coming of these three test. The ranks of different probability distributions were marked from 1 to 17 based on minimum and arrange rank in ascending order. For a particular distribution, total score was obtained by summing up scores of three goodness of fit test. Probability model with the highest score was considered as the best-fitted distribution for a (Sharma and Singh 2010; Olofintoye *et al.* 2009) [16]. Coefficient of determination (R^2) was also calculated to justify the reliability of best-fitted models

2.5 Estimation of Return period

Return period (T): interval is the average interval of time within which any extreme event of given magnitude will be equaled or exceeded at least once**.** The estimation of probability (x mm) of annual rainfall exceeded on average only once in T years**.** Probability of occurrence (p): is expressed as probability of that an (x mm) of annual rainfall of specified magnitude will be equaled or exceeded during a one year period (T). If n is the total number of values and m is the rank of a value in a list ordered descending magnitude ($x1 > x2 > x3... > x_m$), the exceeding probability of the m^{th} largest value, x_m, is

$$P(X \geq x_m) = \frac{m}{n}$$

See Ramachandra Rao and Hamed, page 6-7). A given return level xT with a return period T may be exceeded once in T years. Therefore.

$$P(X \geq x_T) = \frac{1}{T}$$

If the probability model with CDF, F is assumed then on inverting.

$$F(\chi_T) = P(X \leq \chi_T) = 1 - P(X \geq \chi_T) = 1 - \frac{1}{T}$$

And get the general expression

$$\inf\{\chi_T : F(\chi_T) \geq 1 - \frac{1}{T}\} = F^{-1}(1 - \frac{1}{T})$$

3. Result and Discussion

In this study, the highest annual rainfall Parabhani station was received 1117 mm in 1992 years &; lowest 898.5 mm of Jalana station was during 2004 year (table1.). about 23 % variation observed in Hingoli sation. The occurrence of annual rainfall of Hingoli sation was high value of skewness & kurtosis that showed that constant increasing annual rainfall & low value of skewness and kurtosis of Jalna station was showed that constantly deceasing annual rainfall.

Table 1: Descriptive statistics of the annual rainfall at selected stations

Location	Maximum	Minimum	Mean	SD	CV	Skewness	Kurtosis
Aurangabad	1014	501	716.49	157	21.9	0.68	0.39
Beed	941	518	673	116	17.23	0.53	-0.29
Jalna	898.5	458	650	138	20.46	0.14	-0.41
Latur	1148	549	776.9	169	21.7	0.37	-0.4
Nanded	1011	506	741.1	162	19.52	0.67	0.17
Osmanabad	1070	564	727.3	125	17.15	0.54	-0.37
Parbhani	1171	580	845.1	166	19.63	1.8	2.81
Hingoli	1050	510	724.3	170	23.44	0.84	1.05

The test statistic for Kolmogorov-Smirnov test (D), Anderson - Darling test (A^2), and Chi-Squared test (χ^2) for each data set were calculate 17 probability distribution. The probability distribution having the first rank along with their test statistic is presented in (Table.2)

Table 2: Station wise first ranked probability distribution using different goodness of fit test

Location	First position in ranking of different tests					
	Kolmogorov Smirnov		Anderson Darling		Chi-Square	
	Distribution	Statistics	Distribution	Statistics	Distribution	Statistics
Aurangabad	Gen. Extreme Value	0.08749	Log Pearson (3P)	0.35806	Gamma (3P)	0.14475
Beed	Gen. Extreme Value	0.0257	Weibull(3P)	0.579	Log-Pearson 3	0.0719
Jalna	Weibull(3P)	0.139	Weibull(3P)	0.533	Gamma(3P)	1.219
Latur	Lognormal(3P)	0.155	Pearson (3P)	0.678	Weibull	3.603
Nanded	Gamma	0.132	Lognormal (3P)	0.454	Lognormal	2.206
Osmanabad	Weibull	0.1542	Gamma	0.138	Weibull	1.102
Parbhani	Gamma(3P)	0.0786	Gamma(3P)	0.297	Gamma(3P)	0.514

To assign score separately for each goodness of fit test statistic. The score one was assign to minimum test statistic value.. The summing of total test score were obtained for each data set of all 17probability distribution. This was help to identified of best fit probability distribution. The probability distribution was obtain highest score of sum of individual of three goodness of fit test then that distribution indicated that best fit for particular station. The Aurangabad station was observed that best fit for Weibull distribution (score 40). (table.3)

Table 3: Score wise best fit probability distribution

Station	Distribution of highest score	
	Best fit Distribution	Score
Aurangabad	Weibull,	40
Beed	Gen. Extreme Value	42
Jalna	Weibull (3P)	36
Latur	Lognormal (3P)	46
Nanded	Gamma	38
Osmanabad	Weibul	46
Parbhani	Gamma(3P)	51
Hingoli	Log-Pearson 3	48

Maximum likelihood method was used to estimate the parameter of fitted probability distribution. 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 32 years. The parameter of fitted probability distribution are presented in (table. 4)

Table 4: Parameter of best fitted distribution

Station	Distributions	Parameters
Aurangabad	Gen. Extreme Value	k=-0.20427 σ=179.13 μ=657.05
	Log Pearson 3	α =102.16 β=-0.02546 γ=9.1627
	Gamma (3P)	α=5.2266 β=83.248 γ=294.7
	Weibull	α=4.4381 β=789.01
Beed	Gen. Extreme Value	k=-0.06535 α=133.35 β=604.17
	Weibull(3P)	α=1.4709 β=252.33 γ=443.61
	Log-Pearson 3	α=195.0 β=0.01603 γ=3.3618
Jalna	Weibull (3P)	α=3.4975 β=417.3 γ=253.62
	Gamma(3P)	α=84.124 β=13.087 γ=-472.43
Latur	Gamma(3P)	α=8.2583 β=55.224 γ=276.44
	Log Pearson 3	α=737.9 β=-0.008 γ=12.478
	Weibull	α=5.1936 β=783.52
	Lognormal (3P)	σ=0.2072 μ=6.6035 γ=-21.098
Nanded	Gamma	α=17.741 β=37.522
	Lognormal (3P)	σ=0.39608 μ=5.9185 γ=264.34
	Lognormal	σ=0.22938 μ=6.4743
Osmanabad	Weibull	α=6.0716 β=686.75
	Gamma	α=23.74 β=27.531
Parbhani	Gamma(3P)	α=60.049 β=23.045 γ=-575.49
Hingoli	Log Pearson 3	α=9.658 β=-0.08616 γ=7.3303
	Gumbel Max	σ=133.96 μ=608.7

To estimate the return period of rainfall with the help of best fitted probability distribution was estimate for different return periods and probability (%) desired amount of annual rainfall was expected at Marathwada region. Estimate of expected annual rainfall of > 500 mm Aurangabad station by using Weibull probability distribution was very high (89%) occurs every 1 year and >1000 mm was only 1.7 % of occurs of every 56 years

Table 5: Estimate the Return-Periods of Aurangabad station

Return level(mm)	Probability of Occurrence	Return Period (year)
500	0.896002	1
600	0.75329	2
700	0.5318	3
800	0.28586	4
900	0.09719	10
1000	0.017758	56

To estimate the expected annual rainfall of > 500 mm of Nanded station by using Gamma probability distribution was very high (85%) occurs at every 2 year and >1000 mm was only 2.7 % of occurs at every 36 years.

Table 6: Estimate the Return-Periods of Nanded station

Return level(mm)	Probability of Occurrence	Return Period (year)
500	0.857	2
600	0.6368	3
700	0.384	4
800	0.1902	5
900	0.0785	13
1000	0.0277	36

4. Conclusion

In this paper, we identified best fit of probability distribution of annual rainfall of Marathwada region by using goodness of fit test. The present analysis showed that Aurangabad station was best fit Weibull distribution, for Beed it was Lognormal (3P), Latur it was Lognormal (3P), for Nanded it was Gamma Distribution, for Osmanabad the distribution fitted was Weibull and for Parbhani district the fitted distribution was Gamma (3P). Based on this fitted probability distribution

annual rainfall value of different return period are estimate. Estimated of expected annual rainfall of > 500 mm Aurangabad station by using Weibull probability distribution was very high (89%) occurs at every one year and >1000 mm was only 1.7 % of occurs of every 56 years This result is also helpful to prediction and estimation of rainfall for farmers and agriculture department of the Government of Maharashtra for planning cropping pattern of the Marathwada region.

5. References

1. Agarwal MC, Katiyar VS, Ram Babu. Probability analysis of annual maximum daily rainfall of U.P. Himalaya. Indian J. Soil. Cons, 1998; 16(1):35-43.
2. Aksoy H. Use of gamma distribution in hydrological analysis. Turk. J. Eng. Environ. Sci. 2000; 24:419-428.
3. Biswas BC, Khambeta NK. Distribution of short period rainfall over dry farming tract of Maharashtra. J. Mah. Agric. Uni, 1989; 12:157-168.
4. Chakravarti, Laha, Roy. Handbook of Methods of Applied Statistics. John Wiley and Sons. 1967; 1:392-394.
5. Chowdhury JU, Stedinger JR, Lu LH. Goodness-of-fit tests for regional Generalized Extreme Value flood distributions. Wat. Resour. Res. 1991; 27(7):1765-1776.
6. Dinpashoh Y. Selection of variables 5 for the purpose of regionalization of Iran's precipitation climate using multivariate methods. J. Hydrology, 2004; 297:109-123,
7. Fisher RA. The influence of the rainfall on the yield of wheat at Rothamsted. Philosophical transaction of the Royal Society of London. 1924; 213.
8. Gingras D, Adamowski K. Coupling of nonparametric frequency and L-moment analysis for mixed distribution identification. Water Resour. Bull. 1992; 28(2):263-272.
9. Haghighatjou P. Probability distribution functions as applied to monthly and annual precipitation of old station in Iran. J. Agro.Sci. Nature. 2002; 9(3).
10. Keshtkar AR. Assessment and fitting the best statistical distribution to, mean max and min discharge data (central plateau). MSc. Thesis, University of Tehran. 2001.
11. Lee C. Application of rainfall frequency analysis on studying rainfall distribution characteristics of Chia-Nan plain area in Southern Taiwan. J. Crop., Environ. Bioinf. 2005; 2:31-38.
12. Mandal S, Choudhury BU, Mondal M, Bej S. Trend analysis of weather variables in Sagar Island, West Bengal, India: a long-term perspective, 1982-2010-2013; 105(7):947-953.
13. Olofintoye OO, Sule BF, Salami AW. Best-fit Probability distribution model for peak of daily rainfall of selected city in Nigeria. New Yark Sci J. 2009; 2(3):1-11.
14. Phien HN, Ajirajah TJ. Applications of the log-Pearson Type-3 distributions in hydrology. Journal of hydrology. 1984; 73:359-372.
15. Ramachandra Rao A, Hamed KH. Flood Frequency Analysis, CRC Press, Boca Raton, Florida, USA, 2000.
16. Sharma MA, Singh JB. Use of Probability Distribution in Rainfall Analysis. New York Science Journal. 2010; 40-49.
17. Mayooran T, Laheetharam A. The Statistical Distribution of Annual Maximum Rainfall in Colombo District. Sri Lankan Journal of Applied Statistics. 2014; (15-2).
18. Tippet LHC. On the effect of sunshine on wheat yield at Rothamsted Jour. Agril. Sci. 1929; 60(2).
19. Upadhaya A, Singh SR. Estimation of consecutive day's maximum rainfall by various methods and their comparison. Indian Journal of S. Cons. 1998; 26(2):193-201.
20. www.imd.gov.in
21. www.mahaagri.gov.in